

**GENETIC POLYMORPHISMS IN *ACE2* TRANSCRIPTS OF SARS-COV-2
POSITIVE INDIVIDUALS FROM FOUR SELECTED WEST AFRICAN
COUNTRIES**



A RESEARCH PROJECT SUBMITTED TO
THE DEPARTMENT OF BIOCHEMISTRY, CELL & MOLECULAR BIOLOGY,
COLLEGE OF BASIC AND APPLIED SCIENCES, SCHOOL OF BIOLOGICAL SCIENCES,
UNIVERSITY OF GHANA, LEGON,
BY BECKY EWURAMA TETTEH
(11008583)

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER
OF PHILOSOPHY IN MOLECULAR CELL BIOLOGY OF INFECTIOUS DISEASES

NOVEMBER 2024

ABSTRACT

Coronavirus disease 2019 (COVID-19) claimed the lives of millions of people worldwide but with great disparity in the number of cases and deaths across different populations. Multiple hypotheses have been proposed to explain the relatively low COVID-19 in the West African population compared to other parts of the world. The role of host-genetics has been explored in a quest to explain the disparity. Multiple population-specific polymorphisms within receptors which allow SARS-CoV-2 cellular invasion have been identified with varying effects on COVID-19 severity. Using targeted sequencing, this study sought to explore SNPs (Single Nucleotide Variations) within ACE2 (Angiotensin-converting enzyme 2); the primary receptor for SARS-CoV-2 cellular entry, and their impact on its function in COVID-19 severity for four West African countries; Ghana, Nigeria, Benin and Guinea.

RNA (n = 330) was extracted from archived respiratory samples of SARS-CoV-2 positive individuals across the four countries. The RNA was converted to cDNA and PCR optimization for twelve overlapping primers targeting *ACE2* was carried out. Six out of these primers targeting eleven *ACE2* exons were optimized successfully and sequenced on the Oxford Nanopore Technology platform. Following bioinformatic analyses, twelve SNPS in total were identified. Five previously reported *ACE2* missense variants (G353V, T434K, V573A, L585P and S607G) were identified at low frequencies within the sites used in this study. The substitution G352V has been previously described to confer protection against SARS-CoV-2. PolyPhen-2 analysis of the other four previously reported SNPs indicated possible structural changes to ACE2 which may affect its interaction with SARS-CoV-2. Seven novel *ACE2* SNPs; W349*, D350E, D350V, G352W, G352R, G352E, and K353Q were also identified within the four sites. D350G had a high

frequency across all the four West African countries (countries (Benin – 100%, Ghana – 56.2%, Guinea – 84.2% and Nigeria – 77.4%). The modelled structure of ACE2 with each of the three of the novel SNPs, their interactions with Spike receptor binding domain (RBD) and estimated binding affinity was simulated using I-TASSER, ClusPro and PRODIGY respectively. All modelled variants had a relatively lower binding interaction with the Spike protein compared to wild-type ACE2. The lower binding interactions observed potentially suggest a lower rate of SARS-CoV-2 cellular invasion which may be one of the reasons for the low severity of SARS-CoV-2 observed in West Africa.



DEDICATION

This thesis is dedicated to my family for always supporting and believing in me; I could not have come this far without you.



ACKNOWLEDGEMENT

My first thanks go to my supervisors; I am very grateful for their supervision and mentorship. Thank you, Dr Peter Quashie, for challenging and believing in me to take on this project; I am very grateful for the opportunity. The times you would show up in the lab and provide suggestions and encouragement while I worked were very comforting. Thank you, Dr Lily Paemka, for always being available to hear my thoughts on my project, providing solutions and encouraging me to be persistent with your experiences; you always make me feel confident and ready to face my challenges boldly.

I am very grateful for the friends I made along the journey: coursemates, labmates, and other scientists. Thank you Dr Kwamena Abbiw for constantly helping me to structure my work and giving good critiques. Thank you, Mrs. Joyce Ngoi and the NGS team for providing the support I needed during the sequencing step. Thank you to Esther Kyeremateng, Bright Hodogbe, Enoch Langmer, Alberta Serwaa, and Adwoa Ayimadu for helping me out in the lab, even on days when it was not convenient. Thank you to Dr Micheal Yakass and Angela Ayiku for helping me during my optimization and troubleshooting steps. Thank you to Dr Kesego Tapela and Fatima Oyawoye for your contributions on preliminary work leading to this project and constant support.

I also appreciate all participants of the study, partners of the West African Network of Infectious Diseases African Centres of Excellence (WANIDA) and funders - WANIDA and the Gates foundation.

Finally, thank you to God; this has been a long and hard one.

DECLARATION

I, Becky Ewurama Tetteh of the West African Centre of Cell Biology for Infectious Pathogens and the Department of Biochemistry, Cell and Molecular Biology, University of Ghana, do hereby declare that except for the cited references, this thesis was duly carried out by me and the results obtained therein are the true reflection of my project under the supervision of Dr Peter Kojo Quashie and Dr Lily Paemka.

Signature:  Date: 25th November, 2024

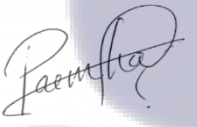
Becky Ewurama Tetteh

(Student)

Signature:  Date: 26th November, 2024

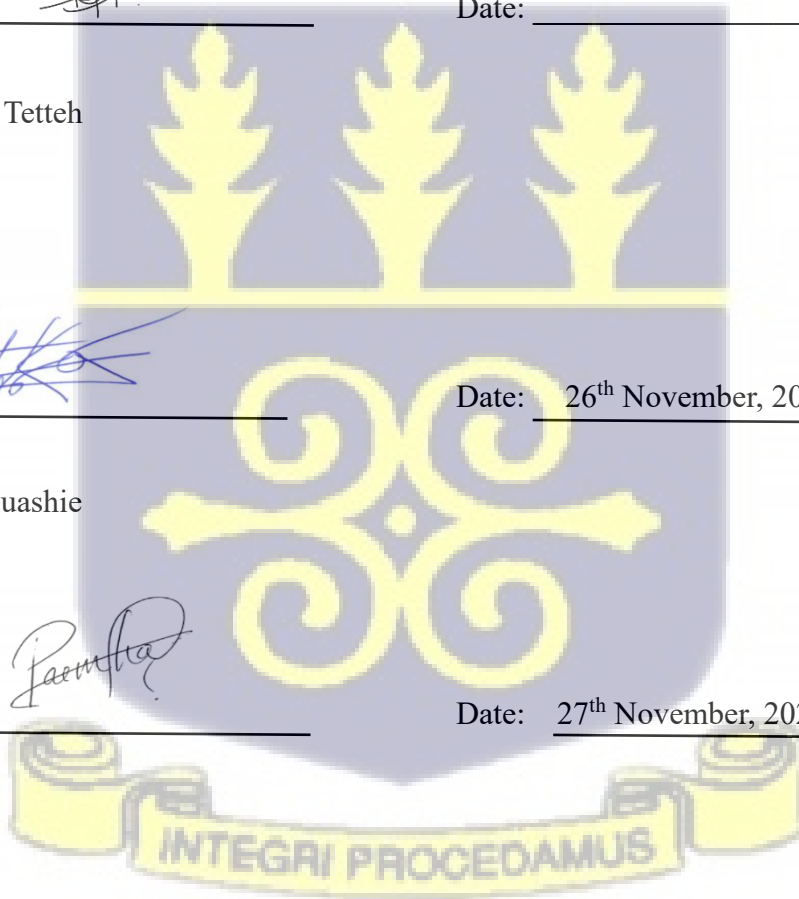
Dr. Peter Kojo Quashie

(Supervisor)

Signature:  Date: 27th November, 2024

Dr. Lily Paemka

(Co-Supervisor)



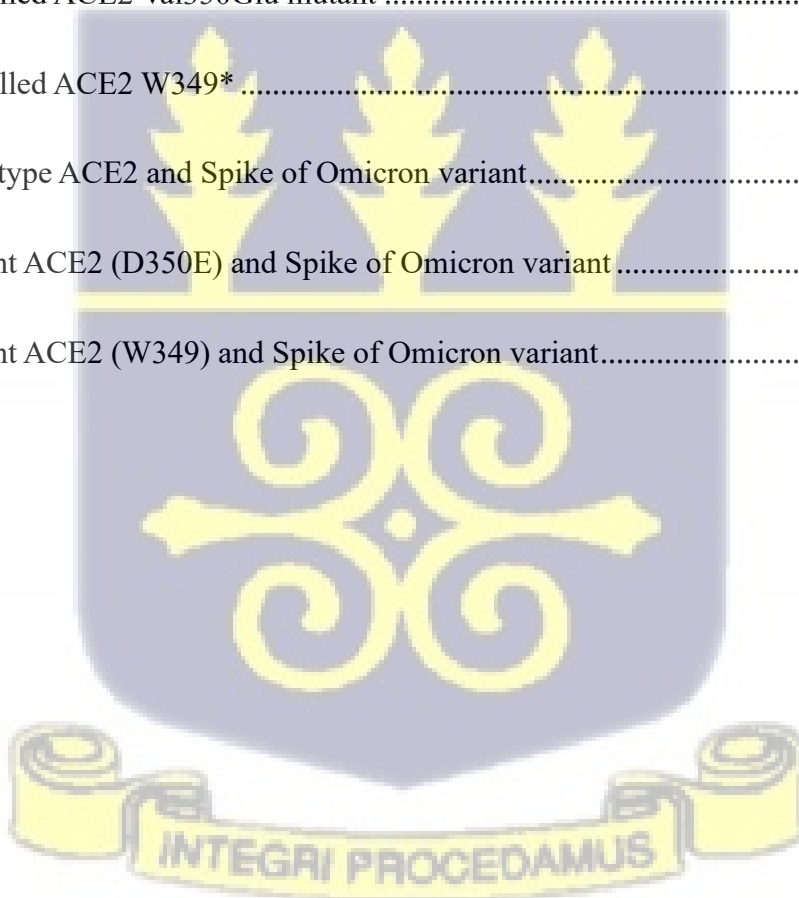
LIST OF TABLES

Table 1: Total number of samples submitted for each WANIDA centre.....	29
Table 2: Master mix reaction for Q5 Hot Start high fidelity 2x master mix and GoTaq ® G2 Hot Start master mix	36
Table 3: PCR condition for Q5 Hot Start high fidelity 2x master mix and GoTaq ® G2 Hot Start master mix.....	37
Table 4: Master mix reaction for One Taq® Quick-Load quick-load 2X master mix with standard buffer master mix	37
Table 5: PCR conditions for ACE2 (primer 2 to 8) using One Taq® Quick-Load 2X master mix with standard buffer	38
Table 6: Number of sample types obtained from each WANIDA centre and number of samples that were sequenced successfully from each centre.....	45
Table 7: Specific gene and amino acid variations identified	51
Table 8: Alpha missense prediction and stability fold using FoldX for known SNPs identified in the Uniprot database	54
Table 9: PolyPhen-2 prediction score, description, sensitivity and specificity of novel SNPs identified	55
Table 10: Summary of binding energy and binding affinity for reference ACE2 and two mutants	61

LIST OF FIGURES

Figure 1: Orthocoronavirinae subfamily classification.....	7
Figure 2: Genome organization of some coronaviruses.....	9
Figure 3: Coronavirus virion structure.....	10
Figure 4: Primary receptors of the seven known human coronaviruses	16
Figure 5: Total number of reported cases retrieved from the WHO dashboard on 30th September, 2024.....	17
Figure 6: Molecular diagnostic methods for COVID-19	21
Figure 7: SARS-CoV-2 cellular invasion.....	28
Figure 8: African map showing the selected West African countries	30
Figure 9: Age classification of samples sequenced.....	32
Figure 10: General workflow for targeted sequencing	33
Figure 11: Schematic of overlapping primers designed for ACE2	35
Figure 12: Region of ACE2 protein variants amplified based on NCBI sequence alignment tool	36
Figure 13: Bioinformatic analysis general workflow	41
Figure 14: In-silico analysis pipeline.....	43
Figure 15: Twelve overlapping ACE2 amplicons after primer optimization.....	47
Figure 16: Gel representation of 85 samples after amplification of the four selected ACE2 primers	48

Figure 17: Gel representation of 24 samples amplified using GoTaq ® G2 Hot Start master mix	49
Figure 18: Number of types of SNPs identified.....	50
Figure 19: Number of known and novel SNPs identified.....	50
Figure 20: Distribution of D350E genotypes across each country	52
Figure 21: Distribution of D350V genotypes across each country	53
Figure 22: Superimposition of reference ACE2 (6M18) and D350V/Glu mutant.....	56
Figure 23: Modelled ACE2 Val350Glu mutant	56
Figure 24: Modelled ACE2 W349*	57
Figure 25: Wild-type ACE2 and Spike of Omicron variant.....	58
Figure 26: Mutant ACE2 (D350E) and Spike of Omicron variant	59
Figure 27: Mutant ACE2 (W349) and Spike of Omicron variant.....	60



LIST OF ABBREVIATIONS

SNPs	Single Nucleotide Variations
COVID-19	Coronavirus disease 2019
CoVs	Coronaviruses
HCoV	Human coronavirus
SARS-CoV	Severe Acute Respiratory Syndrome-Related Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome-Related Coronavirus-2
MERS-CoV	Middle East Respiratory Syndrome-Related Coronavirus
ACE2	Angiotensin-converting enzyme 2
HEL1	Helicase Of Superfamily 1
RBD	Receptor Binding Domain
NTD	N-terminal domain
CTD	C-terminal domain
Nsps	Non-structural proteins
RdRp	RNA-dependent RNA polymerase
DPP4	Dipeptidyl peptidase 4
CD	Cluster of differentiation
ANPEP	Aminopeptidase
BSG	Basigin
EMMPRIN	Extracellular matrix metalloproteinase inducer
TMPRSS2	Transmembrane serine protease 2
CTSL/CTSB	Cathepsin L and Cathepsin B

RT-PCR	Reverse transcription polymerase chain reaction
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
RT	Room temperature
cDNA	Complementary DNA
TAE	Tris-acetate ethylenediaminetetraacetic acid
SNPs	Single Nucleotide Polymorphisms
QC	Quality Control
PRRs	Pattern Recognition Receptors
DAMPs	Damage-Associated Molecular Patterns
PAMPs	Pathogen-Associated Molecular Patterns
CD	Cluster of differentiation
Ig	Immunoglobulin
WHO	World Health Organization
ORF	Open Reading Frame
pp	Polyprotein
RMSD	Root mean square deviation
NCBI	National Centre for Biotechnology Information
BLAST	Basic Local Alignment Search Tool
I-TASSER	Iterative Threading ASSEmbly Refinement
PolyPhen-2	Polymorphism Phenotyping version 2
FFT	Fast Fourier transform
PRODIGY	PROtein binding enerGY prediction

CEA-PCMT	Centre d'Excellence Africain pour le Prevention et le Contrôle des Maladies Transmissibles
ACEGID	African Centre of Excellence for the Genomics of Infectious Diseases
ACEPHAP	African Centre of Excellence for Population Health and Policy
ACENTDFB	African Centre of Excellence for Neglected Tropical Diseases and Forensic Biotechnology
ACEMFS	African Centre of Excellence for Mycotoxins and Food Safety
IRCB/CERPAGE	Institut de Recherche Clinique du Benin/Centre d'Etude et de Recherche sur les Pathologies Associées à la Grossesse et à l'Enfance



CHAPTER ONE

1.0 INTRODUCTION

Coronaviruses of the family *Coronaviridae* are single-stranded, positive-sense RNA viruses with a broad host range causing a plethora of clinical outcomes (Gomes et al., 2020; Mulabbi et al., 2021). Prior to the coronavirus disease 2019 (COVID-19) pandemic, six human coronaviruses (HCoVs) had been identified to cause respiratory illness in humans, with the elderly and immunocompromised persons at risk of severe disease and death (Liu et al., 2021; Mulabbi et al., 2021). The six previously identified HCoVs are HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), and Middle East Respiratory Syndrome-Related Coronavirus (MERS-CoV) (Liu et al., 2021; Mulabbi et al., 2021). The aetiological agent for the recent COVID-19 pandemic; SARS-CoV-2, has been grouped with SARS-CoV and MERS-CoV due to their relatively high pathogenicity (causing severe lower respiratory tract infection) while the other human coronaviruses are typified by mild clinical presentation (15% to 30% of common colds) (Li et al., 2019).

A global count of 775,513,850 confirmed COVID-19 cases had been reported by May 12th, 2024 with 1.2% of these cases representing the African continent (WHO, 2024). The spread of SARS-CoV-2 with an unequal number of cases count and case fatality rates globally (WHO, 2024), raises a concern about viral pathogenesis. Different hypotheses such as frequent exposure to other infectious agents, lower life expectancy, and population mean age, have been proposed to explain the pattern of disease especially for West Africans with a high proportion of asymptomatic cases (Lawal, 2021; Mensah et al., 2022; Oduro-Mensah et al., 2022; Quashie et al., 2021). Amid all the possible hypotheses that have emerged, genetic variations have also stood as a plausible

justification for the lower number of cases observed during the COVID-19 pandemic in Africa (Khayat et al., 2020; Kaltoum, 2021; Sandi et al., 2024).

Infection begins when the Spike protein of SARS-CoV-2 interacts with ACE2 (Angiotensin-converting enzyme 2) expressed on host epithelial cells (Lamers & Haagmans, 2022). After this interaction, TMPRSS2 (Transmembrane Serine Protease 2) is recruited for viral fusion by cleavage of the S2' position of the Spike protein (Lamers & Haagmans, 2022). Entry may also be via the endosomal route of cellular invasion. In this case, Cathepsins L and B are recruited for activation of the Spike protein for RNA release and replication (Gomes et al., 2020; Prasad et al., 2021). Other receptors such as Basigin, Aminopeptidase (ANPEP; receptor for HCoV 229E) and Dipeptidyl peptidase-4 (DPP4; receptor for MERS-CoV) have also been proposed to allow SARS-CoV-2 cellular invasion (Qi et al., 2020). Mutations in all these receptors could have different implications on the level of viral cellular invasion which could be an important factor in the varying clinical phenotype of COVID-19 globally (Duah-Quashie et al., 2024; Jalaledine et al., 2022; Pišlar et al., 2020; Srivastava et al., 2020; Wulandari et al., 2021; Yaghoobi et al., 2023).

Based on the general trend of COVID-19 in West Africa; high asymptomatic cases (Quashie et al., 2021) and lower number of deaths (Parbie et al., 2020; Quashie et al., 2021), it is important to understand the factors driving low severity as the prediction of fatal cases for Africans was not achieved during the pandemic (Njenga et al., 2020; Oduro-Mensah et al., 2022). Immune modulation is a key hypothesis (Tapela et al., 2022; Sandi et al., 2024; Tapela et al., 2024), but it is also important to determine if host genetics play a role in altering the receptors involved in viral entry thereby enhancing or reducing infectivity. Thus, this study sought to identify SNPs present

in *ACE2*; the primary receptor for SARS-CoV-2 cell invasion, and how these variants influence viral entry in four selected West African populations.

1.1 PROBLEM STATEMENT

Early in the COVID-19 pandemic, Africa was predicted to experience high fatalities due to reasons such as poor healthcare systems, malnutrition, and poor surveillance to detect infections on time (Lawal, 2021; Njenga et al., 2020; Tsinda & Mmbando, 2021). Interestingly, the prediction of highly severe cases and deaths was not realized, especially for West Africa. This has led to multiple hypotheses to explain the pattern of COVID-19. Some of these hypotheses include a younger average population, exposure to host gene polymorphisms, endemic diseases like malaria, warm climate, lower life expectancy, and reduced production of pro-inflammatory immune response indicators (Lawal, 2021; Monticelli et al., 2021; Njenga et al., 2020; Oduro-Mensah et al., 2022; Sandi et al., 2024). Studies have identified polymorphisms in the *ACE2* host receptor for SARS-CoV-2 cellular invasion across multiple populations (Ali et al., 2020; Srivastava et al., 2020; Rokni et al., 2022; Duah-Quashie et al., 2024). Recently, unique SNPs have been identified within the first four exons of *ACE2* in Ghanaians (Duah-Quashie et al., 2024). These appear to have the potential to reduce SARS-CoV-2 cellular invasion (Duah-Quashie et al., 2024). However, there are no reports on *ACE2* SNPs in Nigeria, Benin, and Guinea, which also had low COVID-19 severity and deaths during the pandemic. This study focused on identifying SNPs within eleven exons of *ACE2* for individuals in Ghana, Nigeria, Benin and Guinea to further explore the role of host genetics in COVID-19.

1.2 RATIONALE

An important step in viral pathogenesis is receptor interaction. Understanding how a receptor and ligand interact can inform the different clinical presentation of pathogens, immune evasion mechanisms, and interventions such as vaccines and medicines (Everest et al., 2022; Wong & Saier, 2021; F. Li, 2015). Previous *in-vitro* work in our group (Quashie and Paemka laboratories) had identified SNPs of *ACE2* with the potential to affect the infectivity and pathogenicity of SARS-CoV-2. Five SNPs within *ACE2*; S19P, K341R, M383T, D427Y, S692P and L731F reportedly present in Africans, demonstrated a lower infectivity rate with SARS-CoV-2 compared to wild-type *ACE2* (Oyawoye *et. al*, unpublished). This preliminary data suggested that *ACE2* SNPs might play roles in the infectivity and the pathogenicity of SARS-CoV-2. Additionally, a recent report has linked polymorphisms in the first four exons of *ACE2* to the low susceptibility of SARS-CoV-2 in the Ghanaian population (Duah-Quashie et al., 2024). Across the West African region, COVID-19 cases were predominantly mild or asymptomatic compared to other parts of the world. Given the current reports on SNPs identified (Duah-Quashie et al., 2024), there is the need to further investigate the role of host genetics in SARS-CoV-2 cellular invasion. Thus, polymorphisms within the West African region may have contributed to the low severity of COVID-19 within the sub-region.

Since SARS-CoV-2 cellular invasion requires interaction with *ACE2* for an infection to occur, it is necessary to identify SNPs which may alter protein structure and function. These SNPs may play a role in COVID-19 as they can alter SARS-CoV-2 cellular invasion. This will provide more evidence for the disparity in COVID-19 cases and severity observed globally.

89 **1.3 AIM**

90 To identify ACE2 receptor SNPs in persons from Ghana, Nigeria, Guinea and Benin and how these
91 polymorphisms impact virus-host interactions using archived respiratory samples.

92

93

94 **1.4 SPECIFIC OBJECTIVES**

95 1. To identify *ACE2* variants using targeted sequencing.

96 2. To perform *in-silico* validation of the effect of identified ACE2 variants on SARS-CoV-2
97 cellular invasion.

98



99

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Coronaviruses

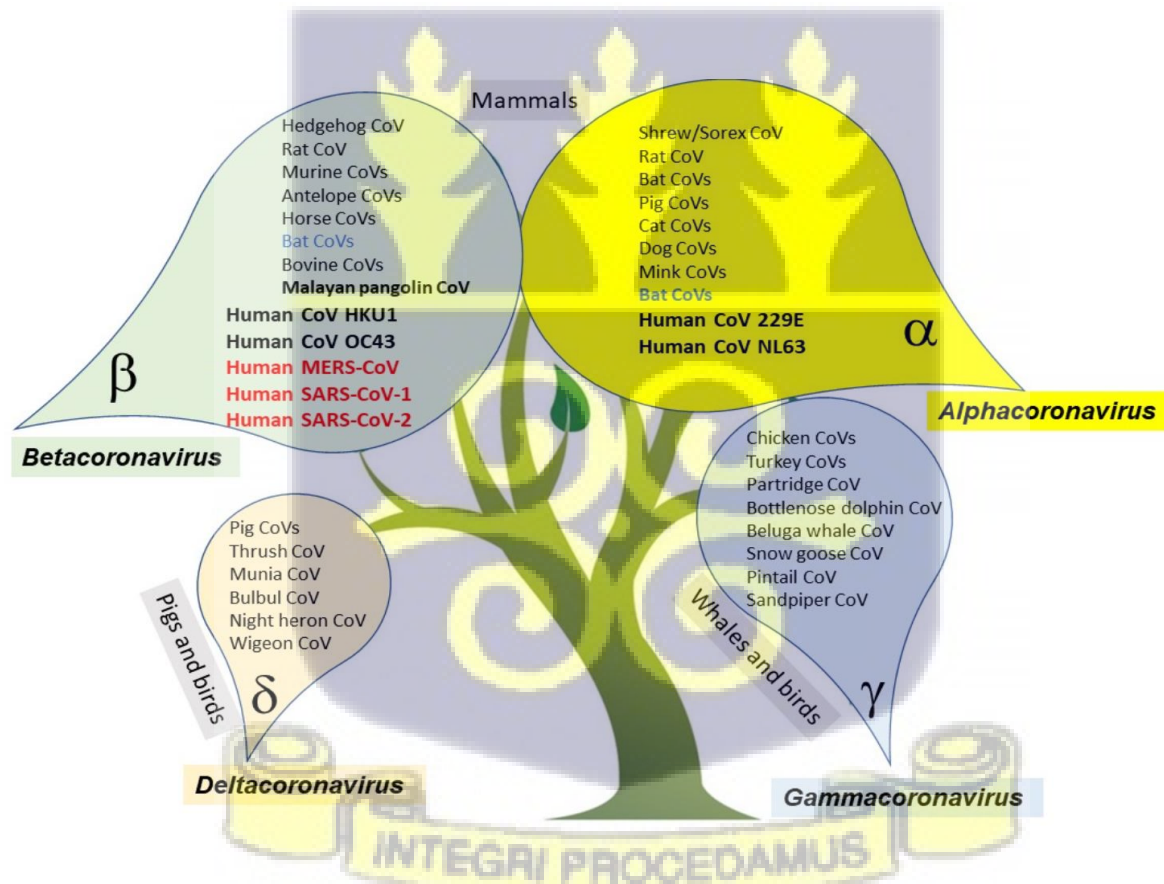
Coronaviruses (CoVs) are a group of viruses with the ability to infect a plethora of hosts globally and can be lethal (X. Li et al., 2019; Schoeman et al., 2022). Natural hosts of CoVs are often animals such as bats, cats, wild birds, pangolins, rodents, and camels (El-Sayed & Kamel, 2021; Z. Zhou et al., 2021). CoVs can survive in different hosts and tissues of the host due to their varying unique genetic flexibility as observed in their capacity to rapidly evolve, making them a menace to public welfare (El-Sayed & Kamel, 2021; Mulabbi et al., 2021). Since the first identification of CoVs in 1993 (McIntosh, 1974; Mulabbi et al., 2021), several CoVs have been identified. Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) is the most recently identified human coronavirus (HCoV) in December 2019 (Shi et al., 2020). Prior to this, HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, and Middle East Respiratory Syndrome-Related Coronavirus (MERS-CoV) were the only known HCoVs in circulation (Mulabbi et al., 2021; Nassar et al., 2021).

