Addis Ababa University

School of Biomedicine and Medical Laboratory Sciences

Department of Microbiology, Immunology and Parasitology

Molecular Characterization of SARS-CoV-2 and Evaluation of Extraction-Free RT-qPCR Detection Methods for the Development of an Alternative and Affordable Diagnostic Tools.

By

Getnet Hailu Alemayehu

December 2025

Addis Ababa, Ethiopia

Molecular Characterization of SARS-CoV-2 and Evaluation of Extraction-Free RT-qPCR Detection Methods for the Development of an Alternative and Affordable Diagnostic Tools.

By

Getnet Hailu Alemayehu

A PhD Dissertation submitted to the Addis Ababa University, College of Health Sciences, School of Biomedicine and Medical Laboratory Sciences, Department of Microbiology, Immunology and Parasitology for the fulfillment of the requirements for the Degree of Doctor of Philosophy in Tropical and Infectious Diseases

Addis Ababa University

School of Biomedicine and Medical Laboratory Sciences

Department of Microbiology, Immunology and Parasitology

December 2025

Addis Ababa, Ethiopia

Addis Ababa University

College of Health Sciences

Department of Microbiology, Immunology and Parasitology

ular Characterization of SARS-CoV-2 and Evaluation of Extraction-Free -qPCR Detection Methods for the Development of an Alternative and Affordable Diagnostic Tools.

By

Getnet Hailu Alemayehu

s submitted to the Department of Microbiology, Immunology and Parasitology; College of Sciences; Addis Ababa University in the fulfilment of the requirements for the Degree of Doctor of Philosophy in Tropical and Infectious Diseases

Approved by Examining Board	Signature	Date
Dr. Gadissa Bedada External Examiner		June 18, 2025
Dr. Wondwossen Amogne Internal Examiner		June 18, 2025
Prof. Mengistu Legesse Supervisor		June 18, 2025
Dr. Brhanu Teka Chairperson		June 18, 2025

June 18, 2025

Dedication

This PhD dissertation is dedicated to my esteemed grandmother, Wudie Shita Workie, who rooted in me a deep appreciation for learning and a strong desire to understand the complexities of our surroundings. Her unwavering support and encouragement throughout my academic journey were instrumental in my success. Furthermore, I would like to extend this dedication to all my altruistic family members. Your selfless sacrifices and unwavering commitment lay the groundwork for many dreams to come true.

Declaration

I, Getnet Hailu Alemayehu, hereby declare that I conducted the work described in this dissertation, entitled " Molecular Characterization of SARS-CoV-2 and Evaluation of Extraction-Free RT-qPCR Detection Methods for the Development of an Alternative and Affordable Diagnostic Tools", and that no portion of the dissertation has been submitted for the award of any academic degree or diploma. This dissertation has properly acknowledged all sources and materials utilized. Intellectual property rights and participants' rights to confidentiality and unrestricted withdrawal from the study were all maintained throughout the research procedure.

Submitted by

Name: Getnet Hailu Alemayehu

Signature: _____

Date: _____

Abstract

Background: SARS-CoV-2, the etiological agent of COVID-19, first emerged in Wuhan, China, and then disseminated globally at an extraordinary pace. It is classified as a single-stranded, positive-sense RNA virus within the *Coronaviridae* family and the *Betacoronavirus* genus. March 2nd, 2025, it has resulted in over 777 million confirmed cases and more than 7 million deaths worldwide. In Africa, over 12 million cases and 256,542 deaths have been reported, with Ethiopia reporting more than 500,000 cases and 7,574 deaths. The primary diagnostic method used for identifying infection is real-time reverse transcription quantitative PCR (RT-qPCR), which is recognized as the gold standard. In Ethiopia the number of SARS-CoV-2 sequences that represented the different pandemic episode and submitted to GISAID is very small compared to other African countries. In Addition, its detection was challenged by limited access to essential laboratory supplies, including RNA extraction kits, alongside the ongoing emergence of new variants. These factors highlight the continuous genomic surveillance of SARS-CoV-2 is essential for tracking of viral mutations, informs public health interventions, and aids in the development of effective diagnostics, vaccines, and therapeutic strategies; and demanding need for the development of an alternative diagnostic methods that are accurate, cost-effective, rapid, and less reliant on imported resources.

Objectives: The study aimed to characterizes circulating SARS-CoV-2 variants during the fifth wave of the pandemic in Ethiopia and identify conserved gene regions through whole genome sequencing, as well as to evaluate the diagnostic performance of an extraction-free RT-qPCR method as a simplified approach for SARS-CoV-2 detection.

Methods: 150 SARS-CoV-2 positive Ct-value of ≤30) samples were retrieved from the national reference repository of Ethiopian Public Health Institute for whole genome sequencing. These samples were primarily collected from COVID-19 suspected cases for clinical management purpose during the fifth wave of the pandemic (June to August 2022). These samples were re-tested, and 70 samples that maintained a Ct value ≤30 were selected for whole genome sequencing on the Illumina NextSeq 550 platform. Of the 70 sequenced samples, 63 yielded high-quality genomes that successfully passed all bioinformatics quality control assessments. These genomes were analyzed for variant classification, single-nucleotide polymorphism (SNP) distances,

phylogenetic relationships, and conserved gene regions identification. The analytical workflow incorporated tools such as FastQC, Cutadapt, BWA, Samtools, Pangolin, Nextclade, SNP-dists, MAFFT, and IQ-TREE. Metadata and sequence data were analyzed using STATA v17.

In addition, three hundred stored nasopharyngeal specimens (190 SARS-CoV-2 positive and 110 negative) were randomly selected from the national repository. The samples were processed using the standard RNA extraction kits (used as the gold standard) and an extraction-free method that involved dilution and heating (EFDH). In the EFDH protocol, samples were diluted at a 1:2 ratio with RNase-free water and subjected to heating at 72°C for 15 minutes. Then, after master mixing, RT-qPCR was performed using DAAN kits targeting the N and ORF1ab genes by the ABI 7500 platform. The results were stratified by viral load: Ct <20, 20–35, and >35 and diagnostic performance was analyzed by using R-Studio and STATA v17. Simultaneously, 150 SARS-CoV-2 positive samples (Ct ≤30) collected during the fifth wave of the pandemic in Ethiopia for genomic analysis.

Results: Among the 70 samples sequenced, 63 met the quality control criteria, exhibiting a genome completeness of 99.8%. The Omicron variant was the predominant strain, accounting for 97% of the samples, while the Delta variant was present in 3% of the samples. Phylogenetic analysis revealed five principal clades. On average, 69 single-nucleotide variations (SNVs) were identified per genome. Mutations were detected within the S, M, and N genes, as well as in non-structural proteins, with notable deletions observed in the S gene. Additional mutations were identified in PLpro and 3CLpro, and the transition/transversion (Ti/Tv) ratio was calculated to be 1.97. The sequences from Ethiopian samples exhibited low divergence (9.78×10^{4}) in comparison to global sequences (1.54×10^{3}). Two conserved regions within the nucleocapsid (N) gene were identified, encompassing amino acids 45–180 and 247–364. Three linear B-cell epitopes, residues 56–57, 160–180, and 255–275, were located within these regions. These epitopes displayed strong antigenic properties, low variability, and surface accessibility, indicating their potential as targets for future diagnostic applications.

The EFDH method demonstrated a sensitivity of 86% and a specificity of 100% when compared to the standard RT-qPCR assay. It accurately identified 163 out of 190 positive samples and all 110 negative samples. The average Ct value recorded for EFDH was 27.7, in contrast to 24.1 for the standard method. A strong correlation between the two methodologies was observed (R^2 =

0.96, p = 0.001). EFDH exhibited 100% sensitivity for samples with Ct values less than 20 and 98% sensitivity for those with Ct values ranging from 20 to 35. However, sensitivity significantly decreased for samples with Ct values greater than 35. The overall diagnostic accuracy of the EFDH method was determined to be 91%, with a kappa value of 0.82 ($p < 0.001$), indicating a strong agreement with the standard method for samples with high viral loads.

Conclusions: The genomic analyses revealed that the Omicron variant was the predominant strain during Ethiopia's fifth wave, suggesting ongoing community transmission and the imperative for sustained surveillance efforts. The nucleocapsid protein was validated as a stable and immunogenic diagnostic target, with identified epitope candidates facilitating future diagnostic advancements. The integration of genomic and structural data may enhance the identification of reliable diagnostic markers in the context of viral evolution. The EFDH method exhibited high sensitivity and specificity for the detection of SARS-CoV-2 in samples with elevated viral loads, demonstrating a strong correlation with the established RT-qPCR method. Nevertheless, its diminished sensitivity in cases of low viral load underscores the necessity for confirmatory testing. EFDH presents an alternative and cost-effective that suitable for resource-limited countries.

Key words: SARS-CoV-2, genetic divergence, whole genome sequencing, EFDH method, Ethiopia.

Acknowledgement

I wish to express my profound gratitude to the numerous exceptional individuals. Whose remarkable contributions have significantly shaped my doctoral studies and positively influenced my academic journey. I extend my sincere appreciation to each of them. Given the extensive network of outstanding individuals involved in my academic experience, I apologize in advance if I inadvertently omit any names in my acknowledgements. Your understanding in this matter is greatly appreciated.

First and foremost, I wish to acknowledge my deepest gratitude to Almighty God and the Virgin Mary. Their presence provided the strength to persevere, their blessings guided me, and their grace alleviated my anxieties. This PhD would not have been attainable without divine support.

I am particularly indebted to Dr. Molalegne Bitew, my PhD advisor at the Bio and Emerging Technology Institute (BETin), where he serves as Lead Researcher and Director of Health Biotechnology. I am thankful for his exceptional guidance, motivational support, enthusiasm, and steadfast belief in my potential during periods of self-doubt. His encouragement has been instrumental in my persistence. Additionally, I am immensely grateful for his insightful feedback, which prompted critical thinking and refinement of my research, ultimately leading to impactful outcomes. His mentorship, guidance, and friendship have left a lasting impression on my academic development. His collaborative spirit and infectious enthusiasm have served as constant sources of inspiration, and his willingness to share his expertise has undoubtedly enhanced my capabilities as a researcher and thinker. I extend my heartfelt thanks for all that he has provided.

I also wish to convey my deepest appreciation to my advisor, Prof. Mengistu Legesse, for his guidance throughout this remarkable journey. From the initial stages of proposal development to the final thesis write-up, Prof. Mengistu has made significant contributions. I recognize that this work would not have been possible without his mentorship. I have greatly benefited from his warm and approachable manner, as well as his valuable feedback and timely comments. I am particularly grateful for his willingness to assist me, even amidst his busy schedule. Prof. Mengistu's unique strengths have fostered an optimal environment for my academic growth. His intellectual rigor and

unwavering support have been pivotal in refining my research questions and navigating uncertainties.

I extend my heartfelt gratitude to my co-advisor, Dr. Girmay Medhin, for his steadfast support and guidance, as well as for providing a complementary perspective to my research. I felt comfortable approaching him with any inquiries or concerns, knowing he would listen patiently and offer insightful advice. His collaboration with my primary advisor has created a holistic support system that has allowed me to thrive. Our discussions were consistently stimulating, encouraging me to refine my arguments and consider alternative interpretations. I appreciate his role in challenging me to become a more well-rounded researcher.

I express my sincere gratitude to my co-advisor, Dr. Ebba Abate, for his encouragement, generosity in sharing knowledge and experience, and insightful suggestions, all of which have significantly contributed to the evolution of my scientific perspective.

I would also like to acknowledge Dr. Andargachew Mulu, Dawit Hailu, Mesfin Mengesha, and Abaysew Ayele. I owe them a deep debt of gratitude for their invaluable guidance and support during my time in the laboratory. Their technical expertise has been a constant source of knowledge, assisting me in navigating complex equipment and troubleshooting challenging experiments. I greatly appreciate their patience in teaching new techniques and their willingness to assist beyond regular hours. Thank you for being the foundation of my scientific endeavours.
I extend my sincere gratitude to Dr. Fifihatsion Tadesse and Dr. Kidist Zealiyas for their exceptional support throughout this academic journey. Your assistance has significantly facilitated the progression of my scholarly pursuits.

I would also like to acknowledge the Ethiopian Public Health Institute for their generous sponsorship and the Aklilu Lemma Institute of Health Research at Addis Ababa University for the invaluable opportunities provided to enhance my educational development. Furthermore, I wish to express my heartfelt appreciation to my esteemed colleagues, Melkeneh, Andargachew, Nebiyou, Habtamu, Achamyeleh, Addisu, Wondimeneh, Abdilfetah, Awad, Gonfa, Dereje, Asdesach, and Dr. Minilik, for their consistent encouragement and unwavering support throughout my academic endeavors.

Additionally, I would like to express my gratitude to my beloved parents, Zenebech Tefera and Hailu Alemayehu. Their prayers have served as a continuous source of strength and inspiration throughout my life, particularly during this critical period.

I am also thankful for the unwavering love and support of my sisters, Abinet, Migbaresenay, Lucy, Meskerem, and Hilina, and my brothers, Feleke, Wossen, and Abel, which have kept me motivated and confident. Lastly, I am profoundly grateful to my cherished wife, Enat Bilaf, and my children, Nathan, Soliyana, and Absalat. Enat, your boundless love, understanding, and encouragement, along with your willingness to assume our family responsibilities, have been invaluable during this journey.

Table of Contents

List of Figures

Lists of Tables

Abbreviations

ACE2	Angiotensin-Converting Enzyme 2
AHRI	Armaure Hansen Research Institute
AIDS	Acquired Immunodeficiency Syndrome
ALIPB	Aklilu Lemma Institute of Pathobiology
ARDS	Acute Respiratory Distress Syndrome
AT	Alveolar Type
BIC	Bayesian Information Criterion
BIP	Backwards Inner Primer
BLOSUM	Blocks Substitution Matrix
BSL	Bulged stem-loop
BSL-3	Biosafety Level 3
CDC	Center for Disease Control and Prevention
cDNA	Complementary Deoxyribonucleic Acid
CFR	Case Fatality Rate
CI	Confidence Interval
CLIA	Chemiluminescent Immunoassay
COVID-19	Coronavirus Diseases 19
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CT	Computed Tomography
Ct	Cycle Threshold
CXR	Chest X-Ray
DECT	Dual-energy CT
DMV	Double-Membrane Vesicles
ds	Double Strand
EFDH	Extraction Free Dilution and Heating
ELISA	Enzyme-Linked Immunosorbent Assay
EM	Electron Microscope
EPHI	Ethiopian Public Health Institute
ER	Endoplasmic Reticulum
ERGIC	Endoplasmic Reticulum-Golgi Intermediate Compartment

FET	Field-Effect Transistor
FIA	Fluorescent Immunoassay
FIP	Forward Inner Primer
FSM	First Stand Mix
GISAID	Global Initiative on Sharing All Influenza Data
HIV	Human Immunodeficiency Virus
HT1	Hybridization Buffer
IC	Internal Control
ICU	Intensive Care Unit
IDs	Identifiers
Ig	Immunoglobulin
IL	Interleukin
INFs	Interferons
IRB	Institutional Review Board
ISGs	Interferon-Stimulated Genes
ITB	Illumina Tune Beads
LFA	Lateral Flow Assay
LOD	Limit of Detection
LUS	Lung Ultrasound
MAFFT	Multiple Alignment using Fast Fourier Transform
MAVS	Mitochondrial Antiviral Signaling Protein
MERS	Middle East respiratory syndrome coronavirus
MHV	Murine Hepatitis Virus
MIS-C	Multisystem Inflammatory Syndrome in Children
MRI	Magnetic Resonance Imaging
MSA	Multiple Sequence Alignment
NBS	Nanoparticle-Based Biosensor
NCBI	National Center for Biotechnology Information
NK	Natural Killer cell
NPIs	Non-Pharmaceutical Interventions
NPV	Negative Predictive Value
NSPs	Nonstructural Proteins

ORF	Open Reading Frame
PAMP	Pathogen-Associated Molecular Pattern
PANGO	Phylogenetic Assignment of Named Global Outbreak Lineages
PCR	Polymerase Chain Reaction
PET-CT	Positron Emission Tomography-Computed Tomography
PLL	Phylogenetic Likelihood Library
PPE	Personnel Protective Equipment
PPV	Positive Predictive Value
PRRS	Pattern Recognition Receptors
RAT	Rapid Antigen Test
RBD	Receptor Binding Domain
RdRP	RNA-dependent RNA Polymerase
RLRs	RIG-I-like receptors
RNA	Ribonucleic Acid
RTC	Replication-Transcription Complex
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
Rt-qPCR	Reverse transcriptase quantitative Polymerase Chain Reaction
RVT	Reverse Transcriptase
SARS-CoV-2	Sever Acute Respiratory Syndrome Corona Virus-2
SARS-CoVs	Severe Acute Respiratory Syndrome Coronavirus
SEM	Standard Error of the Mean
SNNP	Southern Nation Nationality and People
SNVs	Single Nucleotide variations
SPECT	Single-Photon Emission Computed Tomography
ss	Single Strand
TLRs	Toll-like receptors
TMPRSS2	Transmembrane Serine Protease 2
TRS-L	Transcription regulatory sequence leader
TTS	Thrombosis with Thrombocytopenia Syndrome
UND	Undetermined
UNESCO	United nations Educational, Scientific and Cultural Organization
UTR	Untranslated region

VLP	Virus-Like Particles
VOC	Variants of Concern
VoC	Variants of Concern
VoI	Variants of Interest
VTM	Viral Transport Medium
VUMs	Variants under Monitors
WHO	World Health Organization

Lists of Appendices

CHAPTER 1: INTRODUCTION

1.1.Background

The outbreak of COVID-19 emerged in December 2019 in Wuhan City, located in Hubei Province, Chin. The disease is caused by the novel coronavirus SARS-CoV-2 and quickly spread to other regions of China and around the world, leading to a global pandemic. During the time, Fifty-five percent of the infected cases before 1 January 2020 were linked to the Huanan Seafood Wholesale Market. However, the first human-to-human case of SARS-CoV-2 infection was reported on 1 December 2019, who didn't have any exposure to this market (Amodio et al., 2020). The disease was spread from China to other countries via international travelers. Therefore, the first case was confirmed in Thailand on 13 January 2020 (WHO, 2020). Within just one week, the number of confirmed cases surged from 2,798 to 17,391 (from January 27 to February 3), and the number of affected countries doubled from 12 to 24. Given the rapid increase in both cases and affected countries, the WHO declared SARS-CoV-2 a pandemic on March 11, 2020 (WHO Media Brief, 2020).

Globally, as of March 2nd, 2025, there were 777,594,331 confirmed cases of COVID-19, and 7,089,989 deaths (WHO, 2025). The pandemic affects the world with a major outbreak in the United States of America (USA) (103,804,263 cases), China (99.4 million cases) followed by India (45 million cases) and Russia (22,086,064 cases) (Our World in Data, 2025; CDC, 2025). The number of deaths was high in the USA (around 1.2 million), in Brazil (702 thousand) and India (543 thousand) (Our World in Data, 2025; JHU, 2025).

Africa confirmed its first case of COVID-19 in Egypt on the 14th of February 2020. In sub-Saharan Africa, Nigeria was the first country to confirm a case on February 27, 2020 (Lone and Ahmad, A., 2020). As of January 2, 2023, Africa had reported 12,216,748 confirmed COVID-19 cases and 256,542 deaths. The southern region of the continent reported the highest number of cases, with approximately 5.5 million, followed by Northern Africa with 3.8 million cases and Eastern Africa with 1.3 million cases(Africa CDC - COVID-19 Daily Updates, 2020).

Nonetheless, Africa's underlying health issues and weak healthcare systems predicted that the continent would be highly susceptible to the COVID-19 crisis. This was because the continent has been known with a high burden of infectious diseases, hosting about 70% of individuals living with HIV, 25% of

deaths from TB, and 95% of cases of malaria in the world (Kariuki et al., 2022). In addition, most countries in Africa experience weak healthcare systems that include scarce resources, poor infrastructure, and a shortage of health personnel (Dotsey, 2024). The pandemic disorganized international supply chains, leading to shortages of necessary medicine and medical supplies throughout the continent (Dotsey, 2024).

Despite all these challenges, African countries have experienced a comparatively lower COVID-19 burden than other continents such as Europe, North America, Asia and Latin America. Several factors have been proposed to explain this trend, including the continent's younger population and possible cross-immunity from previous exposures to other infectious pathogens (Lawal, 2021; Wamai et al., 2021). However, the true extent of the pandemic's impact in Africa may also be underestimated due to underreporting and limited surveillance systems (Lawal, 2021).

Currently, as of May 2025, COVID-19 continues to pose a significant global health challenge; however, it is now managed as an endemic disease rather than as a public health emergency. The nations have shifted their focus from immediate crisis response to the development of long-term, sustainable disease control strategies (WHO, 2023). This model includes the systematic integration of COVID-19 monitoring, vaccination, and care into national healthcare systems. Updated vaccines, specifically formulated to address emerging variants such as LP.8.1 and LB.1, have been deployed for the 2024-2025 season. These vaccination initiatives primarily target high-risk populations, including older adults and immunocompromised individuals, intending to reduce severe illness and mortality rates (Omotoso et al., 2023).

Global surveillance efforts continue in monitoring the evolution of SARS-CoV-2 variants and the seasonal transmission patterns (WHO, 2023). While the acute phase of the pandemic has reduced, COVID-19 persists as a significant public health challenge that necessitates ongoing vigilance and strategic investment (Wafaa EI-Sadr, 2025).

1.1.1. COVID-19 in the Ethiopian context

Ethiopia confirmed its first case of COVID-19 on 13th March 2020, two days later, the WHO declared a pandemic of the disease (Dikale, 2021). As of March 2nd 2025, the country tested 5,420,453 suspects, of whom 501,304 (9.2%) cases had been confirmed positive and of these, 7,574 died with a case fatality

rate (CFR = 1.5%) (WHO COVID-19 Dashboard, 2025). The first confirmed case in Ethiopia was a 48-year-old Japanese man who had travelled from Burkina Faso to Ethiopia. The second report mentioned three additional cases, two Japanese and one Ethiopian, who had been in contact with the index case. In response, the national government declared a five-month state of emergency on April 8, 2020, while allowing economic activities to continue during the public health crisis (WHO Regional Office of Africa, 2024).

To curve the virus spread, the Federal Government of Ethiopia, along with the National and Regional States and City Administrations, implemented a series of proactive measures. On March 16, 2020, the government restricted public gatherings, including gatherings for religious practices, sporting events, and concerts; ordered school closures; and instructed high-risk civil servants to work from home. While aiming to maintain daily activities, the government introduced containment protocols. Public transportation, including taxis and mass transit, was required to operate at half capacity and adhere to revised working hours. Additionally, all the Regional States and City Administrations imposed travel restrictions. On March 23, 2020, Ethiopia suspended flights to 30 countries affected by the virus, expanding the ban to over 80 countries by March 29. The same day, all land borders were closed, and security forces were deployed to enforce the restrictions. A total of 20,402 prisoners were granted pardons to prevent the spread of the disease in prison. Furthermore, the national parliamentary and presidential elections, initially scheduled for August 29, 2020, were officially postponed (Mohammed et al., 2020).

Ethiopia faced several significant challenges during the COVID-19 pandemic, particularly in healthcare, the economy, and education. The healthcare system, already overstretched due to pre-existing health challenges, struggled to handle the surge in COVID-19 cases (Abagero et al., 2022). The country experienced a shortage of essential medical supplies, including ventilators, protective equipment, and testing kits, which hindered effective response efforts. Additionally, healthcare workers were at high risk due to insufficient personal protective equipment (PPE), exacerbating the situation (Abagero et al., 2022; Grant, K. et al., 2021). The Ethiopian Ministry of Health worked to address this by scaling up local production of medical supplies, but there were still gaps that slowed down response times (Abagero et al., 2022).

Furthermore, Ethiopia's pre-existing health challenges, such as high rates of HIV/AIDS, tuberculosis, and malaria, compounded the risks associated with COVID-19 (Chanda-Kapata et al., 2022). Despite these challenges, the country made strides in responding to the crisis through community-based health interventions and innovative solutions such as mobile health services to monitor and treat patients remotely (Bisrat et al., 2023).

Ethiopia has shifted its efforts from crisis response to the implementation of long-term, sustainable disease control strategies, aligning its approach with that of other nations.

The rapid spread of SARS-CoV-2 highlights the critical importance of diagnostic testing, which is contingent upon the types of tests available, the resources required for testing, and the time necessary to obtain results. Timely identification of suspected cases is imperative for the effective allocation of protective measures and the prevention of further community transmission (Webb et al., 2004).
Consequently, the rapid and accurate detection of COVID-19 cases is essential for controlling the outbreak of the virus within both community and healthcare settings. Research consistently indicates that reverse transcription quantitative polymerase chain reaction (RT-qPCR) serves as the gold standard for COVID-19 diagnosis due to its ability to specifically detect SARS-CoV-2. The development of these molecular diagnostic technologies relies on a comprehensive understanding of the virus's proteomic and genomic structure, as well as the alterations in protein and gene expression that occur during and after infection. Nevertheless, current diagnostic tests face several challenges, including gene mutations, labor-intensive processes, extended testing times, and delays caused by shortages of extraction kits (Younes et al., 2020). Although RT-qPCR provides relatively rapid results (averaging 3-4 hours), there is an urgent need for cost-effective and efficient diagnostic methodologies (Younes et al., 2020).

Therefore, this study aimed to characterizes circulating SARS-CoV-2 variants during the fifth wave of the pandemic in Ethiopia and identify conserved gene regions through whole genome sequencing, towards the development of alternative and affordable diagnostic tools. The study also aimed to evaluate the diagnostic performance of an extraction-free RT-qPCR method as a simplified approach for SARS-CoV-2 detection.

1.2.Rationale of the study

The emergence of new SARS-CoV-2 variants remains a pressing global concern. Recent data reveal that genome sequencing of the virus has been conducted in only a limited number of African countries. Nigeria was the first nation to perform genome sequencing, followed by South Africa, Senegal, and Tunisia. To date, Ethiopia has sequenced and submitted only 824 SARS-CoV-2 samples to GISAID, a figure that is notably low in comparison to other African nations such as South Africa (which has sequenced over 56,886 samples), Kenya (over 13,770 samples), Egypt (approximately 5,146 samples), and Tunisia (about 2,915 samples). Analyzing the genomic sequences of these viruses is crucial for the scientific community, as it enhances our understanding of the virus's virulence and transmissibility and aids in the development of diagnostic and therapeutic strategies. This dissertation, alongside the GISAID/NCBI submissions we have made, provides valuable insights into the status of SARS-CoV-2 variants in Ethiopia during the fifth wave of the pandemic.

The polymerase chain reaction (PCR) is widely regarded as the gold standard for detecting SARS-CoV-2 due to its high sensitivity and specificity. However, this technique depends on RNA extraction kits to isolate viral RNA from biological samples, a critical step that can significantly impact testing efficiency and accessibility. The reliance on these kits not only prolongs the testing process but also depletes essential reagents, thereby creating logistical challenges in laboratory settings.

During the COVID-19 pandemic, the unprecedented demand for SARS-CoV-2 testing resulted in a global shortage of RNA extraction kits, severely undermining diagnostic capabilities. This shortage was particularly detrimental in resource-limited settings, such as many African countries, where access to molecular diagnostic supplies is often constrained. Ethiopia faced considerable challenges in procuring sufficient extraction kits, leading to delays in testing and complicating timely outbreak responses.

Given these limitations, it is important to explore and validate alternative diagnostic protocols that can diminish dependence on conventional RNA extraction kits while maintaining comparable sensitivity and specificity. Targeting conserved gene regions is a crucial strategy in developing rapid and reliable diagnostic tools. These gene regions are stable genomic regions that exhibit minimal variation across different strains or variants of a pathogen, making them ideal targets for monoclonal antibody development. By identifying these gene regions, diagnostic tools can remain effective even when

genetic mutations occur, thereby enhancing their long-term reliability and accuracy. However, currently, there are few antibody-based diagnostic kits approved for large-scale screening in Africa. Consequently, there is a need to develop diagnostic tools that are sensitive, accurate, rapid, and cost-effective, facilitating efficient screening of infected individuals and ensuring timely identification and treatment.

Therefore, this study aimed to explore the genetic variants of SARS-CoV-2 circulated during the fifth wave of the pandemic; to identify conserved gene regions of the virus and to evaluate extraction-free RT-qPCR detection methods towards the development of an alternative and affordable diagnostic tool. The study objectives address these areas to fill existing gaps, and the results provide baseline information for further research.

1.3.Significance of the study

Understanding SARS-CoV-2 variants is crucial for developing public health strategies and maintaining effective control measures. As the virus mutates, changes in its genetic material can affect its ability to spread, cause disease, and evade immunity from previous infections or vaccines. Scientists closely monitor these changes to ensure that vaccines and treatments remain effective, update protective measures as necessary, and maintain accurate diagnostic tools. This surveillance enables rapid responses to new outbreaks and strengthens global health security.

Alternative diagnostic methods that bypass traditional RNA extraction steps offer several advantages. These simplified approaches reduce costs, streamline supply chains, and accelerate testing times. By eliminating the need for expensive RNA extraction kits, which have often been in short supply during the pandemic, these methods improve accessibility, particularly in resource-limited settings. Faster processing times enable quicker outbreak responses and timely isolation of infected individuals.

The identification of specific SARS-CoV-2 target genes is especially important for developing monoclonal antibody-based diagnostic tools, particularly in regions lacking approved antibody tests. Identifying stable genetic sequences across variants facilitates the creation of reliable diagnostic tools that function regardless of the viral strain. In East Africa, where complex PCR testing is limited due to high costs and resource constraints, effective antibody tests could significantly enhance diagnostic

capacity. These simpler, potentially more affordable tests require minimal laboratory infrastructure, making them ideal for widespread use and improved surveillance.

1.4.Research questions

Based on the review of literatures, the following research questions were proposed:

i. What are the predominant genetic variants of SARS-CoV-2 circulated during the last phases of the pandemic in different regions of Ethiopia?
ii. Which SARS-CoV-2 gene regions remain conserved across circulating variants and are suitable targets for monoclonal antibody-based diagnostics?
iii. What is the diagnostic performance (sensitivity, specificity, and accuracy) of RNA extraction free rt-PCR testing compared to standard RNA extraction-based protocols for SARS-CoV-2 detection?

1.5. Research Objectives

1.5.1. General objective

The general objective of this dissertation was to characterizes circulating SARS-CoV-2 variants during the fifth wave of the pandemic in Ethiopia and identify conserved gene regions through whole genome sequencing towards the development of alternative and affordable diagnostic tools. The study also aimed to evaluate the diagnostic performance of an extraction-free RT-qPCR method as a simplified approach for SARS-CoV-2 detection.

1.5.2. Specific objectives

Objective 1: To identify the genetic variants of SARS-CoV-2 circulating in different parts of Ethiopia during the fifth wave of the pandemic.

Objective 2: To identify the conserved gene regions of SARS-CoV-2 through comparative genomic analysis of the sequenced isolates toward the development of an affordable monoclonal antibody-based diagnostic tools.

Objective 3: To evaluate the diagnostic performance of the extraction-free RT-qPCR method in detecting SARS-CoV-2.

CHAPTER 2: LITERATURE REVIEW

2.1. Coronaviruses

Coronaviruses (Latin, corona: a crown) are large, enveloped, positive-sense single-stranded RNA viruses, characterized by their spike glycoproteins (S proteins) that project from the viral envelope, giving them a crown-like appearance under electron microscopy (Commentary: Five Years after COVID, 2025). They can infect a wide range of hosts, including avian, wild, domestic mammalian species, and humans. Coronaviruses are well known for their ability to mutate rapidly, alter tissue tropism, cross the species barrier, and adapt to different epidemiological situations (Amoutzias et al., 2022).

All coronaviruses share similarities in the organization and expression of their genome, in which 16 nonstructural proteins (nsp1 through nsp16), encoded by open reading frame (ORF) 1a/b at the 5′ end, are followed by the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N), which are encoded by other ORFs at the 3′ end. CoVs are separated into four genera based on phylogeny: alpha-CoV (group 1), beta-CoV (group 2), gamma-CoV (group 3) and delta-CoV (group 4) (Durham et al., 1979).

They are one of the largest groups of viruses that belong to the order *Nidovirales*, suborder *Coronavirineae*, and family *Coronaviridae.* The family *Coronaviridae* is classified into two subfamilies, namely, *Letovirinae,* which contains *Alphaletovirus* genus, and *Orthocoronavirina* is further classified based on phylogenetic analysis and genome structure into four genera: *Alphacoronavirus* (*αCoV*), *Betacoronavirus* (*βCoV*), *Gammacoronavirus* (*γCoV*), and *Deltacoronavirus* (*δCoV*), with a prominent distribution of species across these genera: 17 species in *Alphacoronavirus,* 12 in *Betacoronavirus,* 2 in *Gammacoronavirus* and 7 in *Deltacoronavirus* (Temmam et al., 2022).

Alphacoronaviruses (α-CoV) are primarily infect mammals, including humans, bats, pigs, and domestic animals. Outstanding members of Alphacoronaviruses are: Human coronaviruses HCoV-229E and HCoV-NL63: Cause mild upper respiratory tract infections, commonly known as the common cold; Porcine epidemic diarrhea virus (PEDV): A highly pathogenic swine virus leading to severe diarrhea and high mortality in piglets and Feline coronavirus (FCoV): Can mutate into feline infectious peritonitis virus (FIPV), a fatal disease in cats. Alphacoronaviruses generally have smaller genomes

(~27 kb) and show lower recombination rates compared to other genera. Bats are considered their primary reservoirs, with repeated spillover events to other mammalian hosts over time (Woo et al., 2012).

Betacoronaviruses (β-CoV) also primarily infect mammals, including humans, bats, camels, and rodents. The Subgenera and Key Viruses are: Sarbecovirus: Includes SARS-CoV-1 (2002–2004 outbreak) and SARS-CoV-2 (COVID-19 pandemic), both of which originated in bats. Merbecovirus: Includes MERS-CoV, which jumps from camels to humans and has a high fatality rate (~35%). Embecovirus: Includes HCoV-OC43 and HCoV-HKU1, which cause mild respiratory infections, and Murine Hepatitis Virus (MHV), a model for coronavirus research and Nobecovirus and Hibecovirus: Primarily circulate in bats with limited zoonotic potential. Betacoronaviruses have relatively large genomes (~30 kb) and frequently undergo recombination, enabling them to adapt to new hosts. SARS-CoV-2, for example, likely emerged through recombination between bat RaTG13-like coronaviruses and pangolin coronaviruses (Forster et al., 2020). The high adaptability of *betacoronaviruses* makes them the primary concern for future zoonotic spillover events.

Gammacoronaviruses (γ-CoV) are predominantly infect birds, with some reports of infections in marine mammals. Some famous members are: Avian infectious bronchitis virus (IBV): A major poultry pathogen causing respiratory and renal disease, and Beluga whale coronavirus SW1: One of the few *gammacoronaviruses* identified in mammals (Mordecai and Hewson, 2020). They exhibit strong host specificity but can occasionally cross species barriers, as observed with SW1. *Deltacoronaviruses (δ-CoV*) are initially discovered in birds but are now known to infect some mammals, including pigs and wild animals. Some of the Key Viruses: Porcine deltacoronavirus (PDCoV): Emerged in 2012, causing diarrhoea in pigs and raising concerns about potential zoonotic risks and Asian leopard cat coronavirus: Demonstrates the adaptability of *deltacoronaviruses* to new hosts. They have unique spike proteins that enable broad host tropism, increasing their potential for interspecies transmission (Woo et al., 2012).

Among the four genera, *betacoronaviruses* pose the greatest pandemic threat due to their frequent zoonotic spillover and high pathogenicity in humans. The emergence of SARS, MERS, and COVID-19 underscores the importance of surveillance and preparedness. Proactive monitoring of bat and wildlife coronaviruses is essential to prevent future pandemics by identifying high-risk strains before they spill over into human populations (Anthony et al., 2017).

Six human coronaviruses have been reported since the 1960s; four of them (OC43, 229E, NL63, and HKU1) cause mild illness similar to the common cold and gastrointestinal tract infection. The other two, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), have raised significant public health concerns due to their zoonotic emergence and crossing of the species barrier, causing high pathogenicity and mortality in humans (Wu, A. et al., 2020). SARS-CoV and MERS-CoV are believed to have been transmitted from the main host (bats) to intermediate hosts, such as palm civets and romedary camels, respectively, before eventually infecting humans. Both viruses are known for their high pathogenicity (Cui et al., 2019; Drosten et al., 2014).

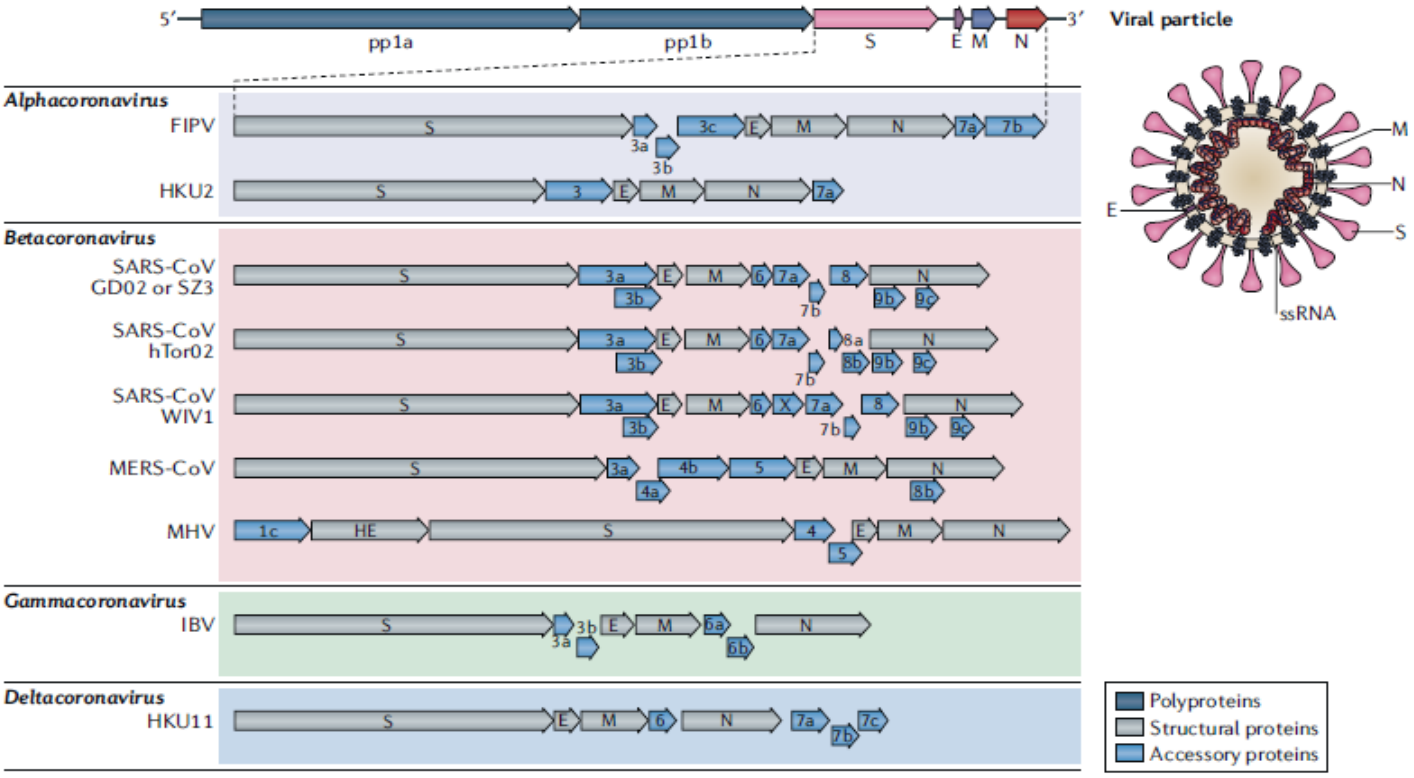


Figure 1.The genomes, genes and proteins of different coronaviruses.
Source from (Cui et al., 2019)

2.2. Emergence of SARS-CoV-2

The emergence of SARS-CoV-2 has undergone extensive scientific examination and argument. While its exact origins remain uncertain, prevailing evidence suggests that the virus likely originated from

natural zoonotic transmission, meaning it was transferred from animals to humans (Andersen et al., 2020). Although alternative explanations, such as the possibility of an accidental laboratory release, have been proposed, they currently lack definitive scientific support (Pekar et al., 2022).

Natural zoonotic origin hypothesis: The leading theory asserts that SARS-CoV-2 originated in wild animal populations. Phylogenetic analyses suggest that the virus most likely originated in wild animals, particularly bats, and may have been transmitted to humans via an intermediate host (Zhou et al., 2020). SARS-CoV-2 belongs to the genus Betacoronavirus, subgenus Sarbecovirus, and shares significant genetic similarity with bat coronaviruses, especially RaTG13 (96.3%), ZXC21 (89%), and SARS-CoV (82%), supporting a bat-related origin (Zhou et al., 2020) (Paraskevis et al., 2020). However, direct bat-to-human transmission is considered unlikely. Instead, an intermediate species likely facilitated viral adaptation before human spillover (Ge et al., 2013).

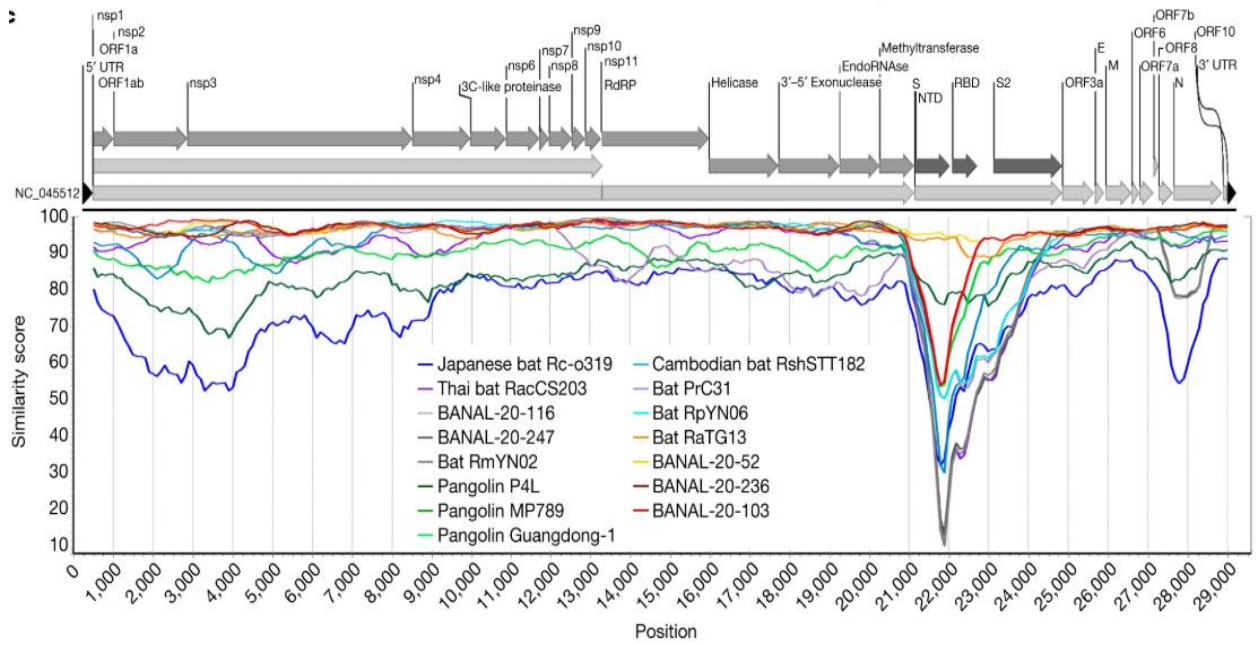


Figure 2. Genomic Similarity between the representative bat and pangolin with the SARS-CoV-2 human prototype strain (NC_045512, Wuhan-Hu-1) used as a reference.
Source: (Temmam et al., 2022).

Pangolins have been anticipated as a potential intermediate host, as some carry coronaviruses with spike proteins similar to those of SARS-CoV-2 (Lam et al., 2020). Despite this, no conclusive evidence has identified the exact intermediary species. Other animals like raccoon dogs, civet cats and minks, all of which are susceptible to coronaviruses and are commonly traded in live animal markets (Worobey

et al., 2022). The Huanan Seafood market in Wuhan, China, was initially suspected as the outbreak's origin due to early cases linked to it and positive SARS-CoV-2 environmental samples from animal stalls (Pekar et al., 2022; Worobey et al., 2022). This supports the likelihood of zoonotic transmission at the market. However, some early cases had no direct connection to the market, suggesting the virus may have been circulating prior to its detection there (Worobey et al., 2022). This has prompted ongoing investigations into whether the market was the virus's true origin or merely an amplification site where human-to-human transmission intensified.

Laboratory origin hypothesis: An alternative theory proposes that SARS-CoV-2 may have accidentally escaped from a research laboratory at Wuhan Institute of Virology (WIV), which has been working on bat coronaviruses. Followers of this theory advocate that the virus could have escaped due to a laboratory incident, possibly involving accidental exposure of a researcher (Ridley and Chan, A., 2021). Concerns have been raised regarding laboratory biosafety protocols and whether gain-of-function research, experiments that enhance a virus's ability to infect hosts, might have contributed to the emergence of SARS-CoV-2 (Ridley and Chan, A., 2021).

However, no credible scientific evidence has been found to support the lab-leak hypothesis. Several genetic studies have shown that SARS-CoV-2 lacks the molecular signatures typical of engineered viruses, indicating that it evolved naturally (Andersen et al., 2020). Additionally, past outbreaks of SARS-CoV and MERS-CoV have confirmed that coronaviruses have a history of natural zoonotic spillover events, making a natural origin more likely than an artificial one (Guan et al., 2003; Zaki et al., 2012).

2.3. Genome structure of SARS-CoV-2

SARS-CoV-2 is an enveloped, single-stranded, positive-sense RNA virus classified under the beta-coronavirus genus. Its genome consists of a non-segmented, large RNA molecule with an approximate length of 29,900 nucleotides. The genome is packed by viral nucleocapsid (N) proteins as a large ribonucleoprotein (RNP) complex and enclosed by an envelope membrane with lipids and viral proteins S (surface or spike), M (membrane) and E (envelope). The SARS-CoV-2 genome is unstable at elevated temperature because of highly enriched A+U content (62%) and reduced G+C content (38%), which is comparable to the hCoV-OC43 genome (63% A+U and 37% G+C) and the hCoV-NL63 genome (66% A+U and 34% G+C) (Brant et al., 2021). This genomic RNA is characterized by the presence of a 5′-

cap structure and a 3′-polyadenylated (poly-A) tail, which play essential roles in translation and genome stability (Sola et al., 2015).

The genome of SARS-CoV-2 is organized in a way that facilitates efficient viral replication and protein synthesis. At the 5' terminus, the genome holds a modified 7-methylguanosine (m7GpppA) cap structure, which assists in protecting the viral RNA from degradation while also facilitating translation initiation (Masters, 2006). This cap structure is followed by a 265-nucleotide 5' untranslated region (UTR), which contains numerous crucial regulatory elements essential for viral gene expression. Among these elements, the leader sequence (72 nucleotides) and the transcription regulatory sequence leader (TRS-L, ACGAAC) play significant roles in viral RNA synthesis and genome stability (Sola et al., 2015).

The SARS-CoV-2 genome encodes a total of 29 proteins, including 25 putative non-structural and accessory proteins, along with four essential structural proteins (Yoshimoto, 2020). Non-structural proteins (NSPs) are critical for viral RNA replication and immune evasion, ensuring efficient viral propagation within host cells. The majority of the genome is dedicated to encoding the replicase gene complex, which consists of two large overlapping open reading frames, ORF1a and ORF1b. These regions account for approximately 70% of the genome and are responsible for encoding the viral non-structural proteins(NSPs) (Sola et al., 2015). A unique ribosomal frameshift mechanism at a specific slippery sequence (-UUUAAAC), regulated by a three-stem pseudoknot structure, ensures the accurate translation of ORF1b. This frameshift occurs with remarkable precision, leading to the production of polyproteins that are subsequently cleaved into 16 distinct NSPs. Each NSPs plays a crucial role in viral replication and immune system evasion (Brierley et al., 2009).

Following the replicase complex, the genome contains structural and accessory genes that contribute to viral assembly, infection, and pathogenicity. The structural genes include those encoding the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, all of which are strategically positioned near the 3' end of the genome (Malone et al., 2022).

The spike protein is essential for host cell recognition and entry, mediating attachment to the angiotensin-converting enzyme 2 (ACE2) receptor. The membrane protein works in coordination with the spike, envelope, and nucleocapsid proteins to facilitate viral assembly and budding. Meanwhile, the

envelope protein forms ion channels in the viral membrane, which influence virus assembly and pathogenicity (Jami et al., 2023). In addition to structural proteins, SARS-CoV-2 encodes several accessory proteins, including ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF9b. These proteins play various regulatory functions that contribute to immune evasion and viral pathogenicity (Zandi et al., 2022). Though their exact roles are still being studied, they are known to interfere with host immune responses, modulate apoptosis, and contribute to viral persistence in infected individuals (Jami et al., 2023).

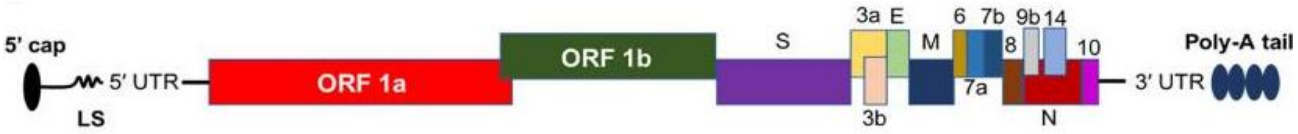


Figure 3. Genomic structure of SARS-CoV-2

This is a representation of 5' capped mRNA that has a leader sequence (LS), poly-A tail at the 3' end, and 5' and 3' UTR. It consists of ORF1a, ORF1b, Spike (S), ORF3a, Envelope (E), Membrane (M), ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid (N), and ORF10. Source:(Naqvi et al., 2020).

The genome ends with a 337-nucleotide 3' UTR, which is significantly longer than those found in related coronaviruses, such as hCoV-OC43 (286 nucleotides) and hCoV-NL63 (287 nucleotides) (Brant et al., 2021). This extended UTR covers specialized elements, including the bulged stem-loop (BSL) and a pseudoknot structure, both of which are crucial for viral RNA synthesis and transcription regulation (Malone et al., 2022). A polyadenylated (poly-A) tail, ranging from 30 to 60 nucleotides, ensures the stability and integrity of the entire genome. This poly-A tail plays a critical role in protecting the viral RNA from degradation by cellular exonuclease, thereby supporting efficient replication and translation (Passmore and Coller, 2022).In conclusion, each component, from the 5' cap to the poly-A tail, is advantageously designed to optimize viral survival, replication, and immune evasion (Wu et al., 2022).

2.4. Variants of SARS-CoV-2

SARS-CoV-2 has undergone continuous evolution since its emergence, leading to the development of multiple variants with distinct characteristics. The high rate of viral replication, dissemination, and prevalence is associated with the emergence of new viral variants. The mutagenesis events, particularly in the S1 subunit of the spike protein, can enhance its pathogenicity, infectivity, and dissemination (Flores-Vega et al., 2022).

The World Health Organization (WHO) recommends using the Greek alphabet for viral classification (n.d.). Other widely used nomenclature systems include those proposed by GISAID (Re3data.Org, 2012), PANGO (O'Toole et al., 2021), and Nextstrain (Hadfield et al., 2018).

The WHO has categorized these variants based on function of their genome mutations, their properties of spreading between susceptible hosts and the disease severity produced, and by the evasion of the immune response elicited by current available vaccines and by the treatments with convalescent plasma or by the use of therapeutic monoclonal antibodies (their impact on public health) into Variants of Concern (VOCs), Variants of Interest (VOIs) and variants under monitoring (VUMs) (WHO, 2023). Variants such as Alpha, Beta, Gamma, Delta, and Omicron are classified as VOCs, whereas Lambda and Mu are designated as VOIs. Additionally, the variants AZ.5 (previously tracked under the parent lineage B.1.1.318), C.1.2, B.1.617 (formerly the VOI Kappa: B.1.617.1), B.1.630, and B.1.640 are monitored as variants under monitoring (VUMs) (WHO, 2023) (Figure 4).

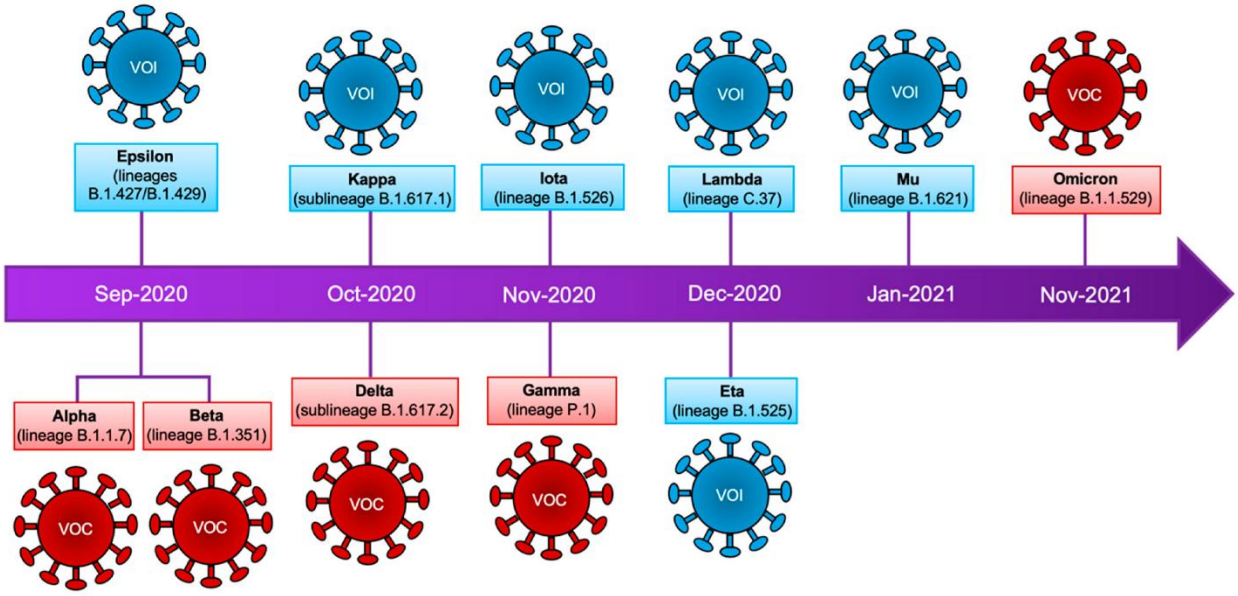


Figure 4. SARS-CoV-2 variants. The timeline summarizes the emergence of SARS-CoV-2 variants.

This Classification was according to the WHO, is shown as well as the lineages of the mutagenic profile. VOC: variants of concern, VOI: variants of interest (Flores-Vega et al., 2022).

