STUDIES ON SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS TYPE-2 IN NORTHERN SOUTH AFRICA

BY

LISA ARRAH MBANG TAMBE (11636268)

A THESIS SUBMITTED IN FULLFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DOCTOR OF PHILOSOPHY DEGREE IN MICROBIOLOGY

TO

THE DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY, FACULTY OF SCIENCE, ENGINEERING AND AGRICULTURE

UNIVERSITY OF VENDA,

PRIVATE BAG X5050

THOHOYANDOU, 0950

SOUTH AFRICA

SUPERVISOR: DR LUFUNO GRACE MAVHANDU-RAMARUMO (SEFAKO MAKGATHO HEALTH SCIENCES UNIVERSITY)

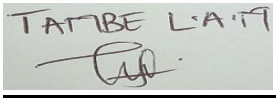
CO- SUPERVISOR: DR DENIS MANGA TEBIT (GLOBAL BIOMED SCIENTIFIC, LLC, VIRGINIA, USA)

CO- SUPERVISOR: PROF. PASCAL OBONG BESSONG (UNIVERSITY OF VENDA)

2025

DECLARATION

I, Tambe Lisa Arrah Mbang, hereby declare that this thesis for the award of a Doctor of Philosophy degree in Microbiology at the University of Venda has not been submitted before for any degree examination at this or any other university. It is my own work and design and execution, and all the reference materials contained therein have been duly acknowledged.

Signature: 

Date: **20 April 2025**

ACKNOWLEDGEMENTS

If it had not been for the Lord who was on my side, I never would have made it. Thank you, Lord!

Additionally, I am grateful for the contribution of the following individuals:

Promoters

My sincere gratitude to my promoters; Dr Lufuno Grace Mavhandu-Ramarumo, Prof Pascal Obong Bessong and Dr Denis Manga Tebit for their mentorship, and constructive criticisms which prompted me to go an extra mile to achieve great outcomes. Without all these, this work would not have been of great value. I am blessed to have been under your tutelage. Thank you so much, and God bless you.

Collaborators

The work would not have been possible without the ceaseless support and commitment of the following:

1) The University of Venda sampling team: Dr. Oisaemi Izevhbekai, Mr. Stephen Muthivhi and Dr. Glynn Pindihama. I am also thankful to Prof. Joshua Edokpayi for all the assistance in better understanding the wastewater terrain.

2) Members of the Biomedical Research and Innovation Platform (BRIP) and Genomics center, at the South African Medical Research Council (SAMRC). Thank you for your endless assistance with protocols for SARS-CoV-2 detection in wastewater and sequencing for genetic characterization. Thank you so much.

Financial support

I am grateful to the Solidarity Fund, the South African Medical Research Council (SAMRC) and the Research and Publication Committee of the University of Venda, South Africa for providing funds to carry out this research project. My sincere gratitude also goes to my parents and siblings for all their financial support.

Colleagues

The entire HIV/AIDS & Global Health Research Programme at the University of Venda for all your support during my training and guidance. I am grateful to Dr. Fri Justine, Dr. Mukhethwa Munzhedzi-Masikhwa and Mr. Phindulo Mathobo who helped me consistently to the point of completion of this thesis.

Family and friends

I would also like to extend my sincere gratitude to my parents (Mr. and Mrs. Tambe) and siblings (Kelly and Edward), for their constant encouragement and support as I embarked on this journey. Words of gratitude are not enough for consistently uplifting me in your prayers; may God continue to bless you abundantly. This PhD is for you.

To Bessong and Diko families, thank you for always being there for me, and all the words of wisdom you shared. God's richest blessings to you all.

The entire Ebobanyah family, thank you for all your calls and love I felt even from a distance, I am forever grateful; may God bless you all abundantly.

Spiritual leaders

I am sincerely grateful to the entire Christ Tabernacle Church, particularly the Pastorate, for the spiritual mentorship, prayers and words of encouragement.

DEDICATION

To my parents, Mr. Tambe Richard and Mrs. Tambe Evelyne

&

My siblings, Tambe Kelly and Tambe Edward

SUMMARY OF THE THESIS

Background: The last three decades have been characterized by the re-emergence of the Coronaviridae family into the human population, causing severe respiratory disease with increased morbidity rates. A dearth of information exists on human coronavirus (HCoV) molecular epidemiology, and circulation in different populations in Africa. As the COVID-19 pandemic progressed across the globe, wastewater-based epidemiology (WBE) was proposed as an alternative tool for assessing and monitoring the occurrence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) at the community level. Additionally, through wastewater-based genomic surveillance of SARS-CoV-2, the evolutionary patterns and distribution of viral types at the population level can be comprehensively characterized. This study systematically reviewed literature published prior to the SARS-CoV-2 outbreak to investigate the prevalence and molecular epidemiology of HCoVs circulating in Africa. Secondly, this study established a wastewater-based surveillance (WBS) system to track the trends of SARS-CoV-2, investigate SARS-CoV-2 variants of concern (VOC) circulating in the population, and determine the prevalence of people infected in the Vhembe and Mopani districts. Thirdly, through WBS, to describe the molecular epidemiology of SARS-CoV-2 and document the respiratory viruses occurring in the Vhembe and Mopani districts.

Methodology: A systematic literature review was conducted according to the PRISMA guidelines, to understand the prevalence and molecular epidemiology of HCoVs in Africa. For the second and third objectives, wastewater influents from seven wastewater treatment plants (WWTPs) and one waste sedimentation pond (WSP) were collected weekly from January 2021 to June 2022. Out of a total of 487 samples collected, about 75% (365/487) were positive for SARS-CoV-2 RNA by qRT-PCR. Of these, 80 met the criteria for allele-specific genotyping (ASG). Positive SARS-CoV-2 RNA detected throughout the surveillance period were compared to 7-day moving average (7D-MA) of clinical cases reported per sub-district. Next, SARS-CoV-2 RNA detected during the surveillance period was normalized using the flow rate and population size. The Spearman's correlation coefficient was used to determine the relationship between non-normalized, and normalized SARS-CoV-2 RNA data when compared to the reported clinical cases. Finally, positive SARS-CoV-2 RNA copies with a standard deviation of less than one (SD <1) were used to predict the prevalence of people infected in the study sites using the Monte Carlo simulation model. This predicted prevalence

was also compared to the 7D-MA clinical cases reported per sub-district, and the correlation between them was determined. Subsequently, samples positive for SARS-CoV-2 were subjected to whole genome sequences (WGS) using the ATOPlex next-generation sequencing method and analyzed for lineage and clade assignment using the Pangolin and Nextclade tool. Relatedness of identified sequences was determined by phylogenetic analysis. VOC was analyzed for prevalence and geographical distribution. Concordance for VOC between ASG and WGS analyses was determined.

Results: Findings from the systematic literature review showed that thirteen out of 54 (24%) African countries had published data on HCoV prevalence and/or genomic epidemiology, from hospitals, clinics, homes, community gatherings, farms, and individuals at airports. The first published data on HCoV was from South Africa in 2008. There was heterogeneity in the type of tests used in determining HCoV prevalence. Two studies reported that risk factors for HCoV include exposure to infected animals or humans. The second objective, establishing a wastewater-based surveillance system was achieved. Briefly, SARS-CoV-2 viral load was detected in wastewater one week prior to increased infection cases reported at the district level during the third and fourth waves, thus serving as an early warning system. Of interest, towards the end of the surveillance period, increased SARS-CoV-2 viral load detected in wastewater were not reflected in the reported clinical cases. Comparing the reported number of cases per district to the predicted prevalence revealed more cases in the Vhembe District than in the Mopani District. Third, SARS-CoV-2 molecular epidemiology and the distribution of respiratory virus were described. A total of 60 SARS-CoV-2 full genomes were analyzed. Delta and Omicron variants were detected as early as January and February 2021, respectively, while the Beta variant was detected in July 2021. Delta variant was significantly predominant at a prevalence of 45%, followed by Omicron (32%), and Beta (5%). Eighteen percent (11/60) of the sequences were assigned a lineage by Pangolin tool, but not a specific WHO variant name. Upon phylogenetic analysis, some of these sequences were seen clustering with the Alpha (2/11) and Delta (2/11) variants, while the remaining sequences clustered with each other. Mutations in the receptor-binding domain (RBD) of the Spike protein (S-protein) were investigated, with some peculiarities observed such as mutation E484K absent in all Beta variant study sequences. Three previously undescribed mutations (A631S, V327I, D427Y) were detected in Delta variant sequences. Concordance in variant

assignments between allele-specific genotyping and WGS was seen in 51.2% of the study sequences. Respiratory virus surveillance revealed year-round circulation of human Adenoviruses (HAdVs), while HCoVs, influenza viruses and human parainfluenza viruses (HPIVs) were mostly detected in winter. Influenza A and B viruses (IAV and IBV) detected in the study site in 2021 were remarkably different from those reported in circulation nationwide by the NICD. Specifically, IAV (H5N1)/Guandong, a highly pathogenic influenza virus, was detected, although at a low frequency.

Discussion and Conclusion: The systematic review revealed that despite the outbreaks of SARS in Southeast Asia in 2002 and MERS in 2012 in the Middle East, the quantum of virologic investigations on HCoV on the African continent was scanty. Pandemic preparedness requires cognizance of disease outbreaks in other continents, establishment of test and surveillance protocols, and infrastructure for eventualities. Regarding the establishment of a wastewater-based surveillance system for SARS-CoV-2 monitoring, this study demonstrates effective surveillance over an extended period in rural settings. This is important because most reports about the application of WBE for monitoring SARS-CoV-2 and circulating variants are predominantly from more urbanized regions in South Africa and other parts of the world. Thus, it reveals applicability of monitoring pathogens in rural areas, despite challenges encountered such as poor or non-existent sewerage systems. Such challenges are common in the African continent, highlighting the need for more of such investigations to strengthen pandemic preparedness measures. The presence of Delta and Omicron VOCs observed prior to other reports in South Africa highlights the importance of population-based approaches in genomic surveillance over approaches that rely on individual samples. Again, it also emphasizes the need for pandemic preparedness efforts to be extended to all geographic regions. Wastewater is known to potentially capture more viral diversity, including SARS-CoV-2 genetic diversity, and could reveal new viruses and VOCs in circulation before they emerge in the wider human population. Thus, continuous surveillance is necessary for documentation of cryptic lineages, which may contribute towards improving vaccine.

Keywords: HCoV prevalence and distribution, Africa, pandemic preparedness, COVID-19, SARS-CoV-2, WBE, early warning, genomic surveillance, VOCs, molecular epidemiology, receptor binding domain, viral evolution, respiratory viruses

SCIENTIFIC OUTPUTS

The following communications have emanated from this study:

Published Articles

Lisa Arrah Mbang Tambe, Phindulo Mathobo, Mukhethwa Munzhedzi, Pascal Obong Bessong, and Lufuno Grace Mavhandu-Ramarumo. Prevalence and Molecular Epidemiology of Human Coronaviruses in Africa prior to SARS-CoV-2 Outbreak: a systematic review Viruses 2023, 15(11), 2146 https://doi.org/10.3390/v15112146

Lisa Arrah Mbang Tambe, Phindulo Mathobo, Nontokozo Daphney Matume, Mukhethwa Munzhedzi, Joshua Nosa Edokpayi, Amsha Viraragavan, Brigitte Glanzmann, Denis M. Tebit, Lufuno Grace Mavhandu-Ramarumo, Renee Street, Rabia Johnson, Craig Kinnear, and Pascal Obong Bessong. Molecular Epidemiology of SARS-CoV-2 in Northern South Africa: Wastewater Surveillance from January 2021 to May 2022. Frontiers in Public Health 11:1309869. https://doi.org/10.3389/fpubh.2023.1309869

Conference Presentations

Lisa Arrah Mbang Tambe, Harry Ngoveni, Phindulo Mathobo, Mukhethwa Munzhedzi, Nontokozo Daphney Matume, Oisaemi Izevbekhai, Joshua Edokpayi, Lufuno Grace Mavhandu-Ramarumo, Rabia Johnson, Renee Street and Pascal Obong Bessong. Wastewater-based Epidemiology as an early warning system for SARS-CoV-2 surveillance in Limpopo Province, South Africa. International Conference on AIDS and Sexually Transmitted Infections in Africa – ICASA 2021, Durban, South Africa. Poster Presentation PEC030.

Lisa Arrah Mbang Tambe, Phindulo Mathobo, Nontokozo D. Matume, Mukhethwa Munzhedzi, Joshua Nosa Edokpayi, Amsha Viraragavan, Brigitte Glanzmann, Denis M. Tebit, Lufuno G. Mavhandu-Ramarumo, Renée Street, Rabia Johnson, Craig Kinnear and Pascal O. Bessong. Genomic Epidemiology of SARS-CoV-2 and Other Respiratory Viruses Detected in Wastewater: 18-Months Surveillance in Limpopo, South Africa. South African Society of Biochemistry and Molecular Biology – SASBMB Congress, Polokwane, South Africa. Oral Presentation, Infectious and Parasitic Diseases, Parallel session 5.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

7D MA	7-day moving average
ACE2	Angiotensin converting enzyme 2
ARI	Acute respiratory infections
ASG	Allele-specific Genotyping
bp	Base pair
CDC	Centre for Disease Control
CI	Confidence interval
CoVs	Coronaviruses
COVID-19	Coronavirus disease 2019
CSG	Coronavirus study group
Ct	Threshold
C_{virus}	Viral concentration
DMV	Double membrane vesicles
E-Protein	Envelope protein
Epi-Week	Epidemiological week
Eq.	Equation
EWS	Early warning system
g.c./day	Genome copies per day
g.c/mL	Genome copies per millilitre
g.c./µl	Genome copies per microlitre
g.c/L	Genome copies per litre
HCoV	Human coronavirus

HE	Hemagglutinase-estarase
ICTV	International Committee on Taxonomy of Viruses
inh.	Inhabitants
J&J	Janssen & Janssen
LMIC	Low- and middle-income countries
MC	Monte Carlo simulation model
M-Protein	Membrane protein
MERS-CoV	Middle eastern respiratory syndrome coronavirus
mL	Millilitre
mL/day	Millilitre per day
N-Protein	Nucleocapsid protein
NGS-SA	Network for Genomic Surveillance in South Africa
NICD	National Institute for Communicable Diseases
nsps	Non-structural proteins
L	Litre
LMICs	Low- and middle-income countries
ORF	Open reading frame
p	*p-value*
PE	Population estimates
PHEIC	Public health emergency of international concern
Pop.	Population
Pp	Polyprotein
Q	Flowrate

r	Spearman's correlation coefficient
R0	Basic reproduction rate
RdRp	RNA-dependent RNA polymerase
RTC	Replication-transcription complex
RVOP	Respiratory Virus Oligo Panel
S-Protein	Spike protein
SARI	Severe acute respiratory illnesses
SAMRC	South African Medical Research Council
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus type 2
SD	Standard deviation
SNP	Single nucleotide polymorphism
sgRNA	Sub-genomic RNA
ssRNA	Single stranded RNA
TMPRSS2	Type II transmembrane serine protease
VOC	Variant of concern
VOI	Variant of interest
VUM	Variant under monitoring
WBE	Wastewater-based epidemiology
WBS	Wastewater-based surveillance
WWS	Wastewater surveillance
WGS	Whole genome sequencing
WHO	World Health Organization

WSP	Waste sedimentation pond
WW	Wastewater
WWTP	Wastewater treatment plant

CHAPTER ONE

General Introduction and Literature Review

1.1) INTRODUCTION

Coronaviruses (CoVs) are positive-sense, single-stranded RNA (ssRNA) viruses which are approximately 32 kb in size, making them the RNA viruses with the largest genome size. They are widely distributed in animals and humans, and result in mild or severe infections. The first CoV to be discovered was the avian infectious bronchitis virus (IBV) in the early 1930s which infected birds. Subsequently, several animal CoVs were discovered. The first human coronavirus (HCoV-229E) was isolated from a medical student in Chicago, USA in 1964 (Mulabbi et al., 2021). After the first discovery of CoVs, seven other HCoVs (HCoV-229E; HCoV-OC43; severe acute respiratory syndrome coronavirus, SARS-CoV; HCoV-NL63; HCoV-HKU1; middle eastern respiratory syndrome coronavirus, MERS-CoV; severe acute respiratory syndrome coronavirus type 2, SARS-CoV-2) have been discovered. Four of these HCoVs (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) circulate throughout the year, causing mild seasonal colds in immunocompetent individuals, but leading to severe respiratory infections in infants, the elderly, and immunocompromised individuals. The other three HCoVs (SARS-CoV, MERS-CoV and SARS-CoV-2), are more pathogenic in nature and were zoonotically introduced to human population leading to severe acute respiratory infections with a high mortality rate (Krishnamoorthy et al., 2020; Poland, 2020; Tang et al., 2022).

The occurrence of the SARS and MERS epidemics in 2002 and 2012, respectively, led to an upscale in studies investigating endemic HCoVs (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) and zoonotic HCoVs (SARS-CoV and MERS-CoV) in the population. The goal of this was to improve understanding of these pathogens to improve transmission prevention and disease management strategies (Shao et al., 2022). A vast amount of information exists on HCoV molecular epidemiology, genetics and evolutionary characteristics of their occurrence in different regions (Woo et al., 2005; Lau et al., 2006; Van Der Hoek et al., 2006; Suwannakarn et al., 2014). This information, though localized to endemic areas of disease, has aided viral surveillance and rapid identification of new HCoV types introduced into the human population. In Africa, however, there is a dearth of information describing the types of HCoVs circulating in the population. Of the few existing studies, most of them only describe the prevalence of HCoV in children (≤15

years) in association with studies of other respiratory viruses that cause severe acute respiratory infections (SARIs). The continuous resurgence of HCoVs in the human population emphasizes the necessity to document not only the prevalence of HCoVs, but also the molecular epidemiology. Knowledge about the frequency, genetic patterns and evolutionary characteristics of HCoVs can help in predicting and managing HCoV infections in different populations and increase pandemic preparedness efforts.

The ability of CoVs to jump species and cause human infection has led to the development of three pathogenic HCoV infections in humans of zoonotic origin. The most recent of these zoonotic pathogenic CoV infections led to the current 2019 coronavirus disease (COVID-19) caused by SARS-CoV-2 (Platto et al., 2021). The main mode of SARS-CoV-2 transmission is through respiratory droplets. Once the virus attaches to the cellular human receptor, there is an incubation period of approximately 2 – 7 days before the onset of any clinical symptoms such as fever, cough, shortness of breath, fatigue, muscle or body aches, headache, new loss of taste or smell, sore throat, congestion or runny nose, nausea or vomiting, and diarrhea. Infection severity varies from asymptomatic individuals to those with comorbidities or patients advanced in age (CDC, 2021).

At the beginning of the COVID-19 pandemic, diagnosis of infected individuals relied on clinical testing of individuals who present with symptoms by using reverse transcription quantitative polymerase chain reaction (RT-qPCR) (Chau et al., 2020). While this method is the gold-standard of testing, one proposition at the start of the pandemic was deployment of testing to every community (Hart and Halden, 2020) to identify infected individuals and prevent transmission. The shortcoming of this method was the availability of resources and personnel, particularly in resource-limited settings. Another drawback was the possibility of super-spreader events (Wei et al., 2020), where simultaneous exposure meant individuals likely present with symptoms about the same time and require medical attention all at once. This can present a major challenge for available resources and the limited healthcare personnel, especially in times of rapidly increasing infection rates. Secondly, asymptomatic individuals are largely unaccounted for since most of them will not be tested in the absence of any symptom. To circumvent this challenge to improve

rapid detection and prevent transmission, monitoring the presence of SARS-CoV-2 in wastewater was proposed since the virus is shed in human feces approximately 3 – 5 days post exposure (Jones et al., 2020). Using wastewater to monitor pathogen occurrences in the environment is known as wastewater-based epidemiology (WBE). This non-invasive technique was initially used to for surveillance of virus outbreaks in the 1950s and 1960, to trace the poliomyelitis epidemics (Wiley et al., 1962). It also proved useful in 2013 in Rahat, Israel, when regular surveillance identified a silent outbreak of polio in the population, which was rapidly intercepted (Brouwer et al., 2018), showing the efficacy of this tool as an early warning system (EWS).

Thus, several studies utilized WBE to investigate the occurrence of SARS-CoV-2 in the community; some of them reported the detection of COVID-19 in wastewater, even before the first confirmed case was reported (Ahmed et al., 2020; Medema et al., 2020b; Westhaus et al., 2021). Subsequently, several other reports around the globe proved the applicability of this method to monitor SARS-CoV-2, which served as an early warning system for identification of potential COVID-19 hotspots (Randazzo et al., 2020; Trottier et al., 2020; Baldovin et al., 2021; Hillary et al., 2021). This tool was thus adapted as an alternative tool for tracking SARS-CoV-2 occurrences in communities, since it accounted for the symptomatic, pre-symptomatic and asymptomatic populations. Simultaneously, SARS-CoV-2 concentrations in wastewater were utilized to estimate the prevalence of those infected within a catchment area (Saththasivam et al., 2021). Such data is relevant because estimates of SARS-CoV-2 infection are essential to understand COVID-19 disease burden and the impact of public health interventions (Morvan et al., 2022). This showed the necessity for establishment of a wastewater surveillance system, not just to identify potential COVID-19 hotspots, but to estimate the number of people infected. This can assist with implementation of public health measures to prevent continuous transmission.

As the pandemic progressed, SARS-CoV-2 constantly evolved, and this is chiefly attributed to the mutation rate of the virus, and its recombination mechanism (Markov et al., 2023). Some of these mutations have led to the spread of variants in the population that were highly transmissible, pathogenic, and often evaded available therapies (Ahmad

et al., 2022). Monitoring of the changes in this virus was predominantly done through high throughput sequencing. Through this genomic epidemiology, five high-risk SARS-CoV-2 variants, and their sublineages were identified and denoted as variants of concern (VOCs) by the World Health Organization (WHO). These variants include the Alpha, Beta, Gamma, Delta, and Omicron VOCs (Walker et al., 2021; Tegally et al., 2021; Nouvelli et al., 2021; Callaway 2021). They were categorized as such due to their increased transmissibility and severity in the population. Thus, monitoring the evolution patterns of the virus is imperative to predict future changes as well as develop efficient vaccines which will cover a wide variety of these continually developing variants. Detection of these variants was mostly achieved through genomic surveillance studies in clinical samples for description of viral lineages occurring in the population (Alsafar et al., 2022; Santiago et al., 2022; Chrysostomou et al., 2023). While this is relevant, it is biased in that it relies only on symptomatic individuals and may not accurately represent population epidemiology. Secondly, such surveillance may not be applicable in resource-limited areas, thus important transmission and evolutionary dynamics may be missed. The advantage of using wastewater for genomic epidemiology is its cost-effectiveness, and total representation of viral dynamics at the population level.

Thus, this study set out to determine the distribution and molecular epidemiology of HCoVs in Africa prior to the SARS-CoV-2 outbreak; to monitor SARS-CoV-2 trends and VOCs circulation; and describe the molecular epidemiology of SARS-CoV-2 in Vhembe and Mopani districts, in the Limpopo Province of South Africa.

1.2) LITERATURE REVIEW

1.2.1) History of Human Coronavirus Discovery

Coronaviruses (CoVs) are positive-sense, single-stranded RNA (ssRNA) viruses with a linear, non-segmented viral genome which is approximately 32 kb in size and packed in a helical nucleocapsid. They are named CoVs because of the crown-like appearance of the surface projections. Based on classification report of the International Committee on Taxonomy of Viruses (ICTV), CoVs belong to the order Nidovirales, suborder Cornidovirineae, family Coronaviridae and subfamily Orthocoronavirinae. The subfamily Orthocoronavirinae is divided into four genera: Alpha- (α), Beta- (β), Gamma- (γ), and Delta- (δ) CoVs. Both α and β coronaviruses exclusively infect mammalian species; their ability to cross-infect humans lead to the emergence of major human pathogens. The γ and δ CoVs on the other hand, have a wider host range including avian species (Woo et al., 1999; Forni et al., 2017). To date, there are seven (7) human CoVs (hCoV) that are known and subdivided into two genera – α and β. Severe acute respiratory syndrome (SARS) CoV – SARS-CoV, Middle Eastern respiratory syndrome (MERS) CoV – MERS-CoV, hCoV-OC43, hCoV-HKU1 and Severe Acute Respiratory Syndrome (SARS) CoV 2 – SARS-CoV-2, belong to the β genera, while hCoVs hCoV-NL63 and hCoV-299E are of the α genera. **Figure 1** shows HCoV classification scheme according to the ICTV. HCoVs were once considered minor human pathogens causing common cold or mild respiratory infections in immunocompetent individuals. CoVs infect mainly mammalian and avian species and cause respiratory and enteric infections with varying severity. The first coronavirus (CoV) was discovered in 1931 and termed the avian infectious bronchitis virus (IBV) (V'kovski et al., 2020). Subsequently, other animal CoVs such as the Murine Hepatitis Virus (MHV), as well as pathogenic CoVs like the transmissible gastroenteritis virus (TGEV), bovine coronavirus (BCV), and the feline infectious peritonitis virus (FIPV) were discovered (Mulabbi et al., 2021).

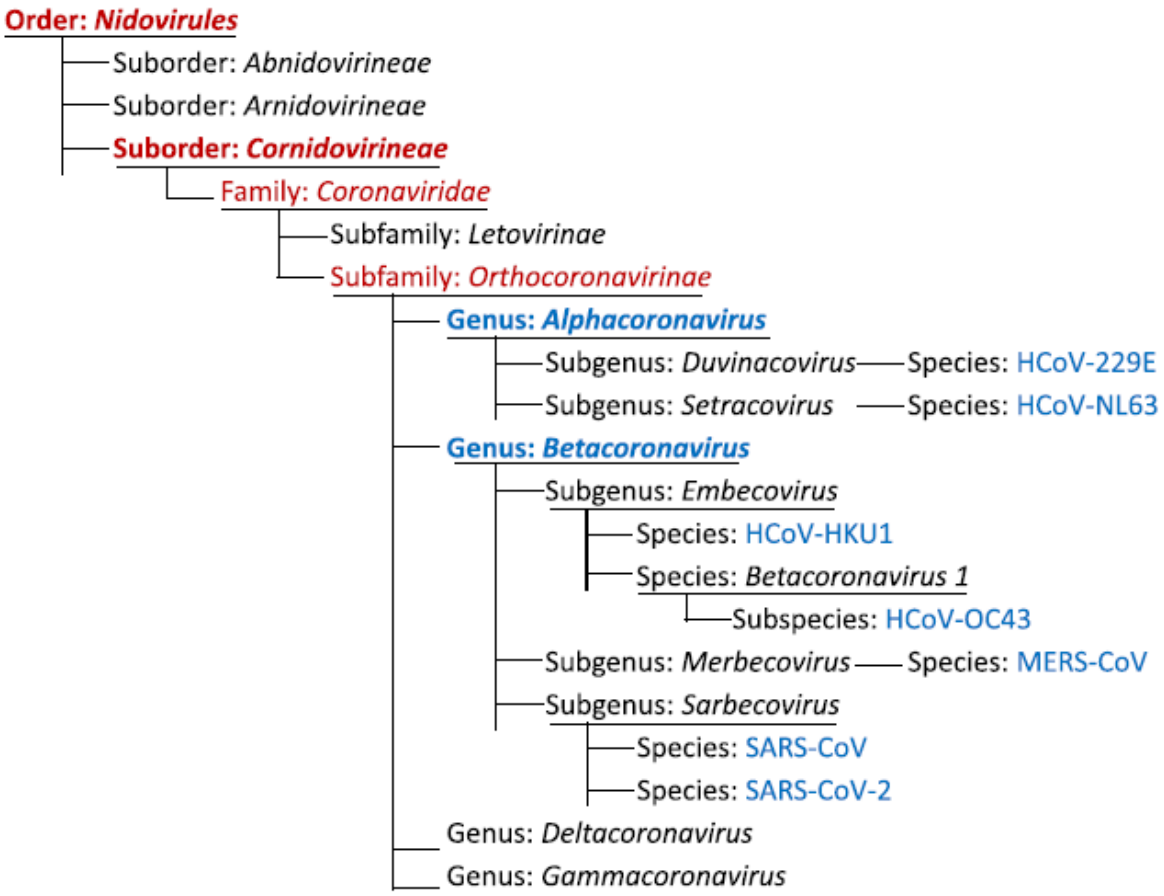


Figure 1: An updated classification scheme per the 10th ICTV report. Adapted from (Liu et al., 2021)

The first HCoV to be isolated was strain B814 which was isolated in 1965 from a patient with a common cold. Since then, more than 30 additional strains were identified, one of them being HCoV-229E isolated from a student specimen coded 229E during a study of respiratory illness amongst medical students in the University of Chicago. HCoV-OC43 (Organ Culture 43) was discovered next using tracheal organ culture and found to be serologically distinct from HCoV-229E (Liu, Liang and Fung, 2021). Up to the 1990s, these two strains were the only HCoV strains studied since they could be cultured. Since HCoV-229E and HCoV-OC43 caused common cold, they were considered relatively harmless. Other HCoV strains such as HCoV-OC16, HCoV-OC37, HCoV-OC38, HCoV-OC44, HCoV-OC48 and strain B814 were all lost, since culturing them was impossible. In 2002, the first HCoV that caused severe acute respiratory illness in humans was discovered and named severe acute respiratory syndrome coronavirus (SARS-CoV)

because it caused severe respiratory infection with a high mortality rate (Wang et al., 2006). SARS-CoV which emerged in the Guangdong Province in South China, was the first pathogenic HCoV, affecting mostly infants, the elderly population, and those with underlying conditions. After SARS-CoV disappeared in 2003, HCoV-NL63 was identified and isolated in the Netherlands from a child suffering from bronchiolitis and conjunctivitis. HCoV-HKU1 was discovered next in 2004 in Hong Kong and isolated from a 71-year-old man who had just returned from Shenzhen, China.

A decade later in 2012, a second pathogenic HCoV strain emerged in Jeddah, Saudi Arabia. This was initially named Human Coronavirus Erasmus Medical Center (HCoV-EMC) but later the international committee on the taxonomy of viruses renamed it 'Middle East Respiratory Syndrome (MERS) CoV'. MERS-CoV causes lower respiratory tract infections, which like SARS-CoV leads to severe acute respiratory failure. It was isolated from a 60-year-old man who suffered from acute pneumonia and later developed respiratory distress, renal failure and eventually died (Zaki et al., 2012). It was later established that the source of this infection was dromedary camels from a farm in Qatar where the first two human cases of MERS-CoV originated (Haagmans et al., 2014). This epidemic although curbed, has spread to countries outside the Middle Eastern regions, with the latest infection reported in March 2020 by Riyadh (WHO, 2021).

In late December 2019, several pneumonia cases of unknown cause were reported by local health facilities in Wuhan, China. Investigation of the cause of these pneumonia cases was linked to a seafood and wet animal local market. After thorough investigations by the Chinese Center for Disease Control (CDC) to identify the source of the pneumonia, they reported the identification of a novel CoV type which was initially named 2019-nCoV (2019 novel Coronavirus) since it formed a clade within the sub-genus sarbecovirus, of the Orthocoronavirinae sub-family (Zhu et al., 2020). This is the third pathogenic CoV of zoonotic origin which affects the upper respiratory tract (URT) and causes acute respiratory distress syndrome (ARDS). When the Coronaviridae study group (CSG) of the ICTV phylogenetically proved that the 2019-nCoV formed a clade with other bat-like SARS and SARS-CoV, the name was changed to severe acute respiratory virus type 2 (SARS-CoV-2) (Coronaviridae study group, 2020).

1.2.2) Human Coronavirus Genome Structure and Organization

The CoVs genome is approximately ~30 kb with a 5′-cap structure and 3′-poly-A tail. Its genomic RNA serves as a template for replication and direct translation of polyprotein 1a/1ab (pp1a/pp1ab). PP1a/1ab codes for non-structural proteins (nsps) to form the replication-transcription complex (RTC) in its double-membrane vesicles (DMVs). This allows a nested set of subgenomic RNAs (sgRNAs) to be synthesized by the RTC in discontinuous transcription, with the subgenomic messenger RNAs (mRNAs) having similar 5′-leader and 3′-terminal sequences. Transcription regulatory sequences which are located between open reading frames (ORFs) are responsible for transcription termination and acquiring a new leader RNA. The minus-strand sgRNA serves as the templates to produce subgenomic mRNAs (V'kovski et al., 2020).

HCoVs have similar genomes and protein organization, with an open reading frame (ORF 1a/b), which encodes 16 nsps (nsp 1 – 16) and makes up approximately 2/3 of the whole genome. The −1 frameshift between ORF1a and ORF1b leads to production of two polypeptides: pp1a and pp1ab which are processed by virally encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro) and about two papain-like proteases into 16 nsps. The other ORFs located near the 3′- terminus on the remaining 1/3 portion of the genome encodes approximately four main structural proteins: spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. The S-protein is responsible for the crown-like appearance of the Spike protein that is seen on the surface of CoVs; it is also primarily important for the binding and entry of the virus into the cell, and it has the most variable sequences. The E protein, a small hydrophobic integral membrane protein, is responsible for virus assembly, while the M protein induces membrane curvature, and is associated with the envelope in all coronaviruses. The N protein is a nonspecific RNA-binding protein that forms the ribonucleocapsid with viral genomic RNA (Zhao et al., 2020).

Sequence analysis SARS-CoV-2 has shown that, like other CoVs, SARS-CoV-2 has a typical CoV genome structure and clusters with other betacoronaviruses such as SARS-CoV, MERS-CoV, Bat SARS-like (SL)-ZC45, and Bat-SL ZXC21. Phylogenetic analysis of CoVs have also shown that SARS-CoV-2 is more related to bat-like CoVs such as bat-

SL-CoV ZC45 and bat-SL-CoV ZXC21, compared to SARS-CoV (Jaimes et al., 2020; Zhou et al., 2020). An illustration of the genomic organization of CoVs is depicted in **Figure 2A**, with a phylogenetic tree showing a closer relationship between SARS-CoV-2 and bat-like CoVs, than with SARS-CoV.

Some CoVs also encode special structural and accessory proteins, such as Hemagglutinin-esterase (HE) protein, 3a/b protein, and 4a/b protein, which are all translated from the sgRNAs. The lower part of **Figure 2B** shows which HCoVs have these special structural and accessory proteins in the genome.

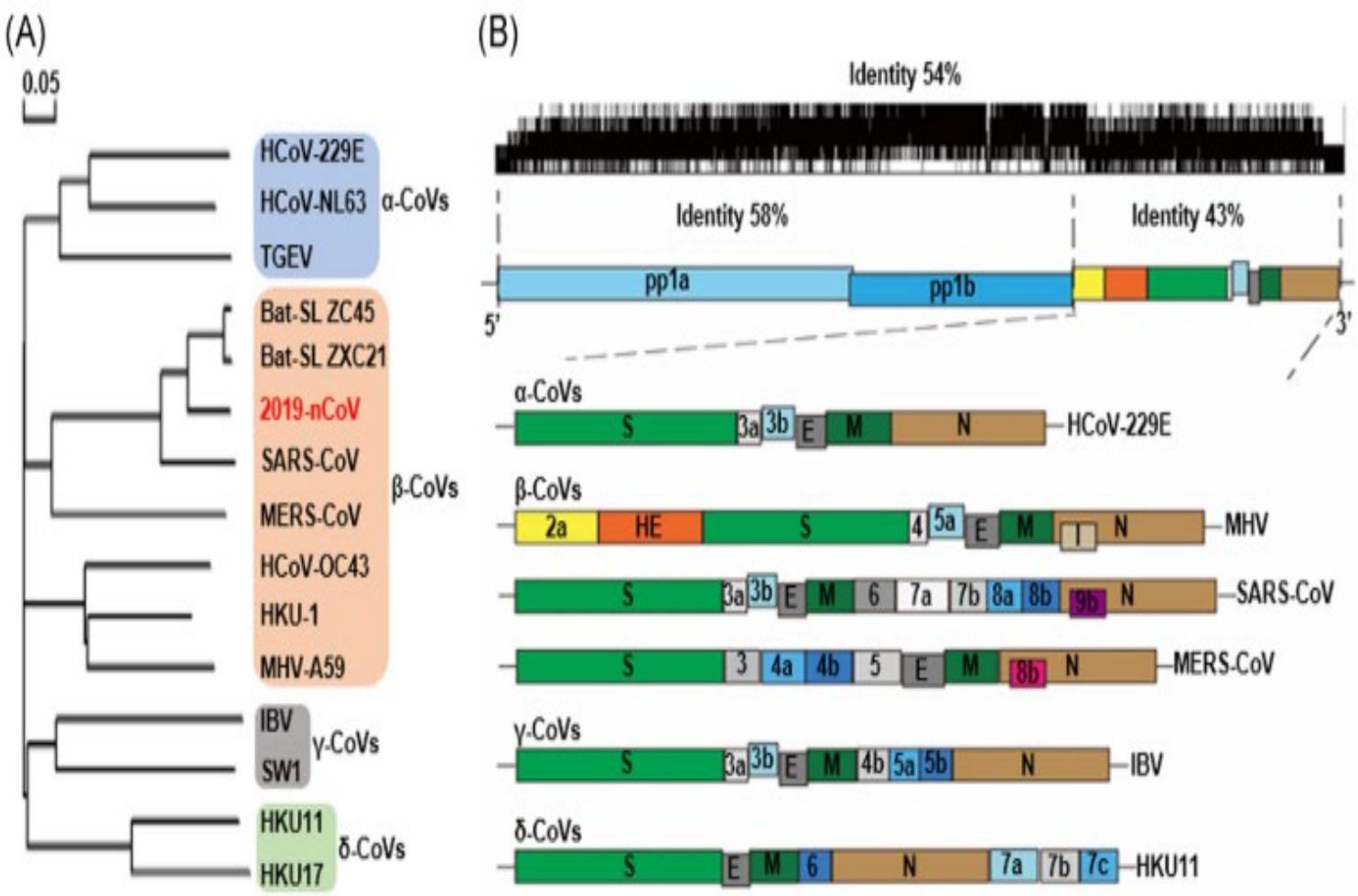


Figure 2: (A) shows a phylogenetic tree of known CoVs, including SARS-CoV-2 (shown in red as 2019-nCoV). **(B)** shows the genomic organization of CoVs with a 54% identity between the genomes of all CoVs. (Adapted from Chen et al., 2020).

During CoVs genome sequence alignment, it has been discovered that the nsp-coding region of all CoVs have a 58% identity, while the structural protein-coding region is 43% similar in all of them; at the whole genome level, these CoVs have a 54% similarity (Chen, Liu and Guo, 2020). The inference is that structural proteins are more diverse, while nsps are more conserved. It has been proposed that CoVs large genome can possibly be maintained by features of the RTC, which contains several RNA processing enzymes such as the exoribonuclease (ExoN) of nsp14 which provides a proofreading function for the RTC, and is unique to CoVs (Chen et al., 2020).

1.2.3) Emergence and Spread SARS-CoV-2 (COVID-19)

In December 2019, cases of pneumonia of an unknown cause were found in patients, who had visited a seafood market in Wuhan Province, China. These patients showed symptoms of viral pneumonia and most reported cases were linked to a wet market located in Wuhan which sells, among other things, live animals. By December 31st, 2019, the Wuhan Municipal Health Commission alerted the World Health Organization (WHO) of a pneumonia outbreak of unidentified cause. Results from investigations by the Chinese CDC released on the 09th of January 2020, reported the identification of a novel betacoronavirus thought to be the cause of this emerging disease (Zhu et al., 2020). As other patients with no history of exposure to this market were identified, who had similar a similar infection pattern, this proved that the novel CoV could be transmitted from person-to-person. Soon after, the virus spread rapidly throughout the province, and to other parts of the country. By the 30th of January 2020, the WHO declared the novel coronavirus outbreak a public health emergency of international concern (PHEIC), and the ICTV officially named this new CoV SARS-CoV-2 on the 11th of February 2020. Subsequently, the WHO named the disease “coronavirus disease 2019, denoted as ‘COVID-19’. Soon after, the virus rapidly spread worldwide leading the WHO to officially confer COVID-19 as a pandemic on the 11th of March 2020 (Wilder-Smith and Osman, 2020). At the time this work was written (November 2023), there have been over 760 million confirmed cases, worldwide. **Figure 3** highlights the main timeline of events in SARS-CoV-2 detection, spread and control by Chinese and global health authorities.

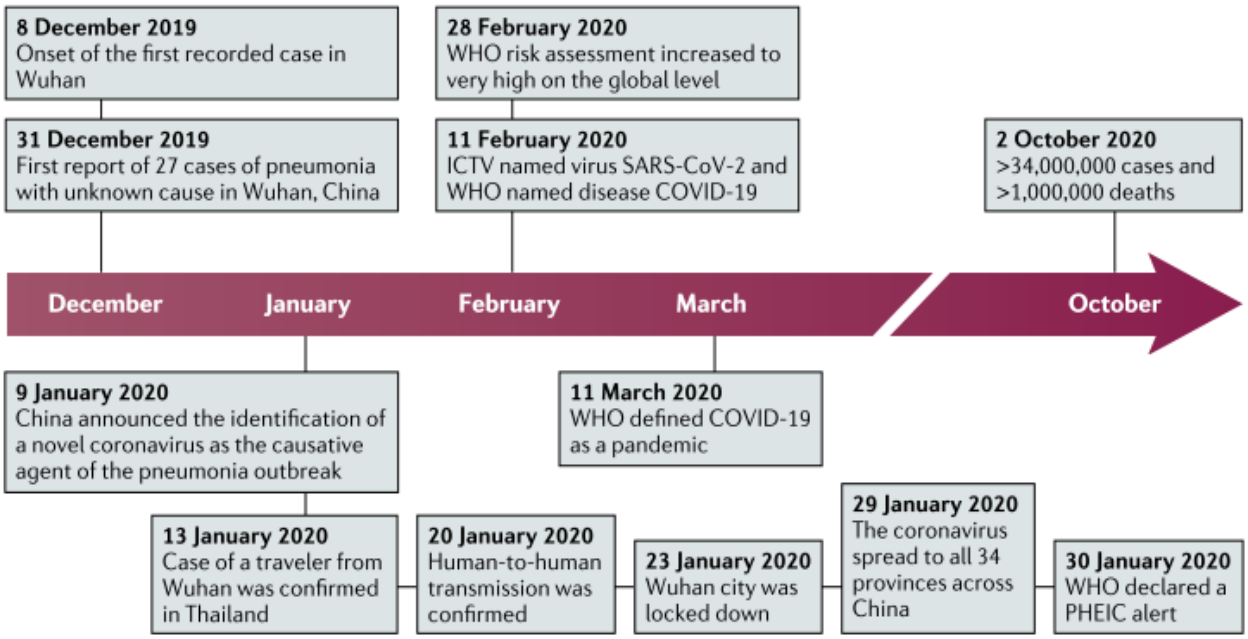


Figure 3: An illustration representing the timeline of events in the COVID-19 outbreak. (Adapted from Hu et al., 2021)

1.2.4) Origins of SARS-CoV-2

When health emergencies arise, investigations towards discovering the cause of pathogen emergence and its mode spread is launched. This information enhances understanding of associated risks, which could be employed by public health specialists for disease prevention, preparedness, and mitigation. To date, some contention remains about the origins of SARS-CoV-2, although two hypotheses have been proposed. The first hypothesis is that SARS-CoV-2 occurred from a natural zoonotic spillover, most likely from the Huanan Seafood Wholesale Market. The second hypothesis is that of a laboratory leak from the Wuhan Institute of Virology (WIV) (Gostin and Gronvall, 2023). The first hypothesis was somewhat justified, while the latter was rejected in a joint technical report published by the WHO and China in March 2021 (WHO, 2021).

Peer-reviewed evidence favours the hypothesis that SARS-CoV-2 emerged from a natural zoonotic spillover into humans (The Lancet Microbe, 2023). Early reports suggested that SARS-CoV-2 originated in bats and was transmitted to humans after passing through an animal reservoir, as was the case for SARS-CoV and MERS-CoV.

Bats are a natural host of alpha- and beta- coronaviruses (Han et al., 2019; Ruiz-Aravena et al., 2022), and several bat species (spp.) have been implicated as progenitors of SARS-CoV-2. First on the list is the Rhinolophus affinis bat (also known as the horseshoe bat) from Yunnan Province, China, whose CoV strain is named 'RaTG13', and genome analysis has shown a 96.2% sequence identity to that of SARS-CoV-2, and phylogenetic analysis has confirmed SARS-CoV-2 cluster to RaTG13 (**Figure 4**). However, RaTG13 only shares one of the six critical residues of SARS-CoV-2 receptor-binding domain (RBD) in the spike which interacts with the human ACE2 receptor. The Rhinolophus malayanus bat from the Yunnan Province from which a coronavirus (RmYN02) was isolated, shares 93.3% identity to SARS-CoV-2. Although RmYN02 contains some similar insertions at the S1/S2 cleavage site in SARS-CoV-2 spike, it has some deletions in the RBD which fails to bind with human ACE2 (Zhou et al., 2020). Other SARSr-CoV-2 viral genome sequences from bats have been reported in Eastern China, Japan, and Southeast Asian countries (Murakami et al., 2020; Wacharapluesadee et al., 2021). However, a progenitor virus that shares >99% identity with SARS-CoV-2 remains unknown.

One study identified a novel lineage of SARSr-CoVs, (RaTG15), as well as seven other viruses, from bats at the same location where RaTG13 was found (Guo et al., 2021). However, RaTG15 and the related viruses share 97.2% amino acid sequence identity with SARS-CoV-2 in the conserved ORF1b region and shows less than 77.6% nucleotide identity to all known SARSr-CoVs, thus forming a distinct lineage in the Sarbecovirus phylogenetic tree. Furthermore, the RBD of RaTG15 can bind to ACE2 from Rhinolophus affinis, Malayan pangolin, which serves as an entry receptor, except in human ACE2. However, it contains a short deletion and has different key residues responsible for ACE2 binding. Another study showed that SARS-CoV-2 genomic diversity prior to February 2020 likely comprised two distinct viral lineages, denoted "A" and "B" (Pekar et al., 2023). The applied phylodynamic rooting methods, coupled with epidemic simulations, revealed that these lineages most probably resulted from at least two separate cross-species transmission events into humans. They postulated that the zoonotic transmission likely began with lineage B viruses about 18 November 2019, followed by a separate introduction of lineage A which likely occurred within weeks of this event. These findings

indicate that it is unlikely that SARS-CoV-2 circulated widely in humans before November 2019. This also defines the period when SARS-CoV-2 first jumped into humans and when the first cases of COVID-19 were reported. The most probable explanation for the introduction of SARS-CoV-2 into humans involves zoonotic jumps from as-yet-undetermined, intermediate host animals at the Huanan market.

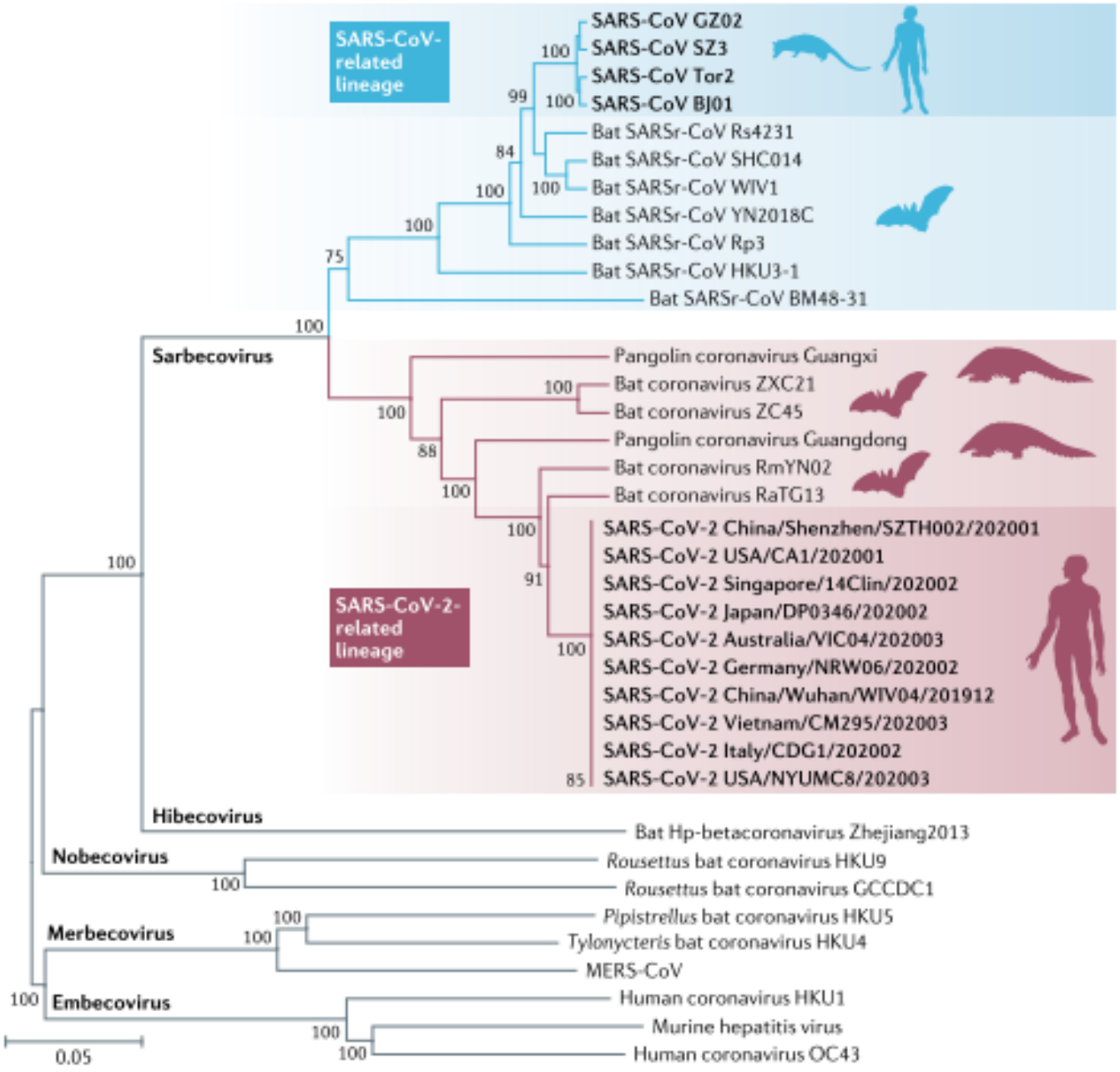


Figure 4: Phylogenetic tree of SARS-CoV-2, SARS-related CoVs and other betacoronaviruses. SARS-CoV-2 clusters with bats and pangolin-related viruses (Adapted from Hu et al., 2020)

1.2.5) Structure, Genomic Organization and Replication of SARS-CoV-2

Structurally, electron micrographs have shown that the virus is spherical in shape, like that of a solar corona or with a sun-like morphology, from which its name Coronavirus, was derived (Alanagreh et al., 2020; Alsobaie, 2021). It is 60–140 nm in diameter, with distinctive spikes of about 9–12 nm, and made up of five types of protein, spike protein (S), envelope protein (E), membrane protein (M) and hemagglutinin-esterase dimer protein (HE) and nucleocapsid protein (N) which are bound to the RNA (**Figure 5**). These S, E, M and HE proteins are on the surface, while the genetic material which permits replication is inside the virion, together with N-protein (Haque et al., 2020). Its genome consists of 14 open reading frames (ORFs) which encode 27 proteins arranged from the 5' to the 3' region of the genome (**Figure 6**).

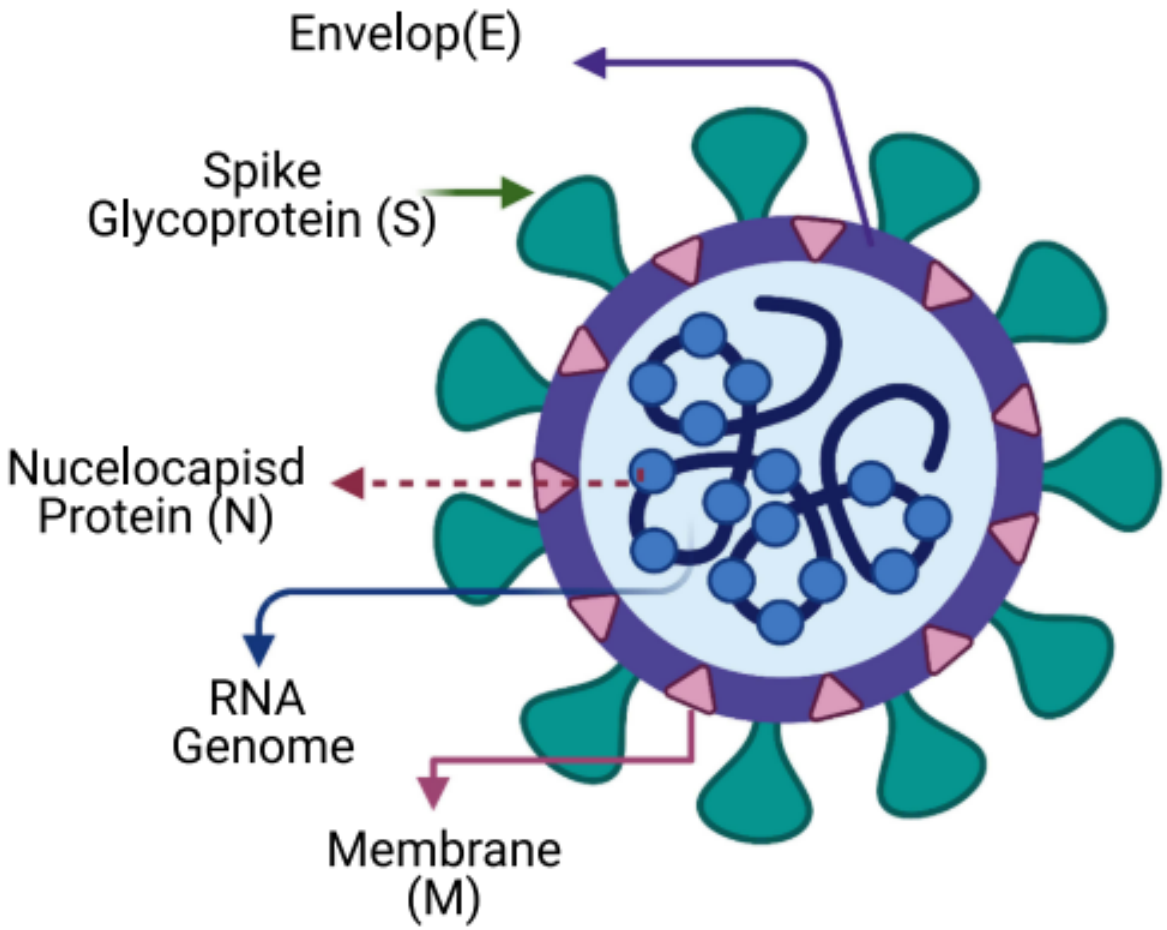


Figure 5: Morphological structure of SARS-CoV-2 showing key structural proteins and the location of its RNA genome. (Adapted from Alsobaie et al., 2021)

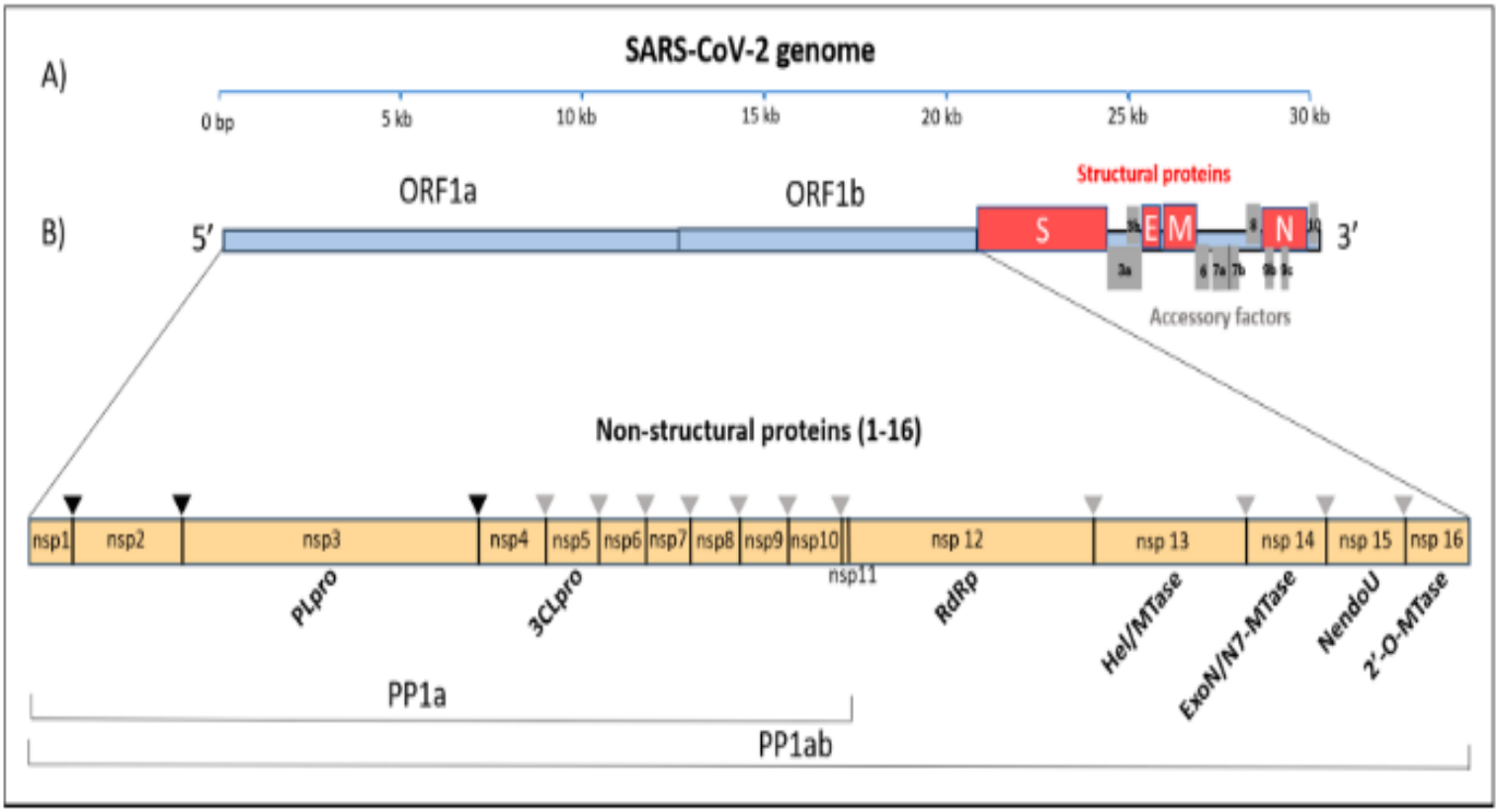


Figure 6: An illustration of the SARS-CoV-2 genome organized in individual ORFs, highlighting PP1a/b which encodes 16 non-structural proteins of the virus. (Adapted from Romano et al., 2020).

In the initial steps of coronavirus infection, the S-protein which has two subunits, S1 and S2, binds to the angiotensin-converting enzyme 2 (ACE2) cellular entry receptors, allowing the virion to penetrate the human cells by a process known as endocytosis (**Figure 7**). The S1 subunit has a receptor-binding domain (RBD) which binds with the receptor-binding motif (RBM) of cell surface receptor while the other part of the S-protein (S2 subunit) facilitates fusion with the host cell membrane. The spike protein is then cleaved by host proteases located on S2 sub-unit to make necessary conformational changes for membrane fusion. This is facilitated by the type II transmembrane serine protease (TMPRSS2). After entering the cell, the genomic RNA is released and uncoated, which allows for immediate translation of two large open reading frames, ORF1a and ORF1b. This forms pp1a and pp1ab which are co-translationally and post-translationally cleaved into individual non-structural proteins (nsps), these nsps form the viral replication and transcription complex (RTC). Once the nsps are expressed, the peri-nuclear double-

membrane vesicles (DMVs), convoluted membranes (CMs) and small open double-membrane spherules (DMSs), which are viral replication organelles create a protective environment for genomic RNA replication, including subgenomic mRNA (sgmRNA) transcription, resulting in a nested set of coronavirus mRNAs. Structural proteins which are translated are then translocated into endoplasmic reticulum (ER) membranes and move through the ER-to-Golgi intermediate compartment (ERGIC). Here, the structural proteins interact with the N-encapsidated, newly produced genomic RNA which leads to budding into the lumen of secretory vesicular compartments. Finally, virions are secreted from the infected cell by exocytosis (Rastogi et al., 2020; V'kovski et al., 2020).

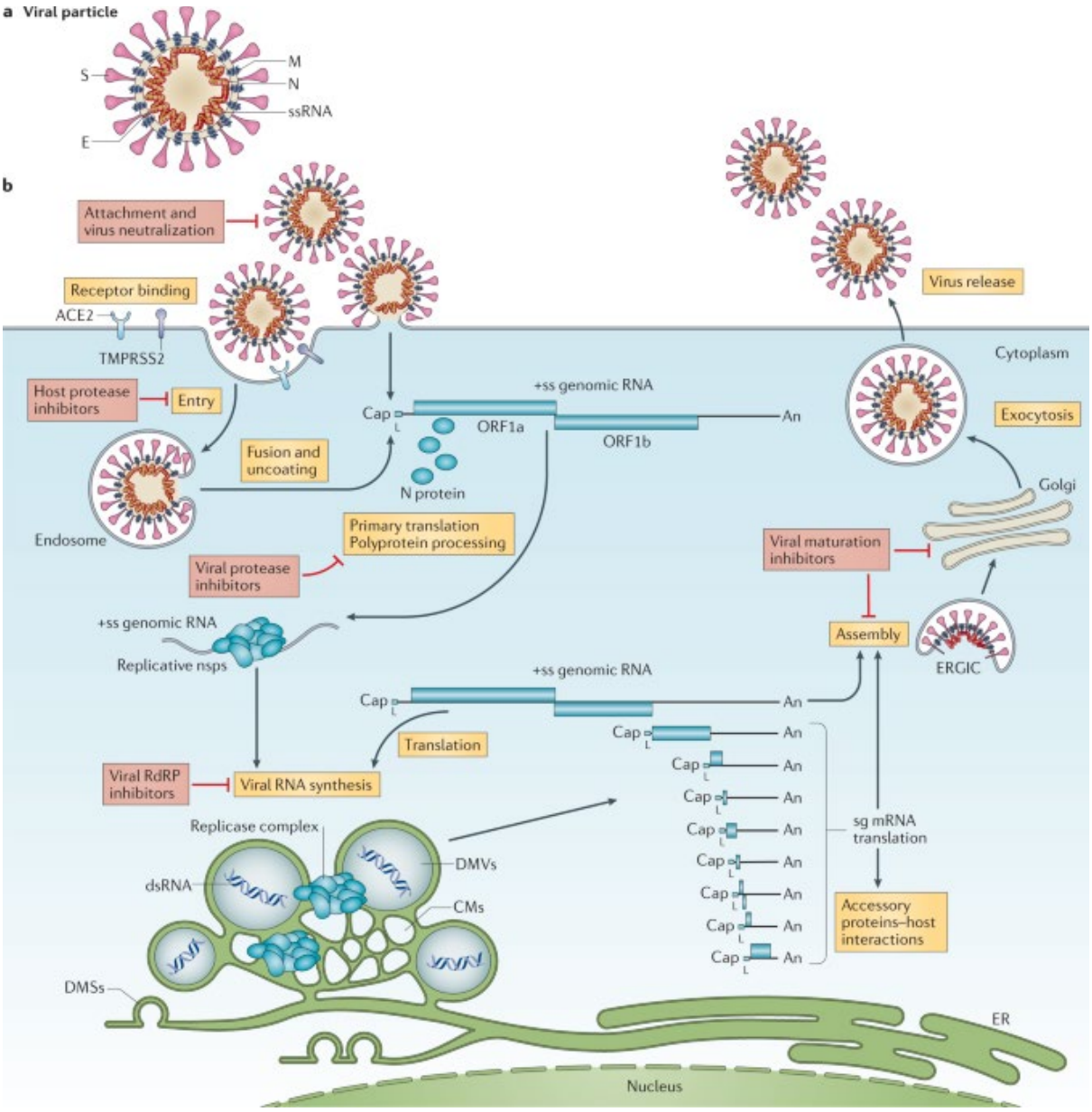


Figure 7: SARS-CoV-2 lifecycle from viral entry though mediation by the S-protein binding to the ACE2 receptor, and cleavage by S2 TMPRSS2 proteases to viral release into the cell (Adapted from V'kovski et al., 2020).

1.2.6) Transmission, Clinical Manifestation and Detection of SARS-CoV-2

Like other coronaviruses, the main route of SARS-CoV-2 transmission is respiratory, through infected respiratory droplets. Viral infection occurs through direct or indirect contact with nasal, conjunctival, or oral mucosa, when respiratory particles are inhaled or deposited on these mucous membranes. This is because the ACE2 target host receptors

are mainly present in the human respiratory tract epithelium (the oropharynx and upper airway). It has also been speculated that the gastrointestinal tract as well as the conjunctiva may also serve as pathways for infection (Meyerowitz et al., 2021).

The dominant route of SARS-CoV-2 transmission is through proximity of contact such as talking, coughing, or sneezing within a 2 meters face-to-face distance for approximately 15 minutes. Viral transmission through close contact is most efficient within households and indoor gatherings (Cevik et al., 2020). Fomite transmission, that is, contact through inanimate surfaces such as stainless steel, plastic, glass is another method through which SARS-CoV-2 is transmitted since the virus can remain viable on these surfaces for days. Hence hand contact with contaminated surfaces to the mucosa of eyes, nose, and mouth could lead to possible transmission (CDC, 2021). To prevent such transmission, the WHO has resolved that disinfectants should be used to clean surfaces and handwashing, and sanitization has been encouraged to prevent transmission since the virus is readily inactivated by disinfectants. Although it has been proven that SARS-CoV-2 is shed in human feces, orofecal transmission is yet to be confirmed (Heneghan et al., 2021).

Post-exposure to the virus, it has been estimated that there is an incubation period (time from exposure to symptom onset) of approximately 2 – 7 days, although ~97.5% of patients will be symptomatic around the 11th day (Wiersinga et al., 2020). Some of these symptoms include: fever, cough, shortness of breath, fatigue, muscle or body aches, headache, new loss of taste or smell, sore throat, congestion or runny nose, nausea or vomiting, and diarrhoea (CDC, 2021). Severity of the infection may vary from asymptomatic individuals to those with comorbidities or patients advanced in age, who may eventually end up having severe cases of pneumonia which can lead to death.

The presence of SARS-CoV-2 is typically detected through molecular, serologic or antigen detection. Nucleic acid amplification tests (NAAT) using the Reverse transcription polymerase chain reaction (RT-PCR) technique has been the golden standard for detection of positive infection, since it can determine if active infection (acute infection). Other NAAT techniques used for testing are the loop-mediated isothermal amplification (LAMP) based and clustered regularly interspaced short palindromic repeats (CRISPR-based) assays. To confirm positive results from a patient suspected to be COVID-19

positive, a nasopharyngeal, oropharyngeal, nasal mid turbinate, or anterior nares swab is collected, and RT-PCR conducted for detection of a specific target of interest. Several testing kits have been approved for use by the food and drugs agency emergency use authorization (FDA EUA) unit (Xu et al., 2020). These kits amplify various genes of the SARS-CoV-2 genome such as the spike (S) protein, various regions of the nucleocapsid (N) protein, the membrane (M) protein, the envelope (E) protein, as well as ORF 1a and 1b (ORF 1a/b).

Serological assays detect the presence of antibodies against SARS-CoV-2 antigens in serum, plasma, or whole blood specimens. This has been an alternative detection method with a short turn-around time that can be used for diagnosis of asymptomatic infections (Lai and Lam, 2021). It has been suggested that these assays be used either in combination with molecular testing or as additional testing for suspected cases with negative nucleic acid results, as this will help improve detection accuracy of COVID-19. Shaw and colleagues conducted a study in Cape Town to determine SARS-CoV-2 seroprevalence and estimate the specificity of anti-SARS-CoV-2 antibody tests (Shaw et al., 2021). They revealed that most positive serology tests were for individuals who reported no symptoms within the last 6 months. These results are in line with suggestions that these assays could be used in combination with molecular testing or as additional means of testing for suspected cases with negative nucleic acid results which will improve COVID-19 detection accuracy. They also corroborate the idea that serological tests could be a useful tool for disease surveillance and epidemiological research to know the actual number of infections and enable more accurate determination of case fatality (Mayne et al., 2020). Point-of-care (POC) diagnostic tests are also used for detection outside laboratory settings, since they have a short turn-around time (Smithgall et al., 2021).

1.2.7) Evolution and Genetic Diversity of SARS-CoV-2

Genetic diversity in SARS-CoV-2, as is the case with other RNA viruses, is critical for its fitness, survival, and probably its pathogenesis. Two main drivers of genetic diversity in SARS-CoV-2 have been identified. These include: its high mutation rate and recombination (Rahimi et al., 2021). In 2020, genomic data available in GISAID by September 2020, estimated the SARS-CoV-2 mutation rate to be approximately 8 × 10-4

nucleotides/genome per year (Rahimi et al., 2021). However, more recent data suggests a mutation rate between 10^{-5} and 10^{-3} (Abavisani et al., 2022). These evolution patterns lead to the development of either a new variant, strain or lineage (Dubey et al., 2022). A new variant can emerge when specific mutations or sets of mutations are selected through numerous rounds of viral replication; a variant could be co-termed a strain when the sequence variation produces a virus with distinctly different phenotypic characteristics. A new lineage is designated from analysis of genetic and phylogenetic sequences and a new variant is detected as a distinct branch on a phylogenetic tree (Mascola, Graham and Fauci, 2021). Developed variants were categorized into variants of interest (VOI), variants of concern (VOC), variants being monitored (VBM), and variants of high consequence (VOHC), by the WHO Technical Advisory Group on SARS-CoV-2 Virus Evolution (TAG-VE). The TAG-VE group was instrumental in characterizing SARS-CoV-2 emerging variants and monitoring their potential impact on vaccines, therapeutics, and diagnostics. Categorizations of these variants was based on their susceptibility existing treatment and vaccines. VOI are those associated with increased transmissibility and decreased viral neutralization by antibodies. VOCs have all characteristics displayed by VOI, as well as increased disease severity resulting in increased hospitalizations, deaths, and disruption of existing diagnostic assays, and reduced efficacy of available vaccines or therapies (CDC, 2023). Throughout the pandemic, five VOCs have been described (Alpha, Beta, Gamma, Delta and Omicron). These variant names were proposed by the WHO to assist with easy communication of new variants, since existing classification bodies (PANGO and Nextclade) had different classification schemes (WHO, 2021).

The Alpha variant (B.1.1.7) was first detected in the United Kingdom (UK) in September 2020 (Hill et al., 2022). This variant is characterized by 11 main mutations, with three key mutations (N501Y, P681H, delH69V70 and delY144/145) in the Spike region. This was closely followed by the Beta variant (B.1.351), first reported December 2020, in South Africa (Tegally et al., 2021) and was predominantly detected in the population during the second wave. It is characterized by eight lineage-defining mutations, with three key mutations (K417N, N501Y, and E484K) in the receptor-binding domain (RBD) of the Spike protein. Next came the Gamma variant (P.1), first identified in Brazil (Faria et al., 2021),

having 11 main mutations and this variant was sparsely detected in Africa (Tegally et al., 2022). This was closely followed by the Delta variant (B.1.617.2), first identified in India in March 2021 (Dhar et al., 2021) and was primarily responsible for the third wave of infections across the globe. It has eight main mutations, with the L452R and P681R mutations being the key mutations. The Omicron VOC (B.1.1.529) identified in Botswana and South Africa, was in circulation in early November 2021 (Viana et al., 2022). This variant, characterized by more than 47 mutations (Standford University Coronavirus Antiviral & Resistance Database, 2023), was predominantly detected in most infections during the fourth wave in South Africa. As at the time of this study (November 2023), the Omicron VOC, and its sub-lineages are still the dominantly circulating variant across the globe.

The first wave in South Africa (15 June – 24 August 2020) was characterized by a mixture of B.1 lineages (Viana et al., 2022). However, the second wave of COVID-19 infections which occurred in South Africa was recorded from 23 November 2020 – 01 February 2021, and characterized by dominance of the Beta VOC. The third wave in South Africa was characterized by the dominance of the Delta VOC. This was noted by May 2021 when 30% of the sequenced COVID-19 samples consisted of the Delta variant. By June 2021, the Delta variant accounted for 45% of the sequenced samples, while Beta variant occurrence in sequenced samples dropped by 31% (NICD, 2021). The fourth wave was dominated by the circulation of the Omicron variant (Vianna et al., 2022). This wave began when a rapid increase in COVID-19 cases was observed by mid-November 2021 in Gauteng province. Sequencing of positive samples revealed the presence a new and genetically distinct lineage of SARS-CoV-2. This fourth wave occurred from 06 December 2021 – 14 February 2022.

1.2.8) Effects of SARS-CoV-2 Mutations

Several studies have documented that SARS-CoV-2 mutations impact viral transmissibility, immune evasion, and clinical severity. As the virus has evolved over time, several variants have arisen, each with notable mutations which have changed the pandemic's dynamics, and subsequent public health responses.

Mutations in the spike protein, particularly in the receptor-binding domain (RBD), have been reported to enhance the viral binding affinity to human ACE2 receptors, which in turn increases viral infectivity. The first documented mutation was mutation D614G, first recorded in March 2020, then spread rapidly by April and May, detected in approximately 70% of sequenced patient samples. D614G is a spike protein mutation which targets the S1/S2 junction region. This mutation is associated with increased transmissibility and infectivity (Zhang et al., 2020; Alquraan, Alzoubi and Rababa'h, 2023). Other mutations found in the Spike protein such as N501Y are associated with increased binding affinity, which contributed to rapid transmission of the Alpha variant (Tian et al., 2021; Liu et al., 2022), while E484K increases immune evasion and reduce vaccine efficiency. As such, variants carrying this E484K mutation (Beta and Gamma) showed reduced susceptibility to neutralization antibodies (Uwamino et al., 2022; Young et al., 2022). This led to updated vaccination booster formulations, which focused on achieving high efficacy. The Omicron variant is also associated with increased immune evasions, which can result in COVID-19 reinfection and breakthrough cases, even among those who have been vaccinated (Mohsin and Mahmud, 2022; Dai et al., 2024).

SARS-CoV-2 mutations are also known to have an impact on clinical severity of those infected. For instance, the Alpha variant was associated with potential increase in disease severity, which may lead to higher hospitalization rates (Florensa et al., 2022). Similarly, the Delta variant has been reported to cause increased transmissibility and clinical severity, which led to higher hospitalizations, as well as severe disease outcomes, particularly among unvaccinated individuals. In South Africa, this variant was responsible for the third wave, leading to increased infections and hospitalizations. This led reinforcement of strict public health measures, where the lockdown phase in South Africa was reinstated from stage one back to stage two. Compared to the Delta variant, Omicron is associated with less disease severity (Esper et al., 2023).