114

2.2 Taxonomy and classification

Currently, CoVs are classified into the order *Nidovirales* based on the presence of identical amino acids of hallmark genes such as the *RdRp* (RNA-dependent RNA polymerase) and *HEL1* (helicase of superfamily 1) (Z. Zhou et al., 2021). In the *Nidovirales* order, they belong to the *Coronaviridae* family and are further classified into *Orthocoronavirinae* and *Letovirinae* subfamilies (Woo et al.,

2023; Z. Zhou et al., 2021). Within the *Orthocoronavirinae* subfamily are the four genera; *Alphacoronavirus* and *Betacoronavirus* which infect humans, and *Gammacoronavirus* and *Deltacoronavirus* which infect birds primarily and some mammals (figure 1) (Woo et al., 2023; Z. Zhou et al., 2021). HCoV-NL63 and HCoV-229E are *Alphacoronaviruses* while HCoV-HKU1, HCoV-OC43, MERS-CoV, SARS-CoV, and SARS-CoV-2 are *Betacoronaviruses* (Schoeman et al., 2022; Woo et al., 2023). SARS-CoV-2, SARS-CoV and MERS-CoV are classified together due to their high pathogenicity; causing severe lower respiratory tract infection, while the other human coronaviruses are considered less pathogenic due to their moderate clinical presentation of 15% to 30% of common colds (X. Li et al., 2019).



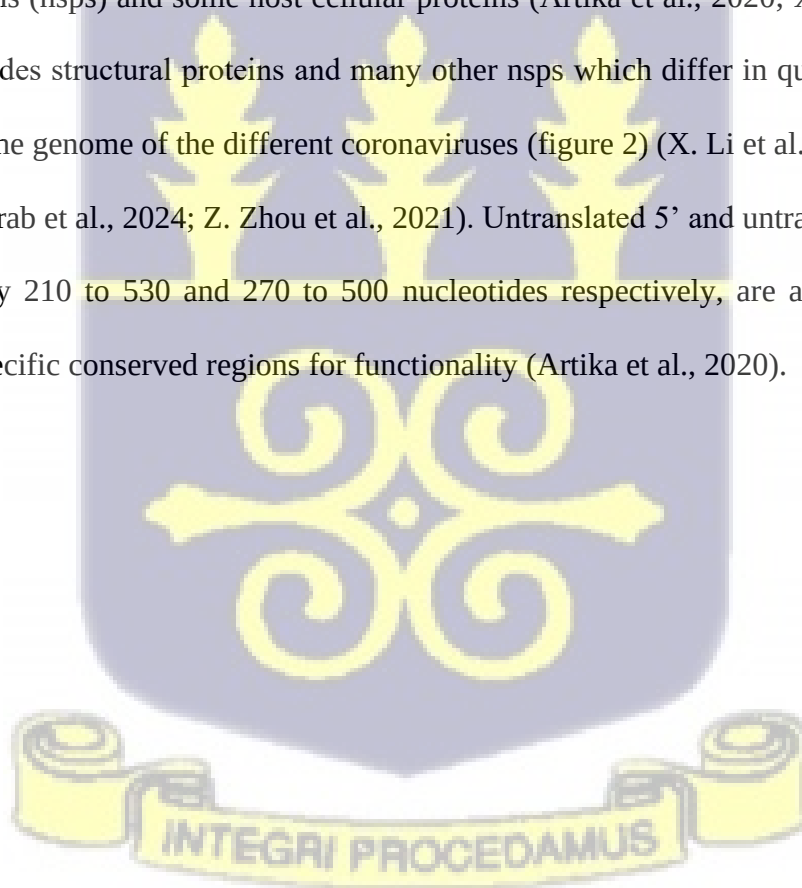
129

130 Figure 1: Orthocoronavirinae subfamily classification

131 *Human coronaviruses are shown in bold font and highly pathogenic human coronaviruses are shown in red font (Source: Shors, 2024)*

2.3 Genome organization

The genome of coronaviruses is approximately 26 to 32 kilobases and is currently the largest known for RNA viruses (X. Li et al., 2019; Schoeman et al., 2022). Coronaviruses have a single-stranded positive-sense RNA genome with a conserved order of a leader sequence and open reading frames at the 5' end which occupies about two-thirds of the genome and a polyadenylated 3' end occupying the other part of the genome, as well as other accessory genes distributed within (X. Li et al., 2019; Schoeman et al., 2022; Z. Zhou et al., 2021). There are two 5' open reading frames which together encode the replicase-transcriptase complex composed of 15 to 16 non-structural proteins (nsps) and some host cellular proteins (Artika et al., 2020; X. Li et al., 2019). The 3' end encodes structural proteins and many other nsps which differ in quantity, magnitude and position in the genome of the different coronaviruses (figure 2) (X. Li et al., 2019; Schoeman et al., 2022; Sohrab et al., 2024; Z. Zhou et al., 2021). Untranslated 5' and untranslated 3' regions of approximately 210 to 530 and 270 to 500 nucleotides respectively, are also present in the genome with specific conserved regions for functionality (Artika et al., 2020).



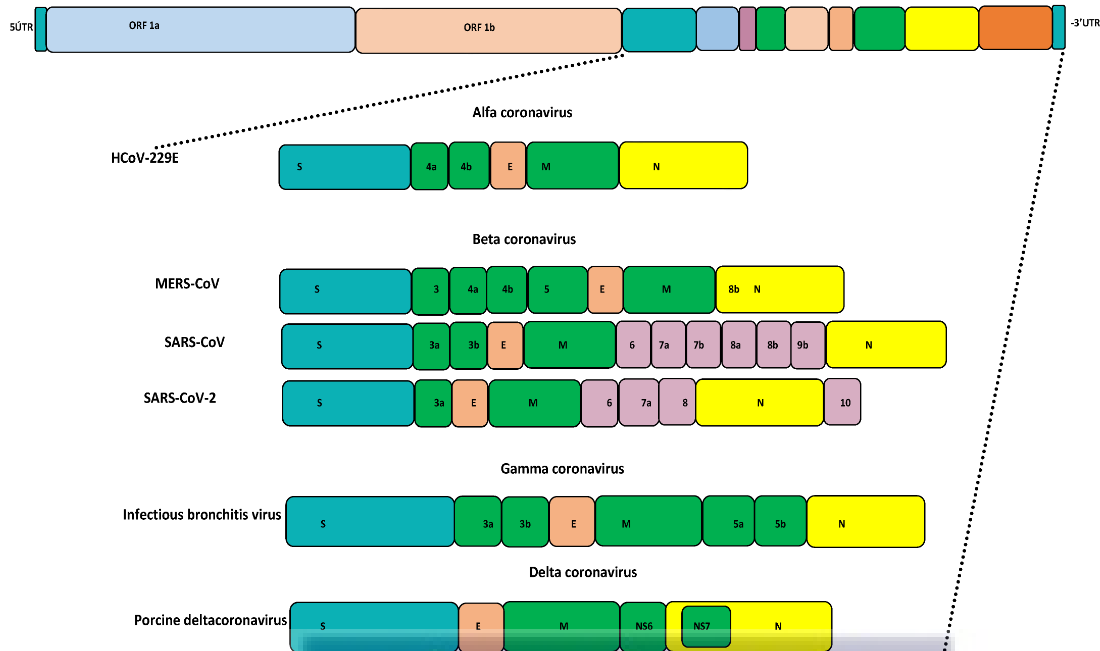


Figure 2: Genome organization of some coronaviruses

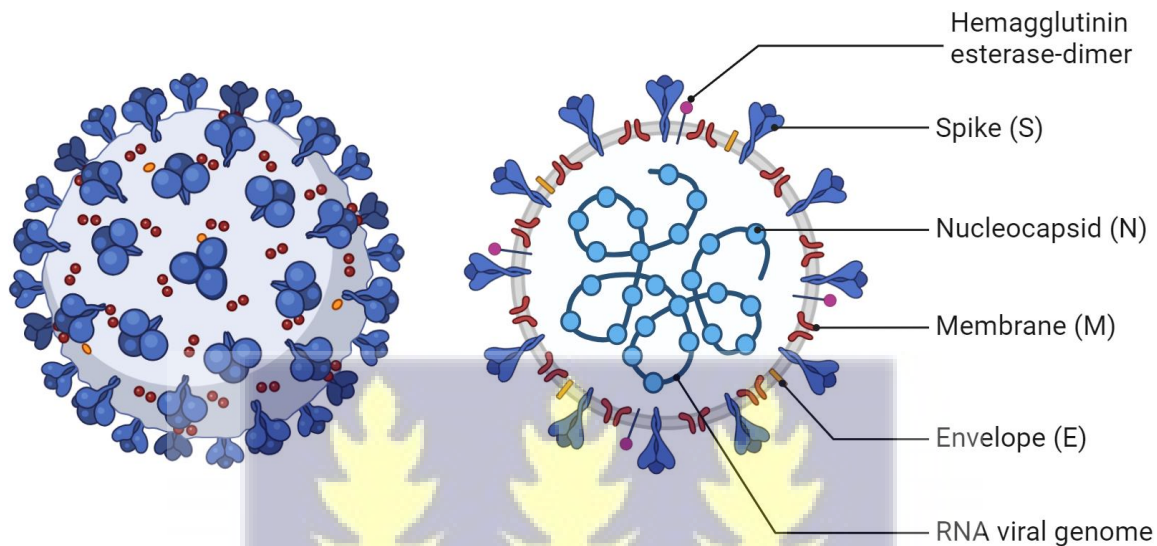
The 3' end of the genome differs significantly across the four coronavirus genera

Source: (Sohrab et al., 2024).

2.4 Structure of the virion

Coronaviruses are spherical pleomorphic viruses with club-like exterior extensions of their Spike protein which gives them a characteristic electron microscope image similar to the solar corona (Rohaim et al., 2021; Woo et al., 2023). They have a diameter range of approximately 80 to 120 nm with a large genome size (Artika et al., 2020; X. Li et al., 2019). Within the membrane of the virions are three or four (depending on the type of CoV) main structural proteins which are Spike, Membrane, Envelope, and Hemagglutinin esterase glycoprotein (figure 3) (specific to the

162 *Embecovirus* subgenus of the *Betacoronavirus* genus) (Florindo et al., 2020; Woo et al., 2023).
163 The Nucleocapsid is also a structural protein within the virion in a helix with the positive sense
164 RNA (Rohaim et al., 2021; Tortorici & Veessler, 2019).



165
166 Figure 3: Coronavirus virion structure

167 *The hemagglutinin esterase glycoprotein is specific to the Embecovirus subgenus*

168 *Adapted from BioRender templates (app.biorender.com/illustrations)*

169

170 2.4.1 Spike protein

171 The Spike protein is a member of class one viral fusion proteins and is essential for cellular
172 invasion by interacting with specific host receptors (X. Li et al., 2019; Tortorici & Veessler, 2019).
173 The Spike protein is essential in defining host specificity, and cellular and tissue tropism (Hulswit
174 et al., 2016; Tortorici & Veessler, 2019; Artika et al., 2020). It comprises three protomers of
175 approximately 1100 to 1600 amino acids, making it the largest transmembrane structural protein

of the virion (X. Li et al., 2019; Tortorici & Veerler, 2019). The Spike protein also plays an important role as immune cells of the host target it for the production of neutralization antibodies (Tortorici & Veerler, 2019). The Spike protein is made up of two subunits; S1 and S2 (Tang et al., 2023; Tortorici & Veerler, 2019). S1 interacts with host receptors influencing tropism and host specificity, while the S2 facilitates host and virus membrane fusion allowing cellular invasion (Tang et al., 2023). The S1 subunit also has the receptor binding domain (RBD) that determines the kind of host receptor used by the virus (X. Li et al., 2019). The different viral species have different RBD positions as seen in SARS-CoV and HCoV-229E at amino acid residues of 318 to 510 and 417 to 547 respectively (X. Li et al., 2019). Also, there may be a cleavage site recognizable by furin-like proteases between the S1 and S2 subunits (Hulswit et al., 2016) which affects the activation of the Spike protein and consequently virus-cellular invasion and infection (Bestle et al., 2020).

2.4.2 Nucleocapsid protein

The Nucleocapsid is an important structural protein that plays a role in genome packaging and replication of the virus (Dong et al., 2023). It has an N-terminal domain (NTD) and a C-terminal domain (CTD) and may undergo post-translational modifications such as phosphorylation and cleavage by proteases (Kim et al., 2022; Liu et al., 2021). Nucleocapsid binds to the viral RNA as dimers within the viral envelope with a characteristic helical symmetry (Kim et al., 2022; Liu et al., 2021). It also has a medial linker domain that is rich in Serine and Arginine of approximately 50 amino acid residues (Wong & Saier, 2021) which may regulate the function of the nucleocapsid (Satarker & Nampoothiri, 2020). The expression of Nucleocapsid is observed in specimens at the

primal stages of infection (Satarker & Nampoothiri, 2020). Other roles of Nucleocapsid include association with membrane protein for virion assembly, promotion of RNA strand separation to increase the efficiency of replication, prevention of the production of interferons, and enhancement of host apoptosis (Wong & Saier, 2021).

2.4.3 Membrane protein

The Membrane protein is approximately 25 to 30 kDa which primarily associates with the other structural proteins (Liu et al., 2021) and homotypic interaction with itself in mobilizing the virion assembly (Artika et al., 2020). The interaction between Membrane and Envelope leads to the formation of viral-like particles for assembly of the viral envelope (Samavati & Uhal, 2020). The Membrane protein spans the viral membrane three times, with its CTD interacting with Nucleocapsid and a short-length NTD at the external portion of the virion (Artika et al., 2020). Interestingly, there is another structural prediction showing both the NTD and CTD on the exterior portion of the virion (Schoeman et al., 2022). The membrane protein may also function in immune evasion using multiple strategies, such as enhancing apoptosis of infected cells and preventing NF κ B production to promote viral replication (Satarker & Nampoothiri, 2020).

2.4.4 Envelope protein

The envelope protein is the least abundant protein of the virion (Schoeman et al., 2022). Its interaction with the membrane protein constitutes the viral envelope (Schoeman et al., 2022). It has a short NTD, an α -helical transmembrane domain, and a CTD domain at the intracellular

portion of the virion (Schoeman et al., 2022; Wong & Saier, 2021). The envelope protein also plays a role in viral pathogenesis via the formation of ion-channel viroporins which change the perviousness of host cell membranes (Schoeman & Fielding, 2020). It may also play a role in maintaining the morphology of the virion, as studies have observed that mutant virions lacking the Envelope show poor structural virion formation, reduced viral plaques and viral titres (Wong & Saier, 2021).

2.4.5 Non-structural proteins

The non-structural proteins (nsps) differ in quantity, magnitude and position in the genome of the different coronaviruses and may provide the virus an added advantage for survival (Schoeman et al., 2022) such as immune evasion (Artika et al., 2020). Nsps are encoded by the 5' end of the genome consisting of two open reading frames (ORFs) 1a and 1b (Rohaim et al., 2021). ORF 1a and 1b are translated into polyproteins (pp) 1a and 1b and processed by its proteases to generate 15 to 16 nsps (X. Li et al., 2019). The interaction between nsps and the host cellular machinery can be explored in the design of therapeutics as they may play a vital function in replication (Rohaim et al., 2021).

2.5 Host interaction and viral entry

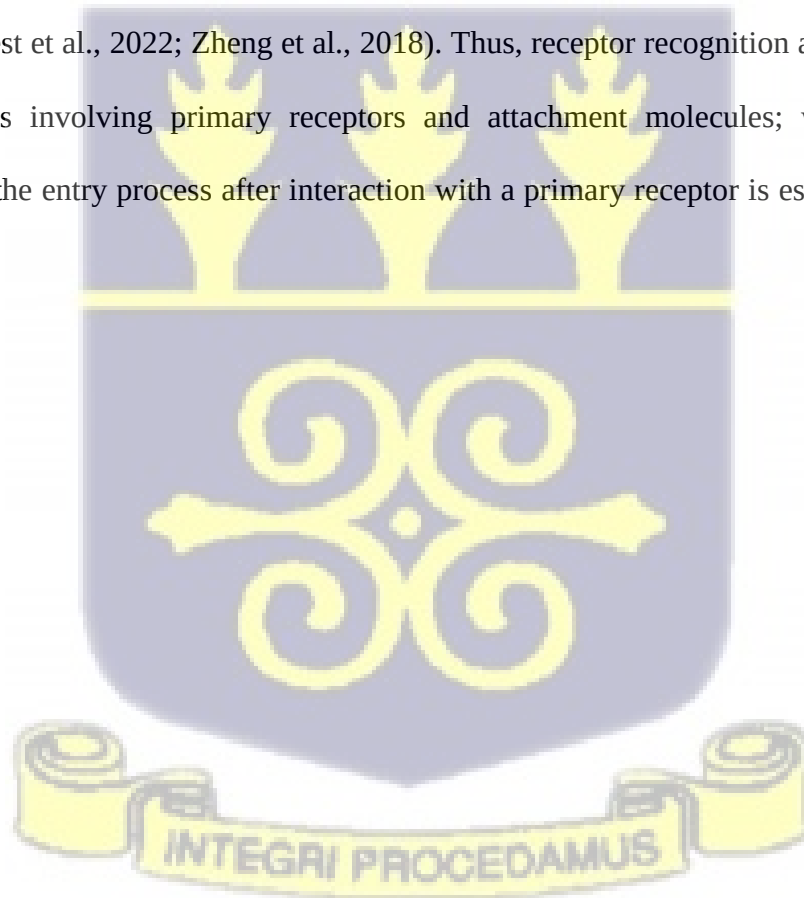
HCoV infection begins when an interaction occurs between host receptors and viral ligands leading to the release of the viral genome into the host cytoplasm (Artika et al., 2020). This interaction occurs between the glycosylated S1 portion of Spike protein and host receptors such as

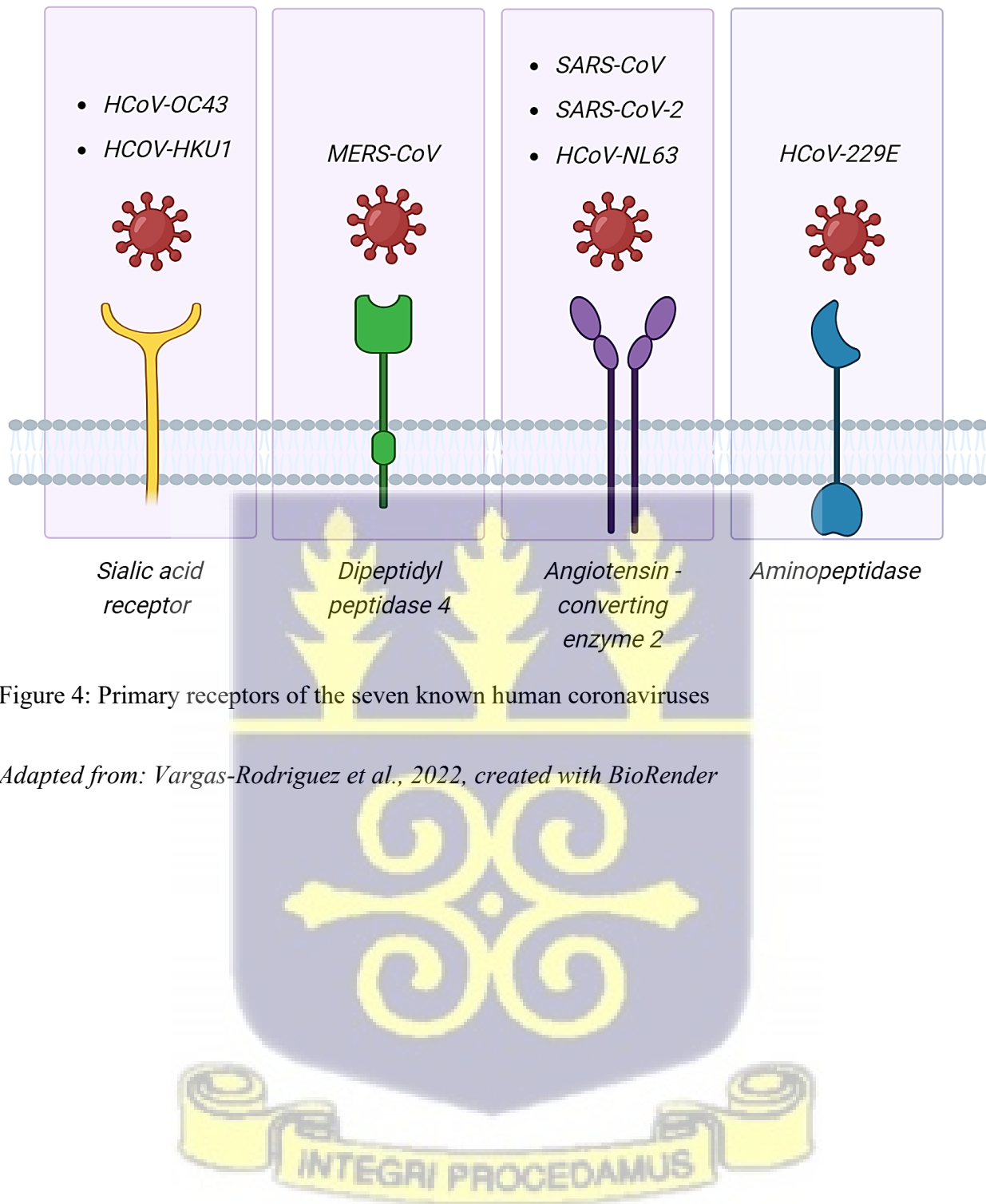
Angiotensin-Converting Enzyme 2 (ACE2) and Dipeptidyl Peptidase 4 (DPP4) (Artika et al., 2020; Everest et al., 2022). Upon recognition of host receptors, the Spike undergoes conformational changes and proteolysis at either both or one of its cleavage sites (the S1/S2 cleavage site and S2' cleavage site) (Hulswit et al., 2016). The host receptor involved in the internalization of the virus is dependent on the viral strain and may be carried out via endocytosis or mediated endocytosis by host receptors such as Transmembrane Serine Protease 2 (TMPRSS2) (Hashimoto et al., 2021; Nassar et al., 2021; Shi et al., 2020). A viral uncoating step occurs after entry, releasing the positive stranded RNA which encodes an RNA-dependent RNA polymerase (Fung & Liu, 2019; Nassar et al., 2021). The RNA-dependent RNA polymerase then utilizes the positive strand RNA as template to produce a negative sense strand, which is used as the template for new positive sense viral RNA (Fung & Liu, 2019; Nassar et al., 2021). A discontinuous transcription of the negative sense strand occurs, producing a nested group of sub-genomic RNA encoding structural and non-structural proteins (Fung & Liu, 2019; Schoeman et al., 2022). Packaging of the proteins then occurs in the endoplasmic reticulum and Golgi apparatus, producing mature virions through the secretory pathway or exocytosis (Fung & Liu, 2019; Schoeman et al., 2022). The replication cycle lasts approximately 10 to 12 hours in an infected cell (Nassar et al., 2021).

2.6 Receptors

Generally, viruses interact with receptors for cellular entry which may involve primary receptors and other modulatory molecules (Everest et al., 2022). The seven HCoVs use only four primary receptors with specific modulatory molecules such as host proteases for viral attachment and entry (Nassar et al., 2021). The four main receptors are Angiotensin-converting enzyme 2 (ACE2),

Dipeptidyl peptidase 4 (DPP4), Aminopeptidase (ANPEP), and 9-O-acetylated-sialic acid (9-O-Ac-Sia) (Nassar et al., 2021; Everest et al., 2022; Vargas-Rodriguez et al., 2022). The envelope-anchored Spike protein of HCoV binds to one of these host receptors causing a conformational change that leads to the fusion between the viral envelope and host cell membrane (W. Ni et al., 2020; Zheng et al., 2018). The Spike protein has S1 and S2 subunits which facilitate viral attachment and fusion, respectively (Kaidashev et al., 2021; Zheng et al., 2018). The function of the Spike subunits is facilitated by host proteases showing up at different viral attachment and entry steps depending on the type of HCoV (Zheng et al., 2018). The HCoV genomic RNA is then deposited into the host cell through endocytosis or membrane fusion post-cleavage by the host proteases (Everest et al., 2022; Zheng et al., 2018). Thus, receptor recognition and viral entry is a complex process involving primary receptors and attachment molecules; which ensure the continuation of the entry process after interaction with a primary receptor is established (Everest et al., 2022).





2.7 Coronavirus disease 2019

2.7.1 Coronavirus disease 2019 epidemiology

Coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) emerged in Wuhan hospitals as an unusual pneumonia in December 2019 (Boban, 2021; Kumar et al., 2021; Meskini et al., 2021). SARS-CoV-2 was identified to have sequence similarity of approximately 80% to SARS-CoV (Kumar et al., 2021; Meskini et al., 2021). Globally, according to the World Health Organization, over 776 million COVID-19 infected persons out of which 7,065,880 million people have lost their lives (assessed 30th September, 2024 (WHO, 2024)). Interestingly, the West African region is one of the less impacted regions compared to other regions like America, Western Pacific, and Europe (figure 5). Many studies have proposed possible reasons for the low number of cases particularly for the African population (Lawal, 2021; Oduro-Mensah et al., 2022; Tsinda & Mmbando, 2021).