The PANGO nomenclature is used to identify currently circulating lineages (O'Toole et al., 2021). This classification, suggested by (Rambaut et al., 2020), employs lineage names that begin with either the letter A or B. Lineages starting with A are directly related to the Wuhan/WH04/2020 variant, while

those starting with B are associated with the Wuhan-Hu-1 variant. New SARS-CoV-2 lineages derived from lineage A or B are assigned numerical values (e.g., lineage A.1 or B.2). Specific lineage designations must adhere to the following criteria: (1) Each successor lineage must demonstrate phylogenetic evidence of emerging from an ancestral lineage of a different geographical population.

To establish this phylogenetic evidence, the lineage must meet these conditions: (a) share one or more nucleotide differences from the ancestral lineage, (b) comprise at least five genomes with over 95% of the genome sequenced, (c) exhibit at least one shared nucleotide change among the genomes of the new lineage, and (d) achieve a bootstrap value greater than 70% for the node representing the new lineage (Rambaut et al., 2020). (2) GISAID classifies new SARS-CoV-2 variants into clades, which are defined by the statistical distribution of viral genome distances within phylogenetic clusters (Han et al., 2019).

This classification involves merging smaller lineages into larger clades. Consequently, viral variants are grouped into eight high-level phylogenetic categories, starting with an early split into S and L, followed by the evolution of L into V and G, and later the evolution of G into GH, GR, and GV, with GR evolving into GRY more recently (Re3data.Org, 2012; Shu and McCauley, 2017).

Nextstrain further categorizes SARS-CoV-2 into 14 major clades: 19A, 19B, and 20A-20L. A clade is established when a new variant reaches a global frequency of 20% at any given time. The clade name corresponds to the year in which the new variant emerges, with the next letter of the alphabet assigned to the clade name. To define a new clade, the variant must have at least two new mutations compared to its parent major clade. Thus, major clades are identified by their emergence year and a corresponding letter, such as 19A, 19B, or 20A (Nextstrain, 2024).

Globally circulating clades include 19A (originating from Asia: China/Thailand), 19B (originating from Asia: China), 20A (originating from North America/Europe/Asia: USA, Belgium, and India), 20B (originating from Europe: UK, Belgium, and Sweden), and 20C (originating from North America: USA) (Table 1) (Alm et al., 2020; Hadfield et al., 2018).

Table 1: Classification of SARS-CoV-2 viral variants identified worldwide.

WHO	PANGO (Lineage)	GISAID (Clade)	Nextstrain (Clade)	Country of Identification
Alpha	B.1.1.7	GRY	20I (V1)	UK
Beta	B.1.351	GH/501Y.V2	20H (V2)	South Africa
Gamma	P.1	GH/501Y.V3	20J (V3)	Japan/Brazil
Delta	B.1.617.2	G/478K.V1	21A	India
Epsilon	B.1.427/B.1.429	GH/452R.V1	21C	USA
Eta	B.1.525	G/484K.V3	21D	Multiple country
Iota	B.1.526	GH/253G.V1	21F	USA
Kappa	B.1.617.1	G/452R.V3	21B	India
Lambda	C.37	GR/452Q.V1	21G	Peru
Mu	B.1.621	GH	21H	Colombia
Omicron	B.1.1.529	BR/484A	21K	Africa

WHO, World Health Organization; PANGO, Phylogenetic Assignment of Named Global Outbreak Lineages; GISAID, Global Initiative on Sharing All Influenza Data; UK, United Kingdom; USA, United States of America.

2.5. Transmission, replication and pathogenesis of SARS-CoV-2

SARS-CoV-2 is primarily transmitted through respiratory droplets and aerosols that are expelled when an infected individual cough, sneezes, speaks, or breathes. These droplets can be inhaled by others or can land on surfaces, where they may subsequently be transferred to the mucous membranes of the eyes, nose, or mouth (Dhand and Li, 2020). Research indicates that aerosol transmission can occur in poorly ventilated indoor environments, facilitating airborne spread over extended distances (Doremalen et al., 2020).

In addition to respiratory transmission, SARS-CoV-2 can also spread through fomite transmission, wherein viral particles remain viable on surfaces such as plastic, metal, and glass for varying durations. However, the actual risk of transmission via contaminated surfaces is considered to be lower than that of airborne spread (Goldman, 2020). Other potential routes of transmission are faecal-oral transmission: SARS-CoV-2 RNA has been detected in stool samples, suggesting that the virus may be shed through the digestive tract. However, its role in human-to-human transmission remains ambiguous (Heneghan et al., 2021; Xiao et al., 2020). Some studies have reported instances of SARS-CoV-2 transmission from mother to fetus during pregnancy, although such occurrences appear to be relatively rare (Schwartz, 2020).

The high expression of ACE2 on the surface of lung alveolar epithelium and enterocytes in the small intestine is believed to facilitate the viral entry of SARS-CoV, a mechanism that SARS-CoV-2 likely also employs (Hamming et al., 2004). In addition to these locations, ACE2 is widely expressed in various human tissues, including the heart, kidneys, and both arterial and venous endothelial cells. The presence of ACE2 in these tissues likely contributes to extrapulmonary manifestations such as diarrhea, acute renal injury, cardiac injury, and vascular endothelial damage, which can ultimately lead to multisystem organ failure (Nadim et al., 2020).

Children exhibit lower ACE2 expression, which may account for their reduced early acquisition rates of COVID-19; however, they demonstrate a higher incidence of multisystem inflammatory syndrome in older children. It is noteworthy that the tissues expressing ACE2 do not participate equally in the pathogenesis of COVID-19, suggesting that additional factors may also contribute to tissue damage (Eun, 2021).

Once SARS-CoV-2 enters the human body, it undergoes a multi-step lifecycle within host cells, which facilitates its replication and the dissemination of new virions (Pampel, 2021). SARS-CoV-2 enters host cells through two pathways, namely the endocytic pathway and transmembrane serine protease 2 (TMPRSS2)-mediated surface pathway (Boechat et al., 2021). The primary steps in this lifecycle include viral entry, RNA replication, protein synthesis, virion assembly, and release (Pampel, 2021). The entry of SARS-CoV-2 into human cells is facilitated by the spike (S) protein, which binds to the angiotensin-converting enzyme 2 (ACE2) receptor located on the surface of host cells. This receptor is expressed at high levels in the lungs, nasal epithelium, intestines, and endothelial tissues (Hoffmann et al., 2020).

The S protein is composed of 2 subunits: S1 and S2. The S1 domain contains the RBD, which is mainly responsible for binding of the virus to the receptor, while the S2 domain mainly contains the HR domain, including HR1 and HR2, which is closely related to virus fusion (Cui et al., 2019). When the S1 subunit attaches to the ACE2 receptor on the host cell, a transmembrane protease, serine 2 (TMPRSS2), cleaves the S protein to reveal the S2 subunit and ACE2. This cleavage activates the S2 subunit trimers to fuse viral and host lipid bilayers, releasing the viral ribonucleoprotein complex into the cell (Lamers and Haagmans, 2022). After binding, the virus becomes enveloped within an endosome formed by the cell membrane and enters the cell through endocytosis(Hernandez et al.,

2022). An alternative entry pathway for the virus involves the endosome, where cathepsins facilitate spike protein cleavage; however, this route is not efficiently utilized in primary epithelial cells (Hoffmann et al., 2020; Lamers et al., 2021).

Following its entry into the host cell, the viral genome (approximately 30 kb positive-sense RNA strand) is released into the cytoplasm. The RNA serves as a direct template for protein synthesis, enabling the host cell's ribosomes to translate two large open reading frames (ORF1a and ORF1b) (V'kovski et al., 2021). These ORFs generate polyproteins pp1a and pp1ab, which are subsequently cleaved into 16 non-structural proteins (NSPs) by viral proteases Mpro and PLpro (Gordon et al., 2020). The NSPs assemble to form the replication-transcription complex (RTC), which is responsible for synthesizing new copies of the viral genome utilizing RNA-dependent RNA polymerase (RdRp; NSP12) and producing sub-genomic RNAs that function as mRNA templates for the synthesis of structural and accessory proteins (V'kovski et al., 2021)

Then the viral structural proteins, including spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins, are synthesized from subgenomic RNAs. The S, M, and E proteins undergo processing in the endoplasmic reticulum (ER) and are subsequently transported to the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) for viral assembly (V'kovski et al., 2021). Meanwhile, the N protein binds to newly replicated viral RNA, forming the ribonucleoprotein complex (Bai, Z. et al., 2021). The M protein plays a critical role in virion morphogenesis through its interactions with the S, E, and N proteins (Brant et al., 2021).

Newly formed virus particles bud from the surface of the host cell, carrying the precisely packaged genome that is ready to initiate another cycle of infection. Throughout this lifecycle, SARS-CoV-2 employs various strategies to evade host immune responses, including the modulation of interferon signalling pathways and the inhibition of host protein synthesis (Maison et al., 2023; V'kovski et al., 2021)

Following receptor binding, the virus must gain access to the cell cytoplasm. The replication and progeny virion assembly process is facilitated by acid-dependent proteolytic cleavage of the S protein by cathepsins, TMPRSS2, or other proteases, followed by the fusion of viral and cellular membranes (Le et al., 2023). Cleavage occurs at two sites within the S2 region of the protein: one site separates the

receptor-binding domain (RBD) from the fusion domains, while the other exposes the fusion peptide (Shang et al., 2020).

Fusion typically occurs within acidified endosomes; however, some coronaviruses are known to fuse at the plasma membrane (Fehr and Perlman, 2015). Upon entry, the viral replicase gene, comprising two open reading frames (ORFs), encodes polyproteins pp1a and pp1ab. These polyproteins contain all the non-structural proteins (NSPs) necessary for intracellular replication (V'kovski et al., 2021). The polyproteins are cleaved by proteases, including papain-like proteases (PLpro) and the main protease (Mpro), resulting in the formation of the replication-transcription complex (RTC) responsible for viral RNA synthesis (Yang and Rao, 2021).

During the assembly of progeny virions, the M protein directs the protein-protein interactions required for virion formation, with assistance from the E protein (Fehr and Perlman, 2015). Co-expression of M and E proteins leads to the formation of virus-like particles (VLPs), indicating their essential role in the formation of the coronavirus envelope. The N protein further enhances VLP formation, suggesting that encapsulated genomes within the ERGIC facilitate particle envelopment. The S protein integrates into virions during this process, although it is not a prerequisite for assembly (Bosch et al., 2005). Subsequently, progeny virions are transported to the cell surface within vesicles and released via exocytosis (Haan and Rottier, 2005).

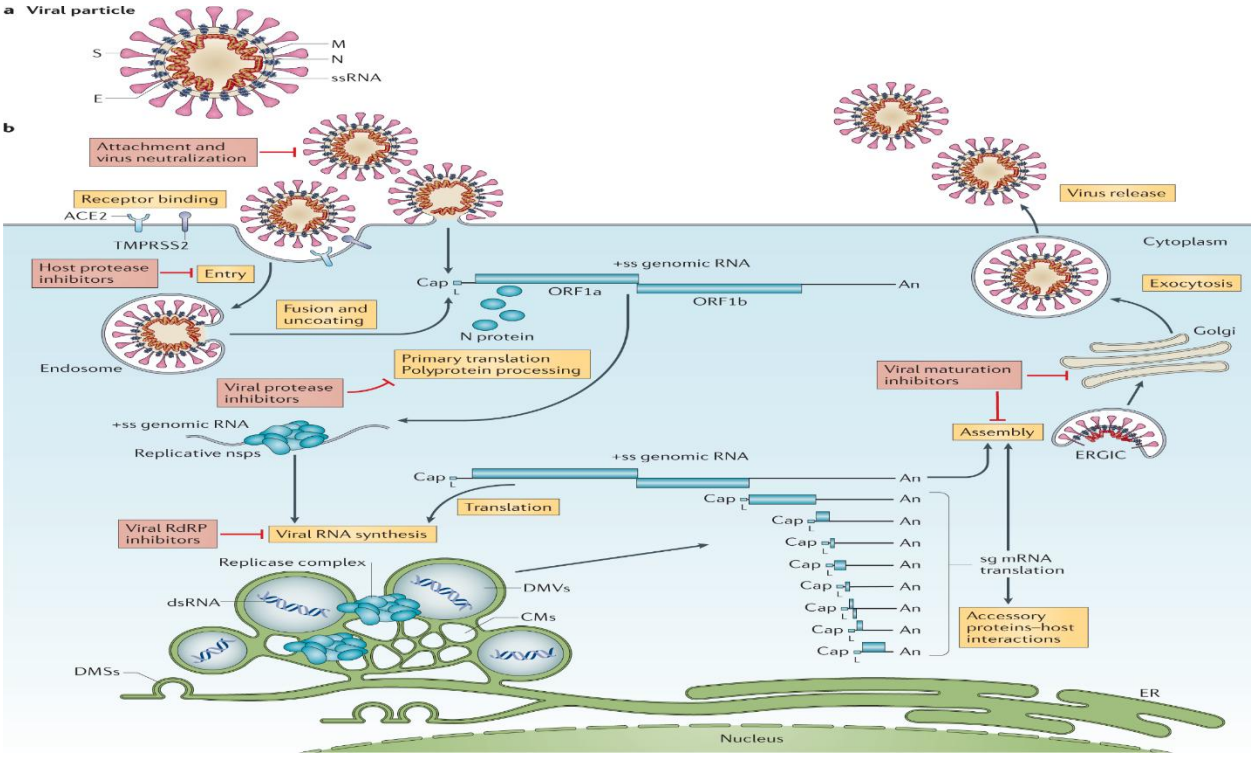


Figure 5. Life cycle of SARS-CoV-2
Source: (Alm et al., 2020)

The replication-transcription complex (RTC) operates within specialized double-membrane vesicles (DMVs) that are derived from the endoplasmic reticulum (ER) (Klein et al., 2020). These structures provide a protected environment for viral RNA synthesis, effectively shielding double-stranded RNA (dsRNA) replication intermediates from recognition by the host's innate immune system (Wolff et al., 2023). Normally, dsRNA serves as a pathogen-associated molecular pattern (PAMP) that is detected by cytoplasmic pattern recognition receptors (PRRs). However, by sequestering dsRNA within DMVs, SARS-CoV-2 evades early immune detection, delaying host antiviral responses (Schneider et al., 2021).

Despite this, some viral RNA fragments eventually escape into the cytoplasm, where they can be detected by melanoma differentiation-associated protein 5 (MDA5), a cytosolic PRR specialized in recognizing long dsRNA molecules (Rebendenne et al., 2021; Sampaio et al., 2021). Once activated, MDA5 triggers the mitochondrial antiviral signaling protein (MAVS) pathway, leading to the phosphorylation of interferon regulatory factor 3 (IRF3) and the subsequent transcription of type I and type III interferons (IFNs) (Akira et al., 2006).

The interferons excreted by infected cells act in both an autocrine and paracrine manner, stimulating nearby epithelial cells to express interferon-stimulated genes (ISGs). These ISGs enhance antiviral defenses, restrict viral replication, and recruit immune cells such as macrophages, neutrophils, and dendritic cells to the site of infection (Acharya, D. et al., 2020; Major et al., 2020). Additionally, immune cells use endosomal Toll-like receptors (TLRs), particularly TLR3, TLR7, and TLR8, to detect viral RNA and amplify the interferon response (Ribero et al., 2020).

However, SARS-CoV-2 employs multiple strategies to antagonize interferon signaling, several viral proteins, including Nsp1, Nsp3, Nsp14, ORF6, and ORF8, have been shown to interfere with interferon production or downstream signaling (Miorin et al., 2020; Xia, H. et al., 2020). As a result, early interferon responses in COVID-19 patients are often delayed or suppressed, allowing for unrestricted viral replication in the initial days of infection (Hadjadj et al., 2020).

If the immune response fails to contain the virus, SARS-CoV-2 migrates from the upper respiratory tract to the lower respiratory tract via aerosolized viral particles or direct dissemination along the tracheobronchial tree (Hou, Y. J. et al., 2020). The virus can also be aspirated into the lungs, leading to alveolar infection. In the alveoli, SARS-CoV-2 preferentially infects alveolar type 2 (AT2) cells, which express high levels of ACE2 (Fan, H. et al., 2024; Mason, R. J., 2020). AT2 cells serve multiple critical functions in the lung, including the secretion of pulmonary surfactant, which prevents alveolar collapse by reducing surface tension. Additionally, AT2 cells act as progenitors for alveolar type 1 (AT1) cells, which mediate gas exchange (Barkauskas et al., 2013). The destruction of AT2 cells by viral replication and immune-mediated damage results in alveolar collapse, impaired oxygen exchange, and lung inflammation, potentially leading to acute respiratory distress syndrome (ARDS) in severe cases (Ackermann et al., 2020; Huang et al., 2020).

The body's physical barriers, including the respiratory mucosa, mucus, and ciliated epithelial cells, are considered the primary host defense mechanism (Alashkar et al., 2020). The virus evades these defenses by binding to angiotensin-converting enzyme 2 (ACE2) receptors and using transmembrane serine protease 2 (TMPRSS2) for cellular entry (Hoffmann et al., 2020; Shang et al., 2020). This interaction affects epithelial integrity and suppresses local immune responses, such as interferon signaling and antigen presentation, thereby facilitating deeper tissue invasion (Robinot et al., 2021).

Upon entry, the virus activates innate immune responses through pattern recognition receptors (PRRs), which lead to the release of interferons; however, these responses are frequently delayed or inhibited by viral proteins (Blanco-Melo et al., 2020). In severe cases, a hyperinflammatory cytokine storm exacerbates tissue damage and contributes to multi-organ failure (Merad and Martin, 2020). The adaptive immune response contains B cells producing antibodies and T cells eliminating infected cells, though these components can become functionally exhausted in severe disease (Premkumar et al., 2020).

Mutations in viral proteins, particularly in the receptor-binding domain (RBD) of the spike protein, increase the virus's ability to evade immune detection (Harvey et al., 2021). Understanding of these mechanisms has informed the development of vaccines and therapeutic interventions, including monoclonal antibodies and immunomodulators.

2.6. Diagnosis of SARS-CoV-2 infection

The diagnosis of COVID-19 involves a combination of physical examination, imaging, and laboratory testing. Clinicians evaluate symptoms such as fever, cough, and shortness of breath, along with vital signs like oxygen saturation (WHO, 2020). Imaging techniques, particularly chest X-rays and CT scans, help identify lung abnormalities, including ground-glass opacities and consolidations, which are common in moderate to severe cases (Bernheim et al., 2020). Laboratory confirmation is primarily achieved through RT-PCR, the gold standard for detecting viral RNA, while rapid antigen tests provide quicker, albeit less sensitive, alternatives (Corman et al., 2020). In some instances, serological tests are employed to detect past infections by identifying antibodies. These diagnostic tools work together to confirm infection, assess disease severity, and guide appropriate clinical management.

2.6.1. Clinical diagnosis of COVID-19

The clinical diagnosis of SARS-CoV-2 infection is critical for effective disease management and public health strategies. Initial assessments typically involve a complete evaluation of patient symptoms, with fever, cough, and shortness of breath serving as hallmark indicators of the infection. However, these clinical signs can overlap with other respiratory illnesses, complicating the diagnostic process (WHO, 2020). The clinical presentation of COVID-19 is highly variable, ranging from asymptomatic infection to severe pneumonia and death (WHO, 2020). The precise proportion of truly asymptomatic infections

remains uncertain; however, some estimates suggest that up to 80% of COVID-19 cases do not show explicit symptoms (Mizumoto et al., 2020). This emphasizes the challenge of identifying and controlling transmission, as asymptomatic carriers may unknowingly spread the virus.

The WHO defines a suspected COVID-19 case based on three key epidemiological and clinical circumstances. First, a patient is considered suspected if they exhibit acute respiratory illness (such as fever, cough, or shortness of breath) and have a history of travel to or residence in an area with community transmission in the 14 days prior to symptom onset. Second, a patient presenting with any acute respiratory symptoms who has had close contact with a confirmed or probable COVID-19 case within the last 14 days is also classified as a suspected case. Lastly, a patient with severe acute respiratory illness requiring hospitalization, in the absence of an alternative diagnosis, is classified as a suspected COVID-19 case (WHO, 2020). If a suspected patient tests positive by laboratory diagnostic methods, they are classified as a confirmed COVID-19 case, irrespective of the severity of their symptoms (Christensen et al., 2022).

The most prevalent symptoms of COVID-19 include fever, dry cough, fatigue, sputum production, dyspnea, sore throat, headache, myalgia (muscle pain), arthralgia (joint pain), and chills. Furthermore, gastrointestinal manifestations such as nausea, vomiting, diarrhea, and abdominal pain have been observed in some patients (WHO, 2020). Notably, olfactory dysfunction (loss of smell) and gustatory dysfunction (loss of taste) have been frequently documented as early and persistent symptoms among individuals diagnosed with COVID-19. A study conducted in Europe reported that 85.6% of patients experienced loss of smell, while 88.0% reported loss of taste; these disturbances often persisted even after the resolution of other symptoms (Lechien et al., 2020). Moreover, COVID-19 has been associated with cutaneous manifestations. A study from Spain identified a range of skin lesions in infected patients, including purpuric, papulovesicular, livedoid, urticarial, maculopapular, and thrombotic-ischemic lesions (Galván Casas et al., 2020).

In children, COVID-19 typically presents as mild or asymptomatic, although some cases exhibit fever (41%), cough (48%), and pharyngeal erythema (46.2%) (Dong et al., 2020). Severe complications in pediatric patients are rare but can occur, such as multisystem inflammatory syndrome in children (MIS-C), a severe inflammatory response triggered by SARS-CoV-2 infection.

A study by Wu & McGoogan, 2020, showed that, among laboratory-confirmed COVID-19 cases, 81% of patients experience mild to moderate disease. Approximately 14% develop severe disease, characterised by dyspnea, tachypnea, low blood oxygen saturation (≤93%), and lung infiltrates affecting more than 50% of lung tissue within 24-48 hours. The remaining 5% progress to critical illness, which includes respiratory failure, septic shock, and/or multiple organ dysfunction (Wu and McGoogan, 2020).

Multiple factors expressively increase the risk of developing severe or fatal COVID-19. Advanced age (≥65 years) is among the most critical predictors of severe illness. Furthermore, individuals with pre-existing medical conditions, including cardiovascular disease, hypertension, diabetes, chronic respiratory diseases, immunodeficiencies, cancer, and obesity, are at a higher risk of complications and mortality (Zhou et al., 2020). Particularly, male sex was independently associated with higher in-hospital mortality and, in general, worse in-hospital outcomes (Jin, et al., 2020; Palaiodimos et al., 2020).

2.6.2. Imaging

Imaging techniques are essential for the detection, assessment, and management of COVID-19-related pulmonary and extrapulmonary complications. Some of the techniques are:

2.6.2.1. Chest X-Ray (CXR)

It is a widely used first-line imaging tool in suspected respiratory infections due to its low cost, rapid results, and accessibility (Kanne et al., 2020). It is used to detect lung abnormalities, assess disease severity, and rule out alternative diagnoses. Pulmonary abnormalities typically become visible on CXR within 12 hours of symptom onset, making it an essential early screening tool (Jacobi et al., 2020). CXR is suggested in cases of clinical suspicion of COVID-19 while awaiting RT-PCR results. It is also used for routine follow-ups, particularly to reduce the burden on CT units (Yoon, S. H. et al., 2020)..

2.6.2.2. Computed Tomography (CT) and High-Resolution CT (HRCT)

It is more sensitive than CXR for detecting early and subtle lung abnormalities. It is beneficial in cases where pneumonia is suspected but CXR findings are normal or inconclusive. CT imaging provides detailed insights into disease extent, complications, and recurrent pulmonary opacities (Ai et al., 2020).

CT is recommended for early diagnosis, particularly when RT-PCR is unavailable or yields false-negative results. It is also useful in cases of clinical deterioration that are not reflected in CXR findings and in severe cases to assess permanent lung damage (Bernheim et al., 2020).

2.6.2.3. Lung Ultrasound (LUS)

LUS has gained recognition as a radiation-free and bedside alternative to CXR and CT in critically ill patients. Initially used for pleural effusion evaluation, LUS has proven useful in diagnosing and monitoring COVID-19 pneumonia (Mongodi et al., 2021). LUS is particularly beneficial for ICU patients, enabling serial monitoring of lung aeration, re-expansion of atelectatic areas, and respiratory support adjustments (Smith et al., 2020).

2.6.2.4. Magnetic Resonance Imaging (MRI)

Although MRI though not a primary imaging modality for COVID-19 lung infections, it plays an essential role in evaluating cardiovascular and neurological complications associated with the disease. Unlike CT, MRI does not expose patients to ionizing radiation, making it suitable for long-term monitoring and younger patients (Rajpal et al., 2021). MRI is particularly beneficial for assessing COVID-19-related myocarditis, as cardiac MRI can detect myocardial oedema, fibrosis, and inflammation, providing insights into cardiovascular complications (Puntmann et al., 2020). Additionally, brain MRI is used to investigate neurological complications, such as stroke, encephalopathy, and microvascular injury in COVID-19 patients (Kremer et al., 2020).

2.6.2.5. Positron Emission Tomography-Computed Tomography (PET-CT)

It particularly uses 18F-FDG (fluorodeoxyglucose), which is an advanced imaging technique that provides quantitative assessments of lung inflammation. It detects increased glucose metabolism in inflamed lung tissues, offering valuable insights into disease activity and multi-organ involvement (Deng, Y. et al., 2020). PET-CT is suggested in early-stage COVID-19, especially when differentiating infectious lesions from malignant lung nodules (Qu, Y.-H. et al., 2021). It is also useful for assessing multi-organ damage in severe COVID-19 cases (Wang, X. et al., 2021).

2.6.3. Laboratory Diagnosis of SARS-CoV-2

Laboratory diagnosis plays a critical role in detecting, managing, and controlling SARS-CoV-2 infections. Accurate and timely diagnosis is essential for patient care, epidemiological surveillance, and

public health interventions. Selecting the appropriate sample for diagnostic testing is essential for the accurate detection of SARS-CoV-2. Various clinical specimens, including nasopharyngeal and oropharyngeal swabs, sputum, tracheal aspirates, and bronchoalveolar lavage, can facilitate detection (Hasan et al., 2020a). Several diagnostic methods, including molecular, serological, and antigen-based tests, are employed to confirm infection, assess disease progression, and monitor immunity (CDC, 2025).

2.6.3.1. Culture-Based Method

The initial isolation of SARS-CoV-2 was accomplished from bronchoalveolar lavage fluid using Vero E6 and Huh7 cell lines, with validation achieved through immunofluorescence, reverse transcription quantitative polymerase chain reaction (RT-qPCR), and metagenomic sequencing (Zhou et al., 2020). SARS-CoV-2 has been effectively cultured in various cell lines, including Vero E6, which demonstrates elevated expression of ACE2, the principal receptor facilitating viral entry (Hoffmann et al., 2020). Also, engineered Vero E6 cells that overexpress TMPRSS2 significantly enhance viral isolation, underscoring the critical role of this protease in SARS-CoV-2 pathogenesis (Zang et al., 2020). Comparative analyses of the susceptibility of different cell lines indicate that both Vero E6 and Vero CCL-81 support productive viral replication, while other lines, such as A549 and EFK3B, exhibit limited or nonexistent replication (Harcourt et al., 2020). A study, conducted by Bullard et al., 2020, evaluating 90 RT-qPCR-confirmed SARS-CoV-2 samples demonstrated viral growth in 28.9% of instances, predominantly among those with lower cycle threshold (Ct) values (<24), emphasizing the importance of viral load in successful isolation (Bullard et al., 2020). Therefore, culture-based methodologies are used in both research and public health contexts and are not advisable for routine diagnostic techniques due to their low sensitivity, protracted time requirements, and necessity for Biosafety Level 3 (BSL-3) containment WHO, 2020).

2.6.3.2. Electron Microscopy (EM)

EM has played an important role in the identification and visualization of SARS-CoV-2. During the initial outbreak in Wuhan, the virus was first seen using EM in cultured cells, confirming it as a coronavirus (Kim et al., 2020; Park et al., 2019; Zhou et al., 2020). EM is unique in its ability to directly visualize intact viral particles, unlike molecular methods that detect nucleic acids or proteins, which may not always represent whole infectious units (Zhu et al., 2020). This makes EM a valuable tool for confirming the presence of viruses in cell culture and for histopathological analysis of infectious

diseases caused by small pathogens like viruses, often in combination with antibodies or molecular probes targeting specific viral components (Bhat et al., 2022). However, analyzing SARS-CoV-2 particles in patient samples through EM poses significant challenges (Bullock et al., 2022; Dittmayer et al., 2020; Goldsmith et al., 2020).

2.6.3.3. Molecular-Based Methods

2.6.3.3.1. *Reverse Transcription quantitative Real-Time PCR (RT-qPCR)*

RT-qPCR, a method for detecting the virus's RNA, is the primary diagnostic tool for identifying SARS-CoV-2 in human patients. This method involves using samples primarily extracted from the respiratory tract, including nasopharyngeal exudate, nasal exudate, tracheobronchial aspirate, and sputum. It amplifies specific viral genes, allowing for real-time monitoring of millions of nucleic acid copies from small sample sizes (Zismann and Nourbakhsh, 2014).

The WHO recommends partial amplification of specific genes, such as the RNA-dependent RNA polymerase (RdRp) gene, to confirm the presence of the virus. Additionally, genes encoding viral structural proteins, including the envelope (E), nucleocapsid (N), and membrane (M) proteins, are often targeted to further verify the virus's presence (Corman et al., 2020). Method optimizations, such as one-step multiplex RT-qPCR or dual-targeting of the E and human RNase P (RP) genes, have improved efficiency and reliability (Coelho et al., 2022; Kudo et al., 2020). Primers are crucial in this reaction, binding to specific target sequences within the DNA strand. To ensure accuracy, primers must be designed with high specificity to prevent non-specific amplification and the formation of unwanted structures, like primer dimers, which can result in false-positive results (Bustin and Nolan, 2020).

Detection of the target gene is facilitated using either DNA probes or fluorescent dyes like SYBR Green. SYBR Green is an intercalating dye that binds to all double-stranded DNA (dsDNA) and emits a fluorescent signal upon binding. In contrast, probes feature a fluorescent reporter at the 5′ end and a quencher dye at the 3′ end, binding to a specific DNA target. When DNA polymerase cleaves the probe during amplification, the reporter is released, generating a fluorescent signal that is detected by the instrument (Zismann and Nourbakhsh, 2014).

The amplification process consists of three primary steps: denaturation, where the temperature rises to separate the double-stranded DNA (dsDNA); annealing, where the temperature decreases to allow primers to bind to their target sequences; and extension, where DNA polymerase facilitates the incorporation of nucleotides into the newly synthesized DNA strand. This cycle is repeated approximately 40 times, producing additional DNA copies with each cycle (Aoki et al., 2021).
Fluorescent signals are emitted as probes are cleaved or as SYBR Green binds to newly synthesized dsDNA. As the number of DNA copies increases, fluorescence intensity rises correspondingly (Aoki et al., 2021).

Despite its benefits, RT-qPCR encounters several challenges, including high costs, the need for specialized technical expertise, and lengthy processing times, which pose obstacles in resource-limited settings (Aoki et al., 2021). False negatives can occur due to low viral loads or high Ct values (>35–40), increasing the risk of undetected transmission (Kudo et al., 2020). The Centers for Disease Control and Prevention (CDC) recommends obtaining at least two negative tests, spaced 24 hours apart, to confirm viral clearance (CDC, 2020). Furthermore, asymptomatic individuals often have lower viral loads but similar durations of viral shedding, highlighting the potential for community transmission (Zhou et al., 2020).

2.6.3.3.2. *Droplet Digital PCR (ddPCR)*

This method was developed to enhance detection sensitivity and minimize false-negative results, especially in samples with low viral loads(Hindson, B. J. et al., 2011). Unlike RT-qPCR, ddPCR partitions samples into thousands of water-oil emulsion droplets, with each droplet serving as an independent PCR reaction. This approach enables absolute quantification of viral RNA using Poisson statistics instead of relying on relative Ct values (Behera, B. C. et al., 2021). A study by Suo et al. (2020) discovered that 26 samples that tested negative by RT-qPCR were positive with ddPCR (Suo et al., 2020). Similarly, another study by Yu et al. (2020) reported enhanced detection in low viral load samples (Yu, F. et al., 2020). These generally show that ddPCR offers superior sensitivity and specificity for detecting SARS-CoV-2. However, ddPCR requires expensive equipment and has a longer turnaround time, limiting its widespread use in routine diagnostics (Behera, B. C. et al., 2021).

2.6.3.3.3. *Reverse Transcription Loop-Mediated Isothermal Amplification (RT-LAMP)*

RT-LAMP is a powerful point-of-care (POC) assay for viruses like SARS-CoV-2. Unlike RT-qPCR, it operates under isothermal conditions, requiring only a single, fixed temperature, making it ideal for low-resources or field settings (Huang et al., 2020). The technique uses four to six specially designed primers that recognize multiple regions on the target sequences, enabling robust amplification (Notomi et al., 2000). Key primers include outer, inner and loop primers, which collectively initiate a process that forms dumbbell-shaped and cauliflower-like DNA structures, allowing exponential amplification without thermal cycling. Simplicity enables RT-LAMP to be performed with basic equipment like a water bath or heating block (Silva et al., 2020).

Detection of amplified product in RT-LAMP can be achieved through various methods such as colourimetric indicators, turbidity measurements, gel electrophoresis, fluorescence, or smartphone-based imaging (Guo et al., 2018; Nagamine et al., 2002). These features make RT-LAMP fast, cost-effective and easy to interpret, especially useful in point-of-care and resource-limited settings (Guo et al., 2018; Nagamine et al., 2002). Despite these advantages, the method has several limitations. The use of multiple primers increases the risk of non-specific amplification and false positives. It is also highly sensitive to contamination, lacks reliable quantification (unlike RT-qPCR), and has limited multiplexing capability (Yan et al., 2020; Zhang et al., 2021). Additionally, the interpretation of colour-based results can be subjective and emerging SARS-CoV-2 variants may affect their sensitivity (Butler et al., 2021). These challenges highlight the need for further optimization to improve the diagnostic reliability of RT-LAMP (Silva et al., 2020).

2.6.3.3.4. *Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-associated (Cas) proteins (CRISPR/Cas) based diagnostic techniques:*

CRISPR-Cas systems provide a rapid, cost-effective, and highly specific and highly specific alternative to traditional methods like RT-qPCR for detecting SARS-CoV-2. These systems utilize Cas proteins (such as Cas9, Cas12 and Cas13) that bind and cleave specific viral nucleic acid sequences, enabling highly sensitive detection, down to 0.4 copies/μL within 40 minutes (Kellner et al., 2019; Li et al., 2018).

Platforms like SHERLOCK (Cas13) and DETECTR (Cas12) generate fluorescent or colourimetric signals, providing quick and accurate results, with specificity rates comparable to RT-qPCR. More recent advances, such as AIOD-CRISPR and Cas14-based systems, further enhance sensitivity and reduce reliance on extensive amplification, making them well-suited for point-of-care use (Patchsung et al., 2020). However, limitations remain, including the need for pre-amplification steps in many protocols, risk of contamination from open-tube reactions and challenges related to regulatory approval and widespread commercial availability compared to the established RT-qPCR (Broughton et al., 2020).

2.6.3.3.5. *Sensor-based diagnostic methodologies*

It offers an alternative detection method that is rapid and high-throughput. Since the emergence of SARS-CoV-2, research groups worldwide have focused on developing alternative sensing modalities based on sensors for diagnosing the virus. These methods aim to minimize the reliance on costly laboratory equipment, reduce the need for trained personnel, and streamline sample preparation, all while providing fast and accurate results(Cui and Zhou, 2020).

One notable advancement is the dual-functional plasmonic biosensor, which combines localized surface plasmon resonance with a field-effect transistor (FET) based sensor, allowing for direct detection from patient samples(Seo et al., 2020). These emerging technologies, adapted from previous virus detection platforms, have significant potential to enhance the accessibility and throughput of SARS-CoV-2 diagnostics.

2.6.3.3.6. *Nanoparticle-based amplification:*

Nanoparticles have been incorporated into nucleic acid amplification techniques to increase the sensitivity and specificity of SARS-CoV-2 detection (Moitra et al., 2020). For instance, a gold nanoparticle (AuNP)-based colourimetric assay using thiol-modified antisense oligonucleotides (ASOs) targeting the SARS-CoV-2 N-gene achieved a limit of detection (LOD) of 0.18 ng/μL (Moitra et al., 2020). Additionally, a nanoparticle-based biosensor (NBS) coupled with RT-LAMP demonstrated a limit of detection(LOD) of 12 copies per test (Zhou et al., 2020). These methods have shown 100% analytical sensitivity in oropharyngeal swabs from COVID-19 patients and 100% specificity when tested on RNA templates from non-COVID-19 patients (Zhu, et al., 2020).

Nanoparticles provide distinct advantages over conventional methods, including lower costs, faster processing times, and enhanced sensitivity, making them ideal for first-line clinical laboratories, especially in resource-limited settings (Zhou et al., 2020). However, these techniques can be more expensive than RT-qPCR, involve complex sample pretreatment, and may experience photobleaching, which can compromise sensitivity and lead to false-negative results (Zhu et al., 2020). Despite these limitations, nanoparticle-based amplification remains a promising diagnostic tool for SARS-CoV-2 detection.

2.6.3.3.7. *Viral genome sequencing*

Genomic sequencing plays a complementary role in SARS-CoV-2 diagnosis, primarily by confirming infections, identifying variants, and monitoring diagnostic accuracy. While RT-PCR and antigen tests remain the primary diagnostic methods, sequencing helps resolve inconclusive cases and detects mutations that may impact test performance (Cosar et al., 2022; Lu et al., 2020).

A significant advantage of sequencing is its ability to differentiate between SARS-CoV-2 variants, which is essential for tracking the emergence of Variants of Concern (VOCs) like Alpha, Delta, and Omicron. Standard PCR tests can detect the virus but cannot identify its strain, whereas sequencing provides a detailed genetic profile, aiding in epidemiological surveillance and vaccine updates (Galloway et al., 2021; Tegally et al., 2021). Additionally, sequencing helps distinguish between reinfections and persistent infections by analyzing viral genetic changes over time. It also supports outbreak investigations, tracing transmission pathways in healthcare and community settings (Deng et al., 2020). Furthermore, it monitors diagnostic tests to ensure their effectiveness as the virus evolves (Fauver et al., 2020).

Despite these advantages, genomic sequencing is not a frontline diagnostic tool due to its high cost, complexity, and longer turnaround time. It requires specialized equipment and trained personnel, making it impractical for rapid clinical diagnosis (Metsky et al., 2019). However, as sequencing technologies continue to advance, they may become more integrated into routine diagnostics, enhancing pandemic response and surveillance (Peeling et al., 2022).

2.6.3.4. Serological methods

2.6.3.4.1. *Antigen testing*

SARS-CoV-2 antigen tests identify the presence of specific viral proteins, including the nucleocapsid (N) protein and the S1 or S2 domains of the spike (S) protein, which are expressed during active viral replication. As a result, these tests are most effective during the early or acute phase of the infection when viral protein concentrations are high (Kumar et al., 2021). Unlike RT-PCR, antigen tests offer a quicker and more cost-effective way to identify active infections, making them valuable for point-of-care use and mass screening in high-risk or resource-limited settings (Dinnes et al., 2020).

There are several formats of antigen tests based on detection methods. The most common is the lateral flow assay (LFA), or rapid antigen test (RAT), which provides visual results within 15-30 minutes through a colour change (Peeling et al., 2022). Another method is the enzyme-linked immunosorbent assay (ELISA)-based antigen test, which utilizes microplates coated with specific antibodies to detect SARS-CoV-2 antigens, producing a colourimetric or fluorescent signal. While this approach is more sensitive than LFAs, it requires laboratory processing. Fluorescent immunoassay (FIA)-based antigen tests use fluorescence for enhanced sensitivity and are frequently utilized in clinical settings (Pekosz et al., 2021). Additionally, more advanced microfluidic-based antigen tests are being developed to improve detection efficiency, although they are not yet widely available (Mina et al., 2020).

The principle of antigen testing is based on the interaction between antigens and antibodies. A respiratory sample, typically collected via a nasopharyngeal or nasal swab, is applied to a test strip or microplate coated with SARS-CoV-2-specific antibodies. If viral antigens are present, they bind to these antibodies, resulting in a detectable signal. This signal may manifest as a colour change (LFA), fluorescence (FIA), or an enzymatic reaction (ELISA). Antigen tests are most effective at detecting viral proteins when viral loads are high, usually within the first few days after symptom onset (Dinnes et al., 2020).

One of the primary advantages of antigen testing is its ability to provide rapid results. Results can be available in as little as 15 minutes, facilitating quicker isolation of infected individuals and containment of outbreaks (Peeling et al., 2022). Additionally, antigen tests are more cost-effective than RT-qPCR, making them more accessible for large-scale testing. They are easy to use and require minimal training

and infrastructure, making them suitable for decentralized testing. Antigen tests are particularly effective at detecting high viral loads, the point at which individuals are most infectious, thereby enhancing their role in reducing transmission (Pekosz et al., 2021).

However, antigen tests also have significant limitations. One major drawback is their lower sensitivity compared to RT-PCR especially in individuals with low viral loads or during the later stages of infection (Dinnes et al., 2020). Consequently, antigen tests are less reliable for identifying asymptomatic infections or cases with minimal viral shedding. Additionally, they have a higher false-positive rate due to potential cross-reactivity with other coronaviruses (Mina et al., 2020). Similarly, antigen tests are less reliable for identifying asymptomatic infections or cases with minimal viral shedding. They have a higher false-positive rate due to potential cross-reactivity with other coronaviruses. Since antigen tests rely on detecting viral proteins, they also have a limited window of detection, making them less useful for retrospective diagnosis or assessing past infections (Pekosz et al., 2021).

2.6.3.4.2. *Antibody testing*

It is used to detect antibodies produced by the immune system in response to SARS-CoV-2 infection or vaccination. Unlike molecular tests that identify the virus itself, antibody tests determine whether an individual has been previously exposed to SARS-CoV-2. These tests are essential for epidemiological surveillance, assessing vaccine efficacy, and understanding immunity (Long et al., 2020).

Antibody tests can be categorized by the antibodies they detect and the methodologies used. The most commonly tested antibodies include immunoglobulin M (IgM), immunoglobulin G (IgG), and immunoglobulin A (IgA). IgM is produced early during infection and indicates recent exposure, while IgG appears later and persists longer, making it useful for determining past infection or vaccine-induced immunity. IgA plays a role in mucosal immunity but is less frequently used in routine testing (Okba et al., 2020).

These tests work by detecting the interaction between viral antigens and antibodies in a patient's blood sample. The spike (S) protein and nucleocapsid (N) protein of SARS-CoV-2 are the most commonly targeted antigens. When a sample containing antibodies is introduced to a test platform, the antibodies bind to these antigens, triggering a detectable signal through colourimetric, fluorescence, or enzymatic

reactions (Long et al., 2020). Typically, IgM can be detected within the first two weeks of infection, while IgG appears later and can remain detectable for months, depending on the individual's immune response (Seow et al., 2020).

Different methodologies are available for antibody detection. The Enzyme-Linked Immunosorbent Assay (ELISA) is a widely used laboratory-based test known for its high sensitivity and specificity, though it requires trained personnel and specialized equipment (Guo, L. et al., 2020). Lateral Flow Immunoassays (LFIA) or rapid antibody tests provide quick results within 10-30 minutes and are easy to use, but they generally have lower accuracy than ELISA (Ibarrondo et al., 2020). Chemiluminescent Immunoassays(CLIA) are another highly sensitive and automated technique suitable for large-scale testing (Cinquanta et al., 2017). Neutralization assays, which assess the ability of antibodies to prevent viral entry into cells, are considered the gold standard for evaluating protective immunity; however, they require biosafety level 3 (BSL-3) laboratories (Khoury et al., 2021).

Antibody testing has several advantages. It can identify past infections, including those in asymptomatic individuals who may not have undergone diagnostic testing at the time of infection. This makes it a valuable tool for seroprevalence studies (Rostami et al., 2021). Furthermore, antibody testing helps to evaluate vaccine-induced immunity, providing insights into the duration of immunity following vaccination (Khoury et al., 2021). The tests are relatively non-invasive, requiring only a blood sample, and some, like rapid tests, provide quick results. Laboratory-based tests, such as ELISA and CLIA, provide high sensitivity and specificity, making them reliable for research and clinical purposes.

However, antibody-based testing does have a significant drawback in diagnosing active infections, as antibodies take time to develop. Therefore, it is not suitable for use in the early stages of COVID-19, when detecting viral RNA through qPCR or antigen testing is more reliable(Long et al., 2020). Furthermore, the presence of antibodies does not necessarily indicate active infection (Khoury et al., 2021). Some individuals, especially those with mild or asymptomatic infections, may produce low levels of antibody, leading to false-negative results (Okba et al., 2020). Additionally, cross-reactivity with other coronaviruses can result in false-positive results, which reduces the specificity of some tests (Guo et al., 2020). Another challenge is the declining of antibody levels over time, complicating the determination of the exact duration of immunity (Seow et al., 2020).

2.6.3.5. Non-Specific Laboratory Tests

Laboratory test results play a crucial role in supporting the diagnosis, prognosis, and monitoring of patients by detecting and measuring various biomarkers (Mehta et al., 2020; Tang et al., 2020). Studies have reported that some biomarkers are nonspecific has association with the infectious process of SARS-CoV-2. Specifically, decreased lymphocyte and platelet counts, reduced serum albumin levels, and elevated serum levels of C-reactive protein, D-dimer, ferritin, lactate dehydrogenase, transaminases, and interleukin-6 have been identified as useful indicators for risk stratification and predicting the severity of COVID-19 (Mehta et al., 2020).

Furthermore, severe cases of COVID-19 have been linked to cytokine storms characterized by elevated levels of IL-2R, IL-6, IL-10, and TNF-α, as well as a decline in absolute CD4+ and CD8+ T lymphocyte counts, potentially leading to cardiovascular collapse, multiple organ failure, and rapid mortality (Del Valle et al., 2020).

2.7. Prevention and Treatment

The prevention of SARS-CoV-2 infection focuses on a combination of vaccination and non-pharmaceutical interventions (NIPs). Since the onset of the pandemic, significant progress has been made in vaccine development. Approved vaccines, including mRNA (Pfizer-BioNTech, Moderna), viral vector (AstraZeneca, Johnson & Johnson), and inactivated virus types (Sinova, Sinopharm), have proven effective in minimizing severe diseases, hospitalization and death (Baden et al., 2021; Polack et al., 2020; Walsh et al., 2020).

Even though rare side effects like thrombosis with thrombocytopenia syndrome (TTS), vaccination is strongly recommended for all populations, including pregnant and lactating women (Shimabukuro et al., 2021). However, vaccines alone are not sufficient. Measurements such as mask wearing, physical distancing, hand hygiene, adequate ventilation and robust testing and contact tracing remain vital, especially in the face of emerging variants (WHO, 2020).

Treatment strategies vary based on disease severity. For mild to moderate cases, antivirals like Remdesivir and Paxlovid are most effective when administered early (Beigel et al., 2020; Hammond et

al., 2022). Monoclonal antibodies are also used for high-risk individuals, though their efficacy may decline with new variants (Dougan et al., 2021).

In severe cases, corticosteroids like dexamethasone reduce inflammation, while immunomodulators such as Tocilizumab and Baricitinib are employed to manage cytokine storms (Rosas et al., 2021). Supportive care, including oxygen therapy, fluid management and anticoagulation, is essential for hospitalized patients. Overall, an integrated approach combining vaccines, NPIs and evidence-based treatment remains the key to managing and mitigating COVID-19's global impact (*IDSA Guidelines, 2020*).

CHAPTER 3: MATERIALS AND METHODS

3.1. Study Design, Period and Setting

A cross-sectional experimental study design was employed to achieve all the study's objectives. The research took place from June to August 2022 at the Virology Reference Laboratory of the Ethiopian Public Health Institute (EPHI) in Addis Ababa, Ethiopia. EPHI, serving as the technical arm of the Federal Ministry of Health, is responsible for leading public health emergency preparedness and response, strengthening national laboratory systems, conducting public health research, and managing health data across the country. EPHI was the first institution in Ethiopia to launch COVID-19 diagnostic testing during the initial phase of the pandemic. Throughout the pandemic, EPHI played a crucial role in expanding SARS-CoV-2 testing capacity by offering technical assistance, mentorship, and training to regional laboratories, universities, and research centers. The Virology Reference Laboratory at EPHI received samples, conducted molecular testing, and communicated results through a national referral network that spans various geographical locations in Ethiopia. Additionally, EPHI served as the primary diagnostic center for government officials, diplomatic missions, and international travellers. The institute also features a dedicated genomics center equipped to support genomic surveillance, clinical diagnostics, and research.

3.2. Source population and study population

The source population comprises all individuals tested for COVID-19. However, SARS-CoV-2-tested individuals during the fifth wave of the pandemic (June to August 2022) were considered as the study population.

3.3. Sample Size Calculation and Sampling Technique

3.3.1. For molecular characterization of SARS-CoV-2 and conserved gene regions identifications:

Taking into account available resources, local sequencing capacity, and our expertise, we deliberately selected 150 previously confirmed positive samples with Ct values of 30 and below, using their unique identifiers from the national electronic databases. These nasopharyngeal (NP) swab samples were collected from clinically suspected cases living in Addis Ababa, Oromia, Amhara, and Southern Nation Nationalities and Peoples (SNNP) regions and were stored at -80 ◦C. Relevant patient's metadata, including demographic information, clinical data, and RT-qPCR results, were anonymously extracted

from electronic databases. The study participants were then grouped based on factors such as age, sex, signs and symptoms, underlying chronic conditions, travel history, and vaccination status.

3.3.2. For evaluation of extraction free RT-qPCR method for SARS-CoV-2 detection

We used a Buderer's formula to calculate the number of positive and negative samples required for the study as follows.

$$n_{pos} = \frac{(Z^2) \cdot Se \cdot (1 - Se)}{d^2} \qquad n_{neg} = \frac{(Z^2) \cdot Sp \cdot (1 - Sp)}{d^2}$$

n_{pos} = Number of positive samples

n_{neg} = Number of negative samples

Se = expected sensitivity of the new method (90%)

Sp = expected specificity of the new method (96%)

d = desired precision (e.g., 0.05)

Z = 1.96 for 95% confidence

The initial calculated sample sizes for positive and negative samples were 139 and 73, respectively. After accounting for 20% poor-quality samples and 15% incomplete data (including Ct-value), the revised sample sizes were 188 and 99. Then we purposely adjusted to 190 positive and 110 negative samples. Therefore, a total of 300 samples were retrospectively obtained from the repository of the National Virology Reference Laboratory, Ethiopian Public Health Institute. These specimens were initially collected for clinical diagnosis of COVID-19 using Viral Transport Media (VTM) (Miraclean Technology, Shenzhen, China). These samples were retrieved by using the unique numbers assigned to each specimen in the laboratory register.

3.4. Inclusion and exclusion criteria

3.4.1. Inclusion criteria

For Molecular Characterization of SARS-CoV-2 and Conserved Gene Region Identifications:

I. Samples collected and tested during the fifth wave of the pandemic in Ethiopia, i.e. from June to August 2022.

II. Those samples that had a Ct-value of less than or equal to 30.

III. Those samples have complete metadata.

For Evaluation of Extraction-Free RT-qPCR Method for SARS-CoV-2 Detection

I. Those samples tested for SARS-CoV-2 and stored for less than a week.

II. Those samples with known Ct-values for both positive and negative test results.

3.4.2. Exclusion criteria

I. Samples of poor quality, including those that were improperly stored, leaked, had illegible labels, or were not labelled correctly.

II. Samples with Incomplete data.

3.5. Study variables

Dependent variable for Molecular Characterization of SARS-CoV-2

- SARS-CoV-2 genetic variants among study participants

Dependent variable for identification of Conserved Gene Regions:

- Conserved SARS-CoV-2 gene regions for monoclonal antibody-based diagnosis

Dependent Variable for Evaluation of Extraction-Free RT-qPCR Method for SARS-CoV-2 Detection:

- Diagnostic Performance (Sensitivity, Specificity, Predictive values and others) of Extraction-free dilution and heating steps for SARS-CoV-2 detection

Independent Variable for Molecular Characterization of SARS-CoV-2 and Conserved Gene Regions Identification

- Sociodemographic data, clinical information, vaccination status, travel history and underlying diseases.
- Cycle threshold value (Ct-value)

3.5.1. Outcome variables

- SARS-CoV-2 genetic variants circulated in the country
- Conserved gene regions for SARS-CoV-2 diagnosis
- Diagnostic performance of the extraction-free dilution and heating method

3.6. Laboratory procedure

3.6.1. Selection of stored specimens

For molecular characterization of SARS-CoV-2 and Conserved gene regions identifications:

A total of 150 nasopharyngeal specimens were retrospectively obtained from the national virology reference laboratory at EPHI. These samples were originally collected in Viral Transport Media (VTM) for clinical COVID-19 diagnosis. All selected specimens tested positive for SARS-CoV-2, with a Ct-value of less than or equal to 30.

For Evaluation of Extraction-Free RT-qPCR Method for SARS-CoV-2 Detection:

A total of 300 nasopharyngeal samples,190 positive and 110 negatives, were retrospectively selected from the national reference laboratory repository at EPHI. Selection of samples was done by using the unique numbers assigned to each specimen in the laboratory register. The study also tried to include samples having different viral load (Ct-values) levels.

3.6.2. SARS-CoV-2 detection (RT-qPCR)

RT-qPCR was performed using the ABI 7500 Fast instrument (Thermo Fisher Scientific) in combination with DAAN SARS-CoV-2 detection kits (Daan Gene Co., Ltd). The thermal cycling protocol included an initial reverse transcription step at 50°C for 15 minutes, followed by a polymerase activation at 95°C for 2 minutes. This was followed by 40 amplification cycles consisting of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 60 seconds (annex 4).

The purpose of this assay was to detect the presence of the SARS-CoV-2 virus using primers and probes that correspond to the N/ORF1a/b gene. For comparison, the standard extraction-based (EB) technique was used as a reference method, alongside the Extraction-free heated (EFHD) method. The laboratory protocol was used to interpret the results. A Ct value greater than 40, along with the absence of an amplification curve in both the FAM and VIC channels, was interpreted as negative for SARS-CoV-2 RNA in the sample. Conversely, the presence of clear amplification curves in both the FAM and VIC channels, with a Ct value of 40 or below, was interpreted as a positive result. If amplification was detected in only one of the two channels (either FAM or VIC) with a Ct value of 40 or less, while the

other channel showed no amplification, the assay was re-tested. The repeated result was considered the final result (annex 4).

At each step of both methods, strict quality control measures were taken. In addition to positive and negative control provided in the kit for the extraction method, we also used an extraction control were also used. All measuring devices in the laboratory were within their calibration date, including the thermal cycler.

3.6.3. RNA extraction, RT-qPCR and whole genome sequencing of SARS-CoV-2 for genomic characterization and conserved gene region identification

3.6.3.1. RNA extraction and RT-qPCR detection prior to cDNA library preparation

Total RNA was extracted from 150 nasopharyngeal swab samples using Mega Bio plus Virus Purification Kit II (BSC87S1E), (Hangzhou Bioer Technology Co., Ltd., Hangzhou, China) on the Bioner Gene Pure Pro fully automatic Nucleic Acid Purification System (NPA-32P) per the manufacturer's instructions (Bioer Technology, 2020).