The impacts of SARS-CoV-2 mutations were also observed on the sensitivity of administered diagnostic tests. Some mutations in the S-gene, such as deletions at position H69V70 (present in Alpha and Omicron variants), were reported to lead to S-Gene Target Failure (SGTF) in certain PCR tests. This SGTF was used as a marker for

identification of these variants (McMillen et al., 2022; Subramoney et al., 2022). Mutations could also impact the accuracy of administered diagnostic tests. This is because PCR and antigen-based diagnostic tests rely on detecting specific viral proteins or genetic sequences. Thus, mutations in these target regions can potentially affect the accuracy and sensitivity of these tests, leading to either false-negative or false-positive results (Wang et al., 2020; Varshney, 2023). Adopting strategies such as continuous genetic surveillance, test optimization, diversifying testing approaches and vaccine monitoring, have been proposed to combat these challenges (Goławski et al., 2022). Some mutations in SARS-CoV-2 variants affect therapeutic efficacy, with some showing reduced susceptibility to some monoclonal antibody therapies. This leads to neutralization of the effects of immunotherapy and escape the protective immunity conferred by vaccines (Yaqinuddin et al., 2021). As a result, new antibody treatments or a combination of treatments are being developed and approved that target a broader range of variants. Similarly, when antiviral drugs were administered to COVID-19 patients, antivirals such as remdesivir and molnupiravir, which target viral replication mechanisms are reported to be less susceptible to viral mutations. Thus, their efficacy is stable. Some studies have reported that molnupiravir retains its efficacy against Alpha, Beta, Gamma, Delta, and Omicron VOCs (Vangeel et al., 2022), with Omicron being highly sensitive to molnupiravir. Other studies have indicated that even though Omicron is sensitive to this antiviral, a need to investigate molnupiravir's effectiveness in treating Omicron infection remains (Li et al., 2022), as well as its effect in breakthrough infections (Tian et al., 2022).

1.2.9) Tropism Impacts of Mutations Detected

Tropism in SARS-CoV-2 refers to the virus' ability to infect specific cell and tissue types within the body. This is largely determined by the distribution of ACE2 receptors, which facilitate SARS-CoV-2 entry into human cells. Understanding SARS-CoV-2 tropism can improve investigation towards prospective treatments, especially for severe or multi-organ complications, since it highlights which tissues and organs may need targeted intervention. Several reports have documented primary tropism which occurs in the upper and lower respiratory tracts. In the upper respiratory tract, SARS-CoV-2 infects cells in the nasal and pharyngeal tract, with high concentration of ACE2 receptors, which allows

the virus to establish infection in the upper respiratory tract (Perrotta et al., 2020; Hu et al., 2021). SARS-CoV-2 infection in the lower respiratory tract and lungs is reported to have strong affinity for alveolar cells in the lung epithelium, especially type II pneumocytes, goblet, nasal epithelial/ciliated and oral mucosal cells (Kaur, Lungarella and Rahman, 2020). Thus, rapid replication may trigger a strong immune response, such as the cytokine storm syndrome, which causes acute respiratory distress syndrome (ARDS) and subsequent respiratory failure, the main cause of death in COVID-19 patients (Huang et al., 2020)

Secondary tropism has also been documented in other organs such as the gastrointestinal tract, heart, central nervous system (CNS), as well as other organs or tissues such as the kidneys and liver cells. The presence of ACE2 receptors in other organs such as the gastrointestinal tract, accounts for observed gastrointestinal symptoms like nausea, vomiting, and diarrhea (Zhang, Garrett and Sun, 2021), with viral genomes also detected in stool samples. Previous studies have indicated that cardiac damage in SARS-CoV-2 infection may be caused by mechanisms such as Spike-ACE2 interaction, cytokine storm, and hypoxemia (Wu et al., 2020). Both ACE2 and TMPRSS2 genes are minimally expressed in cardiac cells, suggesting an alternative viral entry mechanism in SARS-CoV-2 infection within the heart (Behboudi et al., 2024). Another report indicates that the virus may utilize other entry factors like cathepsin L, which is abundant in cardiomyocytes, rather than relying on the TMPRSS2-dependent pathway. This demonstrates the complex ways in which cardiac tissues may be impacted by SARS-CoV-2 (Veluswamy et al., 2021). SARS-CoV-2's neurotropic capability is reported to result from its binding to the transmembrane receptor neuropilin-1 (NRP1), which is highly expressed in the olfactory epithelium, olfactory tubercles, and paraolfactory (Arunachalam et al., 2021). This potentially spreads to the central nervous system (CNS) through the olfactory nerve (Meinhardt et al., 2021), with associated CNS complications such as encephalopathy, progressive dementia, Prion-like disease, Guillain-Barré syndrome after COVID-19 acute infection and vaccination (Rauf et al., 2021; Xia, 2021). Some reports have also documented SARS-CoV-2 entrance into podocytes and tubular epithelial cells and is associated with renal abnormalities, such as hematuria, proteinuria, and acute kidney injury (Martinez-Rojas, Vega-Vega and Bobadilla, 2020). Approximately

36 – 46% of COVID-19 patients are reported to experience acute kidney damage (AKI), and liver damage also reported in severe cases. ACE2 is reported to be expressed in both liver and bile duct cells, with higher expression observed in bile duct cells (Xu et al., 2020; Marques et al., 2021). ACE2 and TMPRSS2 have been detected in trophoblast, hypoblast, and syncytiotrophoblast tissues, which imply the possibility of vertical SARS-CoV-2 transmission (Weatherbee, Glover and Zernicka-Goetz, 2020). Stem cells have also been reported to be potentially infected, which could lead to damage, since ACE2 is expressed in endothelial progenitor cells (EPCs) and hematopoietic stem cells (HSCs). ACE2 and TMPRSS2 are reported to be present in small populations of human umbilical cord blood (UCB)cells, which could differentiate into functional EPCs and HSCs (Ratajczak et al., 2021). The surface protein of red blood cells (RBCs; Band3 protein) has been suggested as a potential entry point for SARS-CoV-2. Some studies have indicated that platelet activation induced during SARS-CoV-2 Spike binding to ACE2 on platelets, may contribute to thrombus formation and inflammatory responses in COVID-19 disease (Cosic, Cosic and Loncarevic, 2020; Lichtenberger and Vijayan, 2021).

Factors such as host receptor distribution, host protease activity and some SARS-CoV-2 variants may influence tropism. The distribution and density of the host's ACE2 receptor plays a significant role in determining viral tropism. Thus, higher expression of ACE2 in certain tissues makes them more susceptible to viral binding. Host protease activity refers to the ability of cellular proteases like TMPRSS2, to facilitate viral entry through fusion with the virus's spike protein. Cells with higher levels of TMPRSS2 are more susceptible to SARS-CoV-2 infection. Certain SARS-CoV-2 variants may also influence tropism by exhibiting enhanced binding affinity for ACE2, which can influence transmission patterns and symptom severity. Mutations such as N501Y, L452R, K417N have been associated with an increased binding affinity to ACE2, leading to higher infectivity and transmissibility (Kumar et al., 2023).

1.2.10) SARS-CoV-2 Treatment and Vaccination

Several antivirals have shown some benefits in reducing the impact of COVID-19 infection, though there have been no proven effective therapies. Various therapies are being evaluated constantly for their efficacy against SARS-CoV-2, which is tracked by the

COVID-vaccine tracker (Biorender, 2021). So far, the drugs being used are those which inhibit key stages in the viral replication cycle. Immunomodulatory agents have been proposed as possible treatment for SARS-CoV-2, however, clinical trials are ongoing around the world for the establishment of their efficacy. Published studies show that immunoglobulin therapy, such as covalescent plasma therapy and monoclonal antibody therapy, are potential therapies for treating COVID-19. However, adverse effects associated with covalescent plasma therapy, as well as the cost of monoclonal antibody production have limited their applicability.

Vaccination has shown to be the most promising method of COVID-19 control since it offers immune protection against the virus. The S-protein has been identified as the immunodominant antigen of the virus, and research reveals that binding and neutralizing antibodies primarily target the RBD domain of the S1 subunit (Creech, Walker and Samuels, 2021). Hence, it is assumed that preventing binding would likely prevent SARS-CoV-2 infection. Therefore, most available vaccines target the S-protein. Several vaccine platforms are being employed for developing vaccines against SARS-CoV-2, such as mRNA in lipid nanoparticles, protein subunits, recombinant vectors, DNA, inactivated virus vaccines and live attenuated virus vaccines. As of April 2021, there are approximately 216 vaccines, with 92 of them in human trials, according to the COVID-19 vaccine and therapeutics tracker (Biorender, 2021). The most utilized vaccines across the globe are double dose vaccines which are administered through intramuscular method (Triggle et al., 2021). Two vaccines are currently approved for administration in South Africa, the Janssen & Janssen (J & J) single dose vaccine and the Pfizer-BioNTech vaccine. The J & J vaccine is an adenovirus serotype-26 vector-based vaccine that expresses the SARS-CoV-2 S-Protein (Ad26.COV2.S). The vector is also replication-incompetent, so there is no risk of infection to the vaccinated individual. Upon vaccination, the individual will generate an immune response against the S-protein and creates antibodies that protects the individual from SARS-CoV-2 infection. This vaccine has an efficacy of 66.9% as reported from clinical trials (Johnson and Johnson, 2021; Sadoff et al., 2021). The Pfizer-BioNTech vaccine is an mRNA vaccine that consists of the full length of the S-protein. Once administered, the S-protein is delivered into host cells, giving the cells instructions to make copies of these proteins. An immune response is then

stimulated, eliciting the production of antibodies and development of memory cells. Clinical trials conducted showed a 95% efficacy of this vaccine to SARS-CoV-2 (Sahin et al., 2021; Rahman et al., 2022).

1.2.11) History of Wastewater-based Epidemiology

For decades, wastewater (WW) has been used to investigate the concentrations of use substances, pharmaceuticals, chemicals, exogenous contaminants and nutrient concentrations. However, these analyses have focused on examining the concentrations of substances that entered the wastewater treatment process, how efficient the wastewater treatment processes were, and evaluation of wastewater effluent as a source of environmental contamination (Choi et al., 2018). The idea of using biomarkers in wastewater to investigate human activity such as illegal drug consumption within a community was first proposed by Daughton (2001), in a theory called wastewater-based epidemiology (WBE). This theory was first tested in 2005 by a group of researchers seeking a more objective approach to estimate the community consumption figures for a common drug of abuse, cocaine. This method was proposed at a time when consumption of illegal drugs was increasing in young people around the globe, and methods used to estimate the trend and magnitude of drug abuse did not give a realistic estimate of drug abuse in local populations. Hence, new methods which could monitor and detect any changes in drug abuse within the said population, in real-time were needed.

This approach was based on the theory that “using non-intrusive drug monitoring methods to investigate sewage could determine collective drug usage parameters at the community level”. The approach assumes that any substance that is excreted by humans and is stable in wastewater can be used to back-calculate the original concentration excreted by the serviced population (Zuccato et al., 2005). From the time of its first application, WBE has been adapted as a tool to monitor biomarkers which indicate a population’s use of licit drugs, pharmaceutical and personal care products, population markers, biological markers (particularly pathogen surveillance), leading to the utilization of this tool for many other fields (Choi et al., 2019). **Figure 8** illustrates the various fields which have adopted the use of WBE methods, and other prospective methods that could be applied in the future.

One such growing field is the application of WBE for the surveillance of infectious pathogens circulating in communities. Previously, analysis in water was done to detect viruses present, monitor the risk of infection, the incidence and behavior of viruses in water, as well as conduct routine water quality checks to test the compliance of water quality to the recommended guidelines. However, a growing amount of research has been geared towards identifying strains in the water environment as a tool for epidemiological studies of waterborne viruses which could give information of the types of viruses circulating in the community which could cause both symptomatic and asymptomatic infections (Bosch et al., 2008). A wide variety of viruses (both RNA and DNA viruses) can be found in human sewage, particularly those which infect the gastrointestinal tract and are shed in human feces, which end up either in the environment or wastewater treatment plants (WWTPs). Of the wide variety of viruses found in wastewater, the most monitored viruses include Enterovirus (Poliovirus, Coxsackie A and B virus, Echovirus), Hepatovirus (Hepatitis A virus), human Rotavirus, Norovirus, Hepevirus (Hepatitis E virus) and Coronavirus (human Coronavirus).

These viruses are widely studied due to their ability to cause serious debilitating effects in humans, with symptoms ranging from mild fever, headaches, diarrhoea to serious and life-threatening symptoms such as acute flaccid paralysis (caused by Poliovirus) and jaundice (Hepatitis) which could lead to death. Thus, WBE methods have been applied for the surveillance of these viruses in various communities around the world (Hovi et al., 2012; Hellmér et al., 2014; Tiwari and Dhole, 2018), which could serve as an early warning system (EWS) in the case of a potential outbreak spreading in the community. Information about these infectious pathogens enables the rapid implementation of interventions to prevent continuous spread within the affected community. One important application of WBE which served as an EWS was seen when regular monitoring of Poliovirus occurrence in a small community in Rahat, Israel in May 2013 revealed a silent outbreak of Polio in the community, years after Israel was declared Polio free. This outbreak was

rapidly contained by early 2014 after the launching of a vaccine program (Brouwer et al., 2018b).

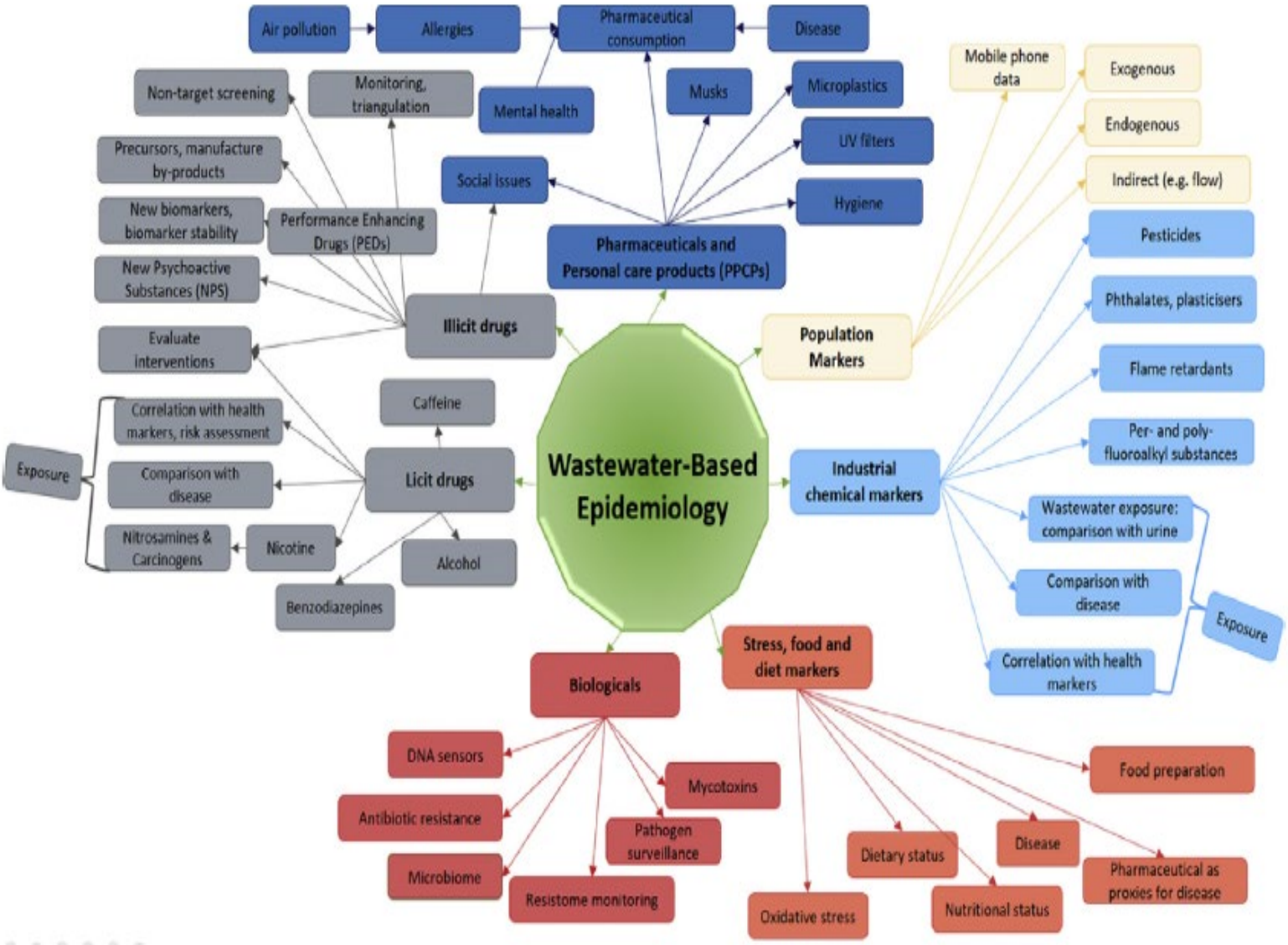


Figure 8: Application of WBE principle to a wide variety of fields (adapted from Choi et al., 2018).

1.2.12) Global Application of WBE for SARS-CoV-2 Surveillance

During the COVID-19 pandemic, WBE was adapted for surveillance of SARS-CoV-2 in several communities around the world. This is based on evidence that SARS-CoV-2 is shed in human fecal matter as well as in urine (though in very minute concentrations) and will end up in sewage systems. Evidence suggests that faecal shedding may occur approximately 3 – 5 days before the onset of other symptoms (Buscarini et al., 2020; He

et al., 2020). Research has also shown that despite the absence of symptoms in infected individuals (asymptomatic, pre-symptomatic and post-symptomatic), viral shedding may still occur (Tang et al., 2020; Wölfel et al., 2020; Zhang et al., 2020a; Zhang et al., 2020b), although asymptomatic individuals may shed a lesser quantity of the virus in feces compared to those showing symptoms and need hospitalization (Polo et al., 2020; Prasek et al., 2022). Based on these clinical observations, municipal wastewater of affected communities might contain the excreted virus from symptomatic and asymptomatic individuals. Early reports documenting molecular detection of SARS-CoV-2 in wastewater samples, and subsequent applicability to monitor infection trends in communities were conducted in the Netherlands and USA (Lodder and de Roda Husman, 2020; Medema et al., 2020). Subsequently, another study conducted in Brisbane, Australia applied WBE to monitor the trends of COVID-19 infections, as well as predict the number of those infected in the catchment area (Ahmed et al., 2020). These findings led to subsequent studies globally, which also utilized wastewater to monitor SARS-CoV-2 occurrences, and identify infection hotspots in studied catchment areas (Wu et al., 2020; Nemudryi et al., 2020; Randazzo et al., 2020). Monitoring of SARS-CoV-2 in wastewater was conducted in various settings; from university/school campuses, neighborhood's catchment areas, assisted care living facilities, hospital wastewater treatment plants, and commercial building centers (Betancourt et al., 2021; Hasan et al., 2021; Karthikeyan et al., 2021; Scott et al., 2021; Yaniv et al., 2021).

In South Africa, the first investigations on application of WBE to detect SARS-CoV-2 was reported from Cape Town. They showed that SARS-CoV-2 viral RNA can be detected from wastewater influent samples and can be used as a surveillance tool in South Africa. This highlighted the potential of WBE as an early warning system (EWS) for monitoring the trends of COVID-19 infections in South Africa, which could facilitate rapid identification of hotspots for evidence-informed interventions (Johnson et al., 2021). Other studies conducted in South Africa were reported from KwaZulu-Natal province, where they applied WBE to monitor COVID-19 infection dynamics in the KwaZulu-Natal province of South Africa (Pillay et al., 2021). Additionally, one study reported surveillance carried out by the South African Collaborative COVID-19 Surveillance System (SACCESS) network, established in 2021. This surveillance was conducted in all nine provinces in South Africa,

across 87 wastewater treatment plants (WWTPs). They found that investing in WBE for SARS-CoV-2 surveillance in low- and middle-income settings (LMICs) can be beneficial as an early warning tool (Iwu-Jaja et al., 2023). As of November 2023, wastewater-based surveillance systems have been set up globally for monitoring SARS-CoV-2 occurrences and have also been adapted for detection of other enteric and respiratory pathogens post-pandemic (Singer et al., 2023).

As the pandemic progressed, various variants were detected from clinical specimens through whole genome sequencing (WGS), which was the gold standard approach (Novelli et al., 2021; Tegally et al., 2021; Walker et al., 2021; Vianna et al., 2022). However, since WGS is expensive, labour intensive and time consuming, other cost-effective approaches were sought for variant monitoring, for application in resource-limited settings. Although several methods were applied to infer the presence of VOCs, in both clinical and wastewater specimen, most studies reported on allele-specific genotyping (ASG) qRT-PCR, since it has a shorter turnaround time (TAT). Utilization of ASG methods for variant monitoring would also be beneficial in regions with few reference laboratories with limited sequencing capabilities (Brito-mutunayagam et al., 2022). Some studies designed single nucleotide polymorphisms (SNP) identification pipeline to identify genetic variation using sequenced SARS- CoV-2 samples (Harper et al., 2021; Lee et al., 2021; Takemae et al., 2022). Other studies developed a digital droplet polymerase chain reaction (ddPCR) assay for rapid identification of circulating variants of concern/interest (VOC/ VOI) using variant-specific mutation combinations in the Spike gene, as well as estimating the relative abundance (Yu et al., 2022; Pernet et al., 2023). Application of the TaqMan SARS-CoV-2 mutation panel molecular genotyping assay was reported. This genotyping assay was used for detection and identification of common variants through specific RT-PCR assays which target SNPs. One study reported from Indiana USA reported the applicability of this assay for surveillance and epidemic control and prevention, since it was able to detect SARS-CoV-2 Alpha, Beta, Delta, Iota, Gamma, Zeta, Kappa, and Epsilon variants in clinical specimen (Neopane et al., 2021). Additionally, some studies utilized qPCR-based melting curve analysis to detect Omicron mutations and discriminate omicron lineages in clinical specimen (Kong et al., 2023).

1.2.13) Limitations Affecting the Implementation of WBE

The pandemic highlighted the unique advantages of utilizing WBS for monitoring trends in SARS-CoV-2 occurrence, which could serve as an early warning system. However, with all its advantages, WBS has peculiar limitations which need to be addressed for better maximization of this tool. This is relevant for better integration of results from WBS with clinical surveillance to improve public health strategies. Some of these limitations are discussed here.

First, sampling for WBS has some challenges in that, the results obtained are “representative” of a particular catchment area / study population. Thus, sampling challenges in both rural and urban settings need to be considered when designing the surveillance program. Examples of these challenges include access to sewage facilities (particularly in rural settings), sampling methods to be used (grab or composite), the decay rate of viruses in different sewer systems, factors that may dilute the concentration of the virus in wastewater (e.g. precipitation), as well as disinfectants and detergents co-entering the sewage network when sampling from hospitals. Comparatively, sampling for clinical surveillance of SARS-CoV-2 is relatively straight forward (Polo et al., 2020).

Once the samples have arrived at the laboratory, the next recommended step to improve results obtained through WBS is optimization of the virus recovery and concentration methods before viral quantification. This is because virus detection in environmental samples requires viral concentration to improve detection limits. Throughout the COVID-19 pandemic, several studies have reported using either one or a combination of processes like ultracentrifugation, filtrations, and polyethylene glycol (PEG) precipitation (Masachessi et al., 2022; Sharif et al., 2021).

Viral quantification is the principal goal of WBE, since high viral concentrations can indicate potential onset of future disease outbreaks. During the pandemic, studies on WBS utilized previously published primer/probe sets targeting SARS-CoV-2 Nucleocapsid, Envelope, and RdRp genes for viral quantification in wastewater (Corman et al., 2020). Although reverse transcription-PCR (RT-PCR) and RT-real time PCR (RT-qPCR) is the gold standard to obtain qualitative and quantitative data, the extraction method used may contain diverse PCR inhibitors which can interfere with the PCR

reaction. Thus, future efforts to improve extraction and purification of nucleic acids from wastewater should be investigated to optimize viral quantification. Utilizing a technique such as digital PCR (dPCR) has been proposed to improve qPCR sensitivity and minimize the effect of PCR inhibitors in complex matrices such as wastewater.

The shedding rate of SARS-CoV-2 and the levels present in wastewater is another potential challenge that may influence the goal of WBE to serve as an early warning system. The shedding rate refers to the rate of virus release from the body; knowing this rate is critical linking wastewater data back to human populations. Factors such as viremia, the duration, severity, age and the stage of the disease can impact the shedding rate of viruses in feces (Chen and Li, 2020). Throughout the pandemic, several studies reported the concentration of SARS-CoV-2 from wastewater, which greatly depends on the number of infected people in the community, and the rate at which the virus is shed in the feces of infected individuals. These viral concentrations are used to establish correlations with the reported COVID-19 clinical cases, to understand the extent of the outbreak in the population.

Another critical step in application of WBS is population normalization, which refers to estimation of the population contributing to wastewater samples. This involves using census and biomarker data, which should be collected independently, and integrated with common units. Endogenous or exogenous human biomarkers have been proposed for estimating the serviced population in an area using statistical modelling. Biomarker quantification can provide context to measured viral concentrations and ensure that differences in viral loads is not attributed to changes in population. Although substances such as creatinine, cholesterol, coprostanol, nicotine, cortisol, androstenedione, and the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA) have been proposed as population biomarkers, some concerns such as differing consumption or disposal habits, stability, and sorption to particulate matter have been raised since they could lead to high uncertainties (Rico et al., 2017). Thus, human nucleic acid matter has been proposed, due to its stability, limited affinity to other species in wastewater, constant excretion by humans, and the possibility of being quantifiable using the same pipelines and platforms as the viral nucleic acid of interest. Examples of these viral indicators of human activity

are pepper-mild mottle virus, crAssphage, and adenovirus, which have been reported as a population biomarker in several WBS studies throughout the pandemic. WBE also uses census data. However, contrary to biomarkers, census data may understate the population. To facilitate inter-city comparisons and ensure that increases in wastewater viral concentration does not coincide with increase in the population within the catchment, population normalization is essential. Thus, WBE data is potentially impacted by population changes which may arise due to commuting or touristic activities. In large populations, these dynamics may not have much of an effect on biomarker levels, but in smaller populations, they may increase uncertainty (Chen et al., 2014; Ort et al., 2014). It is also worth noting that census data may be readily available in higher income nations, which have well established sewage systems as well as frequently updated census data. This may not be the case in low- and middle-income countries (LMICs), where sewage facilities may be absent in some areas or some homes may not be connected to sewage systems, as well as lack of recent census data (Street et al., 2020), may impact on surveillance efforts.

Finally, the aspect of ethical considerations has also been highlighted when applying WBE. This is because, although WBE does not collect data on individuals, expanding its application to include surveillance of other infectious diseases and outbreaks could pose ethical considerations challenges. When monitoring infectious diseases, public health authorities usually implement lockdown measures and/or sampling of specific populations to prevent further spread. However, these actions may lead to stigmatizing practices. This was observed during the Ebola outbreak in West Africa, where stigmatization associated with infected individuals, alongside distrust in health services and treatment centers, led to increased spread of the virus (Polo et al., 2020). As such, there were attempts to conceal cases, administer care at home for infected individuals, which led to increased transmission among family members and members of the community (Shultz et al., 2016). Similar ethical issues have also been observed during the SARS-CoV and influenza outbreaks; during the early phases of the COVID-19 pandemic, social stigma was associated with those infected. This led the WHO to publish the first comprehensive international ethics guideline on public health surveillance in 2017. Their report highlighted the need for appropriate adaptation of ethical guidelines for different social,

economic and epidemiological circumstances (WHO, 2017). Thus, care must be taken in disease reporting, to reduce the media's misinterpretation of published findings.

Taking all these into consideration, although WBE has proven to be effective as an early warning tool, it should be used to complement clinical surveillance, rather than a standalone surveillance method, since it requires robust systems to combine and effectively interpret data from multiple sources.

1.3) Main Aim

Determining the distribution and molecular epidemiology of HCoVs in Africa prior to the SARS-CoV-2 outbreak through a systematic literature review; monitoring SARS-CoV-2 trends and VOCs circulation through wastewater-based surveillance; and utilize wastewater to describe the molecular epidemiology of SARS-CoV-2 in Vhembe and Mopani districts, in the Limpopo Province of South Africa

1.4) Objectives

1.4.1) To systematically review HCoVs distribution and molecular epidemiology in Africa prior to the SARS-CoV-2 outbreak.

1.4.2) To establish a surveillance system for real-time tracking of SARS-CoV-2 trends and description of the VOCs occurring over an 18-month period, in the Limpopo Province.

1.4.3) To describe the molecular epidemiology of SARS-CoV-2 and document the respiratory viruses occurring in the Vhembe and Mopani districts, in Limpopo, South Africa.

1.5) REFERNCES

Ahmad, S.U. et al. (2022) 'A comprehensive genomic study, mutation screening, phylogenetic and statistical analysis of SARS-CoV-2 and its variant omicron among different countries', Journal of Infection and Public Health, 15(8), pp. 878–891. https://doi.org/10.1016/j.jiph.2022.07.002.

Ahmed, W. et al. (2020) 'First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community', Science of the Total Environment, 728, pp. 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Alanagreh, L. et al. (2020) 'The human coronavirus disease covid-19: Its origin, characteristics, and insights into potential drugs and its mechanisms', Pathogens, 9(5), pp. 331. https://doi.org/10.3390/pathogens9050331

Alsafar, H. et al. (2022) 'Genomic epidemiology and emergence of SARS-CoV-2 variants of concern in the United Arab Emirates', Scientific Reports, 12(1), pp. 14669. https://doi.org/10.1038/s41598-022-16967-w

Alsobaie, S. (2021) 'Understanding the Molecular Biology of SARS-CoV- 2 and the COVID-19 Pandemic: A Review', Infection and Drug Resistance, 14, pp. 2259–2268. https://doi.org/10.2147/IDR.S306441

Alquraan, L. et al. (2023) 'Mutations of SARS-CoV-2 and their impact on disease diagnosis and severity', Informatics in Medicine Unlocked, 39(February), p. 101256. Available at: https://doi.org/10.1016/j.imu.2023.101256

Arunachalam, J.P. et al. (2021) 'SARS-CoV-2: The Road Less Traveled: The Respiratory Mucosa to the Brain', ACS Omega, 6(10), pp. 7068–7072. Available at: https://doi.org/10.1021/acsomega.1c00030

Baldovin, T. et al. (2021) 'SARS-CoV-2 RNA detection and persistence in wastewater samples: An experimental network for COVID-19 environmental surveillance in Padua,

Veneto Region (NE Italy)', Science of the total environment, 760, pp. 143329. https://doi.org/10.1016/j.scitotenv.2020.143329

Behboudi, E. et al. (2024) 'SARS-CoV-2 mechanisms of cell tropism in various organs considering host factors', Heliyon, 10(4), pp. e26577. https://doi.org/10.1016/j.heliyon.2024.e26577

Bosch, A. et al. (2008) 'New tools for the study and direct surveillance of viral pathogens in water', Current Opinion in Biotechnology, 19(3), pp. 295–301. https://doi.org/10.1016/j.copbio.2008.04.006

Brito-mutunayagam, S. et al. (2022) 'Rapid detection of SARS-CoV-2 variants using allele-specific PCR', Journal of Virological Methods, 303(114497), pp. 1–4. https://doi.org/10.1016/j.jviromet.2022.114497

Brouwer, A.F. et al. (2018) 'Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data', Proceedings of the National Academy of Sciences of the United States of America, 115(45), pp. E10625–E10633. https://doi.org/10.1073/pnas.1808798115

Buscarini E. et al. (2020) 'GI symptoms as early signs of COVID-19 in hospitalised Italian patients', Gut, 69(8), pp. 1547–1548. https://doi.org/10.1136/gutjnl-2020-321434

Cevik, M. et al. (2020) 'Virology, transmission, and pathogenesis of SARS-CoV-2', BMJ (Clinical Research Ed.), 371, pp. m3862. https://doi.org/10.1136/bmj.m3862

Chau, C.H. et al. (2020) 'COVID-19 Clinical Diagnostics and Testing Technology', Pharmacotherapy, 40(8), pp. 857–868. https://doi.org/10.1002/phar.2439

Chen, Y. et al. (2020) 'Emerging coronaviruses: Genome structure, replication, and pathogenesis', Journal of Medical Virology, 92(4), pp. 418–423. https://doi.org/10.1002/jmv.25681

Choi, P.M. et al. (2018) 'Wastewater-based epidemiology biomarkers: Past, present and future', TrAC - Trends in Analytical Chemistry. pp. 453–469. https://doi.org/10.1016/j.trac.2018.06.004

Choi, P.M. et al. (2019) 'Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology', Proceedings of the National Academy of Sciences of the United States of America, 116(43), pp. 21864–21873. https://doi.org/10.1073/pnas.1910242116

Chrysostomou, A.C. et al. (2023) 'Article Waves and Emergence of the Deltacton Variant : Genomic Epidemiology of the SARS-CoV-2 Epidemic in Cyprus'. Viruses, 15(1), pp. 108. https://doi.org/10.3390/v15010108

Corman, V.M. et al. (2020) 'Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR' Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles, 25(3):2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

Coronaviridae Study group. (2020) 'The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.', Nature Microbiology, 5(4), pp. 536 – 544. https://doi.org/10.1038/s41564-020-0695-z

Cosic, I. et al. (2020) 'RRM prediction of erythrocyte band3 protein as alternative receptor for SARS-CoV-2 virus', Applied Sciences (Switzerland), 10(11), pp. 4053. https://doi.org/10.3390/app10114053

Dai, B. et al. (2024) 'Update on Omicron variant and its threat to vulnerable populations', Public Health in Practice, 7(100), pp. 100494. https://doi.org/10.1016/j.puhip.2024.100494

Dubey, A. et al. (2022) 'Emerging SARS-CoV-2 Variants: Genetic Variability and Clinical Implications', Current Microbiology, 79(1), pp. 1–18. https://doi.org/10.1007/s00284-021-02724-1

Esper, F.P. et al. (2023) 'Alpha to Omicron: Disease Severity and Clinical Outcomes of Major SARS-CoV-2 Variants', Journal of Infectious Diseases, 227(3), pp. 344–352. https://doi.org/10.1093/infdis/jiac411

Florensa, D. et al. (2022) 'Severity of COVID-19 cases in the months of predominance of the Alpha and Delta variants', Scientific Reports, 12(1), pp. 10–15. https://doi.org/10.1038/s41598-022-19125-4

Forni, D. et al. (2017) 'Molecular Evolution of Human Coronavirus Genomes', Trends in Microbiology, 25(1), pp. 35–48. https://doi.org/10.1016/j.tim.2016.09.001

Goławski, M. et al. (2022) 'The Reassessed Potential of SARS-CoV-2 Attenuation for COVID-19 Vaccine Development—A Systematic Review', Viruses, 14(5), pp. 1–28. https://doi.org/10.3390/v14050991

Gostin, L.O. and Gronvall, G.K. (2023) 'The Origins of Covid-19 — Why It Matters (and Why It Doesn't)', The New England Journal of Medicine, 388(25), pp. 2305–2308. https://doi.org/10.1056/NEJMp2305081

Guo, H. et al. (2021) 'Identification of a novel lineage bat SARS-related coronaviruses that use bat ACE2 receptor', Emerging Microbes and Infections, 10(1), pp. 1507–1514. https://doi.org/10.1080/22221751.2021.1956373

Haagmans, B.L. et al. (2014) 'Middle East respiratory syndrome coronavirus in dromedary camels: An outbreak investigation', The Lancet Infectious Diseases, 14(2), pp. 140–145. https://doi.org/10.1016/S1473-3099(13)70690-X

Han, Y. et al. (2019) 'Identification of diverse bat alphacoronaviruses and betacoronaviruses in china provides new insights into the evolution and origin of coronavirus-related diseases', Frontiers in Microbiology, 14(10), 1900. https://doi.org/10.3389/fmicb.2019.01900

Harper, H. et al. (2021) 'Detecting SARS-CoV-2 variants with SNP genotyping', PLoS ONE, 16(2), pp. e0243185. https://doi.org/10.1371/journal.pone.0243185

Haque, S.M. et al. (2020) 'A comprehensive review about SARS-CoV-2', Future Virology, 15(9), pp. 625–648. https://doi.org/10.2217/fvl-2020-0124

Hart, O.E. and Halden, R. U. (2020) 'Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges', Science of the Total Environment, 730, pp. 138875. https://doi.org/10.1016/j.scitotenv.2020.138875

He X. et al. (2020) 'Temporal dynamics in viral shedding and transmissibility of COVID-19', Nature Medicine, 26(5), pp. 672–675. https://doi.org/10.1038/s41591-020-0869-5

Heijnen, L. et al. (2021) 'Droplet Digital RT-PCR to detect SARS-CoV-2 variants of concern in wastewater' Science of the Total Environment, 799, pp. 149456. https://doi.org/10.1101/2021.03.25.21254324

Hellmér, M. et al. (2014) 'Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks', Applied and Environmental Microbiology, 80(21), pp. 6771–6781. https://doi.org/10.1128/AEM.01981-14

Heneghan, C.J. et al. (2021) 'SARS-CoV-2 and the role of orofecal transmission: a systematic review', F1000Research, 10, pp. 231. https://doi.org/10.12688/f1000research.51592.2

Hill, V. et al. (2022) 'The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK', Virus Evolution, 8(2), pp. 1 – 13. https://doi.org/10.1093/ve/veac080

Hillary, L.S. et al. (2021) 'Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK', Water research, 200, pp. 117214. https://doi.org/10.1016/j.watres.2021.117214

Van Der Hoek, L. et al. (2006) 'Human coronavirus NL63, a new respiratory virus', FEMS Microbiology Reviews, 30(5), pp. 760–773. https://doi.org/10.1111/j.1574-6976.2006.00032.x

Hovi, T. et al. (2012) 'Role of environmental poliovirus surveillance in global polio eradication and beyond', Epidemiology and Infection, 140(1), pp. 1–13. https://doi.org/10.1017/S095026881000316X

Hu, B. et al. (2021) 'Characteristics of SARS-CoV-2 and COVID-19', Nature Reviews Microbiology, 19(3), pp. 141–154. https://doi.org/10.1038/s41579-020-00459-7

Huang, C. et al. (2020) 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', The Lancet, 395(10223), pp. 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Iwu-Jaja, C. et al. (2023) 'The role of wastewater-based epidemiology for SARS-CoV-2 in developing countries: Cumulative evidence from South Africa supports sentinel site

surveillance to guide public health decision-making', Science of the Total Environment, 903. https://doi.org/10.1016/j.scitotenv.2023.165817

Jaimes, J.A. et al. (2020) 'Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop', Journal of Molecular Biology, 432(10), pp. 3309–3325. https://doi.org/10.1016/j.jmb.2020.04.009

Johnson, R. et al. (2021) 'Qualitative and quantitative detection of SARS-CoV-2 RNA in untreated wastewater in Western Cape Province, South Africa', South African Medical Journal, 111(3), pp. 198–202. https://doi.org/10.7196/SAMJ.2021.V111I3.15154

Johnson and Johnson. 'Johnson & Johnson Single-Shot COVID-19 Vaccine Phase 3 Data', New England Journal of Medicine. https://www.jnj.com/media-center/press-releases/johnson-johnson-single-shot-covid-19-vaccine-phase-3-data-published-in-new-england-journal-of-medicine.

Johnson, R. et al. (2021) 'Qualitative and quantitative detection of SARS-CoV-2 RNA in untreated wastewater in Western Cape Province, South Africa', South African Medical Journal, 111(3), pp. 198–202. https://doi.org/10.7196/SAMJ.2021.V111I3.15154

Jones, D.L. et al. (2020) 'Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19', Science of the Total Environment, 749, p. 141364. https://doi.org/10.1016/j.scitotenv.2020.141364

Kaur, G. et al. (2020) 'SARS-CoV-2 COVID-19 susceptibility and lung inflammatory storm by smoking and vaping', Journal of Inflammation (United Kingdom), 17(1), pp. 1–8. https://doi.org/10.1186/s12950-020-00250-8

Kong, X. et al. (2023) 'Discrimination of SARS-CoV-2 omicron variant and its lineages by rapid detection of immune-escape mutations in spike protein RBD using asymmetric PCR-based melting curve analysis', Virology Journal, 20(1), pp. 1–15. https://doi.org/10.1186/s12985-023-02137-5

Krishnamoorthy, S. et al. (2020) 'SARS-CoV, MERS-CoV, and 2019-nCoV viruses: an overview of origin, evolution, and genetic variations', Virus Disease, 31(4), pp. 411–423. https://doi.org/10.1007/s13337-020-00632-9

Kumar, R. et al. (2023) 'Understanding Mutations in Human SARS-CoV-2 Spike Glycoprotein: A Systematic Review & Meta-Analysis', Viruses, 15(4). https://doi.org/10.3390/v15040856

Lai, C.K.C. and Lam, W. (2021) 'Laboratory testing for the diagnosis of COVID-19', Biochemical and Biophysical Research Communications, 538, pp. 226–230. https://doi.org/10.1016/j.bbrc.2020.10.069

Lau, S.K.P. et al. (2006) 'Coronavirus HKU1 and other coronavirus infections in Hong Kong', Journal of Clinical Microbiology, 44(6), pp. 2063–2071. https://doi.org/10.1128/JCM.02614-05

Lee, W.L. et al. (2021) 'Quantitative SARS-CoV-2 Alpha Variant B.1.1.7 Tracking in Wastewater by Allele-Specific RT-qPCR', Environmental Science and Technology Letters, 8(8), pp. 675–682. https://doi.org/10.1021/acs.estlett.1c00375

Li, J. et al. (2024) 'Development of multiplex allele-specific RT-qPCR assays for differentiation of SARS-CoV-2 Omicron subvariants', Applied Microbiology and Biotechnology, 108(1), pp. 1–17. https://doi.org/10.1007/s00253-023-12941-2

Li, P. et al. (2022) 'SARS-CoV-2 Omicron variant is highly sensitive to molnupiravir, nirmatrelvir, and the combination', Cell Research, 32(3), pp. 322–324. https://doi.org/10.1038/s41422-022-00618-w

Li, Q. et al. (2021) 'SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape', Cell, 184(9), pp. 2362-2371.e9. https://doi.org/10.1016/j.cell.2021.02.042

Lichtenberger, L.M. and Vijayan, K.V. (2021) 'Is COVID-19–induced platelet activation a cause of concern for patients with cancer?', Cancer Research, 81(5), pp. 1209–1211. https://doi.org/10.1158/0008-5472.CAN-20-3691

Liu, D.X. et al. (2021) 'Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae)', in Encyclopedia of Virology, pp. 428–440. https://doi.org/10.1016/b978-0-12-809633-8.21501-x

Liu, Y. et al. (2022) 'The N501Y spike substitution enhances SARS-CoV-2 infection and transmission', Nature, 602(7896), pp. 294–299. https://doi.org/10.1038/s41586-021-04245-0

Markov, P.V. et al. (2023) 'The evolution of SARS-CoV-2', Nature Reviews Microbiology, 21(6), 361 – 371. https://doi.org/10.1038/s41579-023-00878-2

Marques, F. et al. (2021) 'Acute kidney disease and mortality in acute kidney injury patients with covid-19', Journal of Clinical Medicine, 10(19). https://doi.org/10.3390/jcm10194599

Martinez-Rojas, M.A. et al. (2020) 'Is the kidney a target of SARS-CoV-2?', American Journal of Physiology – Renal Physiology, 318(6), pp. F1454–F1462. https://doi.org/10.1152/AJPRENAL.00160.2020

Masachessi, G. et al. (2022) 'Wastewater based epidemiology as a silent sentinel of the trend of SARS-CoV-2 circulation in the community in central Argentina', Water Research, 219, pp. 118541. https://doi.org/10.1016/j.watres.2022.118541

Mascola, J.R. et al. (2021) 'SARS-CoV-2 Viral Variants - Tackling a Moving Target', Journal of the American Medical Association, 325(13), pp. 1261–1262. https://doi.org/10.1001/jama.2021.2088

Mayne, E.S. et al. (2020) 'The role of serological testing in the SARS-CoV-2 outbreak', South African Medical Journal, 110(9), pp. 842–845. https://doi.org/10.7196/SAMJ.2020.v110i9.15098

McMillen, T. et al. (2022) 'The spike gene target failure (SGTF) genomic signature is highly accurate for the identification of Alpha and Omicron SARS-CoV-2 variants', Scientific Reports, 12(1), pp. 1–8. https://doi.org/10.1038/s41598-022-21564-y

Medema, G. et al. (2020) 'Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in the

Netherlands', Environmental Science and Technology Letters, 7(7), pp. 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Meinhardt, J. et al. (2021) 'Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19', Nature Neuroscience, 24(2), pp. 168–175. https://doi.org/10.1038/s41593-020-00758-5

Meyerowitz, E.A. et al. (2021) 'Transmission of SARS-CoV-2: A Review of Viral, Host, and Environmental Factors', Annals of internal medicine, 174(1), pp. 69–79. https://doi.org/10.7326/M20-5008

Morvan, M. et al. (2022) 'An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence.', Nature communications, 13(1), p. 4313. https://doi.org/10.1038/s41467-022-31753-y

Mohsin, M.D. and Mahmud, S. (2022) 'Omicron SARS-CoV-2 variant of concern A review on its transmissibility, immune evasion, reinfection, and severity', Medicine (Baltimore), 101(19), p. 29165. https://doi.org/10.1097/MD.0000000000029165

Mulabbi, E.N., Tweyongyere, R. and Byarugaba, D. K. (2021) 'The history of the emergence and transmission of human Coronaviruses', Onderstepoort Journal of Veterinary Research, 88(1), pp. e1-e8. https://doi.org/10.4102/ojvr.v88i1.1872

Murakami, S. et al. (2020) 'Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan', Emerging Infectious Diseases, 26(12), pp. 3025–3029. https://doi.org/10.3201/eid2612.203386

Nemudryi, A. et al. (2020) 'Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater.', Cell reports. Medicine, 1(6), p. 100098. https://doi.org/10.1016/j.xcrm.2020.100098

Neopane, P. et al. (2021) 'SARS-CoV-2 variants detection using TaqMan SARS-CoV-2 mutation panel molecular genotyping assays', Infection and Drug Resistance, 14, pp. 4471–4479. https://doi.org/10.2147/IDR.S335583

Pekar, J.E. et al. (2023) 'The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2', Science, 377(6609), pp. 960–966. https://doi.org/10.1126/science.abp8337

Pernet, O. et al. (2023) 'SARS-CoV-2 viral variants can rapidly be identified for clinical decision making and population surveillance using a high-throughput digital droplet PCR assay', Scientific Reports, 13(1), pp. 7612. https://doi.org/10.1038/s41598-023-34188-7

Perrotta, F. et al. (2020) 'Severe respiratory SARS-CoV2 infection: Does ACE2 receptor matter?', Respiratory Medicine, 168, p. 105996. https://doi.org/10.1016/j.rmed.2020.105996

Pillay, L. et al. (2021) 'Monitoring changes in COVID-19 infection using wastewater-based epidemiology: A South African perspective', Science of the Total Environment, 786, pp. 147273. https://doi.org/10.1016/j.scitotenv.2021.147273

Platto, S. et al. (2021) 'History of the COVID-19 pandemic: Origin, explosion, worldwide spreading', Biochemical and Biophysical Research Communications, 538, pp. 14–23. https://doi.org/10.1016/j.bbrc.2020.10.087

Poland, G.A. (2020) 'Another coronavirus, another epidemic, another warning', Vaccine, 38(10), pp. v–vi. https://doi.org/10.1016/j.vaccine.2020.02.039

Polo, D. et al. (2020) 'Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction', Water Research, 186, pp. 116404. https://doi.org/10.1016/j.watres.2020.116404

Prasek, S.M. et al. (2022) 'Population level SARS-CoV-2 fecal shedding rates determined via wastewater-based epidemiology', The Science of the total environment, 838(Pt 4), pp. 156535. https://doi.org/10.1016/j.scitotenv.2022.156535

Rahimi, A. et al. (2021) 'Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection', Genomics, 113(1 Pt 2), pp. 1221–1232. https://doi.org/10.1016/j.ygeno.2020.09.059

Rahman, M.M. et al. (2021) 'A comprehensive review on COVID-19 vaccines: development, effectiveness, adverse effects, distribution and challenges', Virus Disease, 33(1), pp. 1–22. https://doi.org/10.1007/s13337-022-00755-1

Randazzo, W. et al. (2020) 'SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area', Water Research, 181. https://doi.org/10.1016/j.watres.2020.115942

Rastogi, M. et al. (2020) 'SARS coronavirus 2: from genome to infectome', Respiratory Research, 21(1), pp. 318. https://doi.org/10.1186/s12931-020-01581-z

Ratajczak, M.Z. et al. (2021) 'SARS-CoV-2 Entry Receptor ACE2 Is Expressed on Very Small CD45− Precursors of Hematopoietic and Endothelial Cells and in Response to Virus Spike Protein Activates the Nlrp3 Inflammasome', Stem Cell Reviews and Reports, 17(1), pp. 266–277. https://doi.org/10.1007/s12015-020-10010-z

Rauf, M.A. et al. (2021) 'Nano-therapeutic strategies to target coronavirus', View, 2(3), pp. 20200155. https://doi.org/10.1002/VIW.20200155

La Rosa, G. et al. (2021) 'Rapid screening for SARS-CoV-2 variants of concern in clinical and environmental samples using nested RT-PCR assays targeting key mutations of the spike protein', Water Research, 197, pp. 117104. https://doi.org/10.1016/j.watres.2021.117104

Ruiz-Aravena, M. et al. (2022) Ecology, evolution and spillover of coronaviruses from bats, Nature Reviews Microbiology. 20(5), pp. 299 – 314. https://doi.org/10.1038/s41579-021-00652-2

Romano, M. et al. (2020) 'A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping', Cells, 9(5), pp. 1267. https://doi.org/10.3390/cells9051267

Sadoff, J. et al. (2021) 'Interim results of a phase 1–2a Trial of Ad26.COV2.S COVID-19 vaccine', New England Journal of Medicine, 384, pp. 1824–35. https://doi.org/org/10.1056/NEJMoa2034201

Sahin, U. et al. (2021) 'BNT162b2 vaccine induces neutralizing antibodies and poly-specific T cells in humans', Nature, 595(7868), pp. 572–527. https://doi.org/10.1038/s41586-021-03653-6

Santiago, G.A. et al. (2022) 'Genomic surveillance of SARS-CoV-2 in Puerto Rico enabled early detection and tracking of variants', Communications Medicine (London), 2, pp. 100. https://doi.org/10.1038/s43856-022-00168-7

Saththasivam, J. et al. (2021) 'COVID-19 (SARS-CoV-2) outbreak monitoring using wastewater-based epidemiology in Qatar', Science of the Total Environment, 774, pp. 145608. https://doi.org/10.1016/j.scitotenv.2021.145608

Shao, N. et al. (2022) 'Molecular evolution of human coronavirus -NL63, -229E, -HKU1 and -OC43 in hospitalized children in China', Frontiers in Microbiology, 13, pp. 1023847. https://doi.org/10.3389/fmicb.2022.1023847

Sharif, S. et al. (2021) 'Detection of SARs-CoV-2 in wastewater using the existing environmental surveillance network: A potential supplementary system for monitoring COVID-19 transmission', PLoS ONE, 16(6), pp. e0249568. https://doi.org/10.1371/journal.pone.0249568

Shaw, J.A. et al. (2021) 'Higher SARS-CoV-2 seroprevalence in workers with lower socioeconomic status in Cape Town, South Africa', PLoS ONE, 16(2), pp. e0247852. https://doi.org/10.1371/journal.pone.0247852

Shu, Y. and Mccauley, J. (2017) 'GISAID: Global initiative on sharing all influenza data – from vision to reality', Eurosurveillance, 22(13), pp. 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Singer, A.C. et al. (2023) 'A world of wastewater-based epidemiology', Nature Water, 1, pp. 408–415. https://doi.org/10.1038/s44221-023-00083-8

Smithgall, M.C. et al. (2021) 'Types of assays for SARS-COV-2 testing: A review', Lab Medicine. Oxford University Press, 51 (5), pp. e59–e65. https://doi.org/10.1093/LABMED/LMAA039

Street, R. et al. (2020) 'Wastewater surveillance for COVID-19: An African perspective', 743(140719), pp. 1–4. https://doi.org/10.1016/j.scitotenv.2020.14

Subramoney, K. et al. (2022) 'Identification of SARS-CoV-2 Omicron variant using spike gene target failure and genotyping assays, Gauteng, South Africa, 2021', Journal of Medical Virology, 94(8), pp. 3676–3684. https://doi.org/10.1002/jmv.27797

Suwannakarn, K. et al. (2014) 'Prevalence and genetic characterization of human coronaviruses in southern Thailand from July 2009 to January 2011', The Southeast Asian journal of tropical medicine and public health, 45(2), pp. 326–336.

Tang, A. et al. (2020) 'Detection of novel coronavirus by RT-PCR in stool specimen from asymptomatic child, China', Emerging Infectious Diseases, 26 (6): pp. 1337 – 1339. https://doi.org/10.3201/eid2606.200301

Takemae, N. et al. (2022) 'Development of new SNP genotyping assays to discriminate the Omicron variant of SARS-CoV-2', Japan journal of Infectious Diseases, 75(4), pp. 411 – 414. https://doi.org/10.7883/yoken.JJID.2022.007

Tang, G., Liu, Z. and Chen, D. (2022) 'Human coronaviruses: Origin, host and receptor', Journal of Clinical Virology, 155, pp. 105246. https://doi.org/10.1016/j.jcv.2022.105246

The Lancet Microbe (2023) 'Searching for SARS-CoV-2 origins: confidence versus evidence', The Lancet Microbe, 4(4), pp. e200. https://doi.org/10.1016/S2666-5247(23)00074-5

Tian, F. et al. (2021) 'N501Y mutation of spike protein in sars-cov-2 strengthens its binding to receptor ACE2', eLife, 10, e69091. https://doi.org/10.7554/eLife.69091

Tian, L. et al. (2022) 'Molnupiravir and Its Antiviral Activity Against COVID-19', Frontiers in Immunology, 13, pp. 855496. https://doi.org/10.3389/fimmu.2022.855496

Tiwari, S. and Dhole, T. N. (2018) 'Assessment of enteroviruses from sewage water and clinical samples during eradication phase of polio in North India', Virology Journal, 15(1). https://doi.org/ 10.1186/s12985-018-1075-7

Trottier, J. et al. (2020) 'Post-lockdown detection of SARS-CoV-2 RNA in the wastewater of Montpellier, France', One Health, 10, pp. 100157. https://doi.org/10.1016/j.onehlt.2020.100157

Uwamino, Y. et al. (2022) 'The effect of the E484K mutation of SARS-CoV-2 on the neutralizing activity of antibodies from BNT162b2 vaccinated individuals', Vaccine, 40(13), pp. 1928–1931. https://doi.org/10.1016/j.vaccine.2022.02.047

Vangeel, L. et al. (2022) 'Remdesivir, Molnupiravir and Nirmatrelvir remain active against SARS-CoV-2 Omicron and other variants of concern', Antiviral Research, 198(January), pp. 10–12. https://doi.org/10.1016/j.antiviral.2022.105252

Varshney, R.K. (2023) 'State of the Globe: Navigating the Impact of SARS-CoV-2 Mutations on COVID-19 Testing', Journal of Global Infectious Diseases, 15(2), pp. 41–42. https://doi.org/10.4103/jgid.jgid_90_23

Veluswamy, P. et al. (2021) 'The sars-cov-2/receptor axis in heart and blood vessels: A crisp update on covid-19 disease with cardiovascular complications', Viruses, 13(7). https://doi.org/10.3390/v13071346

V'kovski, P. et al. (2020) 'Coronavirus biology and replication: implications for SARS-CoV-2', Nature Reviews Microbiology, 19 (3), pp. 155 – 170. https://doi.org/10.1038/s41579-020-00468-6

Wacharapluesadee, S. et al. (2021) 'Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia', Nature Communications, 12, 972. https://doi.org/10.1038/s41467-021-21240-1

Wang, R. et al. (2020) 'Mutations on COVID-19 diagnostic targets', Genomics, 112(6), pp. 5204–5213. https://doi.org/10.1016/j.ygeno.2020.09.028

Wang, R. et al. (2021) 'Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries', Genomics, 113(4), pp. 2158–2170. https://doi.org/10.1016/j.ygeno.2021.05.006

Weatherbee, B.A.T. et al. (2020) 'Expression of SARS-CoV-2 receptor ACE2 and the protease TMPRSS2 suggests susceptibility of the human embryo in the first trimester', Open Biology, 10(8), pp. 4–7. https://doi.org/10.1098/rsob.200162

Wei, C. et al. (2020) 'A super-spreader of SARS-CoV-2 in incubation period among health-care workers', Respiratory Research, 21(1), pp. 327. https://doi.org/10.1186/s12931-020-01592-w

Westhaus, S. et al. (2021) 'Detection of SARS-CoV-2 in raw and treated wastewater in Germany - Suitability for COVID-19 surveillance and potential transmission risks', The Science of the total environment, 751, pp. 141750. https://doi.org/10.1016/j.scitotenv.2020.141750

Wiley, J.S. et al. (1962) 'Enterovirus in Sewage During a Poliomyelitis Epidemic', Journal (Water Pollution Control Federation), 34(2), pp. 168–178. http://www.jstor.org/stable/25034581

Wilder-Smith, A. and Osman, S. (2020) 'Public health emergencies of international concern: A historic overview', Journal of Travel Medicine, 27(8), taaa227. https://doi.org/10.1093/jtm/taaa227

Wölfel, R. et al. (2020) 'Virological assessment of hospitalized cases of coronavirus disease 2019', Nature, 581(7809), pp. 465 – 469. https://doi.org/10.1038/s41586-020-2196-x

Woo, P.C.Y. et al. (1999) 'Coronavirus Genomics and Bioinformatics Analysis', Viruses, 2(8), pp. 1804–1820. https://doi.org/10.3390/v2081803

Woo, P.C.Y. et al. (2005) 'Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia', Journal of Virology, 79(2), pp. 884–895. https://doi.org/10.1128/jvi.79.2.884-895.2005

Wu, A. et al. (2020) 'Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China', Cell Host Microbe, 27 (3), pp. 325 – 328. https://doi.org/10.1016/j.chom.2020.02.001

Wu, F. et al. (2020) 'SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases', mSystems, 5(4), pp. e00614-20. https://doi.org/10.1128/mSystems.00614-20

Wu, L. et al. (2020) 'SARS-CoV-2 and cardiovascular complications: From molecular mechanisms to pharmaceutical management', Biochemical Pharmacology, 178, pp. 114114. https://doi.org/10.1016/j.bcp.2020.114114

Xia, X. (2021) 'Domains and functions of spike protein in sars-cov-2 in the context of vaccine design', Viruses, 13(1), pp. 109. https://doi.org/10.3390/v13010109

Xu, L. et al. (2020) 'Liver injury during highly pathogenic human coronavirus infections', Liver International, 40(5), pp. 998–1004. https://doi.org/10.1111/liv.14435

Xu, Y. et al. (2020) 'Current approaches in laboratory testing for SARS-CoV-2', International Journal of Infectious Diseases, 100, pp. 7–9. https://doi.org/10.1016/j.ijid.2020.08.041

Yaqinuddin, A. et al. (2021) 'Effect of sars-cov-2 mutations on the efficacy of antibody therapy and response to vaccines', Vaccines, 9(8), pp. 914. https://doi.org/10.3390/vaccines9080914

Young, M. et al. (2022) 'Covid-19: virology, variants, and vaccines', BMJ Medicine, 1(1), pp. e000040. https://doi.org/10.1136/bmjmed-2021-000040

Yu, A.T. et al. (2022) 'Estimating Relative Abundance of 2 SARS-CoV-2 Variants through Wastewater Surveillance at 2 Large Metropolitan Sites, United States', Emerging Infectious Diseases, 28(5), pp. 940–947. https://doi.org/10.3201/eid2805.212488

Zaki, A. M. et al. (2012) 'Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia', New England Journal of Medicine, 367(19), pp. 1814–1820. https://doi.org/10.1056/nejmoa1211721

Zhang W. et al. (2020)a 'Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes', Emerging Microbes & Infections 9(1), pp. 386 – 389. https://doi.org/10.1080/22221751.2020.1729071

Zhang J. et al. (2020)b 'Fecal specimen diagnosis 2019 novel coronavirus-infected pneumonia', Journal of Medical Virology, 92(6), pp. 680 – 682 . https://doi.org/10.1002/jmv.25742

Zhang, J. et al. (2021) 'Gastrointestinal symptoms, pathophysiology, and treatment in COVID-19', Genes and Diseases, 8(4), pp. 385–400. https://doi.org/10.1016/j.gendis.2020.08.013

Zhang, L. et al. (2020) 'Genomic variations of SARS-CoV-2 suggest multiple outbreak sources of transmission', medRxiv. https://doi.org/10.1101/2020.02.25.20027953

Zhao, X. et al. (2020) '2020 Update on Human Coronaviruses: One health, one world', Medicine in Novel Technology and Devices, 8, pp. 100043. https://doi.org/10.1016/j.medntd.2020.100043

Zhou, H. et al. (2020) 'A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein', Current Biology, 30(11), pp. 2196-2203.e3. https://doi.org/10.1016/j.cub.2020.05.023

Zhu, N. et al. (2020) 'A Novel Coronavirus from Patients with Pneumonia in China, 2019', New England Journal of Medicine, 382(8), pp. 727–733. https://doi.org/10.1056/nejmoa2001017

Zuccato, E. et al. (2005) 'Cocaine in surface waters: A new evidence-based tool to monitor community drug abuse', Environmental Health, 4, pp. 14. https://doi.org/10.1186/1476-069X-4-14

WEBSITES

WHO (2021), 'MERS Situation Update' (https://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html) (accessed April 2021)

Standford University Coronavirus Antiviral & Resistance Database (2023) (https://covdb.stanford.edu/variants/omicron_xbb/) (accessed September 2023)

NICD (2021), 'COVID-19 Update: Delta variant in South Africa' (https://www.nicd.ac.za/covid-19-update-delta-variant-in-south-africa/) (accessed October 2024)

CDC (2021), 'Symptoms of COVID-19' (https://www.cdc.gov/covid/signs-symptoms/index.html) (accessed April 2021)

WHO (2021), 'WHO announces simple, easy-to-say labels for SARS-CoV-2 Variants of Interest and Concern' (https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern) (accessed May 2021)

CDC (2023), 'COVID-19 Variant Classification' (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html) (accessed November 2023)

Biorender (2021) 'COVID-19 Vaccine Tracker' (https://biorender.com/covid-vaccine-tracker) (accessed April 2021)

WHO (2021), 'WHO convened global study of origins of SARS-CoV-2: China part' (https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part)

1.6) Thesis Structure

When the COVID-19 pandemic began, challenges such as rapid transmission, and high number of infected individuals, most who required hospitalization and healthcare attention, led to a global shutdown and economic crisis in many nations, particularly those in Africa. With knowledge of the continuous re-emergence of HCoVs in different populations, this study first sought to understand which HCoVs have been detected and described in Africa before the outbreak of SARS-CoV-2. This was done by systematically reviewing published literature. Some important findings from this review revealed that most studies (62.5%) were conducted in a hospital or clinic-facility setting and the rest of them conducted in either in communities, farms or airport settings. The study also showed that even though molecular techniques were employed in the detection of HCoVs, very few molecular epidemiology studies have been conducted in Africa.

Findings from the systematic literature review led to investigation for a cost-effective method that does not rely on obtaining samples from individuals to monitor the trends of SARS-CoV-2 at a population level. With the application of WBE during the pandemic worldwide, particularly in high income settings, this study sought to establish a surveillance system for real-time tracking of SARS-CoV-2 trends and description of the VOCs occurring in the Limpopo Province. In addition to all the findings from this established surveillance system, one key finding was the need for optimization of the applied allele-specific genotyping (ASG) technique for concrete assignment of variants, and the need to confirm determined variants by whole genome sequencing.

To confirm SARS-CoV-2 detected in wastewater samples, positively amplified SARS-CoV-2 samples were subjected to whole genome sequencing. SARS-CoV-2 whole genome sequences were then used to describe the molecular epidemiology of SARS-CoV-2 and investigate the genetic diversity among sequenced samples. The seasonality of respiratory viruses occurring in the study sites was also investigated. Results from SARS-CoV-2 WGS revealed circulation of Delta and Omicron VOCs as early as January and February 2021, respectively. Of the twelve respiratory viruses detected, an H5N1 Influenza A virus, absent from the NICD 2021 clinical surveillance reports was observed

from wastewater analysis. These findings illustrate the benefit of adding wastewater-based surveillance to clinical surveillance for monitoring disease trends in communities.

CHAPTER TWO

Prevalence and Molecular Epidemiology of Human Coronaviruses in Africa prior to SARS-CoV-2 Outbreak: a systematic review

Published Article: ***Viruses*** **2023, 15(11), 2146**

https://doi.org/10.3390/v15112,146

Contribution to Chapter 2: The first author conceptualized the study, performed literature search, carried out analysis, and wrote the original manuscript draft, which was extracted from the main thesis

ABSTRACT

Coronaviruses have been re-emerging in human populations, causing mild or severe acute respiratory diseases, and sometimes in epidemic proportions. This study aimed to systematically review the distribution and molecular epidemiology of human coronavirus (HCoVs) infections in Africa prior to the SAR-CoV-2 pandemic. The first published data on HCoV was from South Africa in 2008 in which NL63 was described in children less than 5 years old. Published data, from 40 studies, on the prevalence or molecular epidemiology of HCoVs was available from 13 of the 54 African countries (24%) prior to the SARS-CoV-2 outbreak. Of the 40 studies, 19 (47.5%) reported exclusively on HCoVs. Eight (four from Kenya) of the 40 studies (20%) reported on the molecular epidemiology of HCoV, in which genotypes of OC43 & NL63 were reported. Endemic HCoV prevalence ranged from 0.0% to 18.2%, while the prevalence of zoonotic MERS-CoV ranged from 0.0% to 83.5%. A prevalence ranging from 0.85% to 10.6% of infection was reported across 15 studies (36.6%) in children ≤13 years old. Specifically, the prevalence for MERS-CoV ranged from 0.18% among a group of camel handlers in Kenya to 83.5% in a population comprising individuals returning from Saudi Arabia and Hospitalized patients in Sudan. Two studies (one each from Kenya and Sudan) investigated SARS-CoV infection for which a prevalence of 0.0% was reported. There was heterogeneity in the type of tests used in determining HCoV prevalence across the different studies. Two studies, one from Côte D'Ivoire and one from Nigeria reported that risk factors for HCoV include exposure to infected animals or humans. Despite the outbreaks of SARS in Southeast Asia in 2002 and MERS in 2012 in the Middle East, the quantum of virologic investigations on HCoV on the African continent was very modest. Pandemic preparedness requires cognizance of disease outbreaks in other continents, establishment of test and surveillance protocols, and infrastructure for eventualities.

Keywords: HCoVs; prevalence; molecular epidemiology; Africa; pandemic preparedness

2.1) INTRODUCTION AND STUDY RATIONALE

Acute respiratory infections (ARIs), including infections with human coronaviruses (HCoV), are the leading cause of morbidity and mortality worldwide. A great burden (43%) of these infections is reported in developing countries, with a death rate above 80% (www.who.int/). Some viruses implicated as causative agents include Human Adenoviruses (hAdV), Influenza viruses, Human Parainfluenza viruses (hPIV) types 1–4, Human Coronaviruses (HCoVs), Human Rhinoviruses (hRV), Human Bocaviruses (hBoV), Enteroviruses (EVs), Parechoviruses, Respiratory syncytial viruses (RSV), human metapneumovirus (hMPV), and Wu Polyomaviruses. These viruses have led to respiratory tract infections, ranging from mild upper respiratory tract infections (URTIs) to severe acute respiratory illness (SARI) affecting the lower respiratory tract (LRTIs). Due to the great disease burden caused by these pathogens, many studies in Africa have investigated the viral etiology of respiratory pathogens that contribute to SARI, with more focus on RSV, hRV, hMPV and hPIVs.

Structurally, Coronaviruses (CoVs) have a protective outer layer, a linear shape, and a single strand of RNA that carries genetic information. They are classified as the Coronaviridae family. The pathogens have the ability to infect both animals and humans (Lim et al., 2016). Coronaviruses possess RNA genomes that are among the largest, measuring between 27 and 33 kilobases (kb). They are categorized into four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus (Fung and Liu, 2019; Chen et al., 2020). So far, a total of seven Human Coronaviruses (HCoV) have been documented. The viruses belong to the Alphacoronavirus (HCoV-NL63 and HCoV-229E) and Betacoronavirus (HCoV-OC43, HCoV-HKU1, severe acute respiratory syndrome coronavirus; SARS-CoV, Middle East respiratory syndrome coronavirus; MERS-CoV, and SARS-CoV-2) families. Seasonally, HCoVs (HKU1, OC43, NL63, and 229E) are present and cause mild upper respiratory tract infections in healthy individuals (Zumla et al., 2016). However, they can also cause more severe lower respiratory tract infections (LRTIs) in infants, young children, immunocompromised individuals, individuals with comorbidities, and the elderly (Vabret et al., 2003; Van Der Hoek, Pyrc and Berkhout, 2006; Gaunt et al., 2010; Mackay et al., 2012). The highly pathogenic Human

Coronaviruses (HCoVs) known as SARS-CoV, MERS-CoV, and SARS-CoV-2 originated from animals and were transmitted to humans, leading to localized epidemics in China (Peiris et al., 2003), the Middle East (Zaki et al., 2012; Raj et al., 2014), and the current global COVID-19 pandemic, respectively. The zoonotic HCoVs (SARS-CoV, MERS-CoV, and SARS-CoV-2) cause more severe illness in comparison to the endemic HCoV variants.

The repeated reintroduction of Human Coronaviruses (HCoVs) into the human population in the past thirty years has increased the need for monitoring these disease-causing agents. Before the COVID-19 pandemic, the majority of studies examining the distribution and prevalence of HCoVs were conducted in countries where the SARS and MERS epidemics were concentrated. Research conducted in these areas made a substantial contribution to enhancing our understanding of the genetic traits, geographic distribution, and evolutionary trends of both native and zoonotic Human Coronaviruses (HCoVs). So far, molecular epidemiology investigations have identified HCoV-OC43 genotypes (A–K) that are regularly found worldwide (Lau et al., 2011; Ren et al., 2015; Oong et al., 2017; Zhang et al., 2018, 2022; Zhu et al., 2018). HCoV-NL63 has been found to have three primary genotypes (A, B, and C) based on genomic studies, which are prevalent globally (Shao et al., 2022). Additionally, these genotypes are further divided into six sub-genotypes (A1–A3 and C1–C3), with sub-genotype C3 being the most recently identified in Chinese pediatric patients (Wang et al., 2020). HCoV-229E has exhibited ongoing genetic changes over time, resulting in the emergence of different genogroups (Genogroup 1–4). Recent research has discovered two new genogroups (Genogroups 5 and 6) in a COVID-19 patient who was also infected with HCoV-229E in Hong Kong (Lau et al., 2021). The classification of HCoV-HKU1 into three genotypes (A, B, and C) is determined through phylogenetic analysis of the RNA-dependent RNA polymerase (RdRp), Spike (S), and Nucleocapsid (N) genes (Woo et al., 2006). The emergence of these genotypes and sub-genotypes is a result of ongoing nucleotide substitution and homologous recombination between circulating strains, which are frequent occurrences in the Coronaviridae family (Su et al., 2016; Forni et al., 2017).

Gaining knowledge about the frequency and genetic patterns of HCoVs can help in predicting and managing HCoV infections in different populations. This systematic research aimed to provide a comprehensive overview of the occurrence and genetic characteristics of Human Coronaviruses (HCoVs) in Africa before the emergence of the SAR-CoV-2 pandemic.

2.1.1) Hypothesis

Human Coronaviruses are sparsely distributed across Africa.

2.1.2) Research Question

What is the distribution and molecular epidemiology of HCoVs in Africa prior to the SARS-CoV-2 outbreak?

2.1.3) Study Aim

To describe the prevalence and molecular epidemiology of HCoVs in Africa prior to the SAR-CoV-2 outbreak

2.1.4) Specific Objectives

a) To determine the prevalence and distribution of HCoVs in Africa prior to the SARS-CoV-2 outbreak.
b) To investigate the methodologies applied for investigating HCoVs occurrence.
c) To determine the molecular epidemiology of HCoVs in Africa prior to the SARS-CoV-2 outbreak.
d) To investigate the risk factors associated with HCoVs infection

2.2) METHODOLOGY

2.2.1) Search Strategy

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach was used. An electronic search was carried out to identify studies that had reported on HCoV occurrence in Africa prior to the SARS-CoV-2 pandemic. PubMed, Web of Science, and Google Scholar databases were used to search articles from 01 January 1966 when the first HCoV was reported until 2019. Articles were searched on all three databases using the following search strategy. For **PubMed**: seroprevalence OR seroepidemiology OR "sero-epidemiology" OR seropositivity OR "sero-epidemiologic studies" OR epidemiology OR prevalence OR incidence OR distribution AND "human coronavirus*" OR "human coronavirus 229E" OR "human coronavirus OC43" OR "human coronavirus NL63" OR "human coronavirus HKU1" OR "severe acute respiratory syndrome coronavirus" OR "middle east respiratory syndrome coronavirus" AND “African country”. **Web of Science**: seroprevalence OR seroepidemiology OR "sero-epidemiology" OR seropositivity OR "sero-epidemiologic studies" OR epidemiology OR prevalence OR incidence OR distribution AND "human coronavirus*" OR "human coronavirus 229E" OR "human coronavirus OC43" OR "human coronavirus NL63" OR "human coronavirus HKU1" OR "severe acute respiratory syndrome coronavirus" OR "middle east respiratory syndrome coronavirus" AND “African country”. **Google Scholar:** seroprevalence OR seroepidemiology OR "sero-epidemiology" OR seropositivity OR "sero-epidemiologic studies" OR epidemiology OR prevalence OR incidence OR distribution AND "human coronavirus*" OR "human coronavirus 229E" OR "human coronavirus OC43" OR "human coronavirus NL63" OR "human coronavirus HKU1" OR "severe acute respiratory syndrome coronavirus" OR "middle east respiratory syndrome coronavirus" AND “African country”.

2.2.2) Inclusion and Exclusion Criteria

Relevant research and case reports on the prevalence and distribution of human coronavirus in African nations were chosen for analysis. The study only considered articles that fulfilled the specified criteria: Studies published prior to 2019 were examined,

specifically those that focused on the viral causes of respiratory viruses, including human coronaviruses, in both community and hospital settings. Additionally, studies that explored surveillance, molecular epidemiology, and genomic sequencing of human coronaviruses were included. Investigations into MERS-CoV in both humans and animals were also considered. Retrospective analyses, studies with multiple study sites in both African and non-African countries, and case reports were also taken into account. The analysis omitted the following studies: those conducted solely on animals, reviews, book chapters, theses, and editorial opinions. **Figure 9** displays the PRISMA flow diagram employed for sourcing, identifying, and choosing studies included in the present study. The relevant information, including the article title and authors, study nation, demography, age range of the study population, year of sample collection, sample size, type of HCoVs identified, technique of detection and genotyping, and prevalence, was retrieved and organized in **Table 1** for the studies that met the inclusion criteria.