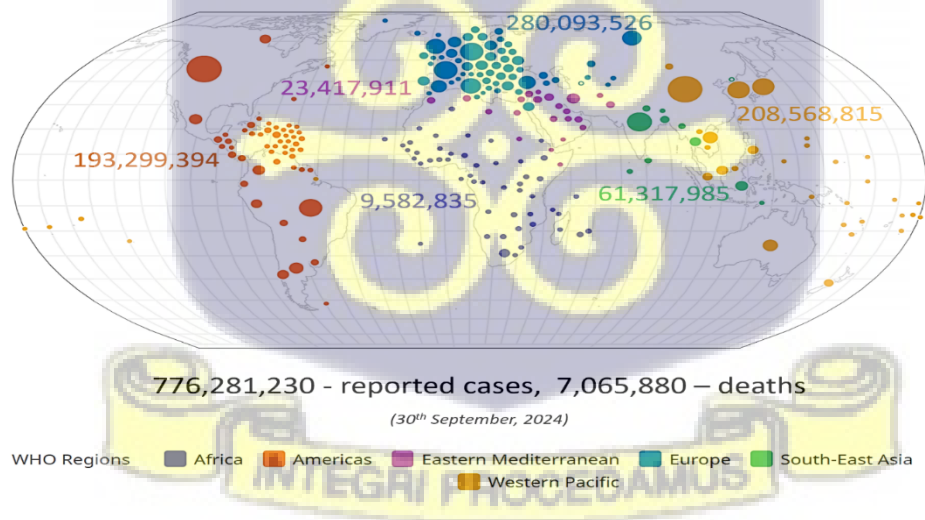


Figure 5: Total number of reported cases retrieved from the WHO dashboard on 30th September, 2024

2.7.2 *The unequal distribution of COVID-19 pandemic*

The COVID-19 pandemic affected countries differently (figure 5). Many plausible hypotheses have emerged to explain the disparity in the number of cases and fatalities observed particularly in West Africa (Njenga et al., 2020; Lawal, 2021; Oduro-Mensah et al., 2022). Early predictions made by experts during the pandemic suggested that Africa will record a high number of COVID-19 infections and related deaths (Oduro-Mensah et al., 2022; Taboe et al., 2020). However, this was not observed especially for West Africa as cases reported were largely mild or asymptomatic (Quashie et al., 2021; Jacobs & Okeke, 2022; Tapela et al., 2024). Some factors that suggest Africa would have been greatly affected by the COVID-19 pandemic include poverty, poor health facilities and personnel, low surveillance and poor diagnostic tools (Njenga et al., 2020; Taboe et al., 2020). Currently, multiple hypotheses; such as exposure to malaria, immune modulation and host genetics have emerged to partly explain the disparity observed with a focus on the African continent (Duah-Quashie et al., 2024; Sandi et al., 2024; Tapela et al., 2024).

2.7.2 *Routes of infection*

COVID-19 spread occurs primarily via viral transmission through inhalation of respiratory aerosols from an infected person or contact with infected surfaces of fomites (Boban, 2021; Kumar et al., 2021). Other transmission routes of infection include air, mother-to-child, sexual intercourse, faecal matter and discharges from the conjunctiva (Jain et al., 2022; Kumar et al., 2021). Transmission through the air is facilitated by the viability of the virus in the air, which is approximately 3 hours to 3 days and necessitating the need for face masks during the pandemic (Kumar et al., 2021; Zhao et al., 2022). Transmission through pregnancy is not fully established

due to conflicting data; some studies identified SARS-CoV-2 in babies between 1 and 1½ days after delivery, in breastmilk, and placenta, and high levels of immunoglobulin M and G in a baby 120 minutes after caesarean birth (Kumar et al., 2021). Despite these results, there is no conclusive indication of vertical transmission between a mother and a child (Kotlyar et al., 2021; Kumar et al., 2021; Yuan et al., 2021). Transmission through sex has also not been fully established although there is evidence of SARS-CoV-2 in semen samples (Kumar et al., 2021). Nevertheless, the possibility of spread via sexual intercourse can be inferred due to the expression of the ACE2 receptor at the points of intercourse (Aitken, 2021; Kumar et al., 2021). Transmission through the orofecal interface and conjunctiva secretions has also been proposed due to the identification of SARS-CoV-2 in faecal matter and tears of infected persons (Heller et al., 2020; Kumar et al., 2021).

2.7.3 Clinical presentation

The clinical presentation of COVID-19 disease varies from person to person based on an interplay between factors such as age, gender, underlying conditions, genetic variation of host and virus and environmental and pathogen exposures (Niemi et al., 2022; Oduro-Mensah et al., 2022; Pradhan et al., 2021). On average, the viral incubation phase after cellular invasion is between 2 to 14 days (Boban, 2021; Mehta et al., 2021), after which symptoms may or may not show. Thus, the clinical signs of COVID-19 can vary between asymptomatic, mild, moderate, and severe or critical categories (Mehta et al., 2021). The virus mainly damages the upper respiratory tract upon infection leading to replication in the infected cells, after which it migrates to the lungs and causes damage (Mehta et al., 2021; Melenotte et al., 2020). About 3% of symptomatic patients progress

to fatal conditions (Marik et al., 2021) involving atypical humoral metabolic processes and pulmonic oedema (Cui et al., 2021). Symptoms of mild COVID-19 involve a plethora of non-specific presentations such as cough, fever, fatigue, headache and pain in the muscles and joints (Schoeman et al., 2022) while severe disease involves shortness of breath, high levels of inflammatory markers, septic shock and multiorgan dysfunction (Lamers & Haagmans, 2022; Boban, 2021; Mehta et al., 2021). Four main phases of clinical progression from mild to critical disease have been described; the first clinical signs occur in the upper and or the lower respiratory tract which manifests as nasal congestion, sneezing, partial or complete loss of smell and loss of taste (Melenotte et al., 2020; Polidoro et al., 2020). Phase two involves infection of the lower respiratory tract leading to lung injury requiring hospitalization (Melenotte et al., 2020; Polidoro et al., 2020). Phase three usually begins from the ninth to the twelfth day (Melenotte et al., 2020) indicated by an abrupt failure of the respiratory system occurring due to a cytokine storm resulting in thrombosis (Melenotte et al., 2020; Polidoro et al., 2020). The fourth phase is characterized by severe distress of the respiratory airway, multiple organ failure and death (Melenotte et al., 2020).

2.7.4 Diagnosis

The gold standard for SARS-CoV-2 detection is reverse transcription polymerase chain reaction (RT-PCR), which is based on the detection of parts of the viral sequence via reverse transcription and amplification (Yüce et al., 2021). The quantification of immunoglobulins in the blood or sputum was also carried out during the pandemic using lateral flow devices and enzyme-linked immunosorbent-based assays indicating the dynamics of immune response to the infection (Tapela et al., 2023; Yüce et al., 2021). Antigen detection kits that detect the presence of the virus without

the need for PCR were also developed and used during the pandemic. These also utilized lateral flow devices and enzyme-linked immunosorbent-based assays (Humbert et al., 2022; Yüce et al., 2021). Other detection platforms like smartphone sensors, biosensors, reverse transcription loop-mediated isothermal amplification (RT-LAMP), nucleic acid hybridization assays, surface plasmon resonance-based assays, CRISPER Cas 12-based lateral flow devices, and genomic analysis using the nanopore technology platform emerged as alternate forms of immunoglobulin and antigen identification (figure 6) (Irkham et al., 2022; Yüce et al., 2021). All of these diagnostic tools offer different levels of sensitivity and specificity on different types of samples (Melenotte et al., 2020).

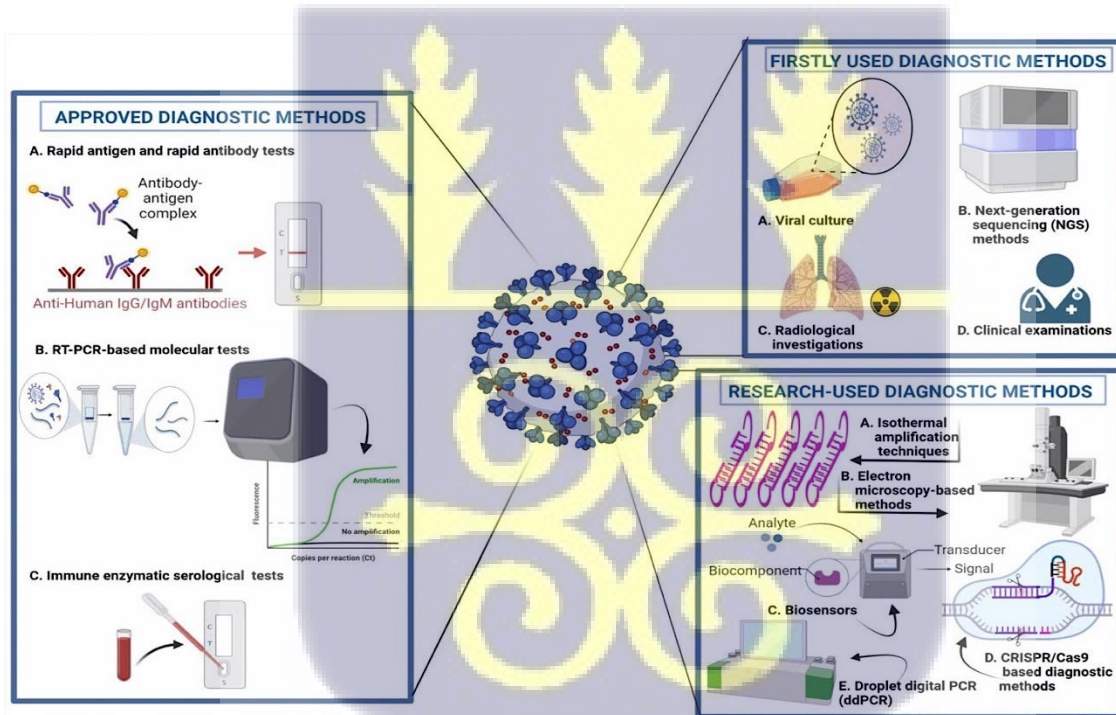


Figure 6: Molecular diagnostic methods for COVID-19

1. Approved Diagnostic Methods, 2. First line of Diagnostic tools, 3. Research-used Diagnostic Methods. Source: (Irkham et al., 2022)

2.7.5 Immune response

The innate immune response is activated upon SARS-CoV-2 infection in a host causing a general defense by proteins such as defensins, metabolites, cytokines and mucins at the epithelial surface to prevent viral replication and spread (Mistry et al., 2022; Smith et al., 2021). As the disease progresses as a result of insufficient general defense, host pattern recognition receptors (PRRs) expressed by innate immune cells detect damage-associated molecular patterns (DAMPs) and pathogen-associated molecular patterns (PAMPs), leading to the production of more defense proteins (Mistry et al., 2022; Tirelli et al., 2023). These defense proteins include type I and III interferon, tumor necrosis factor alpha, and pro-inflammatory cytokines which hinder viral replication (Mohammed et al., 2022; Tirelli et al., 2023). Subsequently, natural killer cells are also activated by the cytokines produced, which also mount a defense against SARS-CoV-2 by degranulation or apoptosis (Mistry et al., 2022).

The adaptive immune response to SARS-CoV-2 is activated a number of days after infection by innate immune molecules to enhance viral clearance (Mistry et al., 2022). Antigen presenting cells such as monocytes, dendritic cells and macrophages process the virus and present them to cluster of differentiation (CD) 4 positive T (CD4⁺) cells for activation and differentiation (Kaklamanos et al., 2021). The CD4⁺ T cells differentiate into T follicular helper cells, interacting with the B lymphocytes to allow differentiation into antibody producing plasma cells or memory B cells targeting surface glycoproteins of the virus (Kaklamanos et al., 2021; Primorac et al., 2022; Silva et al., 2022). Other CD4⁺ T cells which are also produced after differentiation further enhance the adaptive response by providing more signal to enhance B lymphocyte maturation and bridging the interaction between antigen presenting cells and CD8⁺ T cells (Primorac et al., 2022).

Immunoglobulin (Ig) A, IgM, and IgG are successively produced by the plasma cells which enhance viral neutralization (Kaklamanos et al., 2021; Paces et al., 2020). Cytotoxic CD8⁺ T cells upon activation and differentiation, secrete multiple cytokines such as Tumor necrosis factor alpha, and interferon gamma to kill the virus (Mohammed et al., 2022; Primorac et al., 2022).

Despite the host's attempt to prevent viral spread, the virus adopts immune evasion mechanisms to enhance its spread and survival (Minkoff & tenOever, 2023). Some immune evasion mechanisms employed by SARS-CoV-2 include:

1. Inhibiting mitochondrial physiology which interferes with optimal interferon production (Burtscher et al., 2020; Rubio-Casillas et al., 2022).
2. Creation of tunneled nanotubes within the cytoplasm of cells which allows cellular communication between adjacent cells and infection of cells without the known permissible receptors for SARS-CoV-2, preventing extracellular immune recognition (Merolli et al., 2022; Pepe et al., 2022; Rubio-Casillas et al., 2022).
3. Downregulation of the human leukocyte antigen I (HLA-1) molecules required for antigen presentation by the interaction of viral proteins from the 6th and 8th open reading frames of the viral genome (Yoo et al., 2021; Zhang et al., 2021; Rubio-Casillas et al., 2022).
4. The incorporation of SARS-CoV-2 in exosomes which prevent recognition by antibodies, allow cellular invasion of cells without permissible receptors for SARS-CoV-2, and transport of envelope proteins serving as false targets for immune recognition (Troyer et al., 2021; Rubio-Casillas et al., 2022).

2.8 SARS-CoV-2 receptors

2.8.1 Primary Receptors

2.8.1.A Angiotensin-Converting Enzyme 2

Angiotensin-Converting Enzyme 2 (ACE2) is a homolog of ACE, with a diverse set of biological functions expressed in a variety of human organs and tissues (Bosso et al., 2020; W. Ni et al., 2020). ACE2 is a carboxypeptidase and zinc metalloenzyme with one of its main functions being the regulation of circulating angiotensin II levels to maintain blood pressure homeostasis (Bosso et al., 2020). ACE2 has been identified as the receptor for the *betacoronaviruses*; SARS-CoV and SARS-CoV-2 (figure 7), and the *alphacoronavirus* HCoV-NL63 (Everest et al., 2022). Several polymorphisms in the *ACE2* gene that may be linked to severity or susceptibility to COVID-19 in specific populations have been identified (Jalaleddine et al., 2022; Pišlar et al., 2020; Srivastava et al., 2020). For example, the R514G polymorphism has been found in the African/American population which may be associated with susceptibility to COVID-19 and cardiovascular difficulties, I468V has also been found in Asia which leads to an alteration in ACE2 secondary structure that reduces ACE2 function, while K26R, T27A, H34R, and many other variants have also been associated with increased susceptibility (Kaidashev et al., 2021).

Due to the diverse physiological function of ACE2 particularly in hypertension which is associated with increased mortality and severity of COVID-19, identifying polymorphic forms of ACE2 in the West African population will be of great importance, not only in understanding viral entry for SARS-CoV-2 but for other pathogens such as SARS-CoV and HCoV-NL63 and subsequently, aid in the development of therapeutic interventions for West Africans.

2.8.1.B Dipeptidyl peptidase 4

Dipeptidyl peptidase 4 (DPP4) also known as cluster of differentiation 26 (CD 26), is a dimeric type II transmembrane protein and the third protein identified as a receptor for HCoV-229E (Raj et al., 2013). DPP4 plays a role in a wide range of cellular processes such as glucose metabolism, cellular adhesion, regulation of chemotaxis, and regulation of hormones and chemokines (Raj et al., 2013). MERS-CoV interacts with DPP4 at the C-terminal domain for cellular entry (Abbad et al., 2021; Bosch et al., 2014; Everest et al., 2022). DPP4 also interacts with the gp120 envelope of the Human Immunodeficiency Virus by a hyper-sialylation process (Bassendine et al., 2020; Sebastián-Martín et al., 2022). Current data shows a possible interaction between DPP4 and SARS-CoV-2 for cellular invasion causing COVID-19 (figure 7) (Bassendine et al., 2020).

2.8.1.C Aminopeptidase

Aminopeptidase (ANPEP), also described as cluster of differentiation 13 (CD 13), is a type 2 transmembrane metalloprotease that exists in both soluble and insoluble forms (Everest et al., 2022). ANPEP acts as a primary receptor for the human coronavirus 229E and other human viruses (Bosch et al., 2014; Everest et al., 2022). It is also been recently proposed to be one of the receptors for SARS-CoV-2 entry (figure 7) (Qi et al., 2020).

2.8.1.D Basigin

The immunoglobulin family of proteins includes Basigin (basic immunoglobulin - BSG) or cluster of differentiation 147 (CD 147) or extracellular matrix metalloproteinase inducer (EMMPRIN)

(Chen et al., 2005; Everest et al., 2022). BSG plays several roles including *Plasmodium falciparum* cell entry, tumour development, and viral infection (Everest et al., 2022; Shilts et al., 2021). BSG in COVID-19, has been proposed as a receptor for the entry of SARS-CoV-2, however, this has been challenged (Ahmetaj-Shala et al., 2020; Shilts et al., 2021). Chen *et al.* proposed the link between BSG and SARS-CoV-2 to be via mediating cyclophilin A, complexed to the nucleocapsid protein of the virus (Chen et al., 2005). Conversely, after the commencement of clinical trial for meplazumab (humanized monoclonal antibody against BSG), Shilts *et al.*, did not find any evidence to support the invasion of cells via BSG receptor interaction (Shilts et al., 2021).

2.8.2 Attachment or secondary receptors

2.8.2.A Transmembrane Serine Protease 2 (TMPRSS2)

The Spike protein which performs attachment and fusion to the host receptor is cleaved by the Transmembrane Serine Protease 2 (TMPRSS2) to make available the S2 subunit for viral fusion (Everest et al., 2022; Wulandari et al., 2021). TMPRSS2 is a type 2 transmembrane serine protease that performs many functions including proteolytic cleavage activation of membrane proteins and enhancing viral invasion (Everest et al., 2022; Fraser et al., 2022). Activation of TMPRSS2 occurs via an autocleavage process facilitating the endocytosis of MERS-CoV, SARS-CoV, and SARS-CoV-2 in the host (figure 7) (Everest et al., 2022; Wulandari et al., 2021). Some studies have identified SNPs in TMPRSS2 which enhance or decrease viral entry (Wulandari et al., 2021; Rokni et al., 2022; Yaghoobi et al., 2023). These variants of TMPRSS2: V160M, G181R, R240C, G259S, P35L, and G432A, have been identified to cause amino acid substitutions (Wulandari et al., 2021) which may lead to alteration in function.

478 2.8.2.B Cathepsin L and Cathepsin B (CTSL/CTSB)

479 Cathepsins are proteases that recycle cellular proteins within lysosomes at a pH of 4.5 to 5.0 (Gomes
480 et al., 2020). Cathepsin L and B are cysteine cathepsins that are delivered to the endosome for viral
481 entry (Pišlar et al., 2020). They play a role in physiological processes including the presentation of
482 antigens, remodelling of the extracellular matrix, and apoptosis (Gomes et al., 2020). For SARS-
483 CoV and SARS-CoV-2 infection, one of the mechanisms of viral entry is endocytosis and this is
484 mediated by the cleavage of the Spike protein by cathepsin L (figure 7) (Gomes et al., 2020; Pišlar
485 et al., 2020). Thus, when clathrin-mediated endocytosis occurs as a result of low expression of
486 TMPRSS2, cathepsins undergo Spike cleavage for viral internalization (Jackson et al., 2022).
487 Cathepsin B (CTSB) has also been identified to be involved in facilitating the entry of SARS-CoV-
488 2 (figure 7) (Hashimoto et al., 2021).



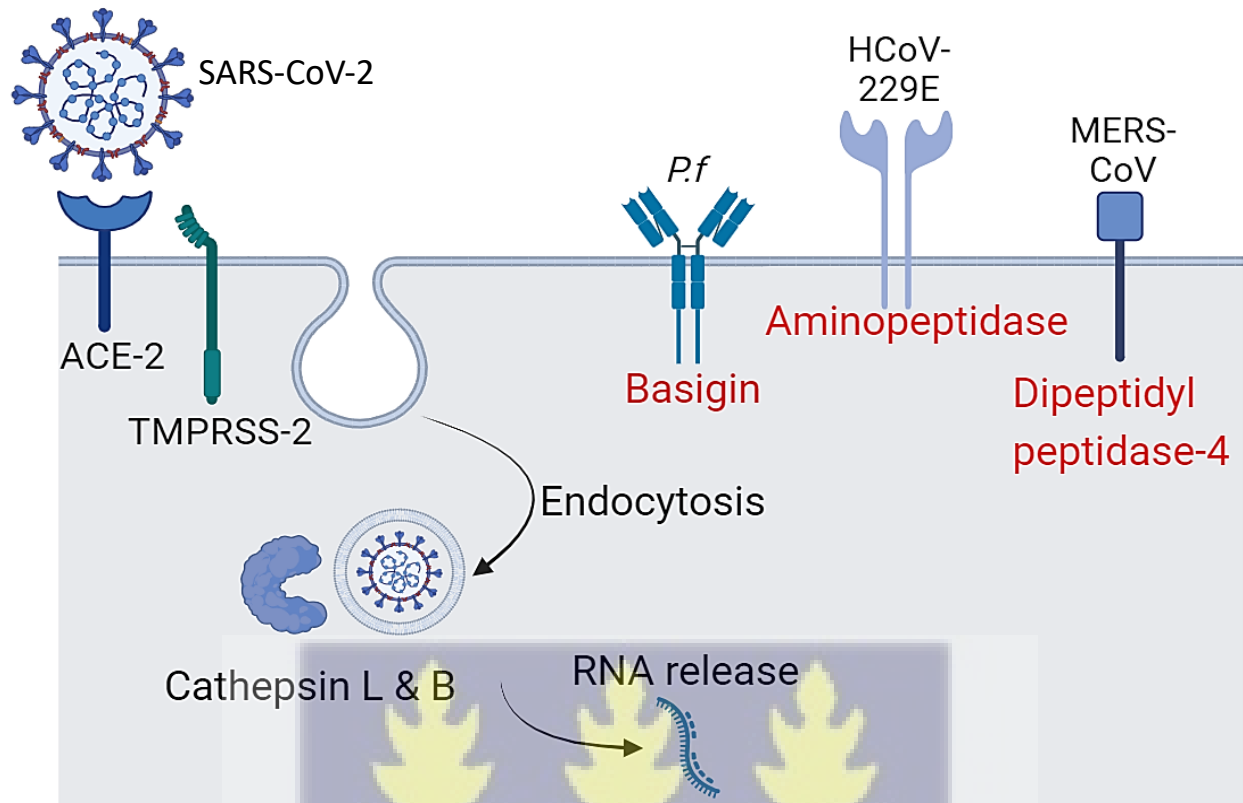


Figure 7: SARS-CoV-2 cellular invasion

Other proposed receptors which may facilitate SARS-CoV-2 cellular invasion are shown in red font. (Created with BioRender).

SARS- CoV-2 - Severe Acute Respiratory Syndrome Coronavirus – 2, ACE2 – Angiotensin Converting Enzyme 2, TMPRSS2 – Transmembrane Serine Protease-2



CHAPTER THREE

3.0 METHODOLOGY

3.1 Study design

Archived respiratory samples from Ghana, Benin, Guinea and Nigeria obtained between 2020 and 2022 during the COVID-19 pandemic; stored at -20 degrees, were used for the study. The Ghanaian samples were sequenced for the identification of SARS-CoV-2 as part of a previous study (Morang'a et al., 2022). Samples from the other countries were obtained as part of a collaborative effort between WACCBIP (West African Centre for Cell Biology of Infectious Pathogens) and other five WANIDA (West African Network of Infectious Diseases African Centres of Excellence) partners (table 1).