Real-time RT-qPCR was subsequently performed to detect SARS-CoV-2 RNA using gene-specific primers and probes targeting the ORF1ab and N genes of the virus. The assay was carried out on the CFX96 Deep Well™ Real-Time PCR Detection System (Bio-Rad Laboratories, Inc., USA). Each sample was also tested for the human RNase P gene as an internal control to ensure the quality of the RNA extraction and amplification. The clinical samples were handled and processed in a Biosafety Laboratory-2 (BSL-2) at the virology laboratory of EPHI. Trained personnel followed strict protocols using personal protective equipment and adapted procedures specifically designed for handling airborne pathogens to ensure safety throughout the process.

Based on the RT-qPCR results, only 70 RNA extracts with a Ct-value of 30 and below were selected for library preparation. The remaining 80 samples, which had Ct values above 30, indicating low viral load, were excluded from the study.

3.6.3.2. Complementary DNA (cDNA) synthesis, DNA library preparation, and whole-genome sequencing

A total of seventy RNA extracts that exhibited low cycle threshold (Ct) values (Ct < 30) in the quantitative real-time PCR assays were selected for downstream library preparation. Double-stranded complementary DNA (cDNA) was synthesized from the extracted RNA using the Illumina COVIDseq Test (RUO) kit (Illumina, Inc., San Diego, CA, USA.), following the manufacturer's protocol (Illumina COVIDseq Test, 2020). For the synthesis of first-strand cDNA, the extracted RNA was incubated with random hexamer primers at 65 ˚C for 3 minutes. Subsequently, First Strand cDNA Master mix was added, consisting of 9 µL of First Strand Mix (FSM) and 1 µL of Reverse Transcriptase (RVT) per reaction (70 total reactions). The reaction was incubated at 25°C for 5 minutes, followed by 50°C for 10 minutes, and finally terminated at 85°C for 5 minutes.

Following cDNA synthesis, paired-end libraries were constructed using the Illumina DNA Prep with Enrichment kit (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions (Illumina DNA Prep, 2020). Briefly, 50 ng of the purified double-stranded cDNA was subjected to tagmentation using bead-linked transposomes. Adapter ligation was performed, and unique dual indexes (IDT for Illumina DNA/RNA UD Indexes) were added through limited-cycle PCR amplification. The resulting amplified products were purified using sample purification beads.

For targeted enrichment of the SARS-CoV-2 genome, indexed libraries were pooled in 12-plex reactions, enabling multiplexing of 12 samples per pool. Hybridization was carried out using biotinylated respiratory virus oligonucleotide probes (Illumina Inc.), which specifically bind to SARS-CoV-2 genomic regions. The hybridized DNA fragments were then captured with streptavidin-coated magnetic beads. Following hybrid capture, the enriched libraries were further amplified and purified using Illumina Tune Beads (ITB).

The final concentration of enriched libraries was quantified using a Qubit® 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Libraries were then denatured with 0.2 N NaOH and diluted using the hybridization buffer (HT1). Sequencing was performed on an Illumina NextSeq system using the NextSeq 550 high-output 550-cycle v3 kit, generating paired-end reads for downstream analysis.

3.6.4. RNA extraction and detection methods of SARS-CoV-2 for the evaluation of EFDH method

3.6.4.1. Extraction-based protocol (reference method)

Extraction was performed manually using the DAAN extraction kits following the standard protocol. The process began with sample lysis under highly denaturing conditions, effectively inactivating RNases to protect the integrity of the viral RNA. These conditions ensured that the RNA remained intact throughout the extraction process. Next, the buffering conditions were carefully adjusted to facilitate the binding of RNA to the specialized membrane of the plates. This step was crucial for capturing the RNA while preventing the co-binding of contaminants.

To ensure the removal of impurities, the bound RNA underwent a series of washes using two wash buffers followed by a final ethanol wash. These steps effectively eliminated contaminants such as proteins, nucleases, and other potential inhibitors, resulting in high-purity RNA. The purified RNA was then eluted using an especially formulated RNase-free buffer, making it immediately available for downstream applications or safe for long-term storage without degradation.

3.6.4.2. Extraction-free (EFHD) protocol

100 µL of each sample was transferred to a heat-resistant 1.5 mL microcentrifuge tube and heated at 72°C for 15 minutes using an Eppendorf Thermomixer. The eluates were then stored at -20°C for up to 4 hours while waiting for other clinical samples to be processed in batches. Afterwards, 10 µL of the eluate was mixed with 20 µL of the master mix, following a similar procedure to the extraction-based eluted samples. Worksheets were prepared with other clinical samples, and detection was conducted using the Applied Biosystems ABI7500 machine, following the laboratory protocol as shown in Figure 6.

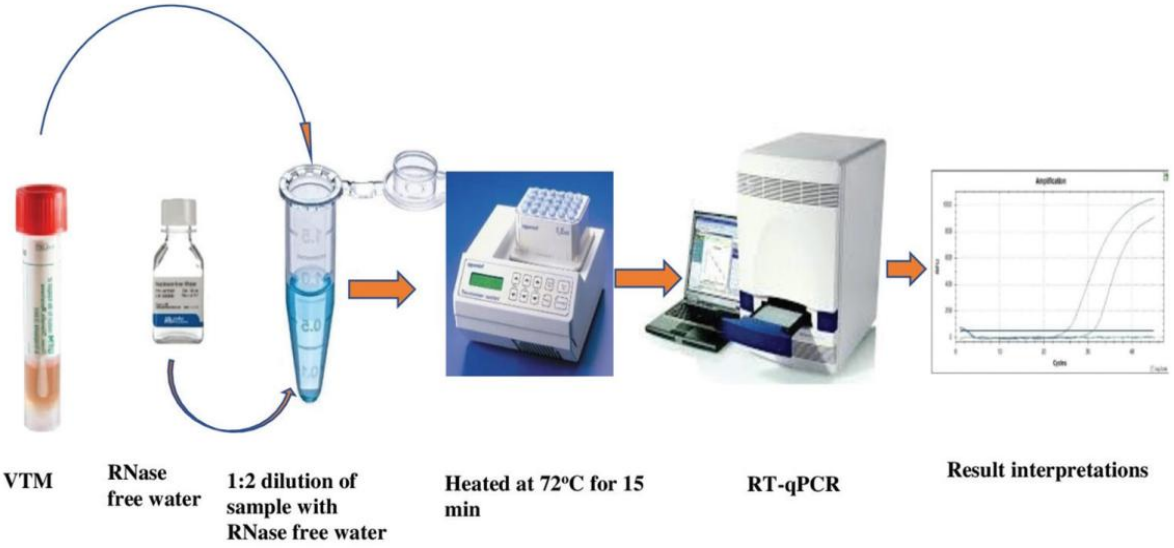


Figure 6. Process for sample dilution, heating, detection, and interpretation of SARS-CoV-2.

3.6.5. Method validation

For Evaluation of Extraction-Free RT-qPCR Method for SARS-CoV-2 Detection:

This experiment was informed by a study conducted in Copenhagen, Denmark, in March 2020, which assessed an alternative molecular detection workflow for SARS-CoV-2 utilizing heat treatment and dilution (Fomsgaard and Rosenstierne, 2020). Additionally, we referenced another study that evaluated an extraction-free RT-qPCR protocol employing a combination of proteinase K and heat inactivation (Genoud et al., 2021). Informed by these methodologies, we selected a specimen exhibiting a cycle threshold (Ct) value of 19, alongside an internal control (IC) labelled with VIC dye, which also presented a Ct value of 19. This specimen was diluted in 1:2 and 1:4 with RNAse-free water.

The diluted specimens were then heated at temperatures of 62°C and 72°C for 10, 15, 20, and 30 minutes, as well as at 96°C for 5, 10, 15, and 20 minutes using an Eppendorf thermomixer. Considering these heated and diluted specimens as an RNA eluate, 10µL of the eluate was added into 20µL of master mix and performed RT-qPCR using the ABI 7500 (Thermo-Fisher Inc.) using DAAN detection kits (Daan Gene Co., Ltd). The target genes for DAAN detection kits for SARS-CoV-2 primarily focused on the N gene and the ORF1ab gene. These genes are crucial for the detection and identification of the SARS-CoV-2 virus through PCR-based assays. The results showed that at a temperature of 72°C, the N/ORF1a/b Ct values were 24 at 15 min. and 26 at 30 min., and the VIC (IC) values were 26 at 15

min and 27 at 30 min, respectively, as shown in Table 2. Compared to the conventional extraction-based methods, a difference of five cycle thresholds (Ct-values) was observed.

Therefore, a promising and comparable cycle threshold was obtained in the 1:2 diluted specimens exposed to 72°C for 15 minutes. Following these experimental results, a total of 190 positive and 110 negative samples were retrieved from the national sample repository.

Table 2. The preliminary experiment shows the different dilution and heating steps with detectable viral RNA.

Original Ct value	Room temperature for 30 min		Undiluted but Heated					Dilution	Heating steps with time											
			99ºC	95ºC	72ºC	65ºC	56ºC		65ºC				72ºC					95ºC		
	Diluted	Undiluted	2 min	5 min	10 min	20 min	30 min		10 min	15 min	20 min	30min	10 min	15 min	20 min	30 min	5 min	10 min	15 min	20 min
19	ND***	ND	ND	Neg	33.1	Neg	ND	1/2	26.6	26	26.8	25	25.6	24.8	26	24.8	24.5	24	26.7	UND*
								IC value	31	31	32	30.8	29.6	26	29.5	25.5	29.6	33	34	UND
			IC**	28	26.7	29.3		1/4	26.8	26.5	24.7	26	26	23.9	24.6	26.6	24.8	25.9	23.8	25.8
								IC value	29.5	30	31.8	31	31	31	29.9	31	31.9	32.8	32.7	32

***Undetermined**
****Internal Control**
***** Not detected**

3.6.6. Quality assurance

For molecular characterization of SARS-CoV-2 and Conserved gene regions identification:

To increase the success rate of whole genome sequencing, an RT-qPCR test was performed again after the samples were thawed at room temperature. The RNA from these samples was re-extracted using the BGI extraction kit, and detection was performed using a BIO-RAD CFX96 Deep Well™ Real-Time PCR Detection system (Bio-Rad Laboratories, Inc., Hercules, CA, USA) following the standard protocol. The clinical samples were handled and processed in a Biosafety Laboratory-2 (BSL-2) at the national virology reference laboratory. Out of the 150 re-analysed samples, 70 had a Ct value less than or equal to 30 and were selected for whole genome sequencing, and others were discarded due to their high Ct values (i.e., >30).

The RNA extracts from these 70 samples were sequenced using the Illumina NextSeq 550 platform, San Diego, CA, USA, at the virology laboratory of EPHI during March 2023. Out of 70 SARS-CoV-2

sequence data, 63 sequences passed the bioinformatics quality control requirements (i.e., with genome coverage greater than 80%, no clustered mutations, and no misplaced stop codons) and were selected for further sequence analysis.

3.7. Data extraction and quality assurance

To address all three objectives:

A data extraction tool was developed based on the key indicators monitored by the Ministry of Health (MOH) and EPHI utilizing both paper-based and/or electronic medical records and SARS-CoV-2 testing records. The tool was used to gather the demographic information, clinical conditions, vaccination status, underlying conditions and travel history of the patients. One data collector was recruited based on their experience in managing the national COVID-19 testing data. Orientation was given to the data collector regarding the objective of the study, ethical issues, data collection procedure and quality of data. The quality of the data was checked throughout the data abstraction period by the investigator.

Every procedure used in the laboratory testing followed the instructions on the manufacturer's test kit package. The reagents, having their quality control to ensure the accuracy of the test, were the ideal situation for molecular testing. For every activity in the lab, a standard operating procedure (SOP) was developed, and a logbook was used to record every laboratory result. Reagent and sample storage were adequately carried out following the company recommendations. To avoid contamination, all necessary laboratory safety measures were implemented.

3.8. Data Entry and Analysis

The statistical analysis of each study objective was addressed as follows:

To explore the genetic variants of SARS-CoV-2 and to identify the conserved gene regions for diagnostic purpose:

The metadata was exported from the national database and converted into Excel format. The data were encoded into STATA v17 and descriptive analyses were performed. Raw sequencing data were obtained in FASTQ format and processed for variant calling and consensus sequence generation. First, the FASTQ sequences' quality was evaluated by the FastQC (v0.11.9) program, and then adapter sequences were trimmed from reads using Cutadapt software v.4.6 (Martin, M., 2011) to remove

adaptor sequences and poor quality reads. An average base quality of Q30 was used for trimming. The trimmed reads were then mapped against the SARS-CoV-2 WuhanHu-1 reference genome sequence (GenBank accession no. NC_045512.2) using the Burrows-Wheeler Alignment BWA mem v0.7.12 (Burrows-Wheeler Aligner, 2023).

The consensus sequence was then generated using Samtools version 1.12. Moreover, SARS-CoV-2 variants were called with locally installed Pangolin (daily updated version (O'Toole et al., 2021)) and mutations were identified with online Nextclade v3.8.2 Clade assignment, mutation calling, and sequence quality checks (Aksamentov et al., 2021). All viral genome sequences obtained from this study were submitted to the Global Initiative on Sharing All Influenza Data (GISAID) with accession IDs from EPI_ISL_19147618 to EPI_ISL_19302657 (GISAID, 2025) and National Center for Biotechnology Information (NCBI) repositories with accession numbers from PQ140559 to PP069552 (NCBI, 2024)

A. Computation of SNP distances in samples to the SARS-CoV-2 reference genome

SNP-dist v0.70 (GitHub, 2024) was used to determine the number of SNPs in the 63 sequences compared to the SARS-CoV-2 reference genome. The hCoV-19/Wuhan/WIV04/2019 sequence from Wuhan, China, which consists of 29,891 base pairs, was used as a reference genome obtained from the GISAID. The snp-dist tool calculates the SNP distance matrix from a dataset of multiple sequence alignments (MSA), where the sequences are of the same length. These distance matrices are commonly used in studies of infectious disease outbreaks. For this analysis, we only considered informative bases (A, T, C and G) and not N's. MSA was conducted using MAFFT v7.526 (K and Dm, 2013), employing the multiple alignment technique with the Fast Fourier Transform. The following parameters were applied: a gap opening penalty of 1.53 for group-to-group alignment, a gap extension penalty of 0.00 for group-to-group alignment, the blocks substitution matrix (BLOSUM) was selected, with a coefficient of 62, and the L-INS-I alignment method of MAFFT was chosen, known for its accuracy in alignments of ≤200 sequences. The T92+G+I model was used to calculate the transition/transversion bias (R).

B. Evolutionary divergence between sequences and overall sequence pairs

To estimate the evolutionary divergence between sequences as well as the average evolutionary divergence over all sequence pairs, MSA of the 63 sequences was performed using MAFFT v7.526 (K

and Dm, 2013). The output of the alignment was used to construct the phylogenetic tree via IQ-tree version v2.3.5-3 (Minh et al., 2020). All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 29,921 positions in the final dataset. The pairwise and overall mean (average) distances were used as proxies for evolutionary divergence.

For comparison purposes, the overall evolutionary divergence among all sequence pairs obtained from the GISAID uploaded during and before the study period was calculated. This analysis included the WIV04 genome and a total of 118 SARS-CoV-2 genome sequences from 62 countries. The sequences encompassed various variants, including former VOCs present in neighbouring countries and different regions worldwide (Table 3).

Table 3. The WIV04 genome, along with 118 SARS-CoV-2 genome sequences from 62 countries, were obtained from the GISAID.

Classifications	Variants	Name	First reported country
VOC	alpha	GRY (B.1.1.7+Q.*)	United Kingdome
VOC	Beta	GH/501Y.V2 (B.1.351+ B.1.351.2+ B.1.351+3)	South Africa
VOC	Gamma	GR/501Y.V3 (P.1+P.1*)	Brazil and Japan
VOC	Omicron	GRA (B.1.1.529+BA.*)	Botswana, South Africa, and Hong Kong
Variant of interest (VOI)	Lambda	GR/452Q.V1 (C.37+C.37.1)	Peru
VOI	Mu	GH (B.1.621+B.1.621.1)	Colombia
VOI		GRA (XBB.1.5+XBB.1.5.*)	Australia, India, and Bangladesh
VOC		GRA (XBB.1.16+XBB.1.1.16.*)	India
Variant under monitoring (VUM)		GRA (XBB+XBB.*, excluding XBB.1.5, XBB.1.16, XBB.1.9.1, XBB.1.9.2, XBB.2.3)	India
VUM		GRA (XBB.1.9.1+XBB.19.1.*)	Indonesia, Singapore, and Israel
VUM		GH/490R (B.1.640+B.1.640.*)	Congo and France. Except for the Gamma VOC

Note: these were sequences that we collected between 1 January and 31 July 2022 before our samples were collected between June, July and August 2022. There was a total of 30,150 positions in the final dataset.

C. Phylogenetic analysis of SARS-CoV-2 genome sequences

The IQ-TREE v2.3.5.3 was used to compute the phylogenetic inference of the 63 SARS-CoV-2 genome sequences (Minh et al., 2020). In addition to these 63 sequences, after aligning the consensus sequences

obtained in this research it was selected the top 118 SARS-CoV-2 genomes and were downloaded from GISAID for phylogenetic analysis. First, an alignment analysis using MAFFT v7.526 was performed (Katoh and Standley, 2013). The MSA, which had 30,058 nucleotide sites, was then imported to IQ-TREE.

The phylogeny involved determining the best-fit DNA substitution model using Model Finder, which tested 88 DNA models. The initial parsimony tree was created by the phylogenetic likelihood library (PLL). The model chosen according to the Bayesian Information Criterion (BIC) score (115,726.5435) was the TIM+F+I+G4. The model of rate heterogeneity was as follows: proportion of invariable sites: 0.8736 and Gamma shape alpha: 0.9901 with 4 categories. The rate parameter (R) was as follows: A-C: 1.0000, A-G: 1.6270, A-T: 0.5506, C-G: 0.5506, C-T: 3.9021, and G-T: 1.0000.

A bootstrap analysis using the UFBoot2 method with 1000 replicates and ML heuristic method was performed. Additionally, branch lengths were optimized through ML on the original alignment. The initial trees for the heuristic search were generated using BioNJ algorithms. From these analyses, we selected the phylogenetic tree with the highest log-likelihood value (-55,910.033). The resulting phylogenetic trees were saved in Newick format and visualized using the Interactive Tree of Life interface (iTOL, 2023). Statistical analysis of the demographic and clinical information was performed using STATA v17. Frequency distribution and tabulation were performed to show the relationship between outcome and independent variables.

D. Conserved gene regions identification

For conserved gene regions, sequence conservation was assessed using Jalview (Waterhouse et al., 2009), where regions with low entropy and high identity were considered highly conserved. Additionally, mutations and variant-specific changes were annotated using the Nextclade (Nextclade, 2023). Following the identification of conserved regions, B-cell linear epitope prediction was conducted using BepiPred 2.0, hosted by the IEDB Analysis Resource (Jespersen et al., 2017).

Epitopes were screened for antigenicity using VaxiJen v2.0 (Doytchinova and Flower, 2007), with a threshold of 0.5, and those scoring above this were further analyzed for allergenicity and toxicity using AllerTOP (Dimitrov et al., 2014) and ToxinPred (Gupta et al., 2013). To ensure epitope specificity, potential cross-reactivity with common human coronaviruses (e.g., HCoV-229E, HCoV-OC43, and

HCoV-NL63) was evaluated using BLASTp (Altschul et al., 1990). For structural validation, three-dimensional models of the SARS-CoV-2 nucleocapsid (N) protein were obtained using AlphaFold predictions (Jumper et al., 2021).

The predicted epitopes were mapped onto the 3D structure using PyMOL (PyMOL, 2024) and UCSF Chimera (Pettersen et al., 2004) to evaluate surface accessibility and structural integrity. Solvent accessibility and flexibility of epitopes were examined using DynaMut (Rodrigues et al., 2018). Molecular docking simulations were then performed using ClusPro (Kozakov et al., 2017) to assess potential binding interactions between monoclonal antibody fragments (Fabs) and the predicted epitopes.

***To evaluate the performance of RT-qPCR on SARS-CoV-2 detection without the RNA extraction process*:**

The data were summarized as mean ± Standard Error of the Mean (SEM) and r^2 (coefficient of determination). The D'Agostino & Pearson test with a 95% confidence interval was used to test for normality. Confidence intervals (CI) for sensitivity, specificity, positive/negative predictive values, and accuracy were calculated using R-studio v.3.6.0. The Wilcoxon signed-rank test was used to compare the two groups. The correlation between the Ct-values of the two techniques was determined using the two-tailed Pearson's correlation coefficient using STATA v17(stata Corp LP, College Station, TX, USA). A p-value of less than 0.05 indicated statistical significance. The Ct values were grouped based on viral load (Ct <20, 20-35, and >35), and performance characteristics in each group were compared.

3.9. Ethical considerations

This study was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were obtained from the Institutional Review Board (IRB) of the Aklilu Lemma Institute of Health Research, Center for Pathobiology Research, Addis Ababa University, under protocol number ALIPB IRB/52/2013/21 dated 19 April 2021. A permission letter was obtained from the EPHI to use stored samples from repositories. Given the retrospective nature of the study and the use of anonymized samples, the requirement for individual patient consent was waived by the Institutional Review Board.

CHAPTER 4: RESULTS

4.1. Identifying the genetic variants of SARS-CoV-2 circulating in different parts of Ethiopia during the fifth wave of the pandemic.

4.1.1. Demographics and clinical characteristics of the study subjects

Out of 150 samples collected, 70 samples which had Ct-value of 30 and below were sequenced. Out of the 70 samples sequenced, 63 samples passed the bioinformatics quality control. The age range of the study participants was from 14 to 75 years (median age of 34 years). Males accounted for 70% (44/63). The majority of the sequenced samples were collected from Addis Ababa (70%; 44/63). In terms of clinical data, 70% (44/63) of the study participants exhibited signs and symptoms of cough, fever and anosmia (Table 4).

Table 4. Socio-demographic and base line data of the study participants

Variables		Frequency (%)
Sex	Female	19 (30%)
	Male	44 (70%)
Age group	<=20	8 (13%)
	21-30	20 (32%)
	31-40	13 (21%)
	41-50	10 (16%)
	51-60	4 (6%)
	>=61	8(13%)
Resident	Oromia	15 (24%)
	Amhara	3 (5%)
	Addis Ababa	44 (70%)
	SNNPR	1 (1%)
Sign and symptoms	No	19 (30%)
	Yes	44 (70%)
Comorbidity	No	36 (57%)
	Yes	27 (43%)
Travel history	No	55 (87%)
	Yes	8 (13%)
Vaccination status	Not vaccinated	38 (60%)
	Vaccinated	25 (40%)
Reason for testing	Contact with cases	10 (16%)
	Follow-Up	2 (3%)
	Suspect	43 (68%)
	Travel Purpose	8 (13%)
Variants of concern	Delta	2 (3%)
	Omicron	61 (97%)

4.1.2. Sequencing and sequence analysis

The majority of the sequences (99.8%) had high quality, with gap sizes ranging from 20 to 114 nt, accounting for approximately 0.4% of the total sequence. On average, each genome had 69.21 single-nucleotide variations (SNVs) in different genes of SARS-CoV-2. The high number of single-nucleotide mutations per clades of SARS CoV-2 was identified in 22A and 21K **(Figure 7).**

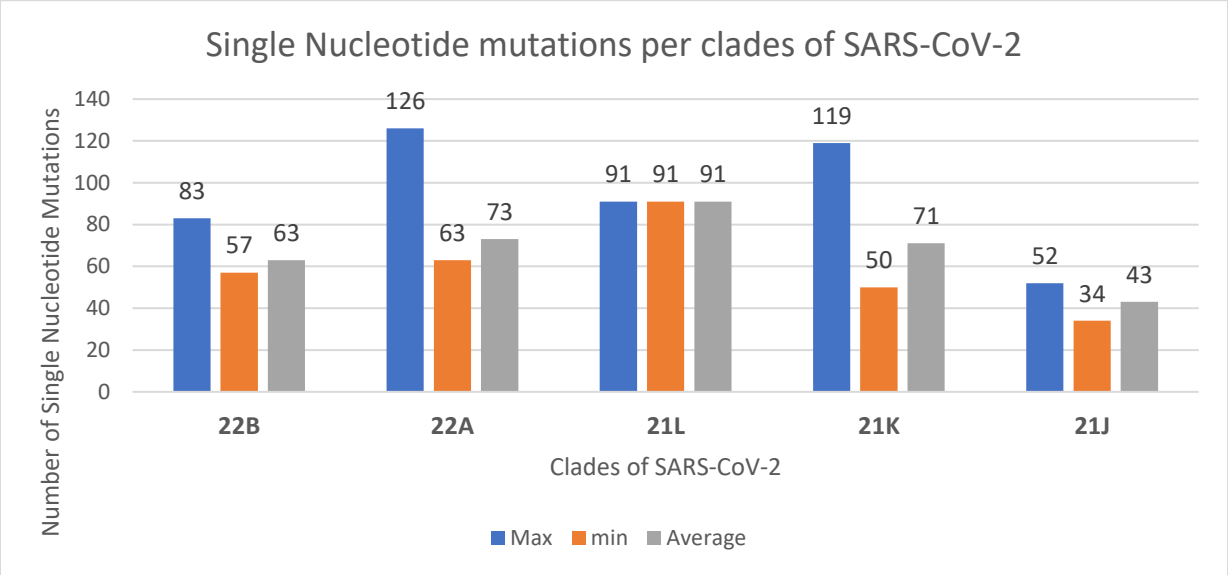


Figure 7.Number of single-nucleotide mutations per clades of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia, 2022.

This is an example of SNVs in different gene position of the SARS-CoV-2 using Nextclad variant analysis Right in Figure 8. This analysis also detected an average of 198 nucleotide deletions, 68 substitutions, and 1 insertion in different genes of SARS-CoV-2 (left in figure 8)

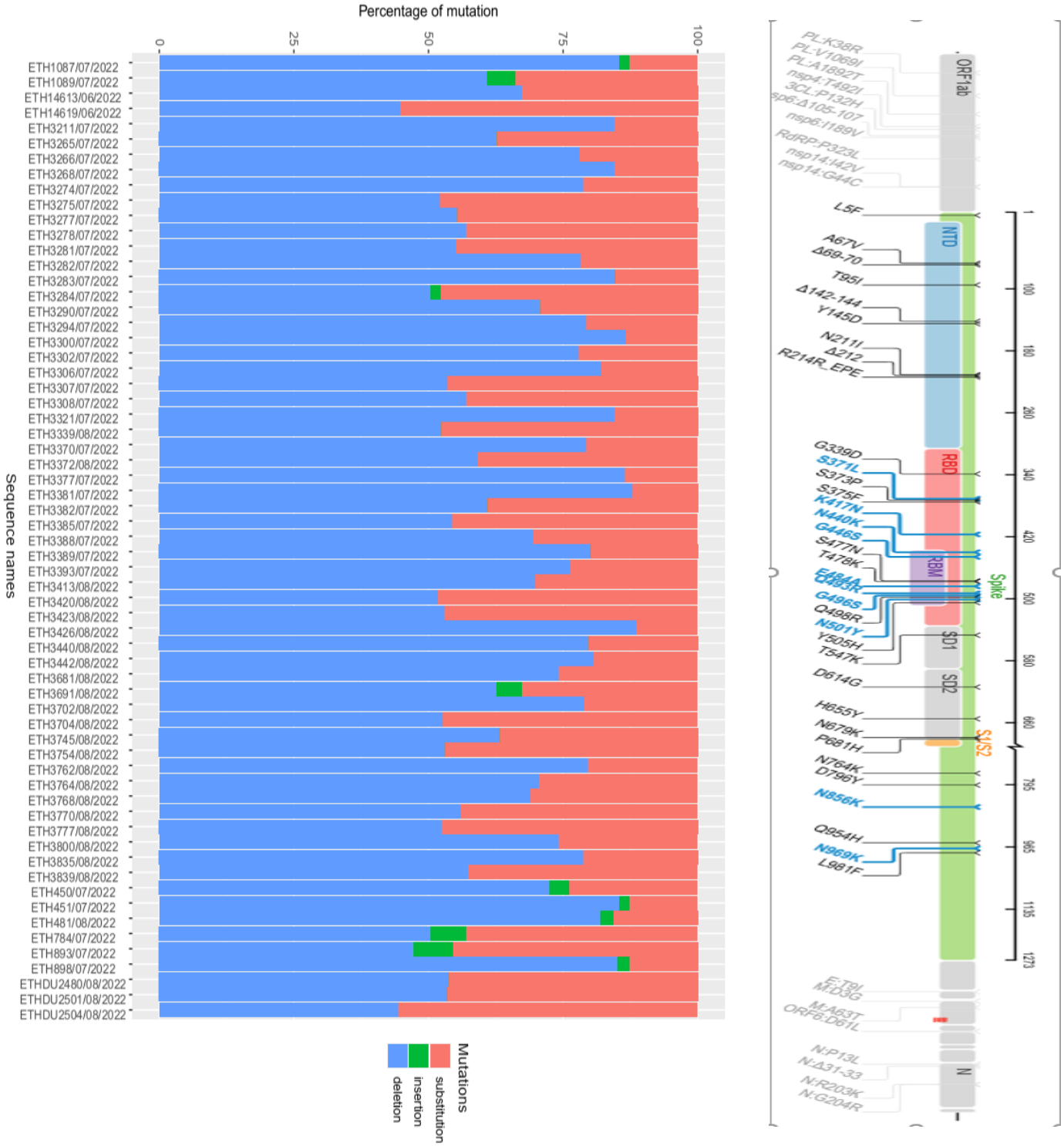


Figure 8. SARS CoV-2 nucleotide substitution, insertion and deletion after comparison of each sequence with the references.

These SARS-CoV-2 nucleotide mutations were observed in different positions of S, M, and N genes and genes coding for non-structural proteins (nsp1, nsp6, nsp9, nsp10, nsp13, ORF3a, and ORF7a) (Table 5). Among them, 371 nucleotide deletions were observed in S gene in the position of 25-27, 69-70, 124-144, 157-158, 212, 374-376, 415-432, 446-464 and 758-767; followed by 177 nucleotide deletion N in the position of 31-33; and 168 deletions in nsp6 gene in the position of 105-108.

Table 5. SARS-CoV-2 nucleotide mutations in different gene positions during the fifth wave of the pandemic in Ethiopia.

Gene	Position	Mutation
S	97, 421-426, 445, 455, 457, 462, 465, 486, 500, 658, 720, 737-742, 744-746, 763-765	S:K97R, S:K97R, S:Y423L, S:K424stop, S:L425I, S:P426T, S:Y421S, S:V445A, S:L455S, S:L455S, S:R457G, S:K462Q, S:E465X, S:F486V, S:F486V. S:F486V, S:T500S, S:N658S, S:N658S,S:I720X, S:D737V, S:C738V, S:T739Q, S:M740C, S:Y741T, S:I742F, S:G744W, S:D745stop, S:S746F, S:N764E, S:L763I, S:R765S, S:N764X, S:R765X.
M	3, 83, 151, 154, 164, 167-169, 198-199, 203-206, 212-213.	M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:D3N, M:A83T, M:I151M, M:I151M, M:I151M, M:H154D, M:L164M, M:Y178X, M:E167A, M:E167K, M:Y179stop, M:R198G, M:Y199C, M:L206stop, M:N203X, M:Y204F, M:K205A, M:N203T, M:Y204N, M:K205A, M:L206stop, M:S212X, M:S212X, M:S213X, M:S213X, M:S212X, M:S212X, M:S212R, M:S213X, M:S213X, M:S212X.
N	379	N:T379N
Nsp1	7	nsp1:G7X
Nsp8	50, 74, 167 and 177	nsp8:D50N, nsp8:A74S, nsp8:A74S, nsp8:A74S, nsp8:A74S, nsp8:A74S, nsp8:A74S, nsp8:V167F, nsp8:S177L
Nsp9	91	nsp9:I91V
Nsp10	75	nsp10:L75M
Nsp13	512	nsp13:V521I
ORF3a	92, 213, 275	ORF3a:S92P, ORF3a:Q213R, ORF3a:L275F
ORF7a	48	ORF7a:P48A
Plpro	684, 906, 932, 1234	PLpro:I684V, PLpro:E906A, PLpro:V932L, PLpro:V1234E
RdRp	930	RdRP:V930I

This study identified several mutations in the PLpro (Papain-like protease), including K38R, V1069I, A1892T, D821N, E906A, T24I, G489S, I684V, S370L, V932L, P985S, T703I, A1179V, P822L, A488S, P1228L, and P1469S. Additionally, the study explores the presence of the P132H mutation in the 3CLpro (Main protease) and the P323L mutation in RdRP.

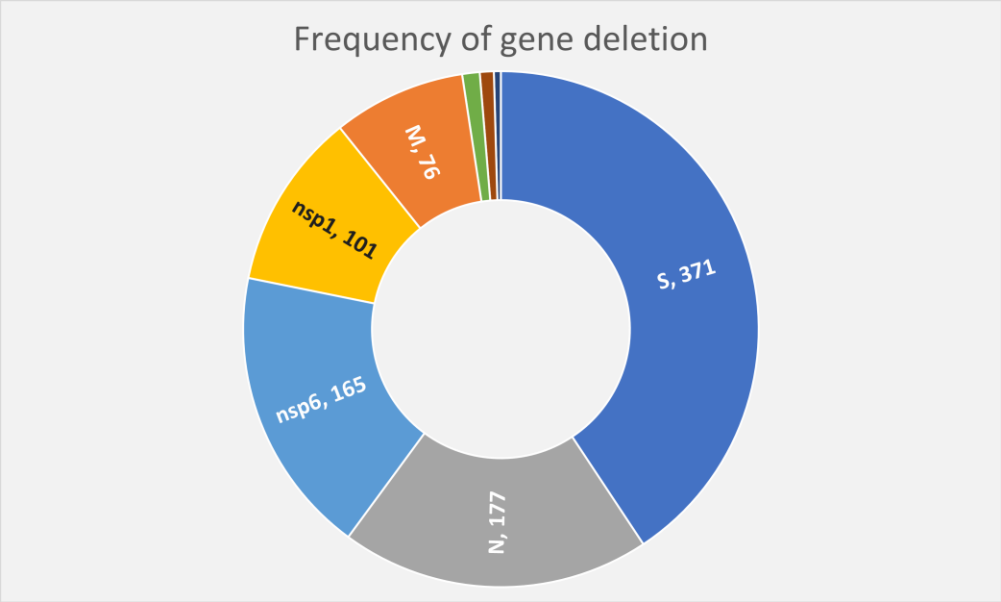


Figure 9. Frequency of gene deletion of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia, 2022

In this study, the transition/transversion bias (R) was 1.693, indicating that transitions occur more frequently than transversions. The base substitution rates (r) for each nucleotide pair were as follows: A=>T (0.06), A=>C (0.03), A=>G (0.07), T=>A (0.05), T=>C (0.17), T=>G (0.03), C=>A (0.05), C=>T (0.29), C=>G (0.03), G=>A (0.11), G=>T (0.06), and G=>C (0.03). These results show that Thymine (T) has the highest substitution rate with cytosine (C) (0.17) and the lowest with guanine (G) (0.03). Cytosine (C) has the highest substitution rate with thymine (T) (0.29), while its rates with adenine (A) and guanine (G) are relatively low (both 0.03). Adenine (A) and guanine (G) have generally similar substitution rates across different nucleotides, with A => G having the highest rate among adenine substitutions and G => A having the highest rate among guanine substitutions. In this study, the average Ti/Tv ratio was 1.9734, ranging from 1.5862 to 2.4737, and the most common nucleotide change was the G>A transition (Figure 10)

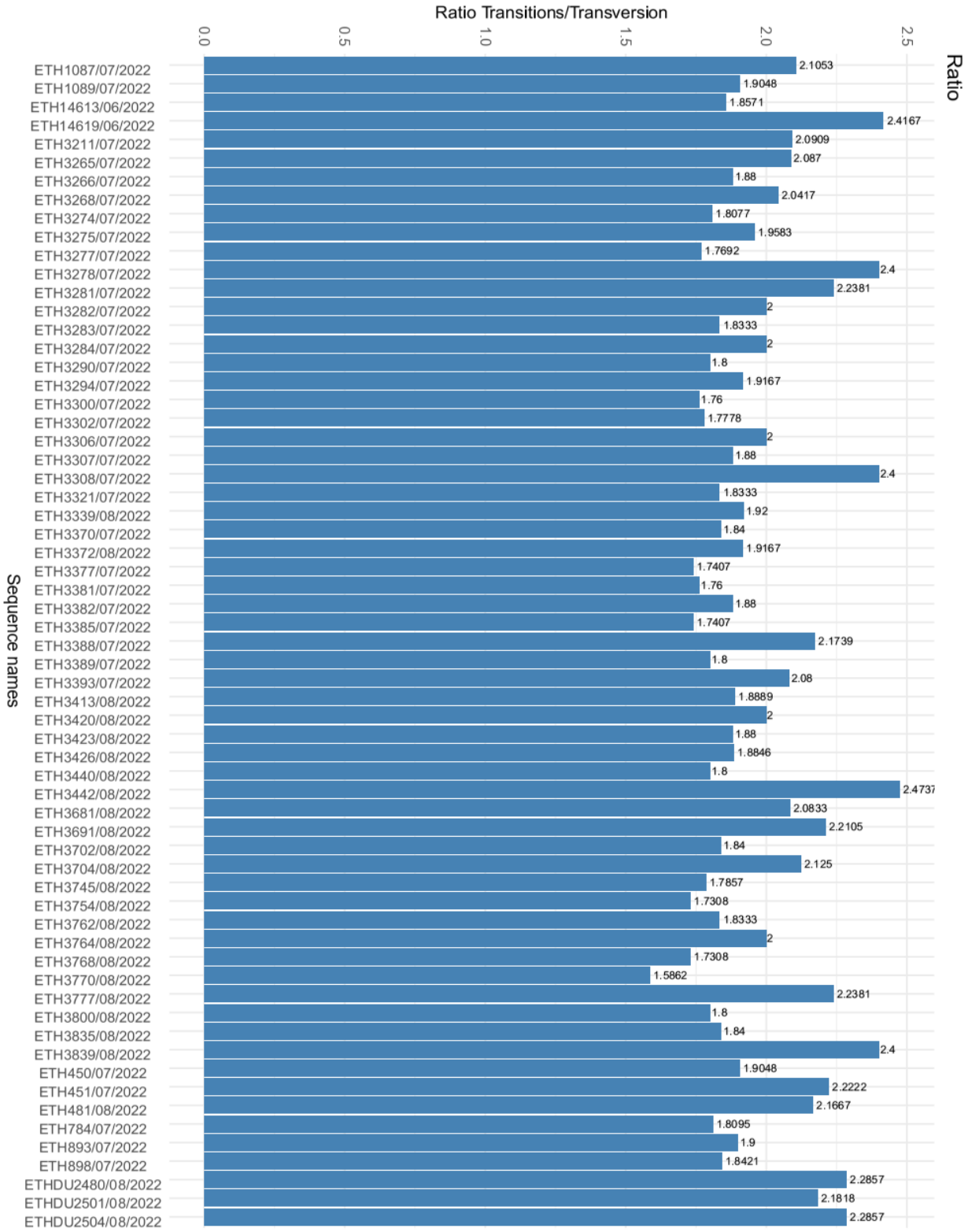


Figure 10. Transition-Transversion ratio of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia, 2022

4.1.3. SARS-CoV-2 genetic variants

The study found that omicron was the dominant variant (97%, 61/63) identified in the total sample sequenced during the fifth wave of the pandemic in selected regions of Ethiopia, followed by the Delta variant (3%, 2/63). All of the identified cases of the Omicron variant belonged to four age groups (<20, 21-30, 31-40, 41-50, 51-60, and >60 years old). In the age groups of 21-30 and 41-50 years old, Delta accounted for 5% and 10%, respectively, of the identified cases, and the remaining infections were caused by the Omicron variant. In addition, all of the 25 individuals who were vaccinated before testing were infected with the Omicron variant. Out of the 38 non-vaccinated study participants, 95% (36/38) were infected with Omicron, and 5.3% (2/38) were identified with the Delta variant.

Table 6. Frequency Distribution of SARS-COV-2 Variants with demographic & Clinical information of patients during the fifth wave of the pandemic in Ethiopia, 2022.

			Variants of SARS COV-2 virus	
			Delta Variant	Omicron Variant
Sex	Female	Number	1	18
		%	5.3	94.7
	Male	Number	1	43
		%	2.3	97.7
Vaccination	No	Number	2	36
		%	5.3	94.7
	Yes	Number	0	25
		%	0.0	100.0
Sign and Symptom	No	Number	0	19
		%	0.0	100.0
	Yes	Number	2	42
		%	4.5	95.5
Travel History	No	Number	2	53
		%	3.6	96.4
	Yes	Number	0	8
		%	0.0	100.0
Ethiopian Nationality	No	Number	0	1
		%	0.0	100.0
	Yes	Number	2	60
		%	3.2	96.8
Age Group	<=20	Number	0	8
		%	0.0	100.0
	21-30	Number	1	19
		%	5.0	95.0
	31-40	Number	0	13
		%	0.0	100.0
	41-50	Number	1	9
		%	10.0	90.0
	51-60	Number	0	4
		%	0.0	100.0
	> 60	Number	0	8
		%	0.0	100.0

According to the Nextstrain classification, the dominant clades identified during the study period were 22A (64%, 40/63) and 22B (18%, 11/63), followed by 21K (Omicron; 14%, 9/63), 21J (Delta, 3.2%, 2/63), and 21L (Omicron; 1.6%, 1/63). Additionally, sub-lineages BA.4.1 (30.2%, 19/63) and BA.4.1.1 (28.5%,18/63) were more prevalent than BA.1.1 (9.5%, 6/63), BA.5.2 (7.9%, 5/63), BF.2 (6.4%, 4/63), and BA.4 (4.8%, 3/63). Interestingly, 90% of the sub-lineages belonged to different strains within the BA lineage, which represents a specific genetic variant or group of closely related variants of the SARS-CoV-2 virus. Additionally, (6.4%, 4/63) of the samples were part of the BF lineage, while both the AY and B lineages each accounted for 1% of the samples.

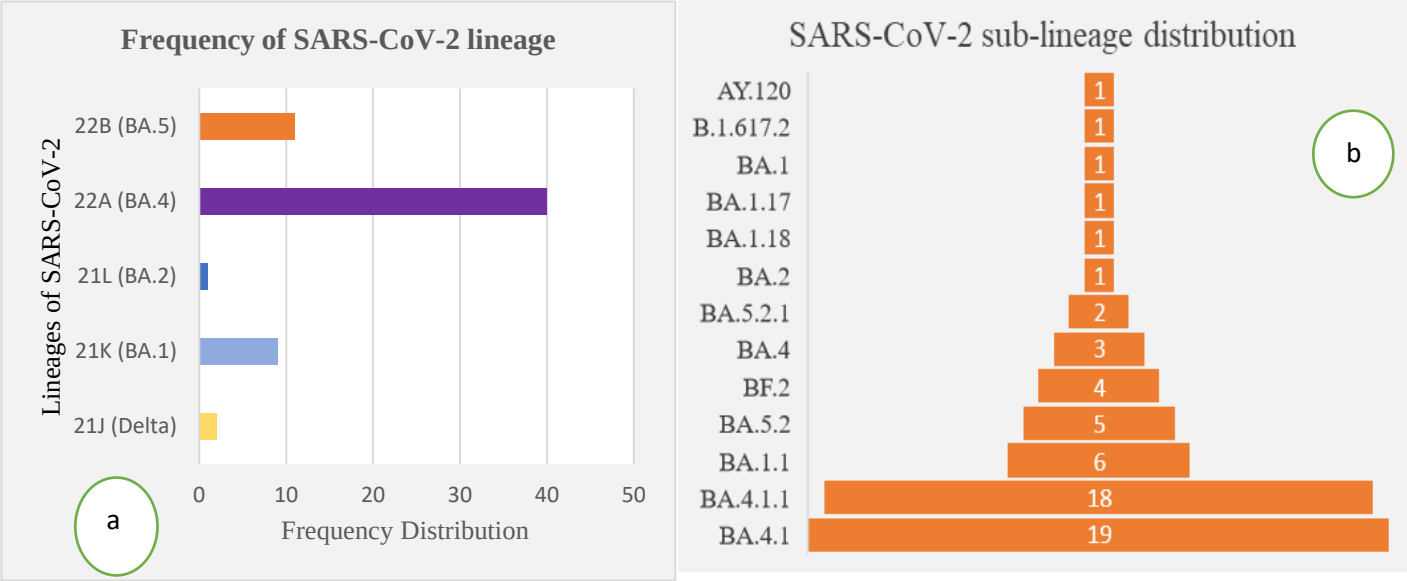


Figure 11. Frequency distribution of SARS-CoV-2 lineage (a) and sub-lineages (b) demonstrated that the relative prevalence of SARS-CoV-2 clades circulated during the fifth wave of the pandemic in Ethiopia from June - August 2022.

Almost a comparable number of vaccinated and unvaccinated participants were infected with the BA.4.1 and BA.4.1.1 sub-lineage of SARS-CoV-2 (Figure 12).

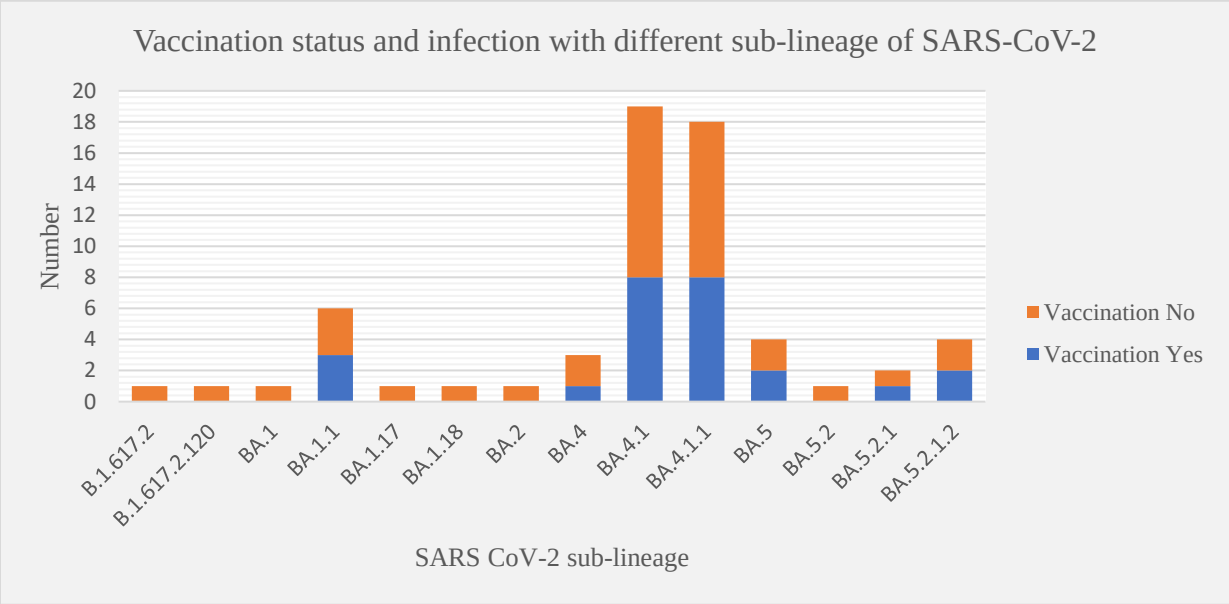


Figure 12. SARS-CoV-2 vaccination status and infection with different sub- lineage of SARS CoV-2 identified during the fifth wave of the pandemic in Ethiopia.

4.1.4. Evolutionary divergence and phylogenetic analysis of SARS-CoV-2

The evolutionary analysis among sequenced data for this study showed there is evidence of a phylogenetic relationship in some of the sequences (Figures 13). The evolutionary divergence between sequences ranges from 0.00 to 3.45x10-3. Looking at the broader analysis, the average evolutionary divergence across all sequence pairs is similarly low. For sequences from Ethiopia, the overall distance is calculated to be 9.78x10-4, with a standard error of 8.47x10-5. In comparison, when considering selected global sequences, the overall distance slightly increases to 1.54x10-3, with a standard error of 1.19x10-4. In addition, the maximum likelihood phylogenetic tree from our study, comprising 63 SARS-CoV-2 genomes, demonstrated seven main dominant clades in the viral population.

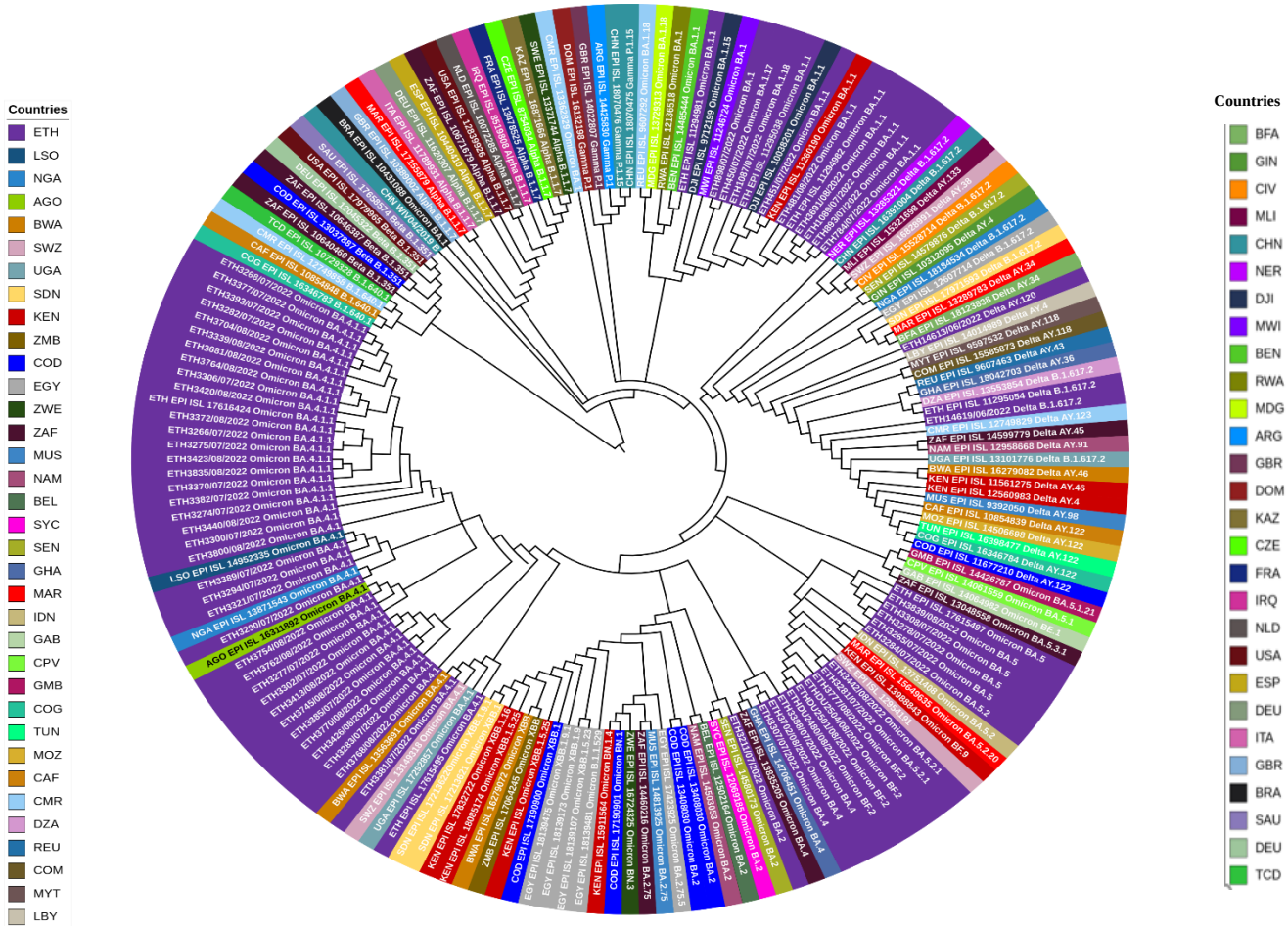


Figure 13. Analysis of the Phylogenetic relatedness of SARS CoV-2 identified in Ethiopia during the fifth wave of the pandemic, 2022 with selected global sequences.

Note: The evolutionary history was inferred from bootstrap phylogeny test (1,000 replicates) using ML phylogenetic reconstruction based on Tamura 3-parameter (T92) model with discrete Gamma distribution and an assumption that a certain fraction of sites is evolutionarily invariable (T92+G+I). All the sequences were rooted on the SARS-CoV-2 reference genome (WIV04), shown with a **bold** broken line for node. The 63 SARS-CoV-2 genome sequences are labelled in purple. Alpha, Beta, Gamma, Delta, and Omicron variants of concern are labelled in light blue, green, gold, red, and blue, respectively. Other variants are in black. Countries from where the SARS-CoV-2 variants were collected are indicated by a three-letter code: AGO: Angola, ARG: Argentina, BEL: Belgium, BEN: Benin, BFA: Burkina Faso, BRA: Brazil, BWA: Botswana, CAF: Central African Republic, CHN: China, CIV: Ivory Coast, CMR: Cameroon, COD: Democratic Republic of Congo, COG: Congo Republic, COM: Comoros, CPV: Cape Verde, CZE: Czech Republic, DEU: Germany, DJI: Djibouti, DOM: Dominican Republic, DZA: Algeria, EGY: Egypt, ESP: Spain, ETH: Ethiopia, FRA: France, GAB: Gabon, GHA: Ghana, GIN: Guinea, GMB: Gambia, GBR: United Kingdom, IDN: Indonesia, IRQ: Iraq, ITA: Italy, KAZ: Kazakhstan, KEN: Kenya, LBY: Libya, LSO: Lesotho, MAR: Morocco, MDG: Madagascar, MLI: Mali, MOZ: Mozambique, MWI: Malawi, MUS: Mauritius, MYT: Mayotte, NAM: Namibia, NER: Niger, NGA: Nigeria, NLD: Netherlands, REU: Réunion, RWA: Rwanda, SAU: Saudi Arabia, SEN: Senegal, SDN: Sudan, SYC: Seychelles, SWE: Sweden, SWZ: Swaziland (Eswatini), TCD: Chad, TUN: Tunisia, UGA: Uganda, USA: United States of America, ZAF: South Africa, ZMB: Zambia, and ZWE: Zimbabwe.

4.2. Exploring Conserver gene regions of SARS-CoV-2 for monoclonal-antibody-based diagnostic tool development.

4.2.1. Target gene regions identification

Mutation profiling using Nextclade highlighted widespread mutations in the spike (S) gene (e.g., D614G, N501Y, P681R, and E484K), indicating huge genetic variability across VOCs. Due to this variability, the S gene was considered less suitable for stable diagnostic targeting (Figure 14).