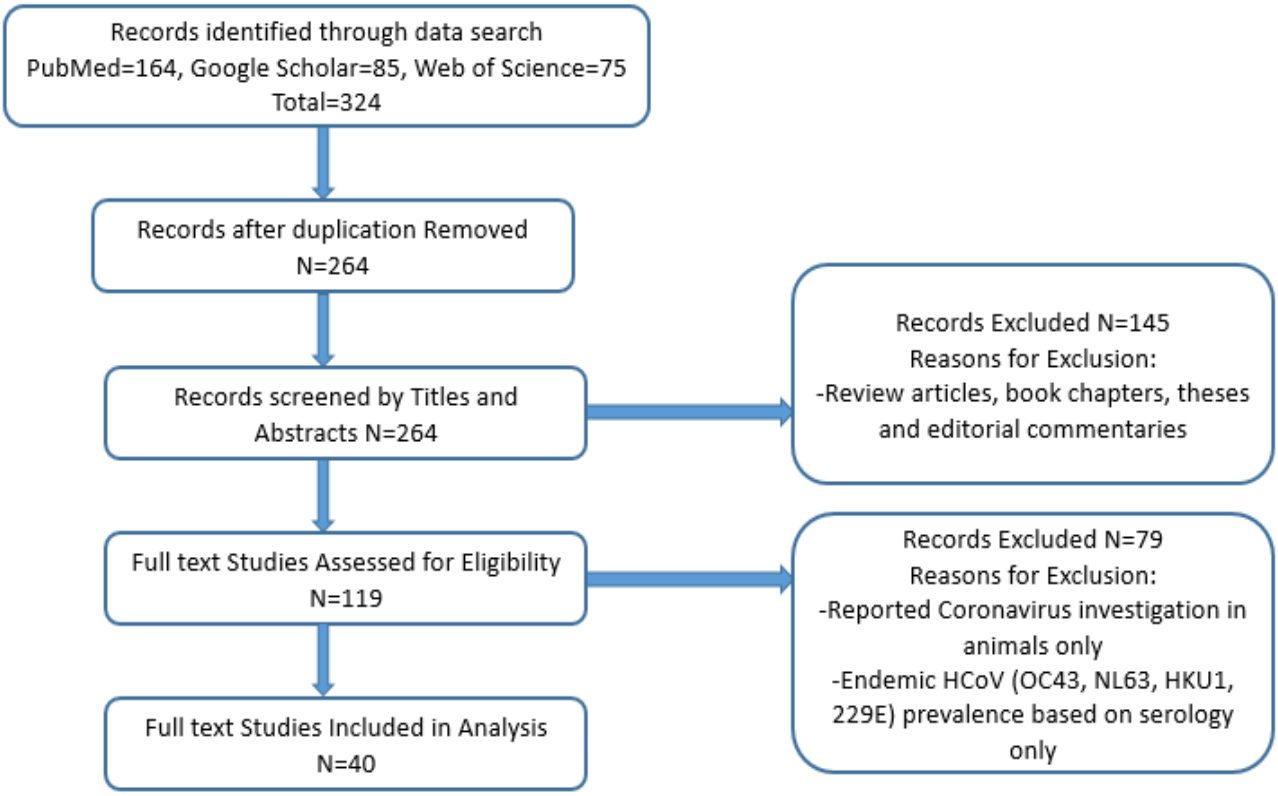


Figure 9: Systematic review PRISMA flowchart showing the guidelines used followed for screening and selecting studies for analysis.

Table 1: Studies that provided information on the occurrence and genetic characteristics of HCoVs in Africa before the outbreak of SARS-CoV-2.

Article Title & Author	Country	Demography	Age Range	Year of Sample Collection	Sample Size	Type of HCoV Investigated	Method of Detection or Genotyping	HCoV Prevalence	Genotypic Characterization
Viral etiology of severe acute respiratory infections in hospitalized children in Cameroon, 2011-2013 (Kenmoe et al., 2016)	Cameroon	Children	0 – 15 years	September 2011 – September 2013	347	OC43, 229E, NL63, HKU1	Multiplex RT-qPCR	5.8%	Not investigated
Viral etiology of influenza-like illnesses in Cameroon, January-December 2009 (Njouom et al., 2012)	Cameroon	Adults and Children	1.2 months – 75 years	January – December 2009	561	OC43, 229E, NL63, HKU1	One-Step RT-qPCR, Multiplex conventional RT-PCR	5.3%	Not investigated
Detection of Non-Influenza Viruses in Acute Respiratory Infections in Children Under 5yrs Old in Cote D'ivoire (January – December 2013) (Kadjo et al., 2018)	Côte D'Ivoire	Children	<5years	January – December 2013	1,059	229E, OC43	Multiplex conventional RT-PCR	3.7%	Not investigated
Investigation of an outbreak of acute respiratory disease in Cote D'ivoire in April 2007 (Ekaza et al., 2014).	Côte D'Ivoire	Animals, Adults and Children	0 – 15+ years	December 2006 – February 2007	104	OC43 & 229E	Multiplex conventional RT-PCR, sequencing (method not specified)	1.9%	Sequenced amplified HCoV-OC43 product (results not mentioned)
Cross-sectional survey and surveillance for influenza viruses	Egypt	Adults & Children	0 – 105 years	2012 – 2015	3,364	MERS-CoV	RT-qPCR	0%	Not investigated

and MERS-CoV among Egyptian pilgrims returning from Hajj during 2012-2015 (Refaey et al., 2017).									
Viral etiology and seasonality of influenza-like illness in Gabon, March 2010 to June 2011 (Lekana-Douki et al., 2014).	Gabon	Adults & Children	10 days – 82 years	March 2010 – June 2011	1,041	NL63, HKU1, 229E, OC43	Multiplex RT-qPCR	6.5%	Not investigated
Human Coronaviruses Associated with Upper Respiratory Tract Infections in Three Rural Areas of Ghana (Owusu et al., 2014)	Ghana	Adults & Children	10+ years	September 2011 – September 2012	1,213	229E, HKU1, NL63, OC43, MERS-CoV	RT-qPCR & Sequencing (method not specified)	12.4% (MERS-CoV not detected)	Similarity between sequenced HCoV strains and refence sequences
High prevalence of common respiratory viruses and no evidence of Middle East respiratory syndrome coronavirus in Hajj pilgrims returning to Ghana, 2013 (Annan et al., 2015).	Ghana	Adults	21 – 85 years	November 2013	839	MERS-CoV	RT-qPCR	0%	Not investigated
Similar virus spectra and seasonality in paediatric patients with acute respiratory disease, Ghana and Germany (Annan et al., 2016)	Ghana & German Children	Children	0 – 13 years	February 2008 – February 2009	1,174	229E, NL63, OC43, HKU1	One-Step RT-qPCR	6.7%	Not investigated
Continuous Invasion by Respiratory Viruses Observed in Rural Households During a Respiratory Syncytial Virus Seasonal Outbreak in Coastal Kenya (Munywoki et al., 2018).	Kenya	Adults & Children	4 – 37 years	December 2009 – June 2010	16,928 samples	OC43, NL63, 229E	Multiplex RT-qPCR	7.5%	Not investigated

Comparison of respiratory pathogen yields from Nasopharyngeal/Oropharyngeal swabs and sputum specimens collected from hospitalized adults in rural Western Kenya (Nyawanda et al., 2019)	Kenya	Adults	18 – 49 years	March 2014 – July 2015	294	NL63, OC43, HKU1, 229E	TaqMan Array Card	6.1%	Not investigated
Viral etiology of severe pneumonia among Kenyan infants and children (Berkley et al., 2010).	Kenya	Children	1 day – 12 years	January – December 2007	759	229E, OC43, NL63, HKU1	RT-qPCR, Sequencing (method not specified)	10%	Results not mentioned
No Serologic Evidence of Middle East Respiratory Syndrome Coronavirus Infection Among Camel Farmers Exposed to Highly Seropositive Camel Herds: A Household Linked Study, Kenya, 2013 (Munyua et al., 2017).	Kenya	Animals Adults & Children	5 – 90 years	2013	760	MERS-CoV	ELISA & plaque-reduction neutralization test (PRNT)	0%	Not investigated
MERS-CoV Antibodies in Humans, Africa, 2013-2014 (Liljander et al., 2016).	Kenya	Adults & Children	5 – 90 years	2013 – 2014	1,122	MERS-CoV	ELISA & plaque-reduction neutralization test (PRNT)	0.18%	Not investigated
Molecular characterization of human coronavirus circulating in Kenya, 2009-2012 (Sipulwa et al., 2016)	Kenya	Adults & Children	2 months – 67 years	January 2009 – December 2012	417	NL63, HKU1, 229E, OC43, MERS-CoV, SARS-CoV	RT-qPCR; Cell culture, Conventional RT-PCR, Sanger sequencing	8.4% (MERS-CoV and SARS-CoV not detected)	Sequenced samples clustered with reference strains. OC43 and NL63 viruses were under negative selection, albeit

									not statistically significant.
Infection patterns of endemic human coronaviruses in rural households in coastal Kenya (Nyaguthii et al., 2021)	Kenya	Adults & Children	4 – 23.4 (IQR)	December 2009 – June 2010	483	OC43, NL63, 229E	Multiplex RT-qPCR	7.5%	Not investigated
Surveillance of Respiratory Viruses in the Outpatient Setting in Rural Coastal Kenya: Baseline Epidemiological Observations (Nyiro et al., 2018)	Kenya	Adults & Children	0 – 100 years	January – December 2016	5,647	OC43, NL63, 229E	Multiplex RT-qPCR	6.8%	Not investigated
Transmission and evolutionary dynamics of human coronavirus OC43 strains in coastal Kenya investigated by partial spike sequence analysis, 2015 – 16 (Abidha et al., 2020)	Kenya	Adults & Children	0 – 100 years	December 2015 – June 2016	3,314	OC43	Multiplex RT-qPCR, Conventional RT-PCR, Sanger sequencing of Spike Gene	2.8%	Sequenced samples clustered with OC43 reference genotypes G (85%) and H (15%)
Surveillance of endemic human coronaviruses (HCoV-NL63, OC43 and 229E) associated with childhood pneumonia in Kilifi, Kenya (Otieno et al., 2020)	Kenya	Children	0 – 4 years	January 2007 – December 2019	7,957	NL63, OC43, 229E	Multiplex RT-qPCR	3.9%	Not investigated
Improved detection of respiratory viruses in pediatric outpatients with acute respiratory illness by real-time PCR using nasopharyngeal flocked swabs (Munywoki et al., 2011)	Kenya	Adults & Children	0 – 12 years	January – April 2009	299	OC43, NL63, 229E	Multiplex RT-qPCR	7.4%	Not investigated

Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya (Kiyuka et al., 2018).	Kenya	Adults & Children	0 – 100 years	February 2008 – May 2014	22,491	NL63	RT-PCR, HiSeq NGS	2.1%	NL63 genotype A and B observed, with six lineages (A0 – A2 and B0 – B2)
Genome sequences of human coronavirus OC43 and NL63, associated with respiratory infections in Kilifi, Kenya (Kamau et al., 2019)	Kenya	Children	2 months – 13 years	2017, 2018	3	OC43, NL63	MiSeq NGS	Retrospective Genomic study	OC43 genomes clustered in distinct genome-based phylogeny branches. NL63 genomes clustered with genotype B
Viral and atypical bacterial etiology .of acute respiratory infections in children under 5 years old living in a rural tropical area of Madagascar (Hoffmann et al., 2012)	Madagascar	Children	2 – 59 months	February 2010 – February 2011	295	NL63, 229E, OC43, HKU1	Multiplex RT-qPCR	8%	Not investigated
Viral etiology of influenza-like illnesses in Antananarivo, Madagascar, July 2008 to June 2009 (Razanajatovo et al., 2011)	Madagascar	Adults & Children	3 months – 77 years	July 2008 – June 2009	313	NL63, 229E, OC43, HKU1	Multiplex RT-qPCR, RT-qPCR	9.6%	Not investigated
Molecular detection of respiratory pathogens among children aged younger than 5 years hospitalized with febrile acute respiratory infections: A prospective hospital-based observational study in Niamey, Niger (Lagare et al., 2019)	Niger	Children	0 – 4 years	January – December 2015	638	OC43, 229E, NL63, HKU1	RT-qPCR	8.0%	Not investigated

Lack of serological evidence of Middle East respiratory syndrome coronavirus infection in virus exposed camel abattoir workers in Nigeria, 2016 (So et al., 2018).	Nigeria	Humans & Animals	Not specified	October 2015 – February 2016	311	MERS-CoV	ELISA, pseudoparticle neutralization assay (ppNT)	0%	Not investigated
Influenza-like illnesses in Senegal: not only focus on influenza viruses (Dia, Diene Sarr, et al., 2014)	Senegal	Adults & Children	0 – 25+ years	May 2012 – June 2013	1,427	OC43, 229E, NL63	Multiplex RT-qPCR	2%	Not investigated
Viral etiology of respiratory infections in children under 5 years old living in tropical rural areas of Senegal: The EVIRA project (Niang et al., 2010).	Senegal	Children	0 – 4 years	July – December 2007	67	OC43, NL63, 229E, HKU1	Multiplex Conventional RT-PCR	7.3%	Not investigated
Respiratory and gastrointestinal infections at the 2017 Grand Magal de Touba, Senegal: a prospective cohort survey (Hoang et al., 2019)	Senegal	Adults & Children	8 months – 75 years	4th – 23rd November 2017	123	NL63, 229E, OC43, HKU1	One Step Duplex RT-PCR	18.2%	Not investigated
Respiratory viruses associated with patients older than 50 years presenting with ILI in Senegal, 2009 to 2011 (Dia, Richard, et al., 2014)	Senegal	Adults	50 – 97 years	January 2009 – December 2011	232	NL63, 229E, OC43	Two-Step RT-qPCR	2.3%	Not investigated
Human coronavirus NL63 infections in infants hospitalised with acute respiratory tract infections in South Africa (Smuts, 2008).	South Africa	Children	13 days – 5 years	2003 – 2004	1055	NL63	Conventional RT-PCR	0.85%	Not investigated

Role of human metapneumovirus, human coronavirus NL63 and human bocavirus in infants and young children with acute wheezing (Smuts, Workman and Zar, 2008).	South Africa	Children	2 months – 6 years	May 2004 – November 2005	242	NL63	Conventional RT-PCR, Sanger sequencing	2.5%	NL63 genotype A and B detected
Human rhinovirus infection in young African children with acute wheezing (Smuts, Workman and Zar, 2011).	South Africa	Children	2 months – 5 years	May 2004 – November 2005	220	NL63	Conventional RT-PCR	1.3%	Not investigated
Contribution of Common and Recently Described Respiratory Viruses to Annual Hospitalizations in Children in South Africa (Venter et al., 2011)	South Africa	Children	0 – 4 years	2006 – 2007	610	NL63, OC43, 229E, HKU1	Multiplex RT-qPCR	4.4%	Not investigated
Clinical epidemiology of bocavirus, rhinovirus, two polyomaviruses and four coronaviruses in HIV-infected and HIV-uninfected South African children (Nunes et al., 2014).	South Africa	Children	1 month – 2 years	February 2000 to January 2002	1460	NL63, OC43, 229E, HKU1	Multiplex RT-qPCR	10.6%	Not investigated
Human bocavirus, coronavirus, and polyomavirus detected among patients hospitalised with severe acute respiratory illness in South Africa, 2012 to 2013 (Subramoney et al., 2018).	South Africa	Adults & Children	<1 – 65+ years	January 2012 – December 2013	680	NL63, HKU1, OC43, 229E	Multiplex RT-qPCR	4.8%	Not investigated
Detection, Identification & Sequencing of Middle East Respiratory Syndrome	Sudan	Adults & Children	<20 – 100 years	2014 – 2017	200	MERS-CoV, Pancoronavirus (229E, OC43,	Conventional One-Step RT-	95.1% (83.5% MERS-CoV;	Sequenced MERS-CoV samples from the hospital and airport

Coronavirus (MERS-CoV) among Sudanese Patients (Ibrahim, Kafi, Musa, Karsani, et al., 2018)						HKU1, NL63, SARS-CoV)	PCR, Sequencing	11.6%Pancoronavirus)	clustered with strains from Thailand and Saudi Arabia, respectively.
Detection of some respiratory viruses by molecular techniques among two Sudanese targets individual (Ibrahim, Kafi, Musa, Karsany, et al., 2018)	Sudan	Adults & Children	<20 – 100 years	2014 – 2017	200	MERS-CoV & Pancoronavirus (229E, OC43, HKU1, NL63, SARS-CoV)	Conventional One-Step RT-PCR (using Pancoronavirus panel)	95.1% (83.5% MERS-CoV; 11.6%Pancoronavirus)	Not investigated
MERS-CoV in camels but not camel handlers, Sudan, 2015 and 2017 (Farag et al., 2019)	Sudan & Qatar	Adults & Animals	Not specified	2015 – 2017	56	MERS-CoV	Spike (S1) protein microarray, S1 protein-based ELISA	0%	Not investigated
Family cluster of Middle East respiratory syndrome coronavirus infections, Tunisia, 2013 (Abroug, Slim, Ouanes-Besbes, Hadj Kacem, et al., 2014).	Tunisia	Adults	30 – 66 years	2013	14	MERS-CoV	RT-qPCR	21%	Sequenced sample clustered with reference sequences from Saudi Arabia and United Arab Emirates.

2.3) RESULTS

2.3.1) Characteristics of Studies Included in the Analysis

Forty articles that fully satisfied the specified criteria were selected and utilized for the analysis. The studies that satisfied the specified criteria were published from 2008 to 2021. Approximately 48% (19 out of 40) of the studies focused on the occurrence or molecular epidemiology of either naturally occurring HCoVs (OC43, NL63, 229E, HKU1) or zoonotic HCoVs (MERS-CoV, SARS-CoV). Approximately half (50%) of the investigations aimed to ascertain the viral etiology, epidemiology, or occurrence pattern of respiratory viruses. The majority of research (62.5%) were carried out in hospital settings or established influenza-surveillance sentinel facilities, where study participants were either hospitalized, consulting, or receiving vaccine. Out of the total 40 research, examination was conducted in communities (including farms and families) in 8 cases, which accounts for 20% of the studies. Two research (5%) were done in an airport context, while the remaining five studies (12.5%) utilized a hybrid strategy, involving either a combination of hospital and community settings or a combination of hospital and airport settings.

2.3.2) HCoV Prevalence and Distribution in Africa

The initial documented information on HCoV was reported in South Africa in 2008, specifically describing NL63 in children under the age of five (**Figure 10**). Prior to the outbreak of SARS-CoV-2, only 24% of African nations, specifically 13 out of 54, had data available on the prevalence of HCoV, as shown in **Table 1**. The prevalence of HCoV, as measured by molecular techniques, was found to be greater (ranging from 0% to 95.1%) compared to the prevalence reported by immunofluorescent assays (ranging from 0% to 0.18%). The occurrence of indigenous HCoVs (OC43, NL63, HKU1, and 229E) varied from 0.85% in children admitted to hospitals in South Africa to 18.2% in a diverse community (including both adults and children) at the Grand Magal de Touba in Senegal. Out of the 40 investigations, 15 (36.6%) specifically examined children aged 0-13 years old. These studies indicated a prevalence of endemic HCoVs (OC43, NL63, HKU1, and 229E) ranging from 0.85% to 10.6%.

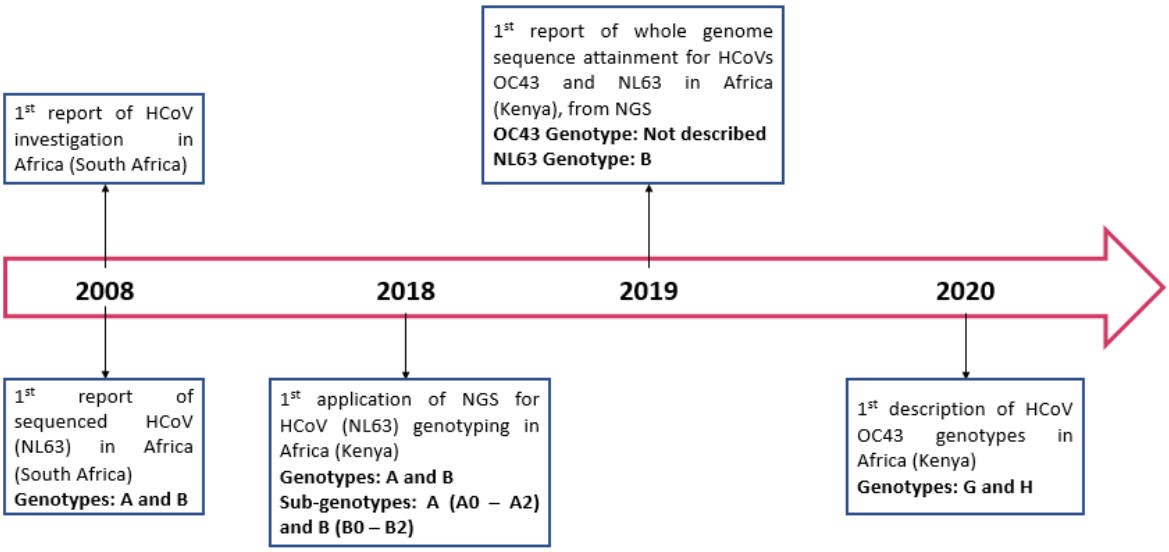


Figure 10: Chronology of research on Human Coronaviruses (HCoVs) in Africa preceding the emergence of SARS-CoV-2.

Overall, the occurrence of MERS-CoV varied from 0% among Egyptian pilgrims who had returned from Hajj to 95.1% in a group consisting of individuals who had returned from Saudi Arabia and hospitalized patients in Sudan. Out of the 11 out of 40 studies (27.5%) that examined the presence of MERS-CoV, 7 out of 11 (63.6%) of them found no cases at all. These investigations examined the prevalence of MERS-CoV in two groups: pilgrims returning to their home countries and animal handlers, including camels. Additionally, the reports also studied communities that had not been previously exposed to MERS-CoV. Out of the 11 investigations conducted, the prevalence of the condition varied between 0.18% in livestock handlers in Garissa and Tana river counties, Kenya, to 83.5% in individuals who had returned to Sudan from Saudi Arabia and hospital patients. The final 1/11 (9.1%) study focused on the presence of MERS-CoV and presented a case report that highlighted a family cluster of MERS-CoV in a father and daughter who had returned to Tunisia from Qatar. Out of a total of 40 research, only two (5%) focused on investigating SARS-CoV infection. These studies were conducted in Kenya and Sudan, and they indicated a prevalence rate of 0.0%.

The prevalence of HCoVs varied throughout different regions of the continent as follows: Southern Africa (0.85–10.6%), Central Africa (5.3–6.5%), West Africa (0–84.3%), East Africa (0–10%), and North Africa (0–95.1%). Among the 13 countries that have released data on

HCoV occurrence, Kenya had the highest proportion of studies published, accounting for 32.5% of the total. South Africa followed with 15%. Senegal accounted for 10% of the retrieved studies. Ghana and Sudan both had reports regarding 7.5% of all published research retrieved, whilst Madagascar, Cote D'Ivoire, and Cameroon each had a prevalence of 5%. The countries with the lowest amount of disclosed data (2.5% each) were Egypt, Gabon, Nigeria, Tunisia, and Niger.

The distribution of published research among African regions, ranked in descending order, is as follows: East Africa (15/40; 37.5%), West Africa (11/40; 27.5%), Southern Africa (6/40; 15%), North Africa (5/40; 12.5%), and Central Africa (3/40; 7.5%). **Figure 11** illustrates the relative amounts of published data, whereas **Figure 12** displays the geographic spread of studies that report the presence of non-SARS-CoV-2 HCoVs in Africa, together with the specific testing methods employed for inquiry.

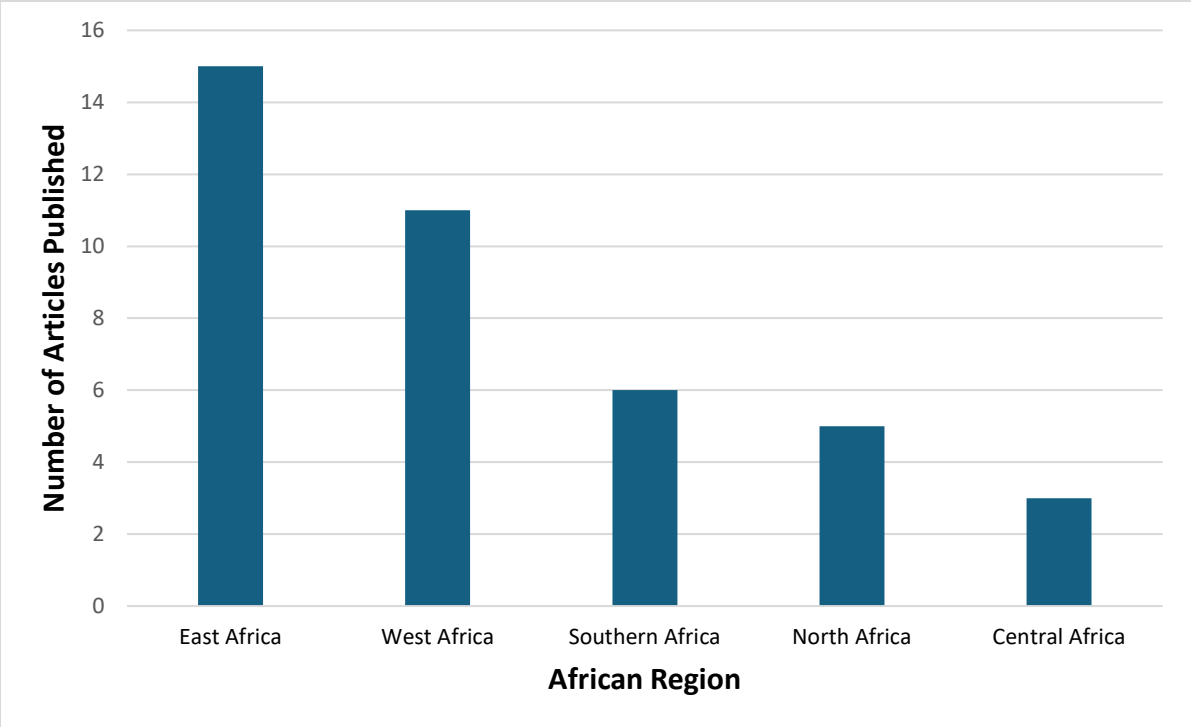


Figure 11: Prevalence of published publications on HCoVs throughout various African regions prior to the emergence of SARS-CoV-2.

Southern Africa- Botswana, Eswatini, Lesotho, Namibia, South Africa, Zimbabwe; Central Africa- Angola, Cameroon, Central Africa Republic, Chad, Congo, Gabon, Democratic Republic of Congo, Equatorial Guinea, Sao Tome & Principle; **West Africa-** Benin, Burkina

Faso, Cabo Verde, Cote D'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leonne, Togo; **East Africa-** Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Rwanda, Seychelles, Somalia, South Sudan, Tanzania, Uganda, Zambia; and **North Africa-** Algeria, Egypt, Libya, Morocco, Sudan, Tunisia.

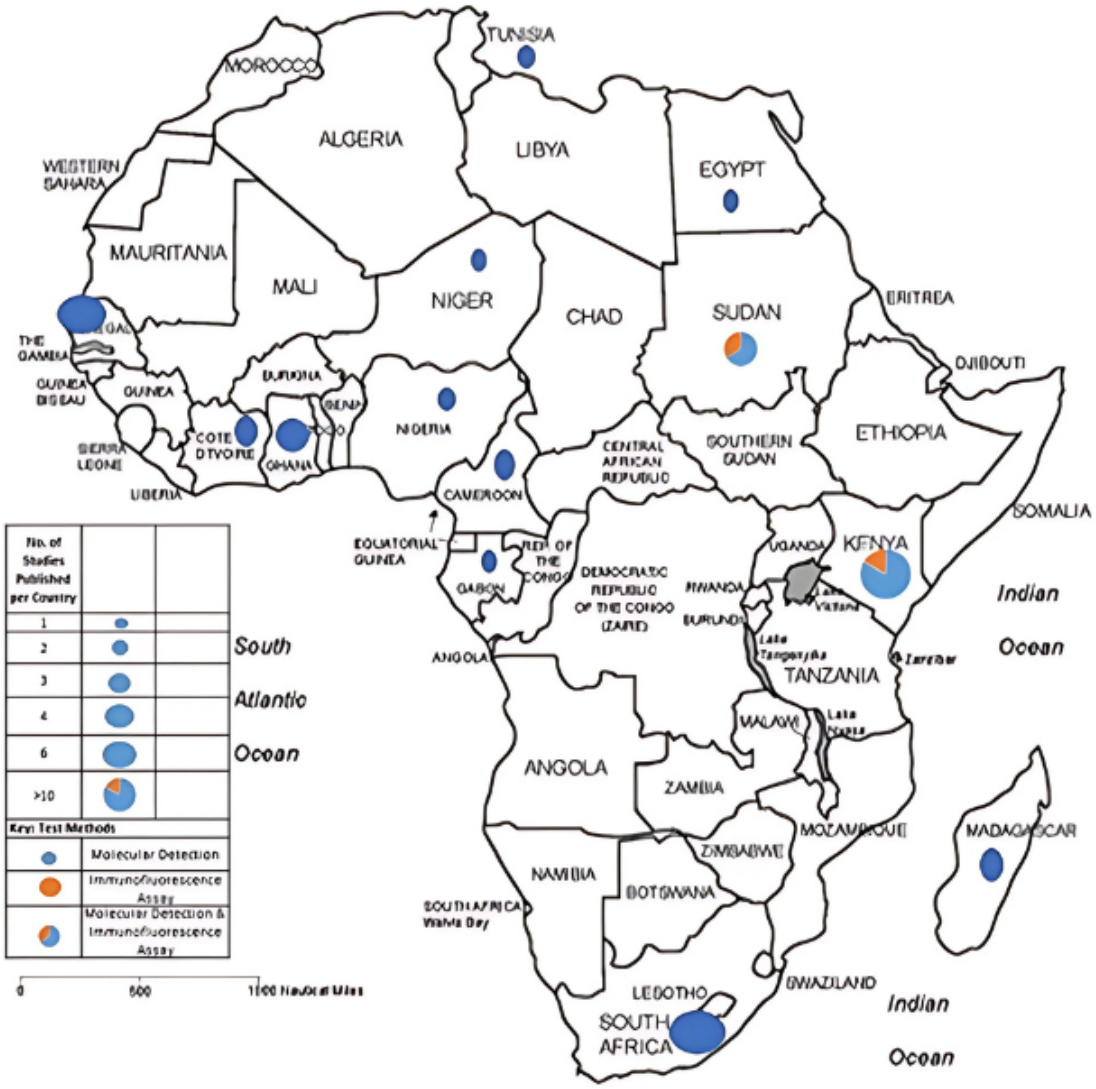


Figure 12: The African countries where research on non-SARS-CoV-2 HCoV had been published before the SARS-CoV-2 outbreak, along with the testing procedures used for inquiry.

2.3.3) Methodologies for HCoVs Detection

Prior to the emergence of SARS-CoV-2, many detection methods were used to ascertain the occurrence of HCoVs in Africa. The methods employed in the study encompassed molecular techniques, immunofluorescence tests (IFA), and culture, as outlined in **Table 1**. Molecular methodologies were employed in 87.5% (35 out of 40) of the evaluated research. The molecular techniques employed were reverse transcription polymerase chain reaction (RT-PCR), real-time reverse transcription polymerase chain reaction (RT-qPCR), multiplex real-time reverse transcription polymerase chain reaction (mRT-qPCR), and TaqMan array card (TAC) approach. The primary use of these molecular approaches was to study endemic HCoVs, which accounted for 70% of the investigations. These methodologies were also utilized in 5 out of 40 studies (12.5%) that specifically examined zoonotic HCoVs, and in 2 out of 40 studies (5%) that investigated both endemic and zoonotic HCoVs. Out of the total of 40 investigations completed in Sudan, only 2 (5%) utilized mRT-qPCR with a pancoronavirus panel. This panel is capable of detecting all types of coronaviruses, including those that affect both humans and animals, except for SARS-CoV and MERS-CoV. A study was conducted using both mRT-qPCR and culture procedures, where mRT-qPCR demonstrated a better sensitivity compared to culture.

Only 10% of the investigations (4 out of 40) utilized serological tests, such as ELISA, plaque-reduction neutralization test (PRNT), and pseudoparticle neutralization assay (ppNT), specifically for the identification of zoonotic MERS-CoV.

2.3.4) Molecular Epidemiology of HCoVs in Africa Prior to SARS-CoV-2 outbreak

Out of the total of 40 papers, 8 (20%) utilized sequencing to investigate the molecular epidemiology of HCoVs. However, only two publications explicitly mentioned that they used the Next Generation Sequencing approach (**Table 1**). Out of the total of eight investigations, half of them (50%) were conducted in Kenya, while the remaining studies were distributed equally among Ghana, South Africa, Sudan, and Tunisia, each accounting for 12.5%. The study from Kenya provided detailed information about the molecular features of locally prevalent human coronaviruses (HCoVs) during the period from 2008 to 2018. Seventy-five percent of the studies conducted in Kenya, namely in Kilifi County, focused on describing the presence of endemic HCoVs in a single rural location. The researchers documented the existence of both genotypes A and B of HCoV-NL63 in Kilifi County, with genotypes G and H

of HCoV-OC43 being the most prevalent among the population. The genotypes of HCoVs 229E and HKU1 were not documented in this area. The remaining study, constituting one-fourth of the total, was carried out in various regions of Kenya, including the Central, Northern, Western, Highlands, and Coastal regions. The study also found that their sequenced endemic HCoV strains were similar to the reference genomes. However, they did not provide genotypes (**Figure 13**).

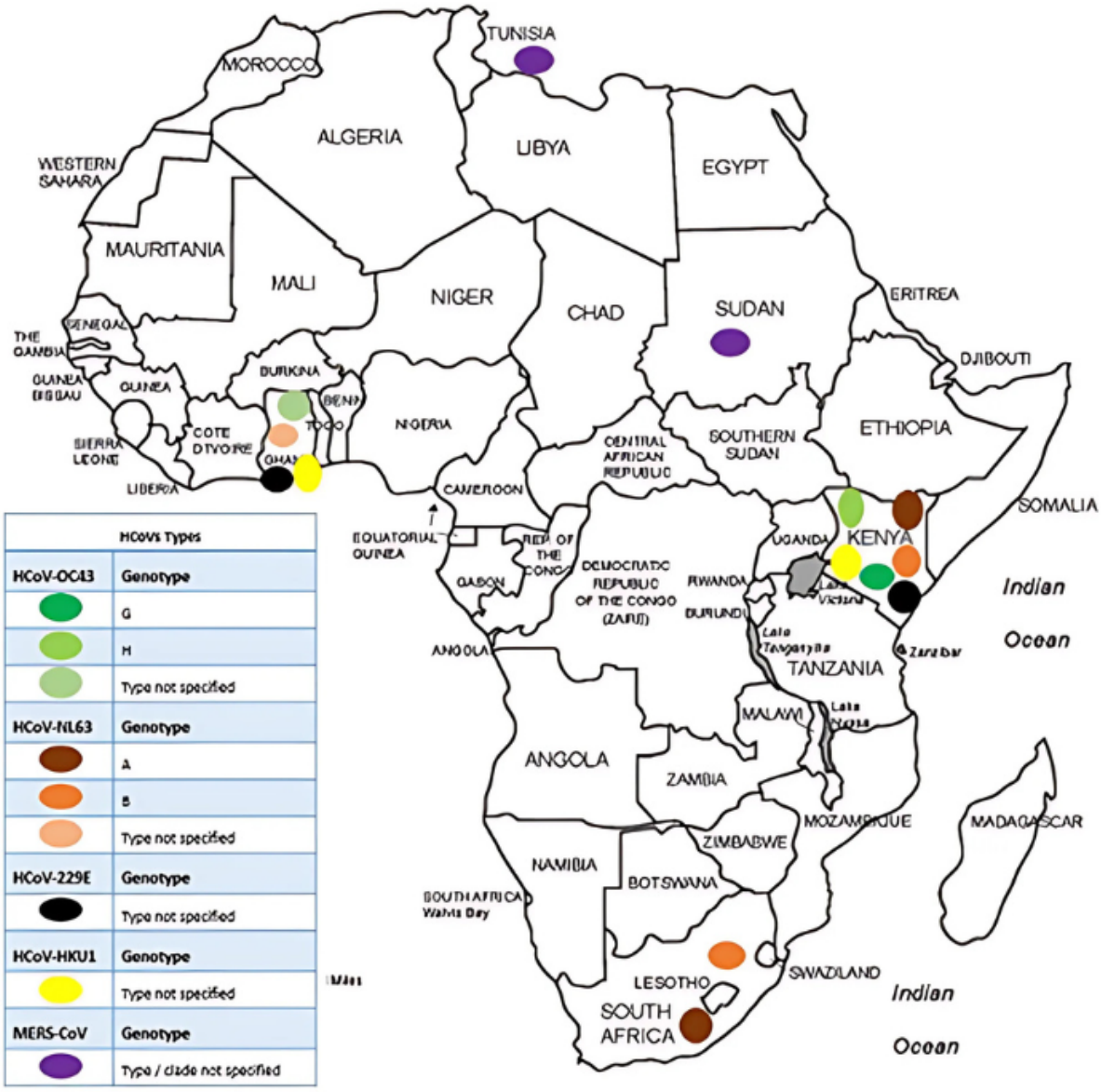


Figure 13: Distribution of HCoVs genotypes in Africa before the emergence of SARS-CoV-2.

The investigations conducted in Ghana and South Africa focused on the genetic traits of indigenous HCoVs, whereas the studies carried out in Sudan and Tunisia examined MERS-CoV. An analysis conducted on samples obtained from rural Ghana between 2011 and 2012 revealed that the HCoV strains found in that area were indistinguishable from the reference sequences. Studies conducted in Cape Town, South Africa, using samples taken from 2004 to 2005, revealed the presence of genotype A and B of HCoV-NL63. The study conducted in Sudan found that samples obtained between 2014 and 2017 from persons who had returned from Saudi Arabia and hospital patients exhibited genetic resemblance to MERS-CoV reference sequences from Saudi Arabia and Thailand, respectively. The samples taken in Tunisia in 2014 were shown to be closely related to MERS-CoV references from Saudi Arabia and the United Arab Emirates, based on their phylogenetic clustering.

2.3.5) Risk Factors Associated with HCoV Infection

Only 2 out of 40 studies (5%) examined risk variables related to HCoV infection. Both investigations done in Côte D'Ivoire and Nigeria utilized questionnaires to determine the probable exposure of the research participants or cases to infections. Data regarding associated risk factors, including exposure to infected animals, individuals (alive or deceased), travel records, and sources of sustenance and water, were gathered during the investigation of an outbreak of acute respiratory illness in Côte D'Ivoire. No correlation was discovered between the origin of exposure and the method of disease transmission. The Nigerian study examined the correlation between occupational exposure (either through direct or indirect contact) to dromedary camels and the transmission of MERS-CoV infection. Although the study participants were exposed to MERS-infected dromedary camels, none of them were infected with MERS-CoV.

2.4) DISCUSSION

Before the emergence of SARS-CoV-2, there was less information available regarding the occurrence, distribution, and prevalence of HCoV in Africa. Nevertheless, the need for ongoing surveillance of Human Coronaviruses (HCoVs) has been proven in the aftermath of the COVID-19 pandemic. Therefore, it is crucial to enhance monitoring efforts, establish uniform testing standards, provide necessary infrastructure, and train workers in order to be well-prepared for a pandemic.

Although endemic HCoVs (OC43, NL63, 229E, and HKU1) generally cause mild infections in individuals with a healthy immune system, they can lead to lower respiratory tract infections (LRTIs) in people with weakened immune systems, children aged 5 or younger, and the elderly, resulting in higher mortality rates (CDC, 2023). Before the emergence of SARS-CoV-2, research conducted in Africa from 2008 to 2021 documented the presence of HCoVs based on samples obtained from February 2000 to December 2019. Prior to the outbreak of SARS-CoV-2, the investigation revealed that the occurrence of endemic HCoVs (OC43, NL63, 229E, and HKU1) throughout the continent ranged from 0.85% to 18.2%. This estimate may be conservative, considering the majority of reports (62.5%) were derived from investigations conducted in hospital settings and specifically targeted children aged 5 years or younger. This population is recognized for bearing the weight of illness and being susceptible to Acute Respiratory Infections (ARIs), including infection with indigenous Human Coronaviruses (HCoVs). On the other hand, people who have a fully functioning immune system and are at least 14 years old are typically found to have moderate or asymptomatic HCoV infections, which are often not detected or diagnosed. Therefore, it is possible that the exact prevalence of endemic HCoV in a population is not known. Enhancing the accuracy of estimating the occurrence of endemic HCoV can be achieved by conducting community-based studies, such as those carried out on farms, in research cohorts, and during community activities. This approach will allow for the inclusion of both symptomatic and asymptomatic individuals, including adults and children. An investigation conducted in Senegal (Hoang et al., 2019) found that the prevalence of endemic HCoVs in the population (8 months–75 years old) was higher (18.2%) compared to the reported prevalence (0.85–10%) in hospital settings in other African locations. Utilizing community-based techniques can be advantageous in aiding

downstream molecular epidemiology investigations, as it allows for the characterization of genotypes present in the population. Additionally, it has the potential to contribute to the enhancement of diagnostic assay development efforts. Sudan saw a greater incidence of the zoonotic MERS-CoV (83.5%) among a group of returning pilgrims and hospitalized patients, as reported by Ibrahim, Kafi, Musa, Karsani, et al. in 2018. The elevated incidence of MERS-CoV could be attributed to extensive transmission that likely took place among pilgrims during the Hajj holiday in Saudi Arabia and subsequently identified following their arrival in Sudan. The COVID-19 pandemic saw an increase in transmission and distribution of variants worldwide, which was attributed to the occurrence of extensive travel and big gatherings (Kucharski et al., 2022). Therefore, the significant occurrence of MERS in Sudan should have prompted the country's public health officials to implement monitoring systems, given that a large number of Sudanese individuals are expected to go to an area where MERS is common for the annual Hajj pilgrimage. A review conducted by Park et al. (2020) found that the prevalence of endemic HCoVs ranged from 0.2% to 18.4% worldwide, indicating similar rates across different regions. Out of the 22 research that were analyzed, most of them were carried out in Asia (14 studies), while the fewest number of studies were conducted in Africa (1 study). Similar to our research, the reported prevalence was predominantly derived from patients (both adults and children) in hospital settings, who had acute respiratory infections (ARIs). This study emphasizes the lack of knowledge on native HCoVs in the continent, while also emphasizing the worldwide requirement for additional investigations conducted outside of hospitals to assess the prevalence of the virus in asymptomatic individuals, as well as the many genotypes that are currently circulating.

Following the COVID-19 pandemic, there is extensive discourse on the topic of pandemic preparedness. Taiwan, which experienced the least impact from the initial wave of COVID-19, provides valuable insights into effective pandemic preparedness (Asia-Pacific on Coronavirus Outbreak, 2023). Taiwan experienced a significantly high mortality rate during the SARS epidemic that occurred from 2002 to 2003. Following the SARS epidemic, Taiwan established a surveillance system that was quickly put into action during the outbreak of SARS-CoV-2. This led to a considerable decrease in infections during the initial wave, and a mitigating effect in following waves.

This review also highlighted the scarcity of molecular epidemiology studies on Human Coronaviruses (HCoVs) in Africa before the emergence of the SARS-CoV-2 outbreak. Foundational and practical investigations in virology are essential elements in the process of preparing for viral pandemics. These efforts involve identifying and evaluating materials for the creation of detection assays, characterizing viral genomes, and identifying epitopes that could be used for future vaccines. In addition to Kenya, where there has been ongoing research on the molecular epidemiology of HCoVs, it is crucial to do further genomic surveillance studies on HCoVs throughout Africa. This is essential in order to quickly detect any novel variants that may emerge. This is especially crucial because the accessibility of worldwide human movement allows for the inconspicuous introduction of new variations among populations. Phylogenetic clustering with MERS-CoV types from Saudi Arabia and UAE was observed in MERS positive patient sequences who returned from the Middle East, as reported in studies from Sudan and Tunisia (Abroug et al., 2014; Ibrahim et al., 2018). The occurrence of travelers introducing HCoV variants into a population was also observed during the COVID-19 pandemic, underscoring the importance of regular surveillance. The prompt detection of novel, potentially harmful genotypes enables swift intervention to halt transmission within the community, hence averting the spread and occurrence of epidemics. No instances of SARS-CoV were identified in any of the studies examined during the research. South Africa reported only one case during the 2002–2003 SARS outbreak, as documented in the World Health Organization's publication on suspected SARS cases from November 1, 2002, to July 31, 2003 (WHO, 2023). Two reasons may have contributed to the lack of additional cases in Africa during the 2002–2003 SARS outbreak. Initially, it has been stated that the transmissibility of SARS-CoV and MERS-CoV is comparatively lower than that of SARS-CoV-2. The transmissibility of SARS-CoV, MERS-CoV, and SARS-CoV-2, as evaluated by the basic reproduction rate (R0), is estimated to be 2.4, 0.9, and 2.5, respectively (Petersen et al., 2020; Pormohammad et al., 2020; Keshta et al., 2021). Furthermore, it has been documented that the primary mode of infection for SARS-CoV and MERS-CoV infections is by nosocomial transmission, as the highest levels of viral shedding occur during the symptomatic phase of the infection. The symptomatic period, during which patients sought medical attention, undoubtedly facilitated the transmission of the disease between patients and healthcare personnel (Zhou et al., 2021). Therefore, it is possible that SARS-CoV was

spread in Africa, but this went unnoticed despite the rise in worldwide mobility. Since its complete elimination in 2003, SARS-CoV has not been observed in the human population.

Prior to the outbreak of SARS-CoV-2, diverse testing methods were used to investigate HCoVs. The utilization of molecular methods was widespread. Regarding pandemic preparedness, this signifies the presence of accessible testing techniques and facilities. Government research institutions throughout Africa have the potential to test and improve current techniques in different environments. These studies will help identify settings that lack sufficient facilities, appropriate infrastructure or equipment, and skilled workers (Dzinamarira et al., 2022). Although whole genome sequencing (WGS) using next generation sequencing (NGS) provides insights into the evolution, variety, transmission, and spread of pathogens in a population, it may not be the most cost-effective approach for genomic surveillance, especially in resource-limited regions like Africa. Once again, the SARS-CoV-2 pandemic provided valuable insights as researchers worldwide employed an allele-specific genotyping (ASG) method for genomic monitoring. This strategy was utilized in several studies conducted by Harper et al. (2021), Wurtzer as al. (2021), and Johnson et al. (2022). This approach demonstrated high precision and cost-effectiveness in identifying genetic variations, making it suitable for widespread use in monitoring HCoVs over the entire continent.

Ultimately, we noticed a lack of available evidence on risk variables linked to HCoV infection. Both studies examining risk variables found no evidence of zoonotic transmission to humans. Africa is home to a wide range of wildlife, bats, and domestic livestock that carry many species of coronaviruses (Geldenhuys et al., 2021). Bats are recognized reservoirs of SARS-CoV and SARS-CoV-2, whereas MERS-CoV is widespread in dromedary camels; both animals are implicated hosts that have led to zoonotic transmission to humans. Studies conducted in living organisms (in vivo), laboratory settings (in vitro), and outside the organism but in a controlled environment (ex vivo) have examined the cause of limited viral transmission in Africa, despite ongoing contact with infected livestock. These studies have found that the MERS-CoV strain prevalent in Africa (Clade C) has a lower potential for transmission compared to the Arabian Clade A and B strains (Chu et al. 2018; Zhou et al. 2021; Rodon et al., 2023). Nevertheless, it is imperative to conduct ongoing phenotypic and molecular epidemiological research in order to closely monitor any potential alterations, especially in light of the persistent livestock trade between Africa and the Middle East. To avoid the spread of Arabian MERS-CoV strains,

it is necessary to confine livestock carrying these strains. This is because these strains have the potential to outperform the African Clade C strains, resulting in a higher risk of transmitting the virus to workers in close contact with the animals. Transmission of the virus has the potential to quickly propagate throughout families and communities, perhaps leading to the emergence of another epidemic.

2.5) CONCLUSION & RECOMMENDATIONS

This systematic review emphasizes the lack of research on Human Coronaviruses (HCoVs) in Africa before the outbreak of the SARS-CoV-2 pandemic. It is desirable that the SARS-CoV-2 pandemic acts as a catalyst for the implementation of surveillance systems to oversee HCoVs species in African populations, encompassing both humans and animals. Although most of Africa lacks sufficient resources, implementing cost-effective surveillance methods using wastewater-based techniques could yield economic benefits by addressing both symptomatic and asymptomatic populations (Ahmed et al., 2020; Nemudryi et al., 2020; Randazzo et al., 2020; Pillay et al., 2021; Wu et al., 2022). This method could be employed in conjunction with allele-specific genotyping to conduct sentinel surveillance in households. Improving and modernizing the current surveillance techniques for the prevalence and molecular epidemiology of HCoVs will strengthen Africa's role in developing diagnostic tests and help in preparing for pandemics.

2.6) REFERENCES

Abidha, C.A. et al. (2020) 'Transmission and evolutionary dynamics of human coronavirus OC43 strains in coastal Kenya investigated by partial spike sequence analysis, 2015 – 16', Virus evolution, 6(1), p. veaa031. https://doi.org/10.1093/ve/veaa031

Abroug, F. et al. (2014) 'Family cluster of middle east respiratory syndrome coronavirus infections, Tunisia, 2013', Emerging Infectious Diseases, 20(9), pp. 1527–1530. https://doi.org/10.3201/eid2009.140378

Ahmed, W. et al. (2020) 'First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community', Science of the Total Environment, 728. https://doi.org/10.1016/j.scitotenv.2020.138764

Annan, A. et al. (2015) 'High prevalence of common respiratory viruses and no evidence of Middle East respiratory syndrome coronavirus in Hajj pilgrims returning to Ghana, 2013', Tropical medicine & international health: TM & IH, 20(6), pp. 807–812. https://doi.org/10.1111/tmi.12482

Annan, A. et al. (2016) 'Similar virus spectra and seasonality in paediatric patients with acute respiratory disease, Ghana and Germany', Clinical Microbiology and Infection, 22(4), pp. 340–346. https://doi.org/10.1016/j.cmi.2015.11.002

Berkley, J.A. et al. (2010) 'Viral etiology of severe pneumonia among Kenyan infants and children.', Journal of the American Medical Association, 303(20), pp. 2051–2057. https://doi.org/10.1001/jama.2010.675

Chen, B. et al. (2020) 'Overview of lethal human coronaviruses', Signal Transduction and Targeted Therapy, 5(1), pp. 89. https://doi.org/10.1038/s41392-020-0190-2

Chu, D.K.W. et al. (2018) 'MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity', Proceedings of the National Academy of Sciences of the United States of America, 115(12), pp. 3144–3149. https://doi.org/10.1073/pnas.1718769115

Dia, N. et al. (2014) 'Influenza-like illnesses in Senegal: not only focus on influenza viruses', PLoS One, 9(3), pp. e93227. https://doi.org/10.1371/journal.pone.0093227

Dia, N. et al. (2014) 'Respiratory viruses associated with patients older than 50 years presenting with ILI in Senegal, 2009 to 2011', BMC Infectious Diseases, 14(1), pp. 1–6. https://doi.org/10.1186/1471-2334-14-189

Dzinamarira, T. et al. (2022) 'Utilization of SARS-CoV-2 Wastewater Surveillance in Africa-A Rapid Review', International journal of environmental research and public health, 19(2), pp. 969. https://doi.org/10.3390/ijerph19020969

Ekaza, E. et al. (2014) 'Investigation of an outbreak of acute respiratory disease in Côte d'Ivoire in April 2007', African Journal of Infectious Diseases, 8(2), pp. 31–35. https://doi.org/10.4314/ajid.v8i2.3

Farag, E. et al. (2019) 'MERS-CoV in camels but not camel handlers, Sudan, 2015 and 2017', Emerging infectious diseases, 25(12), pp. 2333 – 2335. https://doi.org/10.3201/eid2512.190882

Forni, D. et al. (2017) 'Molecular Evolution of Human Coronavirus Genomes', Trends in Microbiology, 25(1), pp. 35–48. https://doi.org/10.1016/j.tim.2016.09.001

Fung, T.S. and Liu, D.X. (2019) 'Human Coronavirus: Host-Pathogen Interaction', Annual Review of Microbiology 8(73), pp. 529–557. https://doi.org/10.1146/annurev-micro-020518-115759

Gaunt, E.R. et al. (2010) 'Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method', Journal of Clinical Microbiology, 48(8), pp. 2940–2947. https://doi.org/10.1128/JCM.00636-10

Geldenhuys, M. et al. (2021) 'Overview of bat and wildlife coronavirus surveillance in Africa: A framework for global investigations', Viruses, 13(5), pp. 1–37. https://doi.org/10.3390/v13050936

Harper, H. et al. (2021) 'Detecting SARS-CoV-2 variants with SNP genotyping', PLoS ONE, 16(2), pp. e0243185. https://doi.org/10.1371/journal.pone.0243185

Hoang, V.T. et al. (2019) 'Respiratory and gastrointestinal infections at the 2017 Grand Magal de Touba, Senegal: a prospective cohort survey', Travel medicine and infectious disease, 32, pp. 101410. https://doi.org/10.1016/j.tmaid.2019.04.010

Van Der Hoek, L. et al. (2006) 'Human coronavirus NL63, a new respiratory virus', FEMS Microbiology Reviews, 30(5), pp. 760–773. https://doi.org/10.1111/j.1574-6976.2006.00032.x

Hoffmann, J. et al. (2012) 'Viral and atypical bacterial etiology of acute respiratory infections in children under 5 years old living in a rural tropical area of Madagascar'. PLoS One, 7(8), pp. e43666. https://doi.org/10.1371/journal.pone.0043666

Ibrahim, H.S. et al. (2018) 'Detection, Identification \& Sequencing of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) among Sudanese Patients', American Journal of Microbiological Research, 6(4), pp. 181–186. https://doi.org/10.12691/ajmr-6-4-6

Ibrahim, H.S. et al. (2018) 'Detections of Some Respiratory Viruses by Molecular Techniques Among Two Sudanese Targets Individuals'. World Journal of Pharmaceutical Research, (7), 18 – 24. http://dx.doi.org/10.20959/wjpr20187-11543

Johnson, R. et al. (2022) 'Tracking the circulating SARS-CoV-2 variant of concern in South Africa using wastewater-based epidemiology.', Scientific reports, 12(1), pp. 1182. https://doi.org/10.1038/s41598-022-05110-4

Kadjo, H.A. et al. (2018) 'Detection of non-influenza viruses in acute respiratory infections in children under five-year-old in Cote D'Ivoire (January – December 2013)', African Journal of Infectious Diseases, 12(2), pp. 78–88. https://doi.org/10.21010/ajid.v12i2.13

Kamau, E. et al. (2019) 'Genome sequences of human coronavirus OC43 and NL63, associated with respiratory infections in Kilifi, Kenya', Microbiology Resource Announcements, 8(46), pp. e00730 – 19. https://doi.org/10.1128/MRA.00730-19

Kenmoe, S. et al. (2016) 'Viral etiology of severe acute respiratory infections in hospitalized children in Cameroon, 2011-2013.', Influenza and other respiratory viruses, 10(5), pp. 386–393. https://doi.org/10.1111/irv.12391

Keshta, A.S. et al. (2021) 'Journal of Infection and Public Health COVID-19 versus SARS: A comparative review'. Journal of Infection and Public Health, 14(7), pp. 967–977. https://doi.org/10.1016/j.jiph.2021.04.007

Kiyuka, P.K. et al. (2018) 'Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya.', The Journal of infectious diseases, 217(11), pp. 1728–1739. https://doi.org/10.1093/infdis/jiy098

Kucharski, A.J. et al. (2022) 'Travel measures in the SARS-CoV-2 variant era need clear objectives', The Lancet, 399(10333), pp. 1367–1369. https://doi.org/10.1016/S0140-6736(22)00366-X

Lagare, A. et al. (2019) 'Molecular detection of respiratory pathogens among children aged younger than 5 years hospitalized with febrile acute respiratory infections: A prospective hospital-based observational study in Niamey, Niger', Health Science Reports, 2(11), pp. e137. https://doi.org/10.1002/hsr2.137

Lau, S.K.P. et al. (2011) 'Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination', Journal of Virology, 85(21), pp. 11325–11337. https://doi.org/10.1128/jvi.05512-11

Lau, S.K.P. et al. (2021) 'Molecular Evolution of Human Coronavirus 229E in Hong Kong and a Fatal COVID-19 Case Involving Coinfection with a Novel Human Coronavirus 229E Genogroup', mSphere, 6(1) pp. e00819-20. https://doi.org/10.1128/msphere.00819-20

Lekana-Douki, S.E. et al. (2014) 'Viral etiology and seasonality of influenza-like illness in Gabon, March 2010 to June 2011', BMC Infectious Diseases, 14, pp. 373. https://doi.org/10.1186/1471-2334-14-373

Liljander, A. et al. (2016) 'MERS-CoV Antibodies in Humans, Africa, 2013-2014', Emerging Infectious Diseases, 22(6), pp. 1086–1089. https://doi.org/10.3201/eid2206.160064

Lim, Y. et al. (2016) 'Human Coronaviruses: A Review of Virus–Host Interactions', Diseases, 4(3), pp. 26. https://doi.org/10.3390/diseases4030026

Mackay, I.M. et al. (2012) 'Co-circulation of four human coronaviruses (HCoVs) in Queensland children with acute respiratory tract illnesses in 2004', Viruses, 4(4), pp. 637–653. https://doi.org/10.3390/v4040637

Munyua, P. et al. (2017) 'No Serologic Evidence of Middle East Respiratory Syndrome Coronavirus Infection Among Camel Farmers Exposed to Highly Seropositive Camel Herds: A Household Linked Study, Kenya, 2013.', The American journal of tropical medicine and hygiene, 96(6), pp. 1318–1324. https://doi.org/10.4269/ajtmh.16-0880

Munywoki, P.K. et al. (2011) 'Improved detection of respiratory viruses in pediatric outpatients with acute respiratory illness by real-time PCR using nasopharyngeal flocked swabs', Journal of clinical microbiology, 49(9), pp. 3365–3367. https://doi.org/10.1128/JCM.02231-10

Munywoki, P.K. et al. (2018) 'Continuous Invasion by Respiratory Viruses Observed in Rural Households During a Respiratory Syncytial Virus Seasonal Outbreak in Coastal Kenya.', Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 67(10), pp. 1559–1567. https://doi.org/10.1093/cid/ciy313

Nemudryi, A. et al. (2020) 'Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater.', Cell reports. Medicine, 1(6), pp. 100098. https://doi.org/10.1016/j.xcrm.2020.100098

Niang, M.N. et al. (2010) 'Viral etiology of respiratory infections in children under 5 years old living in tropical rural areas of Senegal: The EVIRA project.', Journal of medical virology, 82(5), pp. 866–872. https://doi.org/10.1002/jmv.21665

Njouom, R. et al. (2012) 'Viral etiology of influenza-like illnesses in Cameroon, January-December 2009', Journal of Infectious Diseases, 206(SUPPL.1), pp. 29–35. https://doi.org/10.1093/infdis/jis573

Nunes, M.C. et al. (2014) 'Clinical epidemiology of bocavirus, rhinovirus, two polyomaviruses and four coronaviruses in HIV-infected and HIV-uninfected South African children', PLoS ONE, 9(2), pp. e86448. https://doi.org/10.1371/journal.pone.0086448

Nyaguthii, D.M. et al. (2021) 'Infection patterns of endemic human coronaviruses in rural households in coastal Kenya', Wellcome Open Research, 6(27). https://doi.org/10.12688/wellcomeopenres.16508.1

Nyawanda, B.O. et al. (2019) 'Comparison of respiratory pathogen yields from Nasopharyngeal/Oropharyngeal swabs and sputum specimens collected from hospitalized adults in rural Western Kenya', Scientific Reports, 9(1), pp. 112387. https://doi.org/10.1038/s41598-019-47713-4

Nyiro, J.U. et al. (2018) 'Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: baseline epidemiological observations', Wellcome Open Research, 3(89). https://doi.org/10.12688/wellcomeopenres.14662.1

Oong, X.Y. et al. (2017) 'Identification and evolutionary dynamics of two novel human coronavirus OC43 genotypes associated with acute respiratory infections: Phylogenetic, spatiotemporal and transmission network analyses', Emerging Microbes and Infections, 6(1), pp. e3-13. https://doi.org/10.1038/emi.2016.132

Otieno, G.P. et al. (2020) 'Surveillance of endemic human coronaviruses (HCoV-NL63, OC43 and 229E) associated with childhood pneumonia in Kilifi, Kenya', Wellcome Open Research, 5, pp. 150. https://doi.org/10.12688/wellcomeopenres.16037.2

Owusu, M. et al. (2014) 'Human coronaviruses associated with upper respiratory tract infections in three rural areas of Ghana', PLoS one, 9(7), pp. e99782. https://doi.org/10.1371/journal.pone.0099782

Park, S. et al. (2020) 'Global Seasonality of Human Coronaviruses: A Systematic Review', Open Forum Infectious Disease, 7(11), pp. ofaa443. https://doi.org/10.1093/ofid/ofaa443

Peiris, J.S.M. et al. (2003) 'Coronavirus as a possible cause of severe acute respiratory syndrome', Lancet, 361(9366), pp. 1319–1325. https://doi.org/10.1016/S0140-6736(03)13077-2

Petersen, E. et al. (2020) 'Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics', The Lancet Infectious Diseases, 20(9), pp. e238–e244. https://doi.org/10.1016/S1473-3099(20)30484-9

Pillay, L. et al. (2021) 'Monitoring changes in COVID-19 infection using wastewater-based epidemiology: A South African perspective', Science of the Total Environment, 786, pp. 147273. https://doi.org/10.1016/j.scitotenv.2021.147273

Pormohammad, A. et al. (2020) 'Comparison of confirmed COVID-19 with SARS and MERS cases - Clinical characteristics, laboratory findings, radiographic signs and outcomes: A systematic review and meta-analysis', Reviews in Medical Virology, 30(4), pp. e2112. https://doi.org/10.1002/rmv.2112

Raj, V.S. et al. (2014) 'MERS: Emergence of a novel human coronavirus', Current Opinion in Virology, 5(1), pp. 58–62. https://doi.org/10.1016/j.coviro.2014.01.010

Randazzo, W. et al. (2020) 'SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area', Water Research, 181, pp. 115942. https://doi.org/10.1016/j.watres.2020.115942

Razanajatovo, N.H. et al. (2011) 'Viral etiology of influenza-like illnesses in Antananarivo, Madagascar, July 2008 to June 2009', PLoS ONE, 6(3), pp. e17579. https://doi.org/10.1371/journal.pone.0017579

Refaey, S. et al. (2017) 'Cross-sectional survey and surveillance for influenza viruses and MERS-CoV among Egyptian pilgrims returning from Hajj during 2012-2015.', Influenza and other respiratory viruses, 11(1), pp, 57–60. https://doi.org/10.1111/irv.12429

Ren, L. et al. (2015) 'Genetic drift of human coronavirus OC43 spike gene during adaptive evolution', Scientific Reports, 5, pp. 11451. https://doi.org/10.1038/srep11451

Rodon, J. et al. (2023) 'Extended Viral Shedding of MERS-CoV Clade B Virus in Llamas Compared with African Clade C Strain', Emerging Infectious Diseases, 29(3), pp. 585–589. https://doi.org/10.3201/eid2903.220986

Shao, N. et al. (2022) 'Molecular evolution of human coronavirus-NL63, -229E, -HKU1 and -OC43 in hospitalized children in China', Frontiers in Microbiology, 13, pp. 1023847. https://doi.org/10.3389/fmicb.2022.1023847

Sipulwa, L.A. et al. (2016) 'Molecular characterization of human coronaviruses and their circulation dynamics in Kenya, 2009 – 2012', Virology Journal, 13, pp. 18. https://doi.org/10.1186/s12985-016-0474-x

Smuts, H. (2008) 'Human coronavirus NL63 infections in infants hospitalised with acute respiratory tract infections in South Africa.', Influenza and other respiratory viruses, 2(4), pp. 135–138. https://doi.org/10.1111/j.1750-2659.2008.00049.x

Smuts, H.E. et al. (2011) 'Human rhinovirus infection in young African children with acute wheezing', BMC Infectious Diseases, 11, pp. 65. https://doi.org/10.1186/1471-2334-11-65

Smuts, H. et al. (2008) 'Role of human metapneumovirus, human coronavirus NL63 and human bocavirus in infants and young children with acute wheezing.', Journal of medical virology, 80(5), pp. 906–912. https://doi.org/10.1002/jmv.21135

So, R.T. et al. (2018) 'Lack of serological evidence of Middle East respiratory syndrome coronavirus infection in virus exposed camel abattoir workers in Nigeria, 2016', Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 23(32), pp. 1800175. https://doi.org/10.2807/1560-7917.ES.2018.23.32.1800175

Su, S. et al. (2016) 'Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses', Trends in Microbiology, 24(6), pp. 490–502. https://doi.org/10.1016/j.tim.2016.03.003

Subramoney, K. et al. (2018) 'Human bocavirus, coronavirus, and polyomavirus detected among patients hospitalised with severe acute respiratory illness in South Africa, 2012 to 2013', Health Science Reports, 1(8), pp. e59. https://doi.org/10.1002/hsr2.59

Vabret, A. et al. (2003) 'An outbreak of coronavirus OC43 respiratory infection in Normandy, France', Clinical Infectious Diseases, 36(8), pp. 985–989. https://doi.org/10.1086/374222

Venter, M. et al. (2011) 'Contribution of Common and Recently Described Respiratory Viruses to Annual Hospitalizations in Children in South Africa', Journal of Medical Virology, 83(8), pp. 1458–1468. https://doi.org/10.1002/jmv.22120

Wang, Y. et al. (2020) 'Discovery of a subgenotype of human coronavirus NL63 associated with severe lower respiratory tract infection in China, 2018', Emerging Microbes and Infections, 9(1), pp. 246–255. https://doi.org/10.1080/22221751.2020.1717999

Woo, P.C.Y. et al. (2006) 'Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1', Journal of Virology, 80(14), pp. 7136–7145. https://doi.org/10.1128/jvi.00509-06

Wu, F. et al. (2022) 'Making waves: Wastewater surveillance of SARS-CoV-2 in an endemic future', Water research, 219, pp. 118535. https://doi.org/10.1016/j.watres.2022.118535

Wurtzer, S. et al. (2021) 'Monitoring the propagation of SARS CoV2 variants by tracking identified mutation in wastewater using specific RT-qPCR', medRxiv, https://doi.org/10.1101/2021.03.10.21253291

Zaki, A.M. et al. (2012) 'Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia', New England Journal of Medicine, 367(19), pp. 1814–1820. https://doi.org/10.1056/nejmoa1211721

Zhang, S.F. et al. (2018) 'Epidemiology characteristics of human coronaviruses in patients with respiratory infection symptoms and phylogenetic analysis of HCoV-OC43 during 2010-2015 in Guangzhou', PLoS ONE, 13(1), pp. e0191789. https://doi.org/10.1371/journal.pone.0191789

Zhang, Z. et al. (2022) 'Two novel human coronavirus OC43 genotypes circulating in hospitalized children with pneumonia in China', Emerging Microbes and Infections, 11(1), pp. 168–171. https://doi.org/10.1080/22221751.2021.2019560

Zhou, H. et al. (2021) 'A Review of SARS-CoV2: Compared with SARS-CoV and MERS-CoV', Frontiers in Medicine, 8, pp. 628370. https://doi.org/10.3389/fmed.2021.628370

Zhou, Z. et al. (2021) 'Phenotypic and genetic characterization of MERS coronaviruses from Africa to understand their zoonotic potential', Proceedings of the National Academy of Sciences of the United States of America, 118(25), pp. e2103984118. https://doi.org/10.1073/pnas.2103984118

Zhu, Y. et al. (2018) 'A novel human coronavirus OC43 genotype detected in mainland China', Emerging Microbes and Infections, 7(1), pp. 173. https://doi.org/10.1038/s41426-018-0171-5

Zumla, A. et al. (2016) 'Coronaviruses-drug discovery and therapeutic options', Nature Reviews Drug Discovery, 15(5), pp. 327–347. https://doi.org/10.1038/nrd.2015.37

WEBSITES

CDC (2023) 'Common Human Coronaviruses' (https://www.cdc.gov/coronavirus/general-information.html) (accessed on 13 April 2023).

ASIA – PACIFIC, ON CORONAVIRUS OUTBREAK (2023), 'Taiwan's SARS Experience helped it beat COVID-19' (https://www.aa.com.tr/en/asia-pacific/-taiwan-s-sars-experience-helped-it-beat-covid-19-/1830547) (accessed on 11 July 2023).

WHO (2023), 'Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003' (https://www.who.int/publications/m/item/summary-of-probable-sars-cases-with-onset-of-illness-from-1-november-2002-to-31-july-2003) (accessed on 11 July 2023).

CHAPTER THREE

Wastewater-based Surveillance of SARS-CoV-2 and Description of SARS-CoV-2 Variants of Concern in Limpopo Province, South Africa

Paper under revision for resubmission

SSRN: https://dx.doi.org/10.2139/ssrn.4510876 (not peer reviewed)

Contribution to Chapter 3: The first author was responsible for sample receival and processing, qRT-PCR experiments and analysis, allele specific genotyping qRT-PCR experiments and analysis, prevalence estimation modelling and analysis. She also wrote the original draft manuscript which was extracted from the main thesis.

ABSTRACT

Across the globe, wastewater-based epidemiology (WBE) has been utilized as an alternative tool for assessing and managing the COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This sought to establish a wastewater-based surveillance system to monitor the trends of SARS-CoV-2, describe the occurrence of VOCs, and estimate the prevalence of people infected, in Vhembe and Mopani districts, Limpopo Province, South Africa. Grab sampling was employed in weekly collections of wastewater influent from seven wastewater treatment plants and one waste sedimentation pond. Results from this established surveillance system showed potential to serve as an early warning system, since increased viral concentrations preceded clinical data based on reports from the districts. Weekly surveillance data from the study sites was sent to the respective municipal authorities for possible action. Interestingly, towards the end of the surveillance period, increased SARS-CoV-2 viral load detected were not reflected in the reported clinical cases. This could be explained by decreased disease severity of the dominantly circulating Omicron variant, especially with increased vaccine uptake, leading to fewer reported clinical cases. Through allele specific genotyping, Delta and Omicron VOCs were detected in the study sites by 03 May 2021. These findings were confirmed through whole genome sequencing, indicating the circulation of these variants earlier than reported by other studies, thus revealing the importance of community surveillance. The presence of these VOCs was confirmed through whole genome next generation sequencing. Comparing the reported number of cases per district to the predicted prevalence revealed more cases in the Vhembe District, but less in the Mopani District. The latter could be attributed to predominant utilization of non-sewered sanitation systems, thus missing important shedding events of infected inhabitants. WBE application for SARS-CoV-2 surveillance and VOC monitoring have been reported from more urbanized regions in South Africa and other parts of the world. This study shows the applicability of wastewater-based surveillance system in rural areas of Limpopo province, and its contribution to the wastewater-surveillance repository data in Limpopo, South Africa. The study also highlights the challenge of non-existent sewerage systems, which could potentially diminish the advantages of WBE systems in rural settings. This is significant, since the eventual outcomes in these scenarios could potentially impair disease circulation estimation, as well as prevalence evaluation. This may also impact genomic surveillance, since early

detection of variants can contribute to drug development and vaccine efforts to potentially mitigate the detrimental impact of this disease. Utilization of a decentralized WBE approach, which involves detection and quantification of SARS-CoV-2 surface waters, could potentially address issues of poor or non-existent sewerage systems. However, more studies in rural areas are needed to contribute towards addressing these gaps by strengthening pandemic preparedness measures. Finally, this established surveillance system could also be adapted for monitoring and characterization of other infectious pathogens circulating in the population.

Keywords: COVID-19, SARS-CoV-2, wastewater-based epidemiology, early warning, variants of concern, infection prediction models

3.1) INTRODUCTION AND STUDY RATIONALE

When the World Health Organization (WHO) first designated the outbreak of the 2019 Coronavirus disease (COVID-19) as a pandemic in March 2020, approximately 509,164 infections were reported across the globe. However, many more people have been infected with severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) after two years of the pandemic. As of November 02nd, 2023, the WHO has recorded approximately 771,679,618 confirmed COVID-19 cases worldwide, and more than six million mortalities (WHO COVID-19 dashboard, 2023).

At the beginning of the pandemic, rapid identification of infected individuals to prevent further spread proved challenging. Two challenges encountered were the unavoidable wait time from infection to disease manifestation, which was approximately 2 – 14 days (CDC, 2023), and reliance on laboratory tests (the gold standard for detection) for confirmation of infection. Worldwide, these challenges greatly strained the limited resources, healthcare service personnel and facilities available to cater for infected individuals who became symptomatic almost simultaneously, given the time from infection to disease manifestation. Thus, the drawback of determining infection based on clinical manifestation alone was considered to not represent the full picture of infections occurring in the population since those with asymptomatic infections were largely unaccounted for. Therefore, to maximize clinical detection efforts, alternative methods to monitor COVID-19 incidence were needed, which could benefit local decision-making.

For decades, wastewater (WW) has been used to investigate the concentrations of used substances, pharmaceuticals, chemicals, exogenous contaminants and nutrient concentrations (Choi et al., 2018). Wastewater-based epidemiology (WBE) has been applied around the world for the surveillance of enteric viruses which infect the gastrointestinal tract and are shed in human feces which end up either in the environment or wastewater treatment plants (Hovi et al., 2012; Hellmér et al., 2014; Smith, Paddy and Simmonds, 2016; Tiwari and Dhole, 2018). This could serve as an early warning system (EWS) in case of a potential outbreak in the community. The applicability of WBE, as an EWS was demonstrated in May 2013, when a silent Polio outbreak occurring in a small community in Rahat, Israel, was

identified and rapidly intercepted with the launching of a vaccine program, by early 2014 (Brouwer et al., 2018).

The presence of SARS-CoV-2 in saliva, sputum, urine, blood, and feces (Peng et al., 2020; Wölfel et al., 2020; X. Li et al., 2022; Zhao et al., 2022) has been shown. It has also been proven that approximately 3 – 5 days post exposure, viral shedding of SARS-CoV-2 is seen in symptomatic, pre-symptomatic, and asymptomatic individuals (Jones et al., 2020). Such shedding from the saliva, sputum, urine, blood, and feces of infected individuals will eventually end up in wastewater treatment facilities. Thus, it was proposed that wastewater could contain SARS-CoV-2 and could be used as a sensitive alternative tool for surveillance (Lodder and de Roda Husman, 2020; Medema et al., 2020; Wu et al., 2020). Subsequently, several studies around the globe proved the efficacy of WBE in identifying COVID-19 hotspots, monitoring COVID-19 incidences, and serving as an EWS in communities prior to the commencement of a new wave (Gonzalez et al., 2020; Nemudryi et al., 2020; Sherchan et al., 2020; Giuseppina La Rosa et al., 2021; Zhu et al., 2021). Furthermore, SARS-CoV-2 concentrations in wastewater were also used to estimate the number of infected persons in a community (Ahmed et al., 2020). This predictive ability was particularly relevant for estimating total infections and not relying only on COVID-19 positive tests. Thereafter, several studies have utilized other mathematical models to predict the total prevalence and compared the efficacy of their model with the documented clinical cases (Gerrity et al., 2021; Kaplan et al., 2022).