Table 1: Total number of samples submitted for each WANIDA centre

WANIDA Partner	Country	Total number of samples submitted
CEA-PCMT	Guinea Conakry	40
ACEGID	Nigeria	40
ACEPHAP	Nigeria	40
ACENTDFB	Nigeria	40
ACEMFS	Nigeria	40
IRCB/CERPAGE	Benin	60
CEA-ITECH/MTV	Burkina Faso	0
WACCBIP	Ghana	70
Total	-	330

Key:

1. CEA-PCMT : Centre d'Excellence Africain pour le Prevention et le Contrôle des Maladies Transmissibles
2. ACEGID: African Centre of Excellence for the Genomics of Infectious Diseases
3. ACEPHAP: African Centre of Excellence for Population Health and Policy
4. ACENTDFB: African Centre of Excellence for Neglected Tropical Diseases and Forensic Biotechnology
5. ACEMFS: African Centre of Excellence for Mycotoxins and Food Safety
6. IRCB/CERPAGE: Institut de Recherche Clinique du Benin/Centre d'Etude et de Recherche sur les Pathologies Associées à la Grossesse et à l'Enfance

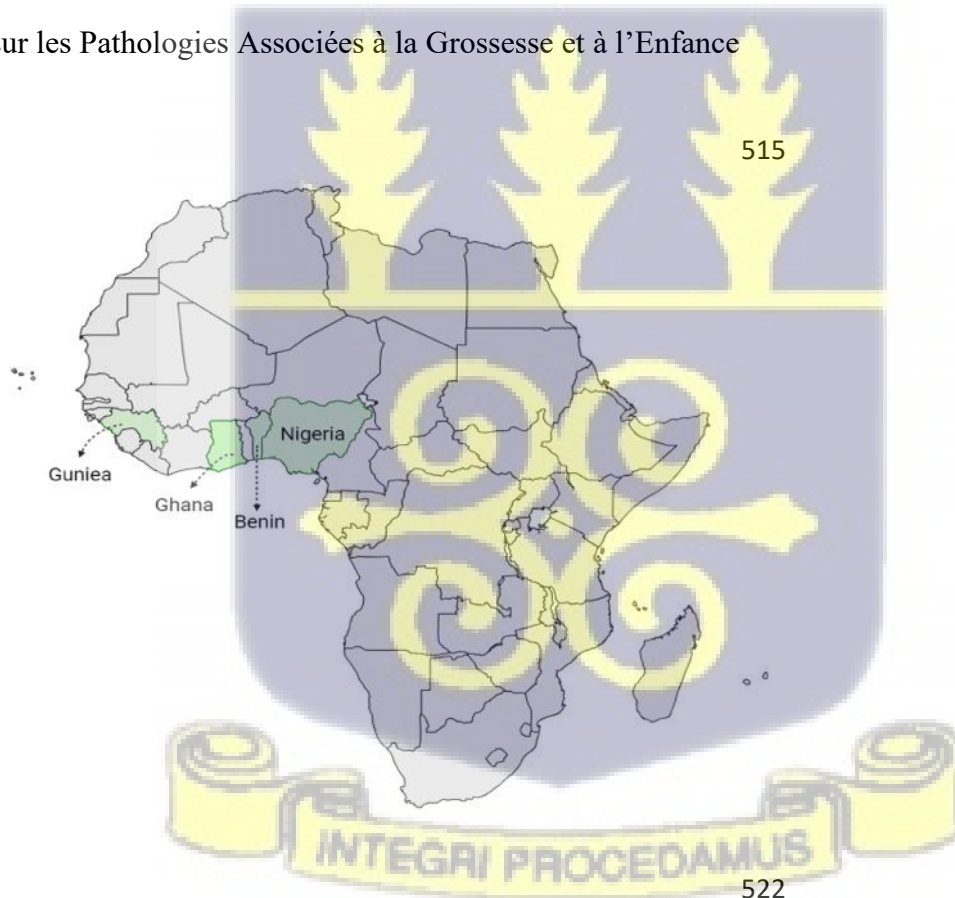


Figure 8: African map showing the selected West African countries

3.2 Inclusion criteria and sample characteristics

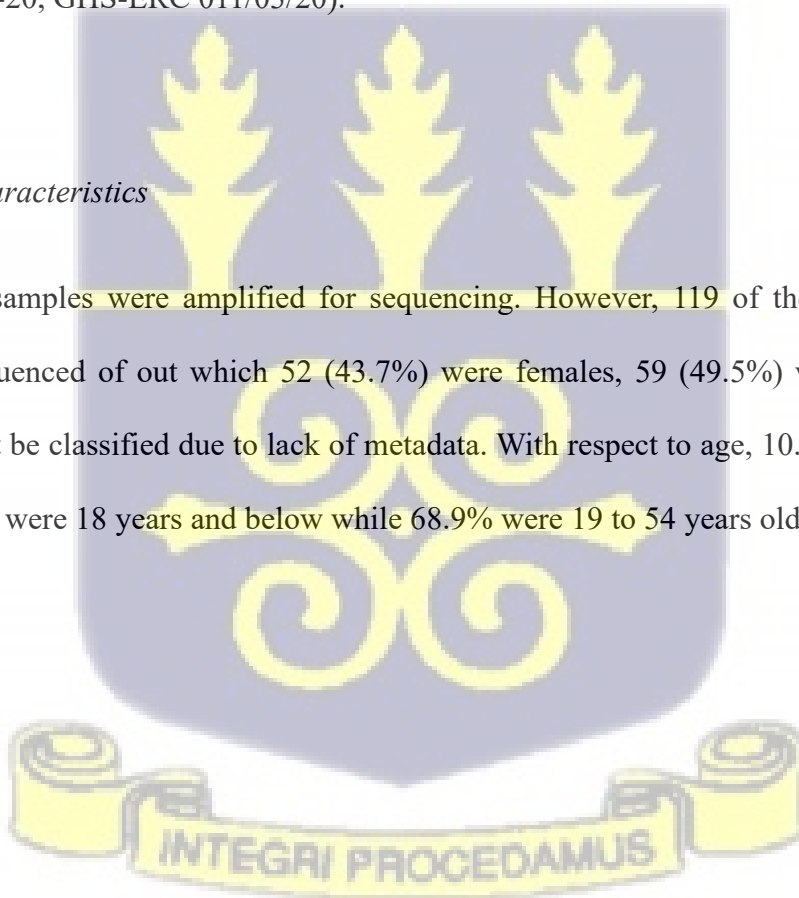
Archived samples from participants who had consented to future studies at the point of recruitment were used in this study.

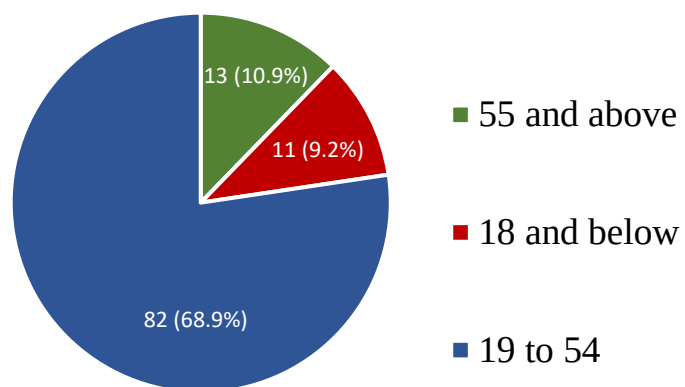
3.2.1 Ethical approval

Ethical clearance had already been obtained from the Ethics Board of the College of Basic and Applied Sciences, University of Ghana, and the Ethics Committee of the Ghana Health Service (ECBAS 063/19-20, GHS-ERC 011/03/20).

3.2.2 Sample characteristics

A total of 330 samples were amplified for sequencing. However, 119 of these samples were successfully sequenced of out which 52 (43.7%) were females, 59 (49.5%) were males and 8 (6.7%) could not be classified due to lack of metadata. With respect to age, 10.9% were 55 years and above, 9.2% were 18 years and below while 68.9% were 19 to 54 years old (figure 9).

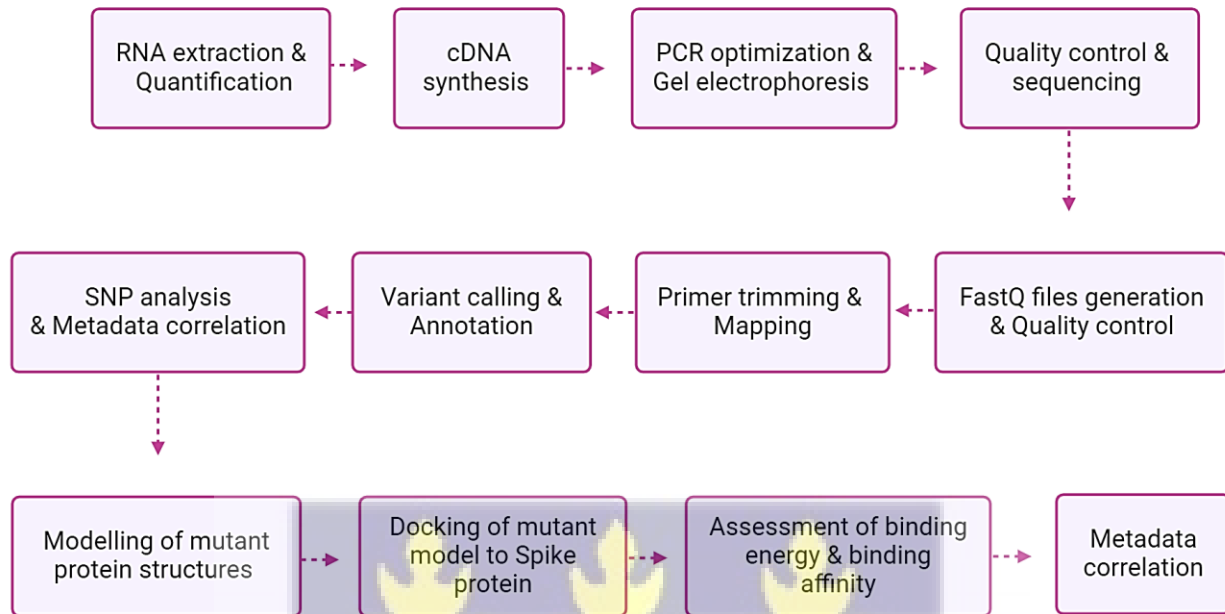




540 Figure 9: Age classification of samples sequenced



3.3 General workflow for targeted sequencing



542

543 Figure 10: General workflow for targeted sequencing

544 *RNA was extracted to exclude intronic sequences; full length ACE2 has over 89 million bases while*
 545 *approximately 3 thousand bases encode the exons (appendix 3).*

546 3.3.1 RNA extraction

547 RNA extraction was carried out with the Zymogen RNA MiniPrep Plus kit (Catalogue number:
 548 R1058) as instructed by the manufacturer with a few modifications. Approximately, 140 μ L of
 549 sputum or viral transport medium containing oropharyngeal or nasal swab was transferred into a
 550 1.5 mL tube. A volume of 15 μ L of proteinase K and 30 μ L of proteinase K digestion buffer were
 551 added to the tube. These were incubated at room temperature (RT) for 40 minutes, vortexed and
 552 centrifuged at 16,000 x g for 2 minutes. Approximately 185 μ L of RNA lysis buffer was added,
 553 vortexed and centrifuged for 2 minutes at 16,000 x g. The lysed samples were then transferred into

the yellow spin-away filter in a collection tube, and centrifuged at 16,000 x g for 2 minutes to eliminate the majority of genomic DNA. The flow through was transferred into a 1.5 mL and about 370 μ L of 99.9% ethanol was added, vortexed and centrifuged for 16,000 x g for 1 minute. The mixture was transferred into the green zymo-spin IIICG column in a collection tube and centrifuged at 16,000 x g for 30 seconds and the flow-through was discarded. The column was washed with 400 μ L of RNA wash buffer, centrifuged at 16,000 x g for 1 minute, and flow-through discarded. DNase I mix was prepared in a nuclease-free tube by adding 5 μ L DNase I (1 U/ μ L) to 75 μ L DNA digestion buffer per sample and 80 μ L of the mixture was added directly unto the column and incubated at RT for 15 minutes. About 400 μ L of RNA prep buffer was then added to the column, centrifuged at 16,000 x g for 1 minute and flow-through discarded. The column was then washed with 700 μ L of RNA wash buffer and centrifuged for 16,000 x g for 2 minutes. The column was transferred into a new 1.5 mL nuclease-free tube, and 40 μ L of nuclease-free water was added to elute RNA, incubated for 1 minute, and centrifuged at 10,000 x g for 1 minute. The RNA was quantified, prepared for complementary DNA (cDNA) synthesis and stored at -80°C.

3.3.2 Complementary DNA synthesis

Complementary DNA (cDNA) was synthesized using Protoscript® II first strand cDNA synthesis kit (catalogue number: E6560L) following the standard protocol. About 2 μ L of RNA was transferred into a 0.2 mL PCR tube, 2 μ L of d(T)₂₃ VN and 4 μ L of nuclease-free water was added. The mixture was incubated at 65°C for 5 minutes, centrifuged briefly and promptly placed on ice. Approximately 10 μ L of ProtoScript II Reaction Mix (2X) and 2 μ L of ProtoScript II Enzyme Mix (10X) were added to the mixture. The mixture was then incubated at 42°C for one hour, and at 80°C to synthesize the DNA and inactivate the enzyme, respectively.

3.3.3 Primer reconstitution and Polymerase Chain Reaction

PCR optimization (Lorenz, 2012) was carried out using 12 previously obtained overlapping primers targeting *ACE2* (figure 11). Lyophilised primers were reconstituted using nuclease-free water following the manufacturer's protocol (appendix 1) to obtain a 100 μ M concentration, after 1:10 dilutions were prepared for each primer and stored at -20°C. Polymerase Chain Reaction (PCR) was carried out using the cDNA samples synthesized as template. The workflow followed the manufacturer's instruction of the Q5 Hot Start high fidelity 2x master mix (catalogue number: M0494L), the GoTaq® G2 Hot Start master mix (catalogue number: 0000264235) (table 2) and the One Taq® Quick-Load quick-load 2X master mix with standard buffer (catalogue number: M0486S) (table 4), with a few modifications. Optimization steps involving varying annealing, denaturation, and extension temperatures and time were carried out for each primer pair. Optimum PCR conditions for each master mix (table 3 and 5) were then used for the amplification of *ACE2*. Each primer pair of *ACE2* was amplified, viewed on a gel and samples successfully amplified were pooled for sequencing.

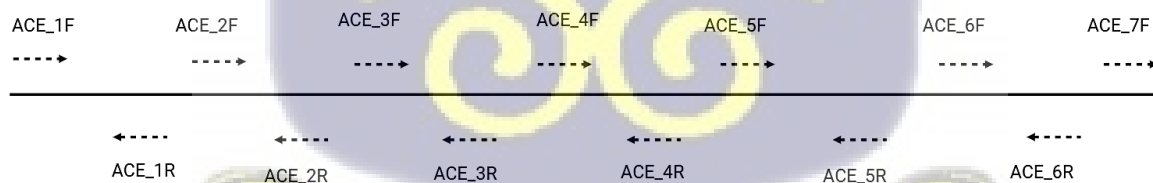


Figure 11: Schematic of overlapping primers designed for *ACE2*

596

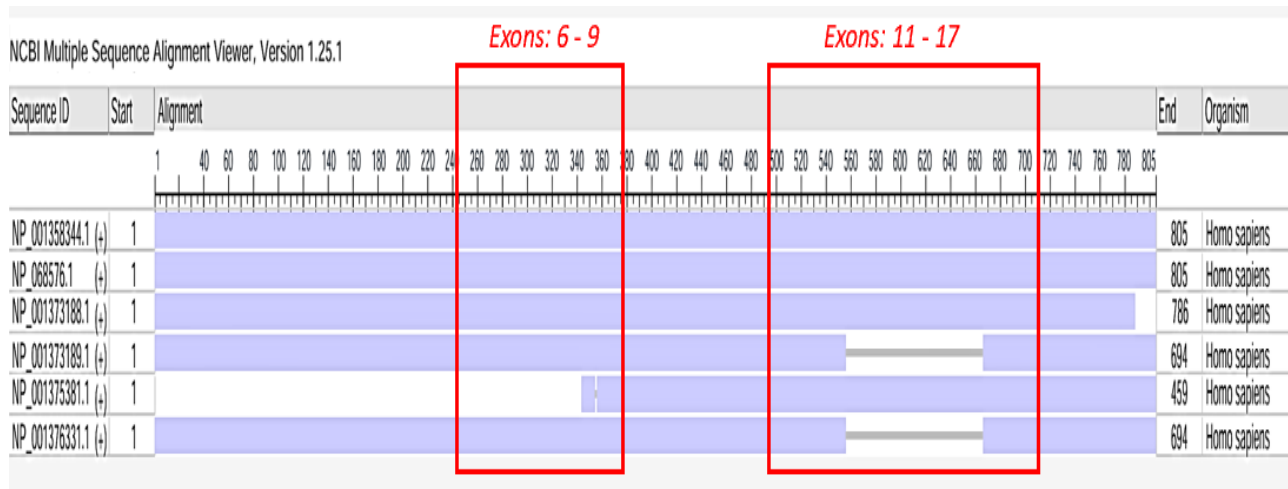


Figure 12: Region of ACE2 protein variants amplified based on NCBI sequence alignment tool

Exons 6 to 9 and 11 to 17 were amplified by the selected primers used in *ACE2* target sequencing.

597

598

599

600

601 Table 2: Master mix reaction for Q5 Hot Start high fidelity 2x master mix and GoTaq® G2 Hot

602 Start master mix

Component	One reaction (μL)
Master mix	12.5
1% DMSO	0.57
cDNA	2.5
Nuclease free water	4.5
F primer (0.27 μM)	0.57
R primer (0.27 μM)	0.57

603

604

Table 3: PCR condition for Q5 Hot Start high fidelity 2x master mix and GoTaq® G2 Hot Start master mix

Condition	Temperature (°C)		Time
	Q5 Hot Start	G2 Hot Start	
Initial denaturation	98	95	45 seconds
Denaturation	98	95	30 seconds
Annealing	59	59	30 seconds
Extension	72	72	25 seconds
Final extension	72	72	2 minutes
Hold	4	4	Infinite
Cycle number = 40			

Table 4: Master mix reaction for One Taq® Quick-Load quick-load 2X master mix with standard buffer master mix

Component	One reaction (µL)
Master mix	17.5
1% DMSO	1
cDNA	2.5
NFH ₂ O	12
F primer (0.29 µM)	1
R primer (0.29 µM)	1

Table 5: PCR conditions for ACE2 (primer 2 to 8) using One Taq® Quick-Load 2X master mix with standard buffer

Condition	Temperature(°C)	Time
Initial denaturation	94	45 seconds
Denaturation	94	30 seconds
Annealing	59	30 seconds
Extension	68	30 seconds
Final extension	68	4 minutes
Hold	4	Infinite

Cycle number = 40

3.3.4 Agarose gel electrophoresis

A two percent gel was prepared using 2 grams of agarose (molecular grade 500 g, catalogue number - ES520-B091370) and 100 mL of 1X tris-acetate ethylenediaminetetraacetic acid (TAE, fisher bioreagents, catalogue number: 207475). The mixture was heated at 500 °C until all the agarose dissolved completely. This was then allowed to cool until warm, 3 µL of GelRed nucleic acid stain (catalogue number: 18G1231) was added, mixed well and poured onto a casting tray with a gel comb fixed and allowed to solidify. Solidified 2% agarose was then transferred into a gel tank and 100 mL of 0.5% TAE was poured over it and the gel comb was removed gently. About 5 uL PCR products of the One Taq® Quick-Load 2X master mix with standard buffer and GoTaq ® G2 Hot Start master mix were loaded into each well (due to the manufacturer's addition of loading dye in the master mix). For PCR products obtained using Q5 Hot Start high fidelity 2x master mix, 2 uL

of gel loading dye (New England Biolabs, catalogue number: B7021) and 3 μ L of PCR products were mixed thoroughly prior to transfer into each well. The electrophoresis tank was set to 100 Volts for 1 hour, after which the gel was viewed using the AmershamTM Imager 600. An expected band size of approximately 400 base pairs indicated amplification of the target for each primer pair.

3.3.5 DNA Library Preparation and Sequencing Using the Oxford Nanopore Technology

Amplicon Preparation

PCR amplicons were purified by an addition of 1X AMPure beads and incubation at room temperature for 5 minutes. The mixture was placed on a magnetic rack to allow for bead pelleting until the supernatant became clear. The supernatant was carefully removed, and the AMPure beads were washed twice with 500 μ L of 80% ethanol. About 22 μ L of elution buffer was then added to the beads and resuspended. An incubation step at room temperature was also carried out for 5 minutes and then positioned on a magnetic rack until a clear supernatant was obtained and transferred into a new tube. The purified amplicons were quantified using a QubitTM Flex Fluorometer (Catalogue number: Q33326). Subsequently, the quantified clean amplicons were diluted with nuclease-free water to achieve a final concentration of 50 ng in a total volume of 12.5 μ L.

Library Preparation

The 12.5 μ L diluted amplicon DNA was transferred into a new Eppendorf tube and 1.75 μ L of Ultra II End-prep reaction buffer and 0.75 μ L of Ultra II End Prep Enzyme Mix (New England Biolabs, catalogue number: E7546S) were added. This was thoroughly mixed and incubated in a thermocycler (Eppendorf master cycler X50s), first at 20°C for 5 minutes and then at 65°C for an

additional 5 minutes to facilitate the preparation of the DNA ends. Following this, the samples were placed on ice for 1 minute to stabilize the reaction.

Native barcoding

Unique barcodes were introduced to each end-prepped DNA sample to allow for distinct identification during sequencing. Native barcodes were thawed, flicked, pulse centrifuged, and kept on ice before added the end-prepped DNA samples. Approximately, 2 μ L of end-prepped DNA and 1.5 μ L of a native barcode were transferred into a new tube and mixed thoroughly. A Blunt/TA Ligase master mix was prepared using nuclease-free water and 6.5 μ L was added to each tube containing the barcoded samples. The components were mixed thoroughly placed in a thermal cycler (Eppendorf master cycler X50s) at 20°C for 20 minutes followed by 65°C for 10 minutes to facilitate ligation of the adaptors. After ligation, 10 μ L from each barcoded fragment was pooled into a new Eppendorf tube and purified using AMPure XP beads as described previously (at the end-prep step). The beads were washed with 80% ethanol to remove contaminants and resuspended in 35 μ L elution buffer.

Adapter ligation and clean-up

Adaptors (5 μ L) specific to the Oxford Nanopore Technology platform were added to the pooled barcoded DNA, as well as 5 μ L NEBNext Quick Ligation Reaction Buffer and 5 μ L of Quick T4 DNA Ligase. This mixture was incubated at room temperature for 20 minutes to allow for efficient ligation. Following this incubation, AMPure XP Beads were added again for purification, removal of unligated adaptors and other contaminants. This was then eluted in 17 μ L of elution buffer and

the concentration of the DNA library was quantified using a Qubit™ Flex Fluorometer (Thermo Fisher Scientific: Q33326). Finally, the library was mixed with 37.5 µL of Sequencing Buffer II and 25.5 µL of Loading Beads II, and introduced into the MinION sequencer for a seventy two-hour sequencing run in real-time high-accuracy mode. The sequence data was then exported and analyzed.

3.3.6 Bioinformatic pipeline



Figure 13: Bioinformatic analysis general workflow

3.3.6. A. Preprocessing and Quality Control

Raw sequencing data were received in FASTQ format and a custom Bash script was used to automate data processing (figure 13). The script merged FASTQ files from previous runs based on a predefined disk usage threshold. If a barcode directory contained less data than the threshold, the script skipped it to optimize storage. Subsequently, NanoPlot (<https://github.com/wdecoster/NanoPlot>) was utilized to perform quality control (QC) analysis on the merged FASTQ files, generating individual QC reports (De Coster & Rademakers, 2023).

3.3.6. B. Adapter Trimming and Quality Filtering

Adapter sequences introduced during library preparation were removed from the FASTQ files using Porechop (<https://github.com/rrwick/Porechop>) (Bonenfant et al., 2023). Following adapter trimming, Nanofilt (<https://github.com/wdecoster/nanofilt>) (De Coster et al., 2018) was employed for quality filtering. This step removed reads shorter than a specified minimum length, longer than a defined maximum length, and those with an average base quality score below a set threshold based on the sequence data for each sample.

3.3.6. C. Read Mapping and Variant Calling

Minimap2 (<https://github.com/lh3/minimap2>) (H. Li, 2018) was used to map the filtered reads against the chromosome X sequence of the human reference genome (Accession number: NC_000023.10). This alignment process allows for the identification of potential variations in the sequences compared to the reference. The resulting SAM files were then converted to sorted and indexed BAM files using SAMtools (<https://github.com/samtools/htslib>) (Bonfield et al., 2021), facilitating efficient storage and analysis.

3.3.6. D. Variant Annotation

Freebayes (<https://github.com/freebayes/freebayes>) (Garrison & Marth, 2012) was employed for variant calling on the aligned reads. This step identified single nucleotide polymorphisms (SNPs) within the mapped RNA-seq data (figure 13). The identified variants were then annotated using SnpEff (<https://pcingola.github.io/SnpEff/>), a software tool that predicts the potential functional

consequences of these variations (Cingolani et al., 2012). SnpEff utilizes reference databases such as dbSNP to provide additional information about known variants. The annotation process enriches the analysis by providing insights into the potential impact of the identified SNPs on gene function.

3.4 General workflow for *in-silico* analysis



Figure 14: In-silico analysis pipeline

I-TASSER provides modelled protein structures, ClusPro predicts binding interaction points between protein and ligand, and PRODIGY allows for the analysis of docked structures.

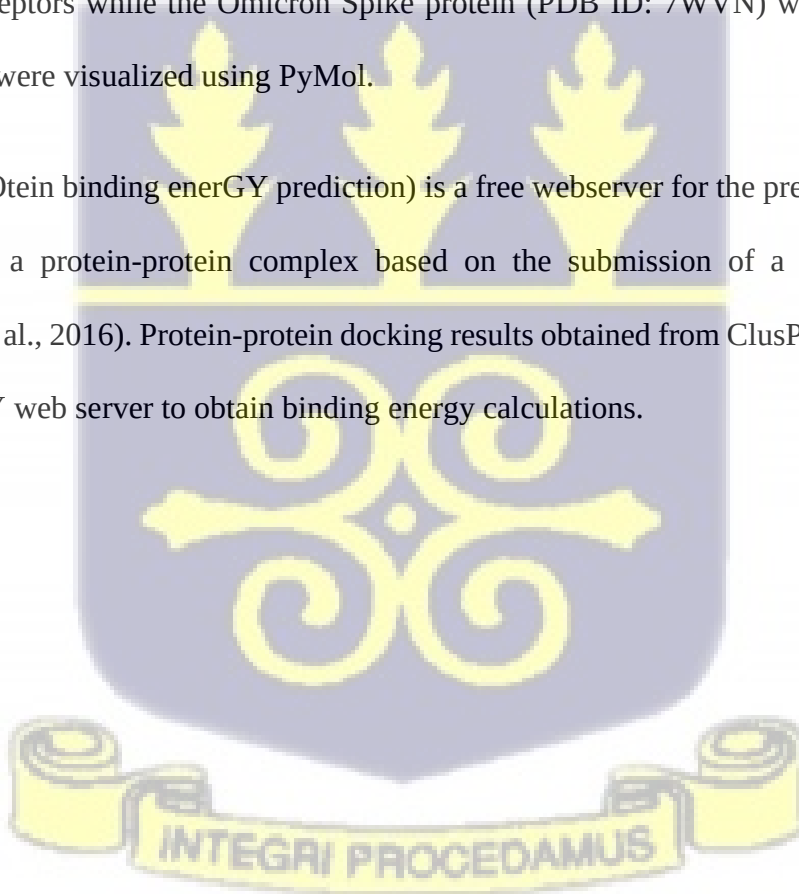
3.4.1 In-silico analysis pipeline

I-TASSER (Iterative Threading ASSEmbly Refinement) is a free online database for predicting protein structure and functional annotation (Yang & Zhang, 2015; X. Zhou et al., 2022). I-TASSER generates a three-dimensional model using an amino acid sequence after several threading orientations and iterative assembly of structural models (Roy et al., 2010). To obtain a three-dimensional model of ACE2, the amino acid sequence information was obtained from the UniProt database (ID: Q9BYF1) and submitted on I-TASSER for a three-dimensional model. Mutant ACE2 amino acid sequences were also submitted. An output from I-TASSER is obtained after a few days (depending on the queue) with a list of tertiary protein structure predictions and confidence score

and, root mean square deviation (RMSD) and TM values. The selected structure is usually the first modelled structure showing a good confidence score and TM values which are visualized using PyMOL or other compatible software (Yang & Zhang, 2015). The ACE2 cryo-electron microscopy structure 6M18 with a 2.9 angstroms resolution was used as template to generate the mutants (Yan et al., 2020).