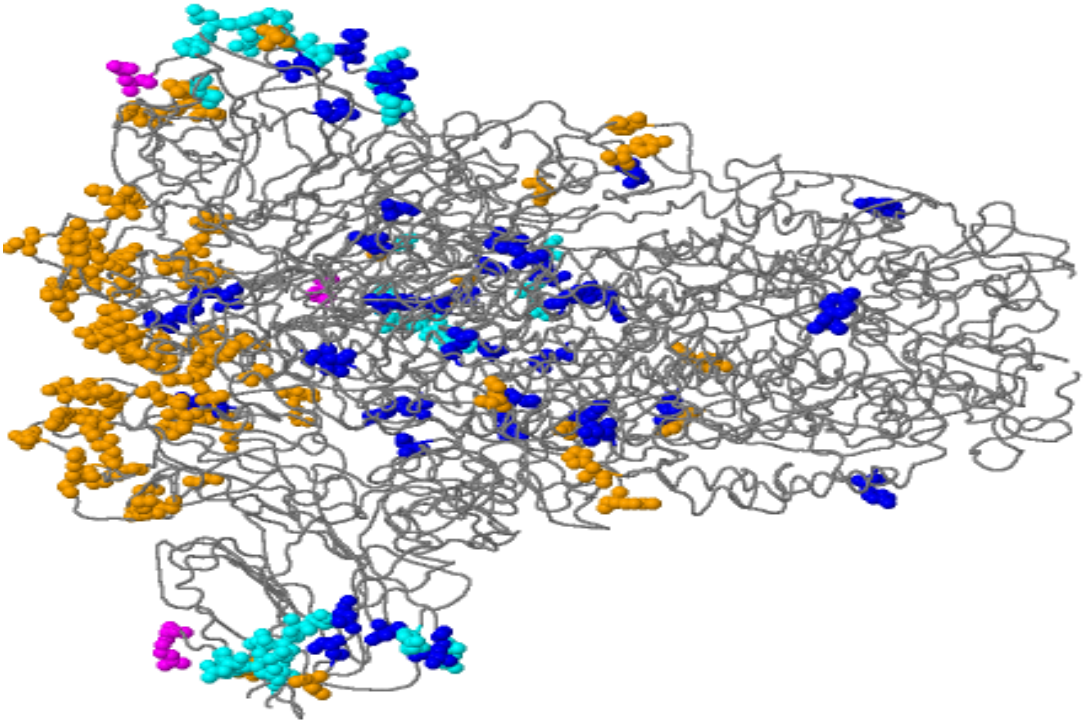


Figure 14. Spike glycoprotein (PDB: 6acc, EM 3.6 Angstrom) with RBD in down conformation
Sources: https://www.epicov.org/epi3/frontend#1a1eca

In contrast, analysis of the N gene showed two major conserved regions: the RNA-binding domain (residues 46-174 and dimerization domain (residues 247-364). Multiple sequence alignment using MAFFT showed that these regions maintained over 98% sequence identity across Alpha, Beta, Delta and Omicron lineages. Entropy analysis using Jelview further confirmed these findings, with low entropy score ranging from 0.0 to 0.2 in N gene compared to higher entropy score in S (1.1-1.2), M (0.3-0.5) and E (0.2-0.3) genes. Mutation frequency analysis using R demonstrated that the N gene

accumulated only 1 to 3 non-synonymous mutations across samples, which reinforces its evolutionary stability and diagnostic potential.

4.2.2. B-cell epitope prediction and antigenicity screening

Subsequent B-cell epitope prediction using BepiPred 2.0 identified 17 potential linear epitopes from the conserved regions of the N protein. Eight of those exhibited antigenicity scores above 0.5 using VaxiJen v2.0, indicating good potential as a diagnostic target. Among them, three epitopes, RRGPEQTQGNFG (residues 152-163), SSSRGTS (66-72) and DPNFKDQVILL (330-341), were prioritized for downstream analysis due to their high antigenicity scores of 0.81, 0.77 and 0.71, respectively, and their location within or near conserved functional domains.

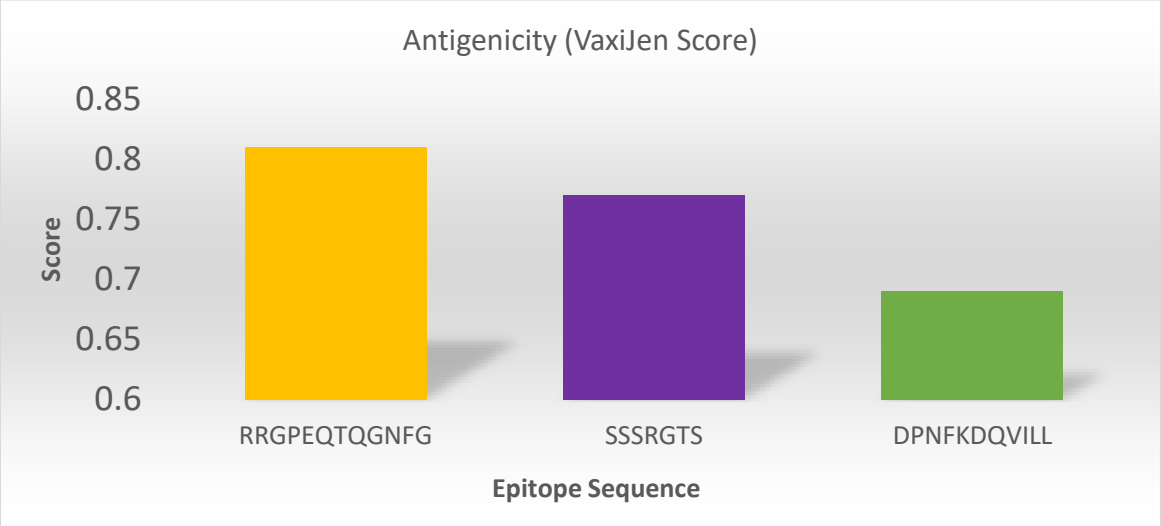


Figure 15. Antigenicity score of the three epitopes

4.2.3. Allergenicity, toxicity and cross-reactivity filtering

All eight antigenic epitopes were evaluated for safety, and five were predicted to be non-allergenic using AllerTOP v2.0 and non-toxic using ToxinPred. To assess the risk of cross-reactivity with endemic human coronaviruses, BLASTp searches against HCoV-229E, HCoV-OC43, HCoVNL63, and HCoV-HKU1 were conducted. Three epitopes (the above three candidates) showed low sequence identity, i.e. <30% and E-values of >0.01. which indicates low risk of cross reactivity and confirms their diagnostic specificity.

4.2.4. Structural mapping and solvent accessibility

For structural validation, the 3D model of the SARS-CoV-2 N protein was obtained from AlphaFold. The three selected epitopes were mapped onto the model using PyMOL and UCSF Chimera. All were found to be surface exposed, and two (RRGPEQTQGNFG and SSSRGTS) were located within flexible coil regions. This location has properties that enhance epitope accessibility and immunogenicity.

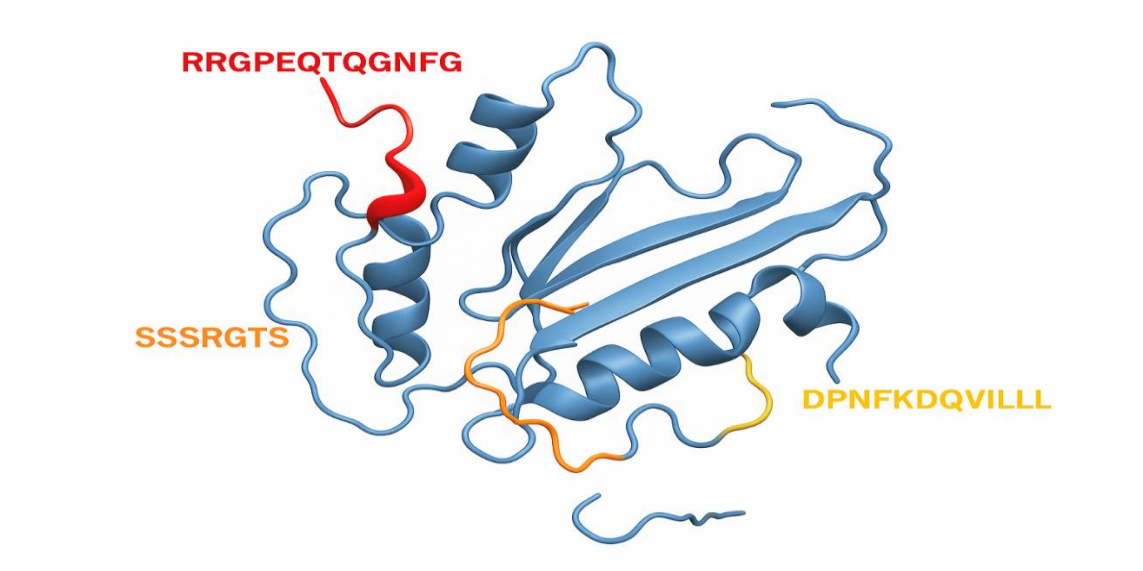


Figure 16. 3D Visualization of SARS-CoV-2 Nucleocapsid
Sources: AI generated, 2025

Solvent accessibility and conformational flexibility analysis conducted by DynMut confirmed that the top candidate epitopes, RRGPEQTQGNFG (residues 152-163), had a solvent accessible surface area (SASA) of 78.4 Å² and a ΔΔG of -0.5 kcal/mol. This indicates favorable stability and surface exposure.

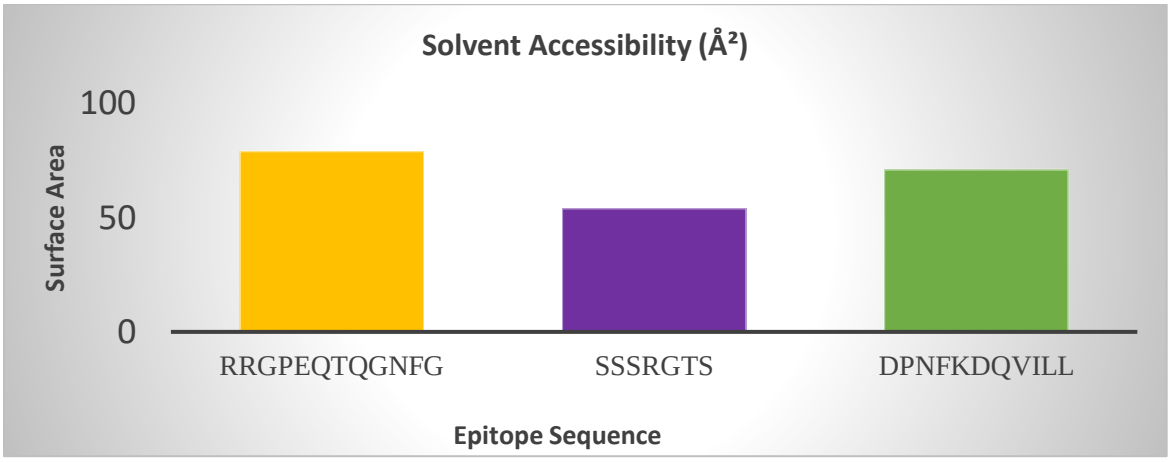


Figure 17. Solvent accessibility of the three epitopes

4.2.5. Molecular docking simulations

Finally, molecular docking simulations were conducted using ClusPro to evaluate the binding interaction between the predicted epitopes and monoclonal antibody fragments (Fabs). The docking results indicated strong affinities, with ClusPro weighted score of -917.4 for RRGPEQTQGNFG, -812.6 for SSSRGTS and -865.2 for DPNFKDQVILL. The RRGPEQTQGNFG epitope formed stable interactions via hydrogen bonds with key residues (R154, K157, D161), supporting its binding potential and utility in diagnostic design.

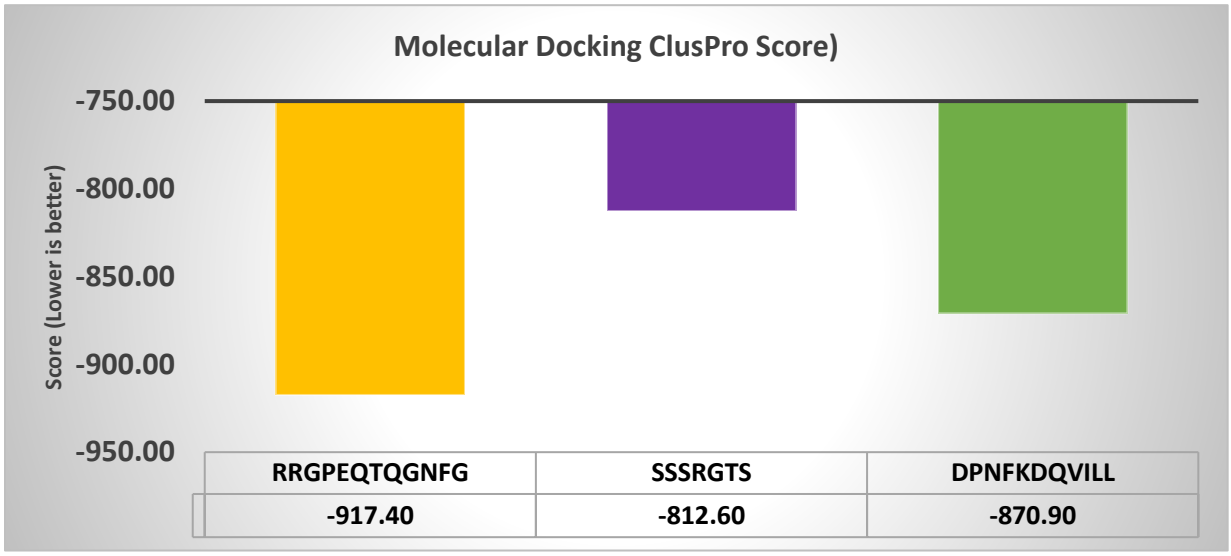


Figure 18. Molecular Docking using ClusPro score of the sequence epitopes

Table 7.

Overall analysis results showed the identification and selection of conserved N gene regions of SARS-CoV-2.

Epitope sequence	Position (N protein)	VaxiJen Score	allergenicity	Toxicity	Cross reactivity	Solvent Accessibility	cluePro Score
RRGPEQTQGNFG	152-163	0.81	Non-allergen	Non-toxic	None	High	-917.4
SSSRGTS	66-72	0.77	Non-allergen	Non-toxic	low	Moderate	-812.6
DPNFKDQVILL	330-341	0.69	Non-allergen	Non-toxic	None	High	-870.9

Note: These results demonstrate that computational epitope prediction, combined with antigenicity screening, structural validation, and docking studies, can effectively identify conserved, non-allergenic, and surface-accessible B-cell epitopes that may serve as candidates for affordable diagnostic or immunological tools targeting SARS-CoV-2.

4.3. Evaluation of Extraction Free RT-qPCR Method for SARS-CoV-2 Detection

4.3.1. Comparative analysis of EFDH and extraction-based detection methods of SARS-CoV-2

The EFHD method successfully detected 85.8% (163/190) of the positive samples and 100% (110/110) of the negative samples. However, there were 15 false negative results, most of which had an average Ct value greater than 36 based on the extraction-based method. Additionally, the EFDH method revealed 12 invalid results. The extraction-based method resulted in a mean Ct-value of 24.1 (95% CI: 23.3-24.8) with a Standard Error of the Mean (SEM) of 0.4. This method covers a range of Ct-values from 16 to 34.1. On the other hand, the extraction-free (EFHD) method had an average CT value of 27.7 (95% CI: 26.8-28.7), with a SEM of 0.5 (annex 6).

The true prevalence among study participants was 63% (190/300) (95% CI: 58–69%). The apparent prevalence, which represents the proportion of true positives out of the total tested, was 53% (163/300) (95% CI: 49–60%). A two-by-two table showing the summary of the performance of two methods, i.e. extraction-based detection versus EFDH detection methods is presented below (Table 8).

Table 8. Two by two table for comparing extraction-based standard method and extraction-free diluted and heated (EFDH) methods.

	Extraction-Based Method		
Extraction-Free Method (EFDH)	**Positive**	**Negative**	**Total**
Positive	163	0	163
Negative	27	110	137
Total	190	110	300

Pearson chi square (p chi) < 0.001.

4.3.2. Diagnostic performance of extraction-free method

The EFDH method showed an overall sensitivity and specificity of 86% (95% CI: 80–90%) and 100% (95% CI: 97–100%), respectively. The positive predictive value was 100% (95% CI: 98–100%), while the negative predictive value was 80% (95% CI: 74–85%). The negative likelihood ratio (NLR), which measures the ability of a test to rule out a disease, was found to be 0.14 (95% CI: 0.1–0.2). The study

also revealed an overall agreement (accuracy) of 91% (95% CI: 87–94%) for the new method, with a kappa value of 0.82 (p-value < 0.001) (Table 9).

Table 9. Performance characteristics of extraction-based and extraction-free heated SARS-CoV-2 detection method, Ethiopia.

Statistic	**Value**	**95% CI***
Apparent prevalence	53%	49% - 60%
True Prevalence	63%	58% - 69%
Sensitivity	86%	80% - 90%
Specificity	100%	97% - 100%
Negative Likelihood Ratio	0.14	0.1 - 0.2
Positive Predictive Value	100%	98% - 100%
Negative Predictive Value	80%	74% - 85%
Accuracy/Agreement/	91%	87% - 94%

**95% Confidence Interval*

The correlation between the two methods is represented as $Y = -0.12 + 1.18x$, $R^2 = 0.96$, and $p = 0.001$, which shows a very strong and significant linear relationship between extraction-based and extraction-free detection methods. For every unit increase in the Ct value by the extraction-based detection methods, the Ct value by the extraction-free detection method increases by approximately 1.18 units (Figure 19). The Wilcoxon signed-rank test results (z = 11.6, p-value = 0.0001) this showed that the two methods do not yield equivalent results, even though they are correlated. It appears that the extraction-based method generally produces lower Ct values compared to the extraction-free (EFDH) method across all observations.

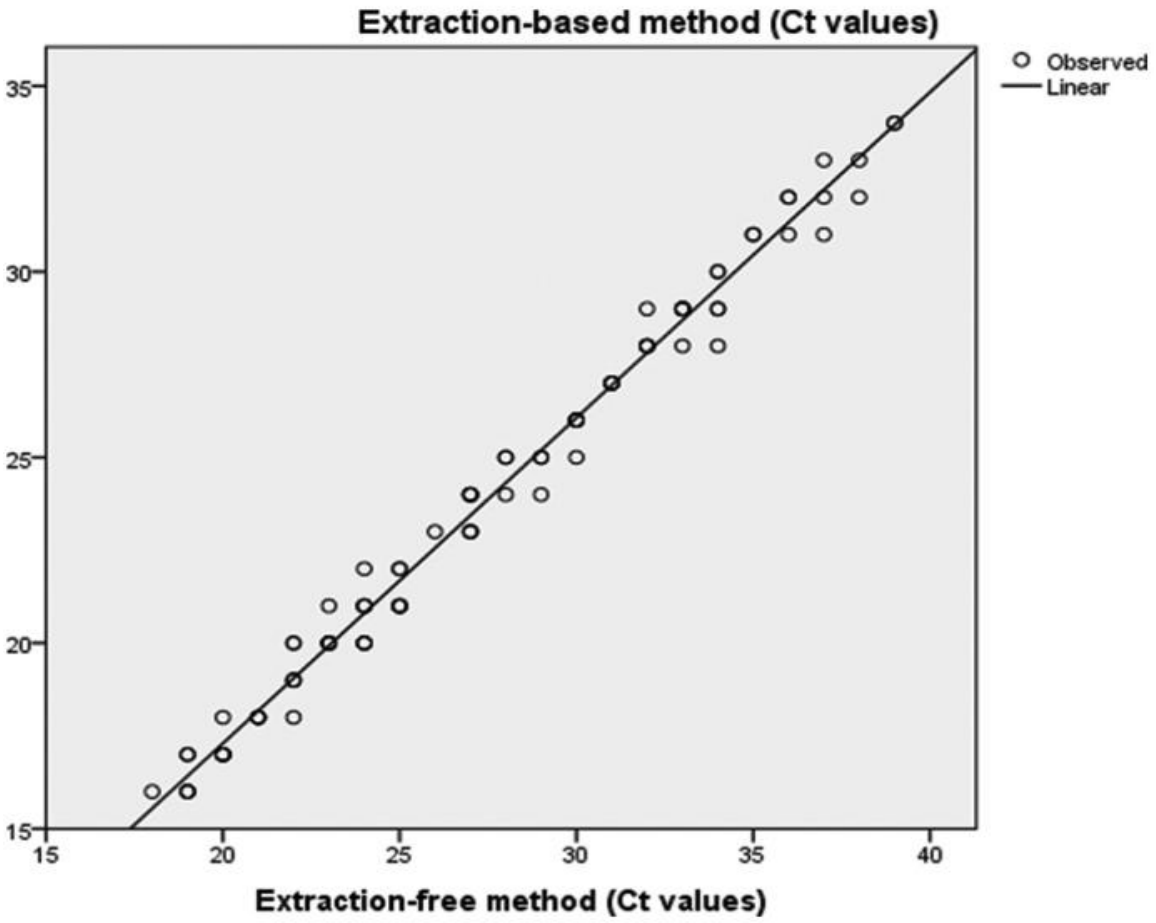


Figure 19. Correlation between extraction-free heating and extraction-based SARS-CoV-2 detection methods, Ethiopia.

The linear line showed a very strong positive relationship. This suggests that the results obtained using the two methods are highly consistent and move in the same direction, implying they are nearly identical in their measurements or outputs.

4.3.3. Comparison of performance characteristics in terms of grouped Ct Value

Regarding the samples with a Ct value less than 20 (38 samples), both techniques correctly identified 100% of the samples as true positives. Out of the 132 results with a Ct value ranging from 20 to 35, 125 (95%) were true positives. For the remaining 130 test results with a Ct value higher than 35, the new method resulted in 20 false negatives. These findings demonstrated a higher level of sensitivity in the Ct < 20 group, as both techniques accurately identified all samples as positive. In the Ct value 20–35 group, both techniques mostly detected positive samples, but the extraction-free method had 5.3% false negativity (seven samples). Lastly, the new technique showed 15.4% of false negativity and low sensitivity in the Ct value greater than 35 group. Therefore, these results suggested that the EFDH method is more reliable for detecting SARS-CoV-2 with a higher viral load (Table 10).

Table 10. Performance of extraction-free and extraction-based methods across different Ct value ranges.

Extraction-Free		**Extraction-Based**		
		Positive	**Negative**	**Total**
Ct < 20)	Positive	38	0	38
	Negative	0	0	0
	Total	38	0	38
Ct = 20–35	Positive	125	0	125
	Negative	7	0	7
	Total	132	0	132
Ct > 35	Positive	0	0	0
	Negative	20	110	130
	Total	20	110	130

CHAPTER 5: DISCUSSION

5.1. Summary and Interpretation of Major Findings

In Ethiopia, limited infrastructure for identifying circulating variants during each phase of the pandemic, and like other developing countries, the shortage of RNA extraction kits for SARS-CoV-2 posed significant challenges to effective response efforts. Therefore, periodic surveillances detect circulating variants in real time along with development of alternative diagnostic protocols with comparable diagnostic accuracy is crucial for both pandemic control and future preparedness. This dissertation aimed to investigate the SARS-CoV-2 variants that circulated during the fifth wave of the pandemic in selected regions of the country and to compare the diagnostic performance of extraction-free heating and extraction-based methods for SARS-CoV-2 detection and

This research provides valuable data on circulating SARS-CoV-2 variants during the fifth wave, which is essential for understanding the virus's evolution and informing public health strategies. This work can enhance testing efficiency, strengthen surveillance, and support more effective pandemic management and preparedness. It also crucial for evaluating alternative diagnostic methods (EFHD) to improve SARS-CoV-2 detection in resource-limited settings, addressing challenges related to the shortage of RNA extraction kits.

The key findings of this study include: (1). Omicron was the predominant variant (97%, 61/63) identified in selected regions of the country, followed by Delta variants (3%, 2/63), with the 22A clade being the most prevalent among Omicron strains, suggesting a significant shift in the viral landscape during this period. (2). A High number of SARS-CoV-2 nucleotide mutations were observed in different positions of S, M, and N genes and genes coding for non-structural proteins (nsp1, nsp6, nsp9, nsp10, nsp13, ORF3a, and ORF7a). These advocates significant viral evolution may influence various aspects of the virus’s behavior. (3). The transition/transversion bias (R) of 1.693 in this study indicates that transitions (substitution of a purine for a purine or a pyrimidine for a pyrimidine) occur more frequently than transversions (substitution between a purine and a pyrimidine). (4). The average evolutionary divergence across all sequences from Ethiopia is 9.78x10-4, with a standard error of 8.47x10-5. This indicates that the virus strains circulating within Ethiopia exhibit relatively low genetic variation, suggesting that local transmission chains may be largely stable, and there has been limited

evolutionary divergence within the Ethiopian population. However, comparing our sequences with the selected global sequences, the overall distance was slightly increasing to 1.54x10-3, with a standard error of 1.19x103. This suggests that the global strains exhibit slightly more genetic variation. (5). This study also identified highly conserved, antigenic, and surface-exposed B-cell epitopes within the SARS-CoV-2 nucleocapsid (N) protein derived from 63 clinical isolates in Ethiopia. In contrast to the variable spike gene, the N gene exhibited significant conservation across variants, thereby establishing it as a stable target for diagnostic purposes. The three principal epitopes identified were characterized as non-allergenic, non-toxic, and demonstrated minimal cross-reactivity with endemic human coronaviruses. Structural and docking analyses substantiated their accessibility and strong potential for antibody binding, thereby supporting their application in the development of robust diagnostics that are resilient to variant changes. (6). The EFDH method has shown high diagnostic performances (sensitivity 86%, 95%CI: (80%-90%); specificity 100%; and positive 100% and negative (80%, 95%CI; 74%-85%) predictive values), especially in a high viral load of SARS-CoV-2. In addition, the overall agreement between extraction-based and EFDH SARS-CoV-2 detection methods was very strong and statistical significance. This suggests that the EFDH method is a reliable alternative to the extraction-based method for SARS-CoV-2 detection.

5.2. Molecular Characterization of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia

In this study, the sequencing analysis revealed a high-quality yield from the RNA extracts, with 99.8% of the sequences being of high quality and only 0.4% comprising gaps. This underscores the robustness of the sequencing methodologies employed, which align with the standards set by various genomic surveillance initiatives (Aksamentov et al., 2021; Hadfield et al., 2018). The genomic analysis of SARS-CoV-2 reveals significant insights into the virus's mutation landscape and its evolutionary dynamics.

This study revealed that on average approximately 69 single nucleotide variations (SNVs) per genome., particularly concentrated in clades 22A and 21K, Which is similar to the study conducted by (Saldivar-Espinoza et al., 2023) that explored the mutational landscape of SARS-CoV-2 for the omicron variant by early June 2022 reported that the average number of SNVs per genome was around 72. In addition, the study by (Harvey et al., 2021) and (Yang et al., 2020) highlighted that the identification of high mutation rates, particularly in certain lineages, suggests that these variants may be under selective pressure, possibly due to host immune responses or therapeutic interventions.

The findings from the Nextclade variant analysis, which identified a total of 4,429 nucleotide mutations, 12,664 deletions, 4,392 substitutions, and 84 insertions across 63 SARS-CoV-2 sequences compared to the Wuhan reference genome. This provides critical insights into the genomic evolution of the virus as indicated by previous study (Aksamentov et al., 2021). This extensive mutational landscape reflects the ongoing adaptive changes that SARS-CoV-2 undergoes in response to selective pressures, including host immune responses and antiviral treatments (Gao et al., 2021).

In this study, significant nucleotide mutations were identified in critical genes such as spike (S), membrane (M), and nucleocapsid (N) genes, as well as non-structural proteins (nsp1, nsp6, nsp9, nsp10, nsp13, ORF3a, and ORF7a. Similar studies have corroborated these findings, highlighting the high mutation rates observed in SARS-CoV-2. For instance, a study conducted in Turkey reported elevated mutation rates in the S and RdRP gene regions, consistent with the findings of this analysis (Sahin et al., 2021). Furthermore, research from Egypt indicated a significant presence of the D614G mutation, which is known to enhance viral infectivity and has been widely documented across various global studies (Roshdy et al., 2022).

Moreover, the study identified specific mutations in the papain-like protease (PLpro) and the main protease (3CLpro). This study is aligned with the study done by (Gao et al., 2021) and (Ismail et al., 2022) that suggested the mutations K38R, V1069I, and P132H in PLpro, along with P323L in RdRP, could potentially alter the enzymatic activities of these proteins, impacting viral replication and immune evasion.

The phylogenetic analysis found in this study indicates that the dominant circulating variants in Ethiopia during the study period (June to August 2022) were primarily Omicron, with a smaller proportion of Delta variants. The Omicron variant (B.1.1.529) of SARS-CoV-2 was first identified in November 2021 in South Africa (Ren et al., 2022). It was classified as a VOC by WHO because of its high transmissibility and evasion from neutralizing antibodies induced by vaccination or natural infection with wild-type virus (WHO, 2021). As of 14 April 2022, the Omicron variant has been confirmed in at least 150 countries around the world (Google My Maps, 2022). From 3 January 2022 to 9 January 2022, the Omicron variant accounted for roughly 62.5% of all SARS-CoV-2 sequences available on the GISAID database (GISAID, 2023).

In this study, the majority of the samples belonged to the BA lineage. Which is consistent with findings reported from North India, 2022, (Zaman et al., 2022) and Philippines, 2022, (Li et al., 2022) during the study period.. The Centers for Disease Control and Prevention (CDC) have developed taxonomic nomenclatures (BA, BF, AY, and B lineages) to categorize different variants of SARS-CoV-2 (CDC, 2020). A study done by (Rauseo and O'Halloran, 2021) suggests understanding the distribution of these variants is crucial for public health interventions, including vaccination strategies and treatment approaches.

This study explored the evolutionary divergence among genetic sequences of SARS-CoV-2 and revealed a spectrum of genetic distances ranging from nearly identical matches to slight variations. Specifically, comparisons within Ethiopian sequences revealed a perfect match between three samples and a very small distance between others, indicating recent community transmission. On a broader scale, the average genetic distance among Ethiopian sequences was low, suggesting high genetic similarity within the country. This finding is similar to a previous study done in Ethiopia through Whole-Genome Sequences of SARS-CoV-2 Isolates from Ethiopian Patients (Alemayehu et al., 2022)

In addition, comparison of the Ethiopian sequences with selected global sequences showed in this analysis that a slightly higher average distance was observed. This indicates slightly more genetic diversity among global sequences. These findings were contrary to the evolution of SARS-CoV-2 was characterized by a steady increase in divergence within major lineages and a stepwise increase associated with each successive new major lineage, leading to a faster overall rate of evolution (Markov et al., 2023). However, another study analyzed 27,977 SARS-CoV-2 sequences from 84 countries obtained throughout the pandemic and showed that SARS-CoV-2 genetic diversity was remarkably low (Rausch et al., 2020).

5.3. Identification of conserved gene regions of SARS-CoV-2 for the development of monoclonal antibody-based diagnostic tools.

This study also examined conserved regions of the SARS-CoV-2 genome as potential targets for monoclonal antibody-based diagnostic tools. Through whole genome sequencing of locally circulating strains, the study identified that while the spike gene accumulated numerous mutations across variants, the N gene remained remarkably conserved. This observation supports a growing body of evidence

indicating that the N protein, particularly its RNA-binding and dimerization domains, is more genetically stable than the S protein and thus a better target for laboratory diagnosis (Chang et al., 2014).

Previous studies have consistently highlighted the instability of the S gene due to immune selection pressure, especially following mass vaccination and widespread natural infections (Harvey et al., 2021). Mutations such as D614G, N501Y and E484K, commonly observed in our results, have been widely reported as key drivers of immune evasion and increased transmissibility (Korber et al., 2020). In contrast, the N gene's lower mutation frequency aligns with earlier genomic surveillance reports from various areas that identified it as a functionally constrained region with limited antigenic drift (Mercatelli and Giorgi 2020; Wu et al., 2020). These features make the N gene an ideal target for diagnostic development.

The conservation analysis, supported by sequence alignment and entropy profiling, reinforced the N gene's stability across multiple variants of concern, including Alpha, Beta, Delta and Omicron. This finding corroborates studies by (Mercatelli and Giorgi 2020) and (Ahmed et al., 2020), who demonstrated that the N protein retains structural integrity and sequence stability across a wide array of SARS-CoV-2 genomes, even in the face of global viral evolution.

Epitope prediction within conserved regions of the N protein further validated their diagnostic potential. Predicted B-cell epitopes showed strong antigenicity, surface accessibility and structural flexibility, key features for effective recognition by host antibodies. These results are consistent with computational and experimental studies from Asia and Europe, where similar N-terminal and central N protein regions were identified as immunodominant and suitable for sero-diagnostic uses (Grifoni et al., 2020; Jespersen et al., 2017). Notably, these studies also emphasized that linear epitopes located in flexible coil regions are more likely to be immunogenic, which aligns with our structural mapping results.

Furthermore, the predicted epitopes demonstrated minimal sequence similarity to common human coronaviruses, suggesting a low risk of cross-reactivity, a major concern in the development of SARS-CoV-2-specific diagnostic tools. This is in agreement with findings by (Naz et al., 2020; Sette and Crotty, 2021), who underscored the need to carefully filter epitopes for cross-reactivity with endemic coronaviruses, particularly in serological assays.

Molecular docking simulations reported the predicted epitopes' ability to bind stably with monoclonal antibody fragments, offering preliminary evidence of their functional relevance. Comparable docking studies by(Johnson et al., 2022; Zhang et al., 2020) have demonstrated similar structural compatibility between predicted N protein epitopes and human immunoglobulins, further validating in silico approaches as a valuable screening tool for diagnostic development.

The implications of these findings are significant for countries like Ethiopia, where access to cost-effective and variant-resistant diagnostics remains limited. By targeting stable regions of the SARS-CoV-2 genome and validating their immunological relevance, this study lays the groundwork for the design of lateral flow assays of ELISAs tailored to the local epidemiological context. Importantly, our approach can also serve as a template for monitoring emerging pathogens beyond COVID-19, especially in settings with limited laboratory infrastructures.

5.4. Performance Evaluation of EFDH and Extraction-Based SARS-CoV-2 Detection Methods

In this study, the performance of the new diagnostic methods showed comparable diagnostic capability with the standard method. It had good sensitivity and specificity. This finding aligns with a study conducted in India, which reported an overall sensitivity of 79% (95% CI: 71%-86%) and specificity of 99% (95% CI: 98%-100% (Jayaprakasam et al., 2021). Another study on extraction-free, multiplexed amplification of SARS-CoV-2 demonstrated a sensitivity of 86% and specificity of 100% (Byrnes et al., 2021). Similarly, a study conducted at the Karolinska University Hospital in Stockholm, Sweden, in 2020, using a heating technique at 95°C for 5 minutes, reported a sensitivity of 96.0% and specificity of 99.8% (Smyrlaki et al., 2020). In contrast, a study conducted in Italy reported lower, 57.3% (95% CI 47.3–66.8%). However, the study reported comparable specificity, 100% (95% CI 94.4%–100%) (Morecchiato et al., 2021).

This study also found that a heightened positive and negative predictive value. This indicated excellent specificity in identifying true positives, thus enhancing the reliability of the method for confirming the presence of the diseases. However, the findings also highlighted the need for improvement in reducing false negatives. This will enhance the method's utility in ruling out the diseases. Similar findings were reported in a study conducted in the United Arab Emirates (UAE) in 2021. It showed a positive predictive value of 92% and a negative predictive value of 91% (Mahmoud et al., 2021). The finding

is also similar to a study done in India in 2021, with a PPV and NPV of 92% and 97%, respectively (Jayaprakasam et al., 2021).

In this study, the overall agreement between extraction-based and EFDH methods was found to be 91%, with a kappa value of 0.82 (p value=0.000). This indicates a high level of reliability in detecting SARS-CoV-2 infections. Similar findings were reported in a study conducted in India, where the agreement was 96.8% (k = 0.83, S.E. = 0.03) (Jayaprakasam et al., 2021). Additionally, a study conducted in Sweden showed an accuracy of 98.8% (95% CI : 97.5–99.5%) (Smyrlaki et al., 2020), while a study in the UAE reported an overall agreement (kappa coefficient) of 0.797 ($p < 0:001$) (Mahmoud et al., 2021).

However, different findings were reported from a study conducted at the clinical laboratory of the Pasteur Institute of M'sila, Algeria, where the overall agreement between extracted and heat-inactivated (65 $^{\circ}$C for 30 min) samples was only 45%, with a 95% confidence interval of 37 to 52% (Mokhtar, 2020). In this study, a near-perfect correlation ($R2 = 0.99$, $p=0.001$) was found between extraction-based and extraction-free heating (EFDH) methods, supporting the consistency and reliability of the EFDH method in detecting SARS-CoV-2. These findings are further supported by a study conducted in Singapore (2020), which showed a correlation coefficient (R^2) of 0.9986 for the detection of SARS CoV-2 without extraction. Similar results were reported from the Karolinska University Hospital in Stockholm, Sweden, with a correlation coefficient of 0.987 (Smyrlaki et al., 2020).

The consistent findings from various studies utilizing an extraction-free method for SARS-CoV-2 detection can be attributed to the method's simplicity, the inherent properties of the SARS-CoV-2 virus, and its validation against established standards. These factors contribute to the method's robustness and reliability in detecting SARS-CoV-2, making it a valuable tool in the fight against the COVID-19 pandemic (Domnich et al., 2021).

Despite the challenges, the overall high performance of the technique supports its continued use in surveillance and outbreak control efforts. However, the potential for false negatives emphasizes the importance of comprehensive testing strategies, including targeted testing of high-risk populations and asymptomatic individuals (Bastola et al., 2020). This method's advantages in terms of speed and simplicity, combined with its high sensitivity and specificity, make it a compelling alternative for

SARS-CoV-2 detection, especially in resource-limited settings or during rapid response scenarios (Hasan et al., 2020).

These findings highlight the strengths and limitations of the extraction-based method in detecting SARS-CoV-2 across different viral load statuses. Notably, the method exhibits high sensitivity and specificity in lower Ct values (< 35). However, its performance decreases in the detection of low viral loads (higher Ct value), as evidenced by an increased rate of false negatives.

A study conducted in Austria to evaluate extraction-free RT-qPCR methods for SARS-CoV-2 diagnostics showed an 8.2% false negative rate for the EFDH method in samples with a Ct value >30 (Domnich et al., 2021). Another finding from Algeria on the detection of SARS-CoV-2 using heat inactivation at 65 °C for 30 min correctly identified 100% of clinical samples with a high viral load (Ct value < 30) (Mokhtar, 2020). Accurate detection of SARS-CoV-2, even at high viral loads (Lower Ct value), is crucial for early case identification, contact tracing, and isolation measures.

5.5. Limitations of the Study

One of the limitations of the study is that the study utilizes a small sample size for genomic sequencing that limits the power to give more conclusive recommendations. Another limitation is limited geographical coverage. Due to financial and infrastructure limitations, the sequencing was done very late however, the data is still innovative for the readers who are interested in the SARS-COV-2 diagnostic alternatives and variant situation in Ethiopia.

Another limitation of the study to identify the conserver gene regions of SARS-CoV-2 that the study's computational nature is a limitation. While the results are promising, they require experimental validation through peptide synthesis, antibody binding assay and clinical performance evaluation. Such validation would be crucial to confirm the predicted epitopes' diagnostic accuracy and to rule out potential false positives due to structural or conformational effects not captured in silico.
Finally, the study did not evaluate the performance of the EFDH method under varying storage conditions and time intervals. Additionally, the study did not establish the limit of detection for the new method.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

The study provides significant insights into an alternative protocol for the detection of SARS-CoV-2 and the genomic variants circulating in selected regions of Ethiopia during the fifth wave of the pandemic. The findings indicate that the EFDH method exhibits considerable potential as a simplified and cost-effective alternative for SARS-CoV-2 detection, particularly in samples with high viral loads. With high sensitivity, perfect specificity, and robust agreement with standard extraction-based RT-qPCR, this method offers a practical solution for contexts where resources, reagents, or time are constrained. Its capacity to yield accurate results without the necessity for RNA extraction can substantially reduce testing time and costs, rendering it especially advantageous during public health emergencies or in low-resource settings. However, the method's diminished sensitivity in low viral load samples, as evidenced by a number of false negatives and invalid results, highlights the necessity for further optimization.

Furthermore, this study elucidates the demographic, clinical, and genomic characteristics of SARS-CoV-2 during the fifth wave of the COVID-19 pandemic in Ethiopia. The findings demonstrate a pronounced predominance of the Omicron variant, particularly sub-lineages BA.4.1 and BA.4.1.1, which accounted for the vast majority of cases, with a minor fraction associated with the Delta variant. Despite vaccination, all vaccinated participants were still infected with Omicron, suggesting potential immune escape and underscoring the variant’s capacity to circumvent vaccine-induced protection. Commonly reported symptoms included cough, fever, and anosmia, with a higher proportion of male participants observed, consistent with global patterns of increased COVID-19 severity among males.

The genomic analysis revealed that clade 22A was the most prevalent, followed by clade 22B, alongside the detection of Omicron variants 21K and 21L and Delta variant 21J. Notably, several key mutations were identified, particularly in the spike (S), nucleocapsid (N), and nsp6 genes, with the spike gene exhibiting the most significant alterations. These mutations may have implications for viral transmissibility, vaccine effectiveness, and immune response.

The relatively low genetic variability among Ethiopian sequences suggests recent and ongoing community transmission, with most sequences displaying close genetic similarity, indicative of a

shared source or simultaneous transmission event. When compared to global sequences, our samples exhibited slightly greater genetic homogeneity, although both local and international datasets indicated minimal overall divergence, suggesting limited evolutionary changes during this wave.

In general, the EFDH method presents potential as a reliable diagnostic tool, although further refinements are required to enhance its accuracy and expand its application. The study highlights the evolving nature of SARS-CoV-2 variants in Ethiopia and emphasizes the critical need for ongoing genomic surveillance. Such efforts are essential for the early detection of new variants and for informing vaccination strategies, public health interventions, and global pandemic preparedness.

Additionally, our findings advocate for the use of the nucleocapsid protein as a stable and immunogenic target for SARS-CoV-2 diagnostics and identify specific epitope candidates for further development. This integrative approach, which combines genomic surveillance with structural immunoinformatic, has the potential to accelerate the identification of reliable diagnostic markers amid evolving viral landscapes.

6.2. Recommendations

Based on the findings of the study and the conclusions drawn, the following recommendations are forwarded. (1) The study recommends increasing the number of sequenced samples and expanding the geographic coverage of genomic sequencing efforts to obtain a more comprehensive understanding of the circulating SARS-CoV-2 variants in Ethiopia. Establishing a national genomic surveillance system is advised to provide real-time data that can inform timely public health interventions. (2) Ensuring the prompt submission of genomic sequencing data to global databases such as GISAID is crucial for supporting international efforts to monitor and respond to emerging variants. This data-sharing initiative will play a vital role in global pandemic preparedness, facilitating more coordinated responses to future waves of COVID-19. (3) Additional studies focus on prioritizing the conserved gene regions of the SARS-CoV-2 nucleocapsid protein as diagnostic targets, particularly for the development of affordable and variant-resilient rapid tests in low-resource settings, and experimental validation of the predicted epitopes are essential to confirm their diagnostic efficacy. (4) The EFDH method could be used as an alternative diagnostic approach for SARS-CoV-2 detection in resource-limited countries like Ethiopia, while further studies are required to improve the EFDH method for better sensitivity in low

viral load cases. (5). In addition, the study suggests that exploring the impact of various storage conditions on the performance of the EFDH method is deemed necessary. Differences in temperature, duration, and handling could potentially affect sample integrity and, consequently, the accuracy of test results. Investigating these factors will ensure the method's robustness, particularly in resource-limited settings where optimal storage conditions may not always be achievable.

REFERENCES

Abagero, A., Ragazzoni, L., Hubloue, I., Barone-Adesi, F., Lamine, H., Addissie, et.al., (2022) A Review of COVID-19 Response Challenges in Ethiopia. *International Journal of Environmental Research and Public Health*, 19 (17) January, p. 11070.

Acharya, D., Liu, G. and Gack, M. U. (2020) Dysregulation of Type I Interferon Responses in COVID-19. *Nature Reviews Immunology*, 20 (7) July, pp. 397–398.

Ackermann, M., Verleden, S. E., Kuehnel, M., Haverich, A., Welte, T., Laenger, et.al., (2020) Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *The New England journal of medicine*, 383 (2) July, pp. 120–128.

Advice for the Public on COVID-19 – World Health Organization (2020) [Online]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public> [Accessed 4 April 2025a].

Africa CDC - COVID-19 Daily Updates (2020) *Africa CDC* [Online blog]. Available from: <https://africacdc.org/covid-19/> [Accessed 5 September 2024].

Ahmed, S. F., Quadeer, A. A. and McKay, M. R. (2020) Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses*, 12 (3) February, p. 254.

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., et.al., (2020) Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, 296 (2) August, pp. E32–E40.

Akira, S., Uematsu, S. and Takeuchi, O. (2006) Pathogen Recognition and Innate Immunity. *Cell*, 124 (4) February, pp. 783–801.

Aksamentov, I., Roemer, C., Hodcroft, E. B. and Neher, R. A. (2021) Nextclade: Clade Assignment, Mutation Calling and Quality Control for Viral Genomes. *Journal of Open Source Software*, 6 (67) November, p. 3773.

Alashkar Alhamwe, B., Miethe, S., Pogge von Strandmann, E., Potaczek, D. P. and Garn, H. (2020) Epigenetic Regulation of Airway Epithelium Immune Functions in Asthma. *Frontiers in Immunology* [Online], 11 August. Available from: <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2020.01747/full> [Accessed 10 April 2025].

Alemayehu, D. H., Adnew, B., Alemu, F., Tefera, D. A., Seyoum, T., Beyene, G. T., et.al., (2022) Whole-Genome Sequences of SARS-CoV-2 Isolates from Ethiopian Patients. *Microbiology Resource Announcements*, 10 (38), pp. e00721-21.

Allotey, J., Kew, T., Fernández-García, S., Gaetano-Gil, A., Yap, M., Sheikh, J., et.al., (2022) SARS-CoV-2 Positivity in Offspring and Timing of Mother-to-Child Transmission: Living Systematic Review and Meta-Analysis. *The BMJ*, 376 March, p. e067696.

Alm, E., Broberg, E. K., Connor, T., Hodcroft, E. B., Komissarov, A. B., Maurer-Stroh, S., et. al., (2020) Geographical and Temporal Distribution of SARS-CoV-2 Clades in the WHO European Region, January to June 2020. *Eurosurveillance*, 25 (32) August, p. 2001410.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215 (3) October, pp. 403–410.

Amodio, E., Vitale, F., Cimino, L., Casuccio, A. and Tramuto, F. (2020) Outbreak of Novel Coronavirus (SARS-CoV-2): First Evidences From International Scientific Literature and Pending Questions. *Healthcare*, 8 (1) February, p. 51.

Amoutzias, G. D., Nikolaidis, M., Tryfonopoulou, E., Chlichlia, K., Markoulatos, P. and Oliver, S. G. (2022) The Remarkable Evolutionary Plasticity of Coronaviruses by Mutation and Recombination: Insights for the COVID-19 Pandemic and the Future Evolutionary Paths of SARS-CoV-2. *Viruses*, 14 (1) January, p. 78.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. and Garry, R. F. (2020) The Proximal Origin of SARS-CoV-2. *Nature Medicine*, 26 (4) April, pp. 450–452.

Anthony, S. J., Johnson, C. K., Greig, D. J., Kramer, S., Che, X., Wells, H., et.al., (2017) Global Patterns in Coronavirus Diversity. *Virus Evolution*, 3 (1) June, p. vex012.

Aoki, K., Takai, K., Nagasawa, T., Kashiwagi, K., Mori, N., Matsubayashi, K., et.al., (2021) Combination of a SARS-CoV-2 IgG Assay and RT-PCR for Improved COVID-19 Diagnosis. *Annals of Laboratory Medicine*, 41 (6) November, pp. 568–576.

Baden, L. R., Sahly, H. M. E., Essink, B., Kotloff, K., Frey, S., et.al., (2021) Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine*, 384 (5) February, pp. 403–416.

Bai, Z., Cao, Y., Liu, W. and Li, J. (2021) The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses*, 13 (6) June, p. 1115.

Bastola, M. M., Locatis, C. and Fontelo, P. (2020) Diagnostic Laboratory Tests for COVID-19 in US: Methodology and Performance. *Research Square*, July, p. rs.3.rs-43374.

Behera, B. C., Mishra, R. R. and Thatoi, H. (2021) Recent Biotechnological Tools for Diagnosis of Corona Virus Disease: A Review. *Biotechnology Progress*, 37 (1), p. e3078.

Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et.al., (2020) Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine*, 383 (19) November, pp. 1813–1826.

Bernheim, A., Mei, X., Huang, M., Yang, Y., Fayad, Z. A., Zhang, N., et.al., (2020) Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology*, 295 (3) June, pp. 685–691.

Bhat, T., Cao, A. and Yin, J. (2022) Virus-like Particles: Measures and Biological Functions. *Viruses*, 14 (2) February, p. 383.

Bioer Technology Product Intro-711_PMT.Pdf (2020) [Online]. Available from: <https://conference.medgenetics.ru/upload/Bioer%20Technology%20Product%20Intro-711_%D0%A0%D0%9C%D0%A2.pdf> [Accessed 21 August 2023b].

Bisrat, H., Manyazewal, T. and Fekadu, A. (2023) Mobile Health–Supported Active Syndrome Surveillance for COVID-19 Early Case Finding in Addis Ababa, Ethiopia: Comparative Study. *Interactive Journal of Medical Research*, 12 (1) August, p. e43492.

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., et.al., (2020) Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, 181 (5) May, pp. 1036-1045.e9.

Boechat, J. L., Chora, I., Morais, A. and Delgado, L. (2021) The Immune Response to SARS-CoV-2 and COVID-19 Immunopathology – Current Perspectives. *Pulmonology*, 27 (5) September, pp. 423–437.

Bosch, B. J., Haan, C. A. M. de, Smits, S. L. and Rottier, P. J. M. (2005) Spike Protein Assembly into the Coronavirion: Exploring the Limits of Its Sequence Requirements. *Virology*, 334 (2) April, pp. 306–318.

Brant, A. C., Tian, W., Majerciak, V., Yang, W. and Zheng, Z.-M. (2021) SARS-CoV-2: From Its Discovery to Genome Structure, Transcription, and Replication. *Cell & Bioscience*, 11 (1) July, p. 136.

Brierley, I., Gilbert, R. J. C. and Pennell, S. (2009) Pseudoknot-Dependent Programmed —1 Ribosomal Frameshifting: Structures, Mechanisms and Models. *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, 24 July, pp. 149–174.

Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Singh, J., Streithorst, J., et.al., (2020) Rapid Detection of 2019 Novel Coronavirus SARS-CoV-2 Using a CRISPR-Based DETECTR Lateral Flow Assay. *medRxiv*, March, p. 2020.03.06.20032334.

Bullard, J., Dust, K., Funk, D., Strong, J. E., Alexander, D., Garnett, L., et.al., (2020) Predicting Infectious SARS-CoV-2 from Diagnostic Samples. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, May, p. ciaa638.

Bullock, H. A., Goldsmith, C. S. and Miller, S. E. (2022) Detection and Identification of Coronaviruses in Human Tissues Using Electron Microscopy. *Microscopy Research and Technique*, 85 (7), pp. 2740–2747.

Burrows-Wheeler Aligner (2020) [Online]. Available from: <https://bio-bwa.sourceforge.net/> [Accessed 7 May 2024c].

Bustin, S. A. and Nolan, T. (2020) RT-qPCR Testing of SARS-CoV-2: A Primer. *International Journal of Molecular Sciences*, 21 (8) April, p. 3004.

Butler, D., Mozsary, C., Meydan, C. *et al.* Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat Commun* **12**, 1660 (2021). https://doi.org/10.1038/s41467-021-21361-7

Byrnes, S. A., Gallagher, R., Steadman, A., Bennett, C., Rivera, R., Ortega, C., et.al., (2021) Multiplexed and Extraction-Free Amplification for Simplified SARS-CoV-2 RT-PCR Tests. *Analytical Chemistry*, 93 (9) March, pp. 4160–4165.

CDC (2020) *Coronavirus Disease 2019 (COVID-19)* [Online]. Centers for Disease Control and Prevention. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html> [Accessed 14 December 2023].

CDC (2025) *Overview of Testing for SARS-CoV-2* [Online]. COVID-19. Available from: <https://www.cdc.gov/covid/hcp/clinical-care/overview-testing-sars-cov-2.html> [Accessed 10 April 2025].

Chanda-Kapata, P., Ntoumi, F., Kapata, N., Lungu, P., Mucheleng'anga, L. A., Chakaya, J., et. al., (2022) Tuberculosis, HIV/AIDS and Malaria Health Services in Sub-Saharan Africa – A Situation Analysis of the

Disruptions and Impact of the COVID-19 Pandemic. *International Journal of Infectious Diseases*, 124 November, pp. S41–S46.

Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D. and Huang, T. (2014) The SARS Coronavirus Nucleocapsid Protein – Forms and Functions. *Antiviral Research*, 103 March, pp. 39–50.

Christensen, B. B., Azar, M. M. and Turbett, S. E. (2022) Laboratory Diagnosis for SARS-CoV-2 Infection. *Infectious Disease Clinics of North America*, 36 (2) June, pp. 327–347.

Cinquanta, L., Fontana, D. E. and Bizzaro, N. (2017) Chemiluminescent Immunoassay Technology: What Does It Change in Autoantibody Detection? *Auto-Immunity Highlights*, 8 (1) June, p. 9.

Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern (2022) [Online]. Available from: <https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern> [Accessed 23 April 2024d].

Coelho, L. E., Luz, P. M., Pires, D. C., Jalil, E. M., Perazzo, H., Torres, T. S., et.al (2022) Prevalence and Predictors of Anti-SARS-CoV-2 Serology in a Highly Vulnerable Population of Rio de Janeiro: A Population-Based Serosurvey. *The Lancet Regional Health – Americas* [Online], 15 November. Available from: <https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(22)00155-7/fulltext> [Accessed 3 April 2025].

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., et.al (2020) Detection of 2019 Novel Coronavirus (2019-nCoV) by Real-Time RT-PCR. *Eurosurveillance*, 25 (3) January, p. 2000045.

Cosar, B., Karagulleoglu, Z. Y., Unal, S., Ince, A. T., Uncuoglu, D. B., Tuncer, G., et. al., (2022) SARS-CoV-2 Mutations and Their Viral Variants. *Cytokine & Growth Factor Reviews*, 63 February, pp. 10–22.

Cui, F. and Zhou, H. S. (2020) Diagnostic Methods and Potential Portable Biosensors for Coronavirus Disease 2019. *Biosensors and Bioelectronics*, 165 October, p. 112349.

Cui, J., Li, F. and Shi, Z.-L. (2019) Origin and Evolution of Pathogenic Coronaviruses. *Nature Reviews Microbiology*, 17 (3) March, pp. 181–192.