WBE was also applied for monitoring SARS-CoV-2 variants of concern (VOCs) at the community level using various methods of sequencing (Bi et al., 2021; Faleye et al., 2021; Gregory et al., 2021; Rothman et al., 2021; Wilhelm et al., 2022). Some reports also indicated the applicability using genotyping methods which are more cost-effective than whole genome sequencing for monitoring SARS-CoV-2 variants in the population (La Rosa et al., 2021; Harper et al., 2021; Neopane et al., 2021; Takemae et al., 2022). Over the course of the pandemic, the WHO classified five SARS-CoV-2 variants as VOCs (WHO, 2023). The Alpha variant (B.1.1.7) was first detected in the United Kingdom (UK) in September 2020 (Hill et al., 2022). This was closely followed by the Beta variant (B.1.351), first reported December 2020, in South Africa (Tegally et al., 2021) and was predominantly circulating in the population during the second wave. Next came the Gamma variant (P.1), first identified in Brazil (Faria et al., 2021), but sparsely detected in Africa (Tegally et al., 2022), followed by the Delta variant

(B.1.617.2), first identified in India in March 2021 (Dhar et al., 2021), which was primarily responsible for the third wave of infections across the globe. The Omicron variant (B.1.1.529), is the last reported VOC till date (November 2023, the time of conclusion of this study), and its lineages and sub-lineages are still dominantly circulating worldwide. This VOC was first identified in Botswana and South Africa, in early November 2021 (Viana et al., 2022). This variant, characterized by more than 47 mutations (Standford University Coronavirus Antiviral & Resistance Database, 2023), was predominantly detected in most infections during the fourth wave in South Africa.

Most studies on wastewater surveillance come from the developed world (COVIDPoops19 Dashboard, 2023), which have established sewer and sanitation systems. Few reports from Africa have listed some challenges encountered with WBE surveillance of SARS-CoV-2 on the continent (Amoah et al., 2022; Dzinamarira et al., 2022; Pillay et al., 2022). In South Africa, WBE systems were established across the country to develop standard methodology for the identification and sequencing of SARS-CoV-2 from wastewater. This was done by the South African Collaborative COVID-19 Surveillance System (SACCESS) network, established in 2021, operating in collaboration with the South African National Institute for Communicable Diseases (NICD) and the South African Medical Research Council (SAMRC) (Johnson et al., 2021; Pillay et al., 2021; Iwu-Jaja et al., 2023).

While studies from the more urban areas of South Africa have shown the applicability of WBE for SARS-CoV-2 surveillance, there remains a necessity for establishment of wastewater-based surveillance (WBS) systems in rural areas of the country, which tend to be ignored. Thus, this study set out to establish a WBS system to monitor the trends of SARS-CoV-2, describe the occurrence of VOCs, and estimate the prevalence of people infected, in Vhembe and Mopani districts, South Africa. Detected trends of infection and variants occurrence will reveal how the COVID-19 pandemic unfolded in this region, compared to other areas globally. In addition, this will contribute towards the nation's repository, as well as highlight some challenges faced by rural regions, which may not be similar to those faced by the more urban areas in South Africa. Such information could be taken into consideration during the establishment of WBS systems in other rural parts of the country, and in other regions of Africa.

3.1.1) Hypothesis

Wastewater-based epidemiology can be used to track SARS-CoV-2 trends in the Limpopo Province of South Africa.

3.1.2) Research Questions

Can wastewater-based epidemiology be used to track the trends of SARS-CoV-2 in the Limpopo Province in South Africa?

3.1.3) Study Aim

To establish a wastewater-based surveillance system for real-time tracking and prevalence estimation of SARS-CoV-2, as well as monitor circulating variants in Limpopo Province, South Africa.

3.1.4) Specific Objectives

a) To determine the spatial and temporal trends of SARS-CoV-2 in the Mopani and Vhembe districts in the Limpopo Province, South Africa, through wastewater-based surveillance.
b) To estimate the prevalence of SARS-CoV-2 in the Mopani and Vhembe districts.
c) To determine SARS-CoV-2 variants of concern circulating in the Mopani and Vhembe districts.

3.2) STUDY DESIGN, MATERIALS AND METHODS

3.2.1) Ethical Considerations

The investigation reported here falls under an umbrella project approved by the The Animal, Environmental and Biosafety Research Ethics Committee (AEBREC) at the University of Venda (SMNS/20/MBY/14/0903). Permission to access wastewater treatment plants and waste sedimentation ponds in the Vhembe and Mopani districts of the Limpopo Province of South Africa, was obtained from the authorities of the respective districts.

3.2.2) Study Design and Study Sites

Wastewater samples were collected from seven wastewater treatment plants (WWTPs) and one waste sedimentation pond (WSP) in the Vhembe (Thohoyandou, Malamulele Siloam, Louis Trichardt) and Mopani (Tzaneen, Giyani, Kgapane, Nkowankowa) districts. These samples were collected at the raw inlet of each WWTP after the coarse screen grit chambers; these grit chambers filter all coarse materials coming into the plant. These municipal WWTPs and WSP which primarily treat domestic wastewater were selected based on their functionality, accessibility and feasibility to collect repeated sampling based on resources available (**Figure 14**).

Wastewater samples were collected from three municipalities in the Vhembe District. The Thohoyandou WWTP is situated in Thohoyandou, in the Thulamela municipality which is home to the University of Venda and has approximately 575,929 inhabitants. This WWTP treats an average of 6 megalitres per day (ML/d) of wastewater coming from surrounding catchment sites. Louis Trichardt is a bustling metropolitan area in the Makhado municipality, with a headcount of about 502,397 people, per the national South African 2022 census. The Louis Trichardt WWTP treats approximately 5 ML/d of wastewater from the communities connected to its sewer system. The Malamulele WWTP is situated in Malamulele town, located within the Collins Chabane municipality, in the Vhembe District with approximately 443,798 inhabitants. This WSP treats an average of 2.5 ML/d. The Siloam WSP treats an average of 20.3 ML/d, and is collected from the Siloam Hospital, situated in the Makhado Municipality; it has about 306 hospital beds, and caters for approximately 16,991 people. However, not all inhabitants of these municipalities have access to sewer systems.

In the Mopani District, wastewater is collected from three major municipalities (Greater Tzaneen, Greater Letaba and Greater Giyani). The Tzaneen and Nkowankowa WWTPs are both located in the Greater Tzaneen Municipality, which is home to about 478,251 people, per the national South African 2022 census. The Tzaneen and Nkowankowa WWTPs treat an average of 24 ML/d and 4.5 ML/d of wastewater, respectively. The Kgapane WWTP treats an average of 5.7 ML/d, and it is situated in Ga-Kgapane, in the Greater Letaba Municipality, with a population of approximately 261,038. The Giyani WWTP is situated in the Greater Giyani Local Municipality, and treats an average of 2.1 ML/d, with a population size of 316,841 in this municipality (Department of Water and Sanitation, South Africa, 2022). Like the Vhembe District, not all inhabitants of these municipalities have access to sewer systems.

At the commencement of the study, wastewater was collected from three sites in the Vhembe District only: Thohoyandou WWTP, Siloam WSP and Malamulele WWTP. This continued for a duration of 13 weeks i.e., from 18 January 2021 till 12 April 2021. However, from 19 April 2021, permission to establish a wastewater surveillance system for monitoring SARS-CoV-2 trends in communities in the Mopani District (Tzaneen WWTP, Nkowankowa WWTPs, Giyani WWTP, and Kgapane WWTP), as well as one additional municipality in the Vhembe District (Louis Trichardt WWTP) was obtained. Thus, sample collection continued for the eight sites throughout the testing period.

A bottle of untreated influent wastewater grab sample (500mL) was collected at the raw inlet from each of the sites once every week on a Monday over an 18-month period (Epidemiological [Epi] calendar Week-03 of year 2021 to Epi Week-26 of year 2022 i.e., 18 January 2021 to 28 June 2022). Samples from each site were assigned specific codes, and these codes were maintained for the duration of the study. Samples were then transported to the laboratory in a 4°C fridge and processed for extraction of total RNA, upon arrival.

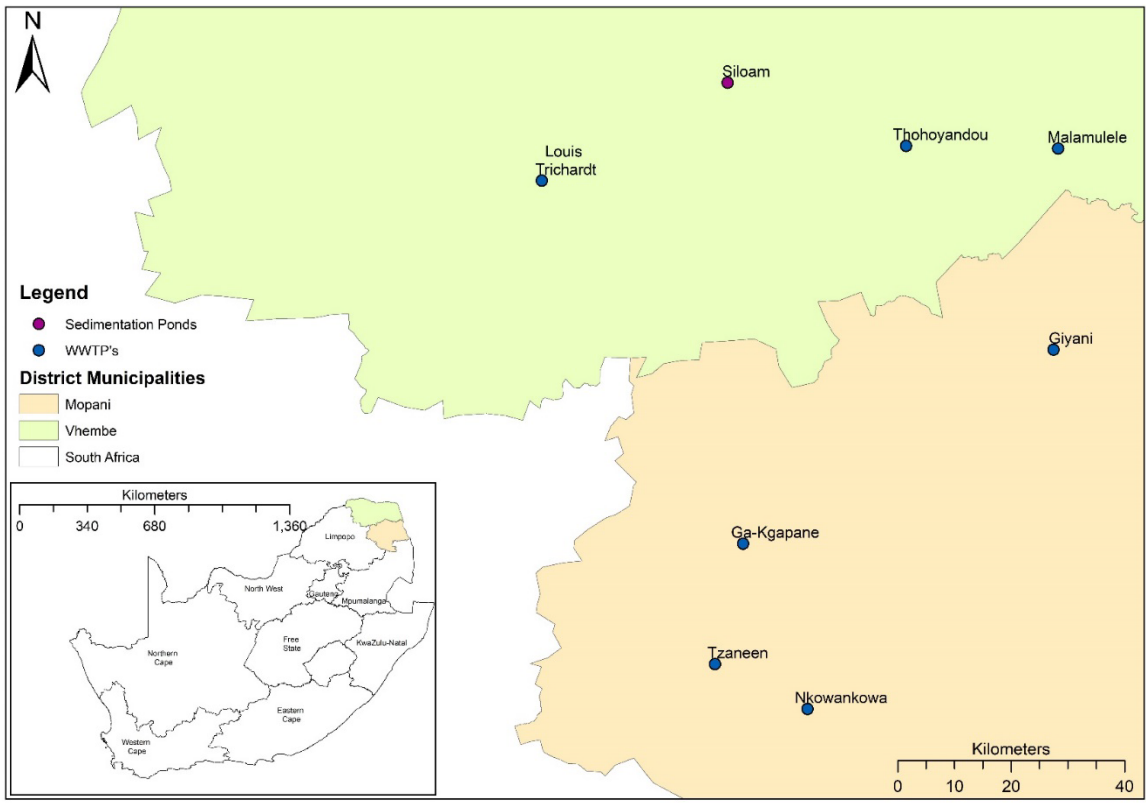


Figure 14: Map of South Africa indicating the wastewater treatment plants and waste sedimentation ponds in the Vhembe and Mopani districts used in the current study.

3.2.3) Extraction of Total RNA from Wastewater Samples

The extraction procedure applied in this study was adapted from a previously published study (Peccia et al., 2020; Johnson et al., 2021). Upon arrival of the wastewater samples to at the laboratory, the following procedures were done to prepare the sample for extraction: depending on the turbidity of the sample, 50 – 300 mL of the sample was aliquoted into 50mL tubes and centrifuged at 2500g for 20 minutes. After centrifugation, the supernatant was discarded, and the resultant pellet (~ 5mL) was taken for downstream extraction.

Total RNA was extracted from the obtained pellet from each sample using a modified protocol of the QIAGEN POWERSOIL total RNA extraction kit (QIAGEN RNeasy ® PowerSoil ® Total RNA Kit Handbook, 2017). Briefly, 5mL of the pellet was added to the 15mL PowerBead tube, which contains silica carbide beads. Next, a lysis buffer, as well as other disruption and precipitation reagents within the kit were added to permit complete cell lysis. Phenol chloroform was also added to maximize lysing efficiency and yield, as well as to denature any proteins in the mixture, but leaving the nucleic acid in solution. Subsequently, the upper

aqueous phase obtained was transferred to a new 15mL tube, and with the addition of a series of chemical agents provided in the kit, the nucleic acid was precipitated, resuspended and purified using the RNeasy JetStar Mini Column. Finally, the purified total RNA was eluted from the column, the pellet airdried, and resuspended 50µl of nuclease-free water (provided in the kit). The resultant total RNA was quantified using a NanoDrop Spectrophotometer to determine the concentration and evaluate the purity of the sample. To ensure that only quality samples were used for downstream experiments and to eliminate any doubts about the results obtained, a quality control (QC) step was implemented. This involved including only samples with total RNA concentration ≥100ng/µl and absorbance ratios (A260/280 and A260/230) ranging between 1.6 – 2.1 for determination of the SARS-CoV-2 RNA copies in each extracted sample.

3.2.4) Quantitative Real-Time Polymerase Chain Reaction for SARS-CoV-2 Amplification

Detection of SARS-CoV-2 in wastewater samples was performed through a quantitative real-time polymerase chain reaction (qRT-PCR), using primer and probe sets approved by the Centres for Disease Control and Prevention (CDC) (Peccia et al., 2020; Sherchan et al., 2020). This primer/probe set targets two distinct regions of the SARS-CoV-2 Nucleocapsid gene (N-gene; N1 and N2) and aligns to the N-gene of the Wuhan SARS-CoV-2 strain (**Figure 15**). The N1 and N2 primer/probe set was purchased from Whitehead Scientific (2019-nCov CDC EUA Kit, Integrated DNA Technologies, USA), and was used for detection of SARS-CoV-2 viral titers in the samples.

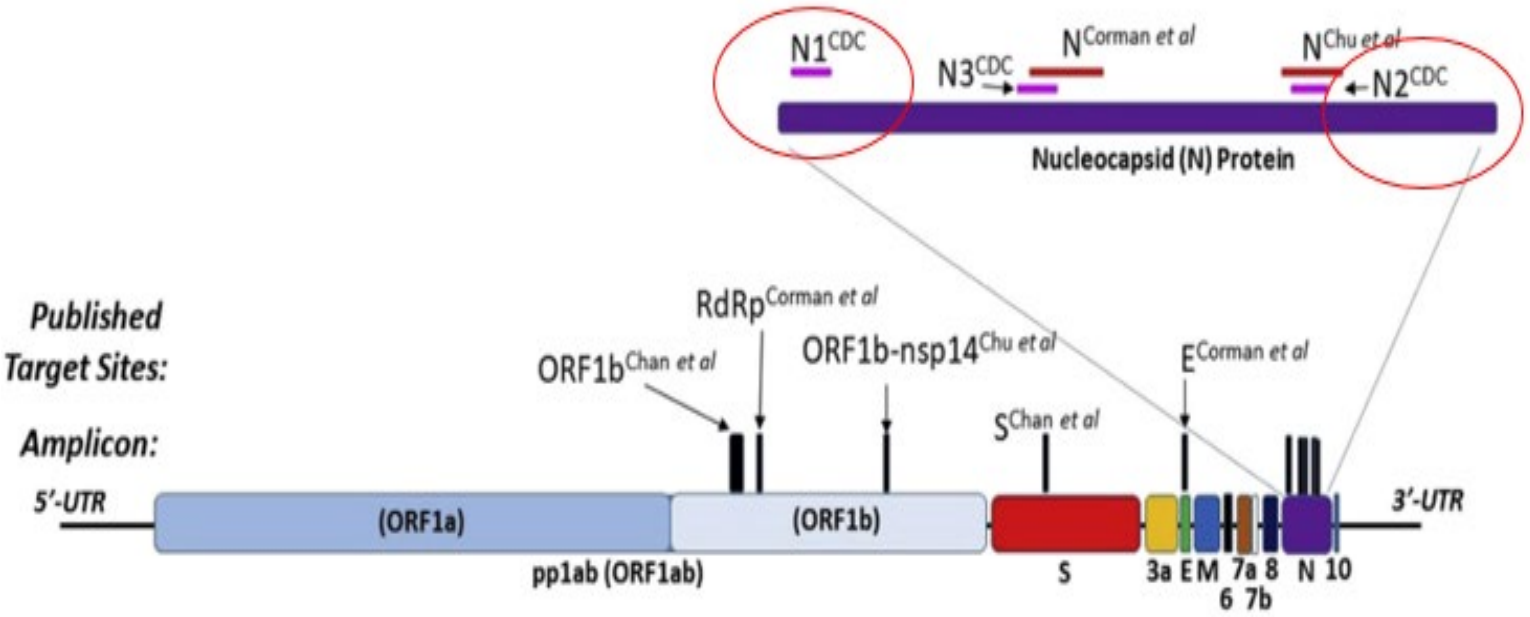


Figure 15: Nucleocapsid gene targeted for molecular detection of SARS-CoV-2, using the N1 and N2 primers (circled portions) and probes (Adapted from Ward et al., 2020).

In addition to the N1 and N2 primer/probe mix, qualitative and quantitative detection of SARS-CoV-2 was achieved through a one-step qRT-PCR using the iTaq Universal probes reaction mix one-step reaction kit, as per the manufacturer's protocol (Bio-Rad Laboratories, Richmond, CA, USA). The master mix reaction mixture consisted of 5µl of iTaq Universal probes reaction mix, 0.25µl of iScript advanced reverse transcriptase, 0.25µl of N1 and N2 Primer/Probe, and 3.25µl of Nuclease free water. A total of 9µl of each master mix (N1 and N2 prepared separately) was added to the PCR plate. Next, all extracted samples which passed QC were standardized to a concentration of 0.2µg/µl and 1µl of RNA was added to each well of the PCR plate, and a negative template control (NTC) was also added for control purposes. To determine the initial starting amount of the target template, the standard curve method was used. This involved doing a 10-fold serial dilution using the SARS-CoV-2 positive control plasmid (2019-nCoV_Positive Control, purchased from Whitehead Scientific), which is supplied at 200,000 genome copies/µl (g.c./µl). Thus, the standard curve concentrations ranged from 20 – 200,000 copies/µl. The SARS-CoV-2 positive control plasmid was also diluted to 200 g.c./µl and included as a positive control for the reaction. All reactions were performed in duplicates, and this was run as a multiplex reaction in the QuantStudio™ 5 Real-

Time PCR System, using the specified cycling conditions. The sequences of the N-gene primers/probes are shown in **Table 2**, while the cycling conditions used for amplification are shown in **Table 3**.

Table 2: SARS-CoV-2 primer-probe sequences and references for its amplification

Organism	**Target**	**Assay Name**	**Target**	**Sequence (5' – 3')**	**References**
SARS-CoV-2	N-protein	2019-nCoV	N1 primer/probe	**F:** 5′-GAC CCC AAA ATC AGC GAA AT-3′ **R:** 5′-TCT GGT TAC TGC CAG TTG AAT CTG-3′ P-FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1	Peccia et al. 2020 Sherchan et al. 2020
			N2 primer/probe	**F:** 5′-TTA CAA ACA TTG GCC GCA AA-3′ **T:** 5′-GCG CGA CAT TCC GAA GAA-3′ P-FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1	

Table 3: Cycling conditions for SARS-CoV-2 Amplification

Real-Time PCR System	Setting	Reverse Transcription Reaction	Amplification	Cycles
QuantStudio 5	Standard	50°C for 10 mins 95°C for 3 mins	95°C for 15 secs 60°C for 1 min	40

3.2.5) Quality Control for RT-qPCR Amplification of SARS-CoV-2

Quality checks for all RT-qPCR amplification of SARS-CoV-2 using the above-described procedure were done for accuracy and reproducibility purposes. This involved inclusion of positive and negative control in each run, as well as analysis of three key components of the standard curve. These key components include: the slope, PCR efficiency and correlation coefficient, which evaluate the quality of the run. The slope of the amplification reaction measures the efficiency of the reaction; accurate and reproducible results should have a slope range of -3.3 to -3.6, which is equivalent to a reaction having an efficiency as close to 100% as possible. The PCR efficiency should be as close to 100% as possible, and this means the template doubles after each thermal cycle during exponential amplification; it always corresponds to the slope obtained by the reaction. The correlation coefficient (R2) measures how well the data fits the standard curve and reflects the linearity of the standard curve; R2 should be as close to 1 as possible. A run which passes QC therefore is one which meets the standard curve criteria, and has amplification of the positive control, but no amplification for the negative control.

3.2.6) Statistical Analysis of SARS-CoV-2 RT-qPCR Amplification Run

Results obtained after completion of the run were analyzed using the QuantStudioTM Design and Analysis software v2.5.1 (Applied Biosystems). Analysis involved setting the amplification threshold (Ct) in the linear phase, ranging from 0.02 – 0.04 to ensure that all samples amplified above the set threshold. Next, the standard curve QC was evaluated by examining the correlation coefficient, PCR efficiency, and slope which always ranged at >0.99, 90 – 100% and -3.3 to -3.6, respectively. A run which passes QC will then be further analyzed to determine the SARS-CoV-2 genome copy number in samples with positive amplification. This

was done using the Shapiro–Wilk test to test the normality of SARS-CoV-2 genome copies per milliliter (g.c./mL).

3.2.7) Genotyping Analysis for Mutation Detection of SARS-CoV-2 Variants of Concern (VOC)

To determine the variants circulating in the communities during the time of surveillance, genotypic analysis through an allele-specific RT-qPCR was performed for mutations pertaining to the Spike gene (S-gene) of SARS-CoV-2. Genotyping analysis, also known as allele-specific genotyping (ASG) is a method used to detect single nucleotide polymorphism (SNP) variants of a target nucleic acid sequence. The preformulated TaqMan™ SNP Genotyping Assays, was used for SARS-CoV-2 variant determination. This assay includes the following components: two sequence-specific primers for amplification of sequences containing the SNP of interest and two allele-specific TaqMan™ probes for Allele 1 and Allele 2 (QuantStudio™ Design and Analysis Software v2 Genotyping Analysis Module User Guide).

For this study, SNP genotyping was done for some signatory mutations belonging to the Alpha, Beta, Delta and Omicron VOCs. In total, 7 TaqMan SARS-CoV-2 Mutation Panels, as well as the TaqPath™ one-step RT-qPCR master mix assay, purchased from ThermoFisher Scientific (Applied Biosystems) were used. These mutation panels include: K417N, L452R, P681H, N501Y, DelH69V70, E484K, and P681R; a combination of these mutations which code for the VOCs are listed in **Table 4**. The SNP genotyping reaction was run in duplicates and the reaction mix consisted of: 2.5µl of TaqPath™ one-step RT-qPCR master mix, 0.25µl of the TaqMan SARS-CoV-2 Mutation Panel (prepared separately for each mutation), and 5µl of SARS-CoV-2 positive sample with viral load ≥1500 g.c./mL. The qRT-PCR reaction was performed in the QuantStudio™ 5 instrument (ThermoFisher Scientific; Applied Biosystems), using thermal cycling conditions specified in **Table 5**.

Table 4: TaqMan SARS-CoV-2 Mutation Panels and combination of mutations which code for VOCs

Organism	Target	TaqMan SARS-CoV-2 Mutation Panel	Signatory Mutations for VOCs
SARS-CoV-2	Spike Protein	S.K417N.AAG.AAT	**Alpha:** N501Y, P681H, delH69V70
		S.L452R.CTG.CGG	
		S.P681H.CCT.CAT	**Beta:** N501Y, E484K, K417N

		S.N501Y.AAT.TAT	**Delta:** L452R, P681R **Omicron:** N501Y, P681H, delH69V70, K417N
		S.delH69V70	
		S.E484K.GAA.AAA	
		S.P681R.CCT.CGT	

Table 5: Cycling conditions for SARS-CoV-2 allele-specific genotyping analysis

Real-Time PCR System	**Setting**	**Pre-Read**	**Reverse Transcription Reaction**	**Amplification**	**Post-Read**	**Cycles**
QuantStudio 5	Standard	60°C for 30 secs	50°C for 10 mins 95°C for 3 mins	95°C for 3 secs 60°C for 30 secs	60°C for 30 secs	45

3.2.8) Analysis of SARS-CoV-2 SNP Genotyping RT-qPCR Run

Results obtained after completion of the run were analyzed using the QuantStudioTM Design and Analysis software v2.5.1 (Applied Biosystems). During genotyping analysis, the software normalizes the fluorescence of the reporter dyes to the fluorescence of the passive reference dye in each well, plots the normalized reporter dye signal of each sample well on an Allelic Discrimination Plot, which contrasts the reporter dye intensities of the allele-specific probes, then algorithmically clusters the sample data, and assigns a genotype call to the samples of each cluster according to its position on the plot (QuantStudio™ Design and Analysis Software v2 Genotyping Analysis Module User Guide).

In the quality control tab of the software, the amplification threshold (Ct) was set to linear phase, ranging from 0.02 – 0.09 to ensure that only samples that amplified above the set threshold were assigned. In the genotyping tab, the allelic discrimination plot displayed assigns samples as amplified for the investigated mutation i.e. homozygous (Allele 1 or Allele 2) or heterozygous (Allele 1/2); in addition, the allelic discrimination plot also indicates whether there is no amplification in a sample, by assigning it as “No amplification” or “Undetermined”. For samples that have a “No amplification” assignment, this means that the mutation was not present. For samples whose duplicates have an “Undetermined” assignment, the reaction was repeated; however, in samples where there was amplification in at least one duplicate for the tested mutation, this was considered as positive for the investigated mutation. **Table 6**

shows signatory mutations used for variant detection alongside the criterion used for variant calling is shown.

Table 6: WHO designated VOC, and their associated TaqMan mutation panels used for investigating their presence, and the criterion for variant calling.

WHO Designated VOC	Associated Mutations for TaqMan Panels	Criterion for Variant Calling
Alpha	delH69V70, N501Y, P681H	Presence of at least two mutations
Beta	E484K, K417N, N501Y	Presence of at least two mutations
Delta	L452R, P681R	Presence of at least one mutation
Omicron	delH69V70, N501Y, P681H, K417N	Presence of K417N and any other mutation

3.2.9) Prediction of the Number of People Infected

The prediction model published by Ahmed et al., (2020) was used to estimate the number of people infected in communities connected to the WWTPs. This model utilizes four main variables to determine the number of people infected. This includes: the number of RNA copies per litre of wastewater, the liters of wastewater received per day into the treatment plant, the grams of feces released per person per day, and the number of RNA copies present per gram of feces released per person. Calculation of the number of people infected with SARS-CoV-2 in each community was done using the Monte Carlo (MC) simulation method. MC simulation is a type of simulation that relies on repeated random sampling and statistical analysis to compute results. The simulation model was run in Microsoft Excel using 10,000 iterations.

Component one (the number of RNA copies per liter of wastewater) was obtained from the average SARS-CoV-2 copies/mL determined by the weekly surveillance. Component two (the liters of wastewater received per day into the treatment plant) was obtained from onsite records of each WWTP and WSP. Component three (the grams of feces released per person per day) was obtained from (Pillay et al., 2021) and modelled as a normal distribution with a mean of 2.07 log10 and a standard deviation of 1.08 $\log^{10}$. The fourth component (the number of RNA copies present per gram of feces released per person) was obtained from (Wölfel et al., 2020) and modelled as a log-uniform distribution with a maximum of 7.67 and a minimum of 2.56.

For weekly prevalence estimates to be calculated, only samples positive for SARS-CoV-2 RNA (>0 copies/mL) were used for analysis. To ensure precision of the output value, the standard deviation (SD) for the average SARS-CoV-2 copies/mL (N1 and N2 average copies) of each sample was calculated. Only samples whose SD was less than one (<1) were used for MC simulation.

Basic statistical analysis such as determination of the Mean, Median, SD, Variance, Skewness, Kurtosis, Coefficient of Variability, Minimum, Maximum, Range Width, and Mean Standard Error of each output was performed to ensure precision and minimize variability. The Maximum value obtained for each sample was then documented and used to plot graphs for each epidemiological week for individual communities. Summary statistics were reported as the median and 95% confidence interval (CI), since the median is less sensitive to extreme values from the input distributions. Basic statistical analysis was performed to ensure precision and minimize variability. The maximum value obtained for each sample was then documented and used to plot graphs for each epidemiological week for individual communities.

3.2.10) Comparison of COVID-19 Clinical Cases to SARS-CoV-2 RNA Levels in the Vhembe and Mopani Districts

Data on the number of people who presented with COVID-19 symptoms and had a confirmed SARS-CoV-2 positive laboratory diagnosis for the duration of the 18-month surveillance period was obtained from the Limpopo Department of Health, South Africa. Using Microsoft Excel, Pivot tables were used to extract data for the sub-districts where the sites are located. In addition, the same software was also used to calculate the 7-day moving average (7D MA) of laboratory-diagnosed COVID-19 cases at sub-district level. This was then used in comparison with the trends of SARS-CoV-2 RNA signals determined through wastewater surveillance.

3.2.11) Comparison of COVID-19 Clinical Cases to Normalized SARS-CoV-2 RNA Levels in the Vhembe and Mopani Districts

SARS-CoV-2 viral RNA obtained throughout the surveillance period was normalized using parameters (flowrate and population statistics) from each site. Firstly, the SARS-CoV-2 daily viral concentration (g.c./day) was normalized with the flowrate using equation 1 (Eq. 1), where

C_{virus} (g.c./L) was multiplied by the daily flow rate (Q, ML/day). Next, SARS-CoV-2 daily viral concentration was normalized to population estimates using equation 2 (Eq. 2), where genome copies per litre (g.c./L), flow rate (Q, ML/day) was normalized to population estimates (PE) census data and multiplied by 1000 inhabitants (Mangwana et al., 2022).

$$SARS - CoV - 2\left(\frac{GC}{day}\right) = \frac{GC}{L} \times Q\left(\frac{ML}{day}\right) \quad \text{(Eq. 1)}$$

$$SARS - CoV - 2_{census}\left(\frac{GC}{day}\right) = \frac{\frac{GC}{L} \times Q\left(\frac{ML}{day}\right)}{PE} \times 1000\ inh \quad \text{(Eq. 2)}$$

The daily flowrate was obtained from each WWTP and WSP for the duration of the surveillance period. However, due to frequent power outages experienced at the sites, daily flowrate data for some days were missing. Thus, to compensate for such events, the average monthly flowrate was computed and used. Population estimates determined by the 2022 national census per sub-district (Department of Statistics South Africa, 2023), was used for computation of the second formula. **Table 7** shows the estimated population for each sub-district where the WWTPs and WSP are located. The normalized SARS-CoV-2 viral RNA trends per site were then compared to the laboratory-diagnosed COVID-19 clinical cases (7D-MA) obtained per sub-district.

Table 7: Sub-district population per the 2022 national South African census used for normalization computation.

WWTP	Municipality	Sub-district Population (2022 National Census)	District
Thohoyandou	Thulamela	575,929	Vhembe
Malamulele	Collins-Chabane	443,798	
Louis Trichardt	Makhado	502,397	
Siloam Hospital	Makhado		
Tzaneen	Greater Tzaneen		

Nkowankowa	Greater Tzaneen	478,251	Mopani
Ga-Kgapane	Greater Letaba	261,038	
Giyani	Greater Giyani	316,841	

3.2.12) Comparison of COVID-19 Clinical Cases to Estimated Prevalence in the Vhembe and Mopani Districts

The 7D MA of clinical cases calculated for the whole surveillance period was used for comparison with the estimated prevalence determined by the MC mathematical model. Since clinical data was not stratified per health clinic for individual sites, the 7D MA clinical data obtained at the sub-district level was used as a proxy for the comparison. In the Vhembe District, the Louis Trichardt and Siloam sites are situated in the Makhado sub-district, while the Thohoyandou and Malamulele sites are situated in the Thulamela and Collins Chabane sub-districts, respectively. For the Mopani District, the Tzaneen and Nkowankowa sites are situated in the Greater Tzaneen sub-district, while the Ga-Kgapane and Giyani sites are situated in the Greater Letaba and Greater Giyani sub-districts, respectively.

3.2.13) Statistical Analysis

To investigate the accuracy of obtained viral loads from N1 and N2 primers used throughout the surveillance period, the Spearman correlation coefficient (r) and p-value were calculated. This showed the correlation between viral loads obtained from both primers. Similar statistical analysis was done to determine the correlation between normalized and non-normalized wastewater data and clinical data.

3.2.14) Definition of Key Terms in the Study

Based on weekly viral load concentrations observed in the study sites, warning of an incoming wave for this study, was taken to be a ≥0.2 Log10/mL (500 – 2500 g.c/mL) increase in SARS-CoV-2 RNA concentration sustained for two weeks. The end of a wave was taken as detection of viral concentration <2.7 Log10/mL. Increases in clinical cases were taken to be a 0.5 Log10 increase observed over a 7D MA sustained for two weeks. Conversely, a value of <1.0 Log10 clinical cases reported over a 7D MA was taken to be the end of a wave. The average SARS-CoV-2 viral load for each wave was also calculated to evaluate the potential impact of each wave. To estimate the magnitude of each wave (potential degree of infections in the

community), SARS-CoV-2 RNA copies were grouped into five categories, and designated as signals (**Figure 16**). A wave with average SARS-CoV-2 copies detected being in category 1 (0 – 500 g.c/mL) was taken to have the least magnitude (number of infected people). Whereas a wave with average SARS-CoV-2 copies detected being in category 5 (≥7501 g.c/mL) was taken to have the highest magnitude. This categorization was adapted for better understanding and interpretation for the municipality public health personnel.

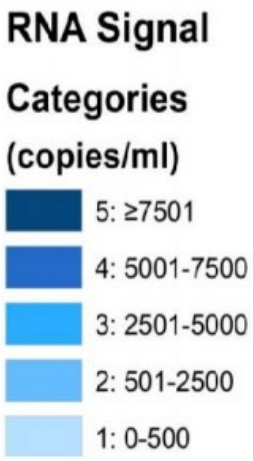


Figure 16: SARS-CoV-2 RNA signal categories developed to evaluate the magnitude of COVID-19 infections in study sites. Adapted from the South African Medical Research Council (SAMRC) COVID-19 and wastewater early warning system team (www.samrc.ac.za/wbe)

3.2.15) COVID-19 Waves in South Africa Throughout the Surveillance Period

According to the National Institute for Communicable Diseases (NICD), a wave documents daily occurrences of COVID-19 across South Africa. Throughout the COVID-19 pandemic, they documented four periods when the country experienced increased transmission from person to person (known as waves). Thus, the proposed definition of a wave was defined as "the period from when COVID-19 weekly incidence is ≥30 cases per 100,000 individuals and ends when the weekly incidence decreases to ≤30 cases per 100,000 persons" (NICD, 2021). As per the NICD weekly reports, the first wave occurred from 15 June – 24 August 2020; the second wave was from 23 November 2020 – 01 February 2021; the third wave from 10 May – 13 September 2021, and the fourth wave from 06 December 2021 – 14 February 2022. This wastewater surveillance system was set up towards the end of the second wave Epi Week-03, 2021 (18 January 2021), and trends of SARS-CoV-2 occurrences in the eight sites in

Vhembe and Mopani districts were monitored for 18 months. The NICD also reported general genomic diversity circulating in the population from SARS-CoV-2 positive samples genotyped by the Network for Genomic Surveillance in South Africa (NGS-SA) (NGS-SA, 2022).

3.3) RESULTS

3.3.1) Wastewater-based Surveillance Trends of SARS-CoV-2 RNA Levels and Variants of Concern Occurrence

Weekly monitoring of SARS-CoV-2 occurrence in three of the four sites in the Vhembe District commenced from Epi Week-03, 2021 (18 January 2021). In the Louis Trichardt site (in the Vhembe District), and all four sites in the Mopani District, sample collection commenced from the Epi Week-16, 2021 (19 April 2021). Surveillance of all eight sites continued till Epi Week-26, 2022 (28 June 2022). SARS-CoV-2 concentrations detected in samples are represented as the average from positive amplification using the N1 and N2 primers. Of the 279 samples collected in the Vhembe District throughout the surveillance period, 183/279 (66%) were positive for SARS-CoV-2, with viral RNA concentrations detected ranging between 2.9 – 142,202.8 g.c./mL (0.6 – 5.2 Log10). In the Mopani District, 240 samples were collected throughout the surveillance period, and 182/240 (75.8%) were positive for SARS-CoV-2, having viral RNA concentrations ranging between 1.2 – 78,190.4 g.c./mL (0.3 – 4.9 Log10).

A) Summary of SARS-CoV-2 Wastewater and COVID-19 Clinical Trends in the Vhembe and Mopani Districts

Throughout the surveillance period, there were three periods when increases in SARS-CoV-2 RNA concentration were observed across all eight sites, indicating the occurrence of a wave. Across all eight sites, the first period which hinted the introduction of the third wave, was observed by Epi Week-16, 2021 (19 April 2021), in the Ga-Kgapane site (Mopani District). Over time, a gradual increase in SARS-CoV-2 viral RNA was seen in the other sites. During this third wave, the average viral load detected in the Mopani district was slightly higher (3307.8 g.c/mL, wave magnitude: category 3) than what was detected in the Vhembe district (2684.9 g.c/mL, wave magnitude: category 3). Thus, the wave magnitude (potential degree of infections in the community), based on this average viral load, indicated that this wave had more impact in the Mopani district compared to the Vhembe district. This third wave, dominated by the Delta VOC, began phasing out in both districts by Epi Week-31, 2021 (02 August 2021), thus lasting for approximately 16 weeks.

The second period, which signaled the beginning of the fourth wave was noted by a gradual increase in viral load detected in Epi Week-46, 2021 (15 November 2021). During this fourth wave, the average viral load detected in the Vhembe district was higher (6134.4 g.c/mL, wave magnitude: category 4) than that in the Mopani district (1792.5 g.c/mL, wave magnitude: category 2). Thus, the wave magnitude based on this average viral load indicated that this wave was more felt in the Vhembe district compared to the Mopani district. This fourth wave was mainly dominated by the Omicron VOC, and this continued for 14 weeks, lasting till Epi Week-08, 2022 (21 February 2022). Towards the end of the surveillance period (June 2022), a spike in viral RNA circulation was observed again in all sites from Epi Week-16, 2022 (18 April 2022). This wave had the shortest occurrence, lasting for 8 weeks, till Epi Week-23, 2022 (07 June 2022), with the Omicron VOC still being in predominant circulation in all sites.

Through allele specific genotyping (ASG), mutations conferring the presence of the Delta and Omicron VOCs were detected as early as Epi Week-18, 2021 (03 May 2021). The presence of these variants was confirmed through whole genome sequencing (WGS). There was concordance in variant assignment for 51.2% of samples evaluated by both ASG and WGS techniques. Analysis of WGS also revealed the earliest circulation of Delta and Omicron VOCs to be January and February 2021 (elaborated in **section 4.3.2**), earlier than what was observed through ASG.

3.3.2) Comparison of SARS-CoV-2 Trends in Wastewater with COVID-19 Clinical Data

Comparison of the wastewater data with clinical data per sub-district showed the efficiency of the set-up surveillance system in signaling the introduction of a new wave, and this was observed approximately one week before increases in the reported clinical cases. Interestingly, spikes in viral RNA circulation observed towards the end of the surveillance period were not translated to increases in reported clinical cases as was characteristic of the previous waves. Details observed per district are described below.

3.3.3) SARS-CoV-2 Trends Compared with COVID-19 Clinical Data in the Vhembe District

In the Vhembe District, a total of 259/479 (54%) samples were collected from the four study sites throughout the surveillance period. Of this total number of samples collected, 168/259

(64.8%) were positive for SARS-CoV-2, and the average SARS-CoV-2 RNA detected for all the surveyed communities throughout the testing period was 6538.6 g.c./mL. However, to accurately compare the trend of SARS-CoV-2 occurrence in the Vhembe District, this was done from Epi Week-16 of year 2021 (19 April 2021) through Epi Week-26 of year 2022 (28 June 2022), since sample collection from the Louis Trichardt WWTP only commenced in Epi Week 16 of year 2021.

A) Summary of SARS-CoV-2 Wastewater and COVID-19 Clinical Trends in the Vhembe District

The third wave was marked by a spike in SARS-CoV-2 concentrations, observed by Epi Week-19, 2021 (10 May 2021), which began in the Thohoyandou site (Thulamela sub-district). This was followed by similar spikes in wastewater viral RNA concentrations in the Malamulele and Louis Trichardt study sites, seen by Epi Week-22, 2021 (31 May 2021), and lastly in the Siloam site (Epi Week-24, 2021). This spikes in wastewater viral RNA concentrations were seen 1 – 3 weeks prior to increases in documented COVID-19 clinical cases. More clinical cases were first documented in the Siloam site by Epi Week-25, 2021, followed by the Thohoyandou, Malamulele, and Louis Trichardt sites three weeks later. This third wave lasted for 10 – 13 weeks in the study sites. The magnitude (potential degree of infections) of this third wave was highest in the Siloam site (category 5), followed by the Malamulele site (category 4). The other two study sites had the lowest wave magnitude (category 2).

The fourth wave was characterized by spiked wastewater viral RNA concentrations seen by Epi Week-46, 2021 (15 November 2021) in both the Thohoyandou and Siloam study sites, which was later observed in the other two study sites in Epi Week-48, 2021. These spikes in wastewater RNA occurred 1 – 3 weeks prior to increases in documented clinical cases. More clinical cases were first documented in the Louis Trichardt site (Epi Week-49, 2021), one week after spikes in wastewater viral RNA concentration was observed. This fourth wave lasted longest in the Siloam site (12 weeks), but with the highest magnitude observed in the Thohoyandou site (category 5).

Towards the end of the surveillance period, a small spike in wastewater RNA concentrations was observed again. This began in the Louis Trichardt site (Epi Week-16, 2021) and was subsequently observed in the Siloam and Malamulele sites by Epi Week-17, 2021, and finally

in the Thohoyandou site (Epi Week-18, 2021). Interestingly, no increase in COVID-19 clinical cases was documented during this last wave observed in wastewater (**Figure 17**). This was the shortest wave (3 – 5 weeks), with all study sites having the same magnitude (category 2). The correlation between viral loads obtained from both N1 and N2 primers used throughout the surveillance period, is shown in **Table 8**. The correlation between normalized and non-normalized wastewater data compared to clinical data is also shown in **Table 8**. All comparisons were statistically significant.

Table 8: A summary of correlation values obtained for comparison of viral loads with N1 and N2 primers, as well as normalized and non-normalized wastewater data compared to clinical data in the Vhembe district study sites.

Study Site	Spearman Correlation between N1 and N2 Viral Loads	P-value of Compared N1 and N2 Viral Loads	Spearman Correlation between Wastewater data and Clinical data	P-value of Wastewater data compared with Clinical data	Spearman Correlation between normalized Wastewater data and Clinical data	P-value of normalized Wastewater data compared with Clinical data
Thohoyandou WWTP	0.773	1e-15	0.41	3e-04	0.41	3e-04
Siloam SP	0.851	2e-15	0.57	<0.0001	0.57	9e-08
Malamulele WWTP	0.556	3e-07	0.56	2e-07	0.56	2e-07
Louis Trichardt WWTP	0.804	1e-14	0.60	3e-07	0.60	3e-07

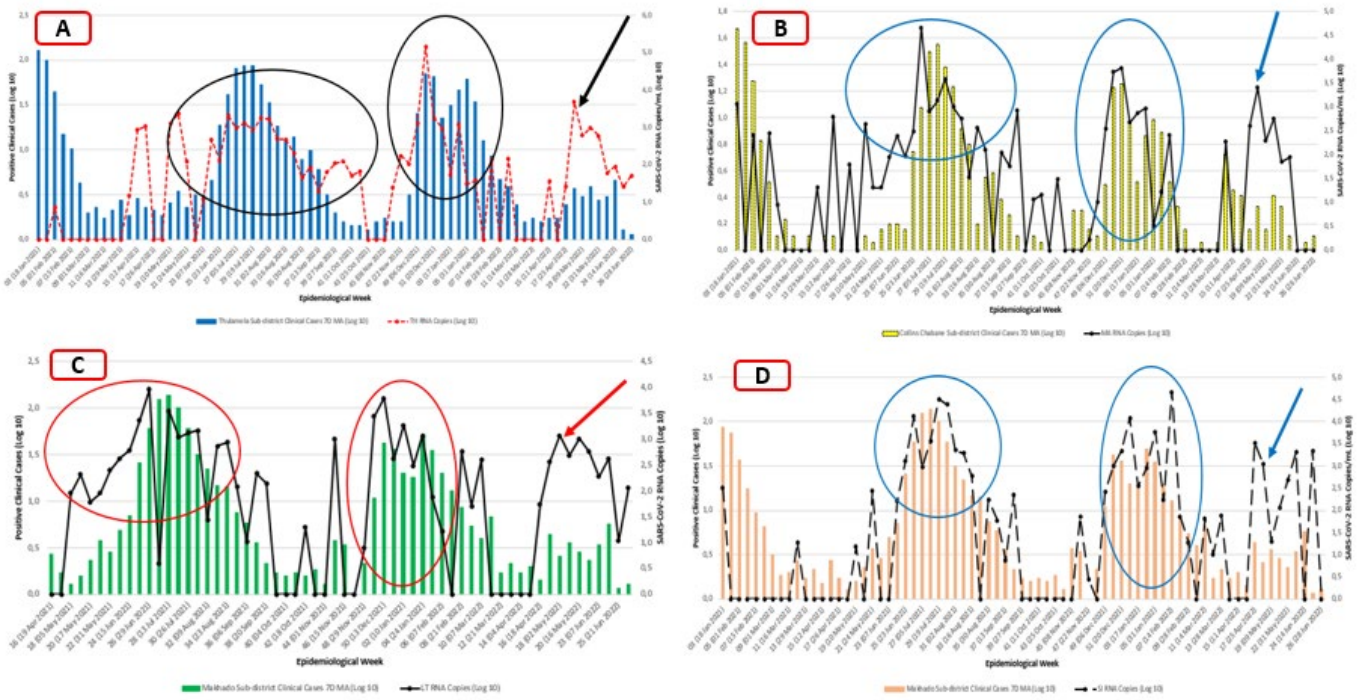


Figure 17: A comparative analysis of SARS-CoV-2 wastewater RNA trends in the study sites compared with the 7D MA of clinical cases in the Vhembe district. **(A)** Thohoyandou site compared to 7D MA clinical cases in the Thulamela sub-district; **(B)** Malamulele site compared to 7D MA clinical cases in the Collins Chabane sub-district; **(C)** Louis Trichardt site compared to 7D MA clinical cases in the Makhado sub-district; and **(D)** Siloam site compared to 7D MA clinical cases in the Makhado sub-district. Spikes in wastewater viral RNA observed towards the end of the surveillance period were not mirrored by reported clinical data.

3.3.3.1) Thohoyandou WWTP

The first signal of a probable incoming wave was noted by 0.2 Log10/mL increase in SARS-CoV-2 RNA concentration observed in Epi Week-19, 2021 (10 May 2021). Three weeks later, in Epi Week-23, 2021 (07 June 2021), there was an increase in reported COVID-19 clinical cases from the Thulamela sub-district, confirming the occurrence of the third wave in this site (**Figure 18 A**). High viral loads were observed continuously for 13 weeks, but began declining by Epi Week-31, 2021 (02 August 2021), with and the average viral concentration observed during this period was 989.6 g.c./mL. Based on the average viral load detected during this period, this wave had a category 2 magnitude (potential degree of infections in the community).

The fourth wave was noted by increases in viral RNA by Epi Week-46, 2021 (15 November 2021). Three weeks later, in Epi Week-49, 2021 (06 December 2021), increases in clinical cases were reported. Over an eight-week period, high viral copies were continually detected, and began declining by Epi Week-04, 2022 (24 January 2022). In the following week, a decline in reported clinical cases was also noted. The average viral load observed during this period was 16,420.1 g.c./mL, which was remarkably higher than what was observed in the third wave. Thus, this wave had the highest magnitude (category 5).

Towards the end of the surveillance period (June 2022), another increase in RNA concentration was observed from Epi Week-18, 2022 (02 May 2022), which continued till Epi Week-22, 2022 (31 May 2022), lasting for four weeks, with an average of 1,718.8 g.c./mL observed. As with the third wave, this wave had a category 2 magnitude. During this last period, no corresponding increases in COVID-19 clinical cases were observed. There was similarity in correlation of non-normalized (**Figure 18**) and normalized wastewater data (**Figure 19 A and B**), when compared to clinical data, all of which were statistically significant (**Table 8**).

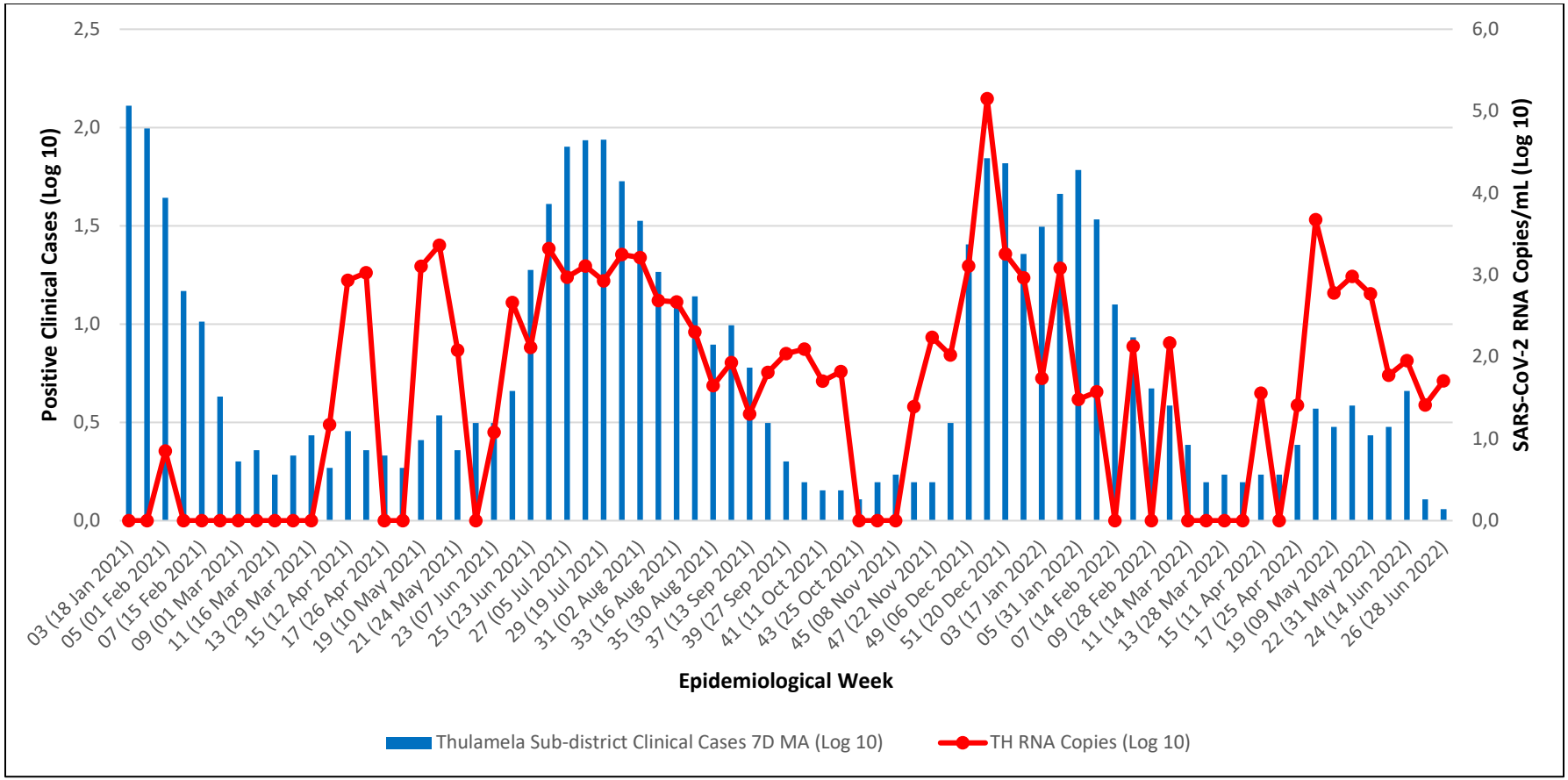


Figure 18: SARS-CoV-2 RNA trends observed throughout the surveillance period in the Thohoyandou (TH) site correlated with the 7D MA of clinical cases in the Thulamela sub-district.

Through WBE, increases in viral RNA copies were observed prior to increases in COVID-19 clinical cases in the third and fourth waves, showing the applicability of the method in this setting.

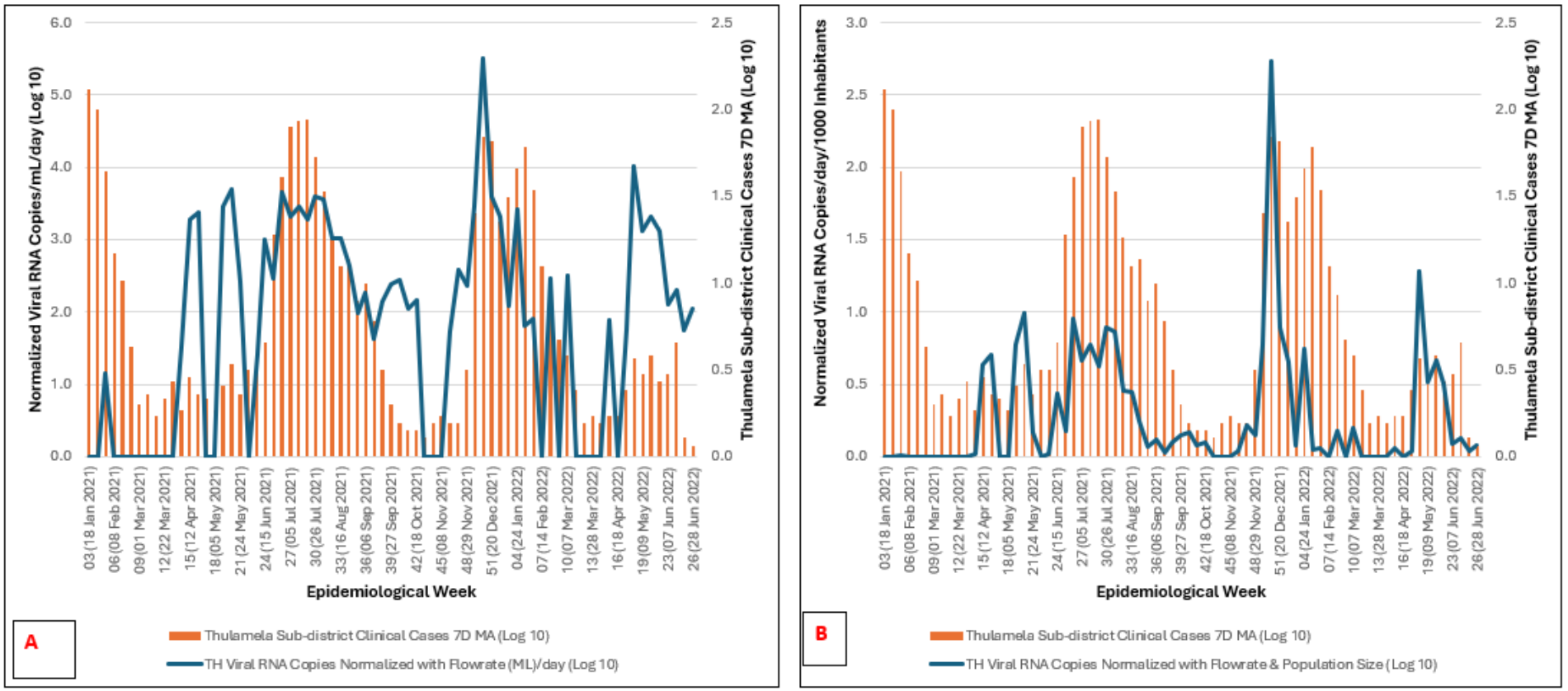


Figure 19 A and B: Normalized SARS-CoV-2 RNA in the Thohoyandou (TH) site correlated with the 7D MA of clinical cases in the Thulamela sub-district.

(B) SARS-CoV-2 viral RNA wastewater data normalized with flow rate and compared to Thulamela sub-district clinical cases.

(C) SARS-CoV-2 viral RNA wastewater data normalized with flow rate and population size and compared to Thulamela sub-district clinical cases.

3.3.3.2) Siloam Hospital SP

This site is also located in the Makhado sub-district like the Thohoyandou site. Wastewater RNA signals were observed throughout the surveillance period, however there were three periods where higher viral copies were detected indicating the commencement of a new wave. Increases in viral concentrations detected from Epi Week-24, 2021 (15 June 2021) indicated the beginning of the third wave in this site. The following week, in Epi Week-25, 2021 (23 June 2021), a sharp increase in COVID-19 clinical cases at the sub-district level were reported, confirming wastewater observations (**Figure 20**). This wave lasted for 10 weeks, i.e. till Epi Week-33, 2021 (16 August 2021), with an average viral load of 8,021 g.c./mL. Based on this average, this wave had a category 5 magnitude for this period.

The start of the fourth wave was marked by increased wastewater RNA signals from Epi Week-46, 2021 (15 November 2021). Three weeks later, by Epi Week-49, 2021 (06 December 2021), increases in reported clinical cases were observed. This fourth wave lasted till Epi Week-07, 2022 (14 February 2022), with an average viral load of 6974.1 g.c./mL (Wave magnitude: category 4) observed throughout the 12-week period. These observations were congruent with the declining clinical cases reported. Towards the end of the surveillance period (June 2022), increases in RNA concentration were observed again from Epi Week-17, 2022 (25 April 2022), which continued till Epi Week-25, 2022 (21 June 2022), lasting for four weeks, with an average of 1,165.2 g.c./mL observed. This last period had the lowest magnitude of infections (category 2), based on the average viral load detected. During this last period, no corresponding increases in COVID-19 clinical cases were observed. There was similarity in correlation of non-normalized (**Figure 20**) and normalized wastewater data (**Figure 21 A and B**), when compared to clinical data, all of which were statistically significant (**Table 8**).

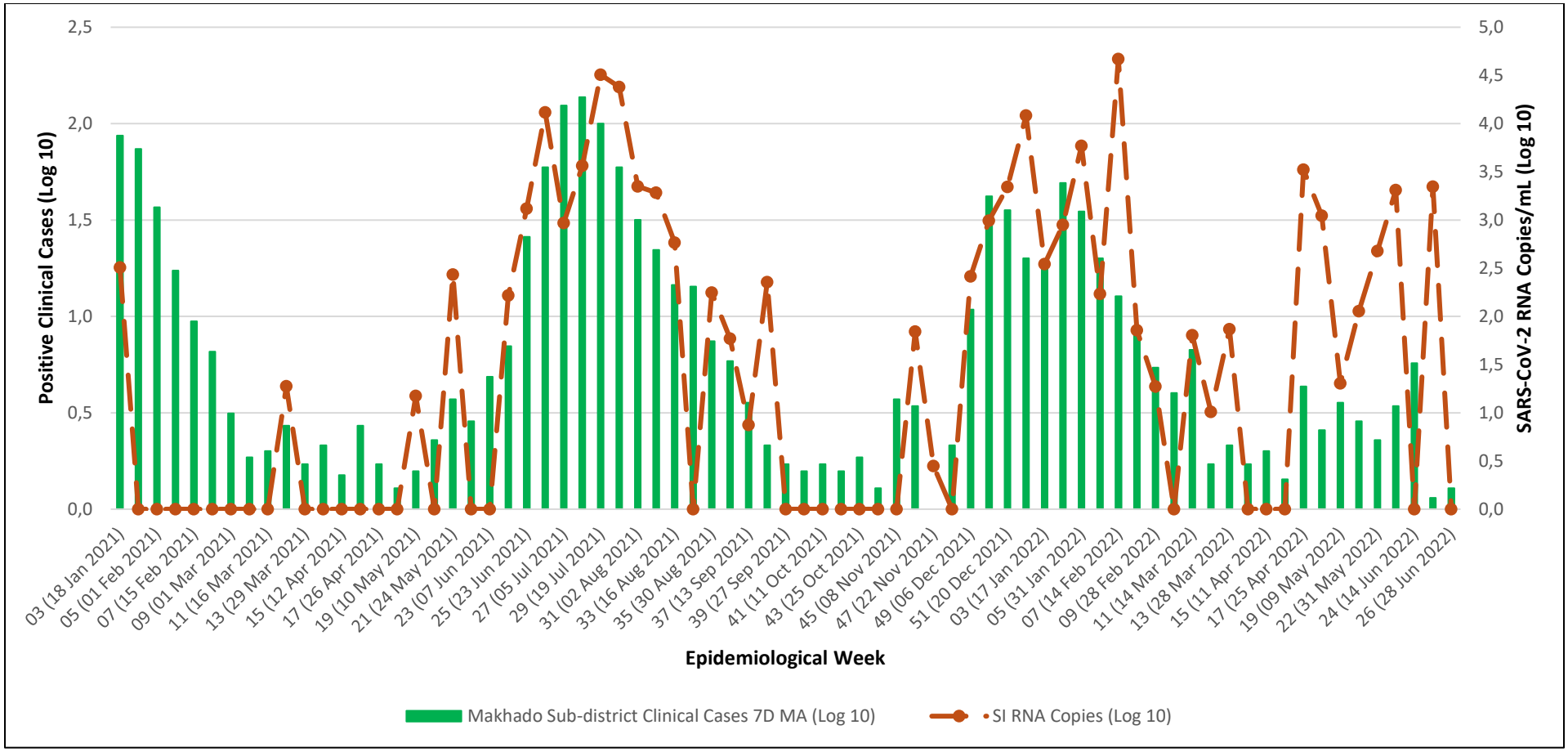


Figure 20: Comparing SARS-CoV-2 trends in the Siloam Hospital (SI) site with 7D MA of clinical cases in the Makhado sub-district.

The wastewater data corroborated with the clinical data throughout the surveillance period, except towards the end of surveillance, where COVID-19 clinical cases remained low, even with increases in viral RNA copies.

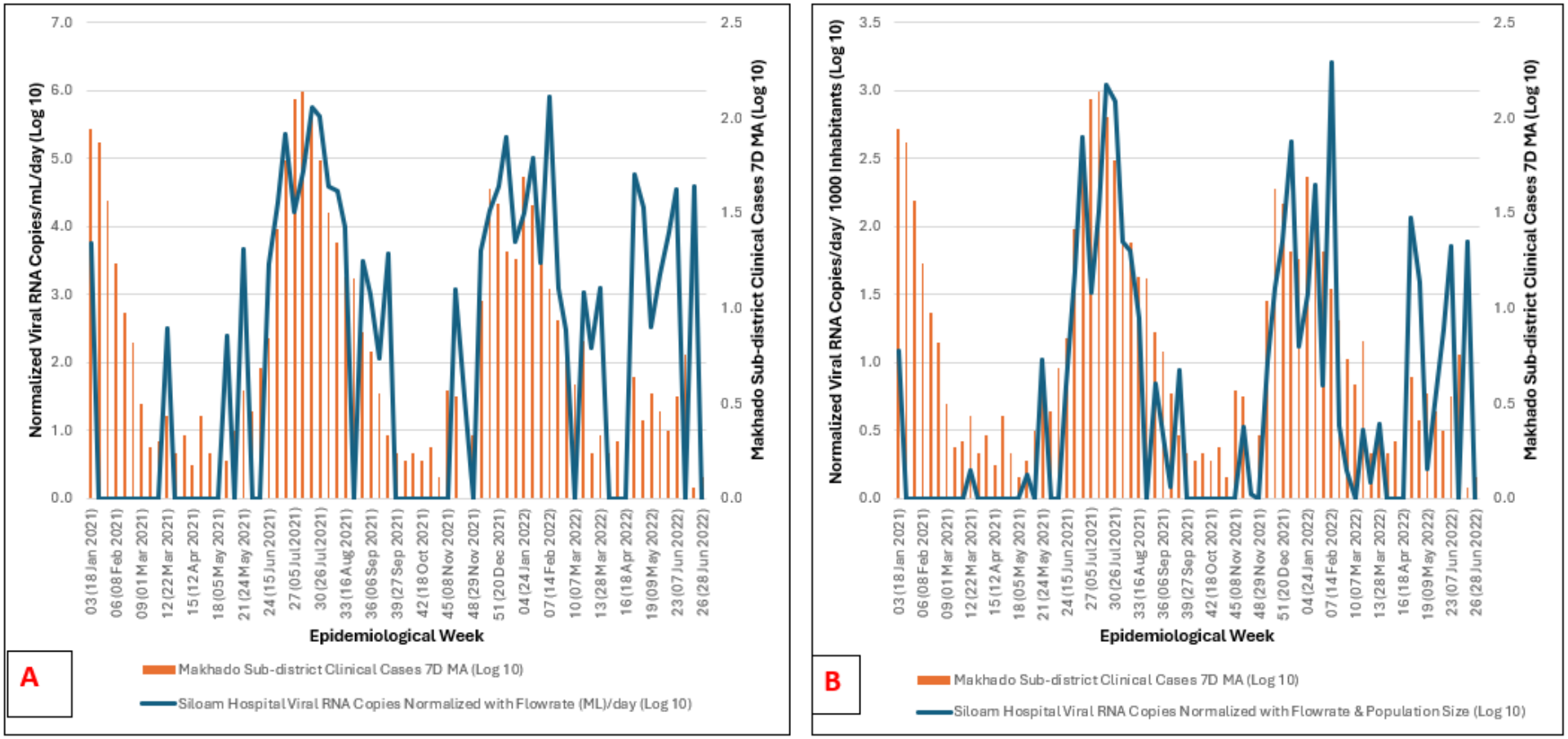


Figure 21 A and B: Normalized SARS-CoV-2 RNA in the Siloam Hospital (SI) site compared with 7D MA of clinical cases in the Makhado sub-district.

(A) SARS-CoV-2 viral RNA detected in wastewater samples normalized with flow rate and compared with the Makhado sub-district clinical cases **(B)** SARS-CoV-2 viral RNA wastewater data normalized with both flow rate and Siloam Hospital estimated population and compared to the Makhado sub-district clinical cases.

3.3.3.3) Malamulele WWTP

From the beginning of surveillance at this site, SARS-CoV-2 RNA concentrations detected had a changeable pattern. However, by Epi Week-22, 2021 (31 May 2021), consistent increases in RNA signals were observed; three weeks later, in Epi Week-25, 2021 (23 June 2021), increases in COVID-19 clinical cases from the Collins Chabane sub-district were reported. This marked the beginning of the third wave in this site (**Figure 22**). This third wave continued for 09 weeks, till Epi week-31, 2021 (02 August 2021), with average viral concentration observed being 5,437.8 g.c./mL (wave magnitude: category 4). As viral RNA concentrations decreased after Epi week-31, 2021, a similar reduction in reported clinical cases were also seen.

Increases in viral RNA signals warning of an incoming fourth wave were seen from Epi Week-48, 2021 (29 November 2021). One week later, in Epi Week-50, 2021 (13 December 2021) increases in clinical cases were observed. This wave lasted for six weeks, till Epi Week-04, 2022 (24 January 2022), with average viral RNA detected being 2,075.9 g.c./mL. The magnitude of this fourth wave was lesser (wave magnitude: category 2) compared to the previous wave

An increase in viral RNA was detected again from Epi Week-17, 2022 (25 April 2022), which continued till Epi Week-20, 2022 (16 May 2022), lasting for three weeks, with average viral concentrations observed to be 601.7 g.c./mL. As with the fourth wave, a category 2 magnitude was observed, based on the detected average viral load. As with the other sites in this district, no increase in COVID-19 clinical cases were reported during this period. There was similarity in correlation of non-normalized (**Figure 22**) and normalized wastewater data (**Figure 23 A and B**), when compared to clinical data, all of which were statistically significant (**Table 8**).

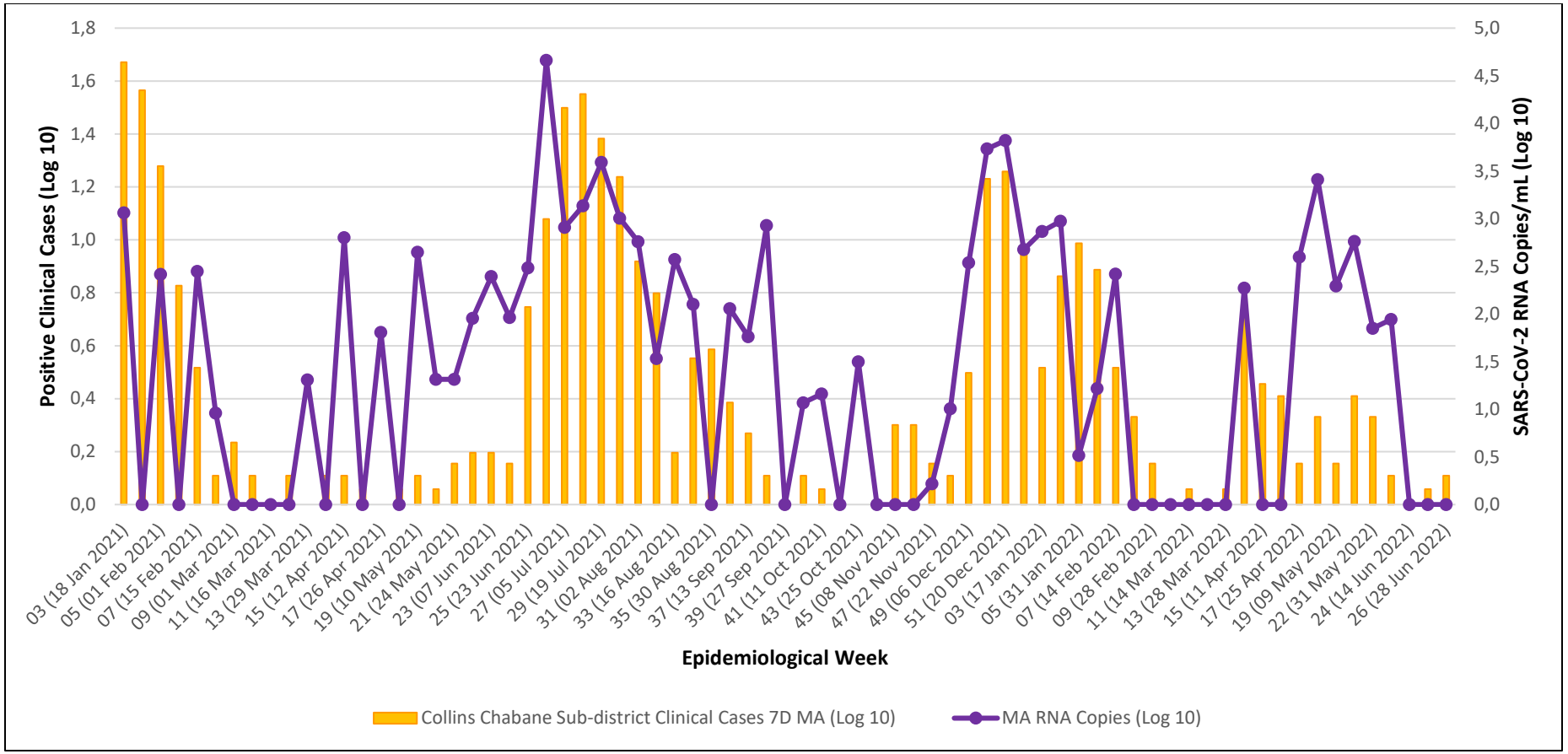


Figure 22: SARS-CoV-2 trends in the Malamulele (MA) site compared to a 7D MA of clinical cases in the Collins Chabane sub-district.

Wastewater RNA signals preceded increases in clinical cases and corroborated with clinical data throughout the surveillance period. However, towards the end of the surveillance period, COVID-19 clinical cases remained low, even with increases in viral RNA copies.

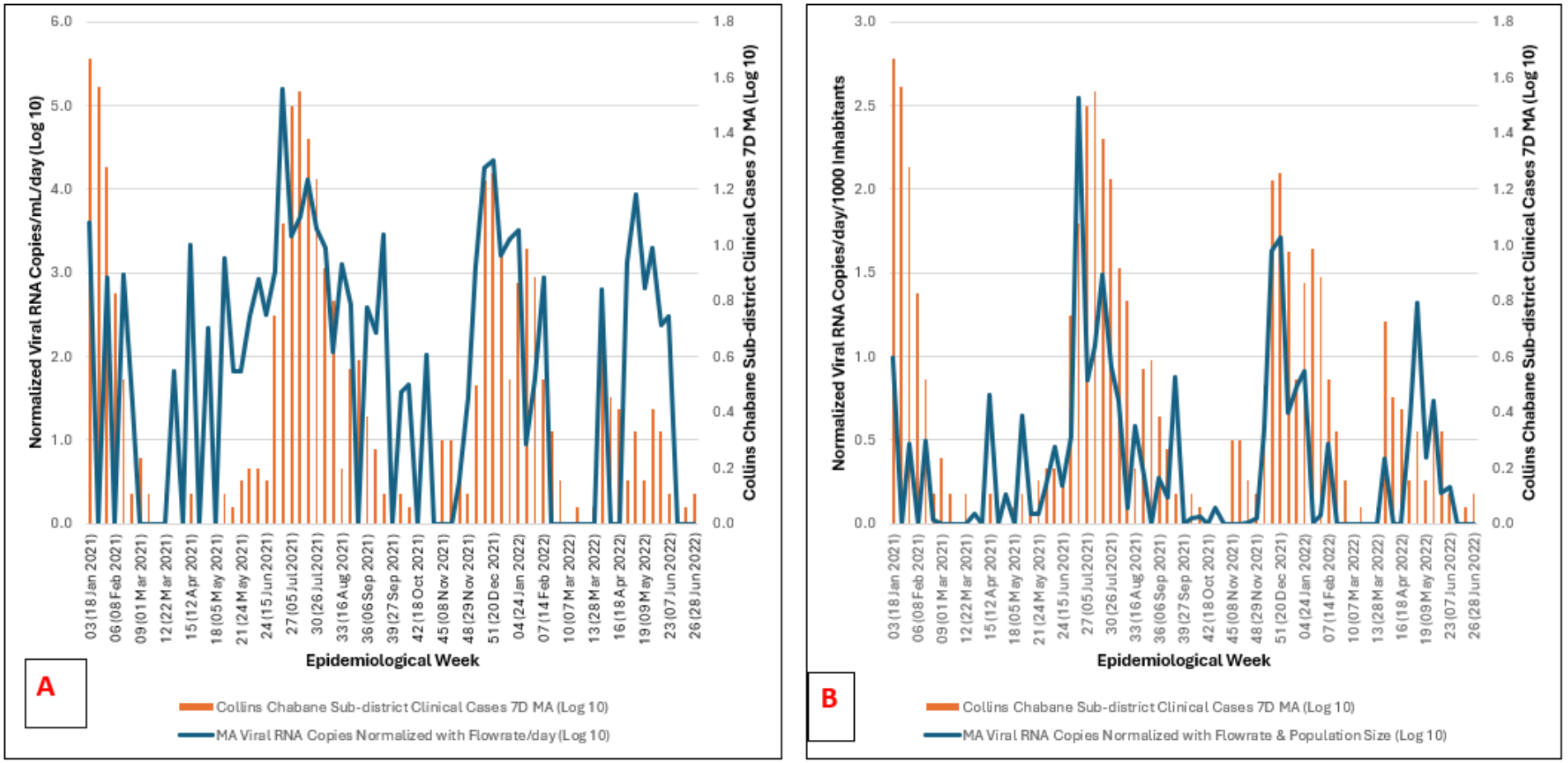


Figure 23 A and B: Normalized SARS-CoV-2 RNA in the Malamulele (MA) site compared with the Collins Chabane sub-district clinical cases.

(A) SARS-CoV-2 viral RNA detected in wastewater normalized with flow rate compared to the clinical cases **(B)** SARS-CoV-2 viral RNA detected in wastewater normalized with both the flow rate and Collins-Chabane estimated population and compared with the sub-district clinical cases.