ClusPro is a free protein-protein docking online tool useful for structural studies based on a rigid-body docking technique (PIPER) and the fast Fourier transform (FFT) correlation (Alekseenko et al., 2020; Jones et al., 2022). Modelled ACE2 mutants (reference ACE2 PDB ID:6M18) were submitted as receptors while the Omicron Spike protein (PDB ID: 7WVN) was used as ligand. Docking results were visualized using PyMol.

PRODIGY (PROtein binding enerGY prediction) is a free webserver for the prediction of binding energy between a protein-protein complex based on the submission of a three-dimensional structure (Xue et al., 2016). Protein-protein docking results obtained from ClusPro were submitted to the PRODIGY web server to obtain binding energy calculations.



CHAPTER FOUR

4.0 RESULTS

4.1 Demography of participants

A total of 330 sputum, nasopharyngeal, or nasal swabs in viral transport media obtained from consented participants were used in this study. Of these, a total of 119 samples were analyzed as the other 330 samples failed to pass quality control for sequencing (table 6).

Table 6: Number of sample types obtained and number of samples that were sequenced successfully from each WANIDA centre

WANIDA Partner	Country	Number of samples				Samples successfully sequenced
		VTM	cDNA	RNA	Sputum	
CEA-PCMT	Guinea Conakry	0	40	0	0	21
ACEGID	Nigeria	40	0	0	0	21
ACEPHAP	Nigeria	30*	5*	10	0	9
ACENTDFB	Nigeria	40	0	0	0	29
ACEMFS	Nigeria	40	0	0	0	12
IRCB/CERPAGE	Benin	40*	0	40	0	4
CEA-ITECH/MTV	Burkina Faso	0	0	0	0	NA
WACCBIP	Ghana	82*	0	0	8*	23

Key: * - Source of successful samples

764

765 *4.2 PCR optimization and gel electrophoresis*

766 PCR optimization was carried out using 12 pairs of overlapping primers targeting *ACE2*. After
767 several steps of optimization for each primer pair, four primers of the *ACE2* gene were selected as
768 most of the other primers failed. Although *ACE2* primer pairs 2 and 3 were amplified successfully,
769 they were not used in this study as they cover the first four exons of *ACE2* which has already been
770 assessed for polymorphisms in the Ghanaian population (Duah-Quashie et al., 2024). The selected
771 *ACE2* primers 4, 6, 7 and 8, target the 3' end of the 6th exon to the 5' end of the 9th exon, and 3'
772 end of the 11th exon to 5' end of the 17th exon of the six different *ACE2* isoforms (figure 10).



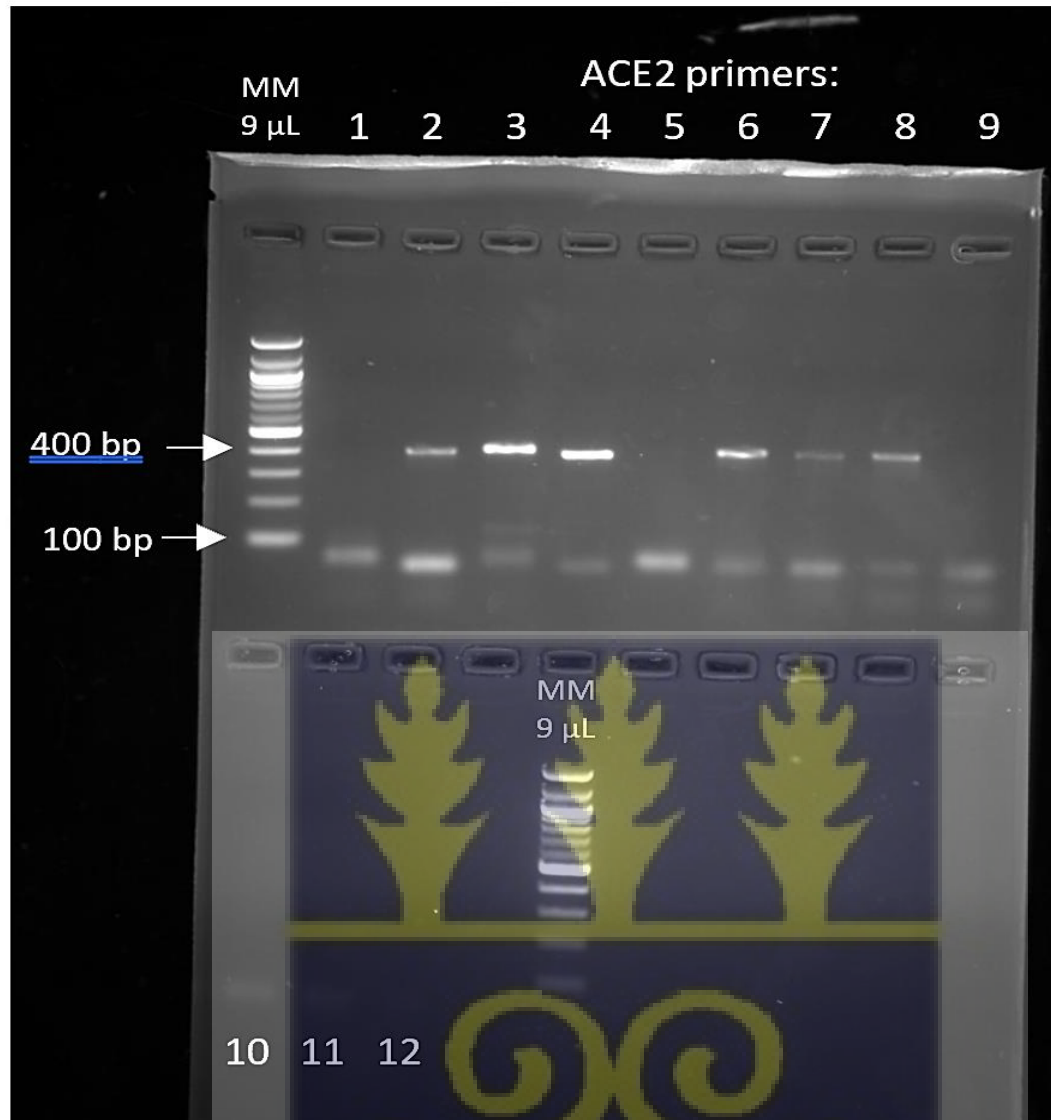


Figure 15: Twelve overlapping ACE2 amplicons after primer optimization

Six primers for ACE2 were optimized successfully at an expected band size of 404 base pairs.

Primer pairs 4,6,7 and 8 were used for selected for amplification and sequencing.

(Ladder size – 100 kb)

774



775

776 Figure 16: Gel representation of 85 samples after amplification of the four selected ACE2 primers

777

778

779

780

781

782

783

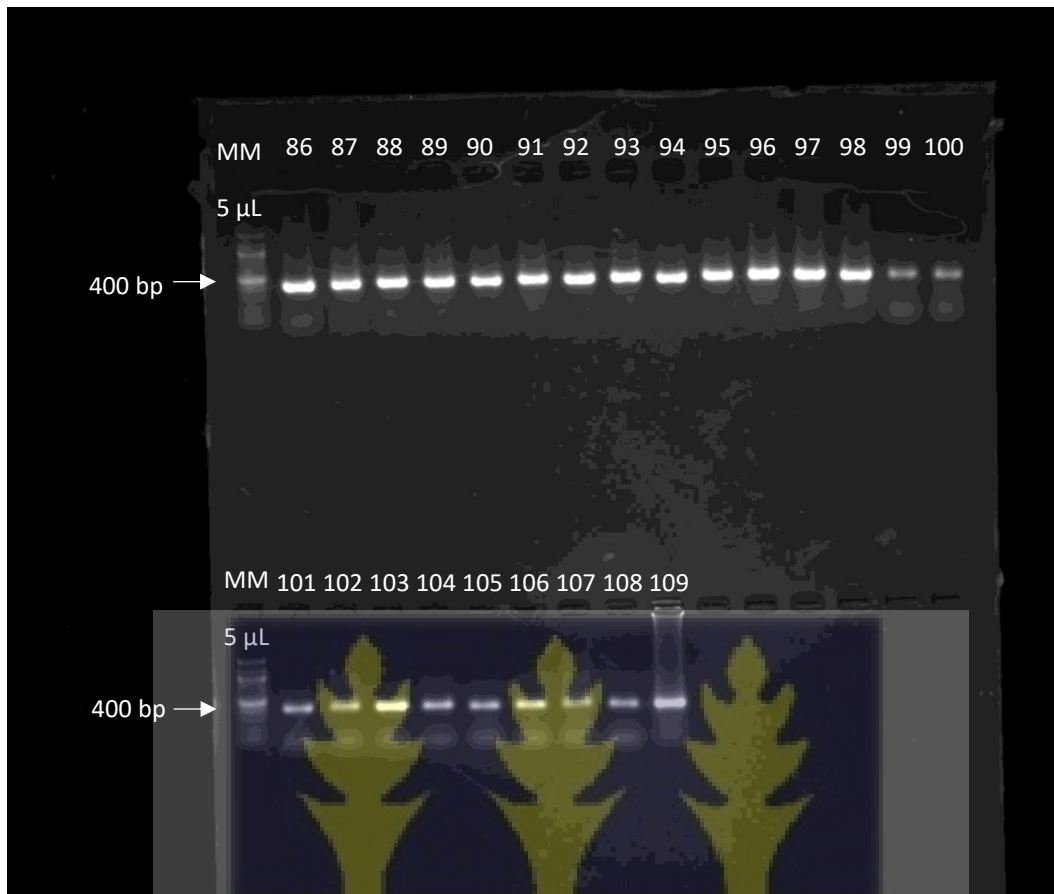
784

785

Sample one to 65 were amplified using the Q5 Hot Start Q5 Hot Start high fidelity 2x master mix while sample 66 to 85 were amplified using the One Taq® Quick-Load quick-load 2X master mix with standard buffer.

(Ladder size – 100 kb)

786



787

788

789

790 Figure 17: Gel representation of 24 samples amplified using GoTaq® G2 Hot Start master mix

791 (Ladder size - 100 kb)

4.3 Bioinformatic analysis

4.3.1 SNP identification

Analysis with SnpEff produced annotated data of twelve missense mutations and one stop codon for all the 119 samples analyzed (figure 16). Out of the thirteen SNPs that were identified, eight of them had been identified while five of them have not been reported in literature (figure 17) based on search in UniProt using the variant viewer option (<https://www.uniprot.org/uniprotkb/Q9BYF1/variant-viewer>).

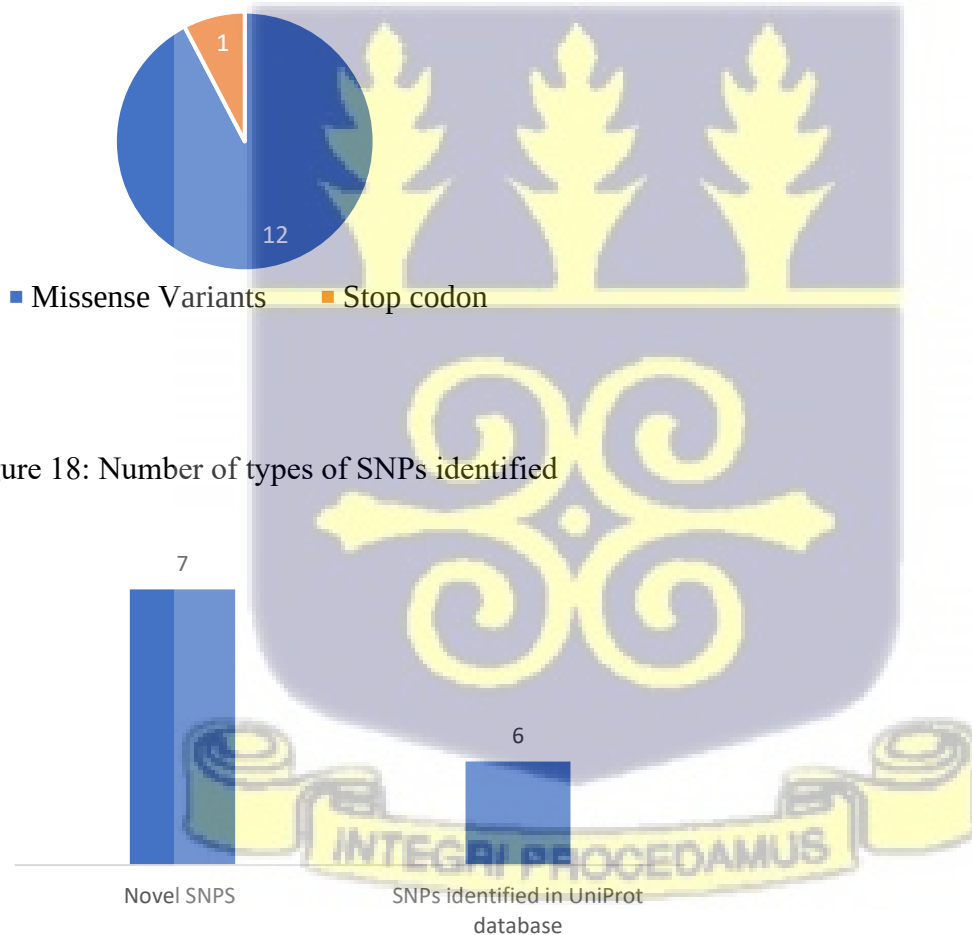


Figure 18: Number of types of SNPs identified

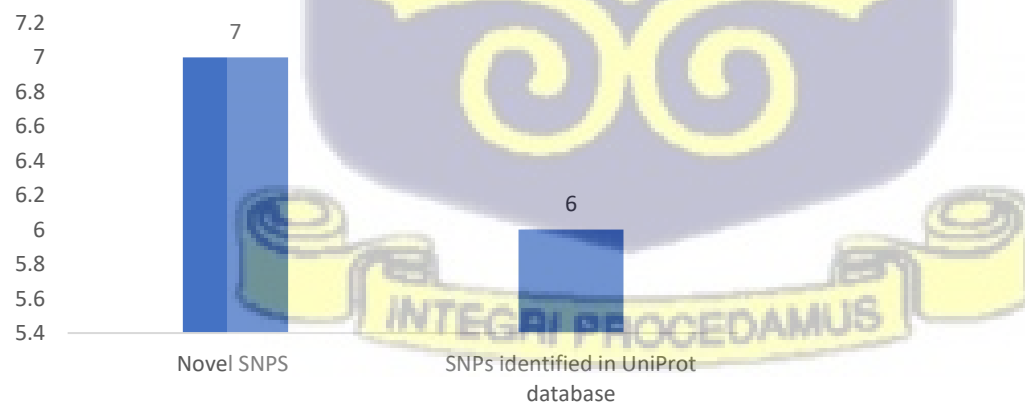


Figure 19: Number of known and novel SNPs identified

802 SNP positions 1055 and 1054 had two alternate alleles coding for different amino acids resulting
 803 in missense mutations (table 7). All other gene positions had one alternate allele identified in the
 804 samples except position 1046 which encodes a stop codon.

805 Table 7: Specific gene and amino acid variations identified

Variant	Gene position	Reference	Alternate	Individuals with SNP in cohort used	Minor allele frequency in Africa	Amino acid change	Protein effect
1	1819	T	C	1	0	p.S607G	Missense
2	1754	A	G	1	0	p.L585P	Missense
3	1721	A	G	1	0	p.V574A	Missense
4	1718	A	G	1	0	p.V573A	Missense
5	1301	G	T	1	0	p.T434K	Missense
6	1055	C	A	2	0	p.G352V	Missense
7	1054	C	A	9	-	p.G352W	Missense
8	1055	C	T	7	-	p.G352E	Missense
9	1057	T	G	2	-	p.K353Q	Missense
10	1054	C	T	9	-	p.G352R	Missense
11	1050	G	C	101	-	p.D350E	Missense
12	1049	T	A	58	-	p.D350V	Missense
13	1046	C	T	8	-	p.W349*	Stop gained

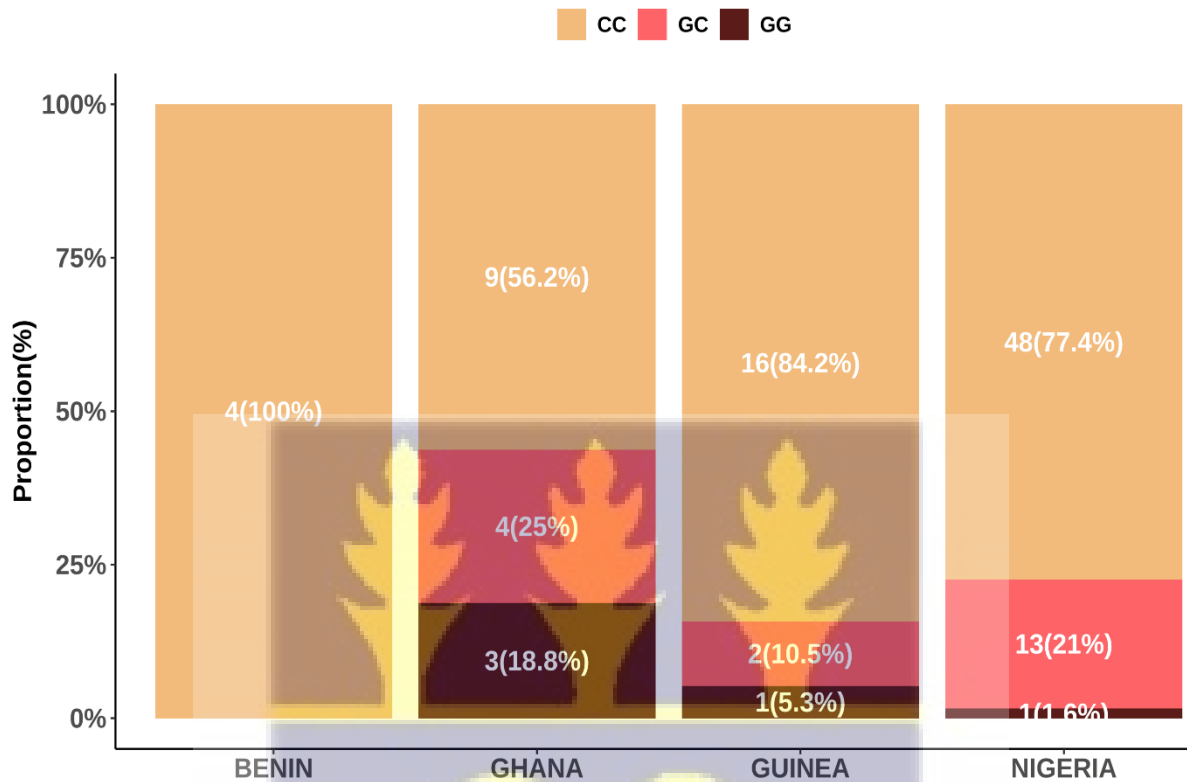
806 Sample 7 to 12 (in the red box) are the seven novel SNPs identified.

807 Distribution of SNPs D350E and D350V across each country:

808

809

C - Alternate allele G – Reference allele



810

Figure 20: Distribution of D350E genotypes across each country

811

812 A high percentage of the alternate CC genotype was observed across each country. The highest

813 CC genotype (77.4%) was found in Nigeria and all four persons sequenced from Benin also had

814 the CC genotype.

815

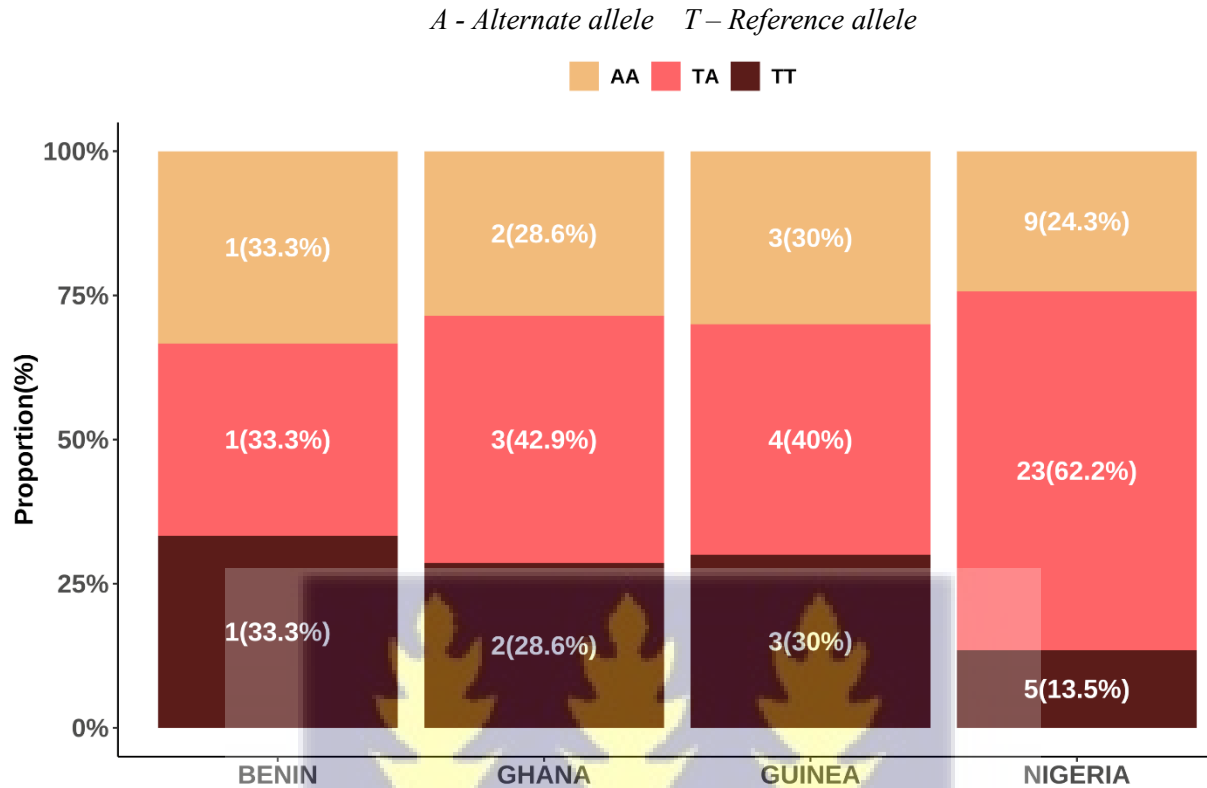


Figure 21: Distribution of D350V genotypes across each country

A high percentage of the heterozygous TA genotype was found across each country. The highest percentage of the heterozygous TA was observed in samples received from Nigeria.

816

817

818



Alpha missense predicted three benign and two pathogenic phenotypes of the known SNPs and stability fold change predicted three likely destabilizing SNPs and two unlikely destabilizing phenotypes. Thus, the resulting phenotype changes could affect overall stability and structure of ACE2.

Table 8: Alpha missense prediction and stability fold using FoldX for known SNPs identified in the Uniprot database

Protein position	Alpha missense prediction	Alpha Missense phenotype	Stability fold change using FoldX ($\Delta\Delta G$)	Stability fold change using FoldX phenotype
G352V	0.74	Pathogenic	3.7	Likely destabilizing
L585P	0.92	Pathogenic	4.7	Likely destabilizing
S607G	0.1	Benign	0.09	Unlikely destabilizing
T434K	0.27	Benign	0.53	Unlikely destabilizing
V573A	0.21	Benign	3.43	Likely destabilizing

PholyPhen-2 analysis of the novel SNPs assigned a probably damaging phenotype to each SNP. For the novel SNP W349*, no PolyPhen-2 prediction was carried out. The sensitivity of each prediction was lower compared to the specificity of each SNP.