Current ICTV Taxonomy Release | ICTV (2019) [Online]. Available from: <https://ictv.global/taxonomy> [Accessed 25 March 2025e].

datadot (2025) *COVID-19 Cases | WHO COVID-19 Dashboard* [Online]. datadot. Available from: <https://data.who.int/dashboards/covid19/cases> [Accessed 22 March 2025a].

datadot (2025) *COVID-19 Deaths | WHO COVID-19 Dashboard* [Online]. datadot. Available from: <https://data.who.int/dashboards/covid19/cases> [Accessed 22 March 2025b].

Del Valle, D. M., Kim-Schulze, S., Huang, H.-H., Beckmann, N. D., Nirenberg, S., Wang, B., et.al., (2020) An Inflammatory Cytokine Signature Predicts COVID-19 Severity and Survival. *Nature Medicine*, 26 (10) October, pp. 1636–1643.

Deng, X., Gu, W., Federman, S., Plessis, L. du, Pybus, O. G., Faria, N. R., et.al., (2020) Genomic Surveillance Reveals Multiple Introductions of SARS-CoV-2 into Northern California. *Science*, 369 (6503) July, pp. 582–587.

Deng, Y., Lei, L., Chen, Y. and Zhang, W. (2020) The Potential Added Value of FDG PET/CT for COVID-19 Pneumonia. *European Journal of Nuclear Medicine and Molecular Imaging*, 47 (7) July, pp. 1634–1635.

Dhand, R. and Li, J. (2020) Coughs and Sneezes: Their Role in Transmission of Respiratory Viral Infections, Including SARS-CoV-2. *American Journal of Respiratory and Critical Care Medicine*, 202 (5) September, pp. 651–659.

Dikale, E. (2021) Insights on Ethiopia's State of Emergency against COVID-19: A Descriptive Overview [Online]. Rochester, NY: Social Science Research Network. Available from: <https://papers.ssrn.com/abstract=3778097> [Accessed 10 April 2025].

Dimitrov, I., Bangov, I., Flower, D. R. and Doytchinova, I. (2014) AllerTOP v.2—a Server for in Silico Prediction of Allergens. *Journal of Molecular Modeling*, 20 (6) May, p. 2278.

Dinnes, J., Deeks, J. J., Adriano, A., Berhane, S., Davenport, C., Dittrich, S., et.al., (2020) Rapid, Point-of-care Antigen and Molecular-based Tests for Diagnosis of SARS-CoV-2 Infection. *The Cochrane Database of Systematic Reviews*, 2020 (8) August, p. CD013705.

Discontinuation of Transmission-Based Precautions and Disposition of Patients with COVID-19 in Healthcare Settings (Interim Guidance) (2020) [Online]. Available from: <https://stacks.cdc.gov/view/cdc/88538> [Accessed 2 April 2025f].

Dittmayer, C., Meinhardt, J., Radbruch, H., Radke, J., Heppner, B. I., Heppner, F. L., et.al., (2020) Why Misinterpretation of Electron Micrographs in SARS-CoV-2-Infected Tissue Goes Viral. *The Lancet*, 396 (10260) October, pp. e64–e65.

Domnich, A., De Pace, V., Pennati, B. M., Caligiuri, P., Varesano, S., Bruzzone, B. and Orsi, A. (2021) Evaluation of Extraction-Free RT-qPCR Methods for SARS-CoV-2 Diagnostics. *Archives of Virology*, 166 (10) October, pp. 2825–2828.

Dong, Y., Mo, X., Hu, Y., Qi, X., Jiang, F., Jiang, Z. and Tong, S. (2020) Epidemiology of COVID-19 Among Children in China. *Pediatrics*, 145 (6) June, p. e20200702.

Doremalen, N. van, Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., et.al., (2020) Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *The New England Journal of Medicine*, March, p. NEJMc2004973.

Dotsey, S. (2024) COVID-19 in Africa: Can the Recent Pandemic Spur Investments in Healthcare Infrastructure Development? *Discover Social Science and Health*, 4 (1) November, p. 63.

Dougan, M., Nirula, A., Azizad, M., Mocherla, B., Gottlieb, R. L., Chen, P., et.al., (2021) Bamlanivimab plus Etesevimab in Mild or Moderate Covid-19. *New England Journal of Medicine*, 385 (15) October, pp. 1382–1392.

Doytchinova, I. A. and Flower, D. R. (2007) VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinformatics*, 8 January, p. 4.

Drosten, C., Meyer, B., Müller, M. A., Corman, V. M., Al-Masri, M., Hossain, R., et.al., (2014) Transmission of MERS-Coronavirus in Household Contacts. *New England Journal of Medicine*, 371 (9) August, pp. 828–835.

Durham, P. J., Stevenson, B. J. and Farquharson, B. C. (1979) Rotavirus and Coronavirus Associated Diarrhoea in Domestic Animals. *New Zealand Veterinary Journal*, 27 (3) March, pp. 30–32.

Ethiopia Confirms First Coronavirus Case (n.d.) [Online]. Available from: <https://www.aa.com.tr/en/africa/ethiopia-confirms-first-coronavirus-case/1765130> [Accessed 15 April 2024g].

Eun, L. Y. (2021) Is Multisystem Inflammatory Syndrome Related with Coronavirus Disease 2019, Kawasaki Disease, and Angiotensin-Converting Enzyme 2 in Children? *Clinical and Experimental Pediatrics*, 64 (5) March, pp. 225–226.

Fan, H., Tian, M., Liu, S., Ye, C., Li, Z., Wu, K. and Zhu, C. (2024) Strategies Used by SARS-CoV-2 to Evade the Innate Immune System in an Evolutionary Perspective. *Pathogens*, 13 (12) December, p. 1117.

Fauver, J. R., Petrone, M. E., Hodcroft, E. B., Shioda, K., Ehrlich, H. Y., Watts, A. G., et.al., (2020) Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell*, 181 (5) May, pp. 990-996.e5.

Fehr, A. R. and Perlman, S. (2015) Coronaviruses: An Overview of Their Replication and Pathogenesis. *Coronaviruses*, 1282 February, pp. 1–23.

Flores-Vega, V. R., Monroy-Molina, J. V., Jiménez-Hernández, L. E., Torres, A. G., Santos-Preciado, J. I. and Rosales-Reyes, R. (2022) SARS-CoV-2: Evolution and Emergence of New Viral Variants. *Viruses*, 14 (4) March, p. 653.

Fomsgaard, A. S. and Rosenstierne, M. W. (2020) An Alternative Workflow for Molecular Detection of SARS-CoV-2 – Escape from the NA Extraction Kit-Shortage, Copenhagen, Denmark, March 2020. *Eurosurveillance*, 25 (14) April, p. 2000398.

Forster, P., Forster, L., Renfrew, C. and Forster, M. (2020) Phylogenetic Network Analysis of SARS-CoV-2 Genomes. *Proceedings of the National Academy of Sciences*, 117 (17) April, pp. 9241–9243.

Galloway, S. E., Paul, P., MacCannell, D. R., Johansson, M. A., Brooks, J. T., MacNeil, A., et.al., (2021) Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR. Morbidity and mortality weekly report*, 70 (3) January, pp. 95–99.

Galván Casas, C., Català, A., Carretero Hernández, G., Rodríguez-Jiménez, P., Fernández-Nieto, D., Rodríguez-Villa Lario, A., et.al., (2020) Classification of the Cutaneous Manifestations of COVID-19: A Rapid Prospective Nationwide Consensus Study in Spain with 375 Cases. *The British Journal of Dermatology*, 183 (1) July, pp. 71–77.

Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J. A., Wang, M. and Cui, S. (2021) Crystal Structure of SARS-CoV-2 Papain-like Protease. *Acta Pharmaceutica Sinica. B*, 11 (1) January, pp. 237–245.

Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S. and Ma, K. (2021) A Systematic Review of Asymptomatic Infections with COVID-19. *Journal of Microbiology, Immunology and Infection*, 54 (1) February, pp. 12–16.

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A. A., Zhu, G., Epstein, J. H., et.al., (2013) Isolation and Characterization of a Bat SARS-like Coronavirus That Uses the ACE2 Receptor. *Nature*, 503 (7477) November, pp. 535–538.

Genoud, V., Stortz, M., Waisman, A., Berardino, B. G., Verneri, P., Dansey, V., et.al., (2021) Extraction-Free Protocol Combining Proteinase K and Heat Inactivation for Detection of SARS-CoV-2 by RT-qPCR. *PLOS ONE*, 16 (2) February, p. e0247792.

GISAID Initiative (n.d.) [Online]. Available from: <https://www.epicov.org/epi3/frontend#28d967> [Accessed 21 August 2023h].

GitHub - Tseemann/Snp-Dists: Pairwise SNP Distance Matrix from a FASTA Sequence Alignment (n.d.) [Online]. Available from: <https://github.com/tseemann/snp-dists> [Accessed 26 October 2023i].

Goldman, E. (2020) Exaggerated Risk of Transmission of COVID-19 by Fomites. *The Lancet Infectious Diseases*, 20 (8) August, pp. 892–893.

Goldsmith, C. S., Miller, S. E., Martines, R. B., Bullock, H. A. and Zaki, S. R. (2020) Electron Microscopy of SARS-CoV-2: A Challenging Task. *The Lancet*, 395 (10238) May, p. e99.

Google My Maps (2023) *Omicron Variant (COVID-19 / SARS-CoV-2) by BNO News* [Online]. Google My Maps. Available from: <https://www.google.com/maps/d/viewer?mid=1lKX8ikHpKAkWCvuXn2g-CVW3OmHbN-k-> [Accessed 23 April 2024].

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et.al., (2020) A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature*, 583 (7816) July, pp. 459–468.

Grant, K., Andruchow, J. E., Conly, J., Lee, D. D., Mazurik, L., Atkinson, P. and Lang, E. (2021) Personal Protective Equipment Preservation Strategies in the Covid-19 Era: A Narrative Review. *Infection Prevention in Practice*, 3 (3) September, p. 100146.

Grifoni, A., Weiskopf, D., Ramirez, S. I., Mateus, J., Dan, J. M., Moderbacher, C. R., et.al., (2020) Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*, 181 (7) June, pp. 1489-1501.e15.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., et.al., (2003) Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science*, 302 (5643) October, pp. 276–278.

Guo, L., Ren, L., Yang, S., Xiao, M., Chang, D., Yang, F., et.al., (2020) Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clinical Infectious Diseases*, 71 (15) July, pp. 778–785.

Guo, X.-G., Zhou, Y.-Z., Li, Q., Wang, W., Wen, J.-Z., Zheng, L. and Wang, Q. (2018) Rapid and Reliable Diagnostic Method to Detect Zika Virus by Real-Time Fluorescence Reverse Transcription Loop-Mediated Isothermal Amplification. *AMB Express*, 8 (1) April, p. 60.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Consortium, O. S. D. D. et.al., and (2013) In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLOS ONE*, 8 (9) September, p. e73957.

Haan, C. A. M. de and Rottier, P. J. M. (2005) Molecular Interactions in the Assembly of Coronaviruses. *Advances in Virus Research*, 64, pp. 165–230.

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et.al., (2018) Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics*, 34 (23) December, pp. 4121–4123.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., et.al., (2020) Impaired Type I Interferon Activity and Inflammatory Responses in Severe COVID-19 Patients. *Science (New York, N.y.)*, 369 (6504) August, pp. 718–724.

Hamming, I., Timens, W., Bulthuis, M., Lely, A., Navis, G. and Goor, H. van (2004) Tissue Distribution of ACE2 Protein, the Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. *The Journal of Pathology*, 203 (2) June, pp. 631–637.

Hammond, J., Leister-Tebbe, H., Gardner, A., Abreu, P., Bao, W., Wisemandle, W., et.al., (2022) Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *New England Journal of Medicine*, 386 (15) April, pp. 1397–1408.

Han, A. X., Parker, E., Scholer, F., Maurer-Stroh, S. and Russell, C. A. (2019) Phylogenetic Clustering by Linear Integer Programming (PhyCLIP). *Molecular Biology and Evolution*, 36 (7) July, pp. 1580–1595.

Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., et.al., (2020) Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases Journal - CDC. [Online]. Available from: <https://wwwnc.cdc.gov/eid/article/26/6/20-0516_article> [Accessed 2 April 2025].

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et.al., (2021) SARS-CoV-2 Variants, Spike Mutations and Immune Escape. *Nature Reviews Microbiology*, 19 (7) July, pp. 409–424.

Hasan, M. R., Mirza, F., Al-Hail, H., Sundararaju, S., Xaba, T., Iqbal, M., et.al., (2020) Detection of SARS-CoV-2 RNA by Direct RT-qPCR on Nasopharyngeal Specimens without Extraction of Viral RNA. *PLOS ONE*, 15 (7) July, p. e0236564.

Hasan, M. R., Mirza, F., Al-Hail, H., Sundararaju, S., Xaba, T., Iqbal, M., et.al., (2020) Detection of SARS-CoV-2 RNA by Direct RT-qPCR on Nasopharyngeal Specimens without Extraction of Viral RNA. *PLOS ONE*, 15 (7) July, p. e0236564.

Heneghan, C. J., Spencer, E. A., Brassey, J., Plüddemann, A., Onakpoya, I. J., Evans, D. H., et.al., (2021) SARS-CoV-2 and the Role of Orofecal Transmission: A Systematic Review. *F1000Research*, 10 November, p. 231.

Hernandez Acosta, R. A., Esquer Garrigos, Z., Marcelin, J. R. and Vijayvargiya, P. (2022) COVID-19 Pathogenesis and Clinical Manifestations. *Infectious Disease Clinics of North America*, 36 (2) June, pp. 231–249.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., et.al., (2011) High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry*, 83 (22) November, pp. 8604–8610.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et.al., (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181 (2) April, pp. 271-280.e8.

Hoffmann, M., Mösbauer, K., Hofmann-Winkler, H., Kaul, A., Kleine-Weber, H., Krüger, N., et.al., (2020) Chloroquine Does Not Inhibit Infection of Human Lung Cells with SARS-CoV-2. *Nature*, 585 (7826) September, pp. 588–590.

Hou, Y. J., Okuda, K., Edwards, C. E., Martinez, D. R., Asakura, T., Dinnon, K. H., et.al., (2020) SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell*, 182 (2) July, pp. 429-446.e14.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et.al., (2020) Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *Lancet (London, England)*, 395 (10223), pp. 497–506.

Huang, W. E., Lim, B., Hsu, C.-C., Xiong, D., Wu, W., Yu, Y., et.al., (2020) RT-LAMP for Rapid Diagnosis of Coronavirus SARS-CoV-2. *Microbial Biotechnology*, 13 (4), pp. 950–961.

Ibarrondo, F. J., Fulcher, J. A., Goodman-Meza, D., Elliott, J., Hofmann, C., Hausner, M. A., et.al., (2020) Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *New England Journal of Medicine*, 383 (11) September, pp. 1085–1087.

IDSA Guidelines on the Treatment and Management of Patients with COVID-19 (2020) [Online]. Available from: <https://www.idsociety.org/practice-guideline/covid-19-guideline-treatment-and-management/> [Accessed 4 April 2025j].

Illumina COVIDseq Test (RUO) Kit - Google Search (2020) [Online]. Available from: https://support.illumina.com/downloads/illumina-covidseq-test-reference-guide-1000000126053.html [Accessed 21 August 2023k].

Illumina DNA Prep with Enrichment Reference Guide (1000000048041) (2020)

Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation | Nucleic Acids Research | Oxford Academic (2020) [Online]. Available from: <https://academic.oup.com/nar/article/49/W1/W293/6246398?login=true> [Accessed 26 October 2023m].

Ismail, G., Abdelghaffar, H., Seadawy, M. G., El-Hosseny, M. F., Gad, A. F., Ageez, A., et.al., (2022) Genome Sequencing Reveals Existence of SARS-CoV-2 B.1.1.529 Variant in Egypt. *Journal of Genetic Engineering & Biotechnology*, 20 May, p. 70.

Jacobi, A., Chung, M., Bernheim, A. and Eber, C. (2020) Portable Chest X-Ray in Coronavirus Disease-19 (COVID-19): A Pictorial Review. *Clinical Imaging*, 64 August, pp. 35–42.

Jami, G., Ataee, M., Esmaeili, V., Chamani, S., Rezaei, A. and Naghizadeh, A. (2023) Characterization of the Angiotensin-Converting Enzyme 2 (ACE2), the Main Receptor for the SARS-CoV-2 Virus. *American Journal of Clinical and Experimental Immunology*, 12 (3) June, pp. 24–44.

Jayaprakasam, M., Aggarwal, S., Mane, A., Saxena, V., Rao, A., Bandopadhyay, B., et.al., (2021) RNA-Extraction-Free Diagnostic Method to Detect SARS-CoV-2: An Assessment from Two States, India. *Epidemiology and Infection*, 149, p. e245.

Jespersen, M. C., Peters, B., Nielsen, M. and Marcatili, P. (2017) BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Research*, 45 (Web Server issue) July, pp. W24–W29.

Jin, J.-M., Bai, P., He, W., Wu, F., Liu, X.-F., Han, D.-M., Liu, S. and Yang, J.-K. (2020) Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Frontiers in Public Health*, 8 April, p. 152.

Johns Hopkins Coronavirus Resource Center (2024) *COVID-19 Map* [Online]. Johns Hopkins Coronavirus Resource Center. Available from: <https://coronavirus.jhu.edu/map.html> [Accessed 5 September 2024].

Johnson, B. A., Zhou, Y., Lokugamage, K. G., Vu, M. N., Bopp, N., Crocquet-Valdes, P. A., et.al., (2022) Nucleocapsid Mutations in SARS-CoV-2 Augment Replication and Pathogenesis. *bioRxiv*, March, p. 2021.10.14.464390.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et.al., (2021) Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596 (7873) August, pp. 583–589.

K, K. and Dm, S. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution* [Online], 30 (4) April. Available from: <https://pubmed.ncbi.nlm.nih.gov/23329690/> [Accessed 26 October 2023].

Kanne, J. P., Little, B. P., Chung, J. H., Elicker, B. M. and Ketai, L. H. (2020) Essentials for Radiologists on COVID-19: An Update—Radiology Scientific Expert Panel. *Radiology*, 296 (2) August, pp. E113–E114.

Kariuki, S., Kering, K., Wairimu, C., Onsare, R. and Mbae, C. (2022) Antimicrobial Resistance Rates and Surveillance in Sub-Saharan Africa: Where Are We Now? *Infection and Drug Resistance*, 15 July, pp. 3589–3609.

Katoh, K. and Standley, D. M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30 (4) April, pp. 772–780.

Kellner, M. J., Koob, J., Gootenberg, J. S., Abudayyeh, O. O. and Zhang, F. (2019) SHERLOCK: Nucleic Acid Detection with CRISPR Nucleases. *Nature protocols*, 14 (10) October, pp. 2986–3012.

Khoury, D. S., Cromer, D., Reynaldi, A., Schlub, T. E., Wheatley, A. K., Juno, J. A., et.al., (2021) Neutralizing Antibody Levels Are Highly Predictive of Immune Protection from Symptomatic SARS-CoV-2 Infection. *Nature Medicine*, 27 (7) July, pp. 1205–1211.

Kim, J.-M., Chung, Y.-S., Jo, H. J., Lee, N.-J., Kim, M. S., Woo, S. H., et.al., (2020) Identification of Coronavirus Isolated from a Patient in Korea with COVID-19. *Osong Public Health and Research Perspectives*, 11 (1) February, pp. 3–7.

Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., et.al., (2020) SARS-CoV-2 Structure and Replication Characterized by in Situ Cryo-Electron Tomography. *Nature Communications*, 11 November, p. 5885.

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et.al., (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182 (4) August, pp. 812-827.e19.

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S. (2017) The ClusPro Web Server for Protein-Protein Docking. *Nature protocols*, 12 (2) February, pp. 255–278.

Kremer, S., Lersy, F., Sèze, J. de, Ferré, J.-C., Maamar, A., Carsin-Nicol, B., et.al., (2020) Brain MRI Findings in Severe COVID-19: A Retrospective Observational Study. *Radiology*, June, p. 202222.

Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. and Lessler, J. (2020) Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Annals of Internal Medicine*, May, pp. M20-1495.

Kudo, E., Israelow, B., Vogels, C. B. F., Lu, P., Wyllie, A. L., Tokuyama, M., et.al., (2020) Detection of SARS-CoV-2 RNA by Multiplex RT-qPCR. *PLOS Biology*, 18 (10) October, p. e3000867.

Kumar, K. S. R., Mufti, S. S., Sarathy, V., Hazarika, D. and Naik, R. (2021) An Update on Advances in COVID-19 Laboratory Diagnosis and Testing Guidelines in India. *Frontiers in Public Health* [Online], 9 March. Available from: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.568603/full> [Accessed 3 April 2025].

Laboratory Testing for 2019 Novel Coronavirus (2019-nCoV) in Suspected Human Cases (2020.) [Online]. Available from: <https://www.who.int/publications/i/item/10665-331501> [Accessed 2 April 2025n].

Lam, T. T.-Y., Jia, N., Zhang, Y.-W., Shum, M. H.-H., Jiang, J.-F., Zhu, H.-C., et.al., (2020) Identifying SARS-CoV-2-Related Coronaviruses in Malayan Pangolins. *Nature*, 583 (7815) July, pp. 282–285.

Lamers, M. M. and Haagmans, B. L. (2022) SARS-CoV-2 Pathogenesis. *Nature Reviews Microbiology*, 20 (5) May, pp. 270–284.

Lamers, M. M., Mykytyn, A. Z., Breugem, T. I., Wang, Y., Wu, D. C., Riesebosch, S., et.al., (2021) Human Airway Cells Prevent SARS-CoV-2 Multibasic Cleavage Site Cell Culture Adaptation. *eLife*, 10, p. e66815.

Lawal, Y. (2021) Africa's Low COVID-19 Mortality Rate: A Paradox? *International Journal of Infectious Diseases*, 102 January, pp. 118–122.

Le, K., Kannappan, S., Kim, T., Lee, J. H., Lee, H.-R. and Kim, K. K. (2023) Structural Understanding of SARS-CoV-2 Virus Entry to Host Cells. *Frontiers in Molecular Biosciences*, 10 November, p. 1288686.

Lechien, J. R., Chiesa-Estomba, C. M., De Siati, D. R., Horoi, M., Le Bon, S. D., Rodriguez, A., et.al. (2020) Olfactory and Gustatory Dysfunctions as a Clinical Presentation of Mild-to-Moderate Forms of the Coronavirus Disease (COVID-19): A Multicenter European Study. *European Archives of Oto-Rhino-Laryngology*, 277 (8) August, pp. 2251–2261.

Li, S.-Y., Cheng, Q.-X., Wang, J.-M., Li, X.-Y., Zhang, Z.-L., Gao, S., et.al., (2018) CRISPR-Cas12a-Assisted Nucleic Acid Detection. *Cell Discovery*, 4 (1) April, pp. 1–4.

Li, Y.-T., Polotan, F. G. M., Sotelo, G. I. S., Dolor, A. Y. M., Tujan, M. A. A., Gomez, M. R. R., et.al., (2022) Lineage BA.2 Dominated the Omicron SARS-CoV-2 Epidemic Wave in the Philippines.

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

Long, Q.-X., Liu, B.-Z., Deng, H.-J., Wu, G.-C., Deng, K., Chen, Y.-K., et.al., (2020) Antibody Responses to SARS-CoV-2 in Patients with COVID-19. *Nature Medicine*, 26 (6) June, pp. 845–848.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et.al., (2020) Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet*, 395 (10224) February, pp. 565–574.

Mahmoud, S. A., Ganesan, S., Ibrahim, E., Thakre, B., Teddy, J. G., Raheja, P. and Abbas, W. Z. (2021) Evaluation of RNA Extraction-Free Method for Detection of SARS-CoV-2 in Salivary Samples for Mass Screening for COVID-19. *BioMed Research International*, 2021 June, pp. 1–6.

Maison, D. P., Deng, Y. and Gerschenson, M. (2023) SARS-CoV-2 and the Host-Immune Response. *Frontiers in Immunology*, 14 June, p. 1195871.

Major, J., Crotta, S., Llorian, M., McCabe, T. M., Gad, H. H., Priestnall, S. L., et.al., (2020) Type I and III Interferons Disrupt Lung Epithelial Repair during Recovery from Viral Infection. *Science (New York, N.y.)*, 369 (6504) August, pp. 712–717.

Malone, B., Urakova, N., Snijder, E. J. and Campbell, E. A. (2022) Structures and Functions of Coronavirus Replication–Transcription Complexes and Their Relevance for SARS-CoV-2 Drug Design. *Nature Reviews Molecular Cell Biology*, 23 (1) January, pp. 21–39.

Markov, P. V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N. I. and Katzourakis, A. (2023) The Evolution of SARS-CoV-2. *Nature Reviews Microbiology*, 21 (6) June, pp. 361–379.

Martin, M. (2011) Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal*, 17 (1) May, pp. 10–12.

Mason, R. J. (2020) Pathogenesis of COVID-19 from a Cell Biologic Perspective. *The European Respiratory Journal*, April, p. 2000607.

Masters, P. S. (2006) The Molecular Biology of Coronaviruses [Online]. In: *Advances in Virus Research*. vol. 66. Academic Press, pp. 193–292. Available from: <https://www.sciencedirect.com/science/article/pii/S0065352706660053> [Accessed 10 April 2025].

Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S. and Manson, J. J. (2020) COVID-19: Consider Cytokine Storm Syndromes and Immunosuppression. *The Lancet*, 395 (10229) March, pp. 1033–1034.

Merad, M. and Martin, J. C. (2020) Pathological Inflammation in Patients with COVID-19: A Key Role for Monocytes and Macrophages. *Nature Reviews Immunology*, 20 (6) June, pp. 355–362.

Mercatelli, D. and Giorgi, F. M. (2020) Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology* [Online], 11 July. Available from: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.01800/full> [Accessed 16 May 2025].

Metsky, H. C., Siddle, K. J., Gladden-Young, A., Qu, J., Yang, D. K., Brehio, P., et.al. (2019) Capturing Sequence Diversity in Metagenomes with Comprehensive and Scalable Probe Design. *Nature Biotechnology*, 37 (2) February, pp. 160–168.

Mina, M. J., Parker, R. and Larremore, D. B. (2020) Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. *New England Journal of Medicine*, 383 (22) November, p. e120.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Haeseler, A. von and Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37 (5) May, pp. 1530–1534.

Miorin, L., Kehrer, T., Sanchez-Aparicio, M. T., Zhang, K., Cohen, P., Patel, R. S., et.al. (2020) SARS-CoV-2 Orf6 Hijacks Nup98 to Block STAT Nuclear Import and Antagonize Interferon Signaling. *Proceedings of the National Academy of Sciences*, 117 (45) November, pp. 28344–28354.

Mizumoto, K., Kagaya, K., Zarebski, A. and Chowell, G. (2020) Estimating the Asymptomatic Proportion of Coronavirus Disease 2019 (COVID-19) Cases on Board the Diamond Princess Cruise Ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25 (10) March, p. 2000180.

Mohammed, H., Oljira, L., Roba, K. T., Yimer, G., Fekadu, A. and Manyazewal, T. (2020) Containment of COVID-19 in Ethiopia and Implications for Tuberculosis Care and Research. *Infectious Diseases of Poverty*, 9 (1) September, p. 131.

Moitra, P., Alafeef, M., Dighe, K., Frieman, M. B. and Pan, D. (2020) Selective Naked-Eye Detection of SARS-CoV-2 Mediated by N Gene Targeted Antisense Oligonucleotide Capped Plasmonic Nanoparticles. *ACS Nano*, 14 (6) June, pp. 7617–7627.

Mokhtar, K. M. (2020) Improved COVID-19 Testing by Extraction Free SARS-Cov-2 RT-PCR [Online]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.08.10.20171512> [Accessed 2 May 2024].

Mongodi, S., De Luca, D., Colombo, A., Stella, A., Santangelo, E., Corradi, F., et.al. (2021) Quantitative Lung Ultrasound: Technical Aspects and Clinical Applications. *Anesthesiology*, 134 (6) June, p. 949.

Mordecai, G. and Hewson, I. (2020) Coronaviruses in the Sea. *Frontiers in Microbiology*, 11 July, p. 1795.

Morecchiato, F., Coppi, M., Baccani, I., Maggini, N., Ciccone, N., Antonelli, A. and Rossolini, G. M. (2021) Evaluation of Extraction-Free RT-PCR Methods for Faster and Cheaper Detection of SARS-CoV-2 Using Two Commercial Systems. *International Journal of Infectious Diseases*, 112 November, pp. 264–268.

Nadim, M. K., Forni, L. G., Mehta, R. L., Connor, M. J., Liu, K. D., Ostermann, M., et.al. (2020) COVID-19-Associated Acute Kidney Injury: Consensus Report of the 25th Acute Disease Quality Initiative (ADQI) Workgroup. *Nature Reviews Nephrology*, 16 (12) December, pp. 747–764.

Nagamine, K., Hase, T. and Notomi, T. (2002) Accelerated Reaction by Loop-Mediated Isothermal Amplification Using Loop Primers. *Molecular and Cellular Probes*, 16 (3) June, pp. 223–229.

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et.al. (2020) Insights into SARS-CoV-2 Genome, Structure, Evolution, Pathogenesis and Therapies: Structural Genomics Approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1866 (10) October, p. 165878.

Naz, A., Shahid, F., Butt, T. T., Awan, F. M., Ali, A. and Malik, A. (2020) Designing Multi-Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 (COVID-19) by Employing Immuno-Informatics Approach. *Frontiers in Immunology*, 11 July, p. 1663.

NCBI Virus (2024) [Online]. Available from: <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202,%20taxid:2697049&utm_source=nuccore&utm_medium=referral&Authors_idx%20q.op%3DAND=Hailu&SubmitterAffil_idx%20q.op%3DOR=Ethiopian%20Public%20Health%20Institute&SubmitterCountry_idx%20q.op%3DOR=Ethiopia> [Accessed 18 September 2024o].

Nextclade (2023) [Online]. Available from: <https://clades.nextstrain.org> [Accessed 21 August 2023p].

Nextstrain (2023) [Online]. Available from: <https://nextstrain.org/> [Accessed 29 March 2025q].

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N. and Hase, T. (2000) Loop-Mediated Isothermal Amplification of DNA. *Nucleic Acids Research*, 28 (12) June, p. e63.

Novel Coronavirus – Thailand (2019) [Online]. Available from: <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON234> [Accessed 22 March 2025r].

Okba, N. M. A., Müller, M. A., Li, W., Wang, C., GeurtsvanKessel, C. H., Corman, V. M., et.al. (2020) SARS-CoV-2 Specific Antibody Responses in COVID-19 Patients [Online]. medRxiv, p. 2020.03.18.20038059. Available from: <https://www.medrxiv.org/content/10.1101/2020.03.18.20038059v1> [Accessed 3 April 2025].

Omotoso, O., Teibo, J. O., Atiba, F. A., Oladimeji, T., Paimo, O. K., Ataya, F. S., et.al. (2023) Addressing Cancer Care Inequities in Sub-Saharan Africa: Current Challenges and Proposed Solutions. *International Journal for Equity in Health*, 22 (1) September, p. 189.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., et.al. (2021) Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evolution*, 7 (2), p. veab064.

Palaiodimos, L., Kokkinidis, D. G., Li, W., Karamanis, D., Ognibene, J., Arora, S., et.al. (2020) Severe Obesity, Increasing Age and Male Sex Are Independently Associated with Worse in-Hospital Outcomes, and Higher in-Hospital Mortality, in a Cohort of Patients with COVID-19 in the Bronx, New York. *Metabolism - Clinical and Experimental* [Online], 108 July. Available from: <https://www.metabolismjournal.com/article/S0026-0495(20)30126-8/fulltext> [Accessed 1 April 2025].

Pampel, J. (2021) *SARS-CoV-2 Life Cycle: Stages and Inhibition Targets* [Online]. Available from: <https://www.antibodies-online.com/resources/18/5410/sars-cov-2-life-cycle-stages-and-inhibition-targets/> [Accessed 29 March 2025].

Paraskevis, D., Kostaki, E. G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G. and Tsiodras, S. (2020) Full-Genome Evolutionary Analysis of the Novel Corona Virus (2019-nCoV) Rejects the Hypothesis of Emergence as a Result of a Recent Recombination Event. *Infection, Genetics and Evolution*, 79 April, p. 104212.

Park, W. B., Kwon, N.-J., Choi, S.-J., Kang, C. K., Choe, P. G., Kim, J. Y., et.al. (2019) Virus Isolation from the First Patient with SARS-CoV-2 in Korea. *Journal of Korean Medical Science* [Online], 35 (7) December. Available from: <https://doi.org/10.3346/jkms.2020.35.e84> [Accessed 2 April 2025].

Passmore, L. A. and Coller, J. (2022) Roles of mRNA Poly(A) Tails in Regulation of Eukaryotic Gene Expression. *Nature reviews. Molecular cell biology*, 23 (2) February, pp. 93–106.

Patchsung, M., Jantarug, K., Pattama, A., Aphicho, K., Suraritdechachai, S., Meesawat, P., et.al. (2020) Clinical Validation of a Cas13-Based Assay for the Detection of SARS-CoV-2 RNA. *Nature Biomedical Engineering*, 4 (12) December, pp. 1140–1149.

Peeling, R. W., Heymann, D. L., Teo, Y.-Y. and Garcia, P. J. (2022) Diagnostics for COVID-19: Moving from Pandemic Response to Control. *Lancet (London, England)*, 399 (10326), pp. 757–768.

Pekar, J. E., Magee, A., Parker, E., Moshiri, N., Izhikevich, K., Havens, J. L., et.al. (2022) The Molecular Epidemiology of Multiple Zoonotic Origins of SARS-CoV-2. *Science (New York, N.y.)*, July, p. eabp8337.

Pekosz, A., Parvu, V., Li, M., Andrews, J. C., Manabe, Y. C., Kodsi, S., et.al. (2021) Antigen-Based Testing but Not Real-Time Polymerase Chain Reaction Correlates With Severe Acute Respiratory Syndrome Coronavirus 2 Viral Culture. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 73 (9) January, pp. e2861–e2866.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. et.al. (2004) UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, 25 (13), pp. 1605–1612.

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et.al. (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*, 383 (27) December, pp. 2603–2615.

Premkumar, L., Segovia-Chumbez, B., Jadi, R., Martinez, D. R., Raut, R., Markmann, A., et.al. (2020) The Receptor Binding Domain of the Viral Spike Protein Is an Immunodominant and Highly Specific Target of Antibodies in SARS-CoV-2 Patients. *Science Immunology*, 5 (48) June, p. eabc8413.

Puntmann, V. O., Carerj, M. L., Wieters, I., Fahim, M., Arendt, C., Hoffmann, J., et.al. (2020) Outcomes of Cardiovascular Magnetic Resonance Imaging in Patients Recently Recovered From Coronavirus Disease 2019 (COVID-19). *JAMA Cardiology*, 5 (11) November, pp. 1265–1273.

[PyMOL] PyMOL 2.5 Released | PyMOL Molecular Graphics System (2022) [Online]. Available from: <https://sourceforge.net/p/pymol/mailman/message/37279385/> [Accessed 15 May 2025t].

Qu, Y.-H., Long, N., Ran, C. and Sun, J. (2021) The Correlation of 18F-FDG PET/CT Metabolic Parameters, Clinicopathological Factors, and Prognosis in Breast Cancer. *Clinical and Translational Oncology*, 23 (3) March, pp. 620–627.

Rajpal, S., Tong, M. S., Borchers, J., Zareba, K. M., Obarski, T. P., Simonetti, O. P. and Daniels, C. J. (2021) Cardiovascular Magnetic Resonance Findings in Competitive Athletes Recovering From COVID-19 Infection. *JAMA Cardiology*, 6 (1) January, pp. 116–118.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., Du Plessis, L. and Pybus, O. G. (2020) A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nature Microbiology*, 5 (11) July, pp. 1403–1407.

Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C. and Kearney, M. F. (2020) Low Genetic Diversity May Be an Achilles Heel of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, 117 (40) October, pp. 24614–24616.

Rauseo, A. M. and O'Halloran, J. A. (2021) What Are the Clinical Implications of the SARS-CoV-2 Variants. *JACC: Basic to Translational Science*, 6 (3) February, pp. 305–308.

Re3data.Org (2012) GISAID [Online]. re3data.org - Registry of Research Data Repositories. Available from: <https://www.re3data.org/repository/r3d100010126> [Accessed 29 March 2025].

Rebendenne, A., Chaves Valadão, A. L., Tauziet, M., Maarifi, G., Bonaventure, B., McKellar, J., et.al. (2021) SARS-CoV-2 Triggers an MDA-5-Dependent Interferon Response Which Is Unable To Control Replication in Lung Epithelial Cells. *Journal of Virology*, 95 (8) March, p. 10.1128/jvi.02415-20.

Ren, S.-Y., Wang, W.-B., Gao, R.-D. and Zhou, A.-M. (2022) Omicron Variant (B.1.1.529) of SARS-CoV-2: Mutation, Infectivity, Transmission, and Vaccine Resistance. *World Journal of Clinical Cases*, 10 (1) January, pp. 1–11.

Ribero, M. S., Jouvenet, N., Dreux, M. and Nisole, S. (2020) Interplay between SARS-CoV-2 and the Type I Interferon Response. *PLOS Pathogens*, 16 (7) July, p. e1008737.

Ridley, M. and Chan, A. (2021) *Viral: The Search for the Origin of COVID-19*. Harper.

Robinot, R., Hubert, M., Melo, G. D. de, Lazarini, F., Bruel, T., Smith, N., et.al. (2021) SARS-CoV-2 Infection Induces the Dedifferentiation of Multiciliated Cells and Impairs Mucociliary Clearance. *Nature Communications*, 12 (1) July, p. 4354.

Rodrigues, C. H., Pires, D. E. and Ascher, D. B. (2018) DynaMut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability. *Nucleic Acids Research*, 46 (Web Server issue) July, pp. W350–W355.

Rosas, I. O., Bräu, N., Waters, M., Go, R. C., Hunter, B. D., Bhagani, S., et.al. (2021) Tocilizumab in Hospitalized Patients with Severe Covid-19 Pneumonia. *New England Journal of Medicine*, 384 (16) April, pp. 1503–1516.

Roshdy, W. H., Kandeil, A., El-Shesheny, R., Khalifa, M. K., Al-Karmalawy, A. A., Showky, S., et.al. (2022) Insight into Genetic Characteristics of Identified SARS-CoV-2 Variants in Egypt from March 2020 to May 2021. *Pathogens*, 11 (8) July, p. 834.

Rostami, A., Sepidarkish, M., Leeflang, M. M. G., Riahi, S. M., Nourollahpour Shiadeh, M., Esfandyari, S., et.al. (2021) SARS-CoV-2 Seroprevalence Worldwide: A Systematic Review and Meta-Analysis. *Clinical Microbiology and Infection*, 27 (3) March, pp. 331–340.

Sahin, E., Bozdayi, G., Yigit, S., Muftah, H., Dizbay, M., Tunccan, O. G., et.al. (2021) Genomic Characterization of SARS-CoV-2 Isolates from Patients in Turkey Reveals the Presence of Novel Mutations in Spike and Nsp12 Proteins. *Journal of Medical Virology*, 93 (10) October, pp. 6016–6026.

Saldivar-Espinoza, B., Garcia-Segura, P., Novau-Ferré, N., Macip, G., Martínez, R., Puigbò, P., et.al. (2023) The Mutational Landscape of SARS-CoV-2. *International Journal of Molecular Sciences*, 24 (10) May, p. 9072.

Sampaio, N. G., Chauveau, L., Hertzog, J., Bridgeman, A., Fowler, G., Moonen, J. P., et.al. (2021) The RNA Sensor MDA5 Detects SARS-CoV-2 Infection. *Scientific Reports*, 11 (1) July, p. 13638.

SARS-CoV-2 Variant Classifications and Definitions (2021) [Online]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html> [Accessed 21 August 2023v].

Schneider, W. M., Luna, J. M., Hoffmann, H.-H., Sánchez-Rivera, F. J., Leal, A. A., Ashbrook, A. W., et.al. (2021) Genome-Scale Identification of SARS-CoV-2 and Pan-Coronavirus Host Factor Networks. *Cell*, 184 (1) January, pp. 120-132.e14.

Schwartz, D. A. (2020) An Analysis of 38 Pregnant Women With COVID-19, Their Newborn Infants, and Maternal-Fetal Transmission of SARS-CoV-2: Maternal Coronavirus Infections and Pregnancy Outcomes. *Archives of Pathology & Laboratory Medicine*, 144 (7) March, pp. 799–805.

Seo, G., Lee, G., Kim, M. J., Baek, S.-H., Choi, M., Ku, K. B., et.al. (2020) Rapid Detection of COVID-19 Causative Virus (SARS-CoV-2) in Human Nasopharyngeal Swab Specimens Using Field-Effect Transistor-Based Biosensor. *ACS Nano*, 14 (4) April, pp. 5135–5142.

Seow, J., Graham, C., Merrick, B., Acors, S., Steel, K. J. A., Hemmings, O., et.al. (2020) Longitudinal Evaluation and Decline of Antibody Responses in SARS-CoV-2 Infection [Online]. medRxiv, p.

2020.07.09.20148429. Available from: <https://www.medrxiv.org/content/10.1101/2020.07.09.20148429v1> [Accessed 3 April 2025].

Sette, A. and Crotty, S. (2021) Adaptive Immunity to SARS-CoV-2 and COVID-19. *Cell*, 184 (4) February, pp. 861–880.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A. and Li, F. (2020) Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature*, 581 (7807) May, pp. 221–224.

Sharifi, A. and Khavarian-Garmsir, A. R. (2020) The COVID-19 Pandemic: Impacts on Cities and Major Lessons for Urban Planning, Design, and Management. *Science of The Total Environment*, 749 December, p. 142391.

Shimabukuro, T. T., Kim, S. Y., Myers, T. R., Moro, P. L., Oduyebo, T., Panagiotakopoulos, L., et.al. (2021) Preliminary Findings of mRNA Covid-19 Vaccine Safety in Pregnant Persons. *New England Journal of Medicine*, 384 (24) June, pp. 2273–2282.

Shu, Y. and McCauley, J. (2017) GISAID: Global Initiative on Sharing All Influenza Data – from Vision to Reality. *Eurosurveillance*, 22 (13) March, p. 30494.

Silva, S. J. R. da, Pardee, K. and Pena, L. (2020) Loop-Mediated Isothermal Amplification (LAMP) for the Diagnosis of Zika Virus: A Review. *Viruses*, 12 (1) January, p. 19.

Smith, M. J., Hayward, S. A., Innes, S. M. and Miller, A. S. C. (2020) Point-of-care Lung Ultrasound in Patients with COVID-19 – a Narrative Review. *Anaesthesia*, 75 (8) August, pp. 1096–1104.

Smyrlaki, I., Ekman, M., Lentini, A., Rufino De Sousa, N., Papanicolaou, N., Vondracek, M., et.al. (2020) Massive and Rapid COVID-19 Testing Is Feasible by Extraction-Free SARS-CoV-2 RT-PCR. *Nature Communications*, 11 (1) September, p. 4812.

Sola, I., Almazán, F., Zúñiga, S. and Enjuanes, L. (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology*, 2 (Volume 2, 2015) November, pp. 265–288.

Strategic Preparedness and Response Plan: April 2023–April 2025 - From Emergency Response to Long-Term COVID-19 Disease Management: Sustaining Gains Made during the COVID-19 Pandemic, 3 May 2023 - World | ReliefWeb (2023) [Online]. Available from: <https://reliefweb.int/report/world/strategic-preparedness-and-response-plan-april-2023-april-2025-emergency-response-long-term-covid-19-disease-management-sustaining-gains-made-during-covid-19-pandemic-3-may-2023> [Accessed 9 May 2025].

Suo, T., Liu ,Xinjin, Feng ,Jiangpeng, Guo , et.al. (2020) ddPCR: A More Accurate Tool for SARS-CoV-2 Detection in Low Viral Load Specimens. *Emerging Microbes & Infections*, 9 (1) January, pp. 1259–1268.

Tang, Y., Zhang, P., Wang, Y., Wang, J., Su, M., Wang, Y., et.al. (2020) The Biogenesis, Biology, and Clinical Significance of Exosomal PD-L1 in Cancer. *Frontiers in Immunology* [Online], 11 April. Available from: <https://www.frontiersin.orghttps://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2020.00604/full> [Accessed 11 April 2025].

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et.al. (2021) Detection of a SARS-CoV-2 Variant of Concern in South Africa. *Nature*, 592 (7854) April, pp. 438–443.

Temmam, S., Vongphayloth, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., et.al. (2022) Bat Coronaviruses Related to SARS-CoV-2 and Infectious for Human Cells. *Nature*, 604 (7905) April, pp. 330–336.

Tombuloglu, H., Sabit, H., Al-Suhaimi, E., Jindan, R. A. and Alkharsah, K. R. (2021) Development of Multiplex Real-Time RT-PCR Assay for the Detection of SARS-CoV-2. *PLOS ONE*, 16 (4) April, p. e0250942.

Tracking SARS-CoV-2 Variants (2022) [Online]. Available from: <https://www.who.int/activities/tracking-SARS-CoV-2-variants> [Accessed 29 March 2025w].

Updated Working Definitions and Primary Actions for SARSCoV2 Variants (2022) [Online]. Available from: <https://www.who.int/publications/m/item/updated-working-definitions-and-primary-actions-for--sars-cov-2-variants> [Accessed 1 April 2024x].

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

Wafaa EI-Sadr; Ayman EI-Mohanders (2025) *Commentary: Five Years after COVID, We Are Sitting Ducks* [Online]. Available from: <https://www.timesunion.com/opinion/article/commentary-five-years-covid-sitting-ducks-20245997.php> [Accessed 9 May 2025].

Walsh, E. E., Frenck, R. W., Falsey, A. R., Kitchin, N., Absalon, J., Gurtman, A., et.al. (2020) Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *New England Journal of Medicine*, 383 (25) December, pp. 2439–2450.

Wamai, R. G., Hirsch, J. L., Van Damme, W., Alnwick, D., Bailey, R. C., Hodgins, S., et.al. (2021) What Could Explain the Lower COVID-19 Burden in Africa despite Considerable Circulation of the SARS-CoV-2 Virus? *International Journal of Environmental Research and Public Health*, 18 (16) January, p. 8638.

Wang, X., Shao, C., Huang, X., Sun, L., Meng, L., Liu, H., Zhang, S., Li, H. and Lv, F. (2021) Histopathological Features of Multiorgan Percutaneous Tissue Core Biopsy in Patients with COVID-19. *Journal of Clinical Pathology*, 74 (8) August, pp. 522–527.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics*, 25 (9) May, pp. 1189–1191.

Webb, G. F., Blaser, M. J., Zhu, H., Ardal, S. and Wu, J. (2004) Critical Role of Nosocomial Transmission in the Toronto Sars Outbreak. *Mathematical biosciences and engineering: MBE*, 1 (1) June, pp. 1–13.

WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020 (n.d.) [Online]. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> [Accessed 22 March 2025y].

WHO, W. H. (2020) Clinical Management of COVID-19: Interim Guidance, 27 May 2020. [Online]. Available from: <https://iris.who.int/handle/10665/332196> [Accessed 1 April 2025].

Wolff, C. M., Singer, D., Schmidt, A. and Bekeschus, S. (2023) Immune and Inflammatory Responses of Human Macrophages, Dendritic Cells, and T-Cells in Presence of Micro- and Nanoplastic of Different Types and Sizes. *Journal of Hazardous Materials*, 459 October, p. 132194.

Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., et.al. (2012) Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, 86 (7) April, pp. 3995–4008.

Worobey, M., Levy, J. I., Malpica Serrano, L., Crits-Christoph, A., Pekar, J. E., Goldstein, S. A., et.al. (2022) The Huanan Seafood Wholesale Market in Wuhan Was the Early Epicenter of the COVID-19 Pandemic. *Science*, 377 (6609) August, pp. 951–959.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., et.al. (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe*, 27 (3) March, pp. 325–328.

Wu, C., Yin, W., Jiang, Y. and Xu, H. E. (2022) Structure Genomics of SARS-CoV-2 and Its Omicron Variant: Drug Design Templates for COVID-19. *Acta Pharmacologica Sinica*, 43 (12) December, pp. 3021–3033.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et.al. (2020) A New Coronavirus Associated with Human Respiratory Disease in China. *Nature*, 579 (7798) March, pp. 265–269.

Wu, Z. and McGoogan, J. M. (2020) Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*, 323 (13) April, pp. 1239–1242.

Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y.-C., Wang, H., et.al. (2020) Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports*, 33 (1) October, p. 108234.

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X. and Shan, H. (2020) Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology*, 158 (6) May, pp. 1831-1833.e3.

Yan, C., Cui, J., Huang, L., Du, B., Chen, L., Xue, G., et.al. (2020) Rapid and Visual Detection of 2019 Novel Coronavirus (SARS-CoV-2) by a Reverse Transcription Loop-Mediated Isothermal Amplification Assay. *Clinical Microbiology and Infection*, 26 (6) June, pp. 773–779.

Yang, H. and Rao, Z. (2021) Structural Biology of SARS-CoV-2 and Implications for Therapeutic Development. *Nature Reviews. Microbiology*, 19 (11), pp. 685–700.

Yang, L., Liu, S., Liu, J., Zhang, Z., Wan, X., Huang, B., Chen, Y. and Zhang, Y. (2020) COVID-19: Immunopathogenesis and Immunotherapeutics. *Signal Transduction and Targeted Therapy*, 5 (1) July, pp. 1–8.

Yoon, S. H., Lee, K. H., Kim, J. Y., Lee, Y. K., Ko, H., Kim, K. H., et.al. (2020) Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. *Korean Journal of Radiology*, 21 (4) April, pp. 494–500.

Yoshimoto, F. K. (2020) The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *The Protein Journal*, 39 (3) June, pp. 198–216.

Younes, N., Al-Sadeq, D. W., AL-Jighefee, H., Younes, S., Al-Jamal, O., Daas, H. I., et.al. (2020) Challenges in Laboratory Diagnosis of the Novel Coronavirus SARS-CoV-2. *Viruses*, 12 (6) May, p. 582.

Yu, F., Yan, L., Wang, N., Yang, S., Wang, L., Tang, Y., et.al. (2021) Quantitative Detection and Viral Load Analysis of SARS-CoV-2 in Infected Patients. [Online]. Available from: <https://dx.doi.org/10.1093/cid/ciaa345> [Accessed 3 April 2025].

Zaki, A. M., Boheemen, S. van, Bestebroer, T. M., Osterhaus, A. D. M. E. and Fouchier, R. A. M. (2012) Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367 (19) November, pp. 1814–1820.

Zaman, K., Shete, A. M., Mishra, S. K., Kumar, A., Reddy, M. M., Sahay, R. R., et.al. (2022) Omicron BA.2 Lineage Predominance in Severe Acute Respiratory Syndrome Coronavirus 2 Positive Cases during the Third Wave in North India. *Frontiers in Medicine*, 9 November, p. 955930.

Zandi, M., Shafaati, M., Kalantar-Neyestanaki, D., Pourghadamyari, H., Fani, M., Soltani, S., et.al. (2022) The Role of SARS-CoV-2 Accessory Proteins in Immune Evasion. *Biomedicine & Pharmacotherapy*, 156 December, p. 113889.

Zang, R., Gomez Castro, M. F., McCune, B. T., Zeng, Q., Rothlauf, P. W., Sonnek, N. M., et.al. (2020) TMPRSS2 and TMPRSS4 Promote SARS-CoV-2 Infection of Human Small Intestinal Enterocytes. *Science Immunology*, 5 (47) May, p. eabc3582.

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et.al. (2020) SARS-CoV-2 Spike-Protein D614G Mutation Increases Virion Spike Density and Infectivity. *Nature Communications*, 11 November, p. 6013.

Zhang, Y., Park, C., Bennett, C., Thornton, M. and Kim, D. (2021) Rapid and Accurate Alignment of Nucleotide Conversion Sequencing Reads with HISAT-3N. *Genome Research*, 31 (7) July, pp. 1290–1295.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et.al. (2020) Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study. *The Lancet*, 395 (10229) March, pp. 1054–1062.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et.al. (2020) A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature*, 579 (7798) March, pp. 270–273.

Zhou, R., Li, F., Chen, F., Liu, H., Zheng, J., Lei, C. and Wu, X. (2020) Viral Dynamics in Asymptomatic Patients with COVID-19. *International Journal of Infectious Diseases*, 96 July, pp. 288–290.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et.al. (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, 382 (8) February, pp. 727–733.

Zismann, V. and Nourbakhsh, M. (2014) Rapid Mapping of RNA 3′ and 5′ Ends [Online]. In: Alvarez, M. L. and Nourbakhsh, M. ed., *RNA Mapping: Methods and Protocols*. New York, NY: Springer, pp. 19–25. Available from: <https://doi.org/10.1007/978-1-4939-1062-5_2> [Accessed 2 April 2025].