3.3.3.4) Louis Trichardt WWTP

Upon surveillance commencement in Epi Week-16, 2021 (19 April 2021), evidence of the presence of the third wave was observed by Epi Week-22, 2021 (31 May 2021). Three weeks later, in Epi Week-25, 2021 (23 June 2021), increases in COVID-19 clinical cases were reported by the Makhado sub-district (**Figure 24**). After Epi Week-34, 2021 (23 August 2021), RNA viral copies detected began declining, thus this third wave lasted for 13 weeks, with average viral RNA detected being 1,659.3 g.c./mL. This wave had a category 2 magnitude, based on the average viral load detected.

An early warning of the incoming fourth wave was seen by Epi Week-48, 2021 (29 November 2021) which occurred one week prior to increases in reported clinical cases in Epi Week-49, 2021 (06 December 2021). As detected viral RNA particles began declining after Epi Week-04, 2022 (24 January 2022), so there was a decline in reported clinical cases. This fourth wave lasted for six weeks, with average viral RNA detected being 1,353.1 g.c./mL. During this wave, there was no change in the magnitude (category 2) of infection in the study site.

Towards the end of the surveillance period (June 2022), an increase in viral RNA was detected again from Epi Week-16, 2022 (18 April 2022). This continued for five weeks till Epi Week-22, 2022 (31 May 2022) with average viral concentrations observed being 601.7 g.c./mL. A similar magnitude of infection (category 2) was observed during this wave. As with the other sites, no increases in COVID-19 clinical cases were reported. There was similarity in correlation of non-normalized (**Figure 24**) and normalized wastewater data (**Figure 25 A and B**), when compared to clinical data, all of which were statistically significant (**Table 8**).

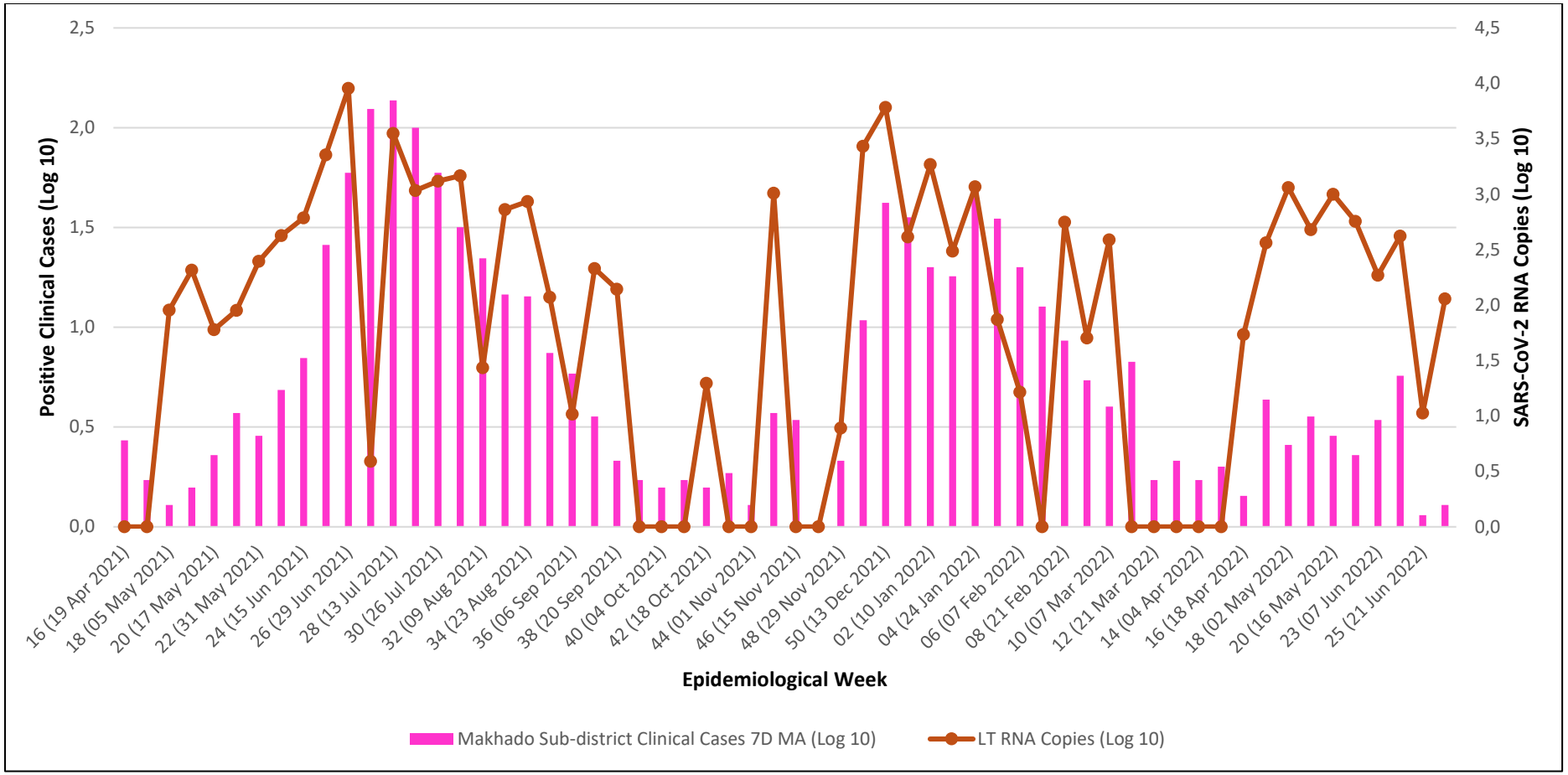


Figure 24: SARS-CoV-2 trends in the Louis Trichardt (LT) site compared to the 7D MA of clinical cases in the Makhado sub-district.

Wastewater RNA signals preceded increases in clinical cases and corroborated with clinical data throughout the surveillance period. However, towards the end of the surveillance period, COVID-19 clinical cases remained low, even with increases in viral RNA copies.

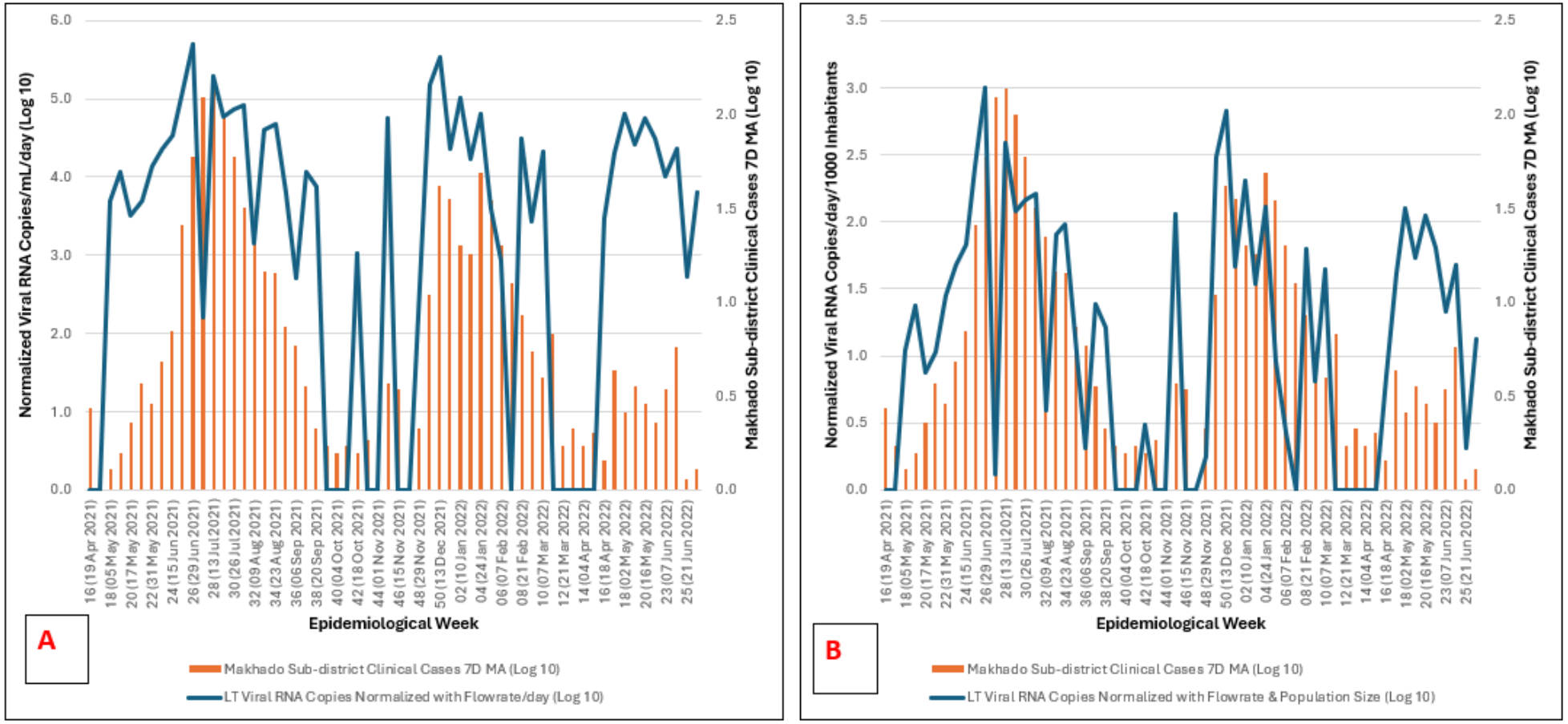


Figure 25 A and B: Normalized SARS-CoV-2 RNA in the Louis Trichardt (LT) site compared to the Makhado sub-district clinical cases.

(A) SARS-CoV-2 viral RNA detected in wastewater samples normalized with flow rate and compared to Makhado sub-district clinical cases **(B)** SARS-CoV-2 viral RNA data normalized with both flow rate and Makhado sub-district estimated population size, then compared to Makhado sub-district clinical cases.

3.3.4) SARS-CoV-2 Trends in Wastewater Compared with COVID-19 Clinical Data in the Mopani District

Sampling in this district began from Epi Week-16 of year 2021 through Epi Week-20 of year 2022. During this period, a total of 220/479 samples (45.9%) were collected, and SARS-CoV-2 was detected in 164/220 (74.5%) of these samples. The average SARS-CoV-2 viral load detected in all four WWTPs throughout the surveillance period was 6805.4 g.c./mL.

A) Summary of SARS-CoV-2 Wastewater and COVID-19 Clinical Trends in the Mopani District

In this district, just like the Vhembe district, there were three main periods when circulation of the virus in the communities was high, implying an increase in infection rates among the population. High SARS-CoV-2 viral concentrations were noticed immediately surveillance commenced (Epi Week-16, 2021; 19 April 2021) in the Ga-Kgapane site (Greater Letaba sub-district), indicating an ongoing wave in this community. Spikes in SARS-CoV-2 concentrations were observed in the Tzaneen site by Epi Week-21, 2021 (24 May 2021), closely followed by the Nkowankowa site (Epi Week-22, 2021), and finally in the Giyani site (Epi Week-23, 2021). These spikes in wastewater viral RNA concentrations were seen 1 – 8 weeks prior to increases in documented COVID-19 clinical cases. Corresponding increases in documented clinical cases were observed in the Nkowankowa, Tzaneen and Giyani sites 1 – 3 weeks after spiked SARS-CoV-2 concentrations were observed in wastewater. Interestingly, only a small increase in COVID-19 cases was recorded from Epi Week-24, 2021 in the Greater Letaba sub-district, eight weeks after spiked SARS-CoV-2 concentrations were observed from the Ga-Kgapane site. This third wave lasted for 17 weeks in the Ga-Kgapane site and had the highest magnitude (category 5) throughout the surveillance period in this district. The magnitude (potential degree of infections) of this third wave in the other study sites was low (category 2).

The fourth wave was marked by increased SARS-CoV-2 RNA signals in wastewater observed in the Giyani site by Epi Week-47, 2021 (22 November 2021). This was followed by similar spikes in wastewater in the Tzaneen and Nkowankowa study sites by Epi Week-48, 2021, and last in the Ga-Kgapane site (Epi Week-49, 2021). These spikes in wastewater RNA occurred 1 – 2 weeks prior to increases in recorded clinical cases. This fourth wave was much

milder in the study sites, lasting for 7 – 10 weeks, with all sites having a low wave magnitude (category 2), which was lower than what was observed in the Vhembe district study sites.

Towards the end of the surveillance period, a small spike in wastewater RNA concentrations was observed again. This began in the Nkowankowa site (Epi Week-16, 2021) and was subsequently observed in all three study sites by Epi Week-17, 2021. Interestingly, increases in COVID-19 clinical cases mirroring wastewater observations were only seen in the Tzaneen and Nkowankowa sites (**Figure 26**). Like the Vhembe district, this was the shortest wave (3 – 6 weeks), with all study sites having the same magnitude (category 2). Table 9 shows the correlation between viral loads obtained from N1 and N2 primers, as well as correlation statistics of normalized and non-normalized wastewater data compared to clinical data.

Table 9: A summary of correlation values obtained for comparison of viral loads with N1 and N2 primers, as well as normalized and non-normalized wastewater data compared to clinical data in the Mopani district study sites

Study Site	Spearman Correlation between N1 and N2 Viral Loads	P-value of Compared N1 and N2 Viral Loads	Spearman Correlation between Wastewater data and Clinical data	P-value of Wastewater data compared with Clinical data	Spearman Correlation between normalized Wastewater data and Clinical data	P-value of normalized Wastewater data compared with Clinical data
Tzaneen WWTP	0.820	1e-15	0.66	7e-09	0.66	7e-09
Nkowankowa WWTP	0.892	9e-22	0.76	<0.0001	0.76	7e-13
Ga-Kgapane WWTP	0.925	3e-26	0.44	4e-04	0.28	3e-02
Giyani WWTP	0.503	4e-05	0.62	9e-08	0.62	9e-08

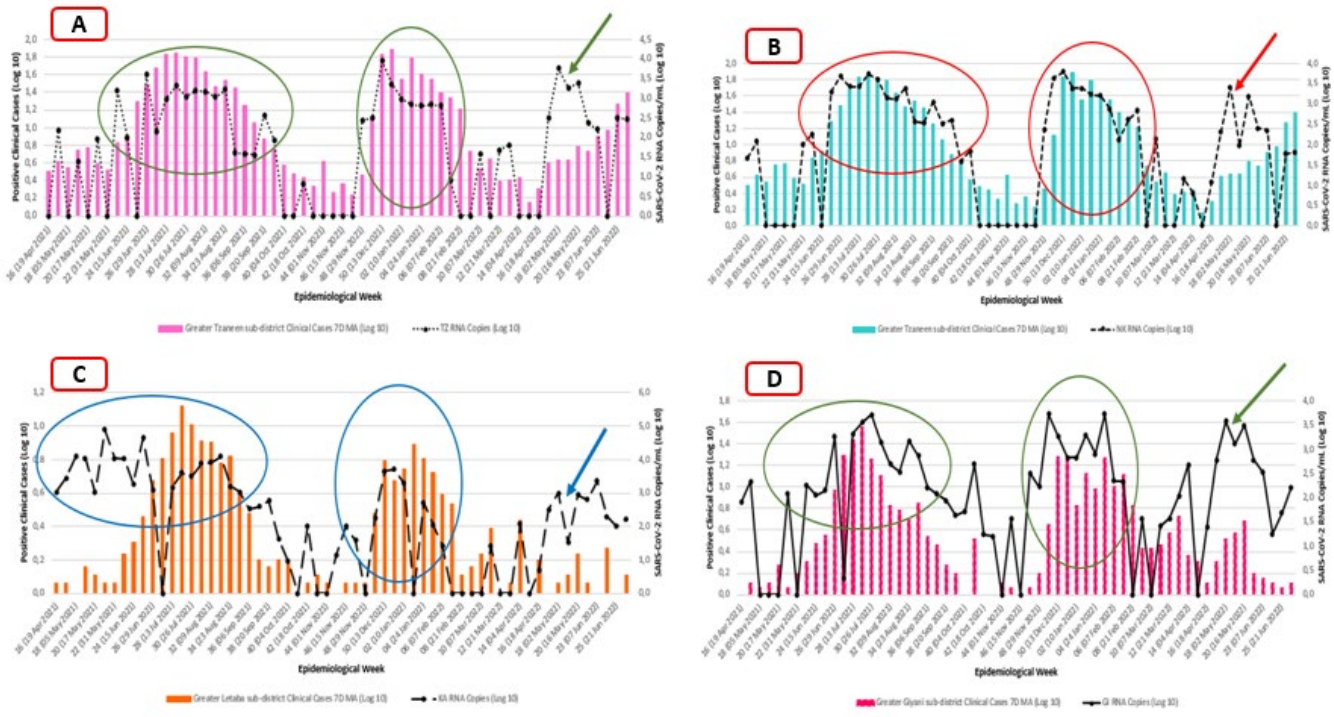


Figure 26: *A comparative analysis of SARS-CoV-2 wastewater RNA trends in the study sites compared with the 7D MA of clinical cases in the Mopani district.* **(A)** Tzaneen site compared to 7D MA clinical cases in the Greater Tzaneen sub-district; **(B)** Nkowankowa site compared to 7D MA clinical cases in the Greater Tzaneen sub-district; **(C)** Ga-Kgapane site compared to 7D MA clinical cases in the Greater Letaba sub-district; and **(D)** Giyani site compared to 7D MA clinical cases in the Greater Giyani sub-district. During the last wave, increased clinical cases were only observed in the Greater Tzaneen sub-district, which mirrored spikes in wastewater viral RNA.

Thus, based on the average viral load detected per wave, the third wave had the most impact in this district, throughout the surveillance period. A detailed description of the trends per WWTPs is elaborated below.

3.3.4.1) Tzaneen WWTP

Post surveillance commencement in Epi Week-16, 2021 (19 April 2021), increase in SARS-CoV-2 RNA signals were observed from Epi Week-21, 2021 (24 May 2021). Two weeks later, in Epi Week-23, 2021 (07 June 2021), increases in COVID-19 clinical cases were reported by the Greater Tzaneen sub-district (**Figure 27**). This third wave continued till Epi Week-34, 2021 (23 August 2021), lasting for 13 weeks, with average viral RNA detected being 1,119.2

g.c./mL. This wave had a category 2 magnitude (potential degree of infections in the community).

An increase in wastewater RNA signals in Epi Week-48, 2021 (29 November 2021) marked the start of the fourth wave in this site; this occurred one week prior to an increase in reported COVID-19 clinical cases. This wave lasted for seven weeks (till Epi Week-06, 2022; 07 February 2022), with average viral RNA detected being 1,714.1 g.c./mL. A similar magnitude of infection (category 2) was observed during this wave.

By Epi Week-17, 2022 (25 April 2022), another steady increase in viral RNA concentration was observed. This lasted for three weeks, i.e. till Epi Week-20, 2022 (16 May 2022), with average viral concentration observed during this period at 2,570.1 g.c./mL. In this last wave, a slightly higher magnitude (category 3) was observed. A small increase in the number of reported clinical cases was also observed by Epi Week-20, 2022 (16 May 2022), corroborating the observations in wastewater. There was similarity in correlation of non-normalized (**Figure 27**) and normalized wastewater data (**Figure 28 A and B**), when compared to clinical data, all of which were statistically significant (**Table 9**).

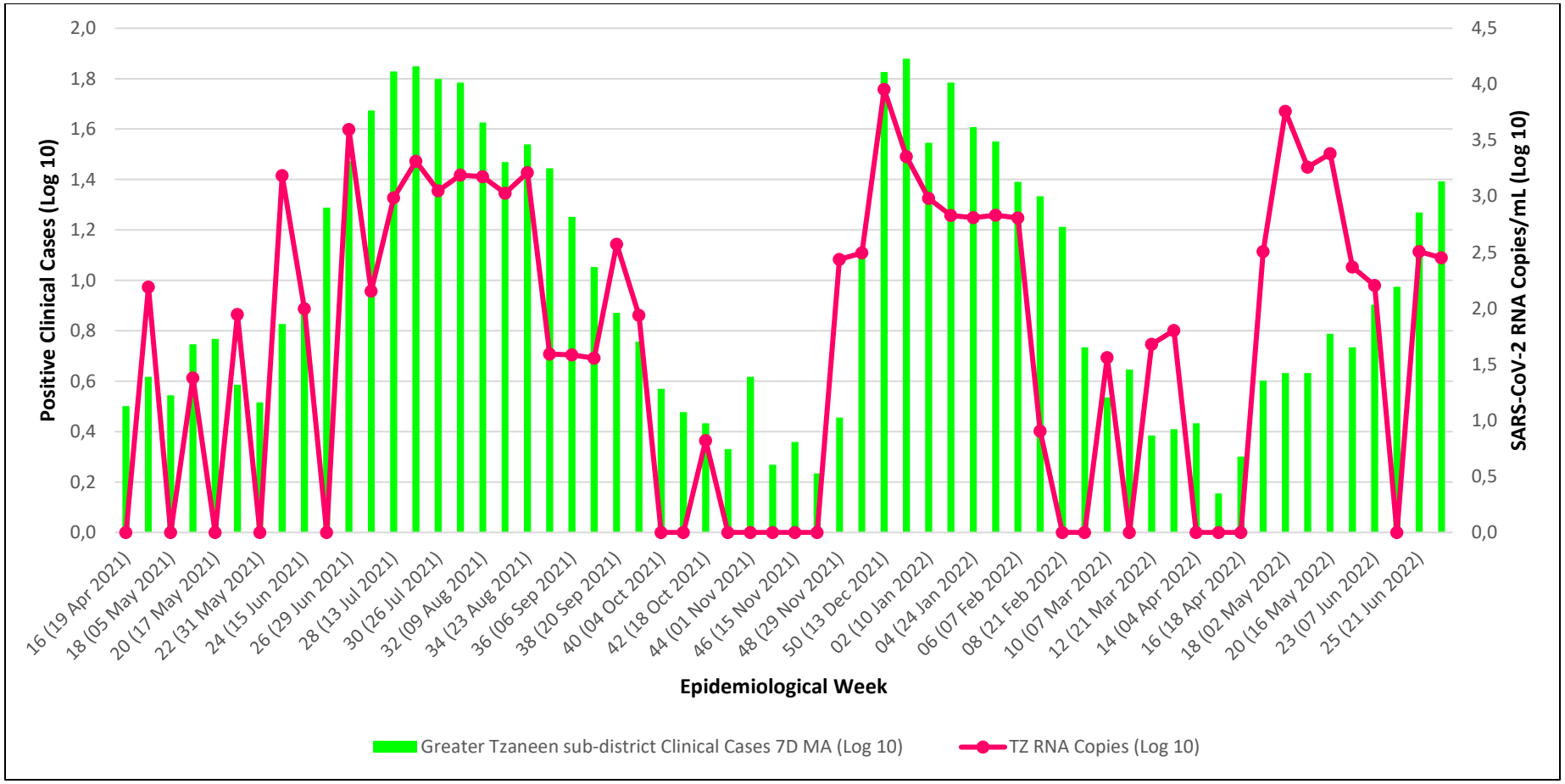


Figure 27: SARS-CoV-2 trends in the Tzaneen (TZ) site compared to the 7D MA of clinical cases in the Greater Tzaneen sub-district. Wastewater RNA signals preceded increases in clinical cases and corroborated with the clinical data throughout the surveillance period.

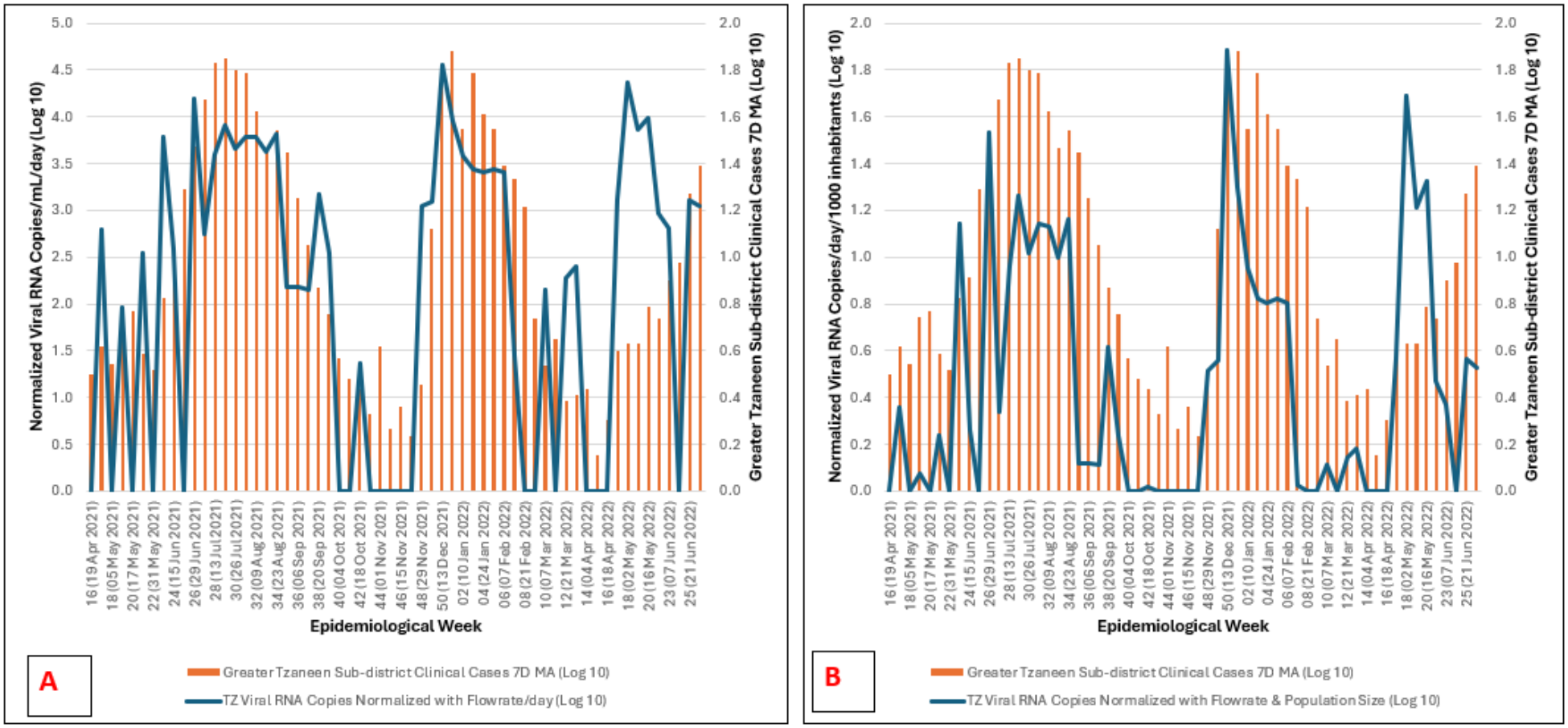


Figure 28 A and B: Normalized SARS-CoV-2 RNA in the Tzaneen (TZ) site compared with the Greater Tzaneen sub-district clinical cases.

(A) SARS-CoV-2 viral RNA detected in wastewater samples normalized with flow rate and compared to clinical cases **(B)** SARS-CoV-2 viral RNA wastewater concentrations normalized with both flow rate and Greater Tzaneen sub-district population size and compared with clinical cases.

3.3.4.2) Nkowankowa WWTP

Like the Tzaneen WWTP, this site is also located in the Greater Tzaneen sub-district. Increases in SARS-CoV-2 RNA signals observed from Epi Week-22, 2021 (31 May 2021) preceded increases in clinical cases reported one week later in Epi Week-23, 2021 (07 June 2021). This confirmed the presence of the third wave in this site **(Figure 29)**. As viral RNA concentrations declined after Epi Week-36, 2021 (06 September 2021), fewer clinical cases were also observed subsequently. Thus, this wave lasted for 14 weeks, with average viral copies detected being 1,927.3 g.c./mL. This wave had a category 2 magnitude, based on the average viral load detected.

The commencement of the fourth wave was marked by an increase in wastewater RNA signals in Epi Week-48, 2021 (29 November 2021), which was one week before the increase in reported COVID-19 clinical cases seen in Epi Week-49 (06 December 2021). This fourth wave continued till Epi Week-08, 2022 (21 February 2022), lasting for 10 weeks, with average viral RNA detected being 1,878.4 g.c./mL. As with the third wave, this wave had a category 2 magnitude.

The last increase in viral RNA concentration was observed in Epi Week-16, 2022 (18 April 2022). This continued till Epi Week-20, 2022 (16 May 2022), lasting for four weeks, and the average viral concentration observed was 876.3 g.c./mL. Similarly, this wave had a category 2 magnitude. As with the Tzaneen site, a small increase in reported COVID-19 clinical cases was noted three weeks later, by Epi Week-20, 2022 (16 May 2022). There was similarity in correlation of non-normalized **(Figure 29)** and normalized wastewater data **(Figure 30 A and B)**, when compared to clinical data, all of which were statistically significant **(Table 9)**.

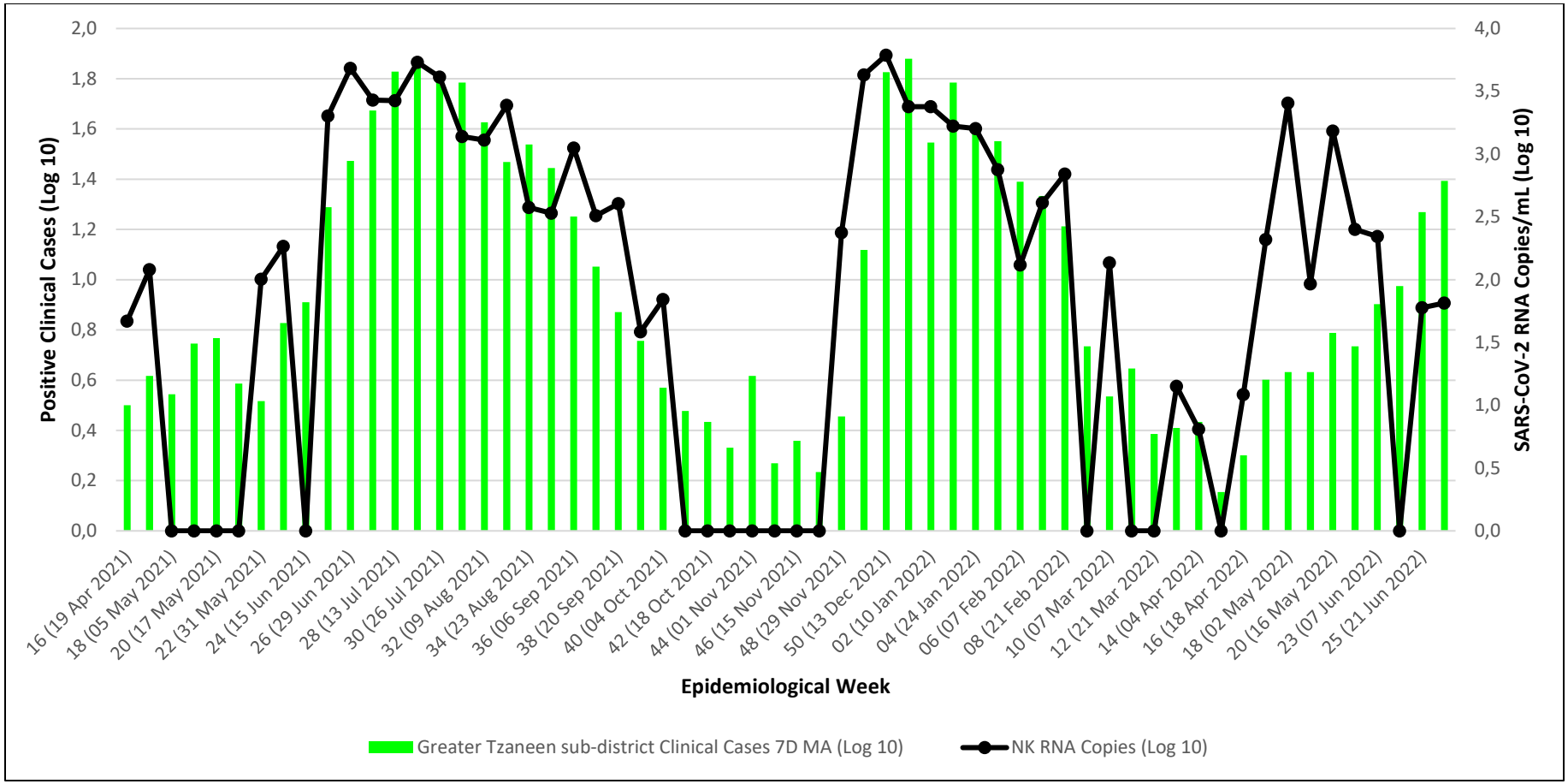


Figure 29: SARS-CoV-2 trends in the Nkowankowa (NK) site compared to the 7D MA of clinical cases in the Greater Tzaneen sub-district. Wastewater RNA signals preceded increases in clinical cases and corroborated with clinical data throughout the surveillance period.

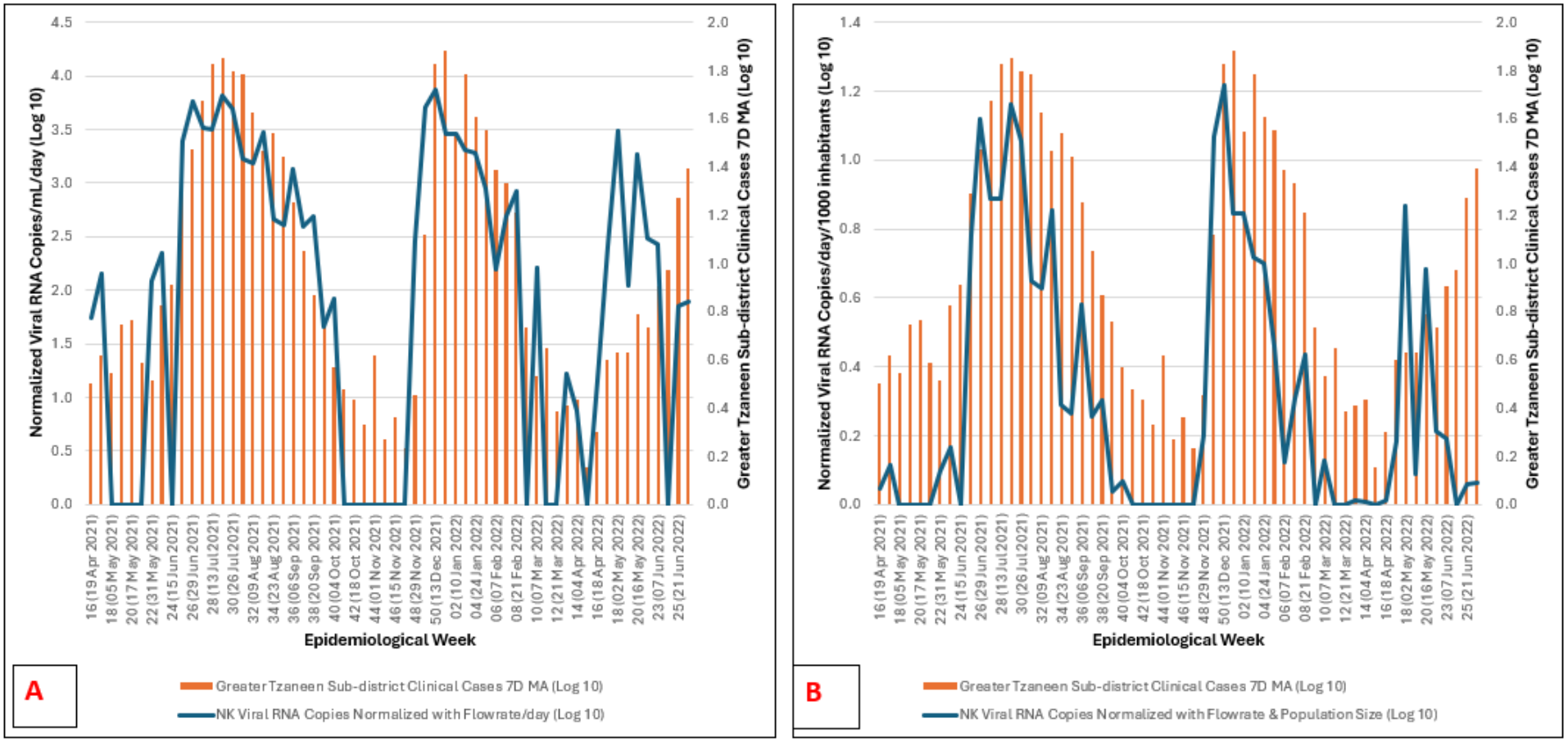


Figure 30 A and B: Normalized SARS-CoV-2 RNA in the Nkowankowa (NK) site compared to the Greater Tzaneen sub-district clinical cases.

(A) SARS-CoV-2 viral RNA detected in wastewater samples normalized with flow rate and compared to the clinical cases **(B)** SARS-CoV-2 viral RNA detected in wastewater normalized with both flow rate and Greater Tzaneen sub-district population size, then compared to the clinical cases.

3.3.4.3) Ga-Kgapane WWTP

High SARS-CoV-2 RNA signals were observed immediately after the surveillance commenced in Epi Week-16, 2021 (19 April 2021), hinting the presence of an ongoing wave. Eight weeks later, by Epi Week-24, 2021 (15 June 2021), a relatively small increase in reported clinical cases from the Greater Letaba sub-district was seen **(Figure 31)**. This continued till Epi Week-33, 2021 (16 August 2021), lasting for 17 weeks, with average viral RNA detected being 10,723 g.c./mL, for the whole period. Based on the average viral load detected, this wave had a category 5 magnitude.

One week prior to increases in reported COVID-19 clinical cases, confirming commencement of the fourth wave, an increase in RNA concentration was observed in Epi Week-49, 2021 (06 December 2021). As with the third wave, very few clinical cases were observed from this sub-district. This fourth wave lasted for five weeks, continuing till Epi Week-04, 2022 (24 January 2022), with average viral RNA detected being 2,065.5 g.c./mL. Based on the average viral load detected, this wave had a category 2 magnitude.

Viral RNA concentrations increased again from Epi Week-17, 2022 (25 April 2022), and this continued for six weeks, till Epi Week-23, 2022 (07 June 2022). The average viral concentration observed during this period was 848.2 g.c./mL. Based on the average viral load detected, this wave had a category 2 magnitude. As with the other sites in the Vhembe District, no increases in COVID-19 clinical cases were observed during this period. There was similarity in correlation of non-normalized **(Figure 31)** and normalized wastewater data **(Figure 32 A and B)**, when compared to clinical data, all of which were statistically significant **(Table 9)**.

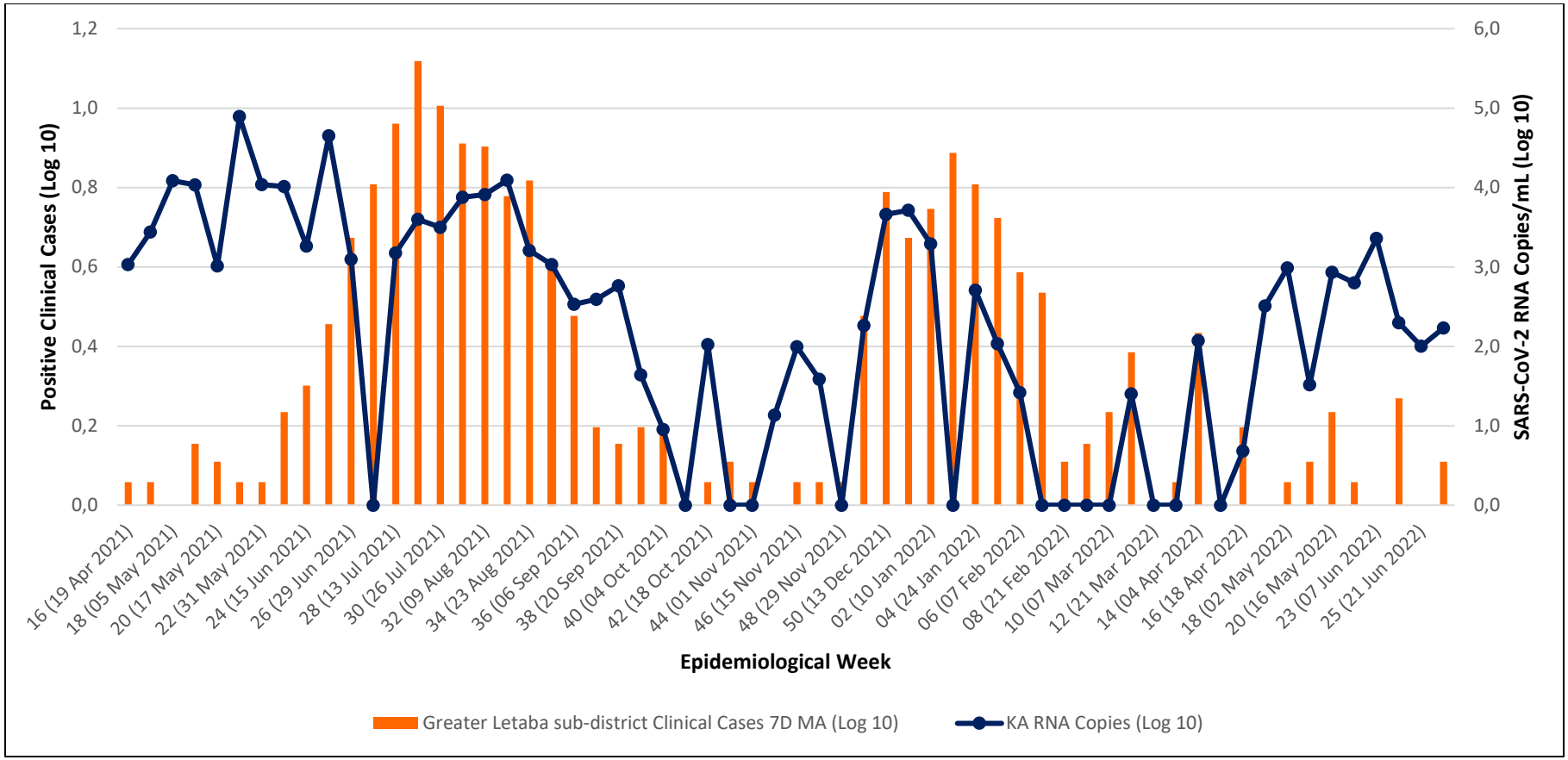


Figure 31: SARS-CoV-2 trends in the Ga-Kgapane (KA) site compared to the 7D MA of clinical cases in the Greater Letaba sub-district. Wastewater RNA signals preceded increases in clinical cases and corroborated with clinical data throughout the surveillance period. However, towards the end of the surveillance period, COVID-19 clinical cases remained low, even with increases in viral RNA copies.

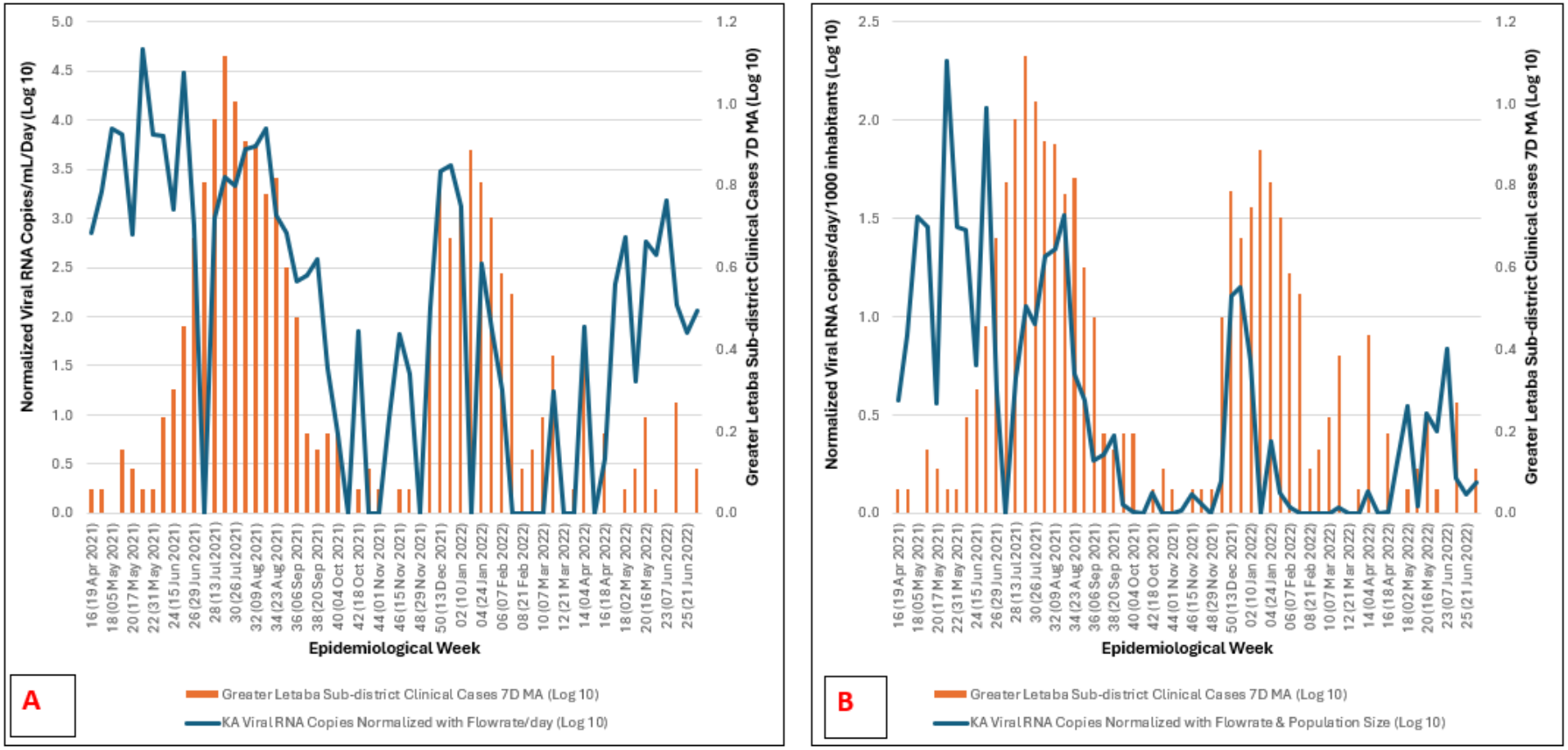


Figure 32 A and B: Normalized SARS-CoV-2 RNA in the Ga-Kgapane (KA) site compared to the Greater Letaba sub-district clinical cases.

(A) SARS-CoV-2 viral RNA detected in wastewater normalized with the flow rate and compared to the clinical cases **(B)** SARS-CoV-2 viral RNA wastewater data normalized with both flow rate and Greater Letaba sub-district population size, compared to clinical cases.

3.3.4.4) Giyani WWTP

After surveillance began in Epi Week-16, 2021 (19 April 2021), the increase in wastewater RNA signals observed from Epi Week-23, 2021 (07 June 2021) hinted the beginning of the third wave. Increases in COVID-19 clinical cases reported by the Greater Giyani sub-district were seen three weeks later by Epi Week-26, 2021 (29 June 2021). This third wave continued till Epi Week-35, 2021 (30 August 2021), lasting for 12 weeks, with average viral RNA of 1,345 g.c./mL being detected **(Figure 33)**. The magnitude for this wave based on the average viral load detected was category 2.

The beginning of the fourth wave was marked by an increase in viral RNA concentrations observed in Epi Week-47, 2021 (22 November 2021). Two weeks later, in Epi Week-49, 2021 (22 November 2021), increases in reported COVID-19 clinical cases were observed. This fourth wave lasted for eight weeks till Epi Week-05, 2022 (31 January 2022) with average viral RNA of 1,906 g.c./mL being detected. The magnitude for this wave based on the average viral load detected was category 2.

Increase in viral RNA circulation was observed again from Epi Week-17, 2022 (25 April 2022) which continued for five weeks till Epi Week-22, 2022 (31 May 2022). The average viral concentration observed during this period was 1,874.3 g.c./mL. Based on the average viral load detected, this wave had a category 2 magnitude. As with the other sites, no increases in COVID-19 clinical cases were reported during this period. There was similarity in correlation of non-normalized **(Figure 33)** and normalized wastewater data **(Figure 34 A and B)**, when compared to clinical data, all of which were statistically significant **(Table 9)**.

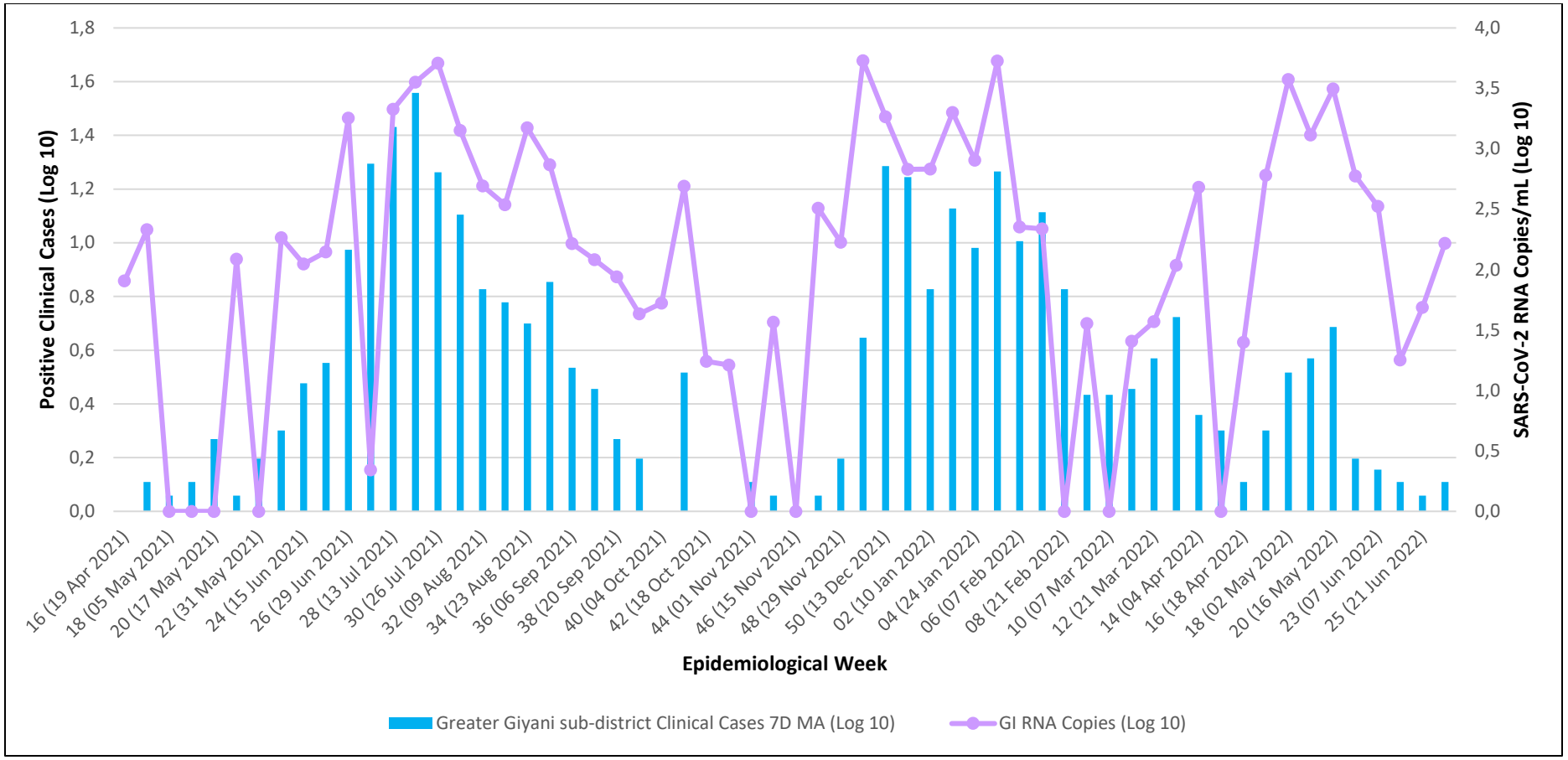


Figure 33: SARS-CoV-2 trends in the Giyani (GI) site compared to the 7D MA of clinical cases in the Greater Giyani sub-district. Wastewater RNA signals preceded increases in clinical cases and corroborated with clinical data throughout the surveillance period. However, towards the end of the surveillance period, COVID-19 clinical cases remained low, even with increases in viral RNA copies.

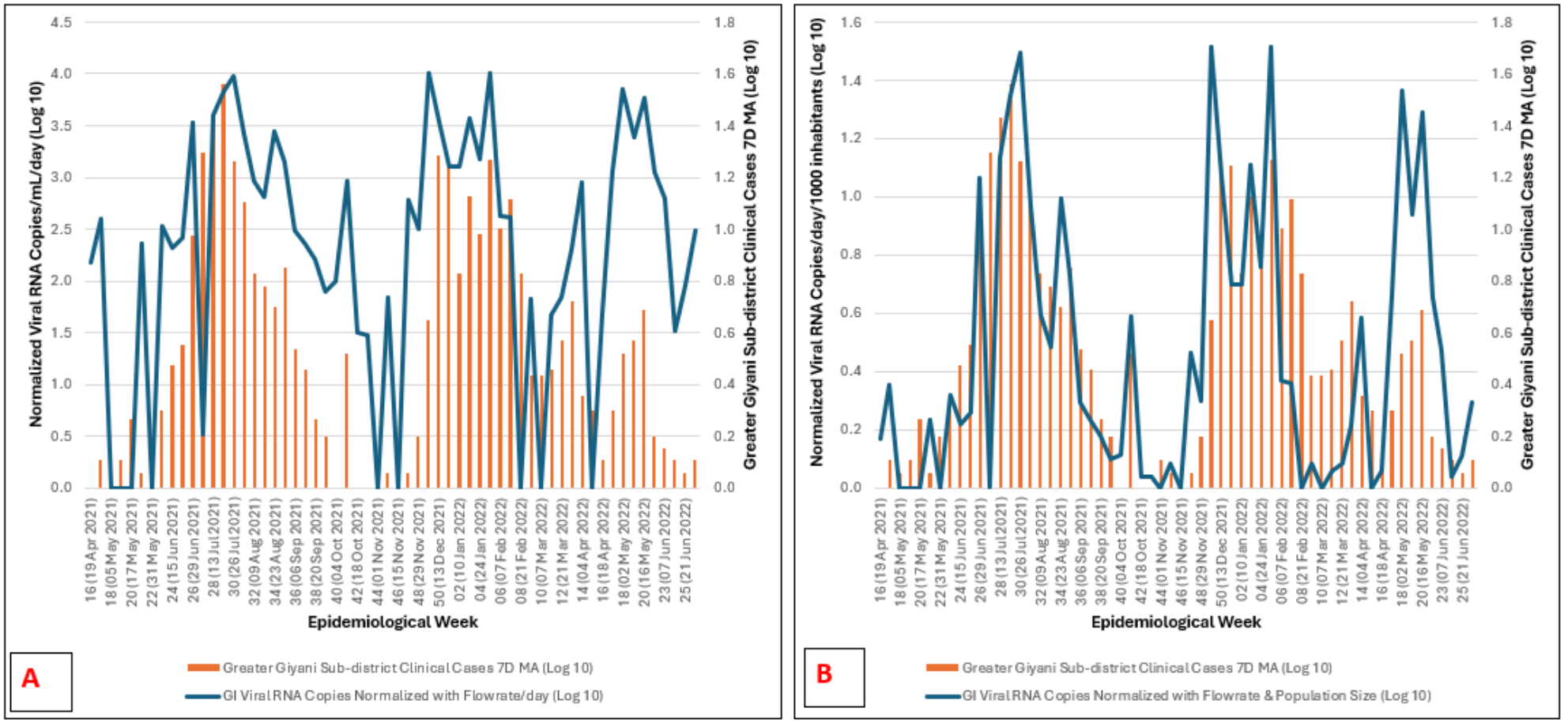


Figure 34 A and B: Normalized SARS-CoV-2 RNA in the Giyani (GI) site compared to the Greater Giyani sub-district clinical cases. **(A)** SARS-CoV-2 viral RNA detected in wastewater normalized with the flow rate compared to the clinical cases **(B)** SARS-CoV-2 viral RNA detected in wastewater normalized with both the flow rate and Greater Giyani sub-district population size and compared to clinical cases.

3.3.5) SARS-CoV-2 Variants of Concern Circulation in Vhembe & Mopani Districts in Limpopo Province, South Africa

With ongoing surveillance of SARS-CoV-2 trends in the study sites, monitoring SARS-CoV-2 variant circulation was simultaneously conducted. Of the 365 SARS-CoV-2 positive samples from both districts, 80 samples (21.9%) met the inclusion criteria (≥1500 g.c./mL) for SARS-CoV-2 VOC determination through ASG. Since very few samples met the inclusion criteria, these ASG results give a general overview of VOCs present in the study sites, based on the detection of mutations in the samples. These samples ranged from Epi Week-18 of year 2021 to Epi Week-25 of year 2022 (03 May 2021 – 21 June 2022), which characterize the period when the third and fourth waves occurred. The Alpha, Beta, and Delta VOCs were in circulation in both districts throughout the surveillance period, though at varying frequencies. Detected VOCs occurred at the following frequencies: Alpha (26%), Beta (24%), Delta (23%), and Omicron precursors (27%), respectively **(Figure 35)**. The Delta VOC was detected as early as Epi Week-18, 2021 (03 May 2021), preceding other reports from South Africa. Similarly, alleles corresponding to Omicron, which could represent Omicron precursors, were also detected by Epi Week-20, 2021 (17 May 2021), prior to other reports from South Africa. Due to the few samples included for ASG, investigation of the most prevalent variant circulating in the study sites during both 3rd and 4th waves were not done.

Whole genome sequencing (WGS) was used to confirm results obtained through ASG. This analysis is elaborated in section 4.3.5. Of the 80 samples subjected to ASG, 41 of them were sequenced (41/80 – 51.3%). Obtained sequences were subjected to the PangoLIN and Nextclade tools for variant assignment. Of the 41 sequenced samples, both PangoLIN and Nextclade assigned the same VOC to 35/41 (85.4%) samples, with frequencies occurring of 4.9%, 46.3% and 34.1% for the Beta, Delta and Omicron VOCs, respectively. For the remaining 6/41 (14.6%) sequenced samples, both PangoLIN and Nextclade tools only assigned a lineage and clade, but not a specific variant. Phylogenetic analysis showed these sequences clustering with downloaded Alpha VOC reference sequences. Concordance between S-gene-defining mutations and variant assignment was investigated in WGS and ASG techniques. There was concordance in S-gene-defining mutations and variant assignment in 21/41 (51.2%) samples evaluated by both techniques. In 13/41 (31.7%) samples, at least one S-gene defining mutation was observed in both techniques, but with a

different variant assignment. For the rest of the 7/41 (17%) samples, no concordance existed between mutations detected by allelic variant genotyping or variant assignment in both techniques.

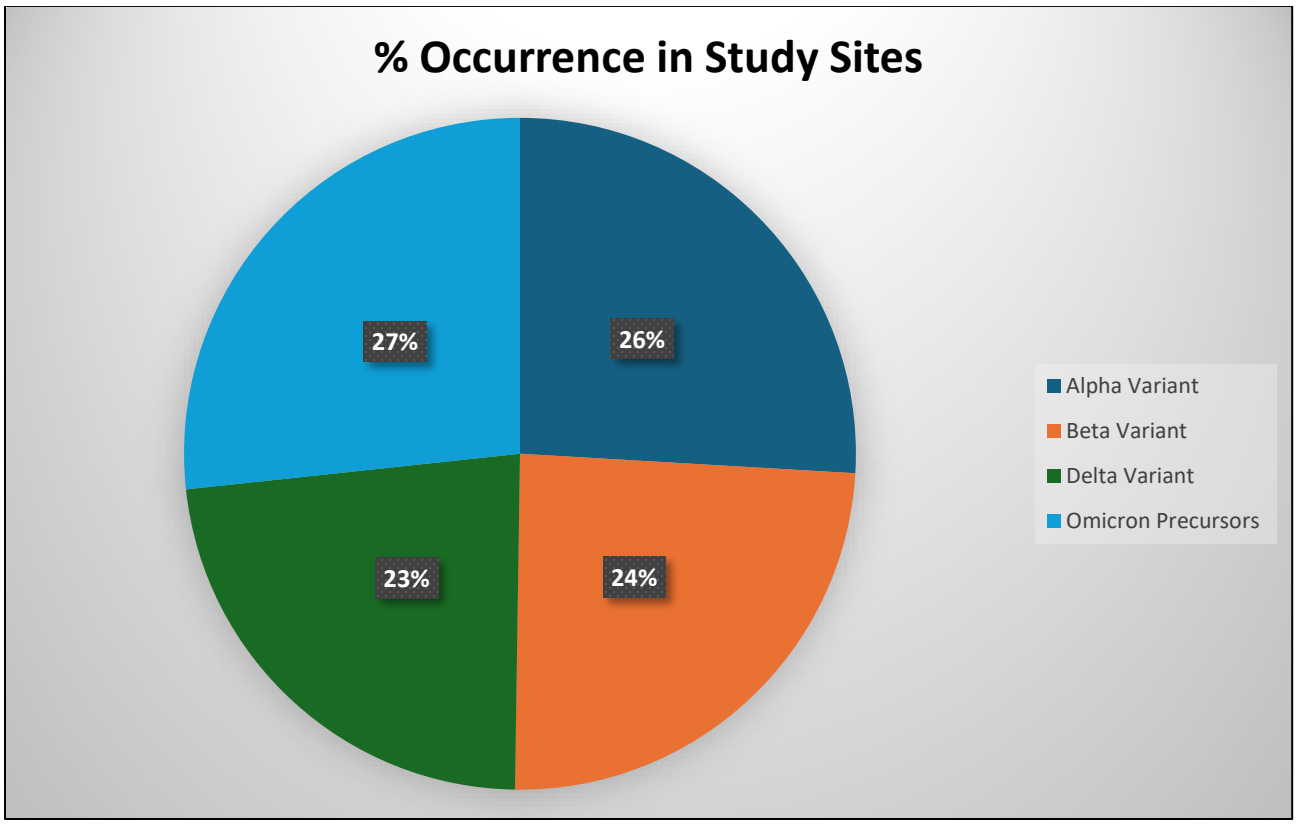


Figure 35: Overall VOC frequency of occurrence, in the study sites, with omicron alleles being mostly detected during the surveillance period.

3.3.6) Comparison of Predicted Prevalence Estimates with Clinical Data in Vhembe and Mopani Districts

The predicted number of people infected throughout the surveillance period is reported as the median and 95% CI, since the median is less sensitive to extreme values from the input distributions. From the four study sites in the Vhembe district, 67.9% SARS-CoV-2 positive samples met the modelling inclusion criteria (SARS-CoV-2 copies/mL with SD <1). A median of 4,437 (95% CI: 34,642 – 53,294; **Table 10**) infections were predicted to have occurred between throughout the surveillance period (18 January 2021 – 28 June 2022) in the Vhembe District. In the Mopani District, 78.5% SARS-CoV-2 positive samples met the modelling inclusion criteria from the four study sites. A median of 232 (95% CI: 1696 – 2334; **Table 11**)

infections were predicted to have occurred between during this surveillance period (19 April 2021 – 28 June 2022).

Table 10: Median number of infections per site in the Vhembe district for SARS-CoV-2 positive samples met the modelling inclusion criteria.

Vhembe District Study Sites	Estimated Prevalence Median	Estimated Prevalence (Lower 95% CI)	Estimated Prevalence (Upper 95% CI)
Thohoyandou Site	725,6	3764,3	5017,4
Malamulele Site	328,9	2273,7	3026,7
Siloam Hospital Site	2549,5	22253,5	36679,0
Louis Trichardt Site	833,7	6351,1	8574,2
Total	4437,7	34642,6	53297,3

Table 11: Median number of infections per site in the Mopani district for SARS-CoV-2 positive samples met the modelling inclusion criteria.

Mopani District Study Sites	Estimated Prevalence Median	Estimated Prevalence (Lower 95% CI)	Estimated Prevalence (Upper 95% CI)
Tzaneen Site	62,0	563,5	788,1
Nkowankowa Site	37,8	265,3	356,2
Ga-Kgapane Site	95,7	664,3	904,2
Giyani Site	36,4	203,6	285,8
Total	231,9	1696,7	2334,3

Per the records obtained from the Limpopo Department of Health, South Africa, the calculated 7D MA showed that approximately 2824 and 2083 clinical cases occurred in the Vhembe and Mopani districts respectively, during this surveillance period. A positive correlation was observed between the predicted number of infections and the determined SARS-CoV-2 viral copies in all study sites. A positive correlation (r range: 0.67 – 0.69) between the predicted number of infections and the reported clinical cases was observed in only 3/8 sites (37.5%). In the other 5/8 sites, a weaker correlation (r range: 0.006 – 0.42) was observed, with the weakest (r = 0.006) being from the Ga-Kgapane study site.

3.3.7) Predicted Prevalence Compared with Clinical Data in Vhembe District

Of all four study sites, more infections were predicted to have occurred in the Siloam Hospital site (median: 2 – 2549) throughout the surveillance period. During the third wave, the model predicted more infections to have occurred in the Louis Trichardt (median: 0.2 – 834) and Collins Chabane (median: 0.8 – 329) study sites, and fewer infections occurring in the other two sites. Similarly, more clinical cases were recorded in to have occurred during the third wave in both sub-districts (7D MA: 0.6 – 136 and 0.3 – 35, respectively).

During the fourth wave, more infections were predicted to have occurred in the Thohoyandou (median: 0.1 – 726) and Siloam Hospital (median: 2 – 1877) sites. Contrarily, reported clinical cases in both sites showed more cases to have occurring during the third wave (7D MA: 2.4 – 86 and 0.6 – 136, respectively), compared to the fourth wave (7D MA: 0.6 – 68.7 and 1.1 – 48.3, respectively).

Interestingly, during the last wave towards the end of the surveillance period (25 April to 21 June 2022), the model predicted a slight increase in infections in all the study sites, which was reflected by the increased viral load concentrations detected. However, no corresponding increase in the number of documented cases at the district level were observed. The model predicted most infections to have occurred in the Siloam hospital site (median: 2 – 2549), which was the highest throughout the surveillance period.

3.3.7.1) Thohoyandou WWTP in the Thulamela Sub-district

Thirty-eight out of fifty (76%) samples met the inclusion criteria (Epi Week-15, 2021 and Epi Week-26, 2022; 12 April 2021 – 28 June 2022). The estimated prevalence determined was compared with the 7D MA of clinical cases from the Thulamela sub-district **(Figure 36)**. The model estimated that the median number of infected people in this site ranged from 0.7 – 726 (95% CI: 0.5 – 5017.4). The model predicted more infections to have occurred during the fourth wave (0.1 – 726 cases) compared to the third wave (0.6 – 10 cases). This corresponded with SARS-CoV-2 concentrations detected during the surveillance period. The highest viral copies (144,202.8 g.c./mL) occurred during the fourth wave in Epi Week-50, 2021 (13 December 2021), and the model also predicted most infections (median 726; 95% CI: 3764 – 5017) to have occurred in this week. Furthermore, a positive Spearman's correlation of 0.92 ($p<0.001$) was seen between the estimated prevalence and the viral copies.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA range: 0.1 – 86 cases) throughout the study period in the Thulamela sub-district. Contrary to the estimated prevalence, more clinical cases were recorded during the third wave (2.4 – 86 cases) compared to the fourth wave (0.6 – 68.7 cases). A median of 0.1 – 20 infections were predicted to have occurred during the last wave towards the end of the surveillance period (25 April to 21 June 2022). However, this was not reflected in clinical cases, since only 0.3 – 3 cases were recorded over a 7D MA during this period. Statistically, a weak correlation ($r = 0.31$; $p = 5e\text{-}02$) was determined to exist between the estimated prevalence and positive clinical cases from the Thulamela sub-district.

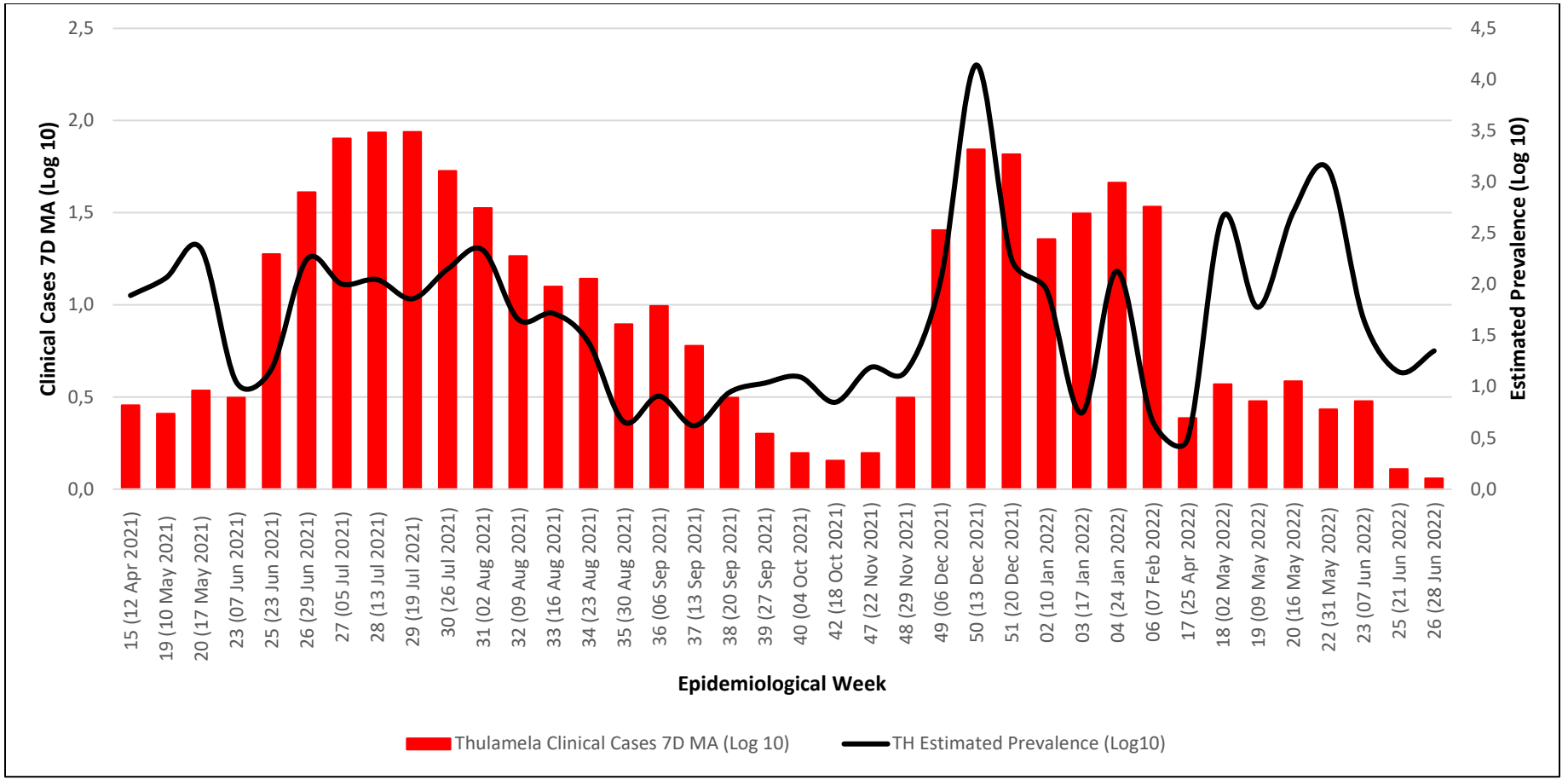


Figure 36: Comparing the number of reported clinical cases in the Thulamela sub-district with the predicted number of infected people in the Thohoyandou site.

More clinical cases were reported in the third wave compared to the fourth wave; however, the model predicted more infections to have occurred during the latter wave.

3.3.7.2) Siloam Hospital SP in the Makhado Sub-district

The Siloam Hospital is also situated in the Makhado municipality, in the Vhembe District and is estimated to serve ~16,000 people with hospital beds, however with ongoing daily consultations without admissions, the number of people served by this facility could be much higher. Twenty-nine out of forty-one (70.7%) samples met the inclusion criteria (Epi Week-03, 2021 to Epi Week-25, 2022; 18 January 2021 – 21 June 2022). The determined estimated prevalence was compared with the 7D MA of clinical cases from the Makhado sub-district **(Figure 37)**. The model estimated that a median of 2 – 2549 (95% CI: 12 – 36,679) people were infected throughout the study period. More infections were predicted to have occurred during the fourth wave (2 – 1877 cases) compared to the third wave (8 – 1001 cases). This corresponded with detected SARS-CoV-2 concentrations observed during the surveillance period. The highest viral copies (46,795.5 g.c./mL) occurred during the fourth wave in Epi Week-07, 2022 (14 February 2022), and the model also predicted most infections (median 1877; 95% CI: 22,253 – 36,679) to have occurred in this week. A strong correlation (r = 0.91; p = 6e-12) was seen between the estimated prevalence and the viral copies detected throughout the surveillance period.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 0.1 – 136 cases) throughout the study period in the Makhado sub-district. Unlike the estimated prevalence more clinical cases were recorded during the third wave (0.6 – 136 cases) compared to the fourth wave (1.1 – 48.3 cases). Although a median of 40 – 2546 infections were predicted to have occurred during the last wave (Epi Weeks 17 to 25, 2022; 25 April to 21 June 2022), no corresponding increases in clinical cases (0.1 – 4.7 cases) were recorded during this period. Similarly, a weak correlation (r = 0.22; p = 3e-01) was seen between the estimated prevalence and positive clinical cases from the Makhado sub-district.

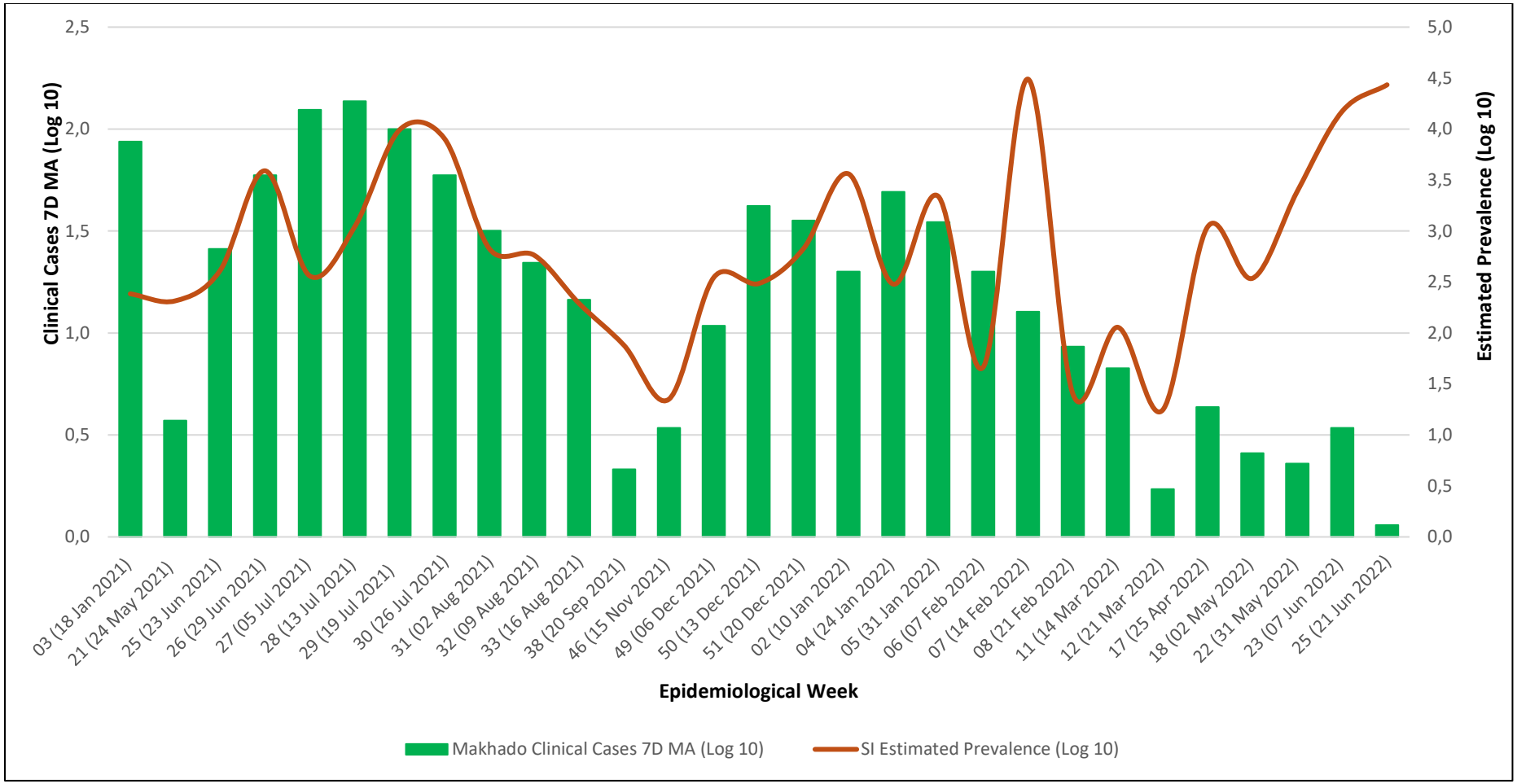


Figure 37: Comparing the number of reported clinical cases in the Makhado sub-district with the predicted number of infected people in the Siloam Hospital site.