834 Table 9: PolyPhen-2 prediction score, description, sensitivity and specificity of novel SNPs
835 identified

836

Protein position	PolyPhen-2 prediction score	PolyPhen-2 Description	PolyPhen-2 Sensitivity	PolyPhen-2 Specificity
G352W	1.000	Probably damaging	0.00	1.00
G352E	0.995	Probably damaging	0.68	0.97
K353Q	0.806	Possibly damaging	0.84	0.93
G352R	0.998	Probably damaging	0.27	0.99
D350E	0.994	Probably damaging	0.69	0.97
D350V	0.994	Probably damaging	0.69	0.97
W349*	-	-	-	-

837 Key:
838

- Benign (less than or equal to 0.446) – not likely to affect protein variant
- Possibly damaging (greater than 0.446 and less than or equal to 0.908) – may affect protein function
- Probably damaging (greater than 0.908) - will affect protein function
- Polymorphism Phenotyping version 2 (PolyPhen-2)

Source: Adzhubei et al., 2010

4.3.2: Protein modelling and analysis

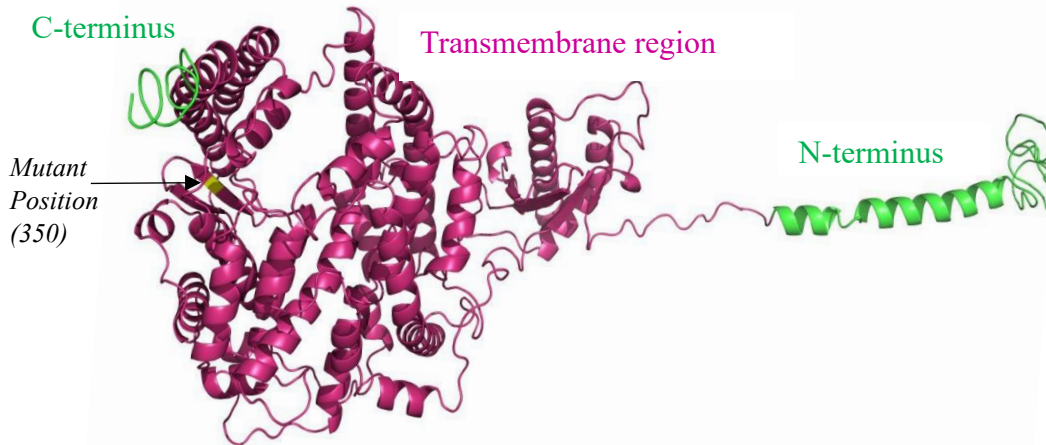


Figure 23: Modelled ACE2 Val350Glu mutant

ACE2 was modelled using iTASSER and Pymol. Mutant position shown in yellow.

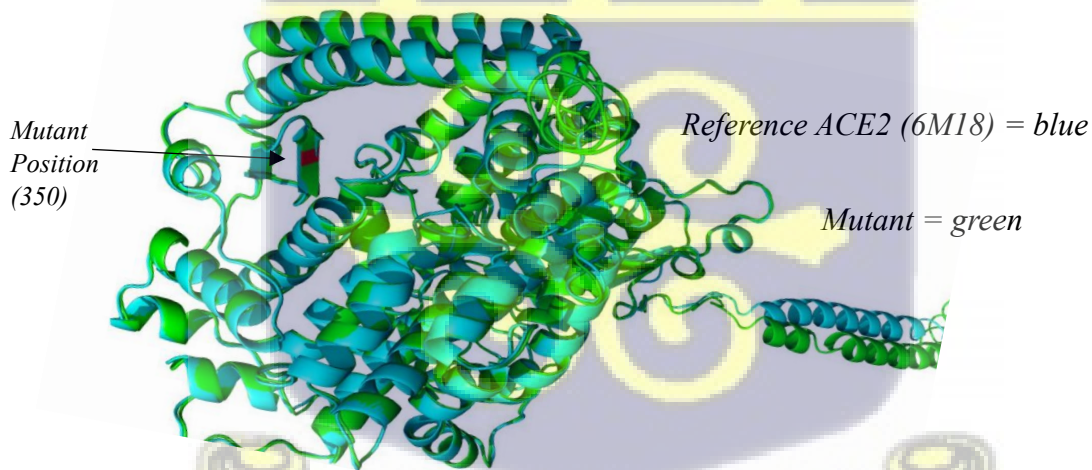


Figure 22: Superimposition of reference ACE2 (6M18) and D350V/E mutant

The root mean square deviation (RMSD) of D350E was 0.4560 while that of D350V was 0.4388.

845

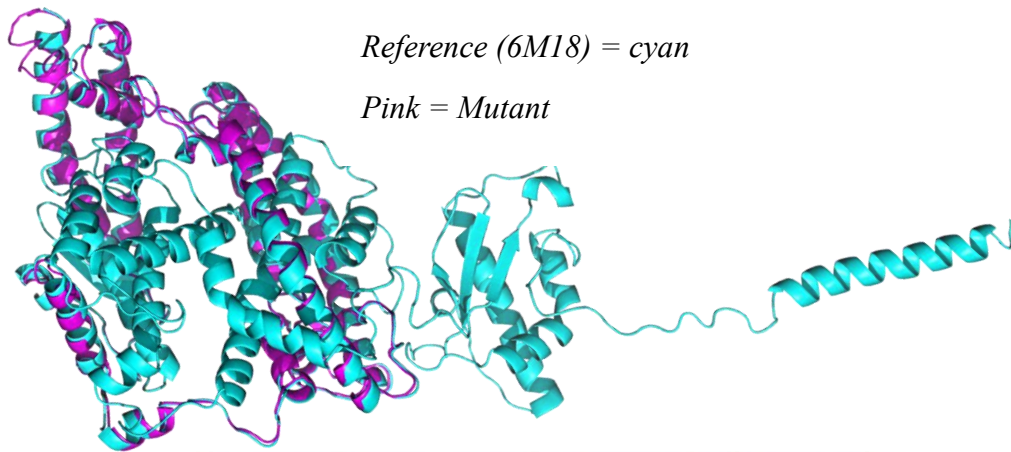


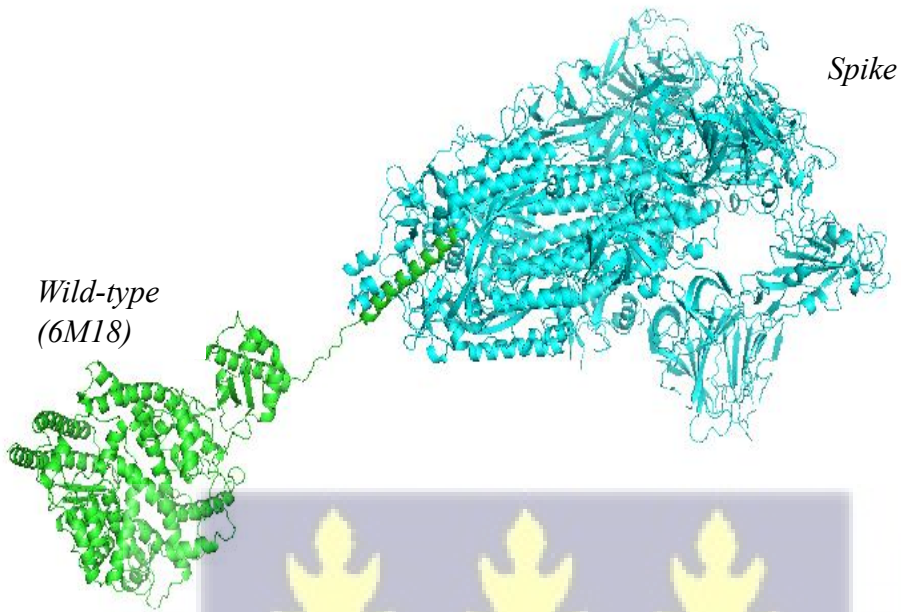
Figure 24: Modelled ACE2 W349*

At position 349, tryptophan produces a stop codon, producing mutant ACE2 with a loss of the transmembrane domain and a portion of the extracellular domain.

The root mean square deviation (RMSD) of W349* was 0.641.



846 4.4.3 Binding affinity analysis of modelled ACE2 and Spike Omicron



847
848 Figure 25: Wild-type ACE2 and Spike of Omicron variant

849 The wild-type ACE2 interacts with the Spike protein of the Omicron variant with a binding
850 energy of -17.6 and a binding affinity of $1.3\text{E}-13$ Kd(M) at 37°C .



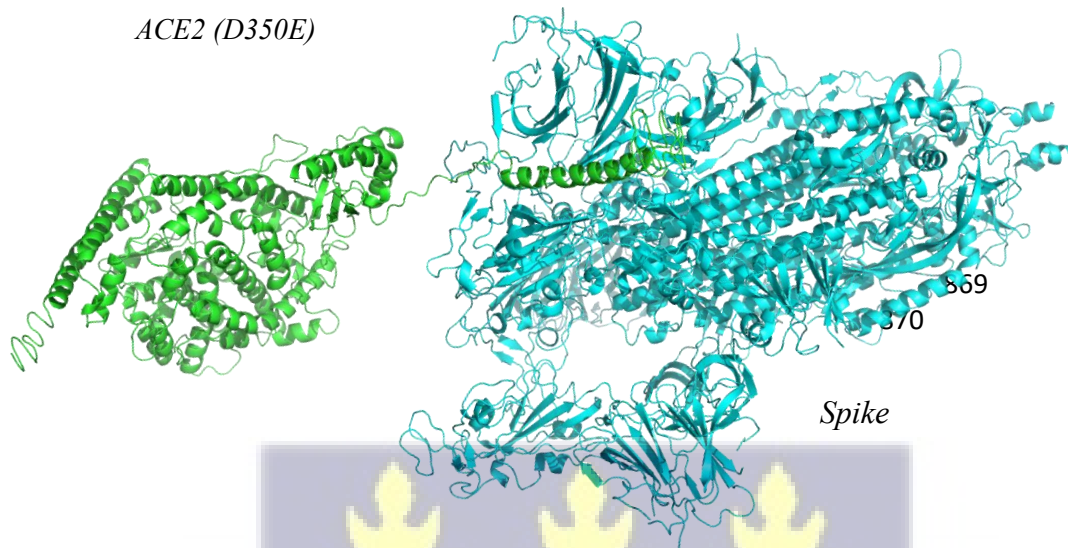


Figure 26: Mutant ACE2 (D350E) and Spike of Omicron variant

Mutant *ACE2* (D350E) interacting with Spike protein of the Omicron variant with a binding energy of -14.1 kcal/mol and a binding affinity of $1.2\text{E}-10$ Kd(M) at 37 °C.



887

888

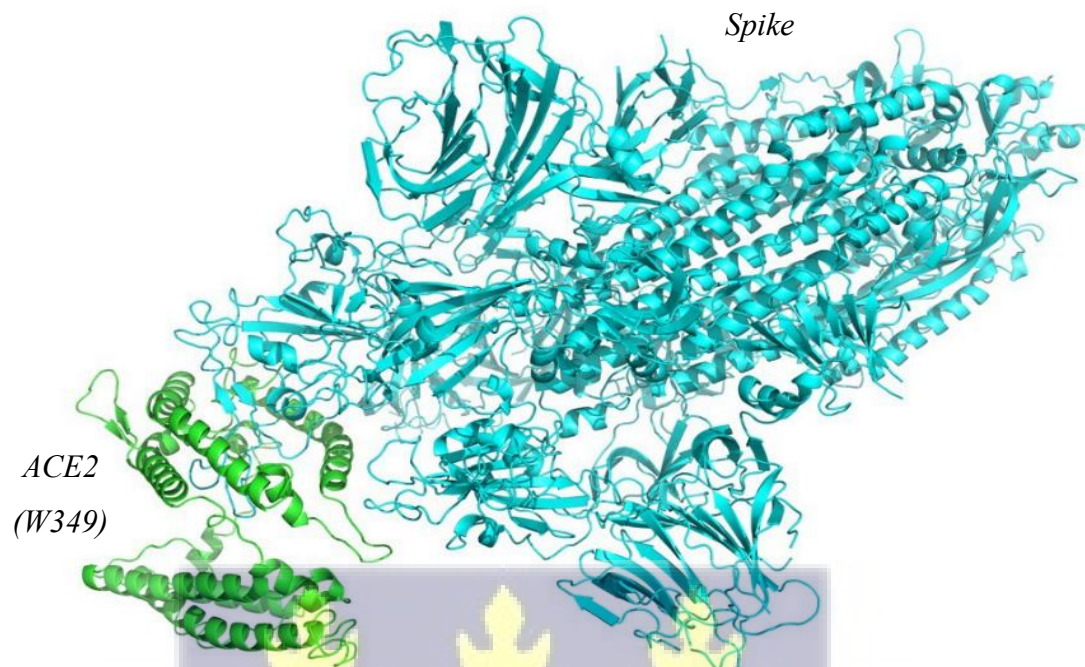


Figure 27: Mutant ACE2 (W349) and Spike of Omicron variant

Mutant ACE2 (W349) interacting with Spike protein of the Omicron variant with a binding energy of -14.7 kcal/mol and a binding affinity of $4.10\text{E-}11$ Kd(M) at 37 °C.

889 Table 10: Summary of binding energy and binding affinity for reference ACE2 and two mutants

ACE2	Binding energy (kcal/mol)	Binding affinity [Kd(M)]
Reference ACE2	-17.6	1.3E-13
Mutant (D350E)	-14.1	1.2E-10
Mutant (Trp349*)	-14.7	4.10E-11

890
891 *A lower binding energy was observed for the reference ACE2 compared to other mutants.*



CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1.1 Introduction to discussion

Multiple hypotheses such as relatively younger population, host genetics, exposure to previous pathogens, and *Plasmodium falciparum* infections had been proposed to explain the relatively low COVID-19 in the West African population (Njenga et al., 2020; Lawal, 2021; Oduro-Mensah et al., 2022; Tapela et al., 2024; Sandi et al., 2024). This study sought to explore the role of SNPs in the primary receptor for SARS-CoV-2 cellular invasion; ACE2, in four West African countries; Ghana, Nigeria, Benin and Guinea, which had low severity of COVID-19.

ACE2 interacts with the RBD of the Spike protein of SARS-CoV-2 for cellular invasion and this interaction can be enhanced or decreased, affecting the onset of infection or severity (Ali et al., 2020; Sorokina et al., 2020). Several polymorphisms in ACE2 have been reported in specific populations which may have played a role in increasing or decreasing COVID-19 severity (Ali et al., 2020; Srivastava et al., 2020; Khayat et al., 2020; Srivastava et al., 2020; Duah-Quashie et al., 2024). ACE2 polymorphisms have been explored within the first four exons in the Ghanaian population; SNPs unique to Ghanaians were identified and proposed to play a role in COVID-19 protection (Duah-Quashie et al., 2024). There are no reports on ACE2 SNPs in Nigeria, Benin, and Guinea, where low COVID-19 severity and deaths were observed during the pandemic. Herein, 11 out of 22 exons of ACE2 were targeted to identify SNPs which may alter its function in SARS-CoV-2 cell entry for individuals from Ghana, Nigeria, Benin and Guinea.

913 *5.1.2 ACE2 target amplification, sequencing and analysis*

914 The *ACE2* targeted sequencing covered exons six to nine and eleven to seventeen of the six *ACE2*
 915 variant proteins produced after intron splicing (Appendix 3). Twelve primer pairs were previously
 916 synthesized for the full-length amplification of *ACE2*. However, only six primers were
 917 successfully amplified in this project due to poor primer design; failure to bind all six *ACE2*
 918 isoforms producing irreproducible data. Out of the six successfully optimized primers, four were
 919 used in this project targeting the 3' end of the 6th exon to the 5' end of the 9th exon, and 3' end of
 920 the 11th exon to 5' end of the 17th exon of the six different *ACE2* isoforms. In figure 14, successful
 921 amplification of the six primer pairs was observed with poor amplification results of primer pairs
 922 1, 5, 9, 10 and 11. These primers were designed to overlap the full length *ACE2* however, some
 923 of the primers could not bind all the *ACE2* isoforms. For example, the reverse primer of *ACE2*
 924 primer pair 1 could not bind the *ACE2* gene sequence template as visualized using Benchling
 925 software (Appendix 5).

926 In figure 15, samples 1 to 65 were amplified using the Q5 Hot Start Q5 Hot Start high fidelity 2x
 927 master mix while sample 66 to 85 were amplified using the One Taq® Quick-Load quick-load 2X
 928 master mix with standard buffer. The varying band sharpness observed could be attributed to
 929 factors such as poor sample quality and the varying chemistry of the polymerases used.
 930 Comparatively, 24 samples amplified using the GoTaq® G2 Hot Start master mix in figure 16
 931 showed sharper bands. Regardless of the type of polymerase used, sharp and faint bands were
 932 observed however, the high-fidelity master mixes; Q5 Hot Start Q5 Hot Start high fidelity 2x
 933 master mix and GoTaq® G2 Hot Start master mix showed sharper bands. In some of the samples,
 934 faint bands were observed at 100 bp which could be unspecific amplification. The amplicons were

purified using AMPure beads to exclude unspecific amplification and other contaminants prior to sequencing.

Analysis of SNPs within this region revealed twelve missense SNPs and one stop codon which truncates the full-length ACE2 (table 7). Out of these, seven were not identified (table 7) in the UniProt database (<https://www.uniprot.org>) (The UniProt Consortium et al., 2023). UniProt is a free open database of protein sequences serving as a centralized hub for other 183 databases such as AlphaMissense, gnomAD, ClinGen and Ensembl (The UniProt Consortium et al., 2023). The five known SNPs identified had a minor allele frequency of zero in the African population (table 7) indicating an absence in Africans based on the UniProt database. Out of these, four SNPs; S607G, L585P, V574A, T434K were identified in one individual each. Thus, the true allele frequencies of these SNPs in the African population are not zero as suggested by the database.

G352V was identified in seven individuals of which five were homozygous for the reference allele while three were heterozygous highlighting the low frequency of the alternate allele in Africans. G352V has also been proposed to provide protection against COVID-19 by decreasing SARS-CoV-2 and ACE2 interaction (Suryamohan et al., 2021). The identification of these SNPs also suggests a low coverage of Africans in genomic databases as the data available is limited to specific African populations only (Gurdasani et al., 2015). A larger sample size will be essential in determining the true prevalence of the five known SNPs that were identified in West Africa.

The novel SNPs that were identified were analyzed using Polymorphism Phenotyping version 2 (PolyPhen-2) (SNP effect on protein function), I-TASSER (protein three-dimensional modelling), CLusPro (protein-protein docking) and Prodigy (assessment of binding efficiency). Seven novel SNPs were identified with 101 individuals having a change from Asp to Glu at position 350; a

large proportion of them had the alternate allele across all four countries (figure 20). All of the novel SNPs identified were predicted to be probably damaging except K353Q which was possibly damaging based on PolyPhen-2 (table 9). Probably damaging indicates a higher chance of the SNP affecting protein function relative to the description of possibly damaging (Adzhubei et al., 2013). PolyPhen-2 prediction was not carried out for the novel W349 stop codon as the algorithm only works for missense mutations only (Adzhubei et al., 2013). Thus, out of the seven novel SNPs identified, five could affect protein function greatly based on PolyPhen-2 prediction which may alter the interaction between SARS-CoV-2 and ACE2 affecting COVID-19 severity.

5.1.4 In-silico analysis of novel mutants

The ACE2 protein sequence (ID: Q9BYF1) was retrieved from UniProt, pasted into Microsoft Word and mutant ACE2 forms were generated by changing amino acids at specific positions. Modelled three-dimensional structure of ACE2 mutants D350V and D350E was carried out and compared to the reference 6M18 (figure 23) (Yan et al., 2020). The reference 6M18 is a wild-type ACE2 solved by electron microscopy with a 2.90 Å (protein structure identification - Q9BYF1 in the UniProt database) (Yan et al., 2020) which makes it a good structure for comparison with modelled ACE2 mutants. ClusPro provides a platform for predicting protein-ligand interaction position at minimum binding energies while providing root mean square deviation (RMSD) values (López-Camacho et al., 2016; Alekseenko et al., 2020).

RMSD is a metric used in discriminating modelled structures by comparing a known ligand-receptor interaction to a computed prediction of ligand position to a receptor (Sargsyan et al., 2017). An RMSD value of 2 and above is considered to have a poor computed docking prediction

while 2 and below is considered highly precise (Sargsyan et al., 2017). Upon comparison of D350E and D350V (figure 21) to the reference 6M18, RMSD values of 0.456 and 0.4388 were obtained indicating good computational models that are not very distinct from the reference. The W349* RMSD value was also 0.641 with the modelled structure showing a loss of the N-terminal encompassing the extracellular and transmembrane domains of ACE2 (figure 22). A loss of the transmembrane domain implies that soluble ACE2 which is not bound to the host cells will be produced. Soluble ACE2 has been described to have the ability to bind to SARS-CoV-2 by acting as a molecular decoy (Kuba et al., 2021). Upon interaction, soluble ACE2 can prevent cellular invasion by neutralizing SARS-CoV-2 such that it is unable to interact with membrane-bound ACE2 leading to an infection (Appendix 4) (Kuba et al., 2021). Soluble ACE2 however is always present in an individual as membrane-bound ACE2 requires cleavage by other host proteases to form active soluble ACE2 which regulates the renin-angiotensin system for blood pressure homeostasis (Yeung et al., 2021).

ClusPro analysis of W349* (producing soluble ACE2) indicated a binding interaction with Spike with a binding energy of -14.7 (kcal/mol) and binding affinity of 4.10×10^{-11} (K_d at 37 °C) (figure 27, table 10). Low binding energy and affinity were also obtained for D350E and D350V (figure 25 & 26, table 10). Binding affinity indicates the strength of the bond between molecules and low binding energy implies a strong bond between molecules being studied (Tallei et al., 2021). Thus, all the structures show a good interaction with the Spike protein of the Omicron variant. Wild-type reference ACE2 (6M18) interaction with the Spike resulted in a binding energy of -17.6 and a binding affinity of 1.3×10^{-13} . A comparison between the wildtype ACE2 and modeled mutants indicates a lesser binding energy for the mutants implying an interaction with lesser strength which

may lead to less cellular invasion causing an infection. The Omicron variant was used in this study as most of the participants were infected with that variant of SARS-CoV-2.

Similar to C-C chemokine receptor type 5 (CCR5) mutation in Human Immunodeficiency Virus (HIV) infection which has been shown restrict HIV (J. Ni et al., 2018), the SNPs identified are host-receptor polymorphisms which may restrict a viral infection. Thus, highlighting the need to research further into host receptor polymorphisms when studying viral infections.

5.2 Conclusion

Twelve missense SNPs and one stop codon were identified within 11 exons of the ACE2 receptor across the four West African countries; Ghana, Benin, Nigeria and Guinea. The SNPs identified showed possible structural changes in ACE2 and lower binding energy and affinity for the Spike protein of SARS-CoV-2 (Omicron variant). Thus, polymorphisms present in the ACE2 receptor may have influenced viral entry, reducing COVID-19 infection and severity in West Africa.

5.3 Recommendations

- The use of fresh respiratory samples for RNA extraction will yield better amplification results as archived samples may have lost their RNA over a long storage period.
- An *in-vitro* validation of mutant models and Spike interaction will be a good option to confirm binding interactions observed after *in-silico* analysis.