List of Publications

Paper I: Hailu G, Legesse M, Mulu A, Medhin G, Tsegaye MM, Alemayehu DH, et al. SARS-CoV-2 Genetic Variants Identified in Selected Regions of Ethiopia Through Whole Genome Sequencing: Insights from the Fifth Wave of COVID-19. Genes. 2025 Mar;16(3):351.https://doi.org/10.3390/genes16030351

Paper II: Hailu G, Bitew M, Medhin G, et al. Diagnostic performance of RNA extraction-free dilution and heating method for the detection of severe acute respiratory syndrome coronavirus 2. Microbes & Immunity. 2025 Dec; x(x). https://doi.org/10.36922/MI025410114

Paper III: Hailu G, Bitew M, Medhin G, Tsegaye MM, Alemayehu DH, et al. Identification of Conserved SARS-CoV-2 Gene Regions for Monoclonal Antibody-Based Diagnostic Tool Development through Whole Genome Sequencing. Submitted manuscript

genes

MDPI

Article

SARS-CoV-2 Genetic Variants Identified in Selected Regions of Ethiopia Through Whole Genome Sequencing: Insights from the Fifth Wave of COVID-19

Getnet Hailu [1,2], Mengistu Legesse [1], Andargachew Mulu [3], Girmay Medhin [1], Mesfin Mengesha Tsegaye [3], Dawit Hailu Alemayehu [3], Abaysew Ayele [3], Atsbeha Gebreegziabxier [2], Adamu Tayachew [1,2], Adimkewu Aguine [2], Haileyesus Dejene [4], Sofonias K. Tessema [5], Harris Onywera [5], Assohoun Egomli Stanislas [6], Ebba Abate [1], Alessandro Marcello [7] and Molalegne Bitew [8,*]

1 Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa P.O. Box 1176, Ethiopia; getnethailu21@gmail.com (G.H.); mengistu.legessed@aau.edu.et (M.L.); gtmedhin@yahoo.com (G.M.); adamutayachew@gmail.com (A.T.); ebbaabate@yahoo.com (E.A.)
2 Ethiopian Public Health Institute, Addis Ababa P.O. Box 1242, Ethiopia; atsbehag@gmail.com (A.G.); adime12ag21@gmail.com (A.A.)
3 Armaur Hansson Research Institute, Addis Ababa P.O. Box 1005, Ethiopia; andargachew.mulu@ahri.gov.et (A.M.); mesfintiru@gmail.com (M.M.T.); dawit.hailu@ahri.gov.et (D.H.A.); abaysew.ayele@ahri.gov.et (A.A.)
4 College of Veterinary and Animal Science, University of Gondar, Gondar P.O. Box 196, Ethiopia; haileyesushdejene@uog.gov.et
5 Africa Centres for Disease Control and Prevention (Africa CDC), Addis Ababa P.O. Box 3243, Ethiopia; sofoniast@africa-union.org (S.K.T.); do79harris@gmail.com (H.O.)
6 Genomics and Metagenomics Platform, Pasteur Institute, Abidjan P.O. Box 75015, Côte d'Ivoire; stanlasso@gmail.com
7 Laboratory of Molecular Virology, International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano 99, 34149 Trieste, Italy; marcello@icgeb.org
8 Bio and Emerging Technology Institute, Addis Ababa P.O. Box 5954, Ethiopia
* Correspondence: molalegne23@yahoo.com

check for updates

Academic Editor: Silvia Turroni

Received: 13 February 2025
Revised: 19 February 2025
Accepted: 23 February 2025
Published: 18 March 2025

Citation: Hailu, G.; Legesse, M.; Mulu, A.; Medhin, G.; Tsegaye, M.M.; Alemayehu, D.H.; Ayele, A.; Gebreegziabxier, A.; Tayachew, A.; Aguine, A.; et al. SARS-CoV-2 Genetic Variants Identified in Selected Regions of Ethiopia Through Whole Genome Sequencing: Insights from the Fifth Wave of COVID-19. *Genes* 2025, *16*, 351. https://doi.org/10.3390/genes16030351

Abstract: Background: The COVID-19 pandemic highlighted SARS-CoV-2 variants with increased transmissibility and immune evasion. In Ethiopia, where cases surged, the understanding of the virus's dynamics was limited. This study analyzed SARS-CoV-2 variants during the fifth wave, crucial for guiding vaccines, therapeutics, diagnostics, and understanding disease severity. Method: From June to August 2022, 150 SARS-CoV-2-positive samples were randomly selected from the Ethiopian Public Health Institute repository. Sixty-three high-quality genome sequences were analyzed. Results: Of the 63 sequences, 70% were from males and 30% from females, with a median age of 34. Omicron dominated (97%, 61/63), primarily clade 22A (64%, 40/63), followed by 22B (18%, 11/63) and 21K (14%, 9/63). Delta accounted for 3.2% (2/63). Omicron was identified in all (25) vaccinated study participants. Ethiopian sequences showed limited evolutionary divergence and lower genetic diversity compared to global sequences. Conclusion: Omicron was the predominant variant during Ethiopia's fifth wave, indicating recent community transmission. Despite minor genetic diversity differences, ongoing surveillance remains critical for tracking variants and informing public health interventions.

Keywords: SARS-CoV-2; genetic divergence; whole genome sequencing; fifth wave; Ethiopia

1. Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which originated in Wuhan City, Hubei

Province, in China [1]. SARS-CoV-2 is transmitted from person to person through the inhalation of respiratory droplets from an infected individual or through direct contact with infected droplets through the eyes, mouth, or nose. However, transmissibility, incubation period, and infectivity vary depending on the variants and individual immunity [2].

Coronaviruses are enveloped RNA viruses belonging to the *Coronaviridae* family, characterized by a positive single-stranded genome exceeding 29,891 nucleotides, encoding for 9890 amino acids [3]. Their genome includes numerous open reading frames (ORFs) that produce structural (SPs) and non-structural proteins (NSPs) essential for the virus's life cycle and pathogenesis [4]. Structural proteins include the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, while non-structural proteins (nsp1–16) facilitate viral metabolism and interaction with the host immune system [5]. The spike glycoprotein of SARS-CoV-2 is cleaved into two subunits, S1 and S2, facilitating virus–cell membrane fusion [6]. The S1 subunit contains receptor-binding and N-terminal domains, crucial for viral entry and potential targets for neutralization. Host protease primes the S2 subunit [6]. Viral mutations can lead to new pathogenic variants affecting transmission, disease severity, and vaccine efficacy [7].

The COVID-19 pandemic was first reported in Africa on 14 February 2020, with the initial confirmed case detected in Egypt [8]. Ethiopia confirmed its first COVID-19 case in March 2020. The wide spread of COVID-19 led the World Health Organization (WHO) to declare the disease as a pandemic [9]. On 8 April 2020, the national government of Ethiopia declared a five-month state of emergency but allowed economic activities to continue during the public health crisis [10]. Ethiopia has experienced multiple pandemic waves, similar to the global trend. The first wave occurred from July to December 2020, followed by the second wave from January to July 2021. The third wave was recorded between July and December 2021, while the fourth wave spanned from December 2021 to April 2022. The fifth wave then emerged from June to August 2022. As of April 2024, from the total of more than 4 million tested, 501,157 positive cases were reported. From these, 488,171 and 7574 (%) individuals recovered and died, respectively. And only 38% of the population received at least one dose of a COVID-19 vaccine [11].

A sero-epidemiological survey conducted in Addis Ababa and Jimma between August 2020 and April 2022 revealed that over 96% of the study group had been exposed to SARS-CoV-2 at least once [12]. This figure indicates a high level of exposure across the population. Similarly, most Ethiopians have had multiple exposures to SARS-CoV-2, leading to high antibody titers with slow decay characteristics. This suggests the development of hybrid immunity due to recurrent infections with different variants and vaccination among many individuals [13].

At the beginning of the pandemic, Ethiopia, like many countries, primarily dealt with the initial SARS-CoV-2 strains. The virus's rapid spread and the emergence of variants were closely monitored as part of global efforts to understand the virus's behavior and adapt public health measures. As the virus continued to circulate, variants began to emerge. Variants of concern (VOCs) like Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529) have been identified in Africa [14,15]. In Ethiopia, similar variants have likely been detected [16]. Unfortunately, SARS-CoV-2 in the country was not sequenced during the actual pandemic period. This was due to a lack of laboratory infrastructure, skilled human resources, and reagents and kits. However, due to the relative improvement of sequencing infrastructure in the country, this project sequenced 63 retrospective samples from the national repository of the Ethiopian Public Health Institute collected during the fifth pandemic wave.

The dynamics of SARS-CoV-2 variants were not consistently monitored as part of the national surveillance systems in Ethiopia. This information is crucial for informing

public health interventions, monitoring the genomic epidemiology of the virus, identifying specific regional challenges, and supporting research and collaboration [13]. Only 824 SARS-CoV-2 samples were sequenced and submitted to GISAID from Ethiopia to date. This is relatively very small compared to other African countries like South Africa (56,886), Kenya (13,770), Egypt (5146), and Tunisia (2915). Out of 824 SARS-CoV-2 samples sequenced and submitted, 63 were for this project. This shows how SARS-CoV-2 sequencing activities were not conducted during the actual outbreak period. So, any information generated in this manuscript is very novel for the Ethiopian condition. The global science community can get information about what was the situation of SARS-CoV-2 variants in Ethiopia during the 5th wave of the pandemic from this manuscript as well as from the GISAID/NCBI submissions we made. This was the main rationale for initiating this study project. Therefore, this study was conducted to investigate the SARS-CoV-2 variants that circulated during the fifth wave of the pandemic in selected regions of Ethiopia.

2. Materials and Methods

2.1. Study Design, Period, and Sample Collection

A retrospective experimental study design was used. The study design was a retrospective experimental design. Sample collection was undertaken during the fifth wave of the pandemic (June to August, 2022) in Ethiopia. A total of 150 SARS-CoV-2-positive samples were collected retrospectively from the national virology reference laboratory repository at the Ethiopian Public Health Institute (EPHI). These samples were nasopharyngeal (NP) swab samples collected from patients residing in Addis Ababa, Oromia, Amhara, and Southern Nation Nationalities and Peoples (SNNP) regions (Figure 1). EPHI is a national reference laboratory for SARS-CoV-2 in Ethiopia, and samples were brought from those geographic areas, as well as from diplomatic communities and international travelers, for viral testing. Samples were stored at −80 °C for long-term storage. Samples for sequencing were purposively selected based on their identifiers (IDs) from the electronic national database. Only the IDs that tested positive and had a cycle threshold (Ct) value less than or equal to 30 were selected. As a result, 150 SARS-CoV-2-positive samples were included in this study due to our limited sequencing resources. The patients' metadata, including demographic information, clinical data, and RT-qPCR results, were anonymously extracted from electronic databases. The study participants were then grouped based on factors such as age, sex, signs and symptoms, pre-existing chronic diseases, travel history, and vaccination status.

2.2. Quality Assurance

To increase the success rate of whole genome sequencing, an RT-qPCR test was performed again after the samples were thawed at room temperature [17]. The RNA from these samples was re-extracted using the BGI extraction kit, and detection was performed using a BIO-RAD CFX96 Deep Well™ Real-Time PCR Detection system (Bio Rad Laboratories, Inc., Hercules, CA, USA) following the standard protocol. The clinical samples were handled and processed in a Biosafety Laboratory-2 (BSL-2) at the national virology reference laboratory. Out of the 150 re-analyzed samples, 70 had a Ct value less than or equal to 30 and were selected for whole genome sequencing, and others were discarded due to their high Ct values (i.e., >30). The RNA extracts from these 70 samples were sequenced using the Illumina NextSeq 550 platform, San Diego, CA, USA, at the virology laboratory of EPHI during March 2023. Out of 70 SARS-CoV-2 sequence data, 63 sequences passed the bioinformatics quality controls requirements (i.e., with genome coverage greater than 80%, no clustered mutations, and no misplaced stop codons) and were selected for further sequence analysis.

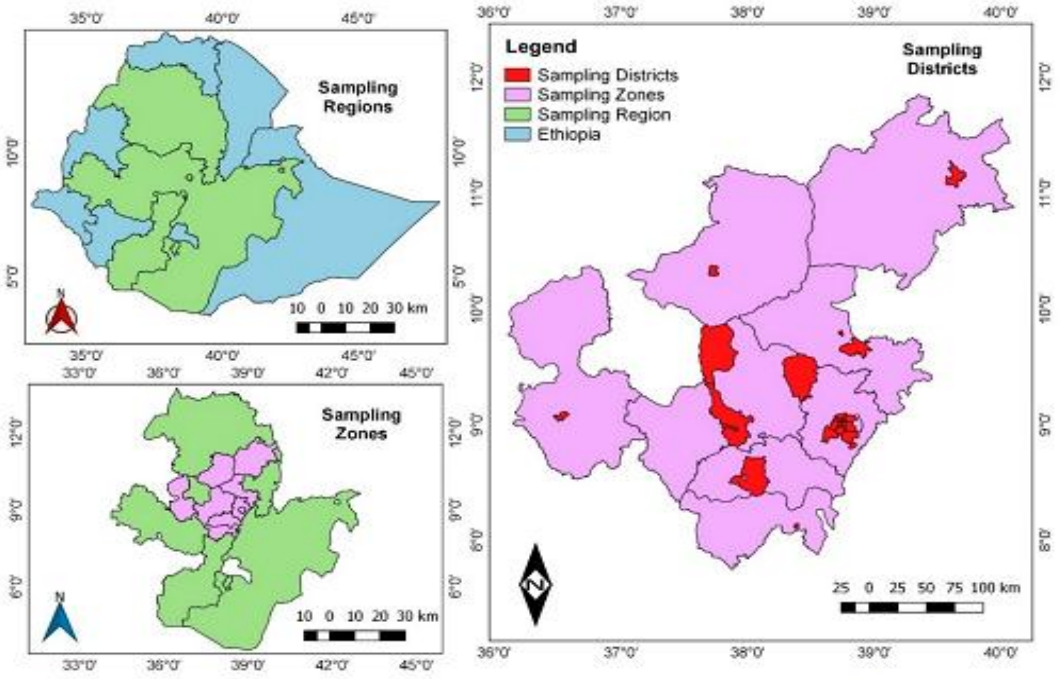


Figure 1. A geographic map of Ethiopia showing the regions and districts where SARS-CoV-2 specimens were collected.

2.3. *RNA Extraction and qRT-PCR*

Total RNA was extracted from 300 µL nasopharyngeal samples using MegaBio plus Virus Purification Kit II (BSC87S1E), (Hangzhou Bioer Technology Co., Ltd., Hangzhou, China), using the Bioner Gene Pure Pro fully automatic Nucleic Acid Purification System (NPA-32P) per the manufacturer's instructions [18]. Real-time qRT-PCR was performed to confirm the presence of the virus in the samples using primers and probes corresponding to SARS-CoV-2 with the QIAamp Viral RNA Mini Kit (Qiagen, Shanghai, China) according to the manufacturer's protocol [19]. qRT-PCR was performed using the BIO-RAD CFX96 Deep Well™ Real-Time PCR Detection system (Bio-Rad Laboratories, Inc., USA). All samples were tested for the human RNase P gene as an internal control.

2.4. *Complementary DNA (cDNA) Synthesis, DNA Library Preparation, and Whole Genome Sequencing*

Seventy RNA extracts, which exhibited low Ct values (Ct < 30) in quantitative real-time PCR assays, were chosen for library preparation. The cDNA synthesis, SARS-CoV-2 sequence enrichment, library amplification, and indexing were performed using the Illumina COVIDSeq test (RUO) (Illumina, USA) following the manufacturer's instructions. The resulting libraries were pooled, normalized, and quantified using a Qubit® 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) [20] before paired-end sequencing was performed on a NextSeq system (Illumina, USA) with a NextiSeq 500/550 cycle V3 kit (Illumina, USA).

2.5. Sequence and Metadata Analysis

The metadata were exported from the national database and converted into Excel format. The data were encoded into STATA v17, and descriptive analyses were performed. Raw sequencing data were obtained in FASTQ format and processed for variant calling and consensus sequence generation. First, the FASTQ sequences' quality was evaluated by the FastQC (v0.11.9) program, and then adapter sequences were trimmed from reads using Cutadapt software v.4.6 [21] to remove adaptor sequences and poor-quality reads. An average base quality of Q30 was used for trimming. The trimmed reads were then mapped against the SARS-CoV-2 WuhanHu-1 reference genome sequence (GenBank accession No. NC_045512.2) using the Burrows–Wheeler Alignment BWA mem v0.7.12 [22]. The consensus sequence was then generated using Samtools version 1.12. Moreover, SARS-CoV-2 variants were called with locally installed Pangolin (daily updated version [23], and mutations were identified with online Nextclade v3.8.2 Clade assignment, mutation calling, and sequence quality checks [24]. All viral genome sequences obtained from this study were submitted to the Global Initiative on Sharing All Influenza Data (GISAID) with accession IDs from EPI_ISL_19,147,618 to EPI_ISL_19302657 [25] and National Center for Biotechnology Information (NCBI) repositories with accession numbers from PQ140559 to PP069552 [26].

2.6. Computation of SNP Distances in Samples to the SARS-CoV-2 Reference Genome

SNP-dist v0.70 [27] was used to determine the number of SNPs in the 63 sequences compared to the SARS-CoV-2 reference genome. The hCoV-19/Wuhan/WIV04/2019 sequence from Wuhan, China, which consists of 29,891 base pairs, was used as a reference genome obtained from the GISAID. The snp-dist tool calculates the SNP distance matrix from a dataset of multiple sequence alignments (MSAs), where the sequences are of the same length. These distance matrices are commonly used in studies of infectious disease outbreaks. For this analysis, we only considered informative bases (A, T, C, and G) and not Ns. MSA was conducted using MAFFT v7.526 [28], employing the multiple alignment technique with the Fast Fourier Transform. The following parameters were applied: a gap opening penalty of 1.53 for group-to-group alignment and a gap extension penalty of 0.00 for group-to-group alignment. The block substitution matrix (BLOSUM) was selected, with a coefficient of 62, and the L-INS-I alignment method of MAFFT was chosen, known for its accuracy in alignments of $\leq$200 sequences.

2.7. Evolutionary Divergence Between Sequences and Overall Sequence Pairs

To estimate the evolutionary divergence between sequences as well as the average evolutionary divergence over all sequence pairs, MSA of the 63 sequences was performed using MAFFT v7.526 [28]. The output of the alignment was used to construct the phylogenetic tree via IQ-tree version v2.3.5-3 [29]. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 29,921 positions in the final dataset. The pairwise and overall mean (average) distances were used as proxies for evolutionary divergence.

For comparison purposes, the overall evolutionary divergence among all sequence pairs obtained from the GISAID uploaded during and before the study period was calculated. This analysis included the WIV04 genome and a total of 118 SARS-CoV-2 genome sequences from 62 countries. The sequences encompassed various variants, including former VOCs present in neighboring countries and different regions worldwide (Table 1).

Table 1. The WIV04 genome, along with 118 SARS-CoV-2 genome sequences from 62 countries, were obtained from the GISAID.

Classifications	Variants	Name	First Reported Country
VOC	alpha	GRY (B.1.1.7 + Q.*)	United Kingdom
VOC	Beta	GH/501Y.V2 (B.1.351 + B.1.351.2 + B.1.351 + 3)	South Africa
VOC	Gamma	GR/501Y.V3 (P.1 + P.1*)	Brazil and Japan
VOC	Omicron	GRA (B.1.1.529 + BA.*)	Botswana, South Africa, and Hong Kong
Variant of interest (VOI)	Lambda	GR/452Q.V1 (C.37 + C.37.1)	Peru
VOI	Mu	GH (B.1.621 + B.1.621.1)	Colombia
VOI		GRA (XBB.1.5 + XBB.1.5.*)	Australia, India, and Bangladesh
VOC		GRA (XBB.1.16 + XBB.1.1.16.*)	India
Variant under monitoring (VUM)		GRA (XBB + XBB.*, excluding XBB.1.5, XBB.1.16, XBB.1.9.1, XBB.1.9.2, XBB.2.3)	India
VUM		GRA (XBB.1.9.1 + XBB.19.1.*)	Indonesia, Singapore, and Israel
VUM		GH/490R (B.1.640 + B.1.640.*)	Congo and France. Except for the Gamma VOC

Note: These were sequences that we collected between 1 January and 31 July 2022 before our samples were collected between June and August 2022. There was a total of 30,150 positions in the final dataset.

2.8. Phylogenetic Analysis of SARS-CoV-2 Genome Sequences

The IQ-TREE v2.3.5.3 was used to compute the phylogenetic inference of the 63 SARS-CoV-2 genome sequences [29]. In addition to these 63 sequences, after the consensus sequences obtained in this research were aligned, the top 118 SARS-CoV-2 genomes were selected and downloaded from GISAID for phylogenetic analysis. First, an alignment analysis using MAFFT v7.526 was performed [28]. The MSA which had 30,058 nucleotide sites was then imported to IQ-TREE. The phylogeny involved determining the best-fit DNA substitution model using Model Finder [6], which tested 88 DNA models. The initial parsimony tree was created by the phylogenetic likelihood library (PLL). The model chosen according to the Bayesian Information Criterion (BIC) score (115,726.5435) was the TIM + F + I + G4. The model of rate heterogeneity was as follows: proportion of invariable sites: 0.8736 and Gamma shape alpha: 0.9901 with 4 categories. The rate parameter (R) was as follows: A-C: 1.0000, A-G: 1.6270, A-T: 0.5506, C-G: 0.5506, C-T: 3.9021, and G-T: 1.0000. A bootstrap analysis using the UFBoot2 method with 1000 replicates and ML heuristic method was performed. Additionally, branch lengths were optimized through ML on the original alignment. The initial trees for the heuristic search were generated using BioNJ algorithms. From these analyses, we selected the phylogenetic tree with the highest log-likelihood value (−55,910.033). The resulting phylogenetic trees were saved in Newick format and visualized using the Interactive Tree of Life interface [30]. Statistical analysis of the demographic and clinical information was performed using STATA v17. Frequency distribution and tabulation were performed to show the relationship between outcome and independent variables.

3. Results

3.1. Demographics and Clinical Characteristics of the Study Subjects

Out of 150 samples collected, 70 samples had a Ct value of less than 30 and were sequenced. Out of 70 samples sequenced, 63 samples passed the bioinformatics quality

control. Of the 63 genome sequences obtained, the samples were collected from patients in the age range of 14 to 75 years, with a median age of 34 years. Of the study participants, 32% (20/63) belong in the age group 21–30 years old, followed by 21% (13/60) in 31–40 years and 16% (10/63) in 41–50 years old. Females accounted for 30% (19/63) and males for 70% (44/63). The majority of the sequenced samples were collected from Addis Ababa (70%; 44/63), followed by Oromia (24%; 15/6), Amhara (5%; 3/63), and Southern Nation Nationalities Region (SNNPR) (2%; 1/63). In terms of clinical data, 70% (44/63) of the study participants exhibited signs and symptoms of cough, fever, and anosmia, and 43% (27/63) of participants reported they had known chronic disease. In addition, 13% (8/63) had a travel history outside Ethiopia, delayed 8–11 days before testing, and approximately 40% (25/63) reported having a vaccination history (Table 2).

Table 2. Socio-demographic characteristics of study participants.

Characteristics		Number (n)	Percentage (%)
Sex	Female	19	30
	Male	44	70
Age Group	≤20	8	13
	21–30	20	32
	31–40	13	21
	41–50	10	16
	51–60	4	6
	>60	8	13
Testing Sites	Oromia	15	24
	Amhara	3	5
	Addis Ababa	44	70
	SNNPR	1	1
Signs and Symptoms	No	19	30
	Yes	44	70
Co-morbidity	No	36	57
	Yes	27	43
Travel History	No	55	87
	Yes	8	13
Vaccination Status	Not vaccinated	38	60
	Vaccinated	25	40
Reason for Testing	Contact with cases	10	16
	Follow-up	2	3
	Suspect	43	68
	Travel purpose	8	13

3.2. Sequencing and Sequence Analysis

The majority of the sequences (99.8%) had high quality, with gap sizes ranging from 20 to 114 nt, accounting for approximately 0.4% of the total sequence. On average, each genome had 69.21 single nucleotide variations (SNVs) in different genes of SARS-CoV-

2. The highest numbers of single nucleotide mutations per clade of SARS-CoV-2 were identified in 22A and 21K (Figure 2).

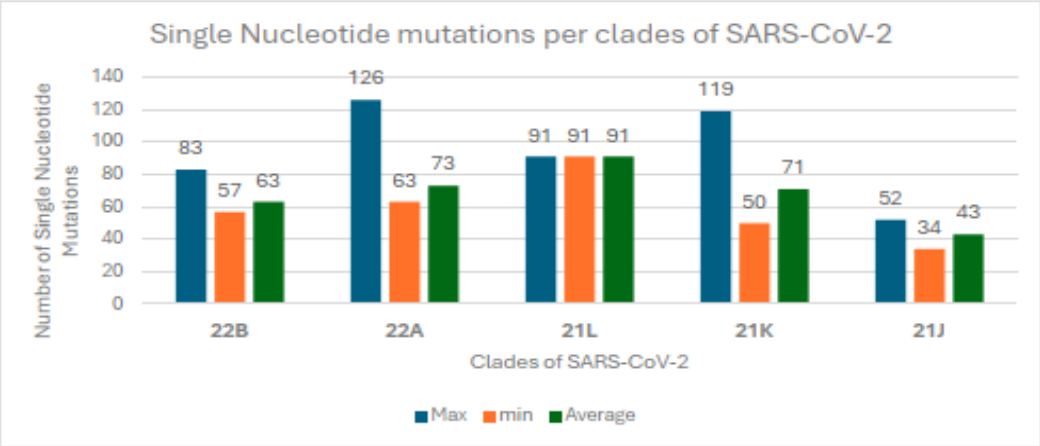


Figure 2. Number of single nucleotide mutations per clades of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia, 2022.

Nextclade variant analysis shows that out of the 63 sequences examined, on average, 69 nucleotide mutations were identified across the samples in comparison to the Wuhan reference genome (Figure 3).

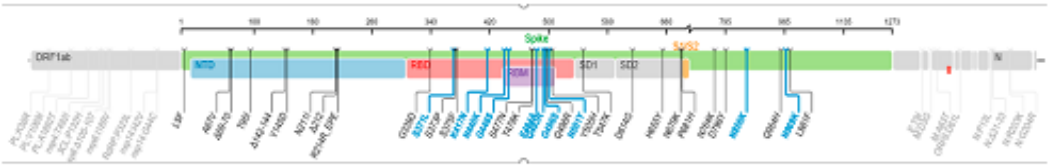


Figure 3. Example of nucleotide mutations in different gene positions of SARS-CoV-2 during the fifth wave of the pandemic in Ethiopia, 2022.

This analysis also detected an average of 198 nucleotide deletions, 68 substitutions, and 1 insertion in different genes of SARS-CoV-2 (Supplementary Table S1). These SARS-CoV-2 nucleotide mutations were observed in different positions of S, M, and N genes and genes coding for non-structural proteins (nsp1, nsp6, nsp9, nesp10, nsp13, ORF3a, and ORF7a) (Supplementary Materials). Among them, 371 nucleotide deletions were observed in the S gene in the positions of 25–27, 69–70, 124–144, 157–158, 212, 374–376, 415–432, 446–464, and 758–767, followed by 177 deletions in N in the position of 31–33 and 168 deletions in the nsp6 gene in the position of 105–108.

This study identified several mutations in the PLpro (papain-like protease), including K38R, V1069I, A1892T, D821N, E906A, T24I, G489S, I684V, S370L, V932L, P985S, T703I, A1179V, P822L, A488S, P1228L, and P1469S. Additionally, this study explores the presence of the P132H mutation in the 3CLpro (main protease), and P323L mutation in RdRP.

In this study, the transition/transversion bias (R) was 1.693, indicating that transitions occur more frequently than transversions. The base substitution rates (r) for each nucleotide pair were as follows: A => T (0.06), A => C (0.03), A => G (0.07), T => A (0.05), T => C

(0.17), T => G (0.03), C => A (0.05), C => T (0.29), C => G (0.03), G => A (0.11), G => T (0.06), and G => C (0.03). These results show that thymine (T) has the highest substitution rate with cytosine (C) (0.17) and the lowest with guanine (G) (0.03). Cytosine (C) has the highest substitution rate with thymine (T) (0.29), while its rates with adenine (A) and guanine (G) are relatively low (both 0.03). Adenine (A) and guanine (G) have generally similar substitution rates across different nucleotides, with A => G having the highest rate among adenine substitutions and G => A having the highest rate among guanine substitutions. In this study, the average Ti/Tv ratio was 1.9734, ranging from 1.5862 to 2.4737, and the most common nucleotide change was the G > A transition.

3.3. SARS-CoV-2 Variants Identified

This study found that omicron was the dominant variant (97%, 61/63) identified in the total sample sequenced during the fifth wave of the pandemic in selected regions of Ethiopia, followed by Delta variants (3%, 2/63). All of the identified cases of the Omicron variant belonged to four age groups (<20, 21–30, 31–40, 41–50, 51–60, and >60 years old). In the age groups of 21–30 and 41–50 years old, Delta accounted for 5% and 10% of the identified cases, respectively, and the remaining infections were caused by Omicron variants. In addition, all of the 25 individuals who were vaccinated before testing were infected with Omicron variants. Out of the 38 non-vaccinated study participants, 95% (36/38) were infected with Omicron, and 5.3% (2/38) were identified with Delta variants (Table 3).

Table 3. Frequency distribution of SARS-CoV-2 variants with demographic and clinical information of patients during the fifth wave of the pandemic in Ethiopia, 2022.

			Variants of SARS-CoV-2 Virus	
			Delta	Omicron
Sex	Female	Number	1	18
		%	5.3	94.7
	Male	Number	1	43
		%	2.3	97.7
Vaccination	No	Number	2	36
		%	5.3	94.7
	Yes	Number	0	25
		%	0.0	100.0
Signs and Symptoms	No	Number	0	19
		%	0.0	100.0
	Yes	Number	2	42
		%	4.5	95.5
Travel History	No	Number	2	53
		%	3.6	96.4
	Yes	Number	0	8
		%	0.0	100.0
Ethiopian Nationality	No	Number	0	1
		%	0.0	100.0
	Yes	Number	2	60
		%	3.2	96.8

Table 3. *Cont.*

			Variants of SARS-CoV-2 Virus	
			Delta	Omicron
Age Group	≤20	Number	0	8
		%	0.0	100.0
	21–30	Number	1	19
		%	5.0	95.0
	31–40	Number	0	13
		%	0.0	100.0
	41–50	Number	1	9
		%	10.0	90.0
	51–60	Number	0	4
		%	0.0	100.0
	>60	Number	0	8
		%	0.0	100.0

Almost comparable numbers of vaccinated and unvaccinated participants were infected with the BA.4.1 and BA.4.1.1 sub-lineages of SARS-CoV-2 (Figure 4). According to the Nextstrain classification, the dominant clades identified during the study period were 22A (64%, 40/63) and 22B (18%, 11/63), followed by 21K (Omicron; 14%, 9/63), 21J (Delta, 3.2%, 2/63), and 21L (Omicron; 1.6%, 1/63). Additionally, sub-lineages BA.4.1 (30.2%, 19/63) and BA.4.1.1 (28.5%, 18/63) were more prevalent than BA.1.1 (9.5%, 6/63), BA.5.2 (7.9%, 5/63), BF.2 (6.4%, 4/63), and BA.4 (4.8%, 3/63). Interestingly, 90% of the sub-lineages belonged to different strains within the BA lineage, which represents a specific genetic variant or group of closely related variants of the SARS-CoV-2 virus. Additionally, 6.4% (4/63) of the samples were part of the BF lineage, while the AY and B lineages each accounted for 1% of the samples.

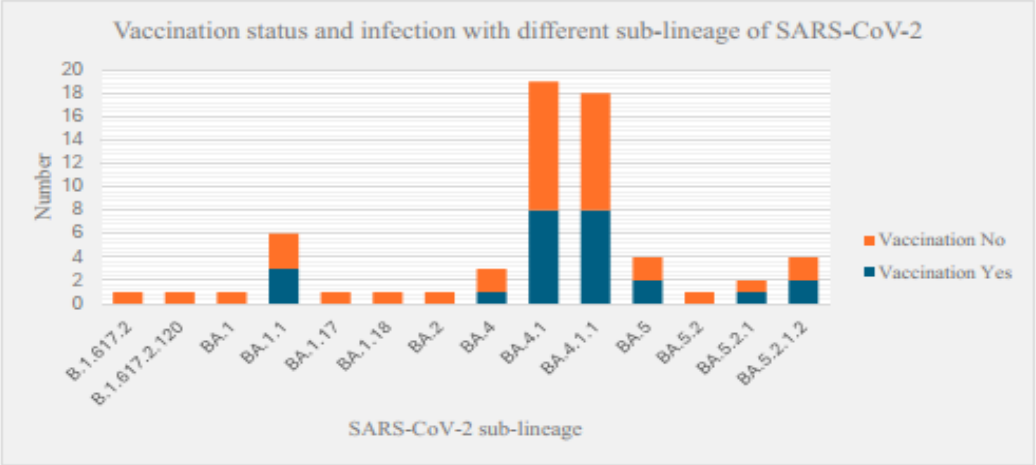


Figure 4. SARS-CoV-2 vaccination status and infection with different sub-lineages of SARS-CoV-2 identified during the fifth wave of the pandemic in Ethiopia.

3.4. Evolutionary Divergence and Phylogenetic Analysis of SARS-CoV-2

The evolutionary analysis among sequenced data for this study showed there is evidence of a phylogenetic relationship in some of the sequences. The evolutionary divergence between sequences ranges from 0.00 to 3.45×10^{-3} substitution per site. Looking at the broader analysis, the average evolutionary divergence across all sequence pairs is similarly low. For sequences from Ethiopia, the overall distance is calculated to be 9.78×10^{-4} substitutions per site, with a standard error of 8.47×10^{-5} substitutions per site. In comparison, when considering selected global sequences, the overall distance slightly increases to 1.54×10^{-3} substitutions per site, with a standard error of 1.19×10^{-4} substitutions per site. In addition, the maximum likelihood phylogenetic tree from our study, comprising 63 SARS-CoV-2 genomes, demonstrated seven main dominant clades in the viral population (Figure 5).

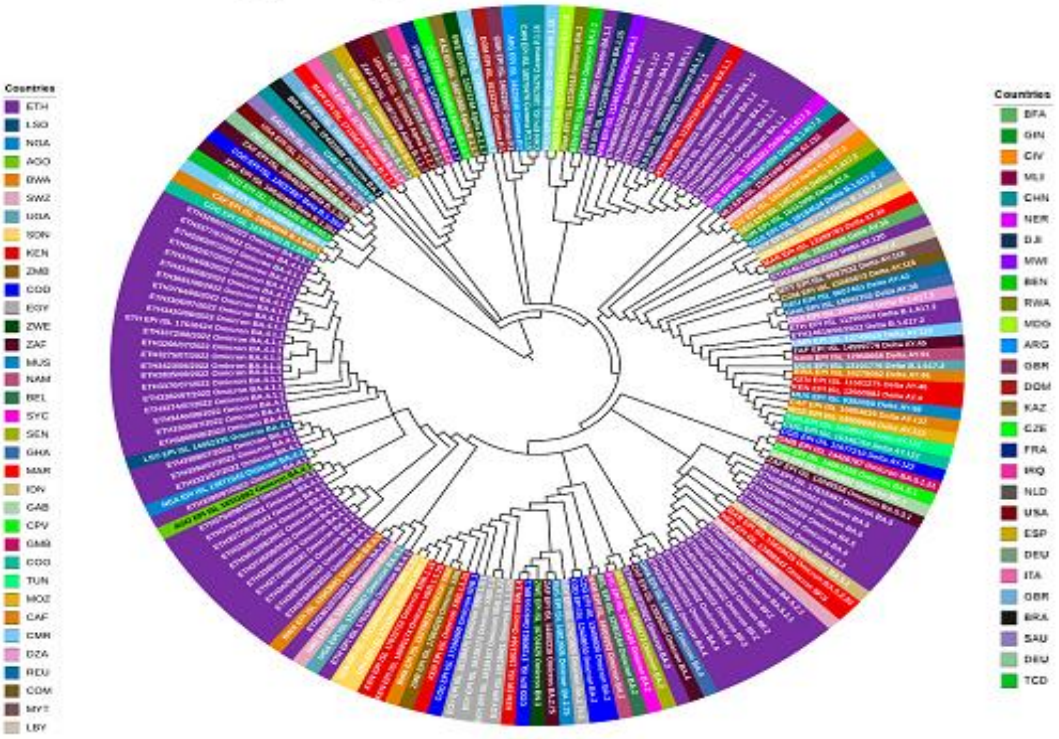


Figure 5. Analysis of the phylogenetic relatedness of SARS-CoV-2 identified in Ethiopia during the fifth wave of the pandemic, 2022, with selected global sequences.

4. Discussion

COVID-19, caused by the SARS-CoV-2 virus, first emerged in Wuhan, China, before spreading globally, including to Ethiopia. SARS-CoV-2 is primarily transmitted through respiratory droplets and direct contact. The spread of the virus is influenced by various factors, including sunlight, climate, temperature, humidity, and population demographics [31]. Asymptomatic and pre-symptomatic individuals play a significant role in the transmission of the virus [32]. It possesses a large RNA genome that encodes various proteins essential for its life cycle and pathogenesis, with the spike protein being particularly important for

viral entry and a key target for neutralization. Mutations in the virus can lead to the emergence of new variants that may impact transmission rates, disease severity, and vaccine efficacy [33].

In Ethiopia, the first COVID-19 case was reported in March 2020, and by April 2024, the country had recorded over 500,000 cases, with approximately 38% of the population vaccinated [34,35]. A sero-epidemiological survey indicated high exposure rates among the population [12]. While Ethiopia initially contended with early virus strains, variants of concern emerged globally, although local sequencing efforts were hampered by infrastructure challenges. This study aims to investigate the SARS-CoV-2 variants circulating during the fifth pandemic wave in selected regions of Ethiopia, addressing a critical gap in the country's genomic surveillance capabilities. The findings from this study provided significant insights into the demographics, clinical characteristics, and SARS-CoV-2 variants circulating in selected regions of Ethiopia during the fifth wave of the pandemic.

In this study, the age distribution of the study participants ranged from 14 to 75 years old, with a mean of 34 years old, revealing significant insights into the demographic distribution of COVID-19 cases within the sample. Notably, 32% of participants were aged 21–30 years, followed by 21% in the 31–40 age group and 16% in the 41–50 age range. This age distribution suggests a predominance of younger individuals in the sample. This finding is aligned with a previous study conducted in Ethiopia by Sisay A. et al. [16]; a high number of study participants infected with SARS-CoV-2 belonged to the age group of 21 to 30. However, different findings by Varnica B. et al. [33] and Carmola L.R. et al. [34] reported that the elderly population was the most significantly afflicted in terms of the number of COVID-19 cases and deaths. These differences might be due to the predominance of younger individuals in our study sample, which has the potential for the underrepresentation of older populations.

The geographical distribution of participants, with the majority from Addis Ababa, highlights the urban concentration of cases, which is consistent with a study that analyzed COVID-19 infections in Zimbabwe and emphasizes that urban areas, characterized by condensed residences, tend to expedite SARS-CoV-2 transmission, leading to a higher total number of confirmed cases [36]. This is echoed by Sharifi and Khavarian-Garmsir [37], who discussed how urban planning and design can significantly influence the infection and mortality rates of COVID-19 and suggested that urban centers are more susceptible to outbreaks due to their structural characteristics. Moreover, they added that the dynamics of urban life, including increased mobility and social interactions, further exacerbate the spread of the virus.

The clinical presentation of symptoms such as cough, fever, and anosmia in 70% of participants underscores the typical manifestations associated with SARS-CoV-2 infections. This finding is consistent with the study by Elavia et al. [38]. They reported that fever was observed in approximately 88% of cases, while cough was presented in about 68% of infected individuals. They suggested that these symptoms are critical for clinicians to recognize, as they are often the first indicators prompting testing for SARS-CoV-2. Furthermore, similar findings demonstrated that fever and cough, along with anosmia, serve as independent predictors of positive SARS-CoV-2 PCR results, reinforcing their significance in clinical assessments [39]. This also aligns with previous findings that highlight the prevalence of these symptoms in COVID-19 patients, emphasizing their importance in clinical diagnosis and public health surveillance, as shown by Sisay A. et al. [16] and Tarris G. et al. [40].

We observed that 70 out of 150 samples had Ct values below 30 after retesting. This is particularly noteworthy, as lower Ct values are indicative of higher viral loads [41]. This study also identified, from 70 samples that had a cycle threshold (Ct) value of less

than 30 and were sequenced, 63 samples passing bioinformatics quality control. This is supported by the study presented in [42] stressing the necessity of rigorous bioinformatics analysis to eliminate low-quality reads that could lead to erroneous conclusions regarding viral evolution. This suggests that the high pass rate of samples in this study highlights that the methodologies employed were robust and effective, which is critical for genomic surveillance efforts.

In this project, the sequencing analysis revealed a high-quality yield from the RNA extracts, with 99.8% of the sequences being of high quality and only 0.4% comprising gaps. This underscores the robustness of the sequencing methodologies employed, which align with the standards set by various genomic surveillance initiatives [26,43]. The genomic analysis of SARS-CoV-2 reveals significant insights into the virus's mutation landscape and its evolutionary dynamics. This study revealed, on average, approximately 69 single nucleotide variations (SNVs) per genome, particularly concentrated in clades 22A and 21K, which is similar to the study conducted by Saldivar-Espinoza et al. [44] that explored the mutational landscape of SARS-CoV-2 for the Omicron variant by early June 2022 and reported that the average number of SNVs per genome was around 72. In addition, the study by Harvey et al. [45] and Yang et al. [46], highlighted that the identification of high mutation rates, particularly in certain lineages, suggests that these variants may be under selective pressure, possibly due to host immune responses or therapeutic interventions.

The findings from the Nextclade variant analysis identified a total of 4429 nucleotide mutations, 12,664 deletions, 4392 substitutions, and 84 insertions across 63 SARS-CoV-2 sequences compared to the Wuhan reference genome. This provides critical insights into the genomic evolution of the virus [26]. This extensive mutational landscape reflects the ongoing adaptive changes that SARS-CoV-2 undergoes in response to selective pressures, including host immune responses and antiviral treatments [47]. In this study, significant nucleotide mutations were identified in critical genes such as the spike (S), membrane (M), and nucleocapsid (N) genes, as well as non-structural proteins (nsp1, nsp6, nsp9, nsp10, nsp13, ORF3a, and ORF7a). Similar studies have corroborated these findings, highlighting the high mutation rates observed in SARS-CoV-2. For instance, a study conducted in Turkey reported elevated mutation rates in the S and RdRP gene regions, consistent with the findings of this analysis [26,48]. Furthermore, research from Egypt indicated a significant presence of the D614G mutation, which is known to enhance viral infectivity and has been widely documented across various global studies [49].

Moreover, this study identified specific mutations in the papain-like protease (PLpro) and the main protease (3CLpro). This study is aligned with the studies conducted by Gao et al. [50] and Ismail et al. [51] that suggested the mutations K38R, V1069I, and P132H in PLpro, along with P323L in RdRP, could potentially alter the enzymatic activities of these proteins, impacting viral replication and immune evasion.

The phylogenetic analysis performed in this study indicates that the dominant circulating variants in Ethiopia during the study period (June to August, 2022) were primarily Omicron, with a smaller proportion of Delta variants. The Omicron variant (B.1.1.529) of SARS-CoV-2 was first identified in November 2021 in South Africa [52]. It was classified as a VOC by the WHO because of its high transmissibility and evasion from neutralizing antibodies induced by vaccination or natural infection with the wild-type virus [53]. As of 14 April 2022, the Omicron variant has been confirmed in at least 150 countries around the world [54]. From 3 January 2022 to 9 January 2022, the Omicron variant accounted for roughly 62.5% of all SARS-CoV-2 sequences available on the GISAID database [25]. In this study, the majority of the samples belonged to the BA lineage. This is consistent with findings reported from North India, 2022 [55], and the Philippines, 2022 [56], during the study period. The Centers for Disease Control and Prevention (CDC) has developed

taxonomic nomenclatures (BA, BF, AY, and B lineages) to categorize different variants of SARS-CoV-2 [57]. A study conducted by Rauseo AM et al. [58] suggests understanding the distribution of these variants is crucial for public health interventions, including vaccination strategies and treatment approaches.

This study explored the evolutionary divergence among genetic sequences of SARS-CoV-2 and revealed a spectrum of genetic distances ranging from nearly identical matches to slight variations. Specifically, comparisons within Ethiopian sequences revealed a perfect match between three samples and a very small distance between others, indicating recent community transmission. On a broader scale, the average genetic distance among Ethiopian sequences was low, suggesting high genetic similarity within the country. This finding is similar to a previous study conducted in Ethiopia through whole genome sequences of SARS-CoV-2 isolates from Ethiopian patients [59].

In addition, in this analysis, a comparison of the Ethiopian sequences with selected global sequences showed that a slightly higher average distance was observed. This indicates slightly more genetic diversity among global sequences. These findings were contrary to the evolution of SARS-CoV-2 being characterized by a steady increase in divergence within major lineages and a stepwise increase associated with each successive new major lineage, leading to a faster overall rate of evolution [60]. However, another study that analyzed 27,977 SARS-CoV-2 sequences from 84 countries obtained throughout the pandemic showed that SARS-CoV-2 genetic diversity was remarkably low [61].

Limitations of This Study

One limitation of our study is the small sample size that limits the power to give more conclusive recommendations. Another limitation is geographical coverage. Due to financial and infrastructure limitations, the sequencing was conducted very late; however, the data are still novel for readers who are interested in the SARS-CoV-2 variant situation in Ethiopia.

5. Conclusions

In conclusion, this study on SARS-CoV-2 variants in Ethiopia during the fifth wave reveals critical insights into the demographics, clinical characteristics, and genomic evolution of the virus, highlighting the predominance of the Omicron variant, particularly sub-lineages BA.4.1 and BA.4.1.1. The findings indicate a significant male predominance among participants, consistent with global trends of higher severity in males, and underscore the typical clinical manifestations of COVID-19, such as cough, fever, and anosmia. The genomic analysis suggests minimal genetic divergence among local sequences, indicating stable transmission dynamics, while the presence of specific mutations raises concerns regarding vaccine efficacy and potential immune evasion. These results emphasize the necessity for ongoing genomic surveillance to monitor variant emergence and inform public health strategies, contributing to a broader understanding of the pandemic's trajectory in Ethiopia and its implications for global health responses.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes16030351/s1, Table S1. SARS-CoV-2 nucleotide mutations in different genes position during the fifth wave the pandemic in Ethiopia.

Author Contributions: The contribution of the authors to the manuscript is as follows: Conceptualization: G.H. and M.L.; data curation: G.H. and D.H.A.; formal analysis and methodology: G.H., M.M.T., A.M. (Anargachew Mulu), D.H.A., A.M. (Alessandro Marcello), A.G., A.T., A.A. (Adimkewu Aguine), S.K.T., H.O., A.E.S., E.A., H.D., M.B., A.A. (Abaysew Ayele), E.A., G.M. and M.L.; investigation: G.H., M.M.T., D.H.A., A.G. and A.T.; supervision: M.L.; writing original draft: G.H.; writing—review and editing: G.H., M.M.T., A.M. (Andargachew Mulu), D.H.A., A.M. (Alessandro Marcello), A.G., A.T.,

A.A. (Adimkewu Aguine), S.K.T., H.O., H.D., A.A. (Abaysew Ayele), A.E.S., M.B., E.A., G.M. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: The sequencing of SARS-CoV-2 was done with the financial support of ICGEB-CRP with the grant code CRP/ETH21-01. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [Grant No. INV-034,799]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Gates Foundation. The first author got financial support for this research from Addis Ababa University.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were obtained from the Institutional Review Board (IRB) of the Aklilu Lemma Institute of Pathobiology at Addis Ababa University, under protocol number ALIPB IRB/52/2013/21 dated 19 April 2021. A permission letter was obtained from the Ethiopian Public Health Institute to use leftover samples from repositories.

Informed Consent Statement: Patient consent was waived due to the retrospective nature of this study, as approved by the Institutional Review Board.

Data Availability Statement: The sequence data are available at GISAID and NCBI.

Acknowledgments: The authors are grateful to the Ethiopian Public Health Institute, Armauer Hansen Research Institute, and Addis Ababa University for their support, coordination, and providing whole genome sequencing laboratories. The Bio and Emerging Technology Institute (BETin) and the International Center for Genetic Engineering and Biotechnology (ICGEB) are highly acknowledged for technical and financial support. We also gratefully thank the authors and the originating and submitting laboratories for their sequences and metadata shared through GISAID. The authors would like to express their special thanks to all who participated in data collection and laboratory activities.

Conflicts of Interest: The authors declare no competing interests.

References

1. Kumar, S.; Tao, Q.; Weaver, S.; Sanderford, M.; Caraballo-Ortiz, M.A.; Sharma, S.; Pond, S.L.K.; Miura, S. An Evolutionary Portrait of the Progenitor SARS-CoV-2 and Its Dominant Offshoots in COVID-19 Pandemic. *Mol. Biol. Evol.* **2021**, *38*, 3046–3059. [CrossRef] [PubMed]
2. WHO Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions. Available online: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions (accessed on 19 February 2025).
3. Farhud, D.D.; Mojahed, N. SARS-CoV-2 Notable Mutations and Variants: A Review Article. *Iran. J. Public Health* **2022**, *51*, 1494–1501. [CrossRef] [PubMed]
4. Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **2020**, *27*, 325–328. [CrossRef]
5. Huang, Y.; Yang, C.; Xu, X.-F.; Xu, W.; Liu, S.-W. Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141–1149. [CrossRef] [PubMed]
6. Cryo-EM Structure of the SARS-Coronavirus Spike Glycoprotein in Complex with Its Host Cell Receptor ACE2 | PLOS Pathogens. Available online: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007236 (accessed on 15 April 2024).
7. Malone, B.; Urakova, N.; Snijder, E.J.; Campbell, E.A. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 21–39. [CrossRef] [PubMed]
8. Africa CDC. Africa Identifies First Case of Coronavirus Disease: Statement by the Director of Africa CDC. Available online: https://africacdc.org/news-item/africa-identifies-first-case-of-coronavirus-disease-statement-by-the-director-of-africa-cdc/ (accessed on 9 June 2024).
9. Cucinotta, D.; Vanelli, M. WHO Declares COVID-19 a Pandemic. *Acta Biomed.* **2020**, *91*, 157–160. [CrossRef]
10. Ethiopia Confirms First Coronavirus Case. Available online: https://www.aa.com.tr/en/africa/ethiopia-confirms-first-coronavirus-case/1765130 (accessed on 15 April 2024).