More clinical cases were reported in the third wave compared to the fourth wave. However, the model predicted more infections to have been in circulation during the fourth wave.

3.3.7.3) Malamulele WWTP in the Collins Chabane Sub-district

Twenty-five out of forty-eight (52%) samples met the inclusion criteria (Epi Week-03, 2021 to Epi Week-23, 2022; 18 January 2021 – 07 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from the Collins Chabane sub-district **(Figure 38)**. The model estimated that a median of 0.8 – 329 (95% CI: 1 – 3,027) people were infected during this study period. More infections were predicted to have occurred during the third wave (0.8 – 329 cases) compared to the fourth wave (0.1 – 37 cases). This corresponded with SARS-CoV-2 concentrations observed during the surveillance period. The highest viral copies (45,995.9 g.c./mL) occurred during the third wave in Epi Week-26, 2021 (29 June 2021), and the model also predicted most infections (median 329; 95% CI: 2274 – 3027) to have occurred in this week. A positive correlation (r = 0.96; p = 3e-14) was observed between the predicted prevalence and the detected viral copies.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 0.3 – 45 cases) throughout the study period in the Collins Chabane sub-district. Like the estimated prevalence, more cases were recorded during the third wave (0.3 – 35) compared to the fourth wave (2.3 – 17.1). Although the model predicted between 1.1 – 14 infections to have occurred during the last wave (Epi Weeks 17 to 23, 2022; 25 April to 07 June 2022), no increase (0.3 – 1.1 cases) in clinical cases were documented during this period. Statistically, a positive correlation (r = 0.69; p = 1e-04) was seen between the estimated prevalence and positive clinical cases from the Collins Chabane sub-district.

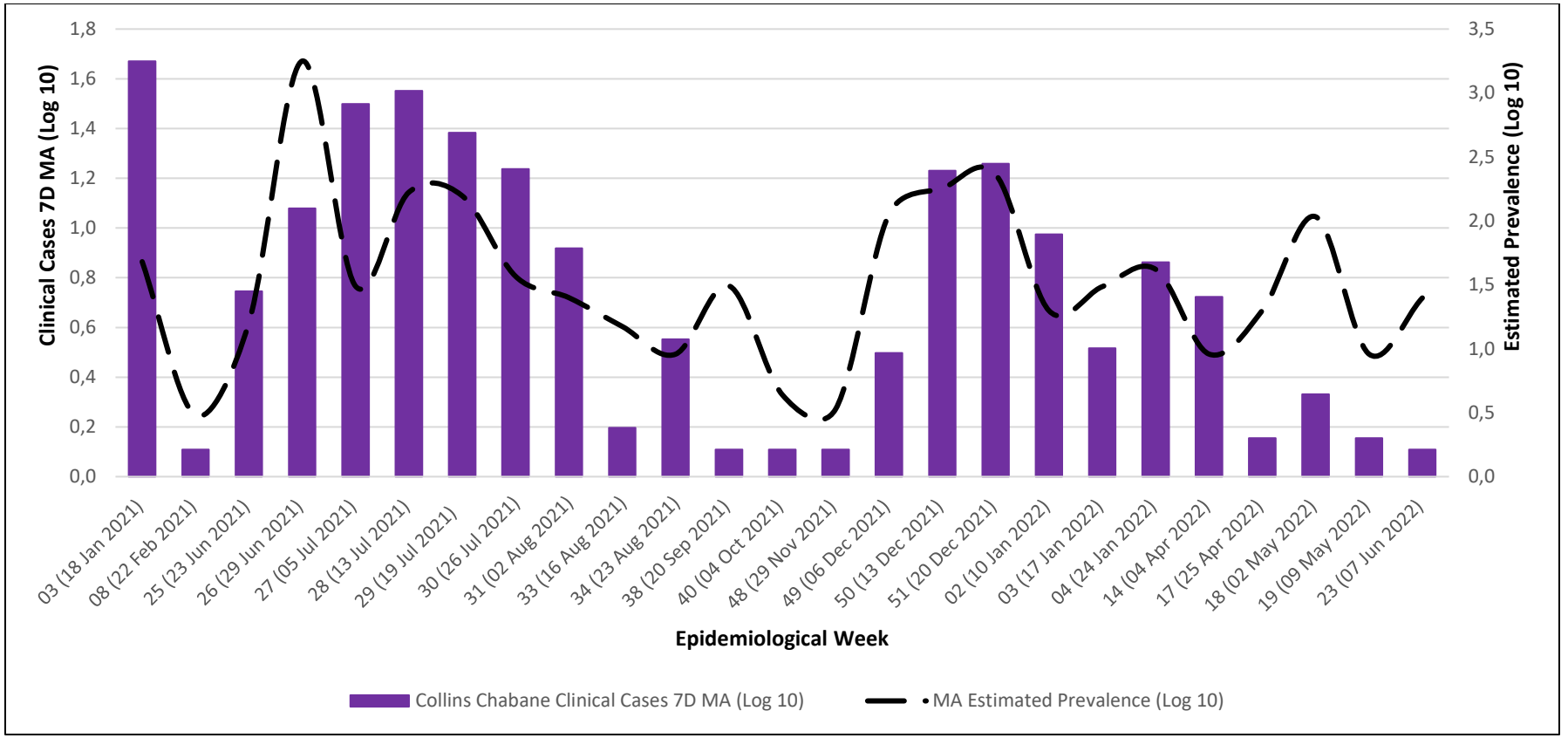


Figure 38: Comparing the number of reported clinical cases with the predicted number of people infected in the Malamulele site. The model predicted most infections to have occurred during the third wave, and this was corroborated with reported clinical cases which documented the most cases occurring during this third wave.

3.3.7.4) Louis Trichardt WWTP in the Makhado Sub-district

Thirty-three out of forty-five (70.7%) samples met the inclusion criteria (Epi Week-19, 2021 to Week-26, 2022; 10 May 2021 – 28 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from the Makhado sub-district **(Figure 39)**. The model estimated that the number of infected people in this community ranged from 0.2 – 834 (95% CI: 2 – 8,574). More infections were predicted to have occurred during the third wave (0.2 – 834 people) compared to the fourth wave (0.6 – 650 people). The highest viral copies (9011.8 g.c./mL) occurred during the third wave in Epi Week-26, 2021 (29 June 2021), and the model also predicted the highest infections (median 834; 95% CI: 6351 – 8574) to have occurred in this week. When the estimated prevalence was compared to the SARS-CoV-2 viral copies, a strong correlation (r = 0.91; p = 6e-12) was seen, thus confirming the precision of the model prediction.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 0.1 – 136 cases) throughout the study period in the Makhado sub-district. Like the estimated prevalence, more clinical cases were recorded during the third wave (0.6 – 136 cases) compared to the fourth wave (1.1 – 48.3 cases). During the last wave (Epi Weeks 17 to 23; 25 April to 07 June 2022), an increase in infections were predicted to have occurred (median: 2 – 156). However, no corresponding increases in clinical cases (7D MA: 0.1 – 4.7) were seen in the sub-district clinical data. A weak correlation (r = 0.44; *p* = 1e-02) existed between the estimated prevalence and positive clinical cases from the Makhado sub-district.

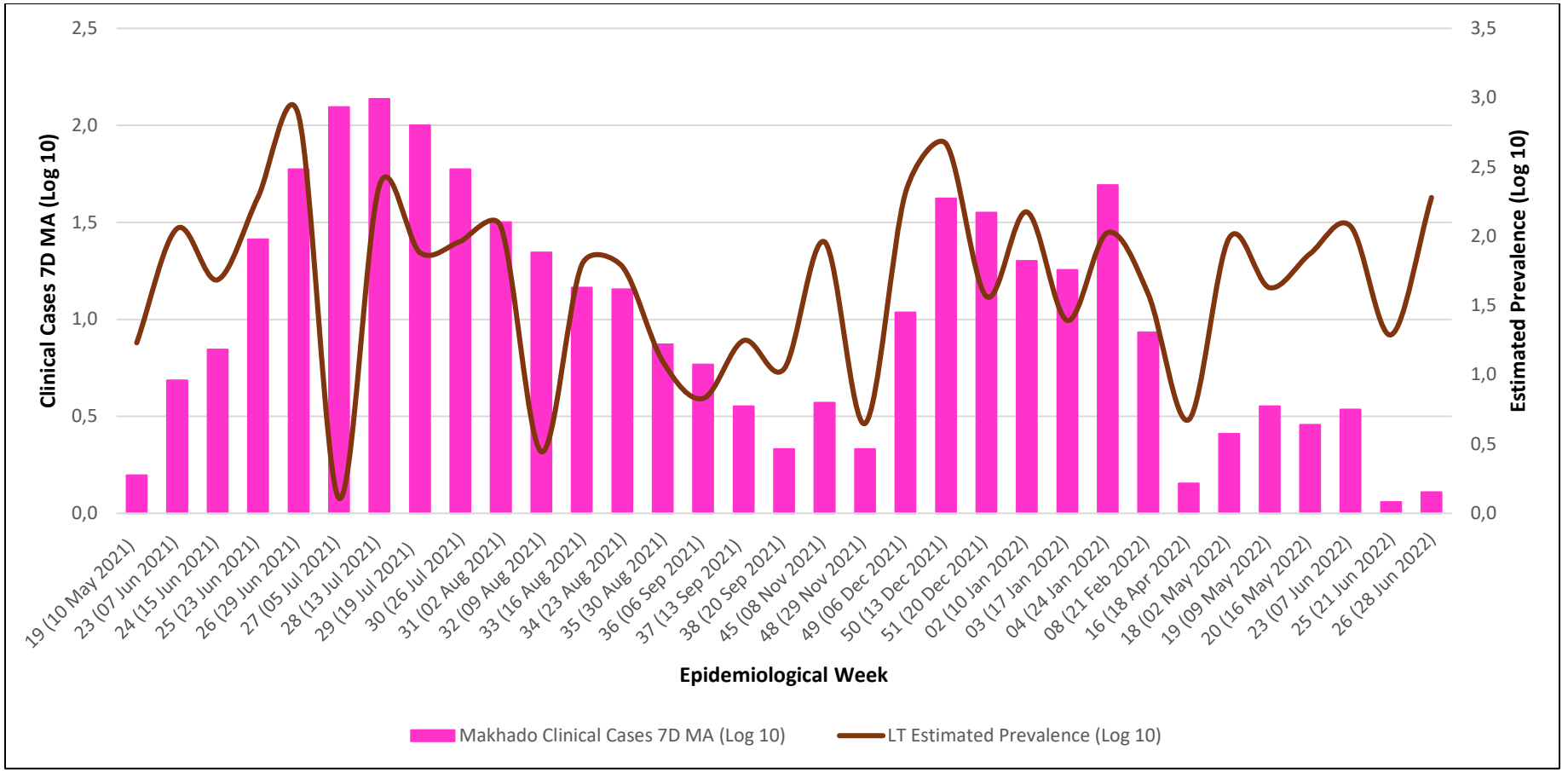


Figure 39: A representation of the predicted number of people infected in the Louis Trichardt site compared to the number of reported clinical cases in the Makhado sub-district.

The model predicted that more cases would occur in the Louis Trichardt site during the third wave compared to the fourth wave and this was confirmed by clinical cases reports from the Makhado sub-district.

3.3.8) Predicted Prevalence Compared to Clinical Data in Mopani District

Of all four study sites, more infections were predicted to have occurred in the Ga-Kgapane site (median: 0.1 – 96) throughout the surveillance period. During the third wave, the model predicted more infections to have occurred in the Ga-Kgapane site only (median: 0.4 – 96). Similarly, more clinical cases were recorded in to have occurred during the third wave (7D MA: 0.1 – 12 cases) in this study site.

For all the other three sites (Tzaneen, Nkowankowa and Giyani), more infections occurred during the fourth wave, as predicted by the statistical model (median: 2 – 62, 0.7 – 38 and 0.5 – 17 respectively). More clinical cases were reported during the fourth wave in Tzaneen and Nkowankowa sites only (7D MA: 2 – 75 and 2 – 75, respectively). Contrastingly, less clinical cases were documented during the fourth wave in the Giyani study site (7D MA: 0.1 – 18.3).

During the last wave towards the end of the surveillance period (25 April to 21 June 2022), the model predicted a slight increase in infections in all the study sites. This corresponded with similar increases in documented cases at the district level for the Tzaneen and Nkowankowa sites only. Contrarily, predicted increases in infection in the Ga-Kgapane and Giyani sites observed from increased SARS-CoV-2 viral load in wastewater was not mirrored by increases in documented cases at the district level for these sites.

3.3.8.1) Tzaneen WWTP in the Greater Tzaneen Sub-district

Thirty out of forty-one (73%) samples met the inclusion criteria (Epi Week-23, 2021 and Epi Week-26, 2022; 07 June 2021 – 28 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from this sub-district **(Figure 40)**. The model predicted a median of 0.3 – 62 (95% CI: 2.3 – 788) infections to have occurred in this site. More infections were predicted to have occurred during the fourth wave (median: 2 – 62; 95% CI: 21 – 788 infections) compared to the third wave (1 – 31; 95% CI: 7 – 347 infections). The model predicted most infections (median: 62; 95% CI: 564 – 788) to have occurred in Epi Week-50, 2021 (13 December 2021), which was also the week where the highest viral copies (8990.9 g.c./mL) were detected. These findings were corroborated by a positive correlation ($r = 0.71$; $p = 1e\text{-}05$) observed between the estimated prevalence and detected SARS-CoV-2 viral copies.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 2 – 75 cases) throughout the study period in the Greater Tzaneen sub-district. Like the estimated prevalence, more clinical cases were recorded during the fourth wave (2 – 75 cases) compared to the third wave (6 – 70 cases). During the last wave (Epi Weeks 17 to 20, 2022; 25 April to 16 May 2022), the model predicted an increase in infections (median: 2 – 58) to have occurred. A corresponding increase in clinical cases (7D MA: 3 – 24) was seen in the sub-district clinical data. This was corroborated by a positive correlation ($r = 0.68$; $p = 3e\text{-}05$) seen between the estimated prevalence and clinical cases.

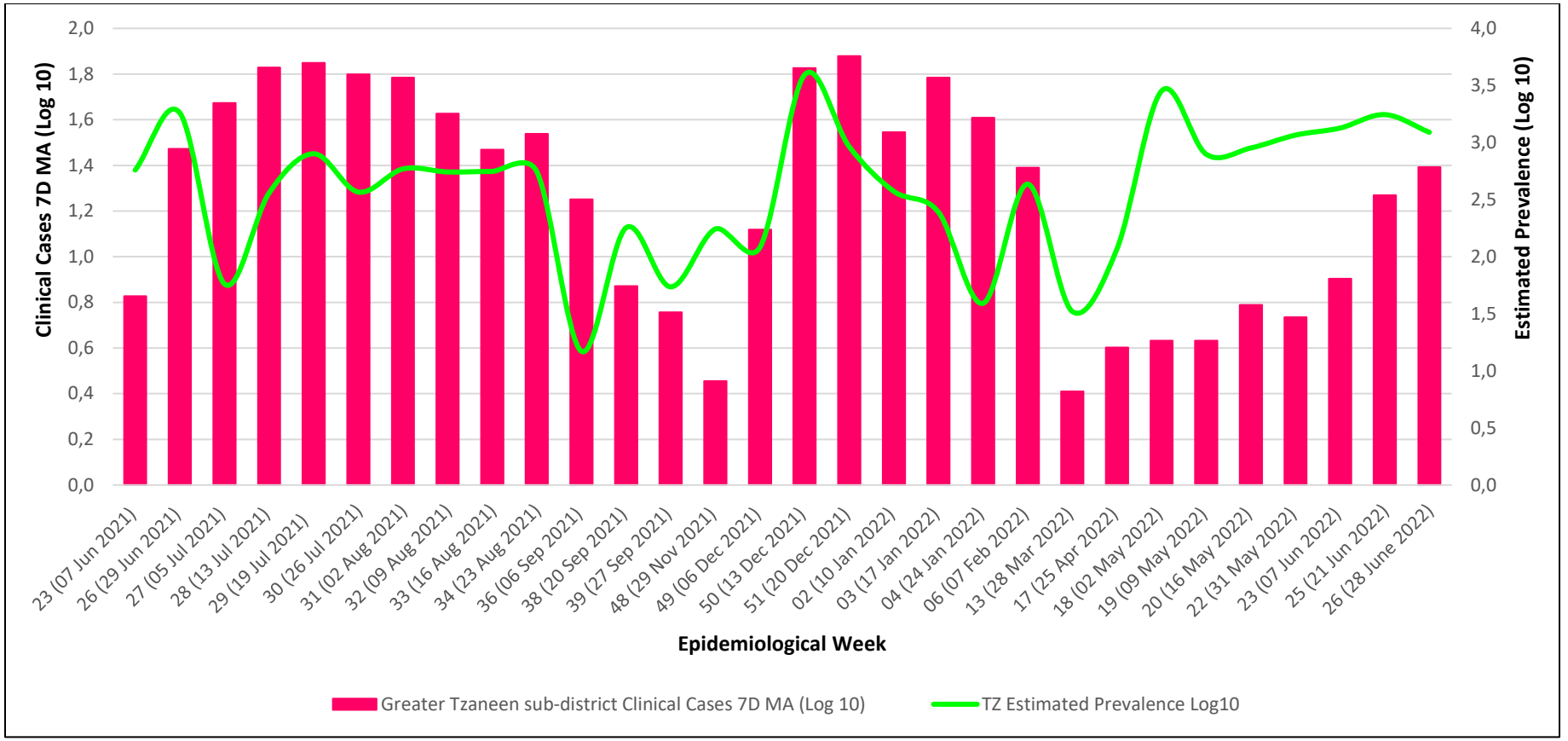


Figure 40: A representation of the predicted number of infected people in the Tzaneen site compared to the reported number of positive clinical cases in the Greater Tzaneen sub-district.

The model predicted that fewer cases occurred in the Tzaneen site during the third wave compared to the fourth wave, and this was similar with reported clinical cases.

3.3.8.2) Nkowankowa WWTP in the Greater Tzaneen Sub-district

Nkowankowa is also situated in the Greater Tzaneen sub-district, in the Mopani District. Per the 2022 national South African census, this sub-district is estimated to have a population of 478,251. Thirty-seven out of forty-three (86%) samples met the inclusion criteria (Epi Week-22, 2021 to Week-26, 2022; 31 May 2021 – 28 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from this sub-district **(Figure 41)**. The model estimated a median 0.1 – 38 (95% CI: 1.4 – 356) infections to have occurred. More infections were predicted to have occurred during the fourth wave (median: 0.7 – 38; 95% CI: 6 – 356 infections) compared to the third wave (0.2 – 29; 95% CI: 2 – 245 infections). The most infections (median: 38; 95% CI: 265 – 356) were predicted to occur in Epi Week-50, 2021 (13 December 2021), which was also the week where the highest viral copies (6143.1 g.c./mL) were detected. A positive correlation ($r = 0.89$; $p = 4e\text{-}14$) was seen between the estimated prevalence and the SARS-CoV-2 viral copies, confirming the precision of the model prediction.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 2 – 75 cases) throughout the study period in the Greater Tzaneen sub-district. Like the estimated prevalence, more clinical cases were recorded during the fourth wave (2 – 75 cases) compared to the third wave (6 – 70 cases). During the last wave (Epi Weeks 17 to 20, 2022; 25 April to 16 May 2022), the model also predicted an increase in infections (median: 2 – 14). A corresponding increase in clinical cases (7D MA: 3 – 24) was seen in the sub-district clinical data, with a positive correlation ($r = 0.67$; $p = 5e\text{-}06$) between the estimated prevalence and positive clinical cases.

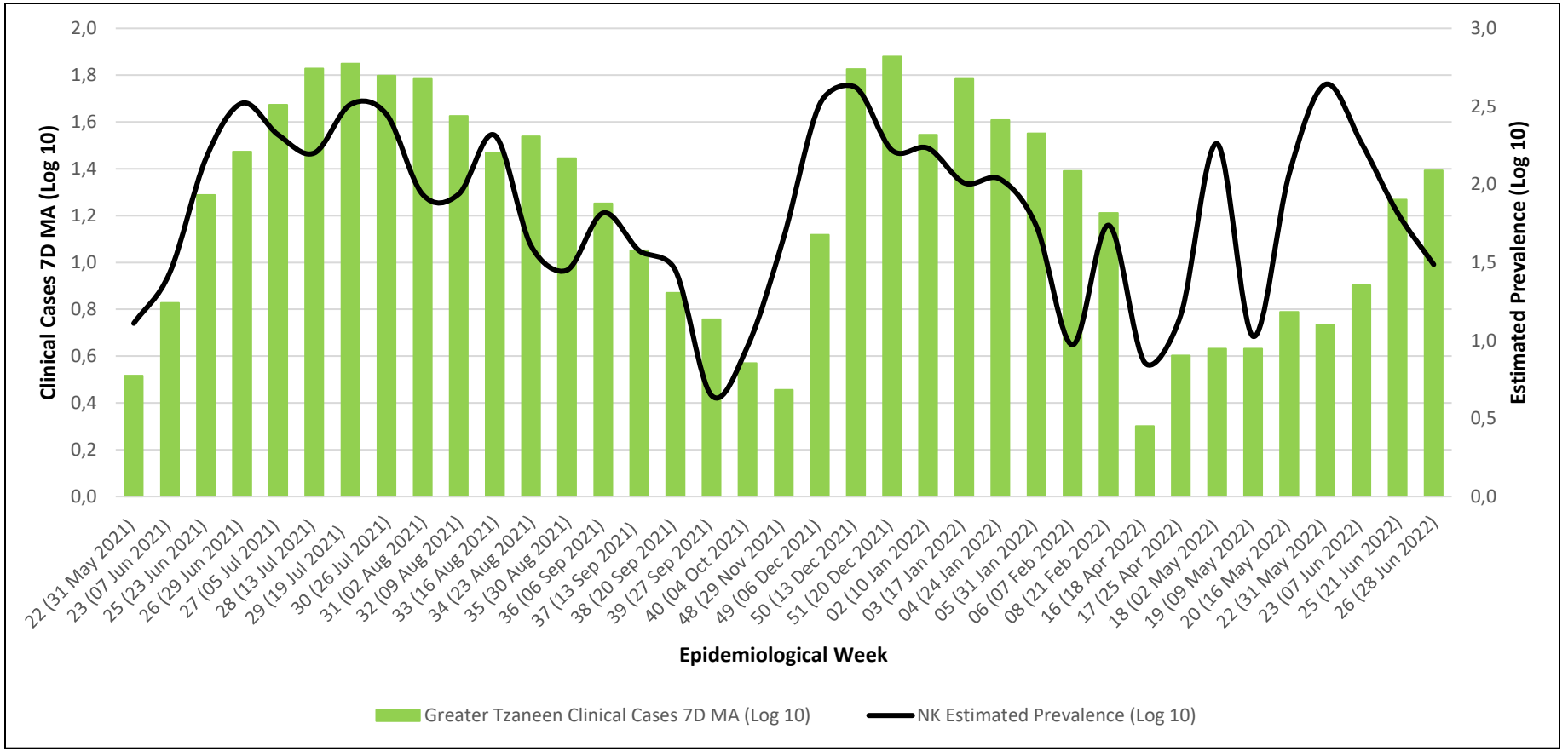


Figure 41: A representation of the predicted number of infected people in the Nkowankowa site compared to the number of positive clinical cases in the Greater Tzaneen sub-district.

The model predicted that fewer cases occurred in the Nkowankowa site during the third wave compared to the fourth wave; this pattern was also seen with reported clinical cases.

3.3.8.3) Ga-Kgapane WWTP in the Greater Letaba Sub-district

Forty out of forty-seven (85%) samples met the inclusion criteria (Epi Week-16, 2021 to Week-26, 2022; 19 April 2021 – 28 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from this sub-district **(Figure 42)**. The median number of people predicted to have been infected in this site ranged from 0.1 – 96 (95% CI: 0.2 – 904). More infections were predicted to occur during the third wave (median: 0.4 – 96; 95% CI: 3 – 904 infections) compared to the fourth wave (0.1 – 6.3; 95% CI: 1 – 65 infections). The highest number of infected cases (median: 96; 95% CI: 664 – 904) were predicted to occur in Epi Week-21, 2021 (24 May 2021), which was also the week where the highest viral copies (78,190.4 g.c./mL) were detected. In addition, a strong correlation ($r = 0.91$; $p = 5e\text{-}17$) was seen between the predicted prevalence and the SARS-CoV-2 viral copies.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 0.1 – 12 cases) in the Greater Letaba sub-district. These were the least number of clinical cases recorded at the sub-district level over a 7D MA for the whole study period, compared to clinical cases recorded by other sub-districts. Like the estimated prevalence, more clinical cases were recorded during the third wave (0.1 – 12 cases) compared to the fourth wave (0.1 – 5 cases). The model also predicted an increase in infections (median: 0.4 – 76) to have occurred during the last wave (Epi Weeks 17 to 23, 2022; 25 April to 07 June 2022). However, no corresponding increases in clinical cases (7D MA: 0.1 – 0.7) were seen in the sub-district clinical data. Similarly, a very poor correlation ($r = 0.006$; $p > 0.05$) existed between the estimated prevalence and positive clinical cases from this sub-district.

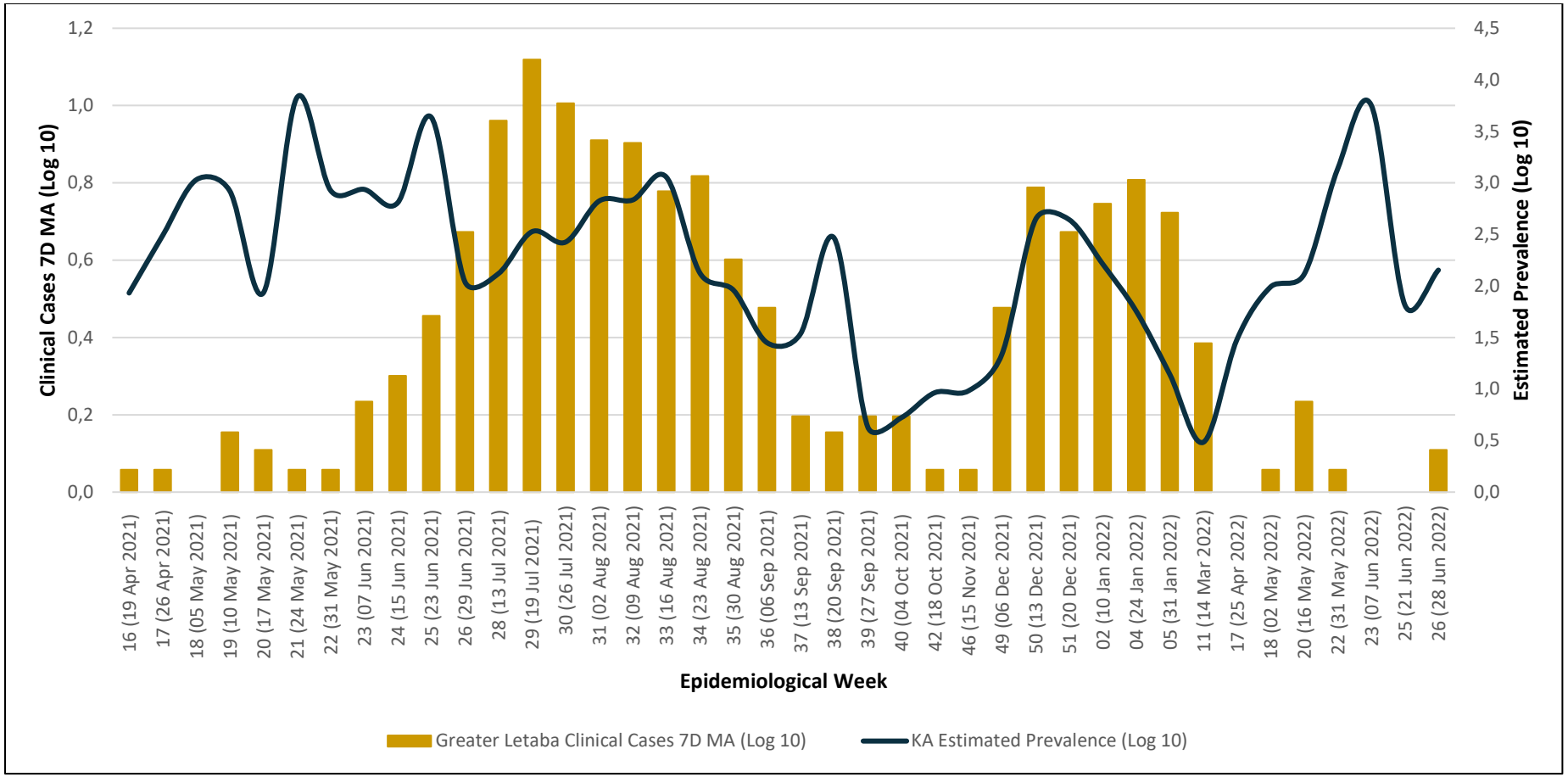


Figure 42: A representation of the predicted number of infected people in the Ga-Kgapane site compared to the number of positive clinical cases in the Greater Letaba sub-district.

The model predicted more cases to have occurred in the Ga-Kgapane site during the third wave compared to the fourth wave. Similar observations were seen with reported cases from the Greater Letaba sub-district.

3.3.8.4) Giyani WWTP in the Greater Giyani Sub-district

Thirty-six out of fifty-one (70.5%) samples met the inclusion criteria Epi Week-26, 2021 and Week-20, 2022 (29 June 2021 – 28 June 2022). The determined predicted prevalence was compared with the 7D MA of clinical cases from this sub-district **(Figure 43)**. The model median number of infected people in this site ranged from 0.1 – 36 (95% CI: 0.5 – 286). The model predicted slightly more infections to have occurred during the fourth wave (Median: 0.5 – 17; 95% CI: 6 – 185) compared to the third wave (Median: 0.5 – 16; 95% CI: 3 – 143). When the estimated prevalence was compared to the SARS-CoV-2 viral copies, a strong Spearman's correlation ($r = 0.99$; $p < 0.001$) was observed, confirming the precision of the model prediction.

Compared to the predicted prevalence, fewer clinical cases were recorded (7D MA: 0.1 – 35 cases) in the Greater Giyani sub-district. Unlike the predicted prevalence, more clinical cases were recorded during the third wave (2.4 – 35 cases) compared to the fourth wave (0.1 – 18.3 cases). The model also predicted the most infections (median: 0.1 – 36) to have occurred during the last wave (Epi Weeks 17 to 22, 2022; 25 April to 31 May 2022). However, the reported clinical cases from this sub-district remained relatively low (7D MA: 0.3 – 4 cases). Similarly, a weak correlation ($r = 0.42$; $p < 0.05$) existed between the estimated prevalence and positive clinical cases from the Greater Giyani sub-district.

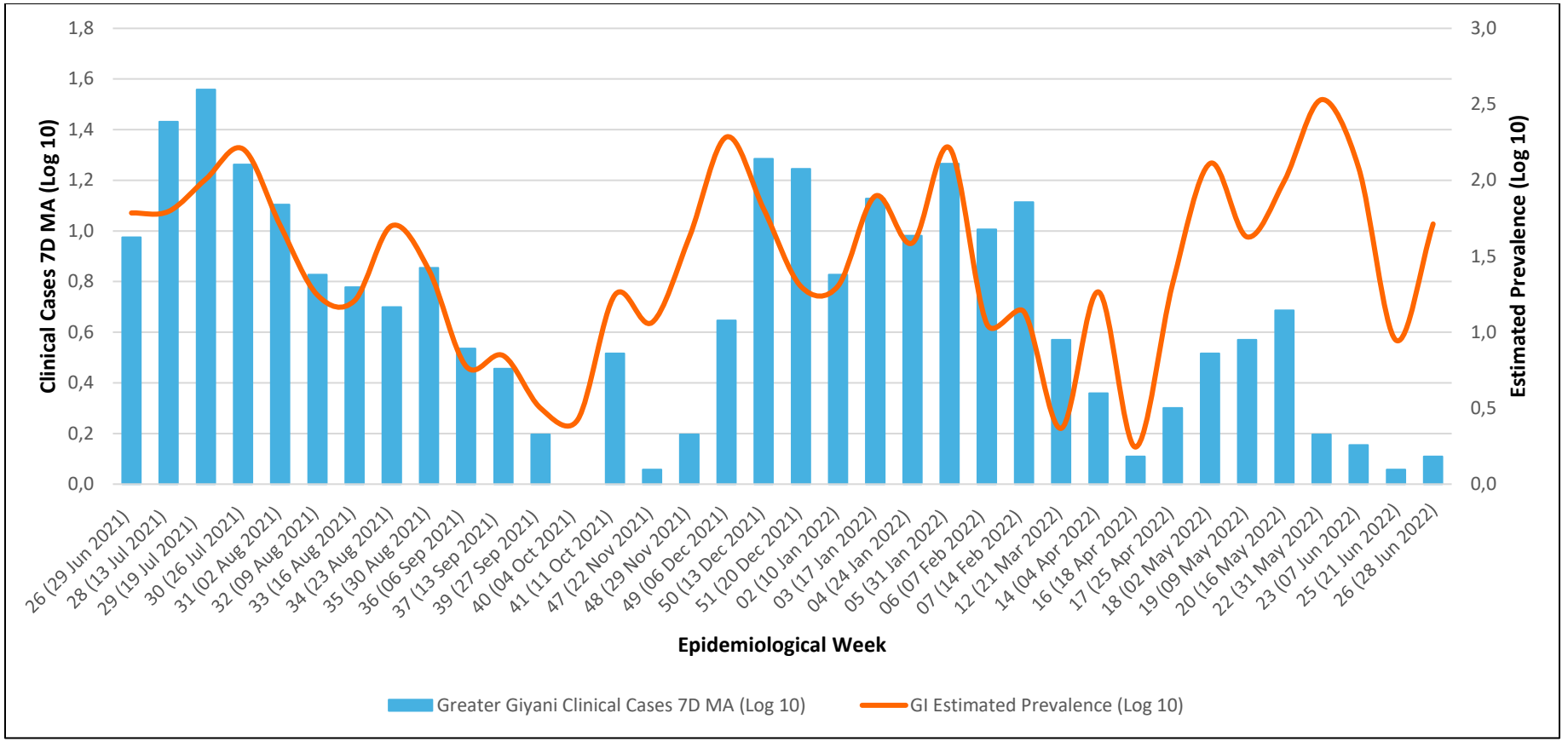


Figure 43: A representation of the predicted number of infected people in the Giyani site compared to the number of positive clinical cases at the Greater Giyani sub-district level.

The model predicted that fewer cases would occur in the Giyani site during the third wave compared to the fourth wave. Similar observations were seen with reported cases from the Greater Giyani sub-district.

3.4) DISCUSSION

Throughout the pandemic, wastewater-based epidemiology (WBE) has been used as a valuable public health tool to complement COVID-19 clinical surveillance efforts. This has been applied worldwide on different scales, from national, regional, city-level, and even building surveillance, to monitor SARS-CoV-2 trends and circulating variants. One major advantage of this tool is its ability to serve as an early warning system, for detection of increased infections (Michael-Kordatou et al., 2020; Bar-Or et al., 2021; Prado et al., 2021; Yaniv et al., 2021) and potential new variants (Crits-Christoph et al., 2021; Gregory et al., 2022). This study sought to establish a wastewater-based surveillance system to monitor the trends of SARS-CoV-2, predict the prevalence those infected, and monitor circulating variants in Limpopo Province, South Africa. This study reports three key findings which highlight the necessity for surveillance systems in rural and peri-urban areas, as a pandemic preparedness measure.

The applied wastewater-based surveillance method for monitoring the trends of SARS-CoV-2 occurrence in the study community is robust and reproducible. The grab sampling method used is cost-effective, compared to composite sampling equipment (Kmush et al., 2022), and can be applied for long-term surveillance in resource-limited settings, where composite samplers may be unaffordable. Additionally, unstable power supply in resource-limited settings may negatively impact functions of the composite sampler and subsequently impair data quality. This is the case with South Africa, which has been experiencing frequent power outages. The applicability of the extraction method used (extracting total RNA from the pellet using the PowerSoil Total RNA extraction kit) has also been demonstrated. First piloted in Connecticut, USA by Peccia et al., (2020), using primary sewage sludge, this method was adapted by Johnson et al., (2021), in Cape Town, South Africa, modified for extraction of total RNA after pellet centrifugation. This shows the applicability of this method in settings with ample resources, as well as limited resources, like our study setting. The recovery efficiency of this protocol has been previously reported (Johnson et al., 2022) and was comparable to what has been documented in other studies (Randazzo, Cuevas-Ferrando, et al., 2020; Randazzo, Truchado, et al., 2020).

The first key finding in this study is the applicability of establishing a WBS system in a rural and peri-urban area. This surveillance system showed potential to serve as an early warning

approximately 2 – 3 weeks prior to commencement of the third and fourth waves in the study sites. This is corroborated by global reports that show similar lead time between incremental detection of SARS-CoV-2 RNA particles in wastewater and corresponding increases in COVID-19 cases. Some studies have reported a lead time of 2 – 14 days (Wurtzer et al., 2020; Feng et al., 2021; Kumar et al., 2021; Olesen, Imakaev and Duvallet, 2021) while others indicate a longer lead time of approximately 5 – 22 days (Kumar et al., 2022; Sangsanont et al., 2022). The 7D-MA is known to be strongly correlated to the SARS-CoV-2 concentration (Swift et al., 2023), and this decreases inherent variability that arises due to differences in sampling time and temperature between samples. These differences may impair the potential of SARS-CoV-2 monitoring to be used as a complimentary tool for public health (Li et al., 2022). In 6/8 (75%) of the study sites, there was a strong correlation between the 7D-MA of clinical cases and the SARS-CoV-2 concentration observed. In the two sites (Thohoyandou and Ga-Kgapane) with a weak correlation, this could have been because the clinical data from the sub-district used as a proxy for comparison, poorly represented cases from the study site. Interestingly, increased SARS-CoV-2 concentrations observed in all study sites during the last wave (18 April – 21 June 2022), showed no corresponding increases in clinical cases in 5/6 sub-districts. This could have been due to the decreased severity of the dominantly circulating Omicron variant, as well as a better immune response to the virus, especially among a highly vaccinated population. As of 29 May 2023, per the COVID-19 Online Resource and News Portal (https://sacoronavirus.co.za/latest-vaccine-statistics/), approximately 891,230 and 814,414 vaccines have been administered in the Vhembe and Mopani districts, respectively. This may explain the no increase in clinical cases during this last wave, due to decreased infection severity, implying that more infections were asymptomatic in nature (Nourbakhsh et al., 2022; Ando et al., 2023). This postulates the hypothesis that SARS-CoV-2 may remain endemic in the population, thus continuous surveillance is necessary for rapid detection of any changes which may arise in the population. Normalizing SARS-CoV-2 wastewater surveillance (WWS) data is recommended, since it addresses variations in fecal strength, and potential dilution in the sewershed. This is mostly done using the daily flowrate, population size, and fecal biomarkers (Maal-Bared et al., 2023). For this study, SARS-CoV-2 concentrations were normalized using the flowrate and population size, and compared to the sub-district clinical cases. No significant changes were observed between normalized wastewater data compared to the sub-district clinical cases, and non-normalized data. Similar

findings indicating that normalization did not improve associations with clinical data have also been reported (Feng et al., 2021; Greenwald et al., 2021). Long-term surveillance is also advantageous, since it provides a full picture of patterns that occurred during the surveillance period. This longitudinal monitoring aided comprehension of the magnitude (potential degree of infections in the community, based on average viral copies) of all three waves that occurred. The third wave (19 April – 06 September 2021), dominated by the Delta VOC was the longest wave (lasting for 21 weeks). However, the fourth wave (15 November 2021 – 21 February 2022), characterized by the omicron wave VOC was much shorter (lasting for 13 weeks), but had the greatest magnitude (category 4: 6134.4 g.c/mL). This categorization was developed to aid public health personnel to better understand wastewater surveillance results, which could be beneficial for policy implementation. Weekly surveillance data from the study sites was relayed to the respective municipal authorities for possible action.

Secondly, data on the proportions of people infected are crucial to monitor the progress of epidemics and guide public health responses, meant to curb transmission. Using the Monte Carlo (MC) mathematical model the number of infected people within the surveyed catchment area was predicted. Approximately 56,792 and 12,010 infections were predicted to have occurred in the surveyed sites of the Vhembe and Mopani districts, respectively. This predicted prevalence had a strong correlation to the viral load detected in all study sites. The determined predicted prevalence is strictly based on the population connected to sewer facilities in the eight study sites, thus, it does not represent the entire population in the Vhembe and Mopani districts. More infections were predicted to have occurred in the four study sites in the Vhembe district (56,792 infections), compared to actual number of reported cases at the district level (22,475 cases). This high predicted prevalence could be a culmination of the number of active, mild and asymptomatic cases in the study area. Those with mild infections, who do not present for testing, as well as those with asymptomatic infections are generally absent from the total number of clinical cases. This showed the relevance of wastewater-based surveillance in showing the real magnitude of the epidemic in the study population. In the Mopani district, 12,010 infections were predicted to have occurred in the four study sites throughout the surveillance period. However, compared to the actual number of clinical cases (16,476 cases) for all three sub-districts where the study sites are located, this predicted number was much lower. This lower predicted number of infections may be explained by the fact that RNA concentrations used for modelling is based solely on homes in the study sites

connected to the sewage system. Thus, important shedding data of those infected within the study sites but not connected to the sewage system was missed. This is corroborated by reports indicating that approximately 37% households in South Africa are not connected to a sewage system, and rely on pit toilets, septic tanks, or open defecation practices (Street et al., 2020). Other factors which may have also influenced the predicted prevalence include shedding rate, population size, in-sewer factors and sampling strategy, which may have contributed to impairment of SARS-CoV-2 RNA concentrations (Bertels et al., 2022; Liu et al., 2022). However, to ensure that shedding events were not missed, samples were collected bi-weekly (regularly on Mondays, and additionally on Wednesdays) for six weeks (16 February 2022 – 23 March 2022). Obtained SARS-CoV-2 viral copies were compared to investigate the difference between both days of collection. We noted there was no marked difference between the viral copies detected on both days. Thus, despite these uncertainties, the results obtained are sufficient to serve as an early warning system. The model was also able to show a difference in infection rates between the third and fourth waves. In the 3/8 (37.5%) sites, the model predicted more infections to have occurred during the fourth wave compared to the third wave. This contradicted clinical data reported during the fourth wave showing fewer cases. This phenomenon has been attributed to lower disease severity of the dominantly circulating variant during the fourth wave (Omicron VOC) compared to the third wave (Delta VOC) (Sigal et al., 2022). This decreased severity is associated with increased immunity, due to more vaccination uptake in various populations, resulting in more mild and asymptomatic cases, which are not tested clinically. Thus, developing strategies to account for homes within surveillance sites not connected to the sewage systems, is necessary. This is particularly important for low-income areas without water-borne sewage systems, which is more common in majority of South Africa, and other low-and middle-income countries (LMIC), worldwide. One way to improve surveillance results, and account for homes not connected to sewage systems, is through decentralized surveillance (Gonçalves et al., 2022). This method involves sample collection from polluted surface waters such as rivers, and sewer interceptors.

Third, we emphasize the necessity for wastewater genomic surveillance in both rural and urban areas. Although genomic surveillance of SARS-CoV-2 was prioritized throughout the pandemic, it was done mainly using clinical samples through next generation sequencing (Li et al., 2020; Aggarwal et al., 2022; Arévalo et al., 2022). Settings with fewer resources need cheaper genotyping methods for regular monitoring and real-time results of circulating

variants, which could be instrumental for decision making. One of such cost-efficient methods is allele-specific genotyping (ASG). Using this ASG method, Delta and Omicron VOCs were detected in the study sites by 03 May 2021. This implies that these variants may have been in South Africa earlier than originally thought. Detection of both variants so early in the study sites may have been due to infected asymptomatic individuals returning to South Africa and subsequently transmitting the virus in the community. Infected individuals traveling from one country to another have been implicated in the distribution of variants across the world (Kucharski et al., 2022). For example, the Limpopo province shares a border with three surrounding nations (Botswana, Zimbabwe, and Mozambique), thus cross-border transmission may have occurred. To our knowledge, this is the first study reporting identification of the Omicron variant as early as May 2021, predating reports of its identification in South Africa in November 2021. However, one report from the Netherlands also indicated identification of the Omicron variant in two previous patient samples, a few days prior to alerts raised by South Africa of its discovery (RVIM Netherlands, 2021). Whole genome sequencing was used to confirm results obtained through allele-specific genotyping. Interestingly, through WGS, the Delta and Omicron VOCs were detected to be circulating even earlier (January and February 2021, respectively) than what was observed through ASG. The Omicron variant is postulated to have arisen due to either its circulation and evolution in a population with little surveillance and sequencing, or it may have developed in a chronically infected COVID-19 patient, and finally it may have evolved from non-human species, which spilled into the human population (Callaway, 2022; Magiorkinis, 2023). Other reports from South Africa which used a similar method of genotyping showed that prior to the third wave, the Beta VOC was mostly predominant, until June 2021, when the Delta variant took over dominance (Johnson et al., 2022; Mangwana et al., 2022). Using whole genome sequencing (WGS), the NICD nationwide wastewater surveillance program, and the Network for Genomic Surveillance in South Africa (NGS-SA) corroborated these findings of the Delta variant dominating the third wave in all nine provinces (NICD, 2021). During the fourth wave, genomic epidemiology from both the national wastewater surveillance program and NGS-SA showed that Omicron principally characterized the fourth wave in South Africa (NGS-SA, 2021, NICD, 2022). Since very few samples met the inclusion criteria, ASG results give a general overview of VOCs present in the study sites, thus accurate comparison of the most dominant variant in the study sites during the third and fourth waves was not possible. Other studies have also reported the

benefit of using ASG for variant detection in both clinical and wastewater specimen (Heijnen et al., 2021; La Rosa et al., 2021; Yu et al., 2022). Just like the current study, one retrospective study conducted in Indianna, USA, reported utilization of the TaqMan SARS-CoV-2 mutation panel for variant determination in patient specimen collected between March – July 2021. A total of 14 different TaqMan SNP genotyping assays which target the Spike and ORF8 genes, as well as seven controls representing the Alpha, Beta, Gamma, Kappa, and SARS-CoV-2 Wuhan-wildtype, were used for assay validation (Neopane et al., 2021). Their results showed that mutations for each control were present; the Alpha, Beta, Delta, Iota, Gamma, Zeta, Kappa, and Epsilon variants were detected in patient specimen. Congruence in variant assignment and selected samples subjected to sequencing was also observed, highlighting the efficacy of the method between. Contrarily, this study utilized only seven TaqMan SNP genotyping assays for variant determination. However, of the samples subjected to WGS, the Delta variant was most dominant (45%) across the study sites, followed by Omicron (31.7%), and Beta VOC occurring at low frequencies (5%).

3.5) LIMITATIONS OF THE STUDY

While wastewater-based surveillance is applicable in this rural setting, this can be seen considering the following limitations. First, the clinical case data used for correlation with detected SARS-CoV-2 RNA copies were obtained from the sub-district level, not the towns which the WWTPs and WSPs were located. This may have been the reason why a low Spearman's correlation was observed for two study sites, when non-normalized SARS-CoV-2 RNA copies was compared to sub-district clinical cases **(section 3.3.3, Figure 18 and section 3.3.4, Figure 31)**. A similarly weak correlation was also obtained when comparing the predicted prevalence per site to sub-district clinical cases **(section 3.3.6)**. These findings highlight some shortcomings that may be experienced in LMIC, where testing facilities are not localized within communities.

Secondly, the prediction model used for computation is highly sensitive to the key input variables. These input variables include: the SARS-CoV-2 RNA copies in stool, followed by the RNA copies in wastewater, and the quantity of feces shed per person per day. Information on the SARS-CoV-2 RNA copies in per gram of feces, and the gram of feces released per

person per day in South Africa is still lacking; thus, obtaining this information is necessary for improved infection estimates. The estimated prevalence of infected people presented here is exclusively based on RNA copies detected in wastewater, which only accounts for homes connected to the WWTP in the study sites. Thus, the total number of infected people in sites not connected to the sewage systems may be underrepresented.

Third, optimization of the allele-specific genotyping protocol is necessary, for adequate genomic surveillance. One shortcoming of this protocol is the inclusion criteria which only caters for SARS-CoV-2 positive samples with concentrations ≥1500g.c./mL. This implies that VOCs present in samples which do not meet this criterion will be unknown, thus, information on circulating variants in times when viral concentrations in the study sites are low will likely be missed. For example, of the 365 SARS-CoV-2 positive samples obtained throughout the surveillance period, only 21.9% of them met the inclusion criteria. Thus, based on this criterion, information on the VOCs presents in the remaining 78.1% will be unknown. This may explain why the ASG results showed similar frequency of occurrence for all investigated VOCs. For example, during the third wave of COVID-19 infections in South Africa, the Delta variant was most dominant clinically, however, ASG results indicate that the Delta variant was only slightly more dominant (29%) than the Alpha, Beta and Omicron VOCs (22%, 24% and 25%, respectively). ASG results were further confirmed by whole genome sequencing using the ATOPlex NGS platform. Based on WGS findings, the Delta and Omicron variants were present in samples collected as early as January and February 2021, respectively (elaborated in section 4.3.2). In addition, the frequency of occurrence of detected VOCs after WGS analysis were markedly different from analysis obtained through ASG. While WGS showed the Alpha, Beta, Delta and Omicron VOCs occurred at frequencies of 5%, 5%, 45% and 32%, respectively, ASG indicated these variants were present at 26%, 24%, 22%, and 28%, respectively. Further comparison was made to determine concordance in variant assignment for both techniques, and concordance was only observed in 51.2% samples (elaborated in section 4.3.5). These differences in variant assignments highlight another shortcoming of the applied ASG method, which may wrongly assign variants, and thus convey incorrect results about variant circulation in communities determined through WBS. Therefore, it is important to include other specific variant-defining or lineage-defining mutations, to improve specificity and reliability of ASG in variant determination. ASG for variant monitoring in wastewater could also be implemented using droplet digital RT-PCR (RT-ddPCR). This technique uses samples

and reagents partitioned in large number (~10000 to 20000) “water in oil” droplets, then the PCR reactions are performed on single molecules in individual closed droplets. Thus, rare mutations can be detected and closely related sequences discriminated through probe binding kinetics (Heijnen et al., 2021). In addition, other qPCR-based approaches can be applied to detect Omicron mutations and discriminate omicron lineages (Kong et al., 2023; Li et al., 2024).

3.6) CONCLUSION AND RECOMMENDATIONS

These findings demonstrate the usefulness of wastewater surveillance for early detection of disease, tracking infection trends, and highlighting near real-time variants circulating in rural and peri-urban settings. Detecting Delta and Omicron VOCs so early in the study sites emphasize the need for pandemic preparedness efforts to be extended to all geographic regions, for comprehensive monitoring and epidemic prevention. This study also shows the impact that non-sewered systems have in skewing surveillance efforts, which could influence the early warning system, and potentially undermine public health policies developed from such surveillance studies. This beckons the need to address these loopholes to efficiently harness the potential of this tool, and enhance pandemic preparedness efforts, particularly in resource constrained settings. This study lays a foundation for implementing wider surveillance studies to track SARS-CoV-2 and other enteric pathogens circulating in this study population.

3.7) REFERENCES

Aggarwal, D. et al. (2022) 'The role of viral genomics in understanding COVID-19 outbreaks in long-term care facilities', The Lancet Microbe, 3(2), pp. e151–e158. https://doi.org/10.1016/S2666-5247(21)00208-1

Ahmed, W. et al. (2020) 'First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community', Science of the Total Environment, 728. https://doi.org/10.1016/j.scitotenv.2020.138764

Amoah, I.D. et al. (2022) 'Effect of selected wastewater characteristics on estimation of SARS-CoV-2 viral load in wastewater.', Environmental Research, 203, pp. 111877. https://doi.org/10.1016/j.envres.2021.111877

Ando, H. et al. (2023) 'Wastewater-based prediction of COVID-19 cases using a highly sensitive SARS-CoV-2 RNA detection method combined with mathematical modeling', Environment International, 173, pp. 107743. https://doi.org/10.1016/j.envint.2023.107743

Arévalo, M.T. et al. (2022) 'A Rapid, Whole Genome Sequencing Assay for Detection and Characterization of Novel Coronavirus (SARS-CoV-2) Clinical Specimens Using Nanopore Sequencing', Frontiers in Microbiology, 13, pp. 910955. https://doi.org/10.3389/fmicb.2022.910955

Bar-Or, I. et al. (2021) 'Regressing SARS-CoV-2 Sewage Measurements Onto COVID-19 Burden in the Population: A Proof-of-Concept for Quantitative Environmental Surveillance', Frontiers in Public Health, 9, pp. 561710. https://doi.org/10.3389/fpubh.2021.561710

Bertels, X. et al. (2022) 'Factors influencing SARS-CoV-2 RNA concentrations in wastewater up to the sampling stage: A systematic review', Science of the Total Environment, 820. https://doi.org/10.1016/j.scitotenv.2022.153290

Bi, C. et al. (2021) 'Simultaneous detection and mutation surveillance of SARS-CoV-2 and multiple respiratory viruses by rapid field-deployable sequencing', Med (New York, N.Y.), 2(6), pp. 689-700.e4. https://doi.org/10.1016/j.medj.2021.03.015

Brouwer, A.F. et al. (2018) 'Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data', Proceedings of the National Academy of

Sciences of the United States of America, 115(45), pp. E10625–E10633. https://doi.org/10.1073/pnas.1808798115

Callaway, E. (2022) 'How months-long COVID infections could seed dangerous new variants', Nature, 606(7914), pp. 452–455. https://doi.org/10.1038/d41586-022-01613-2

Choi, P.M. et al. (2018) 'Wastewater-based epidemiology biomarkers: Past, present and future', TrAC - Trends in Analytical Chemistry, 105, pp. 453–469. https://doi.org/10.1016/j.trac.2018.06.004

Choi, P.M. et al. (2019) 'Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology', Proceedings of the National Academy of Sciences of the United States of America, 116(43), pp. 21864–21873. https://doi.org/10.1073/pnas.1910242116

Crits-Christoph, A. et al. (2021) 'Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants', mBio, 12(1) pp. e02703-20. https://doi.org/10.1128/mBio.02703-20

Dhar, M.S. et al. (2021) 'Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India', Science, 374(6570), pp. 995 – 999. https://doi.org/10.1126/science.abj9932

Dzinamarira, T. et al. (2022) 'Utilization of SARS-CoV-2 Wastewater Surveillance in Africa-A Rapid Review.', International journal of environmental research and public health, 19(2), pp. 969. https://doi.org/10.3390/ijerph19020969

Faleye, T.O.C. et al. (2021) 'Wastewater-Based Epidemiology and Long-Read Sequencing to Identify Enterovirus Circulation in Three Municipalities in Maricopa County, Arizona, Southwest United States between June and October 2020', Viruses, 13(9), pp. 1803. https://doi.org/10.3390/v13091803

Faria, N.R. et al. (2021) 'Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil', Science, 372(6544), pp. 815 – 821. https://doi.org/10.1126/science.abh2644

Feng, S. et al. (2021) 'Evaluation of Sampling, Analysis, and Normalization Methods for SARS-CoV-2 Concentrations in Wastewater to Assess COVID-19 Burdens in Wisconsin

Communities', ACS EST Water, 1(8), pp. 1955–1965. https://doi.org/10.1021/acsestwater.1c00160

Gerrity, D. et al. (2021) 'Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: Methodology, occurrence, and incidence/prevalence considerations', Water Research X, 10, pp. 100086. https://doi.org/10.1016/j.wroa.2020.100086

Gonçalves, J. et al. (2022) 'Centralized and decentralized wastewater-based epidemiology to infer COVID-19 transmission - A brief review.', One health, 15, pp. 100405. https://doi.org/10.1016/j.onehlt.2022.100405

Gonzalez, R. et al. (2020) 'COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology', Water research, 186, pp. 116296. https://doi.org/10.1016/j.watres.2020.116296

Greenwald, H.D. et al. (2021) 'Tools for interpretation of wastewater SARS-CoV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay Area.', Water research X, 12, pp. 100111. https://doi.org/10.1016/j.wroa.2021.100111

Gregory, D.A. et al. (2021) 'Monitoring SARS-CoV-2 Populations in Wastewater by Amplicon Sequencing and Using the Novel Program SAM Refiner.', Viruses, 13(8), pp. 1647. https://doi.org/10.3390/v13081647

Gregory, D.A. et al. (2022) 'Genetic diversity and evolutionary convergence of cryptic SARS-CoV-2 lineages detected via wastewater sequencing', PLoS Pathogens, 18(10), pp. 1–25. https://doi.org/10.1371/journal.ppat.1010636

Harper, H. et al. (2021) 'Detecting SARS-CoV-2 variants with SNP genotyping', PLoS ONE, 16 (2), pp. e0243185. https://doi.org/10.1371/journal.pone.0243185

Hellmér, M. et al. (2014) 'Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks', Applied and Environmental Microbiology, 80(21), pp. 6771–6781. https://doi.org/10.1128/AEM.01981-14

Hill, V. et al. (2022) 'The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK', Virus Evolution, 8(2), pp. 1 – 13. https://doi.org/10.1093/ve/veac080

Hovi, T. et al. (2012) 'Role of environmental poliovirus surveillance in global polio eradication and beyond', Epidemiology and Infection, 140(1), pp. 1–13. https://doi.org/10.1017/S095026881000316X

Iwu-Jaja, C. et al. (2023) 'The role of wastewater-based epidemiology for SARS-CoV-2 in developing countries: Cumulative evidence from South Africa supports sentinel site surveillance to guide public health decision-making', Science of the Total Environment, 903, pp. 165817. https://doi.org/10.1016/j.scitotenv.2023.165817

Johnson, R. et al. (2021) 'Qualitative and quantitative detection of SARS-CoV-2 RNA in untreated wastewater in Western Cape Province, South Africa.', South African medical journal, 111(3), pp. 198–202. https://doi.org/10.7196/SAMJ.2021.v111i3.15154

Johnson, R. et al. (2022) 'Tracking the circulating SARS-CoV-2 variant of concern in South Africa using wastewater-based epidemiology.', Scientific reports, 12(1), pp. 1182. https://doi.org/10.1038/s41598-022-05110-4

Jones, D.L. et al. (2020) 'Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19', Science of the Total Environment, 749, pp. 141364. https://doi.org/1016/j.scitotenv.2020.141364

Kaplan, E.H. et al. (2022) 'Scaling SARS-CoV-2 wastewater concentrations to population estimates of infection', Scientific Reports, 12(1), pp. 10–13. https://doi.org/10.1038/s41598-022-07523-7

Kucharski, A.J. et al. (2022) 'Travel measures in the SARS-CoV-2 variant era need clear objectives', The Lancet, 399(10333), pp. 1367–1369. https://doi.org/10.1016/S0140-6736(22)00366-X

Kumar, M. et al. (2021) 'Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation.', Environmental Research, 196, pp. 110946. https://doi.org/10.1016/j.envres.2021.110946

Kumar, M. et al. (2022) 'Lead time of early warning by wastewater surveillance for COVID-19: Geographical variations and impacting factors', Chemical Engineering Journal, 441, pp. 135936. https://doi.org/10.1016/j.cej.2022.135936

Li, J. et al. (2020) 'Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing', Scientific Reports, 10, pp. 17492. https://doi.org/10.1038/s41598-020-74656-y

Li, L. et al. (2022) 'Longitudinal monitoring of SARS-CoV-2 in wastewater using viral genetic markers and the estimation of unconfirmed COVID-19 cases', Science of the Total Environment, 817, pp. 152958. https://doi.org/10.1016/j.scitotenv.2022.152958

Li, X. et al. (2022) 'SARS-CoV-2 shedding sources in wastewater and implications for wastewater-based epidemiology', Journal of Hazardous Materials, 432, pp. 128667. https://doi.org/10.1016/j.jhazmat.2022.128667

Liu, P. et al. (2022) 'A sensitive, simple, and low-cost method for COVID-19 wastewater surveillance at an institutional level.', The Science of the total environment, 807 (Pt 3), pp. 151047. https://doi.org/10.1016/j.scitotenv.2021.151047

Lodder, W. and de Roda Husman, A.M. (2020) 'SARS-CoV-2 in wastewater: potential health risk, but also data source', The Lancet Gastroenterology and Hepatology, 5(6), pp. 533–534. https://doi.org/10.1016/S2468-1253(20)30087-X

Maal-Bared, R. et al. (2023) 'Does normalization of SARS-CoV-2 concentrations by Pepper Mild Mottle Virus improve correlations and lead time between wastewater surveillance and clinical data in Alberta (Canada): comparing twelve SARS-CoV-2 normalization approaches', Science of the Total Environment, 856 (Pt 1), pp. 158964. https://doi.org/10.1016/j.scitotenv.2022.158964

Magiorkinis, G. (2023) 'On the evolution of SARS-CoV-2 and the emergence of variants of concern', Trends in Microbiology, 31(1), pp. 5–8. https://doi.org/10.1016/j.tim.2022.10.008

Mangwana, N. et al. (2022) 'Sewage surveillance of SARS-CoV-2 at student campus residences in the Western Cape, South Africa', Science of the Total Environment, 851(1), pp. 158028. https://doi.org/10.1016/j.scitotenv.2022.158028

Medema, G. et al. (2020) 'Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands', Environmental Science and Technology Letters, 7(7), pp. 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Michael-Kordatou, I. et al. (2020) 'Sewage analysis as a tool for the COVID-19 pandemic response and management: the urgent need for optimised protocols for SARS-CoV-2 detection and quantification.', Journal of environmental chemical engineering, 8(5), pp. 104306. https://doi.org/10.1016/j.jece.2020.104306

Nemudryi, A. et al. (2020) 'Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater', Cell Reports Medicine, 1(6), pp. 100098. https://doi.org/10.1016/j.xcrm.2020.100098

Neopane, P. et al. (2021) 'SARS-CoV-2 variants detection using TaqMan SARS-CoV-2 mutation panel molecular genotyping assays', Infection and Drug Resistance, 14, pp. 4471–4479. https://doi.org/10.2147/IDR.S335583

Nourbakhsh, S. et al. (2022) 'A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities', Epidemics, 39, pp. 100560. https://doi.org/10.1016/j.epidem.2022.100560

Olesen, S.W. et al. (2021) 'Making waves: Defining the lead time of wastewater-based epidemiology for COVID-19.', Water Research, 202, pp. 117433. https://doi.org/10.1016/j.watres.2021.117433

Peccia, J. et al. (2020) 'SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics', medRxiv, https://doi.org/10.1101/2020.05.19.20105999

Peng, L. et al. (2020) 'SARS-CoV-2 can be detected in urine, blood, anal swabs, and oropharyngeal swabs specimens', Journal of Medical Virology, 92, pp. 1676–1680. https://doi.org/10.1002/jmv.25936

Pillay, L. et al. (2021) 'Monitoring changes in COVID-19 infection using wastewater-based epidemiology: A South African perspective', Science of the Total Environment, 786, pp, 147273. https://doi.org/10.1016/j.scitotenv.2021.147273

Pillay, L. et al. (2022) 'Potential and Challenges Encountered in the Application of Wastewater-Based Epidemiology as an Early Warning System for COVID-19 Infections in South Africa' ACS EST Water, 2(11), pp. 2105–2113. https://doi.org/10.1021/acsestwater.2c00049

Prado, T. et al. (2021) 'Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil.', Water Research, 191, pp. 116810. https://doi.org/10.1016/j.watres.2021.116810

Randazzo, W. et al. (2020) 'Metropolitan wastewater analysis for COVID-19 epidemiological surveillance', International Journal of Hygiene and Environmental Health, 230, pp. 113621. https://doi.org/10.1016/j.ijheh.2020.113621

Randazzo, W. et al. (2020) 'SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area', Water Research, 181, pp. 115942. https://doi.org/10.1016/j.watres.2020.115942

La Rosa, G. et al. (2021) 'Rapid screening for SARS-CoV-2 variants of concern in clinical and environmental samples using nested RT-PCR assays targeting key mutations of the spike protein', Water Research, 197, pp. 117104. https://doi.org/10.1016/j.watres.2021.117104

La Rosa, G. et al. (2021) 'SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring', Science of the Total Environment, 750, pp. 141711. https://doi.org/10.1016/j.scitotenv.2020.141711

QIAGEN RNeasy ® PowerSoil ® Total RNA Kit Handbook, 2017.

QuantStudio™ Design and Analysis Software v2 Genotyping Analysis Module User Guide, 2023, Publication Number MAN0018749 (accessed September 2021).

Rothman, J.A. et al. (2021) 'RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants.', Applied and environmental microbiology, 87(23), pp. e0144821. https://doi.org/10.1128/AEM.01448-21

Sangsanont, J. et al. (2022) 'SARS-CoV-2 RNA surveillance in large to small, centralized wastewater treatment plants preceding the third COVID-19 resurgence in Bangkok, Thailand', Science of the Total Environment, 809, pp. 151169. https://doi.org/10.1016/j.scitotenv.2021.151169

Sherchan, S.P. et al. (2020) 'First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA', The Science of the Total Environment, 743, pp. 140621. https://doi.org/10.1016/j.scitotenv.2020.140621

Sigal, A. et al. (2022) 'Estimating disease severity of Omicron and Delta SARS-CoV-2 infections', Nature Reviews Immunology, 22(5), pp. 267–269. https://doi.org/10.1038/s41577-022-00720-5

Smith, D.B. et al. (2016) 'The use of human sewage screening for community surveillance of hepatitis E virus in the UK', Journal of Medical Virology, 88(5), pp. 915–918. https://doi.org/10.1002/jmv.24403

Street, R. et al. (2020) 'Wastewater surveillance for Covid-19: An African perspective', Science of the Total Environment, 743, pp. 140719. https://doi.org/10.1016/j.scitotenv.2020.140719

Swift, C.L. et al. (2023) 'SARS-CoV-2 concentration in wastewater consistently predicts trends in COVID-19 case counts by at least two days across multiple WWTP scales', Environmental Advances, 11, pp. 100347. https://doi.org/10.1016/j.envadv.2023.100347

Takemae, N. et al. (2022) 'Development of new SNP genotyping assays to discriminate the Omicron variant of SARS-CoV-2', Japanese Journal of Infectious Diseases, 75 (4), pp. 411 – 414. https://doi.org/10.7883/yoken.JJID.2022.007

Tegally, H. et al. (2021) 'Detection of a SARS-CoV-2 variant of concern in South Africa', Nature 592(7854), pp. 438 – 443. https://doi.org/10.1038/s41586-021-03402-9

Tegally, H. et al. (2022) 'The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance', Science, 378(6615), pp. eabq5358. https://doi.org/10.1126/science.abq5358

Tiwari, S. and Dhole, T.N. (2018) 'Assessment of enteroviruses from sewage water and clinical samples during eradication phase of polio in North India', Virology Journal, 15(1), pp. 157. https://doi.org/10.1186/s12985-018-1075-7

Viana R. et al. (2022) 'Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa', Nature, 603(7902), pp. 679 – 686. https://doi.org/10.1038/s41586-022-04411-y

Wilhelm, A. et al. (2022) 'Wastewater surveillance allows early detection of SARS-CoV-2 omicron in North Rhine-Westphalia, Germany.', The Science of the total environment, 846, pp. 157375. https://doi.org/10.1016/j.scitotenv.2022.157375

Wölfel, R. et al. (2020) 'Virological assessment of hospitalized patients with COVID-2019', Nature, 581(7809), pp. 465 – 469. https://doi.org/10.1038/s41586-020-2196-x

Wu, F. et al. (2020) 'SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases', mSystems, 5(4), pp. e00614-20. https://doi.org/10.1128/mSystems.00614-20

Wurtzer, S. et al. (2020) 'Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in wastewater, Greater Paris, France, 5 March to 23 April 2020', Eurosurveillance, 25(50), pp. 2000776. https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000776

Yaniv, K. et al. (2021) 'City-level SARS-CoV-2 sewage surveillance', Chemosphere, 283, pp. 131194. https://doi.org/10.1016/j.chemosphere.2021.131194

Zhao, L. et al. (2022) 'Environmental surveillance of SARS-CoV-2 RNA in wastewater systems and related environments in Wuhan: April to May of 2020.', Journal of environmental sciences (China), 112, pp. 115–120. https://doi.org/10.1016/j.jes.2021.05.005

Zhu, Y. et al. (2021) 'Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks', Science of the Total Environment, 767, pp. 145124. https://doi.org/10.1016/j.scitotenv.2021.145124

WEBSITES

WHO (2023), 'COVID-19 dashboard' (https://covid19.who.int/) (accessed 02 November 2023)

CDC (2023), 'Symptoms of COVID-19' (https://www.cdc.gov/covid/signs-symptoms/index.html) (accessed 20 July 2023)

WHO (2023), 'Tracking SARS-CoV-2 Variants' (https://www.who.int/activities/tracking-SARS-CoV-2-variants) (accessed 15 January 2023)

EMHP (2022), 'Wastewater monitoring of SARS-CoV-2 in England' (https://www.gov.uk/government/publications/monitoring-of-sars-cov-2-rna-in-england-wastewater-monthly-statistics-15-july-2020-to-30-march-2022/emhp-wastewater-monitoring-of-sars-cov-2-in-england-15-july-2020-to-30-march-2022) (accessed May 2022)

COVIDPoops19 Dashboard (2023), 'Summary of Global SARS-CoV-2 Wastewater Monitoring Efforts' (https://www.covid19wbec.org/covidpoops19) (accessed September 2023)

Stanford Coronavirus Antiviral & Resistance Database (2023) (https://covdb.stanford.edu/variants/omicron_xbb/) (accessed September 2023)

Department of Water and Sanitation, South Africa (https://www.dws.gov.za/) (accessed May 2022)

Department of Statistics, South Africa (https://census.statssa.gov.za/#/province/9/2) (accessed July 2023)

SAMRC (2021), 'COVID-19 and wastewater early warning system team' (www.samrc.ac.za/wbe) (accessed 13 July 2021)

NGS-SA (2022), 'SARS-CoV-2 Genomic Surveillance Update' (https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance-update/) (accessed August 2022)

NICD (2021), 'Proposed Definition of a Wave in South Africa' (https://www.nicd.ac.za/wp-content/uploads/2021/11/Proposed-definition-of-COVID-19-wave-in-South-Africa.pdf) (accessed December 2021)

COVID-19 Online Resource and News Portal (2023), 'Latest on Vaccine Statistics' (https://sacoronavirus.co.za/latest-vaccine-statistics/) (accessed May 2023)

RVIM Netherlands (2021), 'Omicron Variant found in two previous test samples' (https://www.rivm.nl/en/news/omicron-variant-found-in-two-previous-test-samples) (accessed February 2022).

WHO (2022), 'TAG-VE Statement on Omicron Sublineages BQ.1 and XBB' (https://www.who.int/news/item/27-10-2022-tag-ve-statement-on-omicron-sublineages-bq.1-and-xbb) (Assessed March 2023)

CHAPTER FOUR

Molecular Epidemiology of SARS-CoV-2 and Description of Other Respiratory Viruses in Wastewater in Northern South Africa

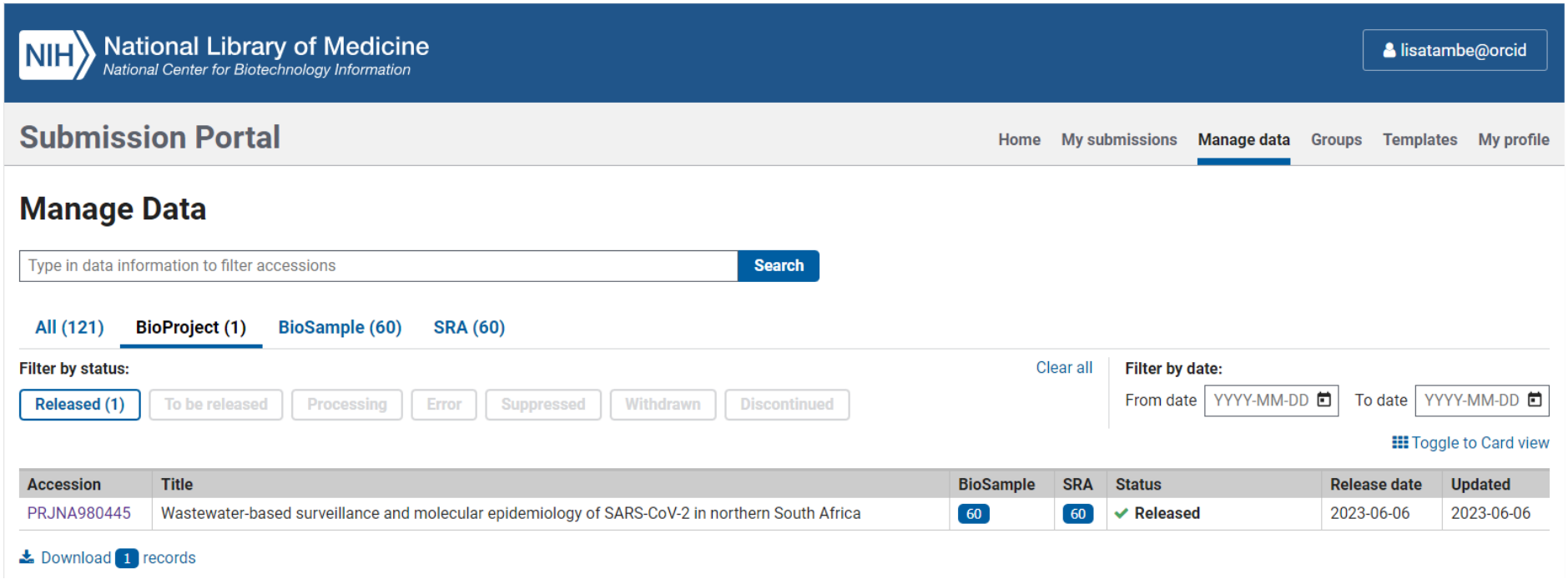
SCIENTIFIC OUTPUTS FROM THIS CHAPTER

1) Published article: Frontiers in Public Health 2023 11:1309869. https://doi.org/10.3389/fpubh.2023.1309869

2) Whole Genome Sequences have been submitted to NCBI SARS-CoV-2 SRA database (https://www.ncbi.nlm.nih.gov/sra/PRJNA980445)

Contribution to Chapter 4: The first author was responsible for sample receival and processing, qRT-PCR experiments and analysis, preparation of samples for sequencing and sequence analysis. She also wrote the original draft manuscript which was extracted from the main thesis.