- 1021 • Whole exome sequencing will be useful to explore mutants present in the other exons of
- 1022 *ACE2* and other proposed receptors which may have also affected viral entry.
- 1023



REFERENCES

- Abbad, A., Anga, L., Faouzi, A., Ioune, N., & Nourlil, J. (2021). Effect of identified non-synonymous mutations in DPP4 receptor binding residues among highly exposed human population in Morocco to MERS-CoV through computational approach. *PLOS ONE*, 16(10), e0258750. <https://doi.org/10.1371/journal.pone.0258750>
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics*, 76(1). <https://doi.org/10.1002/0471142905.hg0720s76>
- Ahmetaj-Shala, B., Vaja, R., Atanur, S. S., George, P. M., Kirkby, N. S., & Mitchell, J. A. (2020). Cardioresenal Tissues Express SARS-CoV-2 Entry Genes and Basigin (BSG/CD147) Increases With Age in Endothelial Cells. *JACC: Basic to Translational Science*, 5(11), 1111–1123. <https://doi.org/10.1016/j.jacbts.2020.09.010>
- Aitken, R. J. (2021). COVID-19 and human spermatozoa—Potential risks for infertility and sexual transmission? *Andrology*, 9(1), 48–52. <https://doi.org/10.1111/andr.12859>
- Alekseenko, A., Ignatov, M., Jones, G., Sabitova, M., & Kozakov, D. (2020). Protein–Protein and Protein–Peptide Docking with ClusPro Server. In D. Kihara (Ed.), *Protein Structure Prediction* (Vol. 2165, pp. 157–174). Springer US. https://doi.org/10.1007/978-1-0716-0708-4_9
- Ali, F., Elserafy, M., Alkordi, M. H., & Amin, M. (2020). ACE2 coding variants in different populations and their potential impact on SARS-CoV-2 binding affinity. *Biochemistry and Biophysics Reports*, 24, 100798. <https://doi.org/10.1016/j.bbrep.2020.100798>
- Artika, I. M., Dewantari, A. K., & Wiyatno, A. (2020). Molecular biology of coronaviruses: Current knowledge. *Heliyon*, 6(8), e04743. <https://doi.org/10.1016/j.heliyon.2020.e04743>

- 1046 Bassendine, M. F., Bridge, S. H., McCaughan, G. W., & Gorrell, M. D. (2020). COVID-19 and
1047 comorbidities: A role for dipeptidyl peptidase 4 (DPP4) in disease severity? *Journal of Diabetes*,
1048 *12*(9), 649–658. <https://doi.org/10.1111/1753-0407.13052>
- 1049 Benchling Biology Software, (2025). Retrieved from <https://benchling.com>.
- 1050 Bestle, D., Heindl, M. R., Limburg, H., Van Lam Van, T., Pilgram, O., Moulton, H., Stein, D. A.,
1051 Harges, K., Eickmann, M., Dolnik, O., Rohde, C., Klenk, H.-D., Garten, W., Steinmetzer, T., &
1052 Böttcher-Friebertshäuser, E. (2020). TMPRSS2 and furin are both essential for proteolytic
1053 activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, *3*(9), e202000786.
1054 <https://doi.org/10.26508/lsa.202000786>
- 1055 Boban, M. (2021). Novel coronavirus disease (COVID-19) update on epidemiology, pathogenicity,
1056 clinical course and treatments. *International Journal of Clinical Practice*, *75*(4).
1057 <https://doi.org/10.1111/ijcp.13868>
- 1058 Bonenfant, Q., Noé, L., & Touzet, H. (2023). Porechop_ABI: Discovering unknown adapters in
1059 Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinformatics*
1060 *Advances*, *3*(1), vbac085. <https://doi.org/10.1093/bioadv/vbac085>
- 1061 Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies,
1062 R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data.
1063 *GigaScience*, *10*(2), giab007. <https://doi.org/10.1093/gigascience/giab007>
- 1064 Bosch, B. J., Smits, S. L., & Haagmans, B. L. (2014). Membrane ectopeptidases targeted by human
1065 coronaviruses. *Current Opinion in Virology*, *6*, 55–60.
1066 <https://doi.org/10.1016/j.coviro.2014.03.011>

- 1067 Bosso, M., Thanaraj, T. A., Abu-Farha, M., Alanbaei, M., Abubaker, J., & Al-Mulla, F. (2020). The
1068 Two Faces of ACE2: The Role of ACE2 Receptor and Its Polymorphisms in Hypertension and
1069 COVID-19. *Molecular Therapy - Methods & Clinical Development*, 18, 321–327.
1070 <https://doi.org/10.1016/j.omtm.2020.06.017>
- 1071 Burtscher, J., Cappellano, G., Omori, A., Koshiba, T., & Millet, G. P. (2020). Mitochondria: In the
1072 Cross Fire of SARS-CoV-2 and Immunity. *iScience*, 23(10), 101631.
1073 <https://doi.org/10.1016/j.isci.2020.101631>
- 1074 Chen, Z., Mi, L., Xu, J., Yu, J., Wang, X., Jiang, J., Xing, J., Shang, P., Qian, A., Li, Y., Shaw, P.
1075 X., Wang, J., Duan, S., Ding, J., Fan, C., Zhang, Y., Yang, Y., Yu, X., Feng, Q., ... Zhu, P. (2005).
1076 Function of HAb18G/CD147 in Invasion of Host Cells by Severe Acute Respiratory Syndrome
1077 Coronavirus. *The Journal of Infectious Diseases*, 191(5), 755–760. <https://doi.org/10.1086/427811>
- 1078 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., &
1079 Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide
1080 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-
1081 3. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- 1082 Cui, X., Chen, W., Zhou, H., Gong, Y., Zhu, B., Lv, X., Guo, H., Duan, J., Zhou, J., Marcon, E., &
1083 Ma, H. (2021). Pulmonary Edema in COVID-19 Patients: Mechanisms and Treatment Potential.
1084 *Frontiers in Pharmacology*, 12, 664349. <https://doi.org/10.3389/fphar.2021.664349>
- 1085 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack:
1086 Visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669.
1087 <https://doi.org/10.1093/bioinformatics/bty149>

- 1088 De Coster, W., & Rademakers, R. (2023). NanoPack2: Population-scale evaluation of long-read
1089 sequencing data. *Bioinformatics*, 39(5), btad311. <https://doi.org/10.1093/bioinformatics/btad311>
- 1090 Dong, H., Zhang, H., Jalin, J., He, Z., Wang, R., Huang, L., Liu, Z., Zhang, S., Dai, B., & Li, D.
1091 (2023). Nucleocapsid proteins from human coronaviruses possess phase separation capabilities
1092 and promote FUS pathological aggregation. *Protein Science*, 32(12), e4826.
1093 <https://doi.org/10.1002/pro.4826>
- 1094 Duah-Quashie, N. O., Opoku-Agyeman, P., Lanza, M., & Rubio, J. M. (2024). Polymorphisms in
1095 the human angiotensin converting enzyme gene (ACE) linked to susceptibility of COVID-19 and
1096 malaria infections in the Ghanaian population. *Infection, Genetics and Evolution*, 119, 105568.
1097 <https://doi.org/10.1016/j.meegid.2024.105568>
- 1098 El-Sayed, A., & Kamel, M. (2021). Coronaviruses in humans and animals: The role of bats in viral
1099 evolution. *Environmental Science and Pollution Research*, 28(16), 19589–19600.
1100 <https://doi.org/10.1007/s11356-021-12553-1>
- 1101 Everest, H., Stevenson-Leggett, P., Bailey, D., Bickerton, E., & Keep, S. (2022). Known Cellular
1102 and Receptor Interactions of Animal and Human Coronaviruses: A Review. *Viruses*, 14(2), 351.
1103 <https://doi.org/10.3390/v14020351>
- 1104 Florindo, H. F., Kleiner, R., Vaskovich-Koubi, D., Acúrcio, R. C., Carreira, B., Yeini, E., Tiram,
1105 G., Liubomirski, Y., & Satchi-Fainaro, R. (2020). Immune-mediated approaches against COVID-
1106 19. *Nature Nanotechnology*, 15(8), 630–645. <https://doi.org/10.1038/s41565-020-0732-3>
- 1107 Fraser, B. J., Beldar, S., Seitova, A., Hutchinson, A., Mannar, D., Li, Y., Kwon, D., Tan, R., Wilson,
1108 R. P., Leopold, K., Subramaniam, S., Halabelian, L., Arrowsmith, C. H., & Bénard, F. (2022).

- 1109 Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. *Nature*
 1110 *Chemical Biology*, 18(9), 963–971. <https://doi.org/10.1038/s41589-022-01059-7>
- 1111 Fung, T. S., & Liu, D. X. (2019). Human Coronavirus: Host-Pathogen Interaction. *Annual Review*
 1112 *of Microbiology*, 73(1), 529–557. <https://doi.org/10.1146/annurev-micro-020518-115759>
- 1113 Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*
 1114 (No. arXiv:1207.3907). arXiv. <http://arxiv.org/abs/1207.3907>
- 1115 Gomes, C. P., Fernandes, D. E., Casimiro, F., Da Mata, G. F., Passos, M. T., Varela, P., Mastroianni-
 1116 Kirsztajn, G., & Pesquero, J. B. (2020). Cathepsin L in COVID-19: From Pharmacological
 1117 Evidences to Genetics. *Frontiers in Cellular and Infection Microbiology*, 10, 589505.
 1118 <https://doi.org/10.3389/fcimb.2020.589505>
- 1119 Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K.,
 1120 Karthikeyan, S., Iles, L., Pollard, M. O., Choudhury, A., Ritchie, G. R. S., Xue, Y., Asimit, J.,
 1121 Nsubuga, R. N., Young, E. H., Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., ... Sandhu, M.
 1122 S. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature*,
 1123 517(7534), 327–332. <https://doi.org/10.1038/nature13997>
- 1124 Hashimoto, R., Sakamoto, A., Deguchi, S., Yi, R., Sano, E., Hotta, A., Takahashi, K., Yamanaka,
 1125 S., & Takayama, K. (2021). Dual inhibition of TMPRSS2 and Cathepsin B prevents SARS-CoV-
 1126 2 infection in iPS cells. *Molecular Therapy - Nucleic Acids*, 26, 1107–1114.
 1127 <https://doi.org/10.1016/j.omtn.2021.10.016>
- 1128 Heller, L., Mota, C. R., & Greco, D. B. (2020). COVID-19 faecal-oral transmission: Are we asking
 1129 the right questions? *Science of The Total Environment*, 729, 138919.
 1130 <https://doi.org/10.1016/j.scitotenv.2020.138919>

- 1131 Hulswit, R. J. G., De Haan, C. A. M., & Bosch, B.-J. (2016). Coronavirus Spike Protein and
 1132 Tropism Changes. In *Advances in Virus Research* (Vol. 96, pp. 29–57). Elsevier.
 1133 <https://doi.org/10.1016/bs.aivir.2016.08.004>
- 1134 Humbert, M. V., Opurum, P. C., Brendish, N. J., Poole, S., He, P., Katis, I., Quaye, J., Bediako, Y.,
 1135 Duriez, P. J., Eason, R. W., Sones, C., Quaye, O., Awandare, G. A., Christodoulides, M., Clark, T.
 1136 W., Quashie, P. K., & McCormick, C. J. (2022). A SARS-CoV-2 nucleocapsid ELISA represents a
 1137 low-cost alternative to lateral flow testing for community screening in LMI countries. *Journal of*
 1138 *Infection*, 84(1), 48–55. <https://doi.org/10.1016/j.jinf.2021.08.049>
- 1139 Irkham, I., Ibrahim, A. U., Nwekwo, C. W., Al-Turjman, F., & Hartati, Y. W. (2022). Current
 1140 Technologies for Detection of COVID-19: Biosensors, Artificial Intelligence and Internet of
 1141 Medical Things (IoMT): Review. *Sensors*, 23(1), 426. <https://doi.org/10.3390/s23010426>
- 1142 Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into
 1143 cells. *Nature Reviews Molecular Cell Biology*, 23(1), 3–20. [https://doi.org/10.1038/s41580-021-](https://doi.org/10.1038/s41580-021-00418-x)
 1144 00418-x
- 1145 Jacobs, E. D., & Okeke, M. I. (2022). A critical evaluation of Nigeria's response to the first wave
 1146 of COVID-19. *Bulletin of the National Research Centre*, 46(1), 44.
 1147 <https://doi.org/10.1186/s42269-022-00729-9>
- 1148 Jain, S., Lamba, B. Y., Sharma, M., & Kumar, S. (2022). Introduction to COVID-19. In *COVID-*
 1149 *19 in the Environment* (pp. 3–16). Elsevier. <https://doi.org/10.1016/B978-0-323-90272-4.00002-6>
- 1150 Jalaeddine, N., Bouzid, A., Hachim, M., Sharif-Askari, N. S., Mahboub, B., Senok, A., Halwani,
 1151 R., Hamoudi, R. A., & Al Heialy, S. (2022). ACE2 polymorphisms impact COVID-19 severity in
 1152 obese patients. *Scientific Reports*, 12(1), 21491. <https://doi.org/10.1038/s41598-022-26072-7>

- 1153 Jones, G., Jindal, A., Ghani, U., Kotelnikov, S., Egbert, M., Hashemi, N., Vajda, S., Padhorny, D.,
1154 & Kozakov, D. (2022). Elucidation of protein function using computational docking and hotspot
1155 analysis by *ClusPro* and *FTMap*. *Acta Crystallographica Section D Structural Biology*, 78(6),
1156 690–697. <https://doi.org/10.1107/S2059798322002741>
- 1157 Kaidashev, I., Shlykova, O., Izmailova, O., Torubara, O., Yushchenko, Ya., Tyshkovska, T., Kyslyi,
1158 V., Belyaeva, A., & Maryniak, D. (2021). Host gene variability and SARS-CoV-2 infection: A
1159 review article. *Heliyon*, 7(8), e07863. <https://doi.org/10.1016/j.heliyon.2021.e07863>
- 1160 Kaklamanos, A., Belogiannis, K., Skendros, P., Gorgoulis, V. G., Vlachoyiannopoulos, P. G., &
1161 Tzioufas, A. G. (2021). COVID-19 Immunobiology: Lessons Learned, New Questions Arise.
1162 *Frontiers in Immunology*, 12, 719023. <https://doi.org/10.3389/fimmu.2021.719023>
- 1163 Kaltoum, A. B. O. (2021). Mutations and polymorphisms in genes involved in the infections by
1164 covid 19: A review. *Gene Reports*, 23, 101062. <https://doi.org/10.1016/j.genrep.2021.101062>
- 1165 Khayat, A. S., De Assumpção, P. P., Meireles Khayat, B. C., Thomaz Araújo, T. M., Batista-Gomes,
1166 J. A., Imbiriba, L. C., Ishak, G., De Assumpção, P. B., Moreira, F. C., Burbano, R. R., Ribeiro-dos-
1167 Santos, A., Ribeiro-dos-Santos, Â. K., Dos Santos, N. P. C., & Dos Santos, S. E. B. (2020). ACE2
1168 polymorphisms as potential players in COVID-19 outcome. *PLOS ONE*, 15(12), e0243887.
1169 <https://doi.org/10.1371/journal.pone.0243887>
- 1170 Kim, J., Kim, M., Kim, D., Park, S., Kang, M., Baek, K., Choi, J.-K., Maharjan, S., Akauliya, M.,
1171 Lee, Y., & Kwon, H.-J. (2022). Targeting the Interaction Between Spike Protein and Nucleocapsid
1172 Protein for Suppression and Detection of Human Coronavirus OC43. *Frontiers in Immunology*,
1173 13, 835333. <https://doi.org/10.3389/fimmu.2022.835333>

- 1174 Kotlyar, A. M., Grechukhina, O., Chen, A., Popkhadze, S., Grimshaw, A., Tal, O., Taylor, H. S., &
1175 Tal, R. (2021). Vertical transmission of coronavirus disease 2019: A systematic review and meta-
1176 analysis. *American Journal of Obstetrics and Gynecology*, 224(1), 35-53.e3.
1177 <https://doi.org/10.1016/j.ajog.2020.07.049>
- 1178 Kuba, K., Yamaguchi, T., & Penninger, J. M. (2021). Angiotensin-Converting Enzyme 2 (ACE2)
1179 in the Pathogenesis of ARDS in COVID-19. *Frontiers in Immunology*, 12, 732690.
1180 <https://doi.org/10.3389/fimmu.2021.732690>
- 1181 Kumar, A., Singh, R., Kaur, J., Pandey, S., Sharma, V., Thakur, L., Sati, S., Mani, S., Asthana, S.,
1182 Sharma, T. K., Chaudhuri, S., Bhattacharyya, S., & Kumar, N. (2021). Wuhan to World: The
1183 COVID-19 Pandemic. *Frontiers in Cellular and Infection Microbiology*, 11, 596201.
1184 <https://doi.org/10.3389/fcimb.2021.596201>
- 1185 Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. *Nature Reviews*
1186 *Microbiology*, 20(5), 270–284. <https://doi.org/10.1038/s41579-022-00713-0>
- 1187 Lawal, Y. (2021). Africa's low COVID-19 mortality rate: A paradox? *International Journal of*
1188 *Infectious Diseases*, 102, 118–122. <https://doi.org/10.1016/j.ijid.2020.10.038>
- 1189 Li, F. (2015). Receptor Recognition Mechanisms of Coronaviruses: A Decade of Structural
1190 Studies. *Journal of Virology*, 89(4), 1954–1964. <https://doi.org/10.1128/JVI.02615-14>
- 1191 Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18),
1192 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>

- 1193 Li, X., Luk, H. K. H., Lau, S. K. P., & Woo, P. C. Y. (2019). Human Coronaviruses: General
1194 Features. In *Reference Module in Biomedical Sciences* (p. B9780128012383957040). Elsevier.
1195 <https://doi.org/10.1016/B978-0-12-801238-3.95704-0>
- 1196 Liu, D. X., Liang, J. Q., & Fung, T. S. (2021). Human Coronavirus-229E, -OC43, -NL63, and -
1197 HKU1 (Coronaviridae). In *Encyclopedia of Virology* (pp. 428–440). Elsevier.
1198 <https://doi.org/10.1016/B978-0-12-809633-8.21501-X>
- 1199 López-Camacho, E., García-Godoy, M. J., García-Nieto, J., Nebro, A. J., & Aldana-Montes, J. F.
1200 (2016). A New Multi-objective Approach for Molecular Docking Based on RMSD and Binding
1201 Energy. In M. Botón-Fernández, C. Martín-Vide, S. Santander-Jiménez, & M. A. Vega-Rodríguez
1202 (Eds.), *Algorithms for Computational Biology* (Vol. 9702, pp. 65–77). Springer International
1203 Publishing. https://doi.org/10.1007/978-3-319-38827-4_6
- 1204 Lorenz, T. C. (2012). Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and
1205 Optimization Strategies. *Journal of Visualized Experiments*, 63, 3998.
1206 <https://doi.org/10.3791/3998>
- 1207 Marik, P. E., Iglesias, J., Varon, J., & Kory, P. (2021). A scoping review of the pathophysiology of
1208 COVID-19. *International Journal of Immunopathology and Pharmacology*, 35,
1209 205873842110480. <https://doi.org/10.1177/20587384211048026>
- 1210 McIntosh, K. (1974). Coronaviruses: A Comparative Review. In W. Arber, R. Haas, W. Henle, P.
1211 H. Hofschneider, N. K. Jerne, P. Koldovský, H. Koprowski, O. Maaløe, R. Rott, H. G. Schweiger,
1212 M. Sela, L. Syruček, P. K. Vogt, & E. Wecker (Eds.), *Current Topics in Microbiology and*
1213 *Immunology / Ergebnisse der Mikrobiologie und Immunitätsforschung* (pp. 85–129). Springer
1214 Berlin Heidelberg. https://doi.org/10.1007/978-3-642-65775-7_3

- 1215 Mehta, O. P., Bhandari, P., Raut, A., Kacimi, S. E. O., & Huy, N. T. (2021). Coronavirus Disease
1216 (COVID-19): Comprehensive Review of Clinical Presentation. *Frontiers in Public Health*, 8,
1217 582932. <https://doi.org/10.3389/fpubh.2020.582932>
- 1218 Melenotte, C., Silvin, A., Goubet, A.-G., Lahmar, I., Dubuisson, A., Zumla, A., Raoult, D., Merad,
1219 M., Gachot, B., Hénon, C., Solary, E., Fontenay, M., André, F., Maeurer, M., Ippolito, G.,
1220 Piacentini, M., Wang, F.-S., Ginhoux, F., Marabelle, A., ... Zitvogel, L. (2020). Immune responses
1221 during COVID-19 infection. *OncImmunology*, 9(1), 1807836.
1222 <https://doi.org/10.1080/2162402X.2020.1807836>
- 1223 Mensah, B. A., Ndong, I. C., Quashie, P. K., Guichet, E., Abuaku, B., Effah-Baafi, Y., Tapela, K.,
1224 Asiedu, K., Appiedu-Addo, S. N. A., Obheng, L. B., Amponsah, J. A., Kusi, K. A., Ofori, M.,
1225 Ayoubu, A., Courtin, D., Tahar, R., Delaporte, E., Awandare, G., & Ndam, N. T. (2022).
1226 Population-based sero-epidemiological investigation of the dynamics of SARS-CoV-2 infections
1227 in the Greater Accra Region of Ghana. *Scientific Reports*, 12(1), 21582.
1228 <https://doi.org/10.1038/s41598-022-25598-0>
- 1229 Merolli, A., Kasaei, L., Ramasamy, S., Kolloli, A., Kumar, R., Subbian, S., & Feldman, L. C.
1230 (2022). An intra-cytoplasmic route for SARS-CoV-2 transmission unveiled by Helium-ion
1231 microscopy. *Scientific Reports*, 12(1), 3794. <https://doi.org/10.1038/s41598-022-07867-0>
- 1232 Meskini, M., Rezghi Rami, M., Maroofi, P., Ghosh, S., Siadat, S. D., & Sheikhpour, M. (2021).
1233 An Overview on the Epidemiology and Immunology of COVID-19. *Journal of Infection and*
1234 *Public Health*, 14(10), 1284–1298. <https://doi.org/10.1016/j.jiph.2021.07.021>
- 1235 Minkoff, J. M., & tenOever, B. (2023). Innate immune evasion strategies of SARS-CoV-2. *Nature*
1236 *Reviews Microbiology*. <https://doi.org/10.1038/s41579-022-00839-1>

- 1237 Mistry, P., Barmania, F., Mellet, J., Peta, K., Strydom, A., Viljoen, I. M., James, W., Gordon, S., &
 1238 Pepper, M. S. (2022). SARS-CoV-2 Variants, Vaccines, and Host Immunity. *Frontiers in*
 1239 *Immunology*, 12, 809244. <https://doi.org/10.3389/fimmu.2021.809244>
- 1240 Mohammed, R. N., Tamjidifar, R., Rahman, H. S., Adili, A., Ghoreishizadeh, S., Saeedi, H.,
 1241 Thangavelu, L., Shomali, N., Aslaminabad, R., Marofi, F., Tahavvori, M., Danshina, S., Akbari,
 1242 M., & Ercan, G. (2022). A comprehensive review about immune responses and exhaustion during
 1243 coronavirus disease (COVID-19). *Cell Communication and Signaling*, 20(1), 79.
 1244 <https://doi.org/10.1186/s12964-022-00856-w>
- 1245 Monticelli, M., Mele, B. H., Andreotti, G., Cubellis, M. V., & Riccio, G. (2021). Why does SARS-
 1246 CoV-2 hit in different ways? Host genetic factors can influence the acquisition or the course of
 1247 COVID-19. *European Journal of Medical Genetics*, 64(6), 104227.
 1248 <https://doi.org/10.1016/j.ejmg.2021.104227>
- 1249 Morang'a, C. M., Ngoi, J. M., Gyamfi, J., Amuzu, D. S. Y., Nuerter, B. D., Soglo, P. M., Appiah,
 1250 V., Asante, I. A., Owusu-Oduro, P., Armoo, S., Adu-Gyasi, D., Amoako, N., Oliver-Commey, J.,
 1251 Owusu, M., Sylverken, A., Fenteng, E. D., M'cormack, V. V., Tei-Maya, F., Quansah, E. B., ...
 1252 Awandare, G. A. (2022). Genetic diversity of SARS-CoV-2 infections in Ghana from 2020-2021.
 1253 *Nature Communications*, 13(1), 2494. <https://doi.org/10.1038/s41467-022-30219-5>
- 1254 Mulabbi, E. N., Tweyongyere, R., & Byarugaba, D. K. (2021). The history of the emergence and
 1255 transmission of human coronaviruses. *Onderstepoort Journal of Veterinary Research*, 88(1).
 1256 <https://doi.org/10.4102/ojvr.v88i1.1872>
- 1257 Nassar, A., Ibrahim, I. M., Amin, F. G., Magdy, M., Elgharib, A. M., Azzam, E. B., Nasser, F.,
 1258 Yousry, K., Shamkh, I. M., Mahdy, S. M., & Elfiky, A. A. (2021). A Review of Human

- 1259 Coronaviruses' Receptors: The Host-Cell Targets for the Crown Bearing Viruses. *Molecules*,
1260 26(21), 6455. <https://doi.org/10.3390/molecules26216455>
- 1261 Ni, J., Wang, D., & Wang, S. (2018). The CCR5-Delta32 Genetic Polymorphism and HIV-1
1262 Infection Susceptibility: A Meta-analysis. *Open Medicine (Warsaw, Poland)*, 13, 467–474.
1263 <https://doi.org/10.1515/med-2018-0062>
- 1264 Ni, W., Yang, X., Yang, D., Bao, J., Li, R., Xiao, Y., Hou, C., Wang, H., Liu, J., Yang, D., Xu, Y.,
1265 Cao, Z., & Gao, Z. (2020). Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19.
1266 *Critical Care*, 24(1), 422. <https://doi.org/10.1186/s13054-020-03120-0>
- 1267 Niemi, M. E. K., Daly, M. J., & Ganna, A. (2022). The human genetic epidemiology of COVID-
1268 19. *Nature Reviews Genetics*, 23(9), 533–546. <https://doi.org/10.1038/s41576-022-00478-5>
- 1269 Njenga, M. K., Dawa, J., Nanyingi, M., Gachohi, J., Ngere, I., Letko, M., Otieno, C. F., Gunn, B.
1270 M., & Osoro, E. (2020). Why is There Low Morbidity and Mortality of COVID-19 in Africa? *The*
1271 *American Journal of Tropical Medicine and Hygiene*, 103(2), 564–569.
1272 <https://doi.org/10.4269/ajtmh.20-0474>
- 1273 Oduro-Mensah, D., Oduro-Mensah, E., Quashie, P., Awandare, G., & Okine, L. (2022). Explaining
1274 the unexpected COVID-19 trends and potential impact across Africa. *F1000Research*, 10, 1177.
1275 <https://doi.org/10.12688/f1000research.74363.2>
- 1276 Paces, J., Strizova, Z., Smrz, D., & Cerny, J. (2020). COVID-19 and the Immune System.
1277 *Physiological Research*, 379–388. <https://doi.org/10.33549/physiolres.934492>
- 1278 Parbie, P. K., Mingle, J. A., Ntiri, M., Adjabeng, M., Bonney, K., Asante, I., Bonney, E. Y.,
1279 Brightson, K., Sarkodie, B., Koram, K., Ohene, S.-A., Ayi, B. S., & Ampofo, W. K. (2020). Human

- 1280 coronaviruses in persons with acute respiratory infections in Ghana. *Health Sciences*
 1281 *Investigations Journal*, 5–11. <https://doi.org/10.46829/hsijournal.2020.6.1.1.5-11>
- 1282 Pepe, A., Pietropaoli, S., Vos, M., Barba-Spaeth, G., & Zurzolo, C. (2022). Tunneling nanotubes
 1283 provide a route for SARS-CoV-2 spreading. *Science Advances*, 8(29), eabo0171.
 1284 <https://doi.org/10.1126/sciadv.abo0171>
- 1285 Pišlar, A., Mitrović, A., Sabotič, J., Pečar Fonović, U., Perišić Nanut, M., Jakoš, T., Senjor, E., &
 1286 Kos, J. (2020). The role of cysteine peptidases in coronavirus cell entry and replication: The
 1287 therapeutic potential of cathepsin inhibitors. *PLoS Pathogens*, 16(11), e1009013.
 1288 <https://doi.org/10.1371/journal.ppat.1009013>
- 1289 Polidoro, R. B., Hagan, R. S., De Santis Santiago, R., & Schmidt, N. W. (2020). Overview:
 1290 Systemic Inflammatory Response Derived From Lung Injury Caused by SARS-CoV-2 Infection
 1291 Explains Severe Outcomes in COVID-19. *Frontiers in Immunology*, 11, 1626.
 1292 <https://doi.org/10.3389/fimmu.2020.01626>
- 1293 Pradhan, M., Shah, K., Alexander, A., Ajazuddin, Minz, S., Singh, M. R., Singh, D., Yadav, K., &
 1294 Chauhan, N. S. (2021). COVID-19: Clinical presentation and detection methods. *Journal of*
 1295 *Immunoassay and Immunochemistry*, 43(1), 1951291.
 1296 <https://doi.org/10.1080/15321819.2021.1951291>
- 1297 Prasad, K., Ahamad, S., Gupta, D., & Kumar, V. (2021). Targeting cathepsins: A potential link
 1298 between COVID-19 and associated neurological manifestations. *Heliyon*, 7(10), e08089.
 1299 <https://doi.org/10.1016/j.heliyon.2021.e08089>