11. Ethiopia COVID—Coronavirus Statistics—Worldometer. Available online: https://www.worldometers.info/coronavirus/country/ethiopia/ (accessed on 7 May 2024).
12. Gudina, E.K.; Ali, S.; Girma, E.; Gize, A.; Tegene, B.; Hundie, G.B.; Sime, W.T.; Ambachew, R.; Gebreyohanns, A.; Bekele, M.; et al. Seroepidemiology and model-based prediction of SARS-CoV-2 in Ethiopia: Longitudinal cohort study among front-line hospital workers and communities. *Lancet Glob. Health* **2021**, *9*, e1517. [CrossRef]
13. Merkt, S.; Ali, S.; Gudina, E.K.; Adissu, W.; Gize, A.; Muenchhoff, M.; Graf, A.; Krebs, S.; Elsbernd, K.; Kisch, R.; et al. Long-term monitoring of SARS-CoV-2 seroprevalence and variants in Ethiopia provides prediction for immunity and cross-immunity. *Nat. Commun.* **2024**, *15*, 3463. [CrossRef] [PubMed]
14. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **2021**, *592*, 438–443. [CrossRef]
15. Viana, R.; Moyo, S.; Amoako, D.G.; Tegally, H.; Scheepers, C.; Althaus, C.L.; Anyaneji, U.J.; Bester, P.A.; Boni, M.F.; Chand, M.; et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **2022**, *603*, 679–686. [CrossRef]
16. Sisay, A.; Tshiabuila, D.; Van Wyk, S.; Tesfaye, A.; Mboowa, G.; Oyola, S.O.; Tesema, S.K.; Baxter, C.; Martin, D.; Lessells, R.; et al. Molecular Epidemiology and Diversity of SARS-CoV-2 in Ethiopia, 2020–2022. *Genes* **2023**, *14*, 705. [CrossRef] [PubMed]
17. Sisay, A.; Abera, A.; Dufera, B.; Endrias, T.; Tasew, G.; Tesfaye, A.; Hartnack, S.; Beyene, D.; Desta, A.F. Diagnostic accuracy of three commercially available one step RT-PCR assays for the detection of SARS-CoV-2 in resource limited settings. *PLoS ONE* **2022**, *17*, e0262178. [CrossRef]
18. Bioer Technology Introduction. Available online: https://conference.medgenetics.ru/upload/Bioer%20Technology%20Product%20Intro-711_%D0%A0%D0%9C%D0%A2.pdf (accessed on 21 August 2023).
19. QIAGEN QIAamp® Viral RNA Mini Handbook, July 2020—Google Search. Available online: https://www.google.com/search?q=QIAGEN+QIAamp%C2%AE+Viral+RNA+Mini+Handbook,+July+2020.&rlz=1C1GCEU_enET1014ET1014&oq=QIAGEN+QIAamp%C2%AE+Viral+RNA+Mini+Handbook,+July+2020.&aqs=chrome.69i57.1026j0j4&sourceid=chrome&ie=UTF-8 (accessed on 21 August 2023).
20. Burrows-Wheeler Alignment (BWA)—Research Computing Documentation. Available online: https://wiki.rc.usf.edu/index.php/Burrows-Wheeler_Alignment_(BWA) (accessed on 4 September 2024).
21. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]
22. Burrows-Wheeler Aligner. Available online: https://bio-bwa.sourceforge.net/ (accessed on 7 May 2024).
23. O'Toole, Á.; Scher, E.; Underwood, A.; Jackson, B.; Hill, V.; McCrone, J.T.; Colquhoun, R.; Ruis, C.; Abu-Dahab, K.; Taylor, B.; et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* **2021**, *7*, veab064. [CrossRef]
24. Aksamentov, I.; Roemer, C.; Hodcroft, E.B.; Neher, R.A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **2021**, *6*, 3773. [CrossRef]
25. GISAID Initiative. Available online: https://www.epicov.org/epi3/frontend#28d967 (accessed on 21 August 2023).
26. NCBI Virus. Available online: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202,%20taxid:2697049&utm_source=nuccore&utm_medium=referral&Authors_idx%20q.op=AND=Hailu&SubmitterAffil_idx%20q.op=OR=Ethiopian%20Public%20Health%20Institute&SubmitterCountry_idx%20q.op=OR=Ethiopia (accessed on 18 September 2024).
27. GitHub—Tseemann/Snp-Dists: Pairwise SNP Distance Matrix from a FASTA Sequence Alignment. Available online: https://github.com/tseemann/snp-dists (accessed on 26 October 2023).
28. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
29. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
30. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation | Nucleic Acids Research | Oxford Academic. Available online: https://academic.oup.com/nar/article/49/W1/W293/6246398?login=true (accessed on 26 October 2023).
31. Matson, M.J.; Yinda, C.K.; Seifert, S.N.; Bushmaker, T.; Fischer, R.J.; Doremalen, N.V.; Lloyd-Smith, J.O.; Munster, V.J. Effect of Environmental Conditions on SARS-CoV-2 Stability in Human Nasal Mucus and Sputum. *Emerg. Infect. Dis.* **2020**, *26*, 2276. [CrossRef] [PubMed]
32. Sirakov, I.; Stankova, P.; Bakalov, D.; Mirani, Y.; Bardarska, L.; Paraskova, G.; Popov, I.; Alexandrova, A.; Dimitrov, G.; Mizgova, G.; et al. Retrospective Analysis of the Spread of SARS-CoV-2 in the Mediterranean Part of Bulgaria, During the First Wave of the Pandemic. *J. Pure Appl. Microbiol.* **2024**, *18*, 438–450. Available online: https://microbiologyjournal.org/retrospective-analysis-of-the-spread-of-sars-cov-2-in-the-mediterranean-part-of-bulgaria-during-the-first-wave-of-the-pandemic/ (accessed on 19 February 2025). [CrossRef]
33. Iqbal, S.; Saki, M.; Sirakov, I.; Akrami, S. Pathogenesis of the SARS-Coronavirus-2 and Potential Therapeutic Strategies. *J. Commun. Dis.* **2022**, 202–209. [CrossRef]

34. Improving Response as Ethiopia Marks Two Years Since First COVID-19 Case | WHO | Regional Office for Africa. Available online: https://www.afro.who.int/countries/ethiopia/news/improving-response-ethiopia-marks-two-years-first-covid-19-case (accessed on 19 February 2025).
35. Anteneh, A.; Getachew, A.; Kenera, M.; Salim, A.; Kedir, F.; Belayihun, B.; Felker-Kantor, E. COVID-19 vaccine hesitancy in Ethiopia: A latent class analysis. *BMC Public Health* **2024**, *24*, 2894. [CrossRef]
36. Silwangani, B.; Munodawafa, D.; Chadambuka, A. COVID-19 Infections and Deaths between Rural and Urban Provinces of Zimbabwe, 2020: A Spatial Variations Analysis. *Texila Int. J. Public Health* **2023**, *11*, 1–13. [CrossRef]
37. Sharifi, A.; Khavarian-Garmsir, A.R. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* **2020**, *749*, 142391. [CrossRef]
38. Elavia, N.; Sharma, N.; Li, S.; Wang, Y.; Milekic, B. An Atypical Presentation of Acute Pulmonary Embolism With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Pneumonia. *Cureus* **2020**, *12*, e8249. [CrossRef]
39. Lan, F.-Y.; Filler, R.; Mathew, S.; Buley, J.; Iliaki, E.; Bruno-Murtha, L.A.; Osgood, R.; Christophi, C.A.; Fernandez-Montero, A.; Kales, S.N. COVID-19 symptoms predictive of healthcare workers' SARS-CoV-2 PCR results. *PLoS ONE* **2020**, *15*, e0235460. [CrossRef] [PubMed]
40. Tarris, G.; de Rougemont, A.; Estienney, M.-A.; Journet, J.; Lariotte, A.-C.; Aubignat, D.; Rebibou, J.-M.; De La Vega, M.F.; Legendre, M.; Belliot, G.; et al. Chronic kidney disease linked to SARS-CoV-2 infection: A case report. *BMC Nephrol.* **2021**, *22*, 278. [CrossRef] [PubMed]
41. Kucirka, L.M.; Lauer, S.A.; Laeyendecker, O.; Boon, D.; Lessler, J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Ann. Intern. Med.* **2020**, *173*, 262–267. [CrossRef] [PubMed]
42. De Maio, N.; Walker, C.R.; Turakhia, Y.; Lanfear, R.; Corbett-Detig, R.; Goldman, N. Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2. *Genome Biol. Evol.* **2021**, *13*, evab087. [CrossRef] [PubMed]
43. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [CrossRef] [PubMed]
44. Saldivar-Espinoza, B.; Garcia-Segura, P.; Novau-Ferré, N.; Macip, G.; Martínez, R.; Puigbò, P.; Cereto-Massagué, A.; Pujadas, G.; Garcia-Vallve, S. The Mutational Landscape of SARS-CoV-2. *Int. J. Mol. Sci.* **2023**, *24*, 9072. [CrossRef] [PubMed]
45. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef]
46. Yang, L.; Liu, S.; Liu, J.; Zhang, Z.; Wan, X.; Huang, B.; Chen, Y.; Zhang, Y. COVID-19: Immunopathogenesis and Immunotherapeutics. *Sig Transduct. Target. Ther.* **2020**, *5*, 128. [CrossRef]
47. Gao, Z.; Xu, Y.; Sun, C.; Wang, X.; Guo, Y.; Qiu, S.; Ma, K. A systematic review of asymptomatic infections with COVID-19. *J. Microbiol. Immunol. Infect.* **2021**, *54*, 12–16. [CrossRef]
48. Sahin, E.; Bozdayi, G.; Yigit, S.; Muftah, H.; Dizbay, M.; Tunccan, O.G.; Fidan, I.; Caglar, K. Genomic characterization of SARS-CoV-2 isolates from patients in Turkey reveals the presence of novel mutations in spike and nsp12 proteins. *J. Med. Virol.* **2021**, *93*, 6016–6026. [CrossRef]
49. Roshdy, W.H.; Kandeil, A.; El-Shesheny, R.; Khalifa, M.K.; Al-Karmalawy, A.A.; Showky, S.; Naguib, A.; Elguindy, N.; Fahim, M.; Abu Elsood, H.; et al. Insight into Genetic Characteristics of Identified SARS-CoV-2 Variants in Egypt from March 2020 to May 2021. *Pathogens* **2022**, *11*, 834. [CrossRef] [PubMed]
50. Gao, X.; Qin, B.; Chen, P.; Zhu, K.; Hou, P.; Wojdyla, J.A.; Wang, M.; Cui, S. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sin. B* **2021**, *11*, 237–245. [CrossRef] [PubMed]
51. Ismail, G.; Abdelghaffar, H.; Seadawy, M.G.; El-Hosseny, M.F.; Gad, A.F.; Ageez, A.; ElShafei, A.; Mohammed, S.S.; Ali, M.S.; El-Ashry, M. abd E.-R. Genome sequencing reveals existence of SARS-CoV-2 B.1.1.529 variant in Egypt. *J. Genet. Eng. Biotechnol.* **2022**, *20*, 70. [CrossRef] [PubMed]
52. Ren, S.-Y.; Wang, W.-B.; Gao, R.-D.; Zhou, A.-M. Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. *World J. Clin. Cases* **2022**, *10*, 1–11. [CrossRef] [PubMed]
53. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. Available online: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (accessed on 23 April 2024).
54. Omicron Variant (COVID-19/SARS-CoV-2) by BNO News. Available online: https://www.google.com/maps/d/viewer?mid=1lKX8ikHpKAkWCvuXn2g-CVW3OmHbN-k- (accessed on 23 April 2024).
55. Zaman, K.; Shete, A.M.; Mishra, S.K.; Kumar, A.; Reddy, M.M.; Sahay, R.R.; Yadav, S.; Majumdar, T.; Pandey, A.K.; Dwivedi, G.R.; et al. Omicron BA.2 lineage predominance in severe acute respiratory syndrome coronavirus 2-positive cases during the third wave in North India. *Front. Med.* **2022**, *9*, 955930. [CrossRef] [PubMed]
56. Li, Y.-T.; Polotan, F.G.M.; Sotelo, G.I.S.; Dolor, A.Y.M.; Tujan, M.A.A.; Gomez, M.R.R.; Onza, J.T.; Chang, A.K.T.; Bautista, C.T.; Carandang, J.C.; et al. Lineage BA.2 dominated the Omicron SARS-CoV-2 epidemic wave in the Philippines. *Virus Evol.* **2022**, *8*, veac078. [CrossRef]

ACCSCIENCE PUBLISHING

Microbes & Immunity

ORIGINAL RESEARCH ARTICLE

Diagnostic performance of RNA extraction-free dilution and heating method for the detection of severe acute respiratory syndrome coronavirus 2

Getnet Hailu[1,2*], Molalegne Bitew[3*], Girmay Medhin[1], Adamu Tayachew[1,2], Nebiyou Yemanebirhan[1,2], Andargachew Gashu[1,2], Samuel Ayele[4], Admikew Agune[2], Assefa Konda[2], and Mengistu Legesse[1]

[1]Department of Tropical and Infectious Diseases, Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia

[2]Department of Infectious Diseases, Ethiopian Public Health Institute, Addis Ababa, Ethiopia

[3]Department of Health Biotechnology, Bio and Emerging Technology Institute, Addis Ababa, Ethiopia

[4]USAID Eliminate TB Project, Addis Ababa, Ethiopia

Abstract

Real-time quantitative polymerase chain reaction remains the gold standard for COVID-19 diagnosis, but RNA extraction is time-consuming, expensive, and associated with increased biosafety requirements. This study evaluated an extraction-free dilution and heating (EFDH) method as a simplified alternative to conventional extraction-based (EB) real-time quantitative polymerase chain reaction for severe acute respiratory syndrome coronavirus 2 detection. A total of 300 archived nasopharyngeal specimens, including 190 positives and 110 negatives, from the National Virology Reference Laboratory at the Ethiopian Public Health Institute, were analyzed. Samples were diluted 1:2 with RNase-free water, heated at 72°C for 15 min, and analyzed using an ABI 7500 Fast instrument. The EFDH method showed a sensitivity of 92%, a specificity of 100%, and an overall accuracy of 85.8%, producing 15 false-negative and 12 invalid results. Agreement with the EB method was high, with 95% concordance and a kappa coefficient of 0.89. Performance was strongest in samples with high viral loads (cycle threshold [Ct] < 20) and declined in those with low viral loads (Ct > 35). A significant correlation was observed between the two methods ($R^2 = 0.99$, p=0.001). These findings indicate that the EFDH approach reliably detects moderate-to-high viral loads and may serve as a practical testing option in resource-limited settings, especially during outbreaks when rapid and simplified workflows are needed.

Keywords: SARS-CoV-2; Extraction free; Heating; Dilution; Diagnostic performance

***Corresponding authors:**
Molalegne Bitew (molalegne.bitew@betin.gov.et)
Getnet Hailu (getnethailu@ephi.gov.et)

Citation: Hailu G, Bitew M, Medhin G, *et al.* Diagnostic performance of RNA extraction-free dilution and heating method for the detection of severe acute respiratory syndrome coronavirus 2. *Microbes & Immunity* doi: 10.36922/MI025410114

Received: October 9, 2025

Revised: November 21, 2025

Accepted: November 25, 2025

Published online: December 10, 2025

1. Introduction

Coronaviruses belong to the *Coronaviridae* family. They are RNA viruses with an enveloped structure and a positive single-stranded genome. Their genome contains over 29,891 nucleotides, which encode for 9,890 amino acids. It also contains numerous

open-reading frames (ORFs) that encode both structural and non-structural proteins. These proteins play crucial roles in the viral life cycle and pathogenic processes.[1] The structural proteins consist of S (spike protein), E (envelope protein), M (membrane protein), and N (nucleocapsid protein). Furthermore, there are 16 non-structural proteins that assist in viral metabolism and interaction with the host immune system.[2] The genetic information of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent for Coronavirus Disease 2019 (COVID-19), plays a multifaceted role in diagnostic tool development, encompassing rapid assay development, enhanced detection sensitivity, variant identification, and informing public health responses.[3,4] Its significance underscores the critical intersection of genetics, diagnostics, and public health in effectively managing infectious disease threats.[4]

In viral outbreaks such as COVID-19, it is crucial to effectively manage and control the pandemics.[5] It is reported that up to 40% of individuals with SARS-CoV-2 infection may be asymptomatic or pre-symptomatic, yet have the potential to spread the virus.[6] Therefore, a dependable laboratory diagnosis is the key to preventing and controlling infections.[7]

COVID-19 can be diagnosed primarily by either immunological or molecular tests. Immunological tests include serological tests that detect antibodies (immunoglobulin [Ig]M and IgG) in blood or viral antigens in respiratory secretions obtained through nasal or throat swabs. These tests can be performed using point-of-care platforms, rapid testing systems that provide results within minutes and are particularly useful in non-laboratory settings. Serological tests are particularly important for diagnosing patients with mild to moderate disease when molecular diagnostics are not available. Serological tests play a crucial role in identifying individuals who have developed immunity against COVID-19.[8] On the other hand, molecular tests detect SARS-CoV-2 RNA in various specimens, primarily nasopharyngeal samples. Unlike serological tests, most molecular tests require specialized laboratory infrastructure, including advanced equipment and highly trained staff. This limitation restricts their widespread use compared to simpler immunological tests.[9] In addition to these primary diagnostic approaches, other laboratory parameters have been utilized to monitor patients with COVID-19.[10-12]

The specimen of choice for SARS-CoV-2 detection can be swabs from the nasopharynx, oropharynx, and anus, as well as saliva and sputum samples. In addition, the virus can be present in other bodily fluids, such as blood, urine, and feces.[13] The nasopharyngeal swab collection technique is among the most commonly used methods, with sensitivities reaching up to 98%.[14] Nasopharyngeal swabbing is regarded as the standard method for safely and conveniently collecting SARS-CoV-2 samples, as it minimizes direct exposure of the healthcare provider to respiratory secretions.[15]

The "gold standard" assay for diagnosing both symptomatic and asymptomatic cases is real-time quantitative polymerase chain reaction (RT-qPCR).[16] The first protocol, recommended by the World Health Organization (WHO), was published in 2020 by the Charité Institute at Berlin University in Germany.[4] This protocol utilizes TaqMan technology and specific primers and probes to detect the RNA-dependent RNA polymerase (*RdRp*), *E*, and *N* genes. In addition, Table 1 lists several in-house methods reported by WHO that are currently being validated in partner laboratories.[4]

Typically, three steps should be completed before conducting quantitative polymerase chain reaction (qPCR). These steps include: (i) purifying total RNA from the sample, (ii) eluting and concentrating the material, and (iii) synthesizing complementary DNA (cDNA) from the template RNA.[17] However, it is important to note that RNA extraction can be a labor-intensive, expensive, and time-consuming process that requires manual handling and carries a risk of exposure to infectious material, which may also introduce experimental errors.[17,18]

The process begins with the isolation of viral RNA from collected specimens, typically performed using commercially available RNA extraction kits. Proper handling during this step is essential to ensure accuracy and to prevent contamination. Once isolated, the RNA undergoes reverse transcription to synthesize cDNA. This

Table 1. List of real-time polymerase chain reaction protocols indicated by the World Health Organization

Institute	Gene targets
China CDC, China	*ORF1ab* and *N*
Institute Pasteur, France	Two targets in *RdRP*
United States CDC, USA	Three targets in the *N* gene
National Institute of Infectious Diseases, Japan	Multiple targets in coronavirus panel
Charité, Germany	*RdRP*, *E*, *N*
Hong Kong University, Hong Kong SAR	*ORF1b-NSP14*, *N*
National Institute of Health, Thailand	*N*

Abbreviations: CDC: Centers for Disease Control and Prevention; SAR: Special administrative region.
Source: https://www.who.int/who-documents-detail/molecular-assays-to-diagnose-COVID-19-summary-table-of-available-protocols.

step uses a master mix containing reverse transcriptase enzyme and is carried out in a thermal cycler to initiate the reaction. The resulting cDNA is then prepared for qPCR by mixing it with primers, distilled water, and either a DNA probe or SYBR Green dye. This mixture is then pipetted into a 96-well plate, which is subsequently placed in an RT-qPCR instrument programmed with the appropriate thermal cycling settings. To ensure reliable results, a negative control is included to confirm the absence of contaminants, and a positive control is used to validate the assay accuracy.[19]

The qPCR amplification process involves three key steps: denaturation, annealing, and extension. During denaturation, the temperature is raised to denature double-stranded DNA (dsDNA) into single strands. Annealing follows, during which the temperature is lowered to allow primers to bind to their specific target sequences on the DNA. In the extension step, DNA polymerase attaches nucleotides to synthesize the cDNA strand, completing the replication process. These three steps are repeated approximately 40 times, doubling the DNA in each cycle. A fluorescent signal is generated when either DNA probes are cleaved or SYBR Green binds to newly formed dsDNA. The intensity of this fluorescence increases with the number of DNA copies, enabling real-time monitoring of the amplification process. The cycle threshold (Ct) value is determined by the number of cycles needed for fluorescence to surpass a predetermined threshold and provides quantitative insights into the amount of target nucleic acid present. Lower Ct values indicate higher viral loads, while higher Ct values indicate lower viral loads.[19,20]

Real-time qPCR was the first standard molecular technique for SARS-CoV-2 detection since the emergence of COVID-19. This molecular diagnostic technique identifies nucleic acids by targeting and amplifying specific genes, producing millions of copies from a small initial sample. This amplification is monitored in real time, enabling precise and timely analysis.[19,21,22] Primers play a critical role in this process by binding specifically to the target gene. To ensure accuracy, primers must be well-designed with high specificity to avoid non-specific amplification or the formation of unwanted structures, such as primer dimers, which can lead to false-positive results.

Despite its effectiveness, RT-qPCR often exhibits variability in performance, largely influenced by the primers and probes used.[23,24] False results in viral polymerase chain reaction (PCR) testing often stem from a mix of biological, technical, and procedural issues. False negatives are especially common when sample quality is poor, viral load is low, or RNA extraction is inefficient. Problems during collection, handling, storage, or transport can further degrade the sample. Contaminants in clinical specimens may inhibit PCR, while cross-contamination during processing can lead to false positives. Mutations in primer or probe binding sites, as well as faulty equipment or reagents, also contribute to inaccurate results.[25,26] On the other hand, false positives may occur due to contamination, cross-reactivity with non-target organisms, or detection of residual viral material post-infection.[27] RNA extraction is a major bottleneck in the workflow due to its labor-intensive and specialized nature, particularly during high-demand periods such as the COVID-19 pandemic. The global shortage of RNA extraction kits has further strained testing capacities. However, numerous countries have adopted in-house RT-qPCR protocols as their standard diagnosis tool. This highlights the need for new diagnostic platforms that are fast, sensitive, and accessible to address this problem.[28]

To address the challenges in conventional RT-qPCR, it is crucial to develop novel testing methods that do not require advanced resources or extended processing time. Ideally, a test that delivers results within 2–3 h would support timely preventive measures, as patients could immediately receive the necessary support and care.[29] Globally, various direct approaches that avoid RNA extraction have been suggested, including heat-processed methods.[30,31] Our method is novel in that it does not use chelating agents, proteinase K, or other additional chemicals. We use heat and dilution to reduce the cost of the RNA extraction step. The cost reduction is particularly important in low-income countries such as Ethiopia when testing large numbers of samples. This consideration provided the main rationale for the present study. The objective of this study is to compare the diagnostic performance of our extraction-free dilution and heating (EFDH) method with that of extraction-based methods for SARS-CoV-2 detection.

2. Materials and methods

2.1. Study population, sample size calculation, and selection of stored specimens

Individuals tested for SARS-CoV-2 during the fifth wave of the pandemic (June–August 2020) were considered the study population. The study involved random selection and retrieval of stored nasopharyngeal specimens. The sample size was calculated using the double-population proportion formula, yielding 300 samples. Accordingly, a total of 300 samples (190 positive and 110 negative) were retrospectively obtained from the National Virology Reference Laboratory at the Ethiopian Public Health Institute in Addis Ababa, Ethiopia, which serves as the first national COVID-19 testing laboratory. They were

initially collected for clinical diagnosis of COVID-19 in viral transport media. The samples were retrieved using the unique numbers assigned to each specimen in the laboratory register. These specimens were retested using the EFDH technique, and the results were compared with those from the extraction-based detection method. The clinical samples were handled and processed in a Biosafety Level 2 laboratory at the virology laboratory.

2.2. Method validation

We intentionally chose six specimens with an average Ct value of 19 and an internal control (with Victoria [VIC] dye) having an average Ct value of 19. These specimens were diluted 1:2 and 1:4 with RNase-free water. The diluted specimens were then heated at 62°C and 72°C for 10, 15, 20, and 30 min, and at 96°C for 5, 10, 15, and 20 min using an Eppendorf thermomixer. Considering these heated and diluted specimens as an RNA eluate, we added 10 µL of the eluate to 20 µL of master mix and performed RT-qPCR on a thermal cycler (ABI 7500, Thermo-Fisher Inc., United States of America) using DAAN detection kits (Cat. no. DA0932, Daan Gene Co., Ltd, China). For the SARS-CoV-2 DAAN detection kits, the primary targets are the *N* and *ORF1ab* genes. The kit targets the *N* and *ORF1ab* genes, with the *N* gene detected on the FAM channel, the *ORF1ab* gene on the VIC channel, and the internal control on the Cy5 channel. These genes are crucial for detecting and identifying SARS-CoV-2 using PCR-based assays. The results show that after 15 min and 30 min incubation at 72°C, the average *N* and *ORF1ab* Ct values were 24 and 26, respectively; and the VIC average values were 26 and 27, respectively. The validation experiment showed that 1:2 diluted specimens exposed to 72°C for 15 min yielded results comparable to the standard RNA extraction-based method for SARS-CoV-2 detection. Therefore, we evaluated this EFDH method on the 300 samples using a 1:2 dilution and exposure at 72°C for 15 min.

2.3. Extraction-based protocol (reference method)

Extraction was performed manually using the QIAamp Viral RNA Kit (Cat. no. 52904, QIAGEN, Germany) following the standard protocol.[32] The process began with sample lysis under highly denaturing conditions, effectively inactivating RNases and protecting the integrity of viral RNA. These conditions ensured that the RNA remained intact throughout the extraction process. Next, the buffering conditions were carefully adjusted to facilitate the binding of RNA to the specialized membrane of the QIAamp plates. This step was crucial for capturing the RNA while preventing the co-binding of contaminants. To ensure the removal of impurities, the bound RNA underwent a series of washes using two wash buffers followed by a final ethanol wash. These steps effectively eliminated contaminants, including proteins, nucleases, and other potential inhibitors, yielding high-purity RNA. The purified RNA was then eluted using a specially formulated RNase-free buffer, making it immediately available for downstream applications or safe for long-term storage without degradation. The unique QIAamp membrane played a pivotal role in the process, ensuring exceptionally high recovery rates of pure and intact RNA. The methodology avoided the use of phenol/chloroform extraction or alcohol precipitation, making it a safe and efficient alternative for obtaining high-quality RNA suitable for sensitive molecular diagnostic procedures.

2.4. PCR reagent preparation (master mix preparation)

The PCR solutions A and B were thawed at room temperature and thoroughly mixed. The solution was then centrifuged at 8,000 rpm for a few seconds. The number of specimens to be tested, along with the number of *ORF1ab* and *N* negative controls and positive controls, was labeled on the PCR tubes. Next, 20 µL of the solution was aliquoted into each PCR tube, then 5 µL of the negative control material, RNA from the specimens to be tested, or positive control material was added to the respective PCR tubes. After securely covering the tubes, they were transferred to the amplification detection area following a brief centrifugation at 8,000 rpm. Finally, the reaction tubes were placed in the instrument's sample sink.

2.5. EFDH protocols (index method)

Each sample (100 µL) was transferred to a heat-resistant 1.5 mL microcentrifuge tube and heated at 72°C for 15 min using a thermomixer (Eppendorf, Germany). The eluates were then stored at −20°C for up to 4 h while waiting for other clinical samples to be processed in batches. Afterward, 10 µL of the eluate was mixed with 20 µL of the master mix, following a procedure similar to that for the extraction-based eluted samples. RT-qPCR was then performed on a thermocycler (ABI7500, Thermo-Fisher Inc., United States of America) following the protocol as shown in Figure 1.

2.6. Detection of SARS-CoV-2 (RT-qPCR)

The detection of SARS-CoV-2 using RT-qPCR was performed using a thermocycler (ABI7500, Thermo-Fisher Inc., United States of America), along with the DAAN kits (Cat. no. DA0932, Daan Gene Co., Ltd, China). The protocol consisted of a single cycle of reverse transcription at 50°C for 15 min, followed by a single cycle of polymerase activation at 95°C for 2 min. This was followed by 40 amplification cycles, each consisting of

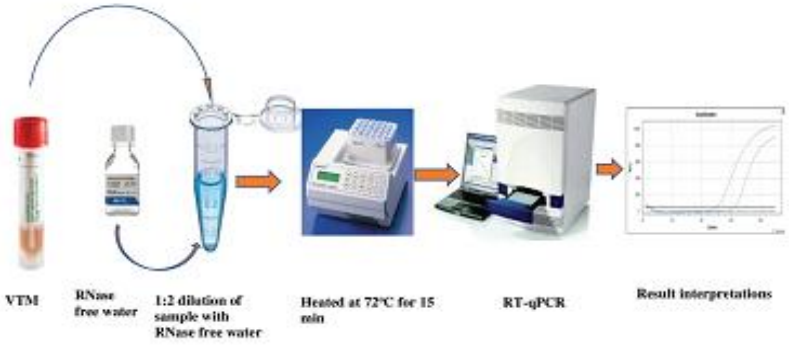


Figure 1. Process of sample dilution and heating for the detection and interpretation of severe acute respiratory syndrome coronavirus 2
Abbreviations: RT-qPCR: Real-time quantitative polymerase chain reaction; VTM: Viral transport media.

15 s at 95°C and 60 s at 60°C. The purpose of this assay was to detect the presence of SARS-CoV-2 using primers and probes targeting the *N* and *ORF1ab* gene. For comparison, the standard EB technique served as the reference method, alongside the EFDH method. The RT-qPCR results were then interpreted. A Ct value greater than 40, or the absence of an amplification curve in both FAM and VIC channels, indicated the absence of SARS-CoV-2 RNA in the sample. Conversely, a clear amplification curve in the VIC channel with a Ct value of 40 or less indicated a positive result. Samples with borderline amplification or ambiguous VIC signals were retested, and the result of the repeat run was considered final. The repeated result was considered the final result. At each step of both methods, strict quality control measures were taken. In addition to the positive and negative controls provided in the kit for the extraction method, we also used an extraction control. All measuring devices in the laboratory, including the thermal cycler, were within their calibration dates.

2.7. Ethics statement

This study was reviewed and approved by the Institutional Review Board of Aklilu Lemma Institute of Pathobiology (ALIPB), Addis Ababa University in Ethiopia, under protocol number ALIPB IRB/52/2013/21. A permission letter was obtained from the Ethiopian Public Health Institute with reference number WG12/19 on April 21, 2021. The studies were conducted in accordance with the local legislation and institutional requirements.

2.8. Statistical analysis

The data were summarized as mean ± standard error of the mean (SEM) and coefficient of determination (R^2). The D'Agostino and Pearson test, using a 95% confidence interval (CI), was used to assess normality. CIs for sensitivity, specificity, positive and negative predictive values, and accuracy were calculated using R software (v.3.6.0, R Foundation for Statistical Computing, Austria). The Wilcoxon signed-rank test was used to compare the two groups. The correlation between the Ct values of the two techniques was determined with the two-tailed Pearson's correlation coefficient using STATA (v17, StataCorp LP, United States of America). A $p<0.05$ indicated statistical significance. The Ct values were grouped accordingly (Ct < 20, 20–35, and >35), and performance characteristics in each group were compared.

3. Results

The EFDH method detected 85.8% (163/190) of the positive samples and 100% (110/110) of the negative samples. However, 15 false-negative results were recorded, most of which had average Ct values >36 by the extraction-based detection method. In addition, the EFDH method yielded 12 invalid results where the internal control did not amplify. To compare the two methods, we excluded the invalid results, using 178 positive and 110 negative results detected by the extraction-based detection method, as shown in Table 2.

The extraction-based method yielded a mean Ct value of 24.1 (95% CI: 23.3–24.8) with an SEM of 0.4. This method detected a range of Ct values from 16.0 to 34.1. On the other hand, the EFDH method yielded an average Ct value of 27.7 (95% CI: 26.8–28.7) with an SEM of 0.5. This method also detected a range of Ct values from 18.0 to 38.9. The true prevalence among study participants was 62% (178/288) (95% CI: 56–67%). The apparent prevalence, defined as the proportion of true positives out of the total tested, was 57% (163/288) (95% CI: 51–62%).

3.1. Diagnostic performance of the EFDH method

The EFDH method showed an overall sensitivity and specificity of 92% (95% CI: 86–95%) and 100% (95% CI: 97–100%), respectively. The positive predictive value was 100% (95% CI: 97–100%), while the negative predictive value was 88% (95% CI: 81–93%). The negative likelihood ratio, which measures the ability of a test to rule out a disease, was found to be 0.08 (95% CI: 0.05–0.14). The study also revealed an overall agreement (accuracy) of 95% (95% CI: 92–97%) for the EFDH method, with a kappa value of 0.89 (p<0.001) (Table 3).

The correlation between the two methods is represented by Equation (1):

$$Y = -0.12 + 1.16\,X \quad (1)$$

With R^2 = 0.99, p=0.001, showing a strong and significant linear relationship between the extraction-based and EFDH detection methods. For every unit increase in the Ct value by the extraction-based detection methods, the Ct value by the EFDH detection method increases by approximately 1.16 units (Figure 2). The Wilcoxon signed-rank test (z = 11.08, p=0.0001) indicated a statistically significant difference in Ct values between the extraction-based and EFDH detection methods. The extraction-based method generally produces lower Ct values compared to the EFDH method across all observations.

Table 2. Comparison of the extraction-based method and the extraction-free dilution and heating (EFDH) methods

EFDH	Extraction-based method		
	Positive	Negative	Total
Positive	163	0	163
Negative	15	110	125
Total	178	110	288

Note: Pearson's χ^2<0.001.

Table 3. Performance characteristics of the extraction-based and extraction-free dilution and heating methods for severe acute respiratory syndrome coronavirus in Ethiopia

Statistic	Value (%)	95% CI
Apparent prevalence	57	51–62%
True prevalence	62	56–67%
Sensitivity	92	86–95%
Specificity	100	85–100%
Negative likelihood ratio	0.08	0.05–0.14
Positive predictive value	100	98–100%
Negative predictive value	88	81–93%
Accuracy/agreement	95	92–97%

Abbreviation: CI: Confidence interval.

3.2. Comparison of performance characteristics in terms of cycle threshold value

Regarding the samples with a Ct value <20 (38 samples), both techniques correctly identified all samples as true positives. Of the 124 results with a Ct value ranging from 20 to 35, 121 (98%) were true positives. For the remaining 119 test results with Ct values higher than 35, the EFDH method yielded nine false negatives. These findings demonstrated that both techniques were highly sensitive to the Ct < 20 group, accurately identifying all samples as positive. In the 20–35 Ct value group, both techniques detected most of the positive samples, but the EFDH method showed 2% false-negative results (3 samples). Finally, the EFDH technique showed low sensitivity in the group with Ct > 35. These results suggested that the EFDH method is more reliable for detecting SARS-CoV-2 in samples with higher viral loads (Table 4).

4. Discussion

We compared the EFDH method to the standard extraction-based SARS-CoV-2 detection method. The EFDH method showed high sensitivity, specificity, and positive and negative predictive values, especially at high SARS-CoV-2 viral loads. In addition, the EFDH method showed a remarkable agreement and correlation with the extraction-based method. The mean Ct values for the EFDH method were slightly higher compared to the extraction-based method; however, they were generally comparable to the gold standard method. The findings

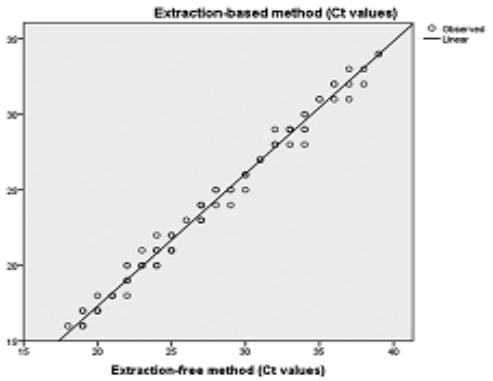


Figure 2. Correlation between extraction-free heating and dilution method and extraction-based severe acute respiratory syndrome coronavirus 2 detection methods in Ethiopia

Table 4. Performance of extraction-free and extraction-based methods across different ranges of cycle threshold value

Extraction-free method	Extraction-based method		
	Positive	Negative	Total
Ct<20			
Positive	38	0	38
Negative	0	0	0
Total	38	0	38
Ct=20–35			
Positive	121	0	121
Negative	3	0	3
Total	124	0	124
Ct>35			
Positive	0	0	0
Negative	9	110	119
Total	9	110	119

Abbreviation: Ct: Cycle threshold.

suggest that the EFDH method is particularly reliable for detecting SARS-CoV-2 in samples with higher viral loads. Despite limitations in samples with low viral loads or high Ct values, the EFDH method demonstrated overall high accuracy and closely aligned with the extraction-based method, indicating a strong performance and reliability.

In this study, the performance of the EFDH method was comparable to that of the extraction-based method. The EFDH method demonstrated good sensitivity and specificity. This finding aligns with a study conducted in India, which reported an overall sensitivity of 79% (95% CI: 71–86%) and specificity of 99% (95% CI: 98–100%).[35] Another study on extraction-free, multiplexed amplification of SARS-CoV-2 demonstrated a sensitivity of 86% and specificity of 100%.[34] Similarly, a study conducted at the Karolinska University Hospital in Stockholm, Sweden, in 2020, using a heating technique at 95°C for 5 min, reported a sensitivity of 96.0% and specificity of 99.8%.[17] In contrast, a study conducted in India and Italy reported lower overall sensitivity, specifically 78.9% (95% CI: 71–86%) and 57.3% (95% CI: 47.3–66.8%), respectively. However, both studies reported comparable specificity, at 99.9% (95% CI: 98–99.6%) and 100% (95% CI: 94.4–100%), respectively.[35]

This study reports higher positive and negative predictive values compared to previous studies. These findings indicate excellent specificity in identifying true positives, thus enhancing the reliability of the method for confirming the presence of the diseases. However, they also highlight the need to reduce false-negative rates, which would further enhance the method's utility for ruling out infection. False viral PCR results are caused by a combination of procedural and biological factors, often due to poor sample quality, low viral load, or inefficient RNA extraction, and may also arise from issues during specimen storage. In addition, contaminants in the sample can inhibit PCR amplification and compromise detection.[29,30] Similar findings were reported in a 2021 study conducted in the United Arab Emirates (UAE), showing positive and negative predictive values of 92% and 91%, respectively.[36] The finding aligns with another 2021 study in India, with positive and negative predictive values of 92% and 97%, respectively.[33]

In this study, the overall agreement between the extraction-based and EFDH methods was found to be 95%, with a kappa value of 0.89 (p=0.000). This indicates a high level of reliability in detecting SARS-CoV-2 infections. Similar findings were reported in a study conducted in India, where the agreement was 96.8% (k = 0.83, SEM = 0.03).[36] In addition, a study conducted in Sweden showed an accuracy of 98.8% (95% CI: 97.5–99.5%),[17] while a study in the UAE reported an overall agreement (kappa coefficient) of 0.797 (p<0.001).[36] However, different findings were reported from a study conducted at the clinical laboratory of the Institut Pasteur of M'sila, Algeria, where the overall agreement between extracted and heat-inactivated (65°C for 30 min) samples was only 45%, with a 95% CI of 37–52%.[37] In the current study, a near-perfect correlation (R^2 = 0.99, p=0.001) was found between the extraction-based and EFDH methods, supporting the consistency and reliability of the EFDH method in detecting SARS-CoV-2. These findings are further supported by a 2020 study conducted in Singapore, which showed an R^2 of 0.9986 for the detection of SARS-CoV-2 without extraction. Similar results were reported from the Karolinska University Hospital in Stockholm, Sweden, with an R^2 of 0.987.[17]

The consistent findings across studies utilizing the extraction-free method for SARS-CoV-2 detection can be attributed to the method's simplicity, the inherent properties of SARS-CoV-2, and its validation against established standards. These factors contribute to the method's robustness and reliability in detecting SARS-CoV-2, making it a valuable tool in overcoming the COVID-19 pandemic.[38] Despite the challenges, the technique's overall high performance supports its continued use in surveillance and outbreak control efforts. However, the potential for false negatives emphasizes the importance of comprehensive testing strategies, including targeted testing of high-risk populations and asymptomatic individuals.[39] This method's advantages in terms of speed and simplicity, combined with its high sensitivity and specificity, make it a compelling alternative for SARS-CoV-2 detection, especially in resource-limited settings or during rapid response scenarios.[40]

These findings highlight the strengths and limitations of the extraction-free method in detecting SARS-CoV-2 across different viral loads. Notably, the method demonstrated high sensitivity and specificity in samples with lower Ct values (<35). However, its performance decreases in detecting samples with low viral loads (higher Ct values), as evidenced by an increased rate of false negatives. A study conducted in Austria to evaluate extraction-free RT-qPCR methods for SARS-CoV-2 diagnostics showed an 8.2% false-negative rate for the EFDH method in samples with a Ct value >30.[36] Another study from Algeria on the detection of SARS-CoV-2 using heat inactivation at 65°C for 30 min correctly identified 100% of clinical samples with a high viral load (Ct value <30).[37] Accurate detection of SARS-CoV-2, even at high viral loads (lower Ct value), is crucial for early case identification, contact tracing, and isolation measures.

5. Limitations

The study utilized previously stored samples instead of freshly collected ones and compared the retest results with the initial findings. It did not assess the performance of the EFDH method across different storage conditions and time intervals. Furthermore, it did not determine the method's limit of detection.

6. Conclusion

The findings suggest that the EFDH method may serve as a reliable option for detecting SARS-CoV-2 in samples with low Ct values, corresponding to higher viral loads. The EFDH method offers a potential advantage in settings where resources are limited or rapid turnaround times are required. Despite the promising results, the study highlights areas for improvement. The number of invalid and false-negative results obtained with the EFDH method necessitates further investigation into the causes of these failures and the development of new strategies to mitigate them.

Acknowledgments

The authors are grateful to the Ethiopian Public Health Institute and the Addis Ababa University for their support and coordination, as well as for providing the testing laboratories. The authors would like to express their special thanks to all who participated in data collection and the laboratory activities.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Getnet Hailu, Molalegne Bitew, Mengistu Legesse

Data curation: Getnet Hailu, Molalegne Bitew, Mengistu Legesse

Formal analysis: Getnet Hailu, Molalegne Bitew, Girmay Medhin, Samuel Ayele, Mengistu Legesse

Investigation: Getnet Hailu, Adamu Tayachew, Nebiyou Yemanebirhan, Andargachew Gashu, Admikew Agune, Assefa Konda, Mengistu Legesse

Methodology: Getnet Hailu, Girmay Medhin, Adamu Tayachew, Nebiyou Yemanebirhan, Andargachew Gashu, Admikew Agune, Assefa Konda, Mengistu Legesse

Writing–original draft: Getnet Hailu, Mengistu Legesse

Writing–review & editing: All authors

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. Ethical review and approval were obtained from the Institutional Review Board of the Aklilu Lemma Institute of Pathobiology at Addis Ababa University, under protocol number ALIPB IRB/52/2013/21. A permission letter was obtained from the Ethiopian Public Health Institute to use leftover samples from repositories.

Consent for publication

Not applicable.

Availability of data

Data used in this work is available from the corresponding author upon reasonable request.

References

1. Wu A, Peng Y, Huang B, *et al.* Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe.* 2020;27(3):325-328. doi: 10.1016/j.chom.2020.02.001
2. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. *Acta Pharmacol Sin.* 2020;41(9):1141-1149. doi: 10.1038/s41401-020-0485-4
3. Harvey WT, Carabelli AM, Jackson B, *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol.* 2021;19(7):409-424. doi: 10.1038/s41579-021-00573-0
4. Corman VM, Landt O, Kaiser M, *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro*

Surveill. 2020;25(3):2000045.

doi: 10.2807/1560-7917.es.2020.25.3.2000045

5. Liu AB, Lee D, Jalihal AP, Hanage WP, Springer M. Quantitatively assessing early detection strategies for mitigating COVID-19 and future pandemics. *Nat Commun.* 2023;14(1):8479.

doi: 10.1038/s41467-023-44199-7

6. Zhang C, Zhou C, Xu W, *et al.* Transmission risk of asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Infect Med (Beijing).* 2022;2(1):11-18.

doi: 10.1016/j.imj.2022.12.001

7. Joynt GM, Wu WK. Understanding COVID-19: What does viral RNA load really mean? *Lancet Infect Dis.* 2020;20(6):635-636.

doi: 10.1016/S1473-3099(20)30237-1

8. Sethuraman N, Jeremiah SS, Ryo A. Interpreting diagnostic tests for SARS-CoV-2. *JAMA.* 2020;323(22):2249-2251.

doi: 10.1001/jama.2020.8259

9. Loeffelholz MJ, Tang YW. Laboratory diagnosis of emerging human coronavirus infections - the state of the art. *Emerg Microbes Infect.* 2020;9(1):747-756.

doi: 10.1080/22221751.2020.1745095

10. Cheng VCC, Lau SKP, Woo PCY, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin Microbiol Rev.* 2007;20(4):660-694.

doi: 10.1128/CMR.00023-07

11. Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M. Routine blood tests as a potential diagnostic tool for COVID-19. *Clin Chem Lab Med.* 2020;58(7):1095-1099.

doi: 10.1515/cclm-2020-0398

12. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, *et al.* Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis.* 2020;34:101623.

doi: 10.1016/j.tmaid.2020.101623

13. Salamanna F, Maglio M, Landini MP, Fini M. Body localization of ACE-2: On the trail of the keyhole of SARS-CoV-2. *Front Med.* 2020;7:594495.

doi: 10.3389/fmed.2020.594495

14. Wang MY, Zhao R, Gao LJ, Gao XF, Wang DP, Cao JM. SARS-CoV-2: Structure, biology, and structure-based therapeutics development. *Front Cell Infect Microbiol.* 2020;10:587269.

doi: 10.3389/fcimb.2020.587269

15. Marty FM, Chen K, Verrill KA. How to obtain a nasopharyngeal swab specimen. *N Engl J Med.* 2020;382(22):e76.

doi: 10.1056/NEJMvcm2010260

16. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am J Infect Control.* 2021;49(1):21-29.

doi: 10.1016/j.ajic.2020.07.011

17. Smyrlaki I, Ekman M, Lentini A, *et al.* Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR. *Nat Commun.* 2020;11(1):4812.

doi: 10.1038/s41467-020-18611-5

18. Bruce EA, Huang ML, Perchetti GA, *et al.* Direct RT-qPCR detection of SARS-CoV-2 RNA from patient nasopharyngeal swabs without an RNA extraction step. *PLoS Biol.* 2020;18(10):e3000896.

doi: 10.1371/journal.pbio.3000896

19. Alhamid G, Tombuloglu H, Rabaan AA, Al-Suhaimi E. SARS-CoV-2 detection methods: A comprehensive review. *Saudi J Biol Sci.* 2022;29(11):103465.

doi: 10.1016/j.sjbs.2022.103465

20. Bustin SA, Nolan T. RT-qPCR testing of SARS-CoV-2: A primer. *Int J Mol Sci.* 2020;21(8):3004.

doi: 10.3390/ijms21083004

21. Aoki MN, Coelho BO, Góes LGB, *et al.* Colorimetric RT-LAMP SARS-CoV-2 diagnostic sensitivity relies on color interpretation and viral load. *Sci Rep.* 2021;11:9026.

doi: 10.1038/s41598-021-88506-y

22. Tombuloglu H, Sabit H, Al-Suhaimi E, Al Jindan R, Alkharsah KR. Development of multiplex real-time RT-PCR assay for the detection of SARS-CoV-2. *PLoS One.* 2021;16(4):e0250942.

doi: 10.1371/journal.pone.0250942

23. Lv DF, Ying QM, Weng YS, *et al.* Dynamic change process of target genes by RT-PCR testing of SARS-Cov-2 during the course of a coronavirus disease 2019 patient. *Clin Chim Acta.* 2020;506:172-175.

doi: 10.1016/j.cca.2020.03.032

24. Tahmasebi S, Khosh E, Esmaeilzadeh A. The outlook for diagnostic purposes of the 2019-novel coronavirus disease. *J Cell Physiol.* 2020;235(12):9211-9219.

doi: 10.1002/jcp.29804

25. Mouliou DS. Managing viral emerging infectious diseases via current molecular diagnostics in the emergency department: The Tricky cases. *Expert Rev Anti Infect Ther.* 2022;20(9):1163-1169.

doi: 10.1080/14787210.2022.2089653

26. Mouliou DS, Gourgoulianis KI. COVID-19 'asymptomatic' patients: An old wives' tale. *Expert Rev Respir Med.* 2022;16(4):399-407.

doi: 10.1080/17476348.2022.2030224

27. Mouliou DS, Gourgoulianis KI. False-positive and false-negative COVID-19 cases: Respiratory prevention and management strategies, vaccination, and further perspectives. *Expert Rev Repir Med*. 2021;15(8):993-1002.

doi: 10.1080/17476348.2021.1917389

28. Pizzol JLD, Hora VP, Reis AJ, *et al*. Laboratory diagnosis for Covid-19: A mini-review. *Rev Soc Bras Med Trop*. 2020;53:e20200451.

doi: 10.1590/0037-8682-0451-2020

29. Zhao Y, Cui C, Zhang K, *et al*. COVID19: A systematic approach to early identification and healthcare worker protection. *Front Public Health*. 2020;8:205.

doi: 10.3389/fpubh.2020.00205

30. Fomsgaard AS, Rosenstierne MW. An alternative workflow for molecular detection of SARS-CoV-2 - escape from the NA extraction kit-shortage, Copenhagen, Denmark, March 2020. *Eur Surveill*. 2020;25(14):2000398.

doi: 10.2807/1560-7917.es.2020.25.14.2000398

31. Villota SD, Nipaz VE, Carrazco-Montalvo A, *et al*. Alternative RNA extraction-free techniques for the real-time RT-PCR detection of SARS-CoV-2 in nasopharyngeal swab and sputum samples. *J Virol Methods*. 2021;298:114302.

doi: 10.1016/j.jviromet.2021.114302

32. QIAGEN. *QIAamp® Viral RNA Mini Handbook*; 2020. Available from: https://www.qiagen.com [Last accessed on 2023 Aug 21].

33. Jayaprakasam M, Aggarwal S, Mane A, *et al*. RNA-extraction-free diagnostic method to detect SARS-CoV-2: An assessment from two states, India. *Epidemiol Infect*. 2021;149:e245.

doi: 10.1017/s0950268821002302

34. Byrnes SA, Gallagher R, Steadman A, *et al*. Multiplexed and extraction-free amplification for simplified SARS-CoV-2 RT-PCR tests. *Anal Chem*. 2021;93(9):4160-4165.

doi: 10.1021/acs.analchem.0c03918

35. Morecchiato F, Coppi M, Baccani I, *et al*. Evaluation of extraction-free RT-PCR methods for faster and cheaper detection of SARS-CoV-2 using two commercial systems. *Int J Infect Dis*. 2021;112:264-268.

doi: 10.1016/j.ijid.2021.09.046

36. Mahmoud SA, Ganesan S, Ibrahim E, *et al*. Evaluation of RNA extraction-free method for detection of SARS-CoV-2 in salivary samples for mass screening for COVID-19. *Biomed Res Int*. 2021;2021:5568350.

doi: 10.1155/2021/5568350

37. Mokhtar KM. *Improved COVID-19 Testing by Extraction Free SARS-Cov-2 RT-PCR*; 2020. Available from: https://medrxiv.org/lookup/doi/10.1101/2020.08.10.20171512 [Last accessed on 2024 May 02].

38. Domnich A, De Pace V, Pennati BM, *et al*. Evaluation of extraction-free RT-qPCR methods for SARS-CoV-2 diagnostics. *Arch Virol*. 2021;166(10):2825-2828.

doi: 10.1007/s00705-021-05165-0

39. Bastola MM, Locatis C, Fontelo P. *Diagnostic Laboratory Tests for COVID-19 in US: Methodology and Performance*. Research Square [Preprint]; 2020.

doi: 10.21203/rs.3.rs-43374/v1

40. Hasan MR, Mirza F, Al-Hail H, *et al*. Detection of SARS-CoV-2 RNA by direct RT-qPCR on nasopharyngeal specimens without extraction of viral RNA. *PLoS One*. 2020;15(7):e0236564.

doi: 10.1371/journal.pone.0236564

Identification of Conserved SARS-CoV-2 Gene Regions for Monoclonal Antibody-Based Diagnostic Development through Whole Genome Sequencing and Bioinformatics

Getnet Hailu [1,2], Mengistu Legesse [1], Andargachew Mulu [3], Girmay Medhin[1], Mesfin Mengesha Tsegaye [3], Dawit Hailu Alemayehu [3], Abaysew Ayele [3], Atsbeha Gebreegziabxier [2], Adamu Tayachew [1,2], Adimkewu Aguine [2], Haileyesus Dejene[4], Sofoniyas K.Tessema [5], Harris Onywera [5], Assohoun Egomli Stanislas [6], Ebba Abate [1], Alessandro Marcello [7] and Molalegne Bitew [8]

[1]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia.
[2]Ethiopian Public Health Institute, Addis Ababa, Ethiopia.
[3]Armaur Hansson Research Institute, Addis Ababa, Ethiopia.
[4]University of Gondar, Gondar, Ethiopia
[5]Africa Centres for Disease Control and Prevention (Africa CDC), Addis Ababa, Ethiopia.
[6]Genomics and metagenomics platform, Pasteur Institute, Abidjan, Cote d'Ivoire.
[7]Laboratory of Molecular Virology, International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy.
[8]Bio and Emerging Technology Institute Addis Ababa, Ethiopia.
*Corresponding Authors: Molalegne Bitew (molalegne23@gmail.com)

Abstract

Background: The high mutation rate of the SARS-CoV-2 spike (S) protein undermines the reliability of S-based diagnostics. The nucleocapsid (N) protein, being more conserved and abundant, offers a stable alternative, especially in resource-limited settings.

Objective: To identify conserved, antigenic regions in the SARS-CoV-2 gene using whole genome sequencing and immunoinformatic for monoclonal antibody-based rapid diagnostic development.

Methods: A retrospective study analyzed 150 SARS-CoV-2 positive samples collected in Ethiopia (June–August 2022). Seventy samples (Ct ≤30) were sequenced using Illumina NextSeq 550. Conserved regions were identified via MAFFT alignment, entropy plots, and mutation profiling. B-cell epitopes were predicted and assessed for antigenicity and cross-reactivity. Structural modeling was conducted using AlphaFold and visualization tools.

Results: High-quality whole genome sequencing of 63 SARS-CoV-2 samples revealed conserved regions in the nucleocapsid (N) gene across major variants. Three surface-exposed, antigenic, and non-allergenic B-cell epitopes were identified with low cross-reactivity to common human coronaviruses. Structural modeling and docking confirmed their accessibility and strong antibody binding, supporting their potential for robust diagnostic applications.

Discussion: Compared to the variable S gene, the conserved and immunogenic N gene epitopes are stable targets for monoclonal antibody development, improving diagnostic reliability across SARS-CoV-2 variants.

Conclusions: Conserved N gene epitopes identified in this study are promising targets for developing robust, variant-resistant rapid diagnostic tools.

Key words: SASR-CoV-2, conserved gene region, diagnostic tool, monoclonal antibody-based, Ethiopia

1. Introduction

The global emergence and persistence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has highlighted the critical need for reliable, rapid, and widely accessible diagnostic tools. While reverse transcription polymerase chain reaction (RT-PCR) is the current gold standard for SARS-CoV-2 detection, its requirements for specialized equipment and skilled personnel limit its deployment in low-resource settings and impede timely diagnosis during outbreaks(1,2). Consequently, antigen-based rapid diagnostic tests (Ag-RDTs) have gained prominence as complementary tools due to their ease of use, cost-effectiveness, and potential for point-of-care deployment(3). Ag-RDTs primarily target the viral spike (S) and nucleocapsid (N) proteins. The S protein, which mediates viral entry via the ACE2 receptor, is highly immunogenic but shows considerable genetic variability, particularly in regions that undergo immune-driven selection(4,5). This variability, observed notably in variants of concern (VOCs) such as Alpha, Beta, Delta, and Omicron, can reduce the sensitivity and specificity of S protein-based diagnostics and vaccines(6,7).

In contrast, the nucleocapsid (N) protein is highly conserved, abundantly expressed, and elicits a strong early antibody response, making it a promising target for diagnostics (1,2). Its structural and functional roles in viral RNA packaging and replication make it less prone to frequent mutations, providing greater stability across circulating variants(8). Due to this high degree of conservation, N protein-based Ag-RDTs may maintain diagnostic performance even as new variants emerge.

Recent advances in immunoinformatics have enabled efficient in silico screening of viral proteins to identify conserved, antigenic, and structurally accessible epitopes that are suitable for monoclonal antibody (mAb) development(9). Such epitope-based approaches allow for the rational design of mAb-targeting strategies that can improve the sensitivity and specificity of diagnostic tools. Computational tools such as BepiPred, IEDB, and structural modeling software (e.g., UCSF Chimera, AlphaFold, PyMOL) have facilitated the prediction of linear B-cell epitopes, assessment of antigenicity scores, and analysis of epitope surface accessibility, streamlining the development of diagnostics that are both robust and broad in variant coverage (10).

The rationale for this study stems from the urgent need to improve the reliability of SARS-CoV-2 diagnostic tools amid ongoing viral evolution. While most rapid diagnostic tests target the spike (S) protein, its high mutation rate across variants has compromised test sensitivity. In contrast, the nucleocapsid (N) protein is highly conserved and abundantly expressed, making it a more stable diagnostic target. This study aimed to identify conserved and antigenic regions within the N gene using whole genome sequencing, mutation profiling, and immunoinformatics, to inform the development of monoclonal antibody-based diagnostics with enhanced sensitivity and resilience to emerging variants. Therefore, the aim of this study was to identify the conserved and antigenic SARS-CoV-2 gene regions for the development of monoclonal antibody-based rapid diagnostic tools for SARS-CoV-2 detection.

2. Materials and Methods

2.1. Study design, period, and sample collection

A retrospective experimental study design was used. The study design was a retrospective experimental design. Sample collection was undertaken during the fifth wave of the pandemic (June to August, 2022) in Ethiopia. A total of 150 SARS-CoV-2 positive samples were collected retrospectively from the national virology reference laboratory repository at the Ethiopian Public Health Institute (EPHI). These samples were nasopharyngeal (NP) swab samples collected from patients residing in Addis Ababa, Oromia, Amhara, and Southern Nation Nationalities and Peoples (SNNP) regions (Figure 1). EPHI is a national reference laboratory for SARS-CoV-2 in Ethiopia and samples were brought from those geographic areas, as well as from diplomatic communities and international travelers for viral testing. Samples were stored at -80°C for long-term storage. Samples for sequencing were purposively selected based on their identifiers (IDs) from the electronic national database. Only the IDs that tested positive and had a cycle threshold (Ct) value less than or equal to 30 were selected. As a result, 150 SARS-CoV-2 positive samples were included in this study due to our limited sequencing resources. The patients' metadata, including demographic information, clinical data, and RT-qPCR results, were anonymously extracted from electronic databases. The study participants were then grouped based on factors such as age, sex, signs and symptoms, pre-existing chronic diseases, travel history, and vaccination status.

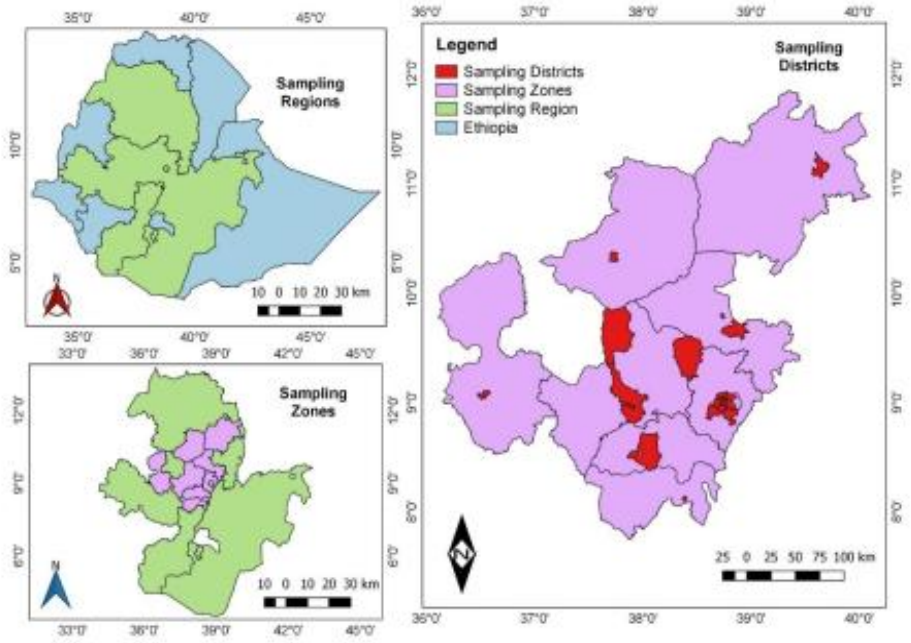


Figure 1. Geographic map of Ethiopia showed the regions and districts where SARS-CoV-2 specimens were collected.