NCBI SARS-CoV-2 SRA database Sequence Submission



ABSTRACT

Wastewater-based genomic surveillance of SARS-CoV-2 provides a comprehensive approach to characterize evolutionary patterns and distribution of viral types in a population. This study documents the molecular epidemiology of SARS-CoV-2 and describes other respiratory viruses present in Mopani and Vhembe districts in the Limpopo Province of South Africa, from January 2021 to June 2022. Out of 487 samples collected weekly from eight wastewater treatment sites, about 75% (365/487) were positive for SARS-CoV-2 RNA by RT-qPCR. Of these, 80 meet the criteria for allele-specific genotyping (ASG) and were genotyped. Sixty whole genome deep sequences (WGS) were obtained and analyzed for variants of concern (VOC) using the Pangolin and Nextclade tools. Relatedness of identified sequences was determined by phylogenetic analysis. VOCs were analyzed for prevalence, genetic characteristics and genetic diversity. Concordance for VOC between ASG and WGS analyses was determined.

Sixty SARS-CoV-2 full genomes were analyzed. Delta and Omicron variants were detected as early as January and February 2021 respectively, while the Beta variant was detected from July 2021. Delta variant was most predominant (45%), followed by the Omicron variant (32%), and Beta VOC having the least prevalence (5%). Eighteen percent (11/60) sequences were assigned lineages and clades only, but not a specific VOC name. Phylogenetic analysis was used to investigate the relationship of these sequences to other study sequences and further characterize them. Mutations in the receptor-binding domain (RBD) of the Spike protein (S-protein) were investigated, and some peculiarities were observed. For instance, in all Beta variant study sequences, mutation E484K was absent. Three previously undescribed mutations were seen in Delta variant study sequences (A361S, V327I, D427Y). A total of 264 mutations were detected in viral genomes of the study sequences. Of these, 72.3% occurred in ORF1ab and Spike protein. Mutations P314L (ORF1b) and D614G (Spike protein) were the most dominant, occurring at frequencies of 100% and 85%, respectively. Except for the absence of some key variant defining mutations in the Spike protein, there was little genetic diversity between the study sequences and reference sequences. Concordance in variant assignment between ASG and WGS was seen only in half (51.2%) of the study samples. Eleven respiratory viruses were identified in circulation, with human Adenoviruses (HAdVs) being the most dominant (86.5%), followed by endemic Human Coronaviruses (51.4%) and

Influenza A and B viruses (IAV & IBV – 32.4%). Of the four endemic human coronaviruses (HCoVs) detected, HCoV-NL63 was most dominant (54.5%), followed by HCoV-OC43 (22.7%). Investigation of the seasonality of these viruses revealed year-round circulation of HAdVs, while HCoVs, influenza viruses and human parainfluenza viruses (HPIVs) were mostly detected in winter. Influenza virus types (IAV and IBV) detected in the study site in 2021 were markedly different from those reported to be in circulation nationwide, as documented by the NICD. Specifically, IAV (H5N1)/Guandong, a highly pathogenic influenza virus, was detected, although at a low frequency.

The presence of Delta and Omicron VOCs prior to other reports in South Africa highlights the importance of population-based approaches in genomic surveillance over approaches that rely on individual samples. Additionally, including non-Spike protein targets could improve the specificity of allele-specific genotyping since all VOCs share similar S-protein mutations. Finally, continuous molecular epidemiology is necessary for documentation mutations whose implications when further investigated could improve treatment, vaccine and antibody development efforts.

Keywords: SARS-CoV-2, WBE genomic surveillance, mutational profile, viral evolution, RBD mutation analysis, respiratory viruses

4.1) INTRODUCTION AND STUDY RATIONALE

Respiratory infections are among the top causes of human death globally (WHO Facts Sheet, 2021). Severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) is an acute respiratory infection (ARI) that has ravaged the world, causing over 673 million infections, with a mortality rate of 6.85 million in the last two years (WHO, 2023). Throughout the Coronavirus disease 2019 (COVID-19) pandemic, genetic characterization has been instrumental in providing information about viral genome organization, mutational profiles, as well as development of drug targets for treatment and vaccines to decrease mortality in those infected with SARS-CoV-2. This was mainly achieved through whole genome sequencing (WGS) of brocheoalveolar lavage (BAL) specimen of patients positive for COVID-19. WGS reveals critical epidemiological facts such as tracing the time of epidemic outbreak, survival time, transmission pathways, and evaluation of possible causes and animal reservoirs (Ahmad et al., 2022). As tons of SARS-CoV-2 sequences were generated, a dynamic nomenclature system, known as the Pango nomenclature was proposed and developed by Rambaut and colleagues to name and track global transmission lineages. In addition to the NextStrain and GISAID nomenclature systems, which are based on classifying SARS-CoV-2 phylogenetic clades, the Pango nomenclature has been instrumental in tracking the evolution of the virus (Rambaut et al., 2020; Toole et al., 2021).

Over the course of 3 years of the pandemic, SARS-CoV-2 has evolved rapidly, due to its high mutation rate, estimated to be between 10^{-5} and 10^{-3} (Abavisani et al., 2022) that significantly impacts viral protein structures, function, and immunogenic characteristics (Grubaugh et al., 2020; Lauring, 2021). These characteristics are strongly associated with the immunological response and clinical outcome in humans. The Spike protein (S-protein) of the virus functions mainly in binding to human cellular entry receptors (angiotensin-converting enzyme 2 – ACE2), which allows infection (V'kovski et al., 2021). Since the beginning of the COVID-19 pandemic, mutations detected in the S-protein have been used to characterize variants of concern (VOCs) and variants of interest (VOI) that arose over time. Both VOCs and VOIs are classified based on their potential impact, with VOCs regarded as posing the highest risk on the population. The WHO has classified five VOCs, which include: Alpha, Beta, Gamma, Delta, and Omicron (WHO, 2023).

By May 2020, the D614G mutation was widely reported to have overtaken the original Wuhan strain and was observed in over 78% of clinical samples worldwide (Korber et al., 2020). As the pandemic progressed, specific key mutations developed in the S-protein of the virus, which led to increased infectivity and transmissibility. Mutations N501Y, DelH69V70, and P681H developed next, and were then classified as the Alpha variant (B.1.1.7), first detected in the UK in September 2020 (Hill et al., 2022; Meng et al., 2021). By December 2020, mutations N501Y, E484K, and K417N were reported, and classified as the Beta VOC (B.1.351). This variant was detected in South Africa (Tegally et al., 2021), and it became the most dominant variant detected in 80% of SARS-CoV-2 genomes in the country. A month later, the Gamma variant (P.1) was reported in Brazil (Faria et al., 2022), as well as travellers from Brazil, arriving in Japan (Ramundo et al., 2021). In May 2021, a more infectious SARS-CoV-2 strain with increased mortality (Dhar et al., 2021; Cherian et al., 2021) spread rapidly through India and was termed the Delta variant (B.1.617.2). By December 2021, the Omicron variant was detected in South Africa (Viana et al., 2021) and rapidly spread around the world. From December 2021 to September 2023, the Omicron variant and its sub-lineages (BA.1, BA.2, BA.3, BA.4, BA.5, XBB, EG.5), including BA.1/BA.2 recombinants (Gu et al., 2022; Shrestha et al., 2022; Tegally et al., 2022) are responsible for current COVID-19 cases worldwide (WHO, 2023). Variant-defining mutations of these VOCs have functional implications with clinical significance which affect treatment and vaccine therapies. Thus, continuous characterization of SARS-CoV-2 in different populations is necessary since such data can be added to genomic repositories and utilized to improve drug design and vaccine therapies.

The COVID-19 outbreak has not had a devastating effect on public health alone, but on social and economic human aspects, as well. For example, to curb SARS-CoV-2 transmission in the masses, public health measures recommended by the world health organization (WHO) implemented around the world included regular handwashing and sanitization, social isolation, use of face masks, closure of schools and child-care facilities, as well as working from home. These preventative and increased hygiene activities reduced the transmission of respiratory viruses, since most of them are spread by droplet transmission between individuals. However, as COVID-19 cases continuously decline, a resurgence in respiratory viruses which cause acute respiratory infections (ARIs) will be seen (Uppala et al., 2022). This resulted in invested efforts employed to understand the seasonal pattern of the respiratory viruses; research from wastewater and human specimen have been reported (Olsen et al.,

2021; Boehm et al., 2023; Di Maio et al., 2024). In addition, comparisons have also been made to determine any pattern similarity in circulation of the respiratory viruses observed pre-COVID-19 (Cho et al., 2024; Foley et al., 2024; Yoshioka et al., 2024). As such, documenting the occurrence and seasonal circulation of these respiratory viruses is important, and could inform public health personnel to offer better directives to policy makers to enact policies which will curb transmission in the population.

One major method implemented in SARS-CoV-2 genetic characterization for detection of new circulating variants has been through genomic surveillance, which has mainly been achieved through the WGS of individual patient clinical samples. However, the drawback of this type of genomic surveillance is that data is only obtained from patients who are tested in healthcare centers. Thus, SARS-CoV-2 genetic diversity in asymptomatic individuals, as well as those who do not seek attention in healthcare facilities, and some communities may be underestimated. Wastewater-based epidemiology (WBE) has proven to be an asset in the identification of COVID-19 hotspots and tracking the trends of infection in the community (Haramoto et al., 2020; Kumblathan et al., 2021; Castiglioni et al., 2022). Applying this population-based approach for SARS-CoV-2 genomic surveillance offers the added advantage of tracking geographical distribution and predicting VOC occurrence in the population. Alongside WGS, allele specific genotyping (ASG) has been utilized as a tool for routine monitoring of SARS-CoV-2 variants in the population (Harper et al., 2021; Takemae et al., 2022). Compared to whole genome sequencing, by next generation sequencing, allele-specific genotyping is less expensive and can be implemented on a larger scale in resource-limited settings. The aim of this objective was to describe the molecular epidemiology and genetic characteristics of SARS-CoV-2 and document the respiratory viruses occurring in the Vhembe and Mopani districts in Limpopo, South Africa.

4.1.1) Hypothesis

4.1.1.1) Single nucleotide variations present in SARS-CoV-2 variants of concern differ across the districts of the Limpopo Province and throughout South Africa.

4.1.1.2) Diverse respiratory viruses circulate seasonally in the Limpopo Province.

4.1.2) Research Questions

4.1.2.1) What is the molecular epidemiology of SARS-CoV-2 viruses in the Limpopo Province from 2021 – 2022?

4.1.2.2) Which respiratory viruses are commonly circulating at the study sites?

4.1.3) Study Aim

To utilize WBE to describe the molecular epidemiology and genetic characteristics of SARS-CoV-2, and to document the respiratory viruses circulating in the Vhembe and Mopani districts in Limpopo, South Africa.

4.1.4) Specific Objectives

a) To utilize WBE to describe the molecular epidemiology of SARS-CoV-2 in the Mopani and Vhembe districts in the Limpopo Province.

b) To compare the diversity between SARS-CoV-2 viruses in Limpopo with other regions of South Africa, and the world.

c) To document the respiratory viruses circulating seasonally in the study sites throughout the surveillance period.

4.2) STUDY DESIGN, MATERIALS AND METHODS

4.2.1) Sample Collection, Extraction and SARS-CoV-2 Quantification

Samples were collected from seven wastewater treatment plants (WWTPs) and one waste stabilization ponds (WSP) in the Vhembe and Mopani districts in Limpopo, South Africa (**section 3.2.2**; **Figure 14**). Selection of WWTPs and WSPs sampling sites are detailed in **section 3.2.2**. Sample collection, extraction and SARS-CoV-2 quantification have also been described previously. Briefly, influent wastewater grab samples (500mL) were collected at the raw inlet after the grid point from each of the sites once every week on a Monday over 17months (January 2021 to May 2022). Samples were transported to the laboratory at 4°C and processed and total RNA extracted as described in **section 3.2.2**. SARS-CoV-2 detection in wastewater samples was achieved by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) as previously described in **section 3.2.4**.

4.2.2) SARS-CoV-2 Whole Genome Sequencing

Extracted RNA samples from wastewater which were positive after RT-qPCR were subjected to whole genome sequencing (WGS). SARS-CoV-2 RNA libraries were produced using the ATOPlex (MGI-Tech) platform protocol as previously described by (Johnson et al., 2022). This is a multiplexed amplicon-based sequencing workflow that is tailor-made for detecting SARS-CoV-2 fragments of approximately 500bp. Total RNA was reverse-transcribed and subjected to target enrichment with the addition of barcodes for sample identification. The amplified products were purified using magnetic beads, and the quality of the libraries was evaluated using the Qubit dsDNA high-sensitivity assay kit (Life Technologies, USA). Quality-assured libraries generated were pooled, and single-stranded circularized DNA (ssDNA) was subsequently synthesized using a dual barcode circularization kit (Cat No: 1000020570). Construction of DNA nanoballs for the ssDNA library and the dual barcoded balance library were conducted in parallel, and mixed at a 3:1 ratio, prior to sequencing on the DNBSEQ-G400 instrument at the SAMRC Genomic Centre.

4.2.3) Quality Control Evaluation of Sequences

Fastq files obtained after sequencing were subjected to the NGS FastQC program to validate the quality of the sequences. Only samples which passed QC were subjected to downstream analysis for molecular epidemiology and genetic characterization.

4.2.4) Genome Assembly, Variant Determination and Lineage Assignment

Sequence data were analyzed using the Geneious v2023.0 software, following the procedure previously described by Matume et al., (2018). Sequence reads were trimmed, aligned and mapped to the SARS-CoV-2 Wuhan sequence (NC_045512.2_Wuhan-Hu-1) downloaded from the NCBI database. This yielded the consensus of the sequence, indicating regions of similarity, and changes from the original SARS-CoV-2 Wuhan sequence; this was repeated for all sequences.

Consensus sequences were subjected to the Nextclade database for SARS-CoV-2 variant calling, clade assignment and mutation determination for the viral genes. In addition, the Phylogenetic Assignment Named Global Outbreak (PANGO) interface is in-built within the Nextclade database, for lineage assignment. Consensus sequences were also subjected to the COVID-19 Lineage Assigner PangoLIN tool for SARS-CoV-2 variant calling and lineage determination. SARS-CoV-2 variant-calling and lineage assignment obtained from both tools were compared to confirm the assignment given

4.2.5) Genetic Diversity of SARS-CoV-2 Viruses in the Study Sites Compared to those around the World

Full-length SARS-CoV-2 sequences from the Limpopo Province, South Africa, other African nations, and around the globe, and classified as Alpha, Beta, Delta and Omicron VOCs were downloaded from the Global Initiative on Sharing Avian Influenza Data (GISAID database). Sequences downloaded were selected from various nations around the world, and must have been collected between January 2021 – May 2022, to match the surveillance period, for accurate comparison. Only high-quality whole genome sequences were downloaded and used for analysis. These SARS-CoV-2 sequences (henceforth referred to as “reference sequences”), were imported to the Geneious v2023.0 software, and aligned with study sequences having similar VOC assignment, using the MAFFT v7.490 parameters. These reference sequences originated from the Limpopo province (n=4), South Africa (n=9), other

African nations (n=29), the Americas (n=2), Europe (n=15), Asia, and the Middle East (n=22). For instance, study sequences classified as Beta VOC were aligned with downloaded reference sequences also classified as Beta VOC. The aligned sequences were subjected to Nextclade software for lineage assignment and mutation determination. Then, the mutations present in the Beta variant in the reference sequences were compared to Beta variant study sequences to determine the genetic diversity between VOCs in the study sites, and other nations. This was done for all four VOCs detected in the study site.

4.2.6) Phylogenetic Analysis

All SARS-CoV-2 genomes from this study were aligned with reference sequences from across the globe as described in **Section 4.2.5** above. Using the Geneious tree builder, a neighbour joining phylogenetic tree was constructed with the Jukes-Cantor Markov model of evolution assumptions using 1000 bootstrap replicates. This model was selected because it assumes that the substitution of a base with any other base occurs with equal probability. All generated phylogenetic trees were rooted with the original SARS-CoV-2 Wuhan sequence (NC_045512.2_Wuhan-Hu-1).

4.2.7) Allele Specific Genotyping for SARS-CoV-2 Variant Determination

To determine VOC circulation, genotypic analysis through qRT-PCR was conducted for mutations conferring resistance in the S-gene of SARS-CoV-2. This study, SNP genotyping was done for some signatory mutations belonging to the Alpha, Beta, Delta, and Omicron VOCs. Only samples with SARS-CoV-2 concentration ≥1,500g.c./mL were included for analysis, using the 7 TaqMan SARS-CoV-2 Mutation Panels, from ThermoFisher Scientific (Applied Biosystems), with the same cycling conditions as previously in **section 3.2.7**.

4.2.8) Comparison Between Allele-Specific Variant Genotyping and WGS in VOC Determination

Samples which were subjected to allele-specific variant genotyping and WGS were compared to infer whether they yielded similar VOC assignment. A combination of mutations in the Spike gene coding for Alpha (N501Y, DelH69V70, P681H), Beta (N501Y, E484K, K417N), Delta (L452R, P681R) and Omicron (N501Y, DelH69V70, P681H, K417N) VOCs, were used for VOC determination (described in **section 3.2.7**, **Table 6**). For samples subjected to WGS,

VOC assignment was determined by Pangolin. To determine whether samples subjected to both techniques had the same variant call, the presence of key mutations in the S-gene (using the allele-specific genotyping criteria) were investigated for both techniques.

4.2.9) Respiratory Virus Sequencing and Sequence Analysis

Determination of respiratory viruses present was done using the Illumina RNA Prep with Enrichment combined with the Illumina Respiratory Virus Oligo Panel v2 (RVOP v2) kit. Illumina RNA Prep with Enrichment used on-bead tagmentation followed by a single hybridization step to generate enriched libraries. Total RNA was transcribed to cDNA and tagmented with reagents from the kit. After amplification, the libraries were enriched as a single reaction using the Illumina Respiratory Pathogen panel. The enriched libraries were then denatured and diluted to a final loading concentration of 2pM, according to the MiniSeq System Denature and Dilute Libraries guide and sequenced on the MiniSeq System to yield approximately 1M of 2 × 75 bp paired end reads using the MiniSeq High Output Reagent Kit. Data analysis was done by uploading the generated FASTQ data files to the DRAGEN RNA Pathogen Detection pipeline, which enables streamlined detection of viral pathogens, including SARS-CoV-2, using coverage and the k-mer-based approach.

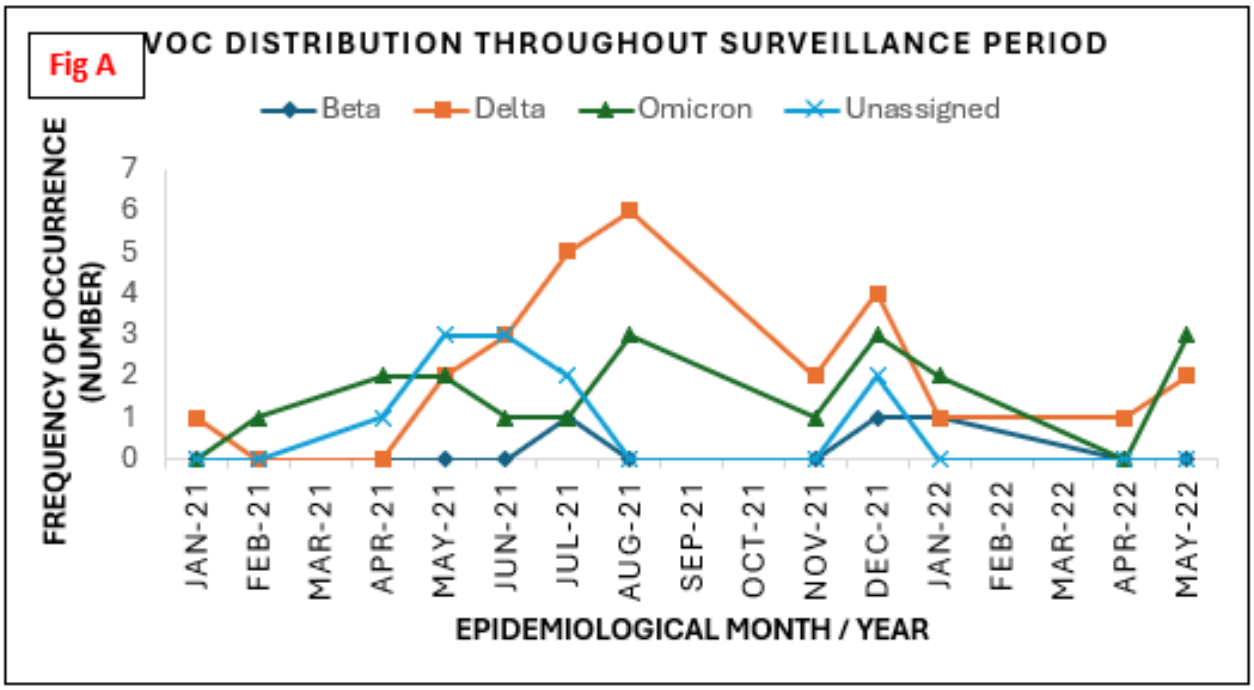
4.3) RESULTS

4.3.1) Molecular Epidemiology of SARS-CoV-2 in the Vhembe and Mopani Districts (January 2021 – June 2022)

Out of 487 samples collected weekly from eight wastewater treatment sites, 365/487 (75%) were positive for SARS-CoV-2 RNA by qRT-PCR. Of these, 80 met the criteria for allele-specific genotyping (ASG) and were genotyped. Seventy-five of 365 SARS-CoV-2 positive samples detected throughout the surveillance period (January 2021 – May 2022) were used for WGS. A total of 60/75 (80%) sequences passed QC and were successfully analyzed using the Nextclade software. Unassembled reads of study sequences (n=57) are submitted to the NCBI SARS-CoV-2 SRA database, under project number: PRJNA980445 (https://www.ncbi.nlm.nih.gov/sra/PRJNA980445), while assembled reads of study sequences (n=3) are submitted to NCBI GenBank database (https://submit.ncbi.nlm.nih.gov/subs/?search=SUB13889355).

Throughout the surveillance period, the Delta variant was most dominant (45%) across the study sites, closely followed by Omicron (31.7%). The Beta VOC occurred at low frequencies (5%), while the Alpha VOC was not detected in the study sites. Both tools (PangoLIN and Nextclade) did not assign a specific VOC name for 18% (11/60) of the study sequences, but assigned the lineage and clade for these sequences, and thus were designated as “unassigned,” for the purpose of classification in this study.

In the study sites, the Beta VOC was only sparsely observed between July and December 2021, as well as in January 2022. Interestingly, during this phasing out of the second wave, Delta and Omicron VOCs were detected. This was observed in January and February 2021 for the Delta and Omicron VOCs, respectively. As surveillance continued, the Delta VOC was continuously in circulation in the study sites and was dominant between April – August 2021. Omicron VOC was also in continuous circulation in the study sites but only became more prominent between December 2021 and January 2022. **Figure 44** illustrates the distribution of the VOCs observed throughout the surveillance period and the overall occurrence of the detected variants.



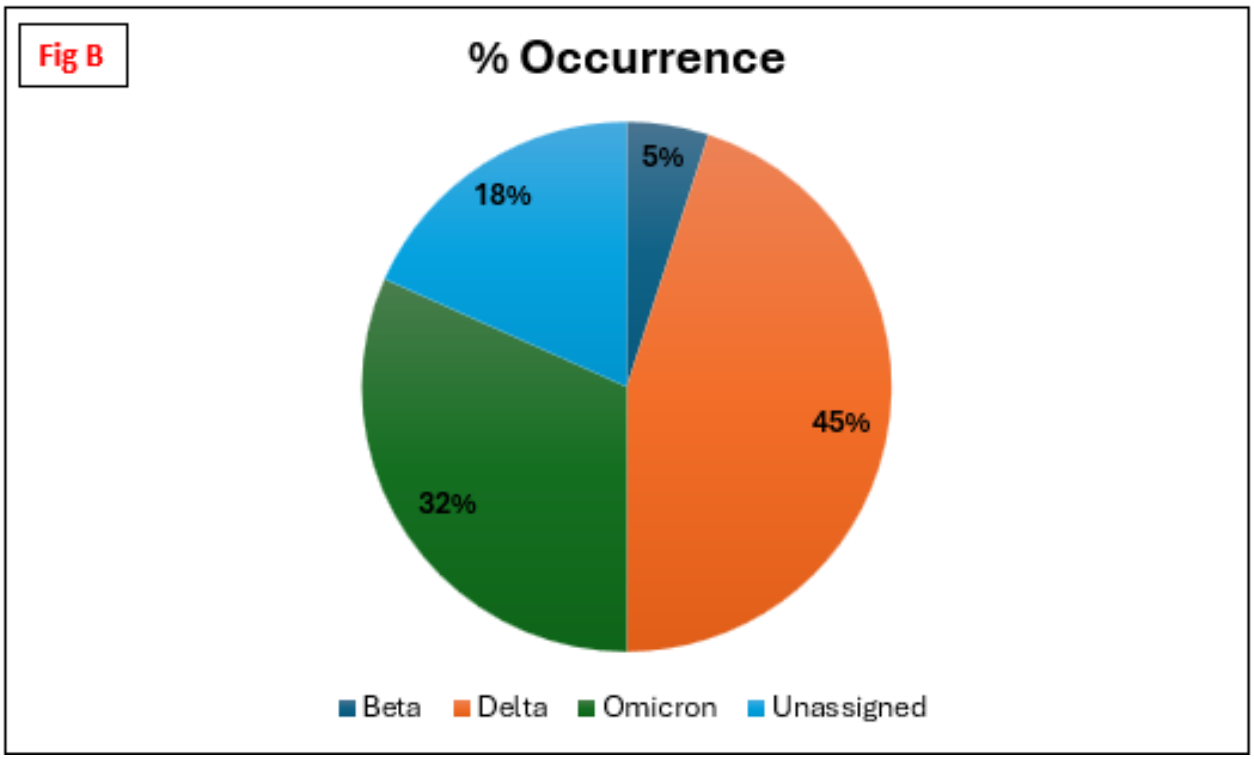


Figure 44: Distribution and overall frequency of SARS-CoV-2 VOCs in the Vhembe and Mopani districts. (**A**) shows the distribution of the variants between January 2021 – May 2022. (**B**) illustrates the general frequency of variant occurrence.

4.3.2) Genetic Characteristics of SARS-CoV-2 in the Study Sites

Near full-length SARS-CoV-2 genomes ranging between 29,842 – 29,903 kilobases (kb) were obtained for 60 viruses. The identified Beta, Delta, Omicron, and unassigned variants belonged to 12 lineages and 11 clades. The lineages detected include: B.1, B.1.1, B.1.1.174, B.1.351, B.1.617, B.1.617.2, B.1.1.529, AY.45, BA.1, BA.2, BA.4, BE.1 (alias BA.5.3.1.1). Lineages associated with the Delta variant (AY.45 and B.1.617.2) were most dominant (30% and 15%, respectively) throughout the surveillance period. Lineage AY.45 was first detected in January 2021 and became dominant in August 2021.

Lineages associated with the Omicron variant (B.1.1.529, BA.1, BA.2, BA.4, and BE.1) were observed at a 30% frequency. Of the 19 Omicron sequences, 9/19 (47.4%) belonged to the BA.1 lineage, 2/19 (10.5%) were of the BA.2 lineage, 3/19 (15.7%) belonged to the BA.3 (also known as B.1.529) lineage, 4/19 (21%) were from the BA.4 lineage, and 1/19 (5.3%) belonged lineage BA.5. The lineage associated with the Beta VOC (B.1.351) was first detected in July 2021, occurring at a 5% frequency throughout the surveillance period. The remaining lineages (B.1, B.1.1, and B.1.1.174) were associated with the unassigned variants.

Clades 20A, 20B, 20H, 21A, 21I, 21J, 21K, 21L, 21M, 22A, and 22B were detected in the study sites throughout the surveillance period. Clade 21A, 21I, and 21J are associated with the Delta variant; clade 21J was the first clade detected in January 2021 and was the most dominant (40%) in the study sites throughout the surveillance period. Clade 20H, first detected in July 2021, is associated with the Beta VOC and was sparsely detected during the period of surveillance. Four clades (21K, 21L, 21M, 22A, and 22B) associated with the Omicron VOC were observed at a frequency of 31.7%, with clade 21K being the most dominant. The remaining clades (20A and 20B) were associated with the unassigned variants. **Figures 45 and 46** illustrate the distribution and frequency of the lineages and clades detected.

Phylogenetic analysis was applied to determine the closest relationship of the 11 unassigned study sequences; 2/11 (18.1%) were seen clustering with Delta variant study and reference sequences. Whereas the remaining 9/11 (81.8%) unassigned sequences clustered with reference sequences of the Alpha variant (**Figure 47**).

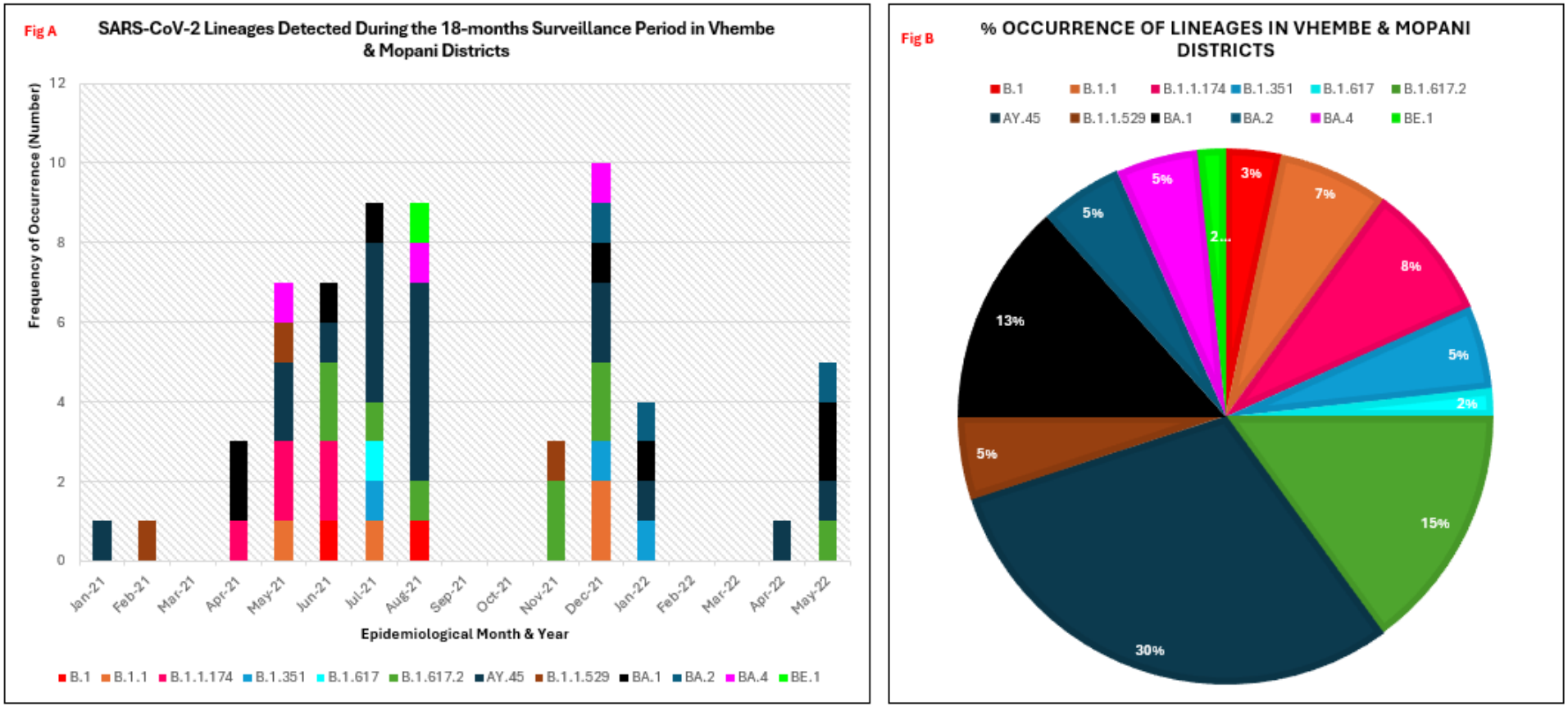
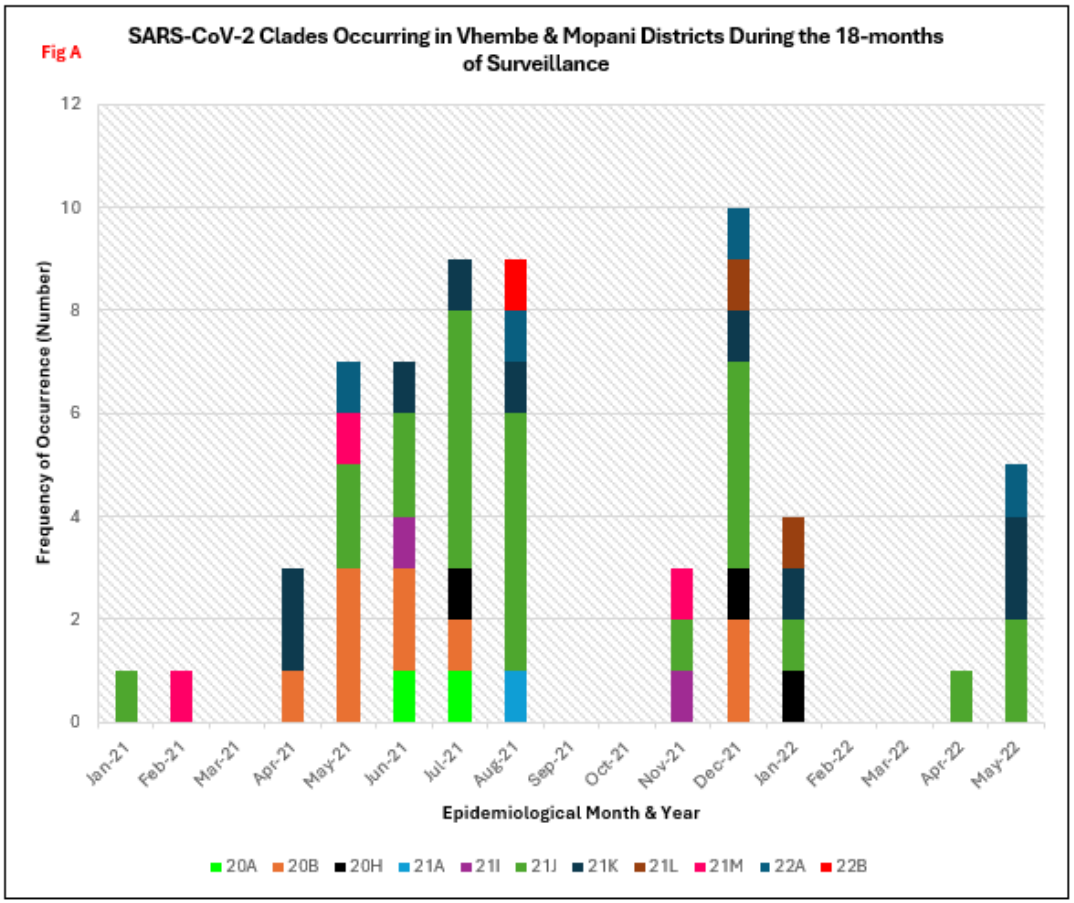


Figure 45: Distribution and percentage occurrence of SARS-CoV-2 lineages detected in the study sites.

Fig A illustrates the diversity of lineages detected at different time points of assessment. **Fig B** highlights the overall percentage occurrence of each of the 12 lineages detected throughout the surveillance period. (**NB:** Sequences were not available for Mar-21, Sep-21, Oct-21, Feb-22, Mar-22).



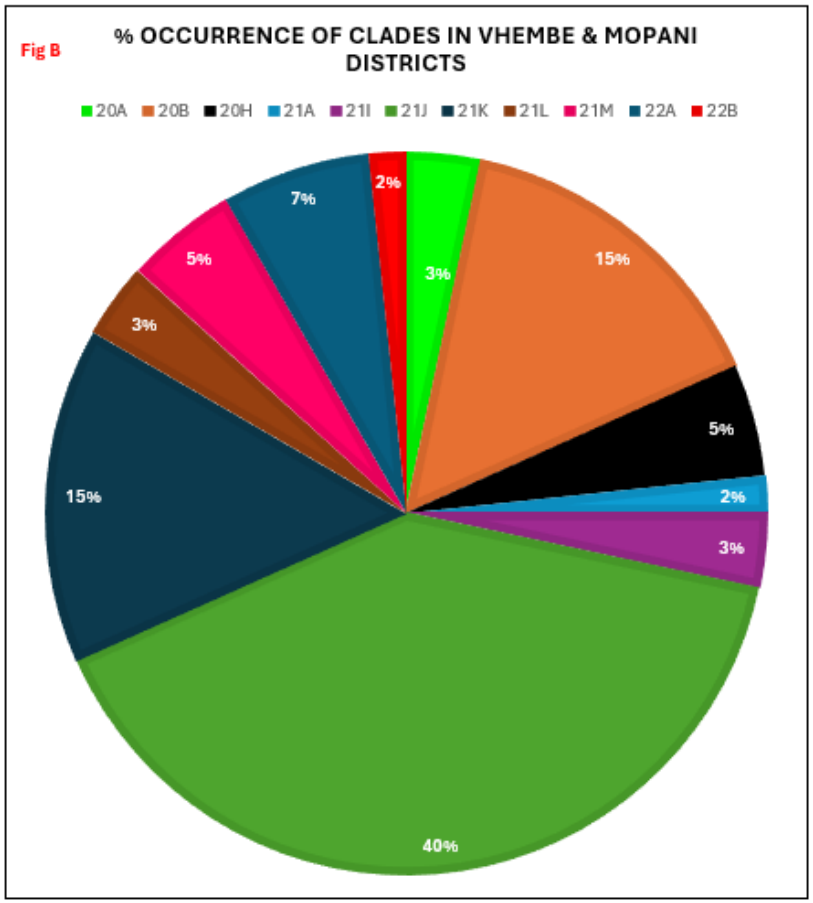


Figure 46: Distribution and percentage occurrence of SARS-CoV-2 clades detected in the study sites.

Distribution and percentage occurrence of SARS-CoV-2 lineages detected in the study sites. **Fig A** illustrates the diversity of lineages detected at different time points of assessment. **Fig B** highlights the overall percentage occurrence of each of the 12 lineages detected throughout the surveillance period. (**NB:** Sequences were not available for Mar-21, Sep-21, Oct-21, Feb-22, Mar-22).

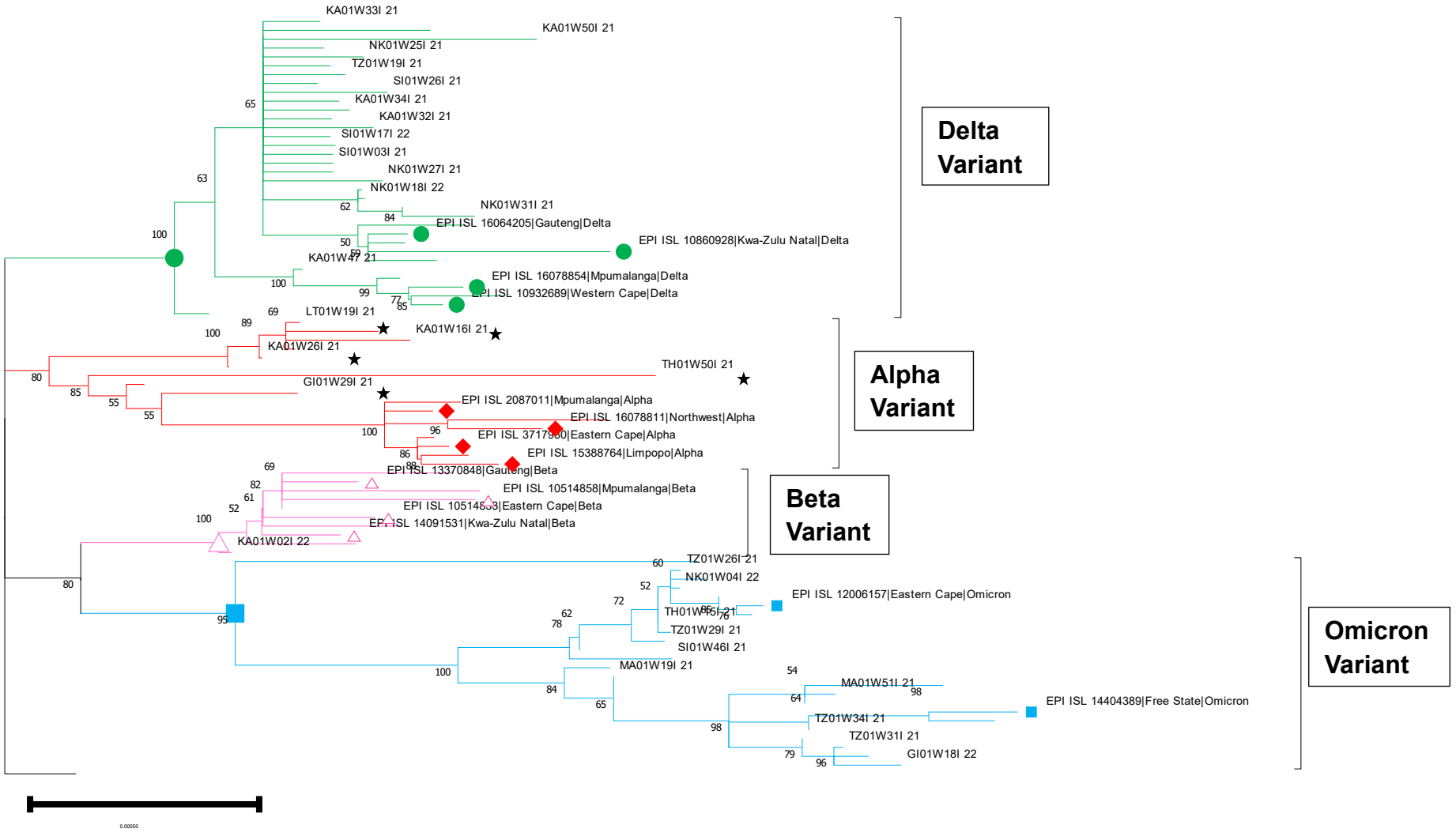


Figure 47: Phylogenetic relationship between study sequences and reference sequences (indicated with colored shapes) from South Africa with 1000 bootstrap iterations.

The green branches highlight all lineages (B.1.617.2, AY.39, AY.45) and clades (21A, 21I, 21J) associated with the Delta variant. The blue branches highlight all lineages (B.1.1.529, BA.1, BA.2, BA.4, BE.1) and clades (21K, 21L, 21M, 22A, 22B) associated with the Omicron variant. The purple branches highlight lineage B.1.351 and clade 20H associated with the Beta variant. The red branches show lineage B.1.1.7 and clade 20I associated with the Alpha variant. Sequences with a black star are those assigned a lineage (B.1, B.1.1, B.1.617, and B.1.1.74) and clade (20A and 20B) by the Nextclade and PangoLIN tools, but without a specific variant name. Phylogenetic analysis shows some of these sequences clustering with the Alpha and Delta variants. However, other “unassigned” sequences still clustered with each other.

4.3.3) Mutations Detected in the Study Sequences

Of the ten coding genes of SARS-CoV-2, mutations were detected in all but the ORF10 gene in the viral genome. A total of 264 missense mutations were detected across the genome, with 133/264 (50.3%) of them occurring above a 3% frequency level. Four nucleotide substitutions were also detected in the 5’ and 3’ UTR non-coding regions (CDS) of the genome, all of which occurred above a frequency of 3%. In the 5’ UTR, these mutations were: C241T (90%) and G210T (43.3%). Nucleotide substitutions G29742T and G29878A were detected at 48.3% and 10% respectively, in the 3’ UTR region. ORF1ab and the Spike protein accounted for the highest number of mutations detected.

ORF1ab protein harboured 100/264 (38%) mutations; only those occurring above a frequency of 3% are reported here; all other mutations and deletions are represented in Appendix 2. In ORF1a, the most common mutations were T3255I (60%), A1306S (40%), P2046L (35%), P3395H (31.6%), V2930L (28.3%) and T3646A (25%). Twelve deletions were detected in the ORF1a gene, with mutations S2083- and L3674- being the most frequently detected at 13%. In the ORF1b gene, mutation P314L was present in all 60 sequences, thus it had the highest frequency. Other frequently occurring mutations in the ORF1b gene were G662S (45%), M115I (35%), I1566V (33.3%), and P1000L (23.3%).

The Spike protein (S-protein) had 91/264 (34.4%) mutations; 68/91 of them were missense mutations, 22/91 were deletions, and 1/91 was an insertion. Of the 68 missense mutations, D614G was the most frequently detected (85%). Deletions E156- and F157- were detected at a 46.6% and 45% frequency, respectively; this was followed by deletion H69- and V70-, occurring in 25% of the sequences. Insertion 214:EPE occurred at a frequency of 13.3%.

Section 4.3.3.1 below describes mutations in the S-protein Receptor-Binding Domain (RBD). All other observed mutations that occurred at a frequency of ≤3% are represented in **Appendix 2**.

4.3.3.1) Mutations in Study Sequence S-protein Receptor-Binding Domain (RBD)

A total of 12 mutations were detected in the RBD of the Beta variant study sequences, with two of them (K417N and N501Y) occurring at a higher frequency. Among the Delta variant study sequences, two previously described RBD mutations (L452R and T478K) occurred at a higher frequency compared to the three novel mutations (A361S, V327I, D427Y) also detected in some sequences. Within the RBD of Omicron study sequences, 18 common mutations were detected. However, D405N and R408S mutations, which are commonly detected in lineages BA.2, BA.4, and BA.5, were completely absent in the study sequences classified to be of BA.2, BA.4, and BA.5 lineages. Only 9/11 (81.8%) of the unassigned study sequences had a mutation the RBD, and mutation Q498H was the most prevalent. Details of the frequency of occurrence of mutations detected in the RBD are presented in **Table 12**.

Table 12: Frequency of occurrence of S-protein RBD mutations in the study sequences.

Beta S-protein RBD		**Delta S-protein RBD**		**Omicron S-protein RBD**		**Unassigned S-protein RBD**	
Mutation	**Frequency**	**Mutation**	**Frequency**	**Mutation**	**Frequency**	**Mutation**	**Frequency**
G339D	1/3 (33.3%)	A361S	2/27 (7.4%)	G339D	19/19 (100%)	A372T	1/11 (9%)
K417N	3/3 (100%)	V327I	1/27 (3.7%)	S371L	11/19 (57.8%)	K417T	2/11 (18.2%)
N440K	1/3 (33.3%)	D427Y	1/27 (3.7%)	S371F	5/19 (26.3%)	S477N	1/11 (9%)
K444-	1/3 (33.3%)	L452R	26/27 (96.3%)	S373P	18/19 (94.7%)	T478K	2/11 (18.2%)
V445-	1/3 (33.3%)	L452W	1/27 (3.7%)	S375F	17/19 (89.4%)	E484Q	1/11 (9%)
S477N	1/3 (33.3%)	T478K	27/27 (100%)	T376A	6/19 (31.5%)	Q493K	1/11 (9%)
T478K	1/3 (33.3%)	Q498H	1/27 (3.7%)	K417N	18/19 (94.7%)	Q498H	6/11 (54.5%)
E484A	1/3 (33.3%)			N440K	11/19 (57.8%)		
Q493R	1/3 (33.3%)			G446S	8/19 (42.1%)		

Q498K	1/3 (33.3%)			L452R	5/19 (26.3%)		
P499S	1/3 (33.3%)			S477N	8/19 (42.1%)		
N501Y	3/3 (100%)			T478K	19/19 (100%)		
				E484A	12/19 (63.2%)		
				F486V	5/19 (26.3%)		
				Q493R	13/19 (68.4%)		
				Q498R	18/19 (94.7%)		
				N501Y	18/19 (94.7%)		
				Y505H	18/19 (94.7%)		

4.3.3.2) Tropism Conferring Mutations in Study Sequences

Several mutations are known to confer viral tropism. However, some mutations have an increased mode of transmissibility (**Figure 48**). Mutation D614G is reported to enhance transmission of the virus and increase virus infectivity due to low S1 domain shedding accompanied by high S-protein integration into the virion (Zhang et al., 2020). The N501Y mutation in the RBD plays a critical role in the transmission and virulence of SARS-CoV-2 (Liu et al., 2022), while K417N is associated with increase in viral infectivity (Li et al., 2021). Mutation L452R is reported to be associated with assisting viral escape (Kumar et al., 2023), while mutation T478K enhances the stability of other mutations in the Delta variant, leading to increased infectivity and transmissibility (Wang et al., 2021).

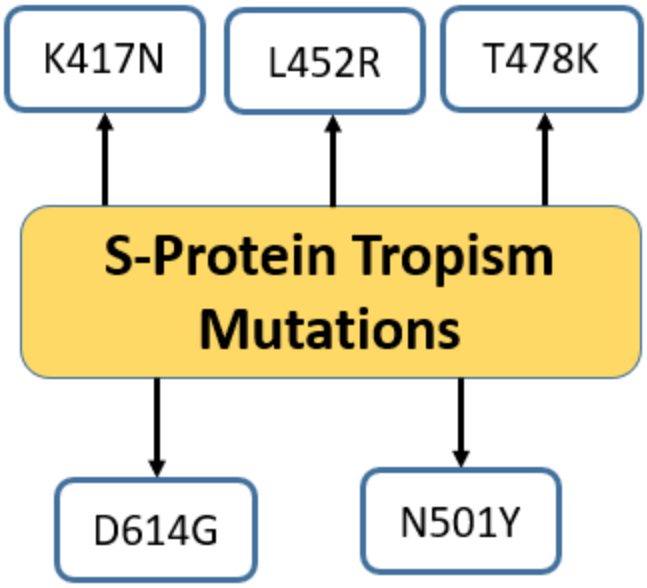


Figure 48: Illustrates five major mutations present in the study sequences which increase infectivity of SARS-CoV-2.

4.3.4) Genetic Diversity Among Study Sequences

Investigation of the genetic diversity among the study sequences belonging to the same variant showed little variability occurring within them. Among the Beta variant sequences, the intra-genetic variability ranged between 0.0003 and 0.0018. Similarly, little diversity was observed among the Delta (0.00 – 0.0012) and Omicron (0.00 – 0.0018) variant study sequences. Among the unassigned study sequences, however, a slightly higher variability (0.00 – 0.0022) was observed.

4.3.4.1) Genetic Diversity Within the S-protein RBD

Beta variant study sequences (3) were compared to Beta variant reference sequences obtained from GISAID. These reference sequences originated from the Limpopo province (n=4), South Africa (n=9), other African nations (n=29), the Americas (n=2), Europe (n=15), Asia, and the Middle East (n=22). Mutation E484K, is associated with reduced neutralizing activity of human polyclonal sera induced in convalescent and vaccinated individuals (Jangra et al., 2021). This mutation was absent in all Beta variant study sequences, although it was present in all reference sequences. The average evolutionary divergence between the sequences was estimated to be 0.0006, showing similarity between them.

Delta variant study sequences (27) were compared to downloaded Delta variant reference sequences (71) from GISAID. These previously published sequences originated from the Limpopo province (n=7), South Africa (n=12), other African nations (n=32), the Americas (n=4), Europe (n=12), Asia, and the Middle East (n=9). In 1/27 (3.7%) of Delta variant study sequences, the Tryptophan (W) amino acid (aa) was detected, in place of the Arginine (R) aa, as seen in position 452. Three previously undescribed novel mutations (V327I, A361S, and D427Y) were detected in the study sequences, but not the reference sequences. The evolutionary divergence between the study and reference sequences was estimated to be 0.0008, showing similarity between the sequences.

Omicron study sequences were compared to 54 Omicron reference sequences obtained from GISAID. The proportion of Omicron lineages downloaded was as follows: 7/54 (12.9%) were of BA.1 lineage, 25/54 (46.3%) for BA.2, 2/54 (3.7%) sequences were of BA.4 lineage and BA.5 occurred at 20/54 (37%). Of the 18 RBD mutations in the Omicron variant, only mutations D405N and R408S, belonging to the BA.2, BA.4, and BA.5 lineages, were completely absent in the study sequences. These mutations evade humoral immunity elicited by Omicron BA.1 infection. However, they were present at high frequencies in the reference sequences. Even with these differences, the average evolutionary divergence (0.0015) between the Omicron study and reference sequences was not much.

Unassigned study sequences which clustered with the Alpha variant (n=9) after phylogenetic analysis (**Figure 47**) were compared to Alpha variant reference sequences originating from the Limpopo province (n=4), South Africa (n=8), other African nations (n=40), the Americas (n=4), Europe (n=13), Asia, and the Middle East (n=13). Mutation N501Y was the only common mutation found in the RBD of the study and reference sequences. This mutation increases ACE2 binding affinity, causing the virus to become more infectious. This mutation was completely absent in the study sequences, but present in high frequency (>60%) in the other populations. The average evolutionary divergence (0.001) between the study and reference sequences was also small.

Table 13: Frequency of occurrence of key mutations defining the Alpha, Beta and Delta VOCs between different populations.

SARS-CoV-2 VOC	S-gene Defining Mutations	Vhembe & Mopani District (%)	Limpopo (%)	South Africa (%)	African Nations (%)	Americas (%)	Europe (%)	Asia & Middle East (%)
Beta Variant	K417N	100	100	100	93.1	100	100	100
	E484K	0	100	100	96.6	100	100	100
	N501Y	100	100	88.9	96.6	100	93.3	90.9
Delta Variant	L452R	96.2	100	100	100	100	100	100
	L452W	3.7	0	0	0	0	0	0
	T478K	96.2	100	100	100	100	100	100
	V327I	3.7	0	0	0	0	0	0
	A361S	7.4	0	0	0	0	0	0
	D427Y	3.7	0	0	0	0	0	0
Alpha Variant	N501Y	0	100	100	97.5	100	100	100

4.3.5) Allele-Specific Variant Genotyping versus WGS in VOC Determination

Of the 80 samples that met the criteria for allelic variant genotyping, 41/80 (51.3%) were subjected to whole genome sequencing. For 21/41 (51.2%) samples evaluated by both techniques, concordance was observed between the S-gene-defining mutations and variant assignment. For 13/41 (31.7%) samples, at least one S-gene defining mutation was observed in both techniques, but with a different variant assignment. Interestingly, there were 7/41 (17%) samples in which no concordance existed between mutations detected by ASG or variant assignment in both techniques (**Table 14**).

Table 14: S-gene mutations and VOC assignment by allele-specific genotyping compared to mutations and VOC assignment observed from WGS.

Sample Name	S-gene Mutations Detected from SNP Genotyping	VOC Assignment	Corresponding S-gene Mutations Observed from WGS	VOC Assignment (by Nextclade)	VOC Assignment based on mutations observed in both Techniques
KA01W21I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	N501Y, delH69V70	Clade 20B (designated "Unassigned" for the purpose of classification in this study)	Mutations present in WGS are not variant specific
KA01W22I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Omicron	N501Y, P681H, delH69V70, K417N, L452R, P681R	Omicron	Omicron
LT01W25I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	delH69V70	Clade 20A (designated "Unassigned" for the purpose of classification in this study)	Mutations present in WGS are not variant specific
TZ01W26I_21	N501Y, P681H, delH69V70, K417N, L452R, P681R	Omicron	N501Y, P681H, delH69V70, K417N	Omicron	Omicron
NK01W27I_21	N501Y, P681H, delH69V70, E484K, K417N	Beta, Omicron	No mutations detected by SNP Genotyping was observed	Delta	No correspondence between mutations detected
LT01W28I_21	P681H, delH69V70, E484K, K417N, P681R	Delta	P681R	Delta	Delta
NK01W28I_21	P681H, delH69V70, E484K, K417N, P681R	Beta, Delta	No mutations detected by SNP Genotyping was observed	Clade 20A (designated "Unassigned" for the purpose of classification in this study)	No correspondence between mutations detected
SI01W28I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
SI01W29I_21	E484K, K417N, L452R, P681R	Delta	P681R	Delta	Delta

GI01W29I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	No mutations detected by SNP Genotyping was observed	Clade 20B (designated "Unassigned" for the purpose of classification in this study)	No correspondence between mutations detected
KA01W29I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	N501Y, K417N	Beta	Mutations present in WGS are not variant specific
TZ01W29I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Omicron	N501Y, P681H, delH69V70, K417N	Omicron	Omicron
TH01W30I_21	P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
KA01W31I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
TH01W31I_21	P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
TZ01W31I_21	delH69V70, K417N, L452R, P681R	Delta	delH69V70, K417N, L452R, P681R	Delta	Delta
NK01W31I_21	N501Y, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
KA01W32I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta
SI01W32I_21	N501Y, delH69V70, E484K, K417N, L452R, P681R	Beta, Omicron	N501Y, K417N	Omicron	Mutations present in WGS are not variant specific
KA01W33I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	L452R	Delta	Mutations present in WGS are not variant specific
TZ01W34I_21	N501Y, delH69V70, E484K, K417N, L452R, P681R	Omicron	N501Y, delH69V70, K417N, L452R	Omicron	Omicron
KA01W34I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Delta	L452R, P681R	Delta	Delta

LT01W49I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	No mutations detected by SNP Genotyping was observed	Clade 20B (designated "Unassigned" for the purpose of classification in this study)	No correspondence between mutations detected
TH01W50I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R	Beta, Omicron	No mutations detected by SNP Genotyping was observed	Clade 20B (designated "Unassigned" for the purpose of classification in this study)	No correspondence between mutations detected
TZ01W50I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Beta, Delta, Omicron	N501Y, K417N	Beta	Mutations present in WGS are not variant specific
LT01W50I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	N501Y, P681H, delH69V70	Omicron	Omicron
KA01W50I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	No mutations detected by SNP Genotyping was observed	Delta	No correspondence between mutations detected
TH01W51I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	L452R	Delta	Mutations present in WGS are not variant specific
TZ01W51I_21	N501Y, P681H, delH69V70, E484K, K417N, L452R	Beta, Omicron	N501Y, P681H, delH69V70, K417N, L452R	Omicron	Omicron
MA01W51I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	No mutations detected by SNP Genotyping was observed	Omicron	No correspondence between mutations detected
NK01W51I_21	N501Y, P681H, delH69V70, K417N, L452R	Omicron	delH69V70, L452R	Delta	Mutations present in WGS are not variant specific
KA01W02I_22	N501Y, P681H, delH69V70, E484K, K417N, L452R	Beta, Omicron	N501Y, K417N	Omicron	Mutations present in WGS are not variant specific
NK01W03I_22	N501Y, P681H, delH69V70, E484K, K417N, L452R, P681R	Omicron	N501Y, P681H, K417N	Omicron	Omicron
NK01W04I_22	N501Y, P681H, delH69V70, K417N	Omicron	N501Y, P681H, K417N	Omicron	Omicron

SI01W05I_22	N501Y, P681H, delH69V70, E484K, K417N, L452R	Beta, Omicron	delH69V70	Delta	Mutations present in WGS are not variant specific
SI01W17I_22	N501Y, P681H, delH69V70, K417N, L452R	Omicron	L452R	Delta	Mutations present in WGS are not variant specific
GI01W18I_22	N501Y, P681H, delH69V70, K417N, L452R	Omicron	N501Y, P681H, delH69V70, K417N, L452R	Omicron	Omicron
TH01W18I_22	N501Y, P681H, delH69V70, K417N, L452R	Omicron	N501Y, P681H, delH69V70, K417N	Omicron	Omicron
TZ01W18I_22	N501Y, P681H, delH69V70, K417N, L452R	Omicron	L452R	Delta	Mutations present in WGS are not variant specific
NK01W18I_22	N501Y, P681H, delH69V70, K417N, L452R	Omicron	L452R	Delta	Mutations present in WGS are not variant specific
TZ01W19I_22	N501Y, P681H, delH69V70, E484K, K417N, L452R	Omicron	N501Y, P681H, delH69V70, K417N	Omicron	Omicron

4.3.6) Description of Respiratory Viruses in Vhembe and Mopani Districts

Forty wastewater samples collected between January and November 2021 from all eight study sites, and positive for SARS-CoV-2, were sequenced using MiniSeq. These samples were selected to represent the whole study period, to have an overview of the respiratory viruses circulating in the population. The obtained sequences were analysed using the Dragen software. The taxonomic content of the different samples varied substantially. The identified respiratory viral sequences belong to the families Adenoviridae (87% of wastewater samples), Coronaviridae (51%), Picornaviridae (43%), Orthomyxoviridae (32%), Paramyxoviridae (30%), Parvoviridae (14%), Pneumoviridae (11%), and Polyomaviridae (5%). **Figure 49** illustrates the frequency of occurrence of the detected viruses during surveillance. Out of 37 sequenced samples, 67.6% had multiple respiratory viruses, 21.6% had single respiratory viruses, and no respiratory virus was detected in 10.8% of the samples. Of the four endemic human coronaviruses (HCoVs) detected, HCoV-NL63 was most dominant (54.5%), followed by HCoV-OC43 (22.7%), and HCoV-HKU1(13.6%). HCoV-229E occurred sparsely (9.1%) throughout the surveillance period. **Table 15** gives a detailed elaboration of the respiratory virus types detected and their frequencies.

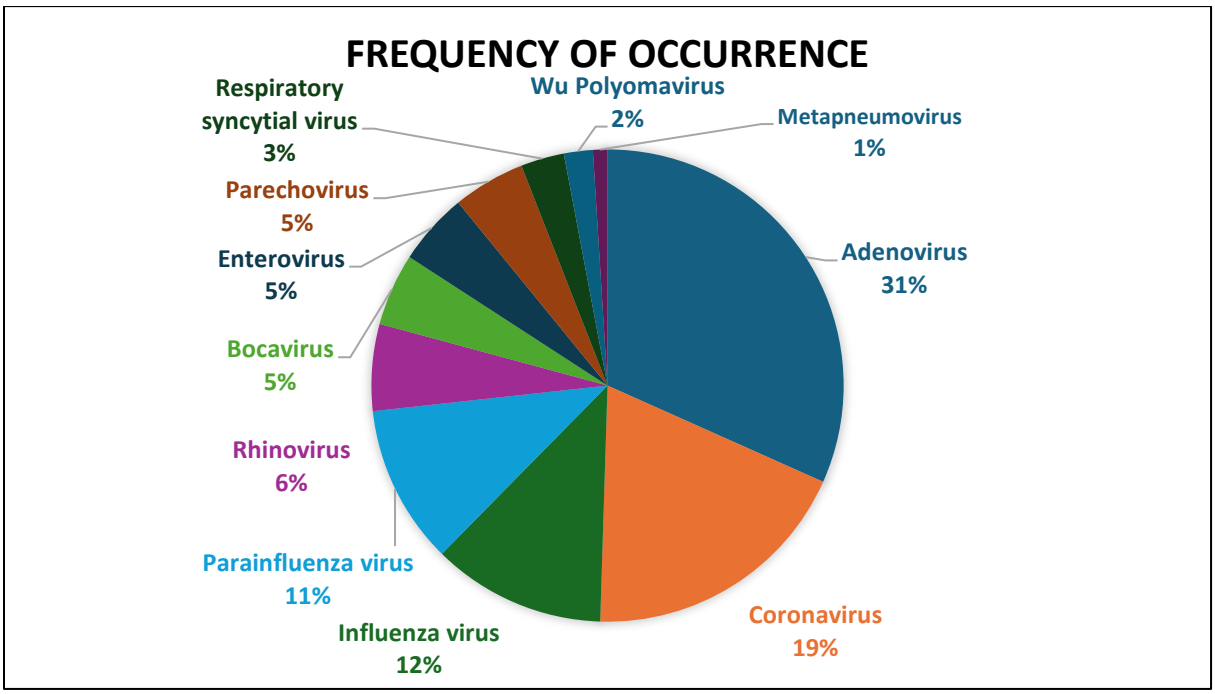


Figure 49: Frequency of occurrence of mammalian respiratory viruses detected in wastewater in Vhembe and Mopani districts.

Table 15: Detailed description of respiratory virus types occurring in the Vhembe and Mopani districts between January and November 2021.

Respiratory Virus Detected	Respiratory Virus Types	Frequency (%)
Human Parechovirus (HPeV)	Human Parechovirus type 1 PicoBank/HPev1/a	100
Human Enterovirus (HEV)	Human Enterovirus C109 isolate NICA08-4327	60
	Human Enterovirus C104 Strain: AK11	40
Human Bocavirus (hBoV)	Human Bocavirus 1 (Primate bocaparvovirus isolate st2)	28,6
	Human Bocavirus 3	42,9
	Human Bocavirus 2c PK isolate PK-5510	14,3
	Human Bocavirus 4 NI Strain HBoV4-NI-385	14,3
Human Rhinovirus (HRV)	Human Rhinovirus A89	33,3
	Human Rhinovirus C (strain 024)	33,3
	Human Rhinovirus B14	33,3
Human Parainfluenza virus (HPIV)	Human Parainfluenza Virus 3	81,8
	Human Parainfluenza virus 4a	9,1
	Human Parainfluenza virus 1	9,1
Influenza A Virus (IAV)	Influenza A Virus (A/goose/Guandong/1/1996(H5N1)	14,3
	Influenza A virus (A/Puerto Rico/8/1934 (H1N1))	42,9
	Influenza A virus (A/Michigan/45/1934 (H1N1))	14,3
	Influenza A virus (A/Texas/50/2012(H3N2))	14,3
	Influenza A virus (A/New York/392/2004 (H3N2))	14,3
Influenza B Virus (IBV)	Influenza B Virus (B/Lee/1940)	75
	Influenza B virus (B/Brisbane/60/20)	25
Human Adenovirus (HAdV)	Human Adenovirus E4	35,4
	Human Adenovirus C2	39,2
	Human Adenovirus B1	25,3
Respiratory syncytial virus (RSV)	Respiratory syncytial virus (type A)	100
Wu Polyomavirus	Wu Polyomavirus	100
Human Metapneumovirus (hMPV)	Human Metapneumovirus (CAN97-83)	100
Human Coronavirus (HCoV)	Human Coronavirus NL63	54,5
	Human Coronavirus OC43	22,7
	Human Coronavirus 229E	9,1
	Human Coronavirus HKU1	13,6

4.3.7) Seasonality of Detected Respiratory Viruses

Analysis was conducted to investigate the variability of respiratory viruses occurring per season. The forty sequenced samples spatially spanning the surveillance period (January – November 2021), were grouped per season as follows: Summer (2/40 – 5%), Autumn (11/40 – 27.5%), Winter (24/40 – 60%) and Spring (3/40 – 7.5%). Thus, more respiratory viral infections were observed during Winter and Autumn, compared to the other two seasons (**Figure 50**). Subsequent analyses were based on this categorization.

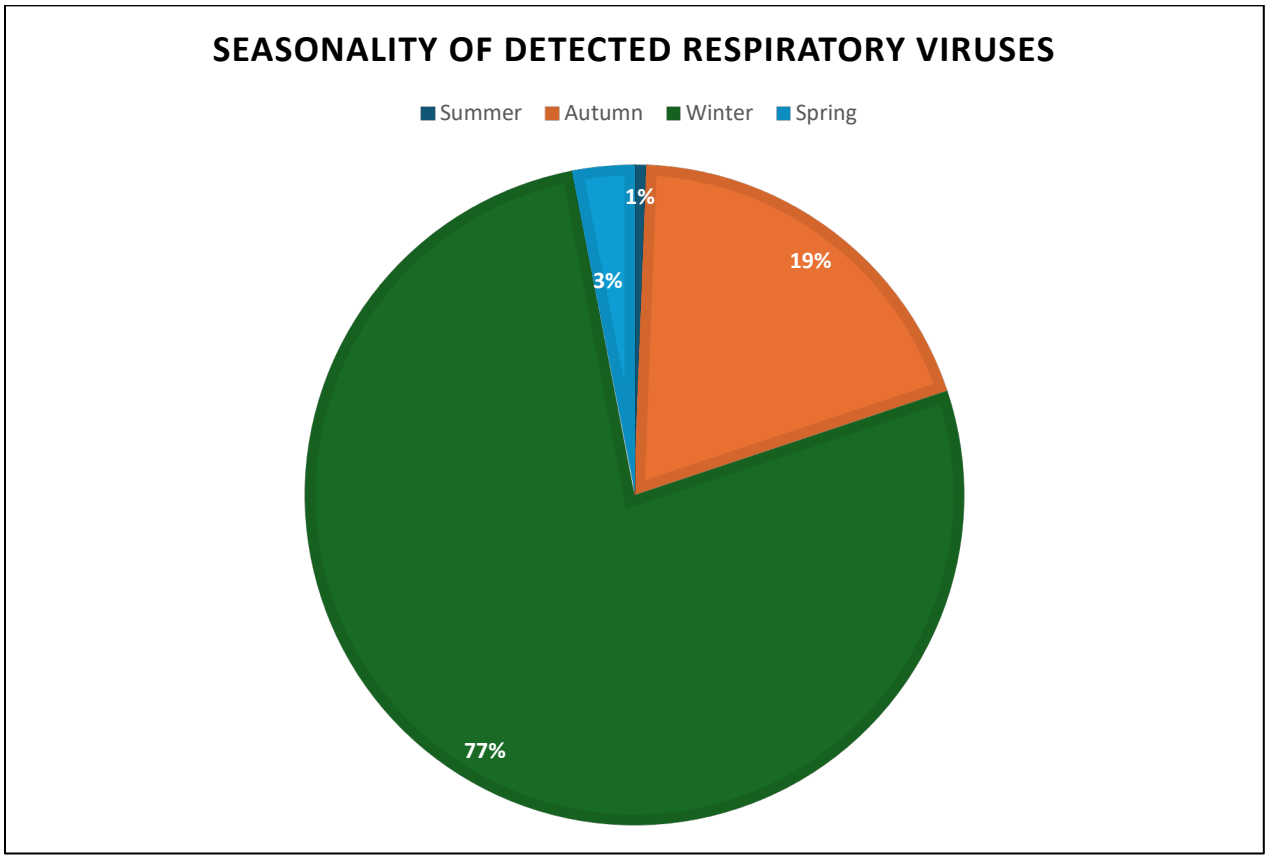


Figure 50: A representation of the seasonality of mammalian respiratory viruses detected in wastewater in Vhembe and Mopani districts. More viruses were circulating during the winter season compared to the other seasons.

Seasonally, human Adenoviruses, particularly type E4, was most dominant in all seasons. In **Figure 50**, the seasonal frequency of circulation of respiratory virus types detected in wastewater during the surveillance period is illustrated. To understand the seasonal occurrence pattern of respiratory viruses detected, analysis on seasonality focused on those with the highest prevalence (HAdV – 49.7%, Influenza viruses – 10.9%, HCoVs – 13.9%,

HPIV – 7.9%). These viruses are known to cause severe acute respiratory illness (SARI) in children, immunocompromised individuals, and the elderly (Murray et al., 2015).

Although HAdVs were detected in occurrence throughout the surveillance period, a higher prevalence was observed in Winter (76.8%) compared to Autumn (18.2%), with type E4 being the most prevalent species detected. More Influenza viruses were detected during winter (85.7%), than autumn (7.1%), with Influenza A (IAV) being more dominant than Influenza B (IBV). HCoVs were more predominant during winter (73.9%), compared to autumn (21.7%) and spring (4.3%); HCoV-NL63 (54.5%) was most dominant in circulation compared to the other three seasonal HCoVs. HPIVs were only detected in autumn and winter, with the highest prevalence observed during winter (84.6%). HPIV 3 was the most predominant species detected.

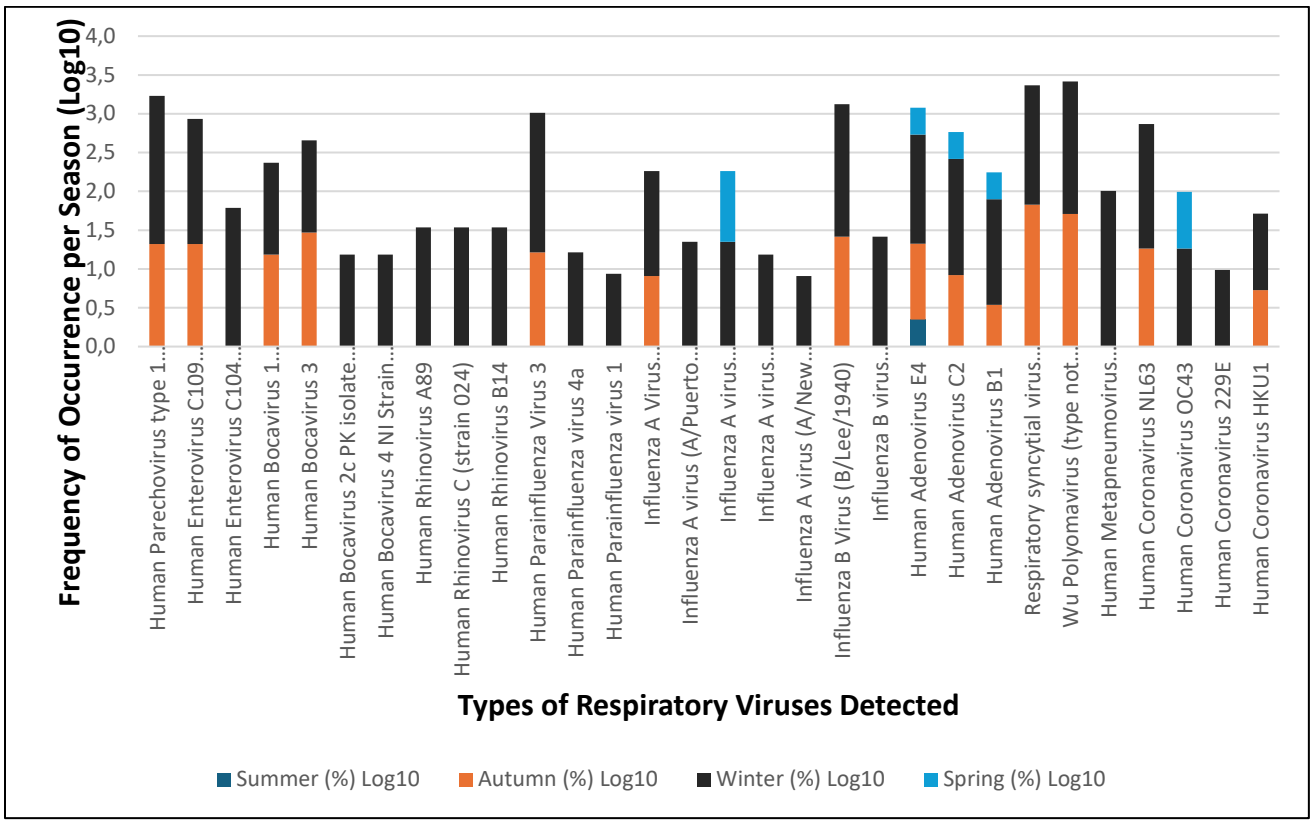


Figure 51: Seasonal frequency of respiratory virus types detected in wastewater surveillance, in the Vhembe and Mopani districts.

While more viruses were detected in the communities during the Winter and Autumn seasons, it is worth noting that viruses known to cause SARIs (HRV, HCoV and Influenza viruses) were mostly predominant during Winter compared to Autumn.

4.4) DISCUSSION

During the COVID-19 pandemic, genomic surveillance was vital in monitoring SARS-CoV-2 evolution, detecting new variants occurring in populations, which enabled rapid sharing of data across the globe. Through such surveillance, the Delta and Omicron VOCs were detected in India and South Africa, by March and December 2021, respectively. Using deep sequencing, the current study confirmed the presence of the Delta variant and the Omicron variant in the Vhembe district by January and February of 2021, respectively. Detection of both variants in the study sites predates reports from the South African National Institute of Communicable Diseases (NICD) which documented the detection of Delta and Omicron variants in the country a few months later, namely May and November 2021, respectively (NICD, 2021). Both variants were detected in the study sites towards the end of the second wave (January to February 2021) when the Beta variant was still predominant in South Africa.

Nine lineages and nine clades were identified at the study sites throughout the surveillance period. Lineage AY.45 or B.1.617.2 (Clade 21J) was the most dominant lineage and mostly predominated during the third wave (May – September 2021) of infections in South Africa, as reported by the South African National Institute for Communicable Disease (NICD) reported (NICD, 2021). The fourth wave in South Africa which began on 06 December 2021 saw the predominance of the Omicron VOC among the population, with lineage BA.1 being responsible for most infections in the population. Earlier reports of the BA.1 lineage occurrence in the population indicate that this lineage spread from the Gauteng province to other provinces in South Africa, and to two regions of Botswana from late October to November 2021 (Viana et al., 2022). Interestingly, this BA.1 lineage was detected in the study population as early as April 2021, and its dominance (47.4%) occurred throughout the surveillance period. Detection of such cryptic lineages in wastewater has also been reported in several locations in the USA (Gregory et al., 2022; Karthikeyan et al., 2022; Smyth et al., 2022; Vo et al., 2023; Shafer et al., 2024). Cryptic lineages are SARS-CoV-2 lineages detected in wastewater but not in clinical specimen (Gregory et al., 2022). In one study, through longitudinal wastewater genomic surveillance, using samples collected between September 2021 and February 2022, in San Diego, USA, the Alpha and Delta VOCs were in circulation 14-days prior their detection via clinical surveillance (Karthikeyan et al., 2022). Similarly, Gregory and colleagues (2022) report the circulation of multiple SARS-CoV-2 RBD

lineages in wastewater samples collected from three states (New York, Missouri, and California) in 2020, which had never been observed in patient samples (Gregory et al., 2022; Smyth et al., 2022).

Cryptic lineages are hypothesized to have occurred through two ways. First, it may be an animal reservoir introducing these viruses into wastewater, since SARS-CoV-2 has a wide host range including household pets, livestock and wildlife. Studies have found evidence of SARS-CoV-2 transmission between human and animal populations; thus, it is plausible that animal reservoirs of cryptic lineages exist undetected, with ongoing virus transmission and exchange between animal species. Alternatively, cryptic SARS-CoV-2 lineages in wastewater are also proposed to be derived from humans with unsampled infections or those with prolonged infections, where viral RNA is no longer detected in the nasopharynx. Such individuals could contribute viral RNA to wastewater even while testing negative via nasal swabs. People with immunocompromising conditions are at high risk for prolonged infections which could select for antigenic variation over the course of infection, driving diversification of SARS-CoV-2 within these hosts. Such selection could account for cryptic lineages, which tend to accumulate high levels of non-synonymous variation in the spike protein while otherwise maintaining the characteristic mutations from viruses that are no longer common in circulation. For instance, one study reported using 12s rRNA approach to resolve the origin of cryptic Omicron lineages detected in their study site and were able to trace its origin to a commercial building (Shafer et al., 2024). Similarly, resolution Omicron's origin in this study site based on current results, will require further analysis to establish certainty of whether these cryptic lineages are of human or animal origin. Furthermore, from the phylogenetic analysis (Figure 39), the study sequences obtained from wastewater surveillance can be seen clustering with the downloaded reference sequences of clinical origin. These results suggest that the obtained sequences are of Delta and Omicron origin. Unfortunately, samples were not collected from the WWTPs prior to the COVID-19 outbreak to determine definite circulation of Omicron for comparison.

Our findings are contrary to other wastewater-based surveillance studies conducted in Cape Town, South Africa, which reported the complete replacement of lineage BA.1 with lineage BA.2 by mid-January 2022 in 31 WWTPs (Johnson et al., 2022). The first appearance of lineage BA.4 likely occurred in mid-December 2021, with phylogeographic analysis indicating

probable dispersal from Limpopo province to Gauteng province, and subsequently to other provinces. Similarly, lineage BA.5 is reported to have emerged in early January 2022 and dispersed from the Gauteng province to other provinces in South Africa (Tegally et al., 2022). In our study, the earliest detection of lineage BA.4 was in May 2021, while lineage BA.5 was observed by August 2021. These observations highlight the advantage of using WBE as a surveillance approach for early detection of lineages that were already circulating in the population but only became dominant in individuals much later. In addition, the little intra-variant genetic diversity between the study sequences and previously published reference sequences further corroborates the silent circulation of these lineages prior to detection in individuals.

In terms of genetic diversity, investigating the Spike gene RBD of the study sequences was chosen because it is associated with the binding of the virus to the human host cell receptor (angiotensin-converting enzyme 2 – ACE2) since this is the target region of most SARS-CoV-2 vaccines. Some peculiarities were observed, such as the absence of mutation E484K in all Beta variant study sequences. Mutation E484K in the RBD of the Beta variant enhances viral binding affinity to human ACE2, as well as reduced antibody neutralizing effect in convalescent and vaccinated individuals (Jangra et al., 2021). This is relevant because the S-protein RBD facilitates SARS-CoV-2 infectivity, transmission, and antibody-mediated neutralization (Shang et al., 2020; Greaney et al., 2021; Harvey et al., 2021; Liu et al., 2022). Thus, the absence of this mutation in our Beta variant study sequences may explain why the Beta variant was sparsely detected (5%) in our study sites. Sparse detection of the Beta variant in this study may also reflect fewer samples sequenced. Due to limited funds, samples were selected to represent the entire surveillance period. Thus, from the 365 SARS-CoV-2 positive samples, only 75 were subjected for WGS. This was done to have a snapshot of variant distribution during the surveillance period. Furthermore, absence of this variant may have been due to the wastewater surveillance system being established just when the second wave in South Africa was phasing out (January – February 2021). Secondly, three novel mutations were detected in the RBD of the Delta variant. Investigating the implication of these mutations is needed to understand their role in viral infectivity and pathogenicity. Next, mutations L425R and T478K in the RBD of the Delta variant are associated with increased affinity with ACE2 (Sun et al., 2022). While these mutations occurred at high frequencies (96.3%) in Delta variant study sequences, a change in amino acid at position 452 (R→W;

L452W) was also observed, though at a lower frequency (3.7%). This new change, alongside the three previously undescribed mutations (V327I, A361S, and D427Y) require further investigation. Of the five mutations reported to have increased tropism effects, mutations T478K and L452R had the highest frequency of occurrence (66% and 46.6%, respectively) in all the study sequences. These mutations are associated with the Delta variant, and are associated with increased viral escape, increased infectivity and transmissibility. This may explain why the Delta variant was most dominant (45%) of all sequenced samples. Reports have shown that, while some neutralizing antibodies are effective against BA.2.12.1, BA.4 and BA.5 Omicron subvariants, mutations S371F, D405N and R408S undermine most sarbecovirus-neutralizing antibodies (Cao et al., 2022). The absence of mutations D405N and R408S in the RBD from all the Omicron sequences from the current study have several implications. First, while this whole genome sequencing data obtained from this study showed presence of the Omicron variant in the study sites as early as February 2021, the absence of these mutations may have probably influenced its continuous, but dormant circulation in the population. Secondly, the absence of these mutations may explain why the fourth wave of COVID-19 infections, characterized by the Omicron VOC had a decreased severity in the study area. Although high SARSCoV-2 viral loads were detected in wastewater in the study sites, fewer clinical cases were reported. This may have been due to an increase in vaccine uptake in these communities.

The S-gene RBD of study sequences, which the PangoLIN and Nextclade tools only assigned lineages and clades revealed the absence of specific variant defining-mutations which are used in classifying SARS-CoV-2 strains belonging to a specific variant. This may have been the reason why they were only assigned lineages and clades, but not a specific variant name. Mutation Q498H was the most common mutation of these “unassigned variants.” The presence of this mutation is associated with increased binding affinity of the viral spike protein to the ACE2 receptor, which facilitates viral entry during (Bate et al., 2022). The presence of this mutation also boosts the binding of other RBD variants, which could imply an increased infectivity for the population in the presence of this mutation.

Utilizing the current data to further investigate minority variants occurring at lower thresholds in the Spike RBD could potentially predict the next nonsynonymous mutations that may generate another lineage, which may occur in the population. This is relevant because,

although the WHO has announced the end of the COVID-19 pandemic, new Omicron subvariants are constantly emerging, with the latest being of lineage (Africa CDC, 2023), as of July 2023. This highlights the need for constant genomic surveillance, at a population level. Additionally, it could also contribute to vaccine development efforts (Grant et al., 2023), as well as facilitate designation of improved ASG panels.

Allele-specific genotyping has been shown to be a cost-effective method for monitoring variants (Kim and Misra, 2007). Our findings indicate that variant assignment determined by allele-specific or single nucleotide polymorphism (SNP) genotyping was 51.2% accurate when compared to results obtained through WGS. This low accuracy could be due to the fact that the presence of at least one mutation does not necessarily prove the occurrence of a variant, since these variants share ≥1 mutation (Yu et al., 2022). In this study, mutations pertaining to the S-gene were used to detect the occurrence of Alpha, Beta, Delta, and Omicron VoCs in the study sites. The N501Y mutation is shared by all variants except Delta; delH69V70 and mutation P681H are common to both Alpha and Omicron variants; K417N is common to both Beta and Omicron, while mutation L452R is present in both the Delta variant and Omicron BA.4 and BA.5 lineages. This could lead to assigning more than one variant per sample, which may not be a true reflection of variant occurrence. To optimize this technique, and improve variant calling, mutations specific to each variant could be included (Lekana-Douki et al., 2022). Additionally, mutation signatures for each variant need to be constantly updated to ensure ample optimization of this ASG technique. For instance, currently circulating Omicron variants reported by the NGS-SA and the US CDC include JN.1, BA.2.86, as well as KP.2 and KP.3 tagged as variants under monitoring (NGS-SA, 2024; Ma et al., 2024). For accurate variant assignment using the ASG technique, mutations specific to each variant will need to be included in the panels.

This study also evaluated the effect of the USCDC N1 and N2 primers are used to target the N-protein of SARS-CoV-2 during RT-qPCR. Mutation evaluation showed the presence of mutation P13L in the N1 primer. The inference is that these mutations may increase the chance of obtaining false-negative results and may negatively impact the early warning ability of the WBS tool (Lesbon et al., 2021). While one study revealed that the P13L mutation has a lower mortality rate (Sant'Anna et al., 2021), another study indicated that this mutation may lead to negative outcomes (Nagy et al., 2021). No mutations that affected the N2 primer were

detected. These findings indicate the importance of using more than one primer for accurate detection and quantification of SARS-CoV-2 in wastewater. Moreover, as more data on genetic characterization is added, accumulation of mutations in the primers may arise, which will require updating the available primers for accurate and efficient detection of the virus post-pandemic.

Over the last two years, SARS-CoV-2 has been mostly predominant in the population, thus the incidence of respiratory viruses is reported to be profoundly decreased in both pediatric and adult populations (Costanza et al., 2022). However, it has been proposed that as the COVID-19 pandemic wanes, seasonally circulating respiratory infections which were suppressed due to changes in human behaviour will re-emerge and circulate in the population (Laurie and Rockman, 2021; Uppala et al., 2022). These seasonally circulating respiratory viruses include: influenza viruses, endemic human coronaviruses, human metapneumovirus (HMPV), human parainfluenza (HPIVs), respiratory syncytial virus (RSV), and human rhinovirus (HRV). These viruses cause acute respiratory infections (ARIs) like SARS-CoV-2, which are known to negatively affect young children, immunocompromised individuals, and the elderly. This study documents respiratory viruses which were in circulation in the Vhembe and Mopani districts over an 11-month period (January – November 2021). Information from this surveillance shows the cost-effectiveness of WBE in tracking the types of respiratory viruses circulating in the community. This could be beneficial to local public health authorities, since limited resources may hinder active surveillance of clinical specimen. This monitoring is important for preparation of adequate therapies which may be needed as respiratory infections eventually increase.

Circulation, distribution, and seasonality of the 12 detected respiratory viruses in the study sites during the surveillance period (January – November 2021) were described. The most prevalent respiratory viruses were, HAdV – 49.7%, Influenza viruses – 10.9%, HCoVs – 13.9%, HPIV – 7.9%, based on the four seasons occurring in South Africa yearly: Summer (December – February), Autumn (March – May), Winter (June – August), Spring (September – November). Annually, NICD conducts Nationwide Syndromic Respiratory Illness Surveillance monitoring respiratory pathogens in public and private health facilities. Influenza, RSV, Bordetella Pertussis and SARS-CoV-2 were monitored in the first two years of the pandemic i.e. 2020 and 2021 (Public Health Bulletin South Africa, 2020 – 2021). These

respiratory pathogens are the main cause of ARTIs that primarily affect children, immunocompromised individuals and the elderly, leading to hospitalizations. In 2021, during the Influenza season in South Africa, which occurs between May and August, the NICD reported few influenza cases, although an increased infection rate was observed during late Spring. Other clinical respiratory virus surveillance studies in South Africa report a similar trend of fewer Influenza infections detected in the population during the first wave of the COVID-19 pandemic lockdown, which continued throughout with the flu season (Refs). However, over time, with lesser lockdowns, more infections were recorded, although not comparable to pre-pandemic prevalence trends (Parsons et al., 2023; Principi et al., 2023; Davids et al., 2024; Graziani et al., 2024; Reddy et al., 2024).

Evidence from studies done around the world allude that this decline in prevalence and circulation of respiratory viruses is attributable to non-pharmaceutical interventions (NPIs) such as mask-wearing, hand hygiene, social distancing, travel restrictions and school closures, all characteristic of curbing the spread of COVID-19. Of note, IAV and IBV types identified in the study population were markedly different from those reported by the NICD. In the study sites, IAV types A(H1N1)/Puerto Rico, A(H1N1)/Michigan, A(H3N2)/Texas, A(H3N2)/New York, and A(H5N1)/Guandong were in circulation, while the NICD reported types A(H1N1)/Pdm09 and A(H3N2) to be occurring nationwide (NICD Reports, 2021). Furthermore, the A (H5N1) has only been reported been reported to be the only highly pathogenic avian influenza virus (HPAIV) type reported in humans. Data from the WHO show that only one A(H5N1) human infection from India was reported worldwide in 2021. Since this strain is more pathogenic than others, continuous surveillance in the population is necessary (CDC, 2024). This was also observed in IBV types circulating in the study sites (B/Brisbane, B/Lee), compared to those reported to be in circulation in South Africa (B/Victoria) by the NICD (NICD Reports, 2021).

HAdVs were detected throughout the surveillance period, which corroborates with clinical respiratory virus surveillance studies around the world that suggest the lack of seasonality of these viruses (Dey et al., 2013; Garcia-Arroyo et al., 2021; Reddy et al., 2024). Although HAdV was unseasonal, it was most detected in the winter season (76.8%), which may require further study to explain the HAdVs seasonality patterns. The current study also observed a higher prevalence of HCoVs occurring during the winter season, with HCoV-NL63 having the highest

frequency of occurrence (54.5%), followed by HCoV-OC43 (22.7%). Comparatively, clinical surveillance from a study conducted in Gauteng reported continuous circulation of HCoVs throughout the study period, with HCoV-OC43 being most dominant (Reddy et al., 2024). Differences in HCoV type detected from wastewater surveillance compared to clinical surveillance was also observed in a report from the United States. They reported HCoV-OC43 to be most dominant from the wastewater surveillance, while HCoV-229E was most dominant clinically (Boehm et al., 2023). Similar observations were reported from a clinical respiratory virus surveillance study in Madagascar, where HCoV-OC43 was the most dominant HCoV during the spring/summer season in 2020 (Razanajatovo et al., 2022). Another clinical surveillance report from the USA showed prevalence of mainly HCoV-OC43 and HCoV-NL63 in 2021 (Olsen et al., 2021). While the seasonal distribution of common HCoVs has been well described in other African nations like Ghana and Kenya pre-COVID-19, as well as other parts of the world, very little is known about its seasonal pattern in South Africa before the pandemic (Subramoney et al., 2018).

This study observed HPIVs circulating in autumn and winter only, with the highest prevalence occurring during winter (84.6%), and HPIV 3 being the most predominant specie. These results are comparable to what was reported in Western Cape, South Africa, where HPIV-3 took over as the dominantly circulating HPIV post-pandemic, than HPIV-4 which was mostly prevalent pre-COVID-19 (Parsons et al., 2023). This shift in dominance was suggested to have occurred due to an increase in the susceptible population not exposed to the virus. This was corroborated by their results which showed HPIV-3 more prevalent in <2 years old, indicating that lockdown restrictions impacted transmission patterns, since children born during the period were only in close proximity with their families. This further highlights the impact of NPIs on the respiratory virus occurrence patterns. Contrarily, another study conducted in South Africa reported a year-round circulation of HPIV, with 50% of the cases recorded in winter of 2020 caused by HPIV-4.

Thus, this study shows the applicability of WBE in monitoring the incidence, prevalence and seasonality of respiratory viruses responsible for acute respiratory tract infections (ARTIs), particularly in vulnerable populations. It also highlights the advantage of WBS in documenting respiratory virus types at a population level. This could assist national sentinel clinical surveillance efforts documenting patterns of pathogen occurrence, since clinical specimens

are only collected from those who visit healthcare facilities. This was observed where IAV and IBV types circulating in the study sites were markedly different from those reported by the NICD, and when the study's finding on HPIV-3 dominance corroborated reports from the Western Cape that showed HPIV-3 dominating circulation in the population post-COVID-19.

4.5) LIMITATIONS OF THE STUDY

Reports in this study should be considered with some limitations. First, detection of fewer Beta VOCs in the study samples may have been due to surveillance commencement in January 2021, towards the phasing out of the second wave, when the Beta variant was mostly dominant. Due to limited funds, samples were selected to represent the entire surveillance period. Thus, from the 365 SARS-CoV-2 positive samples, only 75 were subjected for WGS. This was done to have a snapshot of variant distribution during the surveillance period. For instance, of the 4 samples positive for SARS-CoV-2, that were collected during the second wave (January - February 2021), when the Beta variant was most dominant, only one sample (collected in Epi Week-03, 2021) was included in the sequence run. This sample was designated as the Delta variant by the GISAID, PangoLIN and Nextclade analysis tools.

Secondly, only mutations pertaining to the S-gene receptor-binding domain (RBD) were used by the allele-specific genotyping (ASG) panel to determine the presence of the Alpha, Beta, Delta, and Omicron VOCs. One drawback of this is that some mutations are shared by multiple VOCs. For instance, mutation N501Y is shared by all variants except Delta; the delH69V70 and P681H mutation are present in both Alpha and Omicron variants. Mutation K417N is common to both Beta and Omicron VOCs, while mutation L452R is present in both the Delta variant and Omicron BA.4 and BA.5 lineages. This could lead to assigning more than one variant per sample, which may not be a true reflection of variant occurrence. This may explain why only a 51.2% congruency was obtained when the VOCs determined by ASG and WGS were compared. This may also give an explanation why the Alpha VOC was determined by WGS, but not found in any whole genome sequence. The droplet digital PCR technique has also been proposed to be implemented for ASG. This technique uses samples and reagents partitioned in large number (~10000 to 20000) "water in oil" droplets, then the PCR reactions are performed on single molecules in individual closed droplets. Thus, rare mutations can be detected and closely related sequences discriminated through probe binding kinetics (Heijnen

et al., 2021). In addition, other qPCR-based approaches can be applied to detect Omicron mutations and discriminate omicron lineages (Kong et al., 2023).

Third, respiratory viruses reported here do not span the complete 18-months surveillance period (18 January 2021 – 28 June 2022). Due to limited resources, respiratory virus data was only available for a 11-months (18 January – 22 November 2021), thus only revealing a snapshot of what occurred during the surveillance period. Additionally, these results may be biased in that the 40 sequenced samples used for analysis were mostly from the autumn and winter seasons (27.5% and 60%, respectively). Thus, further investigations should employ cheaper probe-based qRT-PCR assays detection and genotyping methods for confirmation of reported circulation, distribution and seasonality patterns.

4.6) CONCLUSION AND RECOMMENDATION

The current study demonstrates that population-based approaches in genomic surveillance may be advantageous over individual specific approaches. This study shows that Delta and Omicron lineages were in circulation in the population earlier than reports of their detection in other parts of South Africa. Furthermore, genetic characterization of SARS-CoV-2 viruses at the study sites reveals an array of mutations whose implications need further investigation. Knowing the implication of these mutations could assist in developing vaccine and antibody therapy with increased efficiency. Additionally, our findings show that although ASG is cost-effective for variant monitoring, its efficiency needs to be improved. This could be done through targeted sequencing (such as Sanger Sequencing or NGS), for corroboration of ASG results, to improve variant assignment. Finally, although WBS of SARS-CoV-2 has been most prominent in the last four years, similar efforts should be applied to monitoring other respiratory viruses circulating at the community level. This study detected the circulation of a highly pathogenic H5N1 Influenza virus in 2021, which was not detected in South Africa, from documented NICD clinical surveillance reports. Globally, this H5N1 Influenza virus was only reported in one country (India) in 2021. Thus, such respiratory virus surveillance data could be used in combination with those from clinical respiratory pathogen surveillance for potential public health action. This data could also contribute to decision-making policies adopted for prevention, and management public health diseases, while genomic data could contribute towards vaccine development strategies.

4.7) REFERENCES

Abavisani, M. et al. (2022) 'Mutations in SARS-CoV-2 structural proteins: a global analysis', Virology Journal, 19(1), pp. 1–19. https://doi.org/10.1186/s12985-022-01951-7

Ahmad, S.U. et al. (2022) 'A comprehensive genomic study, mutation screening, phylogenetic and statistical analysis of SARS-CoV-2 and its variant omicron among different countries', Journal of Infection and Public Health, 15(8), pp. 878–891. https://doi.org/10.1016/j.jiph.2022.07.002

Bate, N. et al. (2022) 'In vitro evolution predicts emerging SARS-CoV-2 mutations with high affinity for ACE2 and cross-species binding', PLoS Pathogens, 18(7), pp. 1–19. https://doi.org/10.1371/journal.ppat.1010733

Boehm, A.B. et al. (2023) 'Wastewater concentrations of human influenza, metapneumovirus, parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids during the COVID-19 pandemic: a surveillance study', The Lancet Microbe, 4(5), pp. e340–e348. https://doi.org/10.1016/S2666-5247(22)00386-X

Cao, Y. et al. (2022) 'BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection', Nature, 608(7923), pp. 593–602. https://doi.org/10.1038/s41586-022-04980-y

Castiglioni, S. et al. (2022) 'SARS-CoV-2 RNA in urban wastewater samples to monitor the COVID-19 pandemic in Lombardy, Italy (March–June 2020)', Science of the Total Environment, 806 (Pt 4), pp. 150816. https://doi.org/10.1016/j.scitotenv.2021.150816

Cherian, S. et al. (2021) 'Sars-cov-2 spike mutations, l452r, t478k, e484q and p681r, in the second wave of covid-19 in Maharashtra, India', Microorganisms, 9(7), pp. 1542. https://doi.org/10.3390/microorganisms9071542

Costanza, G. et al. (2022) 'Infection Rate of Respiratory Viruses in the Pandemic SARS-CoV-2 Period Considering Symptomatic Patients: Two Years of Ongoing Observations', Biomolecules, 12(7), pp. 987. https://doi.org/10.3390/biom12070987

Davids, M. et al. (2024) 'Changes in Prevalence and Seasonality of Pathogens Identified in Acute Respiratory Tract Infections in Hospitalised Individuals in Rural and Urban Settings in South Africa: 2018–2022', Viruses, 16(3), 404. https://doi.org/10.3390/v16030404

Dey, S. et al. (2013) 'Prevalence, seasonality, and peak age of infection of enteric adenoviruses in Japan, 1995–2009', Epidemiology and Infection, 141 (5), pp. 958 – 960. https://doi.org/10.1017/S0950268812001586

Dhar, M.S. et al. (2021) 'Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India', Science, 374(6570), pp. 995 – 999. https://doi.org/10.1126/science.abj9932

Faria, N.R. et al. (2021) 'Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil', Science, 372(6544), pp. 815 – 821. https://doi.org/10.1126/science.abh2644

Garcia-Arroyo, L. et al. (2021) 'Prevalence and seasonality of viral respiratory infections in a temperate climate region: A 24-year study (1997–2020)', Influenza and other respiratory Viruses, 16(4), pp. 756 – 766. https://doi.org/10.1111/irv.12972

Grant, R. et al. (2023) 'When to update COVID-19 vaccine composition', Nature Medicine, 29(4), pp. 776–780. https://doi.org/10.1038/s41591-023-02220-y

Graziani, A. et al. (2024) 'Circulation and Seasonality of Respiratory Viruses in Hospitalized Patients during Five Consecutive Years (2019–2023) in Perugia, Italy', Viruses, 16(9), pp. 1394. https://doi.org/10.3390/v16091394

Greaney, A.J. et al. (2021) 'Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies', Nature Communications, 12(1), pp. 4196. https://doi.org/10.1038/s41467-021-24435-8

Gregory, D.A. et al. (2022) 'Genetic diversity and evolutionary convergence of cryptic SARS-CoV-2 lineages detected via wastewater sequencing', PLoS Pathogens, 18(10), pp. 1–25. https://doi.org/10.1371/journal.ppat.1010636

Grubaugh, N.D. et al. (2020) 'We shouldn't worry when a virus mutates during disease outbreaks', Nature Microbiology, 5, pp. 529–530. https://doi.org/10.1038/s41564-020-0690-4

Gu, H. et al. (2022) 'Recombinant BA.1/BA.2 SARS-CoV-2 Virus in Arriving Travelers, Hong Kong, February 2022', Emerging Infectious Diseases, 28(6), pp. 1276 – 1278. https://doi.org/10.3201/eid2806.220523

Haramoto, E. et al. (2020) 'First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan', Science of the Total Environment, 737, pp. 140405. https://doi.org/10.1016/j.scitotenv.2020.140405

Harper, H. et al. (2021) 'Detecting SARS-CoV-2 variants with SNP genotyping', PLoS ONE, 16(2), pp. e0243185. https://doi.org/10.1371/journal.pone.0243185

Harvey, W.T. et al. (2021) 'SARS-CoV-2 variants, spike mutations and immune escape', Nature Reviews Microbiology, 19(7), pp. 409 – 424. https://doi.org/10.1038/s41579-021-00573-0

Hill, V. et al. (2022) 'The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK', Virus Evolution, 8(2), pp. 1 – 13. https://doi.org/10.1093/ve/veac080

Jangra, S. et al. (2021) 'SARS-CoV-2 spike E484K mutation reduces antibody neutralisation', The Lancet Microbe, 2(7), pp. e283–e284. https://doi.org/10.1016/S2666-5247(21)00068-9

Johnson, R. et al. (2022) 'Delineating the Spread and Prevalence of SARS-CoV-2 Omicron Sublineages (BA.1 – BA.5) and Deltacron Using Wastewater in the Western Cape, South Africa', The Journal of Infectious Diseases, 226(8), pp. 1418 – 1427. https://doi.org/10.1093/infdis/jiac356

Kim, S. and Misra, A. (2007) 'SNP genotyping: technologies and biomedical applications' Annual Review of Biomedical Engineering 9, pp. 289–320. https://doi.org/10.1146/annurev.bioeng.9.060906.152037

Korber, B. et al. (2020) 'Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus', Cell, 182(4) pp. 812–827.e19. https://doi.org/10.1016/j.cell.2020.06.043

Kumar, R. et al. (2023) 'Understanding Mutations in Human SARS-CoV-2 Spike Glycoprotein: A Systematic Review & Meta-Analysis', Viruses, 15(4), pp. 856. https://doi.org/10.3390/v15040856

Kumblathan, T. et al. (2021) 'Wastewater-Based Epidemiology for Community Monitoring of SARS-CoV-2: Progress and Challenges', ACS Environmental Au, 1(1), pp. 18–31. https://doi.org/10.1021/acsenvironau.1c00015

Laurie, K.L. and Rockman, S. (2021) 'Which influenza viruses will emerge following the SARS-CoV-2 pandemic?', Influenza and other Respiratory Viruses, 15(5), pp. 573–576. https://doi.org/10.1111/irv.12866

Lauring A.S., (2021) 'Genetic Variants of SARS-CoV-2 — What Do They Mean?', JAMA, 325(6), pp. 529–531. https://doi.org/10.1001/jama.2020.27124

Lekana-Douki, S.E. et al. (2022) 'Screening and Whole Genome Sequencing of SARS-CoV-2 Circulating During the First Three Waves of the COVID-19 Pandemic in Libreville and the Haut-Ogooué Province in Gabon', Frontiers in Medicine, 9, pp. 877391. https://doi.org/10.3389/fmed.2022.877391

Lesbon, J.C.C. et al. (2021) 'Nucleocapsid (N) gene mutations of sars-cov-2 can affect real-time RT-PCR diagnostic and impact false-negative results', Viruses, 13(12), pp. 2474. https://doi.org/10.3390/v13122474

Li, Q. et al. (2021) 'SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape', Cell, 184(9), pp. 2362 – 2371.e9. https://doi.org/10.1016/j.cell.2021.02.042

Liu, H. et al. (2022) 'SARS-CoV-2 Variants of Concern and Variants of Interest Receptor Binding Domain Mutations and Virus Infectivity', Frontiers in Immunology, 13, pp. 825256. https://doi.org/10.3389/fimmu.2022.825256

Liu, Y. et al. (2022) 'The N501Y spike substitution enhances SARS-CoV-2 infection and transmission', Nature, 602(7896), pp. 294–299. https://doi.org/10.1038/s41586-021-04245-0

Ma, K.C. et al., (2024) 'Genomic Surveillance for SARS-CoV-2 Variants: Circulation of Omicron XBB and JN.1 Lineages — United States, May 2023–September 2024' MMWR Morbidity and Mortality Weekly Report 73(42), pp. 938 – 945. https://doi.org/10.15585/mmwr.mm7342a1

Matume, N.D. et al. (2018) 'Next generation sequencing reveals a high frequency of CXCR4 utilizing viruses in HIV-1 chronically infected drug experienced individuals in South Africa', Journal of Clinical Virology, 103, pp. 81–87. https://doi.org/10.1016/j.jcv.2018.02.008

Meng, B. et al. (2021) 'Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7', Cell Reports, 35(13), pp. 109292. https://doi.org/10.1016/j.celrep.2021.109292

Murray, J. et al. (2015) 'Determining the provincial and national burden of influenza-associated severe acute respiratory illness in South Africa using a rapid assessment methodology, PLoS ONE, 10 (7), pp. e0132078. https://doi.org/10.1371/journal.pone.0132078

Nagy, A. et al. (2021) 'Different mutations in SARS-CoV-2 associate with severe and mild outcome', International Journal of Antimicrobial Agents, 57(2), pp. 106272. https://doi.org/10.1016/j.ijantimicag.2020.106272

Olsen, S.J. et al. (2021) 'Changes in influenza and other respiratory virus activity during the COVID-19 pandemic—United States, 2020–2021', MMWR Morbidity and Mortality Weekly Reports, 70(29), pp. 1013–1019. https://doi.org/10.15585/mmwr.mm7029a1

Parsons, J. et al. (2023) 'Human Parainfluenza Virus (HPIV) Detection in Hospitalized Children with Acute Respiratory Tract Infection in the Western Cape, South Africa during 2014–2022 Reveals a Shift in Dominance of HPIV 3 and 4 Infections', Diagnostics (Basel, Switzerland), 3(15), pp. 2576. https://doi.org/10.3390/diagnostics13152576

Principi, N. et al. (2023) 'Epidemiology of Respiratory Infections during the COVID-19 Pandemic', Viruses, 15(5), pp. 1160. https://doi.org/10.3390/v15051160

Rambaut, A. et al. (2020) 'A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology', Nature Microbiology, 5(11), pp. 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

Razanajatovo, N.H. et al. (2022) 'Epidemiological Patterns of Seasonal Respiratory Viruses during the COVID-19 Pandemic in Madagascar, March 2020–May 2022', Viruses 15(1), pp. 12. https://doi.org/10.3390/v15010012

Reddy, B. et al. (2024) 'Prevalence and Seasonal Patterns of 16 Common Viral Respiratory Pathogens during the COVID-19 Pandemic in Gauteng Province, South Africa, 2020–2021', Viruses, 16(8), pp. 1325. https://doi.org/10.3390/v16081325

Sant'Anna, F.H. et al. (2021) 'Emergence of the novel SARS-CoV-2 lineage VUI-NP13L and massive spread of P.2 in South Brazil', Emerging Microbes and Infections, 10(1), pp. 1431–1440. https://doi.org/10.1080/22221751.2021.1949948

Shang, J. et al. (2020) 'Structural basis of receptor recognition by SARS-CoV-2', Nature, 581(7807), pp. 221–224. https://doi.org/10.1038/s41586-020-2179-y

Shrestha, B.L. et al. (2022) 'Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission', Reviews in Medical Virology, 32(5), pp. e2381. https://doi.org/10.1002/rmv.2381

Subramoney, K. et al. (2018) 'Human bocavirus, coronavirus, and polyomavirus detected among patients hospitalised with severe acute respiratory illness in South Africa, 2012 to 2013', Health Science Reports, 1(8), pp. e59. https://doi.org/10.1002/hsr2.59

Sun, C. et al. (2022) 'Molecular characteristics, immune evasion, and impact of SARS-CoV-2 variants', Signal Transduction and Targeted Therapy, 7(1), pp. 202. https://doi.org/10.1038/s41392-022-01039-2

Swets, M.C. et al. (2022) 'SARS-CoV-2 co-infection with influenza viruses, respiratory syncytial virus, or adenoviruses', The Lancet, 399(10334), pp. 1463–1464. https://doi.org/10.1016/S0140-6736(22)00383-X

Takemae, N. et al. (2022) 'Development of new SNP genotyping assays to discriminate the Omicron variant of SARS-CoV-2', Japanese Journal of Infectious Diseases, 75(4), pp. 411 – 414. https://doi.org/10.7883/yoken.jjid.2022.007

Tegally, H. et al. (2021) 'Detection of a SARS-CoV-2 variant of concern in South Africa', Nature, 592(7854), pp. 438–443. https://doi.org/10.1038/s41586-021-03402-9

Tegally, H. et al. (2022) 'Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa', Nature Medicine, 28(9), pp. 1785–1790. https://doi.org/10.1038/s41591-022-01911-2

Tempia, S. et al. (2021) 'Decline of influenza and respiratory syncytial virus detection in facility-based surveillance during the COVID-19 pandemic, South Africa, January to October 2020. Eurosurveillance, 26(29), pp. 2001600. https://doi.org/10.2807/1560-7917.ES.2021.26.29.2001600

O'Toole, Á. et al. (2021) 'Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. Wellcome Open Research, 6, pp. 121. https://doi.org/10.12688/wellcomeopenres.16661.2

Trujillo, M. et al. (2021) 'Protocol for safe, affordable, and reproducible isolation and quantitation of SARS-CoV-2 RNA from wastewater', PLoS ONE, 16(9), pp. e0257454. https://doi.org/10.1371/journal.pone.0257454

Uppala, R. et al. (2022) 'Effect of the COVID-19 Pandemic on Lower Respiratory Tract Infection Determinants in Thai Hospitalized Children: National Data Analysis 2015–2020', Tropical Medicine and Infectious Disease, 7(8), pp. 151. https://doi.org/10.3390/tropicalmed7080151

V'kovski, P. et al. (2021) 'Coronavirus biology and replication: implications for SARS-CoV-2', Nature Reviews Microbiology, 19(3), pp. 155–170. https://doi.org/10.1038/s41579-020-00468-6

Viana, R. et al. (2022) 'Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa', Nature, 603(7902), pp. 679–686. https://doi.org/10.1038/s41586-022-04411-y

Wang, R. et al. (2021) 'Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries', Genomics, 113(4), pp. 2158–2170. https://doi.org/10.1016/j.ygeno.2021.05.006

Yu, A.T. et al. (2022) 'Estimating Relative Abundance of 2 SARS-CoV-2 Variants through Wastewater Surveillance at 2 Large Metropolitan Sites, United States', Emerging Infectious Diseases, 28(5), pp. 940–947. https://doi.org/10.3201/eid2805.212488

Zafeiriadou, A. et al. (2023) 'Wastewater surveillance of the most common circulating respiratory viruses in Athens: The impact of COVID-19 on their seasonality', Science of the Total Environment, 900, pp. 166136. https://doi.org/10.1016/j.scitotenv.2023.166136

Zhang, L. et al. (2020) 'Genomic variations of SARS-CoV-2 suggest multiple outbreak sources of transmission', https://doi.org/10.1101/2020.02.25.20027953

WEBSITES

WHO (2023), 'Tracking SARS-CoV-2 Variants' (https://www.who.int/activities/tracking-SARS-CoV-2-variants) (accessed 15 January 2023)

WHO Facts Sheet (2021), 'The Top 10 causes of Death' (https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death) (accessed 20 February 2021)

WHO (2023), 'Leading Causes of Death' (https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death) (accessed 08 January 2023)

NICD (2021), 'SARS-CoV-2 Genomic Surveillance Update' (https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance-update/) (accessed December 2021)

Africa CDC (2023), 'Statement on the new COVID strain, EG.5 SARS-COV-2 subvariant' (https://africacdc.org/news-item/statement-on-the-new-covid-strain-eg-5-sars-cov-2-subvariant/#:~:text=On%209%20August%202023%2C%20the,of%20E.G.5%20cases%20reported) (Accessed 01 October 2023)

Public Health Bulletin South Africa (2021), 'Respiratory Pathogen Epidemiology from the Systematic Influenza-like Illness and Pneumonia Surveillance Programmes'. (https://www.phbsa.ac.za/respiratory-pathogen-epidemiology-from-the-systematic-influenza-like-illness-and-pneumonia-surveillance-programmes/) (Accessed 14 October 2024)

CDC (2024), 'Past Reported Global Human Cases with Highly Pathogenic Avian Influenza A(H5N1) (HPAI H5N1) by Country, 1997-2024' (https://www.cdc.gov/bird-flu/php/avian-flu-summary/chart-epi-curve-ah5n1.html) (accessed October 2024)

NICD (2024), 'Tracking SARS-CoV-2 Variants: SARS-CoV-2 Genomic Surveillance Update. Based on NGS-SA report' (https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/sars-cov-2-genomic-surveillance-update/) (Accessed October 2024).

CHAPTER FIVE

General Conclusions Limitations and Recommendations

5.1) GENERAL CONCLUSIONS

This study delved into HCoVs investigation in Africa, and rural areas in Limpopo province, South Africa. First, it systematically reviews the distribution of HCoVs in Africa prior to the COVID-19 pandemic. Then, using a non-invasive, cost-efficient, and population-based method of observation (WBE), focusing on one HCoV (SARS-CoV-2), this study reports the trends, VOC occurrence, and estimated prevalence in the Vhembe and Mopani districts between January 2021 to June 2022. Finally, this study documented the genomic epidemiology of SARS-CoV-2, and other respiratory viruses in the study sites during the surveillance period. The following conclusions are drawn from the findings in this study.

The systematic review showed that very few investigations were done in Africa pre COVID-19, most of which were concentrated in one region (Kenya, East Africa). While the studies show a low frequency of endemic HCoVs and MERS-CoV in Africa, it also shows the huge need for genomic epidemiology studies to adequately characterize these viruses in the African population. This review brought to light the dearth of information in regard to HCoVs from some nations, and the necessity to establish and maintain surveillance systems to monitor introductions of any more infectious species in the population. It finally emphasizes the need for community-based surveillance studies for comprehensive understanding of HCoV circulation in the African population, since hospital-based surveillance studies may not be applicable to the general population.

Secondly, this study established a wastewater-based surveillance system for monitoring SARS-CoV-2 occurrence, predicting the prevalence of infected individuals, determining the VOCs in the selected study sites. The ability of wastewater surveillance as a useful multi-faceted tool for early warning detection of diseases, tracking trends of infection, and highlighting the variants circulating at a population level in real-time, are also described. Established wastewater surveillance systems are important for pandemic preparedness since are instrumental for identification of infection hotspots, and implementation of necessary measures to curb transmission effectively. Application of a mathematical model to estimate the number of infections in a catchment area could be additionally beneficial to guide public health response policies during periods of sporadic infection epidemics. This study lays a foundation for the implementation of wider surveillance studies that can be used for tracking not only SARS-CoV-2, but other enteric pathogens circulating in this study population. This

information is relevant for tracking pathogen circulation in the region and is mostly important for the interception of disease once outbreaks are identified.

Lastly, the current study demonstrates that population-based approaches in genomic surveillance may be advantageous over individual specific approaches. This study has shown that Delta and Omicron lineages were in circulation in the population earlier than reports of their detection in other parts of South Africa. Furthermore, genetic characterization of SARS-CoV-2 viruses in the study sites has revealed an array of mutations whose implications need further investigation. Knowledge of the implication of these mutations could assist in improving vaccine development and antibody therapy efficiency. Additionally, improving the efficiency of allele specific genotyping is necessary to harness the cost-effectivity of this method. This will be beneficial for variant monitoring, in LMIC, where there are limited resources for NGS approaches. Finally, monitoring other respiratory viruses in the population is relevant. This is evident from the identification of a highly pathogenic H5N1 Influenza virus (A/goose/Guandong/1/1996(H5N1)) detected in the study sites, which was not detected through clinical surveillance in South Africa, and only one clinical case occurring globally in 2021, as reported by the WHO. Such data could contribute to decision-making policies adopted for prevention, and vaccination strategies used for management of public health diseases.

5.2) STUDY LIMITATIONS

While the implementation of WBE has many advantages, it should be seen in line with the following. First, the absence of recent data on SARS-CoV-2 RNA copies in per gram of feces, and the gram of faeces released per person per day in South Africa, as well as lack of adequate home-connected sewage systems may have an impact on the predicted number of people infected. Updating this information is necessary for improved estimates of infected people. Secondly, the allele specific genotyping method used for variant detection only catered for SARS-CoV-2 positive samples with concentrations ≥1500g.c./mL and used similar mutations common to all four variants (Alpha, Beta, Delta and Omicron). Thus, VOCs present in samples which do not meet this concentration criterion will be unknown, and information on circulating variants in times when viral concentrations in the study sites are low will likely be missed. Therefore, optimization of this protocol to include samples with lower viral

concentrations, as well as variant-unique mutations is necessary for adequate genomic surveillance. Another drawback that may hinder effective surveillance in LMIC is the absence of sewage systems. This was observed in the study areas; thus, application of decentralized surveillance methods could potentially circumvent this limitation.

Thirdly, using ASG for variant assignment is advantageous, however, this study noted that when mutations specific to the Spike gene are solely used, variant assignment may not be accurate. This is because some mutations in the Spike gene are shared across variants.

5.3) RECOMMENDATIONS

From this study, the following recommendations could be made: the necessity for long-term surveillance of HCoVs will provide comprehensive data on virus circulation and compensate for year-to-year variation of HCoVs prevalence. Much more needs to be done in terms of genomic surveillance for endemic HCoVs in Africa, as a pandemic preparedness measure. These studies will offer better insight on the phylogeography, genetic diversity, and evolutionary characteristics of HCoVs. Thus, cost-effective ways to achieve this is through allele-specific genotyping, in combination with sanger sequencing methods for confirmation. Such methods could be applied on a wider scale across the continent. To achieve all these in Africa, which has many resource-limited areas, investing in wastewater-based surveillance could be economically beneficial since it is cost-effective, and caters for both symptomatic and asymptomatic populations. One way to achieve this is through targeted surveillance, using a decentralized WBS approach to investigate contaminated surface waters, since most homes in the continent may not be connected to sewage systems. A combination of both WBE and clinical surveillance will be instrumental, not only in pandemic preparedness, but also in arriving at specific solutions for preventing rapid disease dissemination, as was the case with COVID-19. Finally, this study also recommends training personnel, who will be equipped for any sporadic epidemics or pandemics.

APPENDICES

APPENDIX 1: Representative tables (1 – 8) showing estimated prevalence, alongside the 95% confidence intervals. Correlation between the estimated prevalence and the associated 7-day moving average clinical data is also shown

N.B: Weekly estimated prevalence medians represented here were obtained from Monte Carlo model simulations computed with 10,000 iterations in Microsoft Excel. Only samples met the modelling inclusion criteria (SARS-CoV-2 copies/mL with SD <1) were used for computation

Vhembe District

Table 1: Monte Carlo predicted prevalence and 95% confidence intervals in the Malamulele (MA) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Collins Chabane sub-district is also shown.

Epi Week	MA Estimated Prevalence Median	MA Estimated Prevalence (Lower 95% CI)	MA Estimated Prevalence (Upper 95% CI)	Collins Chabane sub-district Clinical Cases 7D MA	Correlation between MA Estimated Prevalence & Collins Chabane sub-district Clinical cases 7D-MA
03 (18 Jan 2021)	6,9	50,0	66,1	45,9	
08 (22 Feb 2021)	0,05	1,6	5,9	0,3	
25 (23 Jun 2021)	2,0	17,7	30,3	4,6	
26 (29 Jun 2021)	328,9	2273,7	3026,7	11,0	
27 (05 Jul 2021)	5,1	30,6	41,9	30,6	
28 (13 Jul 2021)	7,7	54,6	74,8	34,6	
29 (19 Jul 2021)	22,8	145,0	219,0	23,1	
30 (26 Jul 2021)	5,8	38,5	51,6	16,3	
31 (02 Aug 2021)	3,2	23,9	32,6	7,3	
33 (16 Aug 2021)	2,3	14,5	19,7	0,6	
34 (23 Aug 2021)	0,8	8,3	14,0	2,6	
38 (20 Sep 2021)	3,7	31,7	43,6	0,3	
40 (04 Oct 2021)	0,1	1,9	6,7	0,3	
48 (29 Nov 2021)	0,1	1,4	2,5	0,3	
49 (06 Dec 2021)	1,9	52,8	167,0	2,1	
50 (13 Dec 2021)	28,2	204,3	278,5	16,0	
51 (20 Dec 2021)	36,6	249,4	334,3	17,1	
02 (10 Jan 2022)	3,1	21,2	28,5	8,4	
03 (17 Jan 2022)	4,4	33,9	46,6	2,3	
04 (24 Jan 2022)	5,8	40,7	56,3	6,3	
14 (04 Apr 2022)	1,2	9,6	13,4	4,3	
17 (25 Apr 2022)	2,3	18,8	28,9	0,4	
18 (02 May 2022)	13,7	103,2	142,6	1,1	
19 (09 May 2022)	1,1	9,0	12,1	0,4	
23 (07 Jun 2022)	3,6	25,2	34,1	0,3	

Coefficient (rs)	0,6943482
N:	25
T Statistic:	4,6272874
DF:	23
p value:	1E-04

Table 2: Monte Carlo predicted prevalence and 95% confidence intervals in the Thohoyandou (TH) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Thulamela sub-district is also shown.

Epi Week	TH Estimated Prevalence Median	TH Estimated Prevalence (Lower 95% CI)	TH Estimated Prevalence (Upper 95% CI)	Thulamela sub-district Clinical Cases 7D MA	Correlation between TH Estimated Prevalence & Thulamela sub-district Clinical cases 7D-MA
15 (12 Apr 2021)	8,3	29,2	41,7	1,9	
19 (10 May 2021)	9,9	42,4	57,3	1,6	
20 (17 May 2021)	7,9	59,3	81,3	2,4	
23 (07 Jun 2021)	0,0	1,1	2,7	2,1	
25 (23 Jun 2021)	0,4	3,5	4,9	17,9	
26 (29 Jun 2021)	6,9	45,7	62,4	39,9	
27 (05 Jul 2021)	3,0	29,2	43,6	78,9	
28 (13 Jul 2021)	4,0	29,6	39,7	85,3	
29 (19 Jul 2021)	2,5	17,8	24,0	85,7	
30 (26 Jul 2021)	5,6	37,5	51,3	52,3	
31 (02 Aug 2021)	5,6	52,6	88,4	32,6	
32 (09 Aug 2021)	1,5	11,5	15,6	17,4	
33 (16 Aug 2021)	1,7	12,6	17,5	11,6	
34 (23 Aug 2021)	0,7	5,9	10,4	12,9	
35 (30 Aug 2021)	0,1	1,0	1,3	6,9	
36 (06 Sep 2021)	0,2	1,7	2,4	8,9	
37 (13 Sep 2021)	0,1	0,9	2,2	5,0	
38 (20 Sep 2021)	0,2	1,9	3,2	2,1	
39 (27 Sep 2021)	0,4	2,7	3,8	1,0	
40 (04 Oct 2021)	0,4	2,9	4,1	0,6	
42 (18 Oct 2021)	0,2	2,3	2,3	0,4	
47 (22 Nov 2021)	0,5	4,3	5,7	0,6	
48 (29 Nov 2021)	0,4	3,1	4,7	2,1	
49 (06 Dec 2021)	4,4	30,7	41,1	24,4	
50 (13 Dec 2021)	725,6	3764,3	5017,4	68,7	
51 (20 Dec 2021)	6,6	48,9	67,1	64,7	
02 (10 Jan 2022)	2,8	22,4	30,7	21,7	
03 (17 Jan 2022)	0,2	1,2	1,6	30,3	
04 (24 Jan 2022)	3,9	29,8	41,9	45,0	
06 (07 Feb 2022)	0,1	1,0	1,4	33,1	
17 (25 Apr 2022)	0,1	0,5	0,7	1,4	
18 (02 May 2022)	19,9	122,5	162,6	2,7	
19 (09 May 2022)	1,9	15,0	20,3	2,0	
20 (16 May 2022)	3,6	93,2	195,1	2,9	
22 (31 May 2022)	25,5	292,3	458,8	1,7	
23 (07 Jun 2022)	1,4	11,3	17,3	2,0	
25 (21 Jun 2022)	0,3	3,0	5,0	0,3	
26 (28 Jun 2022)	0,8	5,9	7,9	0,1	

Coefficient (rs)	0,3061515
N:	43
T Statistic:	2,0592033
DF:	41
***p* value:**	5E-02

Table 3: Monte Carlo predicted prevalence and 95% confidence intervals in the Louis Trichardt (LT) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Makhado sub-district is also shown.

Week	LT Estimated Prevalence Median	LT Estimated Prevalence (Lower 95% CI)	LT Estimated Prevalence (Upper 95% CI)	Makhado sub-district Clinical Cases 7D MA	Correlation between LT Estimated Prevalence & Makhado sub-district Clinical cases 7D-MA
19 (10 May 2021)	16,6	123,1	174,8	0,6	
23 (07 Jun 2021)	40,1	745,9	1460,8	3,9	
24 (15 Jun 2021)	63,8	448,0	613,0	6,0	
25 (23 Jun 2021)	309,2	2020,8	2756,5	24,9	
26 (29 Jun 2021)	833,7	6351,1	8574,2	58,4	
27 (05 Jul 2021)	0,2	2,4	3,8	123,3	
28 (13 Jul 2021)	339,9	2133,9	2894,8	136,1	
29 (19 Jul 2021)	92,5	642,1	861,7	99,0	
30 (26 Jul 2021)	105,4	903,3	1248,3	58,4	
31 (02 Aug 2021)	162,4	1169,4	1546,0	30,7	
32 (09 Aug 2021)	2,9	19,0	25,4	21,1	
33 (16 Aug 2021)	96,8	646,9	865,8	13,6	
34 (23 Aug 2021)	72,1	576,4	786,7	13,3	
35 (30 Aug 2021)	10,8	105,5	175,5	6,4	
36 (06 Sep 2021)	1,1	29,0	98,4	4,9	
37 (13 Sep 2021)	17,6	174,7	240,2	2,6	
38 (20 Sep 2021)	15,0	99,0	139,7	1,1	
45 (08 Nov 2021)	95,6	847,4	1247,8	2,7	
48 (29 Nov 2021)	0,7	18,1	32,1	1,1	
49 (06 Dec 2021)	279,7	1938,2	2686,9	9,9	
50 (13 Dec 2021)	650,9	4279,2	5784,5	41,0	
51 (20 Dec 2021)	38,1	353,6	510,4	34,6	
02 (10 Jan 2022)	191,9	1411,0	1963,8	19,0	
03 (17 Jan 2022)	31,7	214,6	296,6	17,0	
04 (24 Jan 2022)	136,0	1019,1	1406,5	48,3	
08 (21 Feb 2022)	49,1	303,3	413,0	7,6	
16 (18 Apr 2022)	4,2	33,2	46,5	0,4	
18 (02 May 2022)	128,5	835,2	1127,5	1,6	
19 (09 May 2022)	52,7	388,7	551,9	2,6	
20 (16 May 2022)	91,9	617,2	831,5	1,9	
23 (07 Jun 2022)	156,1	1069,4	1454,6	2,4	
25 (21 Jun 2022)	2,3	55,8	241,8	0,1	
26 (28 Jun 2022)	76,8	1667,5	3384,5	0,3	

Coefficient (rs)	0,4434601
N:	33
T Statistic:	2,7547669
DF:	31
p value:	1E-02

Table 4: Monte Carlo predicted prevalence and 95% confidence intervals in the Siloam Hospital (SI) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Makhado sub-district is also shown.

Week	SI Estimated Prevalence Median	SI Estimated Prevalence (Lower 95% CI)	SI Estimated Prevalence (Upper 95% CI)	Makhado sub-district Clinical Cases 7D MA	Correlation between SI Estimated Prevalence & Makhado sub-district Clinical cases 7D-MA
03 (18 Jan 2021)	10,6	143,8	288,8	128,1	
21 (24 May 2021)	7,9	89,9	321,3	1,3	
25 (23 Jun 2021)	51,2	350,1	466,3	17,9	
26 (29 Jun 2021)	442,2	3301,9	4490,9	39,9	
27 (05 Jul 2021)	33,5	282,5	440,3	78,9	
28 (13 Jul 2021)	151,5	932,9	1270,3	85,3	
29 (19 Jul 2021)	1001,2	8241,3	11418,9	85,7	
30 (26 Jul 2021)	941,0	6808,4	9322,5	52,3	
31 (02 Aug 2021)	83,8	572,5	767,2	32,6	
32 (09 Aug 2021)	74,2	489,7	666,8	17,4	
33 (16 Aug 2021)	21,7	162,4	225,0	11,6	
38 (20 Sep 2021)	8,8	63,5	86,0	2,1	
46 (15 Nov 2021)	2,2	18,0	24,8	0,6	
49 (06 Dec 2021)	8,9	217,8	481,8	24,4	
50 (13 Dec 2021)	35,7	253,7	355,2	68,7	
51 (20 Dec 2021)	72,7	565,9	778,4	64,7	
02 (10 Jan 2022)	476,0	3132,9	4149,1	21,7	
04 (24 Jan 2022)	35,5	253,7	347,0	45,0	
05 (31 Jan 2022)	245,9	1877,1	2504,6	59,7	
06 (07 Feb 2022)	5,0	38,2	52,0	33,1	
07 (14 Feb 2022)	1877,2	22253,5	36679,0	11,6	
08 (21 Feb 2022)	2,9	20,3	28,1	7,6	
11 (14 Mar 2022)	1,8	75,1	150,6	1,4	
12 (21 Mar 2022)	0,4	11,7	21,3	0,6	
17 (25 Apr 2022)	120,0	928,2	1261,2	1,4	
18 (02 May 2022)	39,8	291,5	394,0	2,7	
22 (31 May 2022)	222,9	1859,7	3048,3	1,7	
23 (07 Jun 2022)	1714,1	12734,1	17133,1	2,0	
25 (21 Jun 2022)	2549,5	20947,5	33415,6	0,3	

Coefficient (rs)	0,2131329
N:	29
T Statistic:	1,1335158
DF:	27
p value:	3E-01

Mopani District

Table 5: Monte Carlo predicted prevalence and 95% confidence intervals in the Tzaneen site (TZ). The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Greater Tzaneen sub-district is also shown.

Week	TZ Estimated Prevalence Median	TZ Estimated Prevalence (Lower 95% CI)	TZ Estimated Prevalence (Upper 95% CI)	Greater Tzaneen sub-district Clinical Cases 7D MA	Correlation between TZ Estimated Prevalence & Greater Tzaneen sub-district Clinical cases 7D-MA
23 (07 Jun 2021)	13,4	95,7	132,5	5,7	
26 (29 Jun 2021)	31,4	231,5	346,7	28,7	
27 (05 Jul 2021)	1,0	7,3	10,1	46,1	
28 (13 Jul 2021)	6,5	57,1	77,9	66,4	
29 (19 Jul 2021)	14,4	110,8	149,9	69,7	
30 (26 Jul 2021)	7,5	53,0	69,7	61,9	
31 (02 Aug 2021)	12,7	98,6	134,3	59,9	
32 (09 Aug 2021)	11,4	91,0	120,2	41,3	
33 (16 Aug 2021)	7,3	75,7	125,0	28,4	
34 (23 Aug 2021)	11,3	85,4	114,3	33,6	
36 (06 Sep 2021)	0,3	2,3	3,1	16,9	
38 (20 Sep 2021)	2,8	27,3	38,7	6,4	
39 (27 Sep 2021)	0,5	5,4	9,6	4,7	
48 (29 Nov 2021)	1,8	21,3	36,8	1,9	
49 (06 Dec 2021)	2,4	18,9	26,8	12,1	
50 (13 Dec 2021)	62,0	563,5	788,1	66,0	
51 (20 Dec 2021)	16,6	136,3	186,0	74,7	
02 (10 Jan 2022)	6,1	58,0	80,2	34,1	
03 (17 Jan 2022)	4,7	36,5	52,3	59,9	
04 (24 Jan 2022)	4,9	33,5	46,0	39,6	
06 (07 Feb 2022)	4,7	41,3	62,4	23,6	
13 (28 Mar 2022)	0,5	4,7	7,3	1,6	
17 (25 Apr 2022)	2,0	16,2	22,2	3,0	
18 (02 May 2022)	57,6	388,5	547,3	3,3	
19 (09 May 2022)	14,8	116,0	159,9	3,3	
20 (16 May 2022)	20,5	117,4	155,3	5,1	
22 (31 May 2022)	18,8	160,4	238,7	4,4	
23 (07 Jun 2022)	8,6	107,0	165,5	7,0	
25 (21 Jun 2022)	31,4	244,2	345,1	17,6	
26 (07 Feb 2022)	26,5	223,3	324,4	23,7	

Coefficient (rs)	0,6835026
N:	35
T Statistic:	5,3790445
DF:	33
p value:	6E-06

Table 6: Monte Carlo predicted prevalence and 95% confidence intervals in the Nkowankowa (NK) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Greater Tzaneen sub-district is also shown.

Week	NK Estimated Prevalence Median	NK Estimated Prevalence (Lower 95% CI)	NK Estimated Prevalence (Upper 95% CI)	Greater Tzaneen sub-district Clinical Cases 7D MA	Correlation between NK Estimated Prevalence & Greater Tzaneen sub-district Clinical cases 7D-MA
22 (31 May 2021)	0,5	6,1	11,7	2,3	
23 (07 Jun 2021)	1,0	12,4	22,7	5,7	
25 (23 Jun 2021)	9,8	80,9	116,2	18,4	
26 (29 Jun 2021)	28,8	182,9	245,2	28,7	
27 (05 Jul 2021)	14,2	113,6	167,3	46,1	
28 (13 Jul 2021)	15,3	99,4	132,8	66,4	
29 (19 Jul 2021)	24,2	186,6	259,6	69,7	
30 (26 Jul 2021)	25,4	173,6	236,6	61,9	
31 (02 Aug 2021)	7,5	54,7	72,1	59,9	
32 (09 Aug 2021)	6,1	46,3	62,8	41,3	
33 (16 Aug 2021)	12,0	132,8	197,5	28,4	
34 (23 Aug 2021)	2,1	22,8	33,2	33,6	
35 (30 Aug 2021)	1,6	11,6	16,5	26,9	
36 (06 Sep 2021)	5,5	39,5	54,9	16,9	
37 (13 Sep 2021)	1,6	19,1	35,5	10,3	
38 (20 Sep 2021)	2,1	16,0	21,8	6,4	
39 (27 Sep 2021)	0,2	1,4	2,0	4,7	
40 (04 Oct 2021)	0,3	3,5	9,5	2,7	
48 (29 Nov 2021)	1,2	12,0	69,8	1,9	
49 (06 Dec 2021)	20,7	164,9	254,6	12,1	
50 (13 Dec 2021)	37,8	265,3	356,2	66,0	
51 (20 Dec 2021)	29,3	210,2	286,1	74,7	
02 (10 Jan 2022)	15,8	96,5	130,0	34,1	
03 (17 Jan 2022)	8,2	48,9	65,6	59,9	
04 (24 Jan 2022)	8,0	58,1	79,2	39,6	
05 (31 Jan 2022)	4,4	27,5	37,3	34,6	
06 (07 Feb 2022)	0,7	4,6	6,2	23,6	
08 (21 Feb 2022)	3,0	25,4	38,4	15,3	
16 (18 Apr 2022)	0,1	2,5	4,9	1,0	
17 (25 Apr 2022)	1,0	7,0	9,8	3,0	
18 (02 May 2022)	13,5	96,4	139,6	3,3	
19 (09 May 2022)	0,6	6,1	9,5	3,3	
20 (16 May 2022)	8,6	57,4	81,3	5,1	
22 (31 May 2022)	11,6	166,8	276,4	4,4	
23 (07 Jun 2022)	13,7	93,7	128,1	7,0	
25 (21 Jun 2022)	1,8	31,7	65,4	17,6	
26 (28 Jun 2022)	2,0	16,9	23,6	23,7	

Coefficient (rs)	0,6748074
N:	37
T Statistic:	5,4095488
DF:	35
p value:	5E-06

Table 7: Monte Carlo predicted prevalence and 95% confidence intervals in the Ga-Kgapane (KA) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Greater Letaba sub-district is also shown.

Week	KA Estimated Prevalence Median	KA Estimated Prevalence (Lower 95% CI)	KA Estimated Prevalence (Upper 95% CI)	Greater Letaba subdistrict Clinical Cases 7D MA	Correlation between KA Estimated Prevalence & Greater Letaba sub-district Clinical cases 7D-MA
16 (19 Apr 2021)	1,3	8,1	10,9	0,1	
17 (26 Apr 2021)	2,7	25,0	36,8	0,1	
18 (05 May 2021)	15,8	101,5	137,9	0,0	
19 (10 May 2021)	12,3	92,9	124,3	0,4	
20 (17 May 2021)	1,3	9,0	12,2	0,3	
21 (24 May 2021)	95,7	664,3	904,2	0,1	
22 (31 May 2021)	12,1	90,3	121,6	0,1	
23 (07 Jun 2021)	12,7	96,0	128,7	0,7	
24 (15 Jun 2021)	9,9	72,4	97,8	1,0	
25 (23 Jun 2021)	65,9	524,3	704,0	1,9	
26 (29 Jun 2021)	1,5	10,0	13,7	3,7	
28 (13 Jul 2021)	2,1	13,7	18,3	8,1	
29 (19 Jul 2021)	4,8	32,1	43,3	12,1	
30 (26 Jul 2021)	4,0	26,2	35,3	9,1	
31 (02 Aug 2021)	9,9	61,4	82,5	7,1	
32 (09 Aug 2021)	9,3	63,5	86,3	7,0	
33 (16 Aug 2021)	13,8	109,1	154,8	5,0	
34 (23 Aug 2021)	1,9	14,4	19,4	5,6	
35 (30 Aug 2021)	1,1	8,5	11,6	3,0	
36 (06 Sep 2021)	0,4	2,9	3,8	2,0	
37 (13 Sep 2021)	0,5	3,1	4,3	0,6	
38 (20 Sep 2021)	0,6	23,2	47,1	0,4	
39 (27 Sep 2021)	0,1	0,3	0,5	0,6	
40 (04 Oct 2021)	0,009	0,3	0,6	0,6	
42 (18 Oct 2021)	0,1	1,0	1,3	0,1	
46 (15 Nov 2021)	0,1	0,8	1,0	0,1	
49 (06 Dec 2021)	0,3	1,9	2,8	2,0	
50 (13 Dec 2021)	6,1	41,6	57,2	5,1	
51 (20 Dec 2021)	6,3	47,7	64,7	3,7	
02 (10 Jan 2022)	2,2	18,3	24,7	4,6	
04 (24 Jan 2022)	0,6	4,8	8,1	5,4	
05 (31 Jan 2022)	0,1	1,2	1,9	4,3	
11 (14 Mar 2022)	0,03	0,2	0,3	1,4	
17 (25 Apr 2022)	0,4	3,2	4,5	0,0	
18 (02 May 2022)	1,3	10,3	14,7	0,1	
20 (16 May 2022)	0,9	9,7	16,4	0,7	
22 (31 May 2022)	11,5	116,1	189,8	0,1	
23 (07 Jun 2022)	77,5	493,1	680,8	0,0	
25 (21 Jun 2022)	0,9	6,3	8,4	0,0	
26 (28 Jun 2022)	1,6	12,8	17,3	0,3	

Coefficient (rs)	0,0057589
N:	40
T Statistic:	0,0355007
DF:	38
p value:	1E+00

Table 8: Monte Carlo predicted prevalence and 95% confidence intervals in the Giyani (GI) site. The correlation between the estimated prevalence and 7-day moving average (7D-MA) of clinical cases in the Greater Giyani sub-district is also shown.

Week	GI Estimated Prevalence Median	GI Estimated Prevalence (Lower 95% CI)	GI Estimated Prevalence (Upper 95% CI)	Greater Giyani sub-district Clinical Cases 7D MA	Correlation between GI Estimated Prevalence & Greater Giyani sub-district Clinical cases 7D-MA
26 (29 Jun 2021)	5,7	45,9	60,8	8,4	
28 (13 Jul 2021)	6,2	46,8	63,6	26,0	
29 (19 Jul 2021)	9,7	69,0	94,2	35,1	
30 (26 Jul 2021)	15,8	106,3	142,8	17,3	
31 (02 Aug 2021)	5,9	43,5	58,0	11,7	
32 (09 Aug 2021)	1,6	10,7	15,0	5,7	
33 (16 Aug 2021)	1,1	9,9	14,6	5,0	
34 (23 Aug 2021)	5,0	40,4	55,4	4,0	
35 (30 Aug 2021)	2,6	18,1	25,2	6,1	
36 (06 Sep 2021)	0,5	2,9	4,0	2,4	
37 (13 Sep 2021)	0,3	4,1	7,2	1,9	
39 (27 Sep 2021)	0,1	1,5	3,4	0,6	
40 (04 Oct 2021)	0,2	1,1	1,5	0,0	
41 (11 Oct 2021)	1,9	12,5	17,4	2,3	
47 (22 Nov 2021)	1,0	7,6	11,0	0,1	
48 (29 Nov 2021)	0,5	17,4	39,6	0,6	
49 (06 Dec 2021)	17,1	129,7	183,6	3,4	
50 (13 Dec 2021)	5,8	44,6	62,0	18,3	
51 (20 Dec 2021)	1,9	14,3	18,9	16,6	
02 (10 Jan 2022)	2,0	13,8	18,4	5,7	
03 (17 Jan 2022)	6,6	59,6	88,0	12,4	
04 (24 Jan 2022)	2,7	23,9	35,6	8,6	
05 (31 Jan 2022)	14,1	116,0	160,3	17,4	
06 (07 Feb 2022)	0,8	6,4	9,7	9,1	
07 (14 Feb 2022)	0,6	7,7	13,2	12,0	
12 (21 Mar 2022)	0,1	0,9	1,3	2,7	
14 (04 Apr 2022)	1,7	13,1	18,5	1,3	
16 (18 Apr 2022)	0,1	0,6	0,8	0,3	
17 (25 Apr 2022)	1,9	14,5	20,0	1,0	
18 (02 May 2022)	10,7	83,5	113,2	2,3	
19 (09 May 2022)	3,5	27,6	38,1	2,7	
20 (16 May 2022)	7,9	67,7	94,2	3,9	
22 (31 May 2022)	36,4	203,6	285,8	0,6	
23 (07 Jun 2022)	16,6	87,1	115,1	0,4	
25 (21 Jun 2022)	0,7	4,9	6,8	0,1	
26 (28 Jun 2022)	4,3	37,5	51,0	0,3	

Coefficient (rs)	0,419033
N:	36
T Statistic:	2,6910122
DF:	34
p value:	1E-02

APPENDIX 3: Frequency of occurrence of mutations detected in the 60 near-full length SARS-CoV-2 genomes in the Vhembe and Mopani districts throughout the surveillance period (January 2021 – May 2022)

ORF1a Mutations	Freq (%)	ORF1b Mutations	Freq (%)	S-gene Mutations	Freq (%)	ORF3a Mutations	Freq (%)	E-gene Mutations	Freq (%)	M-gene Mutations	Freq (%)	ORF6 Mutations	Freq (%)	ORF7a Mutations	Freq (%)	ORF7b Mutations	Freq (%)	ORF8 Mutations	Freq (%)	N-gene Mutations	Freq (%)
T3255I	60	P314L	100	D614G	85	S26L	41.6	T19I	33.3	I82T	45	S41P	10	T120I	45	T40I	40	D119-	45	R203K	46.6
A1306S	40	G662S	45	T478K	65	T223I	15	P71L	3.3	A63T	31.6	D61L	8.3	V82A	40	L11F	6.6	F120-	45	G204R	46.6
P2046L	35	M115I	35	P681R	65	H182D	11.6			Q19E	30	S43P	1.6	E95A	1.6	E3*	1.6	L84-	10	D63G	45
P3395H	31.6	I1566V	33.3	G142D	60	S171L	6.6			G3G	15	D53Y	1.6	F63H	1.6	F13-	1.6	P85-	10	D377Y	45
V2930L	28.3	P1000	23.3	L452R	46.6	Q57H	5			G6S	8.3			E22D	1.6			F6-	6.6	R203M	43.3
T3646A	25	R1315C	15	E156-	46.6	Q185H	5			C86S	8.3			F101-	1.6			G66-	3.3	G215C	43.3
G1307S	16.6	T2163I	15	F157-	45	F105L	1.6			D3N	3.3			L102-	1.6			S67-	3.3	E31-	31.6
T3090I	15	A1918V	10	R158G	45	H247Y	1.6			K15N	1.6							K68E	3.3	R32-	31.6
K1795Q	13.3	P1427L	3.3	G339D	28.3	I32S	1.6											F120V	3.3	S33-	31.6
S135R	13.5	P2256S	3.3	K417N	28.3	G174D	1.6											I121L	3.3	P13L	31.6
K856R	13.3	F2314L	3.3	N501Y	28.3	I179T	1.6											Q27*	1.6	S413R	15
A2710T	13.3	T730I	1.6	Q954H	26.6	I123T	1.6													P151S	6.6
S2083-	13.3	T1511I	1.6	H69-	25	E102D	1.6													T205I	5
L3674-	13.3	R1226H	1.6	V70-	25	I10L	1.6													L221V	1.6
L2084I	11.6	L293F	1.6	S373P	25	S180L	1.6														
S3675F	11.6	M617I	1.6	D796Y	25	Q245K	1.6														
T842I	11.6	T1739A	1.6	P681H	25	P267L	1.6														
D2980G	10	L1220F	1.6	H655Y	25	P25S	1.6														
P3359S	10	S220N	1.6	N969K	25																
Q3729K	10	D870A	1.6	Q498R	23.3																
K14-	6.6	M1156I	1.6	Y505H	23.3																
S142-	6.6			S375F	23.3																
F143-	6.6			E484A	18.3																
S3675-	6.6			G142-	18.3																
I3758V	6.6			V143-	15																
V3718A	6.6			Y144-	15																
P2018L	5			Q493R	16.6																

V1056M	5			A67V	16.6																
ORF1a Mutations	**Freq (%)**	**ORF1b Mutations**	**Freq (%)**	**S-gene Mutations**	**Freq (%)**	**ORF3a Mutations**	**Freq (%)**	**E-gene Mutations**	**Freq (%)**	**M-gene Mutations**	**Freq (%)**	**ORF6 Mutations**	**Freq (%)**	**ORF7a Mutations**	**Freq (%)**	**ORF7b Mutations**	**Freq (%)**	**ORF8 Mutations**	**Freq (%)**	**N-gene Mutations**	**Freq (%)**
P2287S	5			A27S	16.6																
T265I	3.3			214: EPE	13.3																
T224I	3.3			S477N	13.3																
K1655N	3.3			S371L	13.3																
T3750I	3.3			T574K	13.3																
P1640L	3.3			L981F	13.3																
G82-	1.6			T547K	13.3																
H83-	1.6			N440K	13.3																
V84-	1.6			G446S	13.3																
F2780L	1.6			T376A	11.6																
G2126S	1.6			T95I	11.6																
S3384L	1.6			Y145D	11.6																
N3725S	1.6			L212I	11.6																
L1688W	1.6			N856K	11.6																
L1130F	1.6			G496S	11.6																
I2198V	1.6			N211-	11.6																
M85V	1.6			L24-	11.6																
V86F	1.6			P25-	11.6																
F3677L	1.6			P26-	11.6																
K292N	1.6			F486V	8.3																
A3209V	1.6			A67-	6.6																
C2165F	1.6			I68-	6.6																
T4217I	1.6			L141-	6.6																
L552V	1.6			D215G	5																
K564N	1.6			Q498H	5																
S2822P	1.6			T572N	5																
T951I	1.6			A701V	5																
G963C	1.6			P26Q	5																
T1637I	1.6			D80A	3.3																

A2098V	1.6			K854Q	3.3																
ORF1a Mutations	**Freq (%)**	**ORF1b Mutations**	**Freq (%)**	**S-gene Mutations**	**Freq (%)**	**ORF3a Mutations**	**Freq (%)**	**E-gene Mutations**	**Freq (%)**	**M-gene Mutations**	**Freq (%)**	**ORF6 Mutations**	**Freq (%)**	**ORF7a Mutations**	**Freq (%)**	**ORF7b Mutations**	**Freq (%)**	**ORF8 Mutations**	**Freq (%)**	**N-gene Mutations**	**Freq (%)**
N2201H	1.6			G72R	3.3																
H2357Y	1.6			L241-	3.3																
L3201F	1.6			L242-	3.3																
L3606F	1.6			A243-	3.3																
Y3685C	1.6			S151-	1.6																
T1241I	1.6			H66-	1.6																
T265S	1.6			K444-	1.6																
T770I	1.6			V445-	1.6																
				F898S	1.6																
				D1118H	1.6																
				V62I	1.6																
				H66Q	1.6																
				A372I	1.6																
				E484Q	1.6																
				A688V	1.6																
				P792Q	1.6																
				W152R	1.6																
				T20I	1.6																
				K41R	1.6																
				D88H	1.6																
				N1178D	1.6																
				S50L	1.6																
				L828	1.6																
				K854E	1.6																
				L18F	1.6																
				W64C	1.6																
				C136F	1.6																
				P499S	1.6																