2.2. Quality Assurance

To increase the success rate of whole genome sequencing, an RT-qPCR test was performed again after the samples thawed at room temperature (11). The RNA from these samples was re-extracted using the BGI extraction kit, and detection was performed using BIO-RAD, CFX96 Deep Well™ Real-Time PCR Detection system (Bio Rad Laboratories, Inc., United States) following the standard protocol. The clinical samples were handled and processed in a Biosafety Laboratory-2 (BSL-2) at the national virology reference laboratory. Out of the 150 re-analyzed samples, 70 had a Ct value less than or equal to 30 and were selected for whole genome sequencing and others were discarded due to their high Ct values (i.e., >30). The RNA extracts from these 70 samples were sequenced using the Illumina NextSeq 550 platform at the virology laboratory of EPHI during March 2023. Out of 70 SARS-CoV-2 sequence data, 63 sequences were passed the bioinformatics quality controls requirements (i.e., with genome coverage greater than 80%, no clustered mutations, and no misplaced stop codons) and selected for further sequence analysis.

2.3.Comparative Genome Analysis

Raw sequencing data were obtained in FASTQ format and processed for variant calling and consensus sequence generation. First, the FASTQ sequences' quality was evaluated by the FastQC (v0.11.9) program, and then adapter sequences were trimmed from reads using Cutadapt software v.4.6 (12) to remove adaptor sequences and poor quality reads. In addition, dataset of globally 118 representative SARS-CoV-2 sequences from 62 countries (including the major variants) was downloaded from GISAID. Multiple sequence alignment of local and global genomes was carried out using MAFFT. Mutation analysis was performed using Nextclade, CoV-GLUE, and SnpEff to identify variant-specific and conserved regions across genomes. All viral genome sequences obtained from this study were submitted to the Global Initiative on Sharing All Influenza Data (GISAID) with accession ID from EPI_ISL_ 19147618 to EPI_ISL_19302657 (13) and National Center for Biotechnology Information (NCBI) repositories with accession numbers from PQ140559 to PP069552 (14).

To identify conserved gene regions, multiple sequence alignment (MSA) was performed between the generated sequences and a diverse dataset of global SARS-CoV-2 sequences downloaded from the GISAID database, representing major variants of concern (VOCs) including Alpha, Beta, Delta, and Omicron. Alignment was conducted using MAFFT (v7), and sequence conservation was assessed using Jalview and entropy plots, where regions with low entropy and high identity were considered highly conserved. Additionally, mutations and variant-specific changes were annotated using the Nextclade.

Following the identification of conserved regions, B-cell linear epitope prediction was conducted using BepiPred 2.0 hosted by the IEDB Analysis Resource. Epitopes were screened for antigenicity using VaxiJen v2.0, with a threshold of 0.5, and those scoring above this were further analyzed for allergenicity and toxicity using AllerTOP and ToxinPred, respectively. Cross-reactivity with common human coronaviruses (e.g., HCoV-229E, HCoV-OC43, and HCoV-NL63) was assessed via BLASTp to ensure specificity.

For structural validation, three-dimensional models of the SARS-CoV-2 nucleocapsid (N) protein were obtained using AlphaFold predictions. The predicted epitopes were mapped onto the 3D structure using PyMOL and UCSF Chimera to evaluate surface accessibility and structural integrity. Solvent accessibility and flexibility of epitopes were examined using DynaMut.

Molecular docking simulations were then performed using ClusPro to assess potential binding interactions between monoclonal antibody fragments (Fabs) and the predicted epitopes.

Table 1. The WIV04 genome, along with 118 SARS-CoV-2 genome sequences from 62 countries, were

Classifications	Variants	Name	First reported country
VOC	alpha	GRY (B.1.1.7+Q.*)	United Kingdome
VOC	Beta	GH/501Y.V2 (B.1.351+ B.1.351.2+ B.1.351+3)	South Africa
VOC	Gamma	GR/501Y.V3 (P.1+P.1*)	Brazil and Japan
VOC	Omicron	GRA (B.1.1.529+BA.*)	Botswana, South Africa, and Hong Kong
Variant of interest (VOI)	Lambda	GR/452Q.V1 (C.37+C.37.1)	Peru
VOI	Mu	GH (B.1.621+B.1.621.1)	Colombia
VOI		GRA (XBB.1.5+XBB.1.5.*)	Australia, India, and Bangladesh
VOC		GRA (XBB.1.16+XBB.1.1.16.*)	India
Variant under monitoring (VUM)		GRA (XBB+XBB.*, excluding XBB.1.5, XBB.1.16, XBB.1.9.1, XBB.1.9.2, XBB.2.3)	India
VUM		GRA (XBB.1.9.1+XBB.19.1.*)	Indonesia, Singapore, and Israel

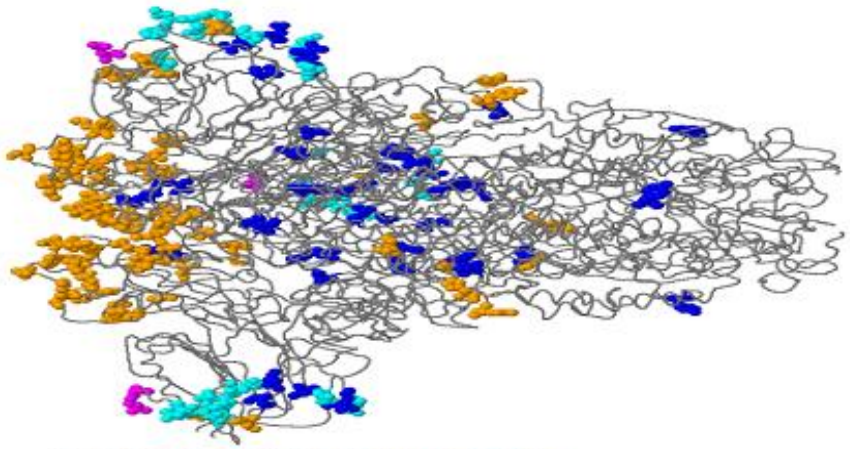
3. Results

3.1.Whole Genome Sequencing and Genome Quality

Whole genome sequencing was successfully performed on SARS-CoV-2 isolates from 63 clinical samples collected between June to August 2022. The sequencing yielded high-quality reads with an average depth of >99× and genome coverage of over 99%. Quality assessment of raw reads showed minimal adapter contamination and low base-calling error rates. The final assembled genome sequences showed a high degree of similarity to the reference SARS-CoV-2 genome (NC_045512.2) with a consensus sequence identity of >99.8%. The assembled genomes were deposited in GISAID under accession ID from EPI_ISL_ 19147618 to EPI_ISL_19302657 (13) and National Center for Biotechnology Information (NCBI) repositories with accession numbers from PQ140559 to PP069552 (14).

3.2.Identification of Conserved Gene Regions of SARS-CoV-2

Mutation profiling using Nextclade highlighted widespread mutations in the spike (S) gene (e.g., D614G, N501Y, P681R, and E484K), indicating huge genetic variability across VOCs. Due to this variability, the S gene was considered less suitable for stable diagnostic targeting (Figure 1).



Sources: https://www.epicov.org/epi3/frontend#1a1eca

Figure 1. Spike glycoprotein (PDB: 6acc, EM 3.6 Angstrom) with RBD in down conformation

In contrast, analysis of the N gene showed two major conserved regions: the RNA -binding domain (residues 46-174 and dimerization domain (residues 247-364). Multiple sequence alignment using MAFFT showed that these regions maintained over 98% sequence identity across Alpha, Beta, Delta and Omicron lineages. Entropy analysis using Jelview further confirmed these findings, with low entropy score ranging from 0.0 to 0.2 in N gene compared to higher entropy score in S (1.1-1.2), M (0.3-0.5) and E (0.2-0.3) genes. Mutation frequency analysis using R demonstrated that the N gene accumulated only 1 to 3 non-synonymous mutations across samples which reinforce its evolutionary stability and diagnostic potential.

3.3.B-cell Epitope Prediction and Antigenicity Screening

Subsequent B-cell epitope prediction using BepiPred 2.0 identified 17 potential linear epitopes from the conserved regions of N protein. Eight of those exhibited antigenicity scores above 0.5 using VaxiJen v2.0, indicating good potential as a diagnostic target. Among them, three epitopes, RRGPEQTQGNFG (residues 152-163), SSSRGTS (66-72) and DPNFKDQVILL (330-341), were prioritized for downstream analysis due to their high antigenicity scores of 0.81, 0.77 and 0.71 respectively, and their location within or near conserved functional domains.

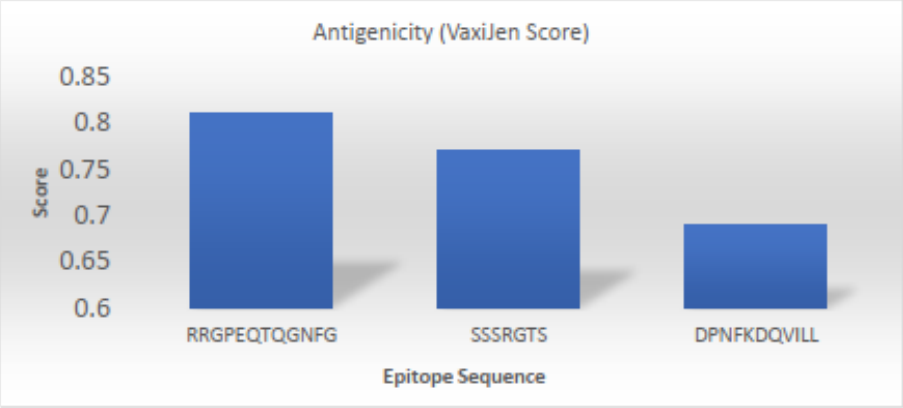


Figure 2. Antigenicity (VaxiJen Score) of the sequenced epitopes

3.4.Allergenicity, Toxicity and Cross reactivity Filtering

All Eight antigenic epitopes were evaluated for safety and five were predicted to be non-allergenic using AllerTOP v2.0 and non-toxic using ToxinPred. To assess the risk of cross reactivity with endemic human coronaviruses, BLASTp searches against HCoV-229E, HCoV-OC43, HCoVNL63, HCoV-HKU1 were conducted. Three epitopes (the above three candidates) showed low sequence identity i.e. <30% and E-values of >0.01. which indicate low risk of cross reactivity and confirming their diagnostic specificity.

3.5.Structural mapping and Solvent Accessibility

For structural validation, the 3D model of the SARS-CoV-2 N protein was obtained from AlphaFold. The three selected epitopes were mapped onto the model using PyMOL and UCSF Chimera. All were found to be surface exposed and two (RRGPEQTQGNFG and SSSRGTS) were located within flexibile coil regions. This location has a properties that enhance epitope accessibility and immunogenicity.

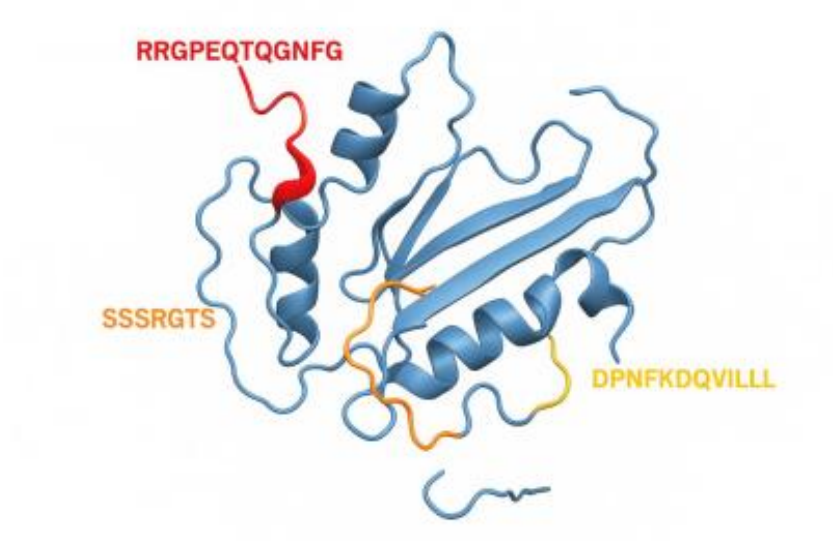


Figure 3. 3D Visualization of SARS-CoV-2 Nucleocapsid

Solvent accessibility and conformational flexibility analysis conducted by DynMut confirmed that the top candidate epitopes, RRGPEQTQGNFG (residues 152-163), had a solvent accessible surface area (SASA) of 78.4 Å² and a ΔΔG of -0.5 kcal/mol. This indicates that favorable stability and surface exposure.

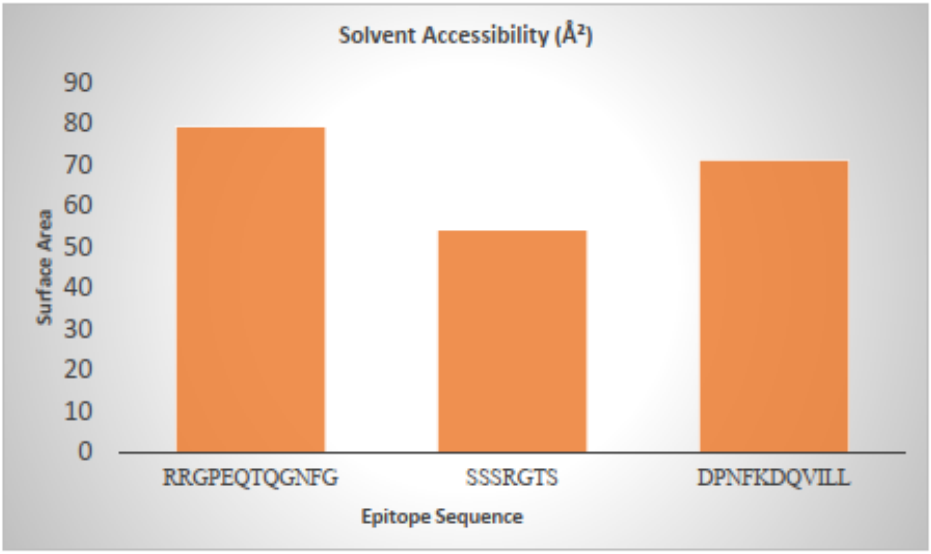


Figure 4. Solvent accessibility of sequenced epitopes

3.6.Molecular Docking Simulations

Finally, molecular docking simulations were conducted using ClusPro to evaluate the binding interaction between the predicted epitopes and monoclonal antibody fragments (Fabs). The docking results indicated strong affinities, with ClusPro weighted score of -917.4 for RRGPEQTQGNFG, -812.6 for SSSRGTS and -865.2 for DPNFKDQVILL. The RRGPEQTQGNFG epitope formed stable interactions via hydrogen bonds with key residues (R154, K157, D161), supporting its binding potential and utility in diagnostic design.

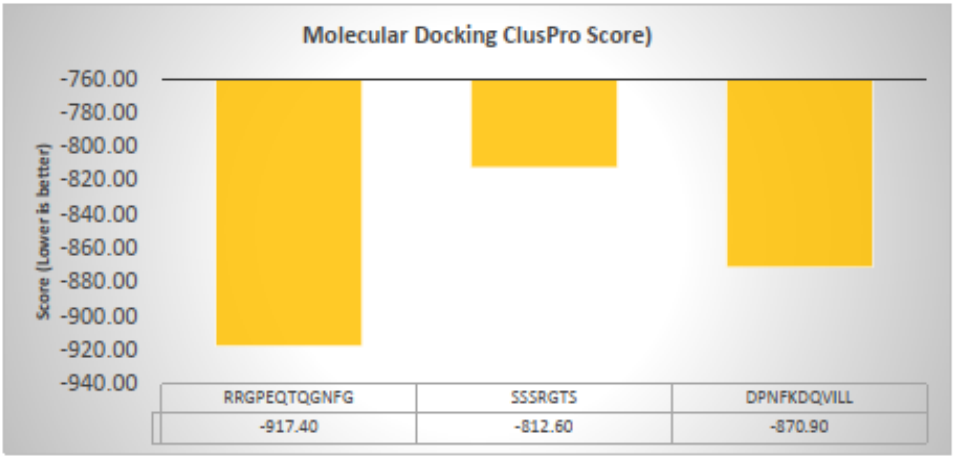


Figure 5. Molecular Docking ClusPro Score of each epitope

Table 2. Overall measurements/scores of epitope sequences

Epitope sequence	Position (N protein)	VaxiJen Score	allergenicity	Toxicity	Cross reactivity	Solvent Accessibility	cluePro Score
RRGPEQTQGNFG	152-163	0.81	Non-allergen	Non-toxic	None	High	-917.4
SSSRGTS	66-72	0.77	Non-allergen	Non-toxic	low	Moderate	-812.6
DPNFKDQVILL	194-204	0.69	Non-allergen	Non-toxic	None	High	-870.9

Note: These results demonstrate that computational epitope prediction, combined with antigenicity screening, structural validation, and docking studies, can effectively identify conserved, non-allergenic, and surface-accessible B-cell epitopes that may serve as candidates for affordable diagnostic or immunological tools targeting SARS-CoV-2.

4. Discussion

In this study, we explored conserved regions within the SARS-CoV-2 genome as potential targets for monoclonal antibody-based diagnostic tools. Through whole genome sequencing of locally circulating strains, the study identified that while the spike gene accumulated numerous mutations across variants, the N gene remained remarkably conserved. This observation supports a growing body of evidence indicating that the N protein, particularlyits RNA-binding and dimerization domains, is moregenetically stable than the S protein and thus a better target for laboratory diagnosis (15). Previous studies have consistently highlighted the instability of the S gene due to immune selection pressure, especially following mass vaccination and widespread natural infections (16). Mutations such as D614G, N501Y and E484K, commonly observed in our results, have been widely reported as key drivers of immune evasion and increased transmissibility (17). In contrast, the N gene's lower mutation frequency aligns with earlier genomic surveillance reports from various areas that identified it as a functionally constrained region with limited antigenic drift(18,19). These futures make N gene an ideal target for diagnostic development.

The conservational analysis, supported by sequence alignment and entropy profiling, reinforced the N gene's stability across multiple variants of concern, including Alpha, Beta, Delta and Omicron. This finding corroborates studies by (19) and (20), who demonstrated that the N protein retains structural integrity and sequences stability across a wide array of SARS-CoV-2 genomes, even in the face of global viral evolution.

Epitope prediction within conserved regions of N protein further validated their diagnostic potential. Predicted B-cell epitopes showed strong antigenicity, surface accessibility and structural flexibility, key features for effective recognition by host antibodies. These results are consistent with computational and experimental studies from Asia and Europe, where similar N-terminal and central N protein regions were identified as immunodominant and suitable for serodiagnostic uses

(21,22). Notably, these studies also emphasized that linear epitopes located in flexible coil regions are more likely to be immunogenic, which aligns with our structural mapping results.

Furthermore, the predicted epitopes demonstrated minimal sequence similarity to common human coronaviruses, suggesting a low risk of cross reactivity, a major concern in the development of SARS-CoV-2 specific diagnostic tools. This is in agreement with findings by (8,23), who underscored the need to carefully filter epitopes for cross-reactivity with endemic coronaviruses, particularly in serological assays.

Molecular docing simulations reported the predicted epitopes' ability to bind stably with monoclonal antibody fragments, offering preliminary evidence of their functional relevance. Comparable docking studies by(24,25) have demonstrated similar structural compatibility between predicted N protein epitopes and human immunoglobulins, further validation in silico approaches as a valuable screening tool for the diagnostic development.

The implications of these findings are significant for countries like Ethiopia, where access to cost-effective and variant resilient diagnostics remains limited. By targeting stable regions of the SARS-CoV-2 genome and validating their immunological relevance, this study lays the groundwork for the design of lateral flow assays of ELISAs tailored to local epidemiological context. Importantly, our approach can also serve as a template for monitoring emerging pathogens beyond COVID-19, especially in setting with limited laboratory infrastructures.

Limitation of the study

Nevertheless, the study's computational nature is a limitation. While the results are promising, they require experimental validation through peptide synthesis, antibody binding assay and clinical performances evaluation. Such validation would be crucial to confirm the predicted epitopes' diagnostic accuracy and to rule out potential false positives due to structural or conformational effects not captured in silico.

5. Conclusion

The study findings support the use of the nucleocapsid protein as a stable and immunogenic target for SARS-CoV-2 diagnostics and highlight specific epitope candidates for further development. This integrative approach, combining genomic surveillance with structural immunoinformatic, can accelerate the identification of reliable diagnostic markers amid evolving viral landscapes.

6. References

1. Zhou R, Li F, Chen F, Liu H, Zheng J, Lei C, et al. Viral dynamics in asymptomatic patients with COVID-19. International Journal of Infectious Diseases. 2020 Jul 1;96:288–90.
2. Jiang HW, Zhang HN, Meng QF, Xie J, Li Y, Chen H, et al. SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. Cell Mol Immunol. 2020 Sep;17(9):998–1000.
3. Chen Z, Zhang Z, Zhai X, Li Y, Lin L, Zhao H, et al. Rapid and Sensitive Detection of anti-SARS-CoV-2 IgG, Using Lanthanide-Doped Nanoparticles-Based Lateral Flow Immunoassay. Anal Chem. 2020 May 19;92(10):7226–31.
4. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature. 2021 Aug;596(7871):276–80.
5. Zahradnik J, Marciano S, Shemesh M, Zoler E, Harari D, Chiaravalli J, et al. SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. Nat Microbiol. 2021 Sep;6(9):1188–98.
6. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021 Apr;592(7854):438–43.
7. Xie C, Jiang L, Huang G, Pu H, Gong B, Lin H, et al. Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. International Journal of Infectious Diseases. 2020 Apr 1;93:264–7.
8. Naz A, Shahid F, Butt TT, Awan FM, Ali A, Malik A. Designing Multi-Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 (COVID-19) by Employing Immuno-Informatics Approach. Front Immunol. 2020 Jul 10;11:1663.
9. Chen J, Wang R, Wang M, Wei GW. Mutations Strengthened SARS-CoV-2 Infectivity. Journal of Molecular Biology. 2020 Sep 4;432(19):5212–26.
10. Cosar B, Karagulleoglu ZY, Unal S, Ince AT, Uncuoglu DB, Tuncer G, et al. SARS-CoV-2 Mutations and their Viral Variants. Cytokine Growth Factor Rev. 2022 Feb;63:10–22.

11. Sisay A, Abera A, Dufera B, Endrias T, Tasew G, Tesfaye A, et al. Diagnostic accuracy of three commercially available one step RT-PCR assays for the detection of SARS-CoV-2 in resource limited settings. PLoS One. 2022;17(1):e0262178.

12. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

13. GISAID Initiative [Internet]. [cited 2023 Aug 21]. Available from: https://www.epicov.org/epi3/frontend#28d967

14. NCBI Virus [Internet]. [cited 2024 Sep 18]. Available from: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202,%20taxid:2697049&utm_source=nuccore&utm_medium=referral&Authors_idx%20q.op%3DAND=Hailu&SubmitterAffil_idx%20q.op%3DOR=Ethiopian%20Public%20Health%20Institute&SubmitterCountry_idx%20q.op%3DOR=Ethiopia

15. Chang C ke, Hou MH, Chang CF, Hsiao CD, Huang T huang. The SARS coronavirus nucleocapsid protein – Forms and functions. Antiviral Res. 2014 Mar;103:39–50.

16. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021 Jul;19(7):409–24.

17. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020 Aug 20;182(4):812-827.e19.

18. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020 Mar;579(7798):265–9.

19. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol [Internet]. 2020 Jul 22 [cited 2025 May 16];11. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.01800/full

20. Ahmed SF, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses. 2020 Feb 25;12(3):254.

21. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell. 2020 Jun 25;181(7):1489-1501.e15.

22. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017 Jul 3;45(Web Server issue):W24–9.

23. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. Cell. 2021 Feb 18;184(4):861–80.

24. Johnson BA, Zhou Y, Lokugamage KG, Vu MN, Bopp N, Crocquet-Valdes PA, et al. Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. bioRxiv. 2022 Mar 28;2021.10.14.464390.

25. Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat Commun. 2020 Nov 26;11:6013.

Appendices

Appendix 1: Data abstraction tool

ID	Lab SN	Testing Laborator	Sex	Age	Permanet Residence Area Region	Zone/Sub-city	Woreda	Nationality	Location of Sample Collection	Organisation unit	Specific site of sample collection	Symptom compatible with COVID-19? (cough, fever, shortness of breath, sorethroat, headache, easy fatigue)	Co-Morbidity	If yes,	Travel History (Yes/No)	If "Yes" Travel Histrory From Where	Vaccination status

Appendix 2: Standard Operating Procedure for Manual RNA Extraction by DaAn Gene kit

Principle of the procedure: Lysis of solution contains strong protein denaturant, which can quickly dissolve protein and make nucleic acid dissociated; under the existence of lysis solution and ethyl alcohol, the dissociated nucleic acid compositions can be combined on the silicon membrane; the by actions of inhibitor remover and deionized solution, remove the protein, inorganic salt ions and many organic impurities. Then use eluent to elute pure nucleic acid.

Materials and Reagents:

1. DaAn RNA extraction kit containing;
    a. Spin columns
    b. Collection tubes
    c. Deionized solution (concentrated solution)
    d. Eluent
    e. Proteinase K
    f. Carrier RNA (dry powder)
2. Ethanol (96-100%)
3. Micro-centrifuge tubes (1.5mL)
4. Micropipette 5ul, 10ul, 20ul, 200ul, 1000ul capacity
5. Sterile, RNase-free pipette tips with aerosol barrier (filter)
6. Permanent marker pen
7. Powder free Disposable gloves, gown
8. Micro-centrifuge
9. Vortex
10. Masks
11. Timer
12. RNaseZAP, 70%alcohol, paper towel (napkin) -for cleaning and disinfection
13. Micro centrifuge racks 2-3

Reagents stability and storage:

All materials from the DaAn Gene extraction kit are stable at room temperature

Left over reconstituted carrier RNA are stored -20°C and stable for 1wk/month

Environmental and Safety Control:

- Wear dedicated lab coats which should remain in the RNA extraction and purification rooms.
- Use dedicated lab equipment for room
- Use Biosafety level II cabinet and
- Use biohazard bags for proper waste disposal/segregation

Quality Control:

Control preparation:

- Preparation of negative control is similar with extraction of patient specimens and for each batch of extraction one negative control(mock) is prepared
- RP gene (sample collection and extraction control) is prepared along with RNA

Detailed Procedure:

Procedure for Preparation of carrier RNA: Before conducting the test, lysis solution must be mixed with carrier RNA to be configured in to lysis working solution:

	1 test	20 tests	24 tests	48 tests
Lysis solution	200 μl	4 ml	4.8 ml	9.6 ml
Carrier RNA	4 μl	80 μl	96 μl	192 μl

Respiratory Sample Extraction procedure

1. Label micro centrifuge tubes with correct patient information
2. Add50 μl of proteinase K into sterile a 1.5mL micro-centrifuge tube.
3. Add 200 uL of specimen to the micro-centrifuge tube.
4. Add 200 μl lysis working solution (i.e lysis solution containing carrier RNA) fasten down the tube cover, oscalate it in vortex for 15 seconds to mix the solution sufficiently, then conduct high speed centrifugation for 10 seconds.
5. Incubate at temperature of 72 ^{o}C for 10 minutes at the same time, preheat the eluent with temperature of 72^{o}C.
6. Add 250uL of ethanol (96-100%) to the tube and mix by pulse-vortexing for 15sec. Briefly centrifuge the 1.5mL micro-centrifuge tube to remove drops from the inside of the lid at 8000 rpm for 1 minute.
7. Add 500 μl inhibitor remover into the spin column, conduct centrifugation at 8000 rpm for 1 minute at room temperature then fit the spin column in a new collection tube.
8. Add 500 μl deionized solution into the spin column, conduct centrifugation at 8000 rpm for 1 minute at room temperature then fit the spin column in a new collection tube.
9. Add 500 μl deionized solution into the spin column once again, conduct centrifugation at 8000 rpm for 1 minute at room temperature then fit the spin column in a new collection tube.
10. Place spin column and collection tube at room temperature, conduct centrifugation at14000 rpm for 3 minutes in order to remove residual ethanol.
11. Take out spin column; fit it in a new 1.5 ml centrifuge tube. Open the cover of spin column lay it aside for 2 minutes at temperature of 72 ^{o}C.
12. Carefully add 50μl eluent preheated at 72^{o}C right above the membrane of spin column, fasten down the tube cover, after standing for 1 minute at room temperature, and conduct centrifugation at 14,000 rpm for 1 minute. Then the solution in the centrifuge tube is nucleic acid solution. It is better to use it immediately, if you want to preserve it, put it at temperature of -20^{o}C.

Appendix 3: Automated RNA extraction using MegaBio plus Virus RNA Purification kit II (BSC87S1E)

Principles: Nucleic acid in serum, plasma, swabs and other samples is released by using Lysis Buffer. Released virus RNA is bound exclusively and specifically to the magnetic beads. The virus RNA boundto magnetic particles is captured by magnetic material; contaminants are removed by washing with Wash Buffer. The nucleic acid is then eluted from the practices with an Elusion Buffer.

Required equipment and reagents

- **Lysis Buffer** – contain surfactants and Tris buffer
- **Wash Buffer I** – contain High salt solution
- **Wash Buffer II** – contain Low salt solution
- **Elusion Buffer** – contain Dnase/RNase free water
- **MegaBio Reagent** – contain Magnetic particles coated with silicon
- **Magnetic rack or Bioer NPA**-32P nucleic acid purification instrument
- **Water bath or dry bath**
- **Vortex mixer**

Procedures

1. Put the 96 well pre-loaded reagent at room temperature. Invert 96-well plat upside down for three times, and tear off the plastic bag. Centrifuge the pre-loaded reagent for a few seconds (or swing by hand a few times) to avoid reagent adhering to the wall of the tubes. Tear off the almunium foil film of 96 well plate and identify the direction of the plate (magnetic beads in column #6 & #12.
2. Add 300uL sample to the 96 well plat Lysis Buffer strip (column #1 & #7), please avoid cross contamination.
3. Place 96 deep well plate to the instrument, install the 8-strip tips on the instrument
4. Run the program according to the following procedures:

Step	Well	Name	Waiting time (min:ss)	Waiting time (min:ss)	Waiting time (min:ss)	Adsorption	Speed	Volume (uL)
1	6	Beads	00:00	00:00	00:15	Normal	M	200
2	1	Binding	00:00	03:00	00:35	Strong	F	700
3	2	Wash 1	00:00	00:30	00:20	Strong	F	500
4	3	Wash 2	00:00	00:30	00:20	Strong	F	500
5	5	Elusion	01:00	02:00	00:25	Strong	F	70
6	6	Discard	00:00	00:00	00:00	Normal	M	200

Elution temperature: 80oC, Elution starts heating at step 5.

5. After the automated purification is over, transfer the Elution Buffer in column 5 and 11 to a clean nuclease free 0.5mL centrifuge tube; if not using it immediately, please store at -20 degree.

Appendix 4: Standard Operating Procedure For RT-PCR for Detecting SARS-COV-2 by DaAn Kit

Principle: This kit is based on one-step RT-PCR technique. In practice, 2019 Novel Coronavirus (2019-nCoV) ORF1ab and N genes are selected as amplification target regions. Specific primers and fluorescent probes are designed for the detection of 2019 Novel Corona virus RNA in the specimens. This kit also includes an endogenous internal standard detection system, which is used for monitoring over the processes of specimen collection, RNA and PCR amplification, thereby reducing false negative results.

Materials and Reagents:

- NC (ORF1ab/N) PCR reaction solution
- NC(ORF1ab/N) PCR reaction solution B
- Specific primers, probes, tris (hydroxymethyl) aminomethane-hydrochloric acid buffer
- Hot start TaqDNA polymerase, c-MMLV reverse transcriptase
- NC(ORF1ab/N) negative control
- NC(ORF1ab/N) positive control
- Gloves,
- Laboratory coats,
- Discard container with plastic bag
- Cold Rack for 96 well plate
- Rack for PCR ependorf tubes
- Sterile filtered tips, 200µl
- Sterile filtered tips, 10µl
- Disinfectants, ethanol, RNase ZAP
- Clear adhesive tape
- 70% ethanol
- Marker
- PCR workstation or hood

Sample Type and Storage:

Applicable sample types: throat swabs, nasopharyngeal swab.

Sample storage and transportation: The specimens for virus isolation and RNA detection should be tested as soon as possible. The specimens that can be detected within 24 hours can be stored at 4 ℃; those that cannot be detected within 24 hours should be stored at -70℃ or below (if no -70℃ storage conditions, then temporarily stored in -20℃ refrigerator). The specimens should avoid repeated freezing and thawing during transport. Specimens should be sent to the laboratory as soon as possible after collection. If the specimens need to be transported over long distances, it is recommended to use dry ice for storage.

Environmental and safety controls:

- Wear dedicated lab coats which should remain in the RNA extraction and RNA addition rooms.
- Wear and use dedicated lab coat and equipment and perform preparation in PCR work station in master mix preparation room
- Use Biosafety level II cabinet, lab coat and equipment for master mix preparation area
- Use ethanol alcohol and RNase ZAP for decontamination of PCR work station and equipment and use mini plastic waste bag(biohazard) inside the station for used pipette tips large biohazard bags outside the station for paper towel and large waste disposal

Quality Control:

- NC (ORF1ab/N) negative control: no obvious amplification curve for FAM and VIC detection channels, and Cy5 channel's Ct value<35;
- NC (ORF1ab/N) positive control: obvious amplification curves for FAM and VIC detection channels and Ct value ≤ 32, and amplification curve or no amplification curve for Cy5 channel;
- The above requirements must be met at the same time in each experiment; otherwise, the experiment is invalid and needs to be carried out again.
- Note: FAM channel for N gene detection; VIC channel for ORF1ab gene detection; Cy5 channel for internal standard.

Detailed Procedure:

PCR reagent preparation (Master Mix preparation):

Take out the NC (ORF1ab/N) PCR reaction solution A and NC (ORF1ab/N) PCR reaction solution B from the kit. After thawing at room temperature, shake and mix. Centrifuge at 8,000 rpm for a few seconds before use. Take N (N = number of specimens to be tested + NC (ORF1ab/N) negative control + NC (ORF1ab/N) positive control) PCR reaction tube. A NC single-reaction amplification system is prepared as follows:

NC(ORF1ab/N) PCR reaction solution A	NC(ORF1ab/N) PCR reaction solution B	Amplification system
17μL	3μL	20μL

After thoroughly mix the components, centrifuge for a short time to cause all the liquid on the tube wall to fall to the bottom of the tube, and then aliquot 20 μL of the amplification system into the PCR tubes. Add 5μL each of the negative control material, the RNA of specimens to be tested, and the positive control material processed into the PCR reaction tubes, cover the tubes tightly, and transfer them to the amplification detection area after centrifugation at 8,000 rpm for several seconds.

Place the reaction tube in the sample sink of the instrument.

Setting of ABI 7500 Instrument:

Open the "Setup" window, set the negative control (NTC), positive control and labelling Unknown specimen in the corresponding order, and set the specimen name in the column of "Sample Name"; the

probe detection modes are set as: Reporter1: FAM, Quencher 1: NONE; Reporter2: VIC, Quencher2: NONE; Reporter3: Cy5, Quencher3: NONE; Passive Reference: NONE.
Open the "Instrument" window and set the cycle conditions as follows:

Stage	Reps	Target	Duration	Data Collection
1	1 Cycle	50°C	00:15:00	N
2	1 Cycle	95 °C	00:15:00	N
3	45 Cycles	94 °C	00:00:15	N
		55 °C	00:00:45	Y

When finish, save the results. Adjust the Start value, End value and Threshold value of Baseline according to the image after analysis (the user can adjust them according to the actual conditions, the Start value can be set at 3~15 and the End value at 5~20, adjust the Threshold value at the Log chart window, enabling the Threshold value line to be at the log phase, the amplification curve of the negative control to be straight or lower than the threshold line), click Analysis to obtain the analysis result automatically, and read the test result in the "Report" window.

Performance Characteristics: The analytical sensitivity of this kit is 500copies/ml. Cross-reaction: no cross-reaction with other pathogens such as seasonal influenza A (H1N1) virus, novel influenza A (H1N1-2009) virus, influenza A H3N2, H5N1, H7N9, influenza B Yamagata, influenza B Victoria, RSV A, RSV B, Para influenza I, Para influenza II, Para influenza III, adenovirus types 1, 2, 3, 4, 5, 7 & 55, enterovirus types A, B, C and D, hMPV (human metapneumovirus), EB virus, measles virus, human cytomegalovirus, rotavirus, norovirus, mumps virus, varicella zoster virus, mycoplasma pneumoniae, chlamydia pneumoniae, legionella, Bordetella pertussis, Hemophilus influenzae, staphylococcus aureus, streptococcus pneumoniae, streptococcus pyogenes, Klebsiella pneumoniae, mycobacterium tuberculosis, aspergillus fumigatus, candida albicans, candida glabrata, Cryptococcus neoformans, coronavirus (HKU1, OC43, NL63, 229E), SARS corona virus, MERS corona virus, and human genomic DNA that are similar to 2019 Novel Corona virus or cause similar symptoms.

Result Reporting:

- If the test sample has no amplification curve in the FAM and VIC channels, there is amplification curve in the Cy5 channel, it can be judged that there is no 2019 Novel Coronavirus (2019-nCoV) RNA in the sample;
- If the test sample has obvious amplification curve in the FAM and VIC channels and Ct value $\leq$ 40, it can be judged that the sample is positive for 2019 Novel Coronavirus (2019-nCoV);
- If the test sample only has the Ct value $\leq$ 40 in a single channel of FAM or VIC, and there is no amplification curve in the other channel, the results need to be re-tested. If the re-test results are consistent, the sample can be judged to be positive for 2019 Novel Coronavirus (2019-nCoV). If the re-test results are negative, it can be judged that no 2019 Novel Coronaviruses (2019-nCoV) RNA has been detected.

Result Interpretation:

- Negative and positive control materials should be tested in each experiment. Only when the control materials meet the quality control requirements can the test results be determined;

- When the FAM and VC detection channels are positive, the result from the Cy5 channel (internal standard channel) may be negative due to the competition of the system;
- When the internal standard result is negative, if the FAM and VC detection channels of the test tube are also negative, it indicates that the system is inhibited or the operation is wrong, the test is invalid. Therefore, the sample needs to be re-tested;

The report is recommended to be in the following format: The format of negative result report: no 2019 Novel Coronavirus (2019-nCoV) RNA was detected in the specimens, and the concentration was lower than the sensitivity of the kit;

The format of the positive result report: 2019 Novel Coronavirus (2019-nCoV) RNA was detected in the specimens.

Appendix 5. cDNA Library Preparation and Sequencing Procedures

Annealing RNA

During this process, the extracted RNA is annealed using random hexamers to prepare for cDNA synthesis.

Preparation

1. Prepare the following consumables

Reagent	Storage	Instructions
EPH3	-25°C to -15°C	Thaw at room temperature, and then invert to mix.

2. Save the following COVIDSeq Anneal program on the thermal cycler:
    - Choose the preheat lid option
    - Setthereactionvolumeto17µl
    - 65°Cfor3minutes
    - Holdat4°C

Procedure

1. Label new PCR plate CDNA1.
2. Add 8.5µl EPH3 to each well.
3. Add 8.5µl eluted sample to each well.
4. Seal and shake at 1600rpm for1minute.
5. Centrifuge at 1000 ×gfor1minute.
6. Place on the preprogrammed thermocycler and run the COVIDSeq Anneal program.

Synthesize First Strand cDNA

Preparation

1. Prepare the following consumables

Reagent	Storage	Instructions
FSM (First Strand Mix)	-25°C to-15°C	Thaw and bring to room temperature. Invert to mix, and then keep on ice.
RVT (Reverse Transcriptase)	-25°C to-15°C	Invert to mix before use. Keep on ice.

2. Save the following COVIDSeq FSS program on the thermocycler:
    - Choose the preheat lid option
    - Set the reaction volume to 25µl

- 25°C for 5minutes
- 50°C for 10minutes
- 80°C for 5minutes
- Hold at 4°C

Procedure

1. In a 1.7 ml tube, combine the following volumes to prepare First Strand cDNA Master Mix. Multiply each volume by the number of samples.
    - FSM(9µl)
    - RVT(1µl)

Reagent overage is included to account for small pipetting errors.

2. Add8µlmastermixtoeachwelloftheCDNA1plate.
3. Sealandshakeat1600rpmfor1minute.
4. Centrifuge at 1000 ×gfor1minute.
5. Place on the preprogrammed thermocycler and run the COVIDSeq FSS program.

SAFE STOPPING POINT

If you are stopping, seal the plate and store at-25°C to-15°C for up to 7 days.

Amplify cDNA

This step uses two separate PCR reactions to amplify cDNA

Preparation

1. Prepare the following consumables

Reagent	Storage	Instructions
C5P1	-25°C to-15°C	Thaw at room temperature. keep on ice until use
C5P2	-25°C to-15°C	Thaw at room temperature. keep on ice until use
IPM	-25°C to-15°C	Thaw at room temperature, and then invert to mix. Keep on ice until use

2. Save the following COVIDSeq PCR program on the thermocycler:
    - Choose the preheat lid option
    - Set the reaction volume to 25µl
    - 98°C for 3minutes
    - 35 cycles of:
        - 98°C for 15 seconds
        - 63°C for 5minutes
    - Hold at4°C

Procedure

1. Label two new PCR plates COV1 and COV2.

   The plates represent two separate PCR reactions on each sample and control in the CDNA1 plate.

2. In a 15ml tube, combine the following volumes to prepare COVIDSeq PCR1MasterMixand COVIDSeq PCR2 Master Mix. Multiply each volume by the number of samples.
   Reagent overage is included to account for small pipetting errors.

Reagent	COVIDSeq PCR 1 Master Mix (µl)	COVIDSeq PCR 2 Master Mix (µl)
IPM	15	15
C5P1	4.3	N/A
C5P2	N/A	4.3
Nuclease-free water	4.7	4.7

3. Add 20 µl COVIDSeq PCR 1 Master Mix to each well of the COV1 plate corresponding to each well of the CDNA1 plate.
4. Add 5 µl first strand cDNA synthesis from each well of the CDNA1 plate to the corresponding well of the COV1 plate.
5. Add 20 µl COVIDSeq PCR 2 Master Mix to each well of the COV2 plate corresponding to each well of the CDNA1 plate.
6. Add 5 µl first strand cDNA synthesis from each well of the CDNA1 plate to the corresponding well of the COV2 plate.
7. Seal and shake at 1600 rpm for 1 minute.
8. Centrifuge at 1000 x g for 1 minute.
9. Place in the preprogrammed thermal cycler and run the COVIDSeq PCR program.

SAFE STOPPING POINT

If you are stopping, seal the plate and store at -25°C to -15°C for up to 3 days.

Tagment PCR Amplicons

This step uses EBLTS to tagment PCR amplicons, which is a process that fragments and tags the PCR amplicons with adapter sequences.

Preparation

1. Prepare the following consumables:

Reagent	Storage	Instructions
EBLTS	2°C to 8°C	Bring to room temperature. Vortex thoroughly before use.
TB1	-25°C to -15°C	Bring to room temperature. Vortex thoroughly before use.

2. If COV1 and COV2 plates were stored frozen, prepare as follows.
   a. Thaw at room temperature.
   b. Check seals, and then shake at 1600 rpm for 1 minute.

c. Centrifuge at 1000 x g for 1 minute.

3. Save the following COVIDSeq TAG program on the thermal cycler:
   - Choose the preheat lid option
   - Set the reaction volume to 50 µl
   - 55°C for 5 minutes
   - Hold at 10°C

Procedure

1. Label a new PCR plate TAG1.
2. Combine COV1 and COV2 as follows.
   a. Transfer 10 µl from each well of the COV1 plate to the corresponding well of the TAG1 plate.
   b. Transfer 10 µl from each well of the COV2 plate to each well of the TAG1 plate containing COV1.
3. In a 15 ml tube, combine the following volumes to prepare Tagmentation Master Mix. Multiply each volume by the number of samples.
   – TB1 (12 µl)
   – EBLTS (4 µl)
   – Nuclease-free water (20 µl)
4. Add 30 µl master mix to each well in TAG1 plate.
5. Seal and shake at 1600 rpm for 1 minute.
6. Place on the preprogrammed thermal cycler and run the COVIDSeq TAG program.

Post Tagmentation Clean Up

This step washes the adapter-tagged amplicons before PCR amplification.

Preparation

1. Prepare the following consumables:

Reagent	Storage	Instructions
ST2	Room temperature	Vortex before use.
TWB	2°C to 8°C	Vortex before use.

Procedure

1. Centrifuge the TAG1 plate at 500 x g for 1 minute.
2. Add 10 µl ST2 to each well of the TAG1 plate.
3. Seal and shake at 1600 rpm for 1 minute.
4. Incubate at room temperature for 5 minutes.
5. Centrifuge at 500 × g for 1 minute.
6. Place on the magnetic stand and wait until the liquid is clear (~3 minutes).

7. Inspect for bubbles on the seal. If present, centrifuge at 500 x g for 1 minute, and then place on the magnetic stand (~3 minutes).
8. Remove and discard all supernatant.
9. Wash beads as follows.
    a. Remove from the magnetic stand.
    b. Add 100 µl TWB to each well.
    c. Seal and shake at 1600 rpm for 1 minute.
    d. Centrifuge 500 × g for 1 minute.
    e. Place on the magnetic stand and wait until the liquid is clear (~3 minutes).
    f. For first wash only, remove and discard all supernatant from each well.
10. Wash beads a **second** time.
    Leave supernatant in plate for second wash to prevent beads from over drying.

Amplify Tagmented Amplicons

This step amplifies the tagmented amplicons using a PCR program. The PCR step adds prepaired 10 base pair Index 1 (i7) adapters, Index 2 (i5) adapters, and sequences required for sequencing cluster generation.

Preparation

1. Prepare the following consumables:

Reagent	Storage	Instructions
EPM	-25°C to -15°C	Invert to mix. Keep on ice until use.
Index adapters	-25°C to -15°C	Thaw at room temperature. Vortex to mix, and then centrifuge at 1000 × g for 1 minute.

2. Open each prepared index adapter plate seal as follows. Use a new PCR plate for each different index set.
    a. Align a new 96-well PCR plate above the index adapter plate, and then press down to puncture the foil seal.
    b. Discard the PCR plate.
3. Save the following COVIDSeq TAG PCR program on the thermal cycler:
    - Choose the preheat lid option and set to 100°C
    - Set the reaction volume to 50 µl
    - 72°C for 3 minutes
    - 98°C for 3 minutes
    - 7 cycles of:
    - 98°C for 20 seconds
    - 60°C for 30 seconds
    - 72°C for 1 minute

- 72°C for 3 minutes
- Hold at 10°C

Procedure

1. In a 15 ml tube, combine the following volumes to prepare PCR Master Mix. Multiply each volume by the number of samples.
   – EPM (24 µl)
   – Nuclease-free water (24 µl)
2. Vortex PCR Master Mix to mix.
3. Keep the TAG1 plate on magnetic stand and remove TWB.
4. Use a 20 µl pipette to remove any remaining TWB.
5. Remove the TAG1 plate from the magnetic stand.
6. Add 40 µl PCR Master Mix to each well.
7. Add 10 µl index adapters to each well of the PCR plate.
8. Seal and shake at 1600 rpm for 1 minute.
9. If liquid is visible on the seal, centrifuge at 500 x g for 1 minute.
10. Inspect to make sure beads are resuspended. To resuspend, set your pipette to 35 µl with the plunger down, and then slowly pipette to mix.
11. Place on the preprogrammed thermal cycler and run the COVIDSeq TAG PCR program.

Pool and Clean Up Libraries

This step combines libraries from each 96-well sample plate into one 1.7 ml tube. Libraries of optimal size are then bound to magnetic beads, and fragments that are too small or large are washed away.

Preparation

1. Prepare the following consumables:

Reagent	Storage	Instructions
ITB or IPB	Room temperature	Vortex thoroughly to mix.
RSB	2°C to 8°C	Let stand for 30 minutes to bring to room temperature.
		Vortex and invert to mix.

2. Prepare 2.5 ml 80% EtOH from absolute EtOH for each tube of pooled libraries.

Procedure for Illumina COVIDSeq Test (RUO)

The following steps describe the procedure for the Illumina COVIDSeq Test (RUO) kit.

1. Centrifuge the TAG1 plate at 500 × g for 1 minute.
2. Place on the magnetic stand and wait until the liquid is clear (~3 minutes).
3. To pool libraries, do as follows. Repeat the steps for each additional sample plate.

a. Use a 20 µl eight-channel pipette to transfer 5 µl library from each well of the TAG1 plate to a PCR 8-tube strip, resulting in 60 µl pooled library per row. Change tips after each column.
b. Label a new 1.7 ml tube Pooled ITB.
c. Transfer 55 µl pooled library from each well of the PCR 8-tube strip into the Pooled ITB tube. For each sample plate, these volumes result in 440 µl pools of pooled libraries.

If processing 3072 samples, these steps result in 32 Pooled ITB tubes.

4. Vortex the Pooled ITB tubes to mix, and then centrifuge briefly.
5. Vortex ITB to resuspend.
6. Add ITB using the resulting volume of Pooled ITB tube volume multiplied by 0.9. For example, for 96 samples, add 396 µl ITB to each tube.
7. Vortex to mix.
8. Incubate at room temperature for 5 minutes.
9. Centrifuge briefly.
10. Place on the magnetic stand and wait until the liquid is clear (~5 minutes).
11. Remove and discard all supernatant.
12. Wash beads as follows.
    a. Keep on the magnetic stand and add 1000 µl fresh 80% EtOH to each tube.
    b. Wait 30 seconds.
    c. Remove and discard all supernatant.
13. Wash beads a **second** time.
14. Use a 20 µl pipette to remove all residual EtOH.
15. Add 55 µl RSB.

    Note: Due to library yield excess, the RSB volume does not impact batches with a small number of samples.

16. Vortex to mix, and then centrifuge briefly.
17. Incubate at room temperature for 2 minutes.
18. Place on the magnetic stand and wait until the liquid is clear (~2 minutes).
19. Transfer 50 µl supernatant from each Pooled ITB tube to a new microcentrifuge tube.

SAFE STOPPING POINT

If you are stopping, cap the tube and store at -25°C to -15°C for up to 30 days.

Quantify and Normalize Libraries

1. Analyze 2 µl library pool using a Qubit dsDNA HS Assay kit.
   If libraries are outside the standard range, dilute to 1:10 concentration, and analyze again.
2. Calculate the molarity value using the following formula.
   - Use 400 bp as the average library size.

$$\frac{Library\ concentration\ ng/\mu l}{660\ \frac{g}{mol}\ x\ average\ library\ size\ (bp)} x10^6 = Molarity\ (nM)$$

3. Dilute each library pool to a minimum of 30 µl at a normalized concentration 4 nM using RSB.

Pool and Dilute Libraries

After diluting to the starting concentration of 4 nM, libraries are ready to be denatured and diluted to the final loading concentration.

1. Transfer the designated volume of normalized libraries containing the appropriate index adapter sets to a new microcentrifuge tube for each number of samples specified in for v3 primers and Table 1.
   If you have multiple normalized pools, combine the designated volume of each normalized pool in the tube. Doing so produces a final pool of samples diluted to a starting concentration of 4 nM. Do not combine pools with the same index adapter set.

   Table 1 Normalized Pool Volumes and Sample Numbers for Denature and Dilution by Instrument

Sequencing System	Volume of Normalized Libraries Used per Run (µl)	Samples per Final Pool of Normalized Libraries	Samples per Flow Cell
NextSeq 500/550 or 550Dx HO Flow Cell	25	384	384

   2. Follow the denature and dilute instructions for your system to dilute to the final loading concentration.
   - For the NextSeq 500/550 Sequencing System and NextSeq 550Dx Sequencing System, see the *NextSeq System Denature and Dilute Libraries Guide (document # 15048776)*.

3. Use the following loading concentrations for your system.
   Table 2 Loading Concentrations by Instrument

Sequencing System	**Starting Concentration (nM)**	**Final Loading Concentration (pM)**
NextSeq 500/550 or 550Dx HO flow cell	4	1.4

Prepare for Sequencing

The Illumina COVIDSeq Test (RUO) is compatible with the NovaSeq 6000 Sequencing System SP and S4 flow cells, the NextSeq 2000 Sequencing System, the NextSeq 500/550 Sequencing Systems, and the NextSeq 550DX instrument.

Read Recommendations for Sequencing

Read length and depth can be optimized based on sample type and workflow needs. Viral titer and quality of extracted RNA may impact the read configuration necessary to achieve optimal consensus genome calling. Increasing read depth may improve coverage.

Read length of 2 × 150 is recommended for all runs. Shorter read lengths down to 2 × 75 may be used.

For low complexity high-titer samples (for example nasopharyngeal swabs) a minimum read depth of 0.5M fragments (1M total reads) is recommended.

For high complexity low-titer samples (for example wastewater) a minimum read depth of 4M fragments (8M total reads) is recommended.

Set Up Sequencing Run for Illumina COVIDSeq Test (RUO)

Refer to the documentation for your sequencing system and the following information to set up your run.

If using the NextSeq 500/550 or NexSeq 550Dx, refer to the *NextSeq 500 System Guide (document # 15046563)*, *NextSeq 550 System Guide (document # 15069765)*, or *NextSeq 550Dx Instrument Reference Guide (document # 1000000009513)*.

- If using Local Run Manager, refer to the *Local Run Manager Software Guide (document # 1000000111492)*.
- Enter **Paired End** as the Read Type.
- Enter **10** as the value for Index 1 and Index 2.

Analysis Software

After sequencing completes, analysis either takes place locally using installed pipeline software or in Base Space Sequence Hub.

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Getnet Hailu
Assignment title:	Dissertation
Submission title:	Dissertation
File name:	Getnet_Dissertation_GH_Final.docx
File size:	30.11M
Page count:	231
Word count:	51,165
Character count:	292,087
Submission date:	31-May-2025 10:30AM (UTC+0300)
Submission ID:	2688966370

Molecular Characterization of SARS-CoV-2 and Evaluation of Extraction-Free RT-qPCR Detection Methods for the Development of an Alternative and Affordable Diagnostic Tools.

PhD Dissertation

By

Getnet Hailu Alemayehu

Addis Ababa University

School of Graduate Studies

Tropical and Infectious Diseases Program

May, 2025

Addis Ababa, Ethiopia

Dissertation

ORIGINALITY REPORT

22%	19%	16%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	support.illumina.com.cn Internet Source	4%
2	gcris.iyte.edu.tr Internet Source	3%
3	emea.support.illumina.com Internet Source	2%
4	Verónica Roxana Flores-Vega, Jessica Viridiana Monroy-Molina, Luis Enrique Jiménez-Hernández, Alfredo G. Torres et al. "SARS-CoV-2: Evolution and Emergence of New Viral Variants", Viruses, 2022 Publication	1%
5	www.esp.org Internet Source	1%
6	Submitted to Addis Ababa University Student Paper	1%
7	www.wjgnet.com Internet Source	1%
8	www.frontiersin.org	