- 1300 Primorac, D., Vrdoljak, K., Brlek, P., Pavelić, E., Molnar, V., Matišić, V., Erceg Ivkošić, I., &
 1301 Parčina, M. (2022). Adaptive Immune Responses and Immunity to SARS-CoV-2. *Frontiers in*
 1302 *Immunology*, 13, 848582. <https://doi.org/10.3389/fimmu.2022.848582>
- 1303 Qi, F., Qian, S., Zhang, S., & Zhang, Z. (2020). Single cell RNA sequencing of 13 human tissues
 1304 identify cell types and receptors of human coronaviruses. *Biochemical and Biophysical Research*
 1305 *Communications*, 526(1), 135–140. <https://doi.org/10.1016/j.bbrc.2020.03.044>
- 1306 Quashie, P. K., Mutungi, J. K., Dzabeng, F., Oduro-Mensah, D., Oporum, P. C., Tapela, K.,
 1307 Udoakang, A. J., WACCBIP COVID-19 Team, Asante, I., Paemka, L., Kumi-Ansah, F., Quaye,
 1308 O., Amoako, E., Armah, R., Kilba, C., Boateng, N. A., Ofori, M., Kyei, G. B., Bediako, Y., ...
 1309 Awandare, G. A. (2021). *Trends of SARS-CoV-2 antibody prevalence in selected regions across*
 1310 *Ghana* [Preprint]. *Epidemiology*. <https://doi.org/10.1101/2021.04.25.21256067>
- 1311 Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. A., Dijkman, R., Muth, D.,
 1312 Demmers, J. A. A., Zaki, A., Fouchier, R. A. M., Thiel, V., Drosten, C., Rottier, P. J. M., Osterhaus,
 1313 A. D. M. E., Bosch, B. J., & Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional
 1314 receptor for the emerging human coronavirus-EMC. *Nature*, 495(7440), 251–254.
 1315 <https://doi.org/10.1038/nature12005>
- 1316 Rohaim, M. A., El Naggar, R. F., Clayton, E., & Munir, M. (2021). Structural and functional
 1317 insights into non-structural proteins of coronaviruses. *Microbial Pathogenesis*, 150, 104641.
 1318 <https://doi.org/10.1016/j.micpath.2020.104641>
- 1319 Rokni, M., Heidari Nia, M., Sarhadi, M., Mirinejad, S., Sargazi, S., Moudi, M., Saravani, R.,
 1320 Rahdar, S., & Kargar, M. (2022). Association of TMPRSS2 Gene Polymorphisms with COVID-
 1321 19 Severity and Mortality: A Case-Control Study with Computational Analyses. *Applied*

- 1322 *Biochemistry and Biotechnology*, 194(8), 3507–3526. <https://doi.org/10.1007/s12010-022-03885->
- 1323 w
- 1324 Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein
- 1325 structure and function prediction. *Nature Protocols*, 5(4), 725–738.
- 1326 <https://doi.org/10.1038/nprot.2010.5>
- 1327 Rubio-Casillas, A., Redwan, E. M., & Uversky, V. N. (2022). SARS-CoV-2: A Master of Immune
- 1328 Evasion. *Biomedicines*, 10(6), 1339. <https://doi.org/10.3390/biomedicines10061339>
- 1329 Samavati, L., & Uhal, B. D. (2020). ACE2, Much More Than Just a Receptor for SARS-COV-2.
- 1330 *Frontiers in Cellular and Infection Microbiology*, 10, 317.
- 1331 <https://doi.org/10.3389/fcimb.2020.00317>
- 1332 Sandi, J. D., Levy, J. I., Tapela, K., Zeller, M., Yeboah, J. A., Saka, D. F., Grant, D. S., Awandare,
- 1333 G. A., Quashie, P. K., Andersen, K. G., & Paemka, L. (2024). Upper Airway Epithelial Tissue
- 1334 Transcriptome Analysis Reveals Immune Signatures Associated with COVID-19 Severity in
- 1335 Ghanaians. *Journal of Immunology Research*, 2024, 1–13. <https://doi.org/10.1155/2024/6668017>
- 1336 Sargsyan, K., Grauffel, C., & Lim, C. (2017). How Molecular Size Impacts RMSD Applications
- 1337 in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 13(4), 1518–
- 1338 1524. <https://doi.org/10.1021/acs.jctc.7b00028>
- 1339 Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory Syndrome
- 1340 Coronavirus-2. *Archives of Medical Research*, 51(6), 482–491.
- 1341 <https://doi.org/10.1016/j.arcmed.2020.05.012>

- 1342 Schoeman, D., & Fielding, B. C. (2020). Is There a Link Between the Pathogenic Human
1343 Coronavirus Envelope Protein and Immunopathology? A Review of the Literature. *Frontiers in*
1344 *Microbiology*, *11*, 2086. <https://doi.org/10.3389/fmicb.2020.02086>
- 1345 Schoeman, D., Gordon, B., & Fielding, B. C. (2022). Coronaviruses. In *Encyclopedia of Infection*
1346 *and Immunity* (pp. 241–258). Elsevier. <https://doi.org/10.1016/B978-0-12-818731-9.00052-5>
- 1347 Sebastián-Martín, A., Sánchez, B. G., Mora-Rodríguez, J. M., Bort, A., & Díaz-Laviada, I. (2022).
1348 Role of Dipeptidyl Peptidase-4 (DPP4) on COVID-19 Physiopathology. *Biomedicines*, *10*(8),
1349 2026. <https://doi.org/10.3390/biomedicines10082026>
- 1350 Shi, Y., Wang, G., Cai, X., Deng, J., Zheng, L., Zhu, H., Zheng, M., Yang, B., & Chen, Z. (2020).
1351 An overview of COVID-19. *Journal of Zhejiang University-SCIENCE B*, *21*(5), 343–360.
1352 <https://doi.org/10.1631/jzus.B2000083>
- 1353 Shilts, J., Crozier, T. W. M., Greenwood, E. J. D., Lehner, P. J., & Wright, G. J. (2021). No evidence
1354 for basigin/CD147 as a direct SARS-CoV-2 spike binding receptor. *Scientific Reports*, *11*(1), 413.
1355 <https://doi.org/10.1038/s41598-020-80464-1>
- 1356 Silva, M. J. A., Ribeiro, L. R., Lima, K. V. B., & Lima, L. N. G. C. (2022). Adaptive immunity to
1357 SARS-CoV-2 infection: A systematic review. *Frontiers in Immunology*, *13*, 1001198.
1358 <https://doi.org/10.3389/fimmu.2022.1001198>
- 1359 Smith, N., Goncalves, P., Charbit, B., Grzelak, L., Beretta, M., Planchais, C., Bruel, T., Rouilly,
1360 V., Bondet, V., Hadjadj, J., Yatim, N., Pere, H., Merkling, S. H., Ghoulane, A., Kernéis, S., Rieux-
1361 Laucat, F., Terrier, B., Schwartz, O., Mouquet, H., ... Di Santo, J. P. (2021). Distinct systemic and
1362 mucosal immune responses during acute SARS-CoV-2 infection. *Nature Immunology*, *22*(11),
1363 1428–1439. <https://doi.org/10.1038/s41590-021-01028-7>

- 1364 Sohrab, S. S., Alsaqaf, F., Hassan, A. M., Tolah, A. M., Bajrai, L. H., & Azhar, E. I. (2024).
 1365 Genomic Diversity and Recombination Analysis of the Spike Protein Gene from Selected Human
 1366 Coronaviruses. *Biology*, 13(4), 282. <https://doi.org/10.3390/biology13040282>
- 1367 Sorokina, M., M. C. Teixeira, J., Barrera-Vilarmau, S., Paschke, R., Papasotiriou, I., Rodrigues, J.
 1368 P. G. L. M., & Kastiris, P. L. (2020). Structural models of human ACE2 variants with SARS-CoV-
 1369 2 Spike protein for structure-based drug design. *Scientific Data*, 7(1), 309.
 1370 <https://doi.org/10.1038/s41597-020-00652-6>
- 1371 Srivastava, A., Bandopadhyay, A., Das, D., Pandey, R. K., Singh, V., Khanam, N., Srivastava, N.,
 1372 Singh, P. P., Dubey, P. K., Pathak, A., Gupta, P., Rai, N., Sultana, G. N. N., & Chaubey, G. (2020).
 1373 Genetic Association of ACE2 rs2285666 Polymorphism With COVID-19 Spatial Distribution in
 1374 India. *Frontiers in Genetics*, 11, 564741. <https://doi.org/10.3389/fgene.2020.564741>
- 1375 Suryamohan, K., Diwanji, D., Stawiski, E. W., Gupta, R., Miersch, S., Liu, J., Chen, C., Jiang, Y.-
 1376 P., Fellouse, F. A., Sathirapongsasuti, J. F., Albers, P. K., Deepak, T., Saberianfar, R., Ratan, A.,
 1377 Washburn, G., Mis, M., Santhosh, D., Somasekar, S., Hiranjith, G. H., ... Seshagiri, S. (2021).
 1378 Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2.
 1379 *Communications Biology*, 4(1), 475. <https://doi.org/10.1038/s42003-021-02030-3>
- 1380 Taboe, H. B., Salako, K. V., Tison, J. M., Ngonghala, C. N., & Glèlè Kakaï, R. (2020). Predicting
 1381 COVID-19 spread in the face of control measures in West Africa. *Mathematical Biosciences*, 328,
 1382 108431. <https://doi.org/10.1016/j.mbs.2020.108431>
- 1383 Tallei, T. E., Fatimawali, Yelnetty, A., Idroes, R., Kusumawaty, D., Emran, T. B., Yesiloglu, T. Z.,
 1384 Sippl, W., Mahmud, S., Alqahtani, T., Alqahtani, A. M., Asiri, S., Rahmatullah, M., Jahan, R.,
 1385 Khan, Md. A., & Celik, I. (2021). An Analysis Based on Molecular Docking and Molecular

- 1386 Dynamics Simulation Study of Bromelain as Anti-SARS-CoV-2 Variants. *Frontiers in*
1387 *Pharmacology*, 12, 717757. <https://doi.org/10.3389/fphar.2021.717757>
- 1388 Tang, X., Qian, Z., Lu, X., & Lu, J. (2023). Adaptive Evolution of the Spike Protein in
1389 Coronaviruses. *Molecular Biology and Evolution*, 40(4), msad089.
1390 <https://doi.org/10.1093/molbev/msad089>
- 1391 Tapela, K., Oporum, P. C., Nuokpem, F. Y., Tetteh, B., Siaw, G. K., Humbert, M. V., Tawiah-Eshun,
1392 S., Barakisu, A. I., Asiedu, K., Arhin, S. K., Manu, A. A., Appiedu-Addo, S. N. A., Obheng, L.,
1393 Quansah, D., Languon, S., Anyigba, C., Dosoo, D., Edu, N. K. O., Oduro-Mensah, D., ... Quashie,
1394 P. K. (2023). Development of an Affordable ELISA Targeting the SARS-CoV-2 Nucleocapsid and
1395 Its Application to Samples from the Ongoing COVID-19 Epidemic in Ghana. *Molecular Diagnosis*
1396 *& Therapy*, 27(5), 583–592. <https://doi.org/10.1007/s40291-023-00655-0>
- 1397 Tapela, K., Oyawoye, F. O., Olwal, C. O., Oporum, P. C., Amponsah, J. A., Segbedzi, K. A. L.,
1398 Tetteh, B., Kumi-Ansah, F., Mutungi, J. K., Obodai, E., Amoako, E., Agyemang, S., Ndam, N. T.,
1399 Ampofo, W. K., Rayner, J. C., Awandare, G. A., Paemka, L., Bediako, Y., & Quashie, P. K. (2022).
1400 Probing SARS-CoV-2-positive plasma to identify potential factors correlating with mild COVID-
1401 19 in Ghana, West Africa. *BMC Medicine*, 20(1), 370. [https://doi.org/10.1186/s12916-022-02571-](https://doi.org/10.1186/s12916-022-02571-2)
1402 2
- 1403 Tapela, K., Prah, D. A., Tetteh, B., Nuokpem, F., Dosoo, D., Coker, A., Kumi-Ansah, F., Amoako,
1404 E., Assah, K. O., Kilba, C., Nyakoe, N., Quansah, D., Languon, S., Anyigba, C. A., Ansah, F.,
1405 Agyeman, S., Owusu, I. A., Schneider, K., Ampofo, W. K., ... Bediako, Y. (2024). Cellular immune
1406 response to SARS-CoV-2 and clinical presentation in individuals exposed to endemic malaria. *Cell*
1407 *Reports*, 43(8), 114533. <https://doi.org/10.1016/j.celrep.2024.114533>

- 1408 The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi,
1409 E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer,
1410 T., Fan, J., Garmiri, P., Da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., ... Zhang, J. (2023).
1411 UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–
1412 D531. <https://doi.org/10.1093/nar/gkac1052>
- 1413 Tirelli, C., De Amici, M., Albrici, C., Mira, S., Nalesso, G., Re, B., Corsico, A. G., Mondoni, M.,
1414 & Centanni, S. (2023). Exploring the Role of Immune System and Inflammatory Cytokines in
1415 SARS-CoV-2 Induced Lung Disease: A Narrative Review. *Biology*, 12(2), 177.
1416 <https://doi.org/10.3390/biology12020177>
- 1417 Tortorici, M. A., & Veeler, D. (2019). Structural insights into coronavirus entry. In *Advances in*
1418 *Virus Research* (Vol. 105, pp. 93–116). Elsevier. <https://doi.org/10.1016/bs.aivir.2019.08.002>
- 1419 Troyer, Z., Alhusaini, N., Tabler, C. O., Sweet, T., De Carvalho, K. I. L., Schlatzer, D. M., Carias,
1420 L., King, C. L., Matreyek, K., & Tilton, J. C. (2021). Extracellular vesicles carry SARS-CoV-2
1421 spike protein and serve as decoys for neutralizing antibodies. *Journal of Extracellular Vesicles*,
1422 10(8), e12112. <https://doi.org/10.1002/jev2.12112>
- 1423 Tsinda, E. K., & Mmbando, G. S. (2021). Recent updates on the possible reasons for the low
1424 incidence and morbidity of COVID-19 cases in Africa. *Bulletin of the National Research Centre*,
1425 45(1), 133. <https://doi.org/10.1186/s42269-021-00589-9>
- 1426 Vargas-Rodriguez, J. R., Garza-Veloz, I., Flores-Morales, V., Badillo-Almaraz, J. I., Rocha-Pizaña,
1427 M. R., Valdés-Aguayo, J. J., & Martinez-Fierro, M. L. (2022). Hyperglycemia and Angiotensin-
1428 Converting Enzyme 2 in Pulmonary Function in the Context of SARS-CoV-2 Infection. *Frontiers*
1429 *in Medicine*, 8, 758414. <https://doi.org/10.3389/fmed.2021.758414>

- 1430 Wong, N. A., & Saier, M. H. (2021). The SARS-Coronavirus Infection Cycle: A Survey of Viral
1431 Membrane Proteins, Their Functional Interactions and Pathogenesis. *International Journal of*
1432 *Molecular Sciences*, 22(3), 1308. <https://doi.org/10.3390/ijms22031308>
- 1433 Woo, P. C. Y., De Groot, R. J., Haagmans, B., Lau, S. K. P., Neuman, B. W., Perlman, S., Sola, I.,
1434 Van Der Hoek, L., Wong, A. C. P., & Yeh, S.-H. (2023). ICTV Virus Taxonomy Profile:
1435 Coronaviridae 2023: This article is part of the ICTV Virus Taxonomy Profiles collection. *Journal*
1436 *of General Virology*, 104(4). <https://doi.org/10.1099/jgv.0.001843>
- 1437 Wulandari, L., Hamidah, B., Pakpahan, C., Damayanti, N. S., Kurniati, N. D., Adiatmaja, C. O.,
1438 Wigianita, M. R., Soedarsono, Husada, D., Tinduh, D., Prakoeswa, C. R. S., Endaryanto, A.,
1439 Puspaningsih, N. N. T., Mori, Y., Lusida, M. I., Shimizu, K., & Oceandy, D. (2021). Initial study
1440 on TMPRSS2 p.Val160Met genetic variant in COVID-19 patients. *Human Genomics*, 15(1), 29.
1441 <https://doi.org/10.1186/s40246-021-00330-7>
- 1442 Xue, L. C., Rodrigues, J. P., Kastitis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: A
1443 web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*,
1444 32(23), 3676–3678. <https://doi.org/10.1093/bioinformatics/btw514>
- 1445 Yaghoobi, A., Lord, J. S., Rezaiezhadeh, J. S., Yekaninejad, M. S., Amini, M., & Izadi, P. (2023).
1446 TMPRSS2 polymorphism (rs12329760) and the severity of the COVID-19 in Iranian population.
1447 *PLOS ONE*, 18(2), e0281750. <https://doi.org/10.1371/journal.pone.0281750>
- 1448 Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020a). Structural basis for the recognition
1449 of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), 1444–1448.
1450 <https://doi.org/10.1126/science.abb2762>

- 1451 Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020b). Structural basis for the recognition
1452 of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), 1444–1448.
1453 <https://doi.org/10.1126/science.abb2762>
- 1454 Yang, J., & Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Current*
1455 *Protocols in Bioinformatics*, 52(1). <https://doi.org/10.1002/0471250953.bi0508s52>
- 1456 Yeung, M. L., Teng, J. L. L., Jia, L., Zhang, C., Huang, C., Cai, J.-P., Zhou, R., Chan, K.-H., Zhao,
1457 H., Zhu, L., Siu, K.-L., Fung, S.-Y., Yung, S., Chan, T. M., To, K. K.-W., Chan, J. F.-W., Cai, Z.,
1458 Lau, S. K. P., Chen, Z., ... Yuen, K.-Y. (2021). Soluble ACE2-mediated cell entry of SARS-CoV-
1459 2 via interaction with proteins related to the renin-angiotensin system. *Cell*, 184(8), 2212-
1460 2228.e12. <https://doi.org/10.1016/j.cell.2021.02.053>
- 1461 Yoo, J.-S., Sasaki, M., Cho, S. X., Kasuga, Y., Zhu, B., Ouda, R., Orba, Y., De Figueiredo, P.,
1462 Sawa, H., & Kobayashi, K. S. (2021). SARS-CoV-2 inhibits induction of the MHC class I pathway
1463 by targeting the STAT1-IRF1-NLRC5 axis. *Nature Communications*, 12(1), 6602.
1464 <https://doi.org/10.1038/s41467-021-26910-8>
- 1465 Yuan, J., Qian, H., Cao, S., Dong, B., Yan, X., Luo, S., Zhou, M., Zhou, S., Ning, B., & Zhao, L.
1466 (2021). Is there possibility of vertical transmission of COVID-19: A systematic review.
1467 *Translational Pediatrics*, 10(2), 423–434. <https://doi.org/10.21037/tp-20-144>
- 1468 Yüce, M., Filiztekin, E., & Özkaya, K. G. (2021). COVID-19 diagnosis—A review of current
1469 methods. *Biosensors and Bioelectronics*, 172, 112752. <https://doi.org/10.1016/j.bios.2020.112752>
- 1470 Zhang, Y., Chen, Y., Li, Y., Huang, F., Luo, B., Yuan, Y., Xia, B., Ma, X., Yang, T., Yu, F., Liu, J.,
1471 Liu, B., Song, Z., Chen, J., Yan, S., Wu, L., Pan, T., Zhang, X., Li, R., ... Zhang, H. (2021). The
1472 ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I.

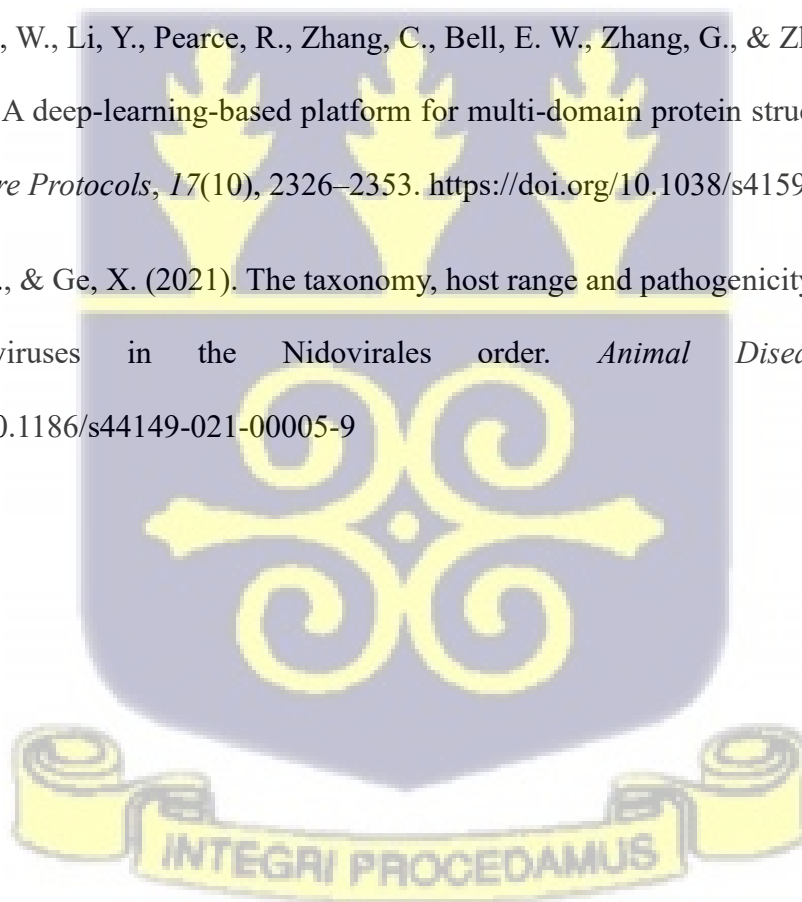
1473 *Proceedings of the National Academy of Sciences*, 118(23), e2024202118.
1474 <https://doi.org/10.1073/pnas.2024202118>

1475 Zhao, X., Liu, S., Yin, Y., Zhang, T. (Tim), & Chen, Q. (2022). Airborne transmission of COVID-
1476 19 virus in enclosed spaces: An overview of research methods. *Indoor Air*, 32(6).
1477 <https://doi.org/10.1111/ina.13056>

1478 Zheng, Y., Shang, J., Yang, Y., Liu, C., Wan, Y., Geng, Q., Wang, M., Baric, R., & Li, F. (2018).
1479 Lysosomal Proteases Are a Determinant of Coronavirus Tropism. *Journal of Virology*, 92(24),
1480 e01504-18. <https://doi.org/10.1128/JVI.01504-18>

1481 Zhou, X., Zheng, W., Li, Y., Pearce, R., Zhang, C., Bell, E. W., Zhang, G., & Zhang, Y. (2022). I-
1482 TASSER-MTD: A deep-learning-based platform for multi-domain protein structure and function
1483 prediction. *Nature Protocols*, 17(10), 2326–2353. <https://doi.org/10.1038/s41596-022-00728-0>

1484 Zhou, Z., Qiu, Y., & Ge, X. (2021). The taxonomy, host range and pathogenicity of coronaviruses
1485 and other viruses in the Nidovirales order. *Animal Diseases*, 1(1), 5.
1486 <https://doi.org/10.1186/s44149-021-00005-9>



APPENDIX

1490

1491 Appendix 1: Primer name, sequence and nuclease-free water used in a dilution of lyophilized

1492 product

Primer name	Primer sequence	Nuclease-free water added
ACE2_1F	<i>CTCTGGCAATGAGGACACTGAG</i>	801.99
ACE2_1R	<i>TCGGCTTCGTGGTTAAACTTGT</i>	738.50
ACE2_2F	<i>CTTGTTGCTGTAACTGCTGCTC</i>	667.76
ACE2_2R	<i>AGCATTCTTGTTGGATTATCTGGGT</i>	769.58
ACE2_3F	<i>ACAAGAGCAAACGGTTGAACAC</i>	809.66
ACE2_3R	<i>GCCCTCACATAGGCATGAAGAT</i>	868.73
ACE2_4F	<i>CGGCCAGTTGATTGAAGATGTG</i>	824.17
ACE2_4R	<i>AAGCTGTGGGATGGCAGACT</i>	856.62
ACE2_6F	<i>CATTGGTCTTCTGTCACCCGAT</i>	914.76
ACE2_6R	<i>ACAGTTTCTGTCCAGCTTCTGT</i>	653.60
ACE2_7F	<i>AGAAGCACTTTGTCAAGCAGCT</i>	877.37
ACE2_7R	<i>TTTACTTTTAAAAAGTACTGCCTCATAGC</i>	854.02
ACE2_8F	<i>TCAGCTCTTGGAGATAAAGCATATGA</i>	886.35
ACE2_8R	<i>TCAGGATGACAATGCCAACCAC</i>	853.7
ACE2_9F	<i>TAACCAGCCCCCTGTTTCCATA</i>	563.35
ACE2_9R	<i>TGTCAATGAAGATGCTCTCTCCTT</i>	845.74
ACE2_10F	<i>CTTGAGGTGATTTTGTGTATGTAAATGT</i>	771.01
ACE2_10R	<i>CAATCCTTGCAGCTACACCAGT</i>	761.73
ACE2_11F	<i>TGGAGCAAGTGTTGGATCTTGT</i>	896.8
ACE2_11R	<i>TGGAATTTCCATTCACTCCACAGT</i>	756.91
ACE2_12F	<i>AGCTCACTTTCAATTAAATCCATTGTCAA</i>	649.66
ACE2_12F	<i>TTGAGTGTCTGCTTTGTGCTCA</i>	604.80

1493

1494

1495

1496

1497 Appendix 2: Selected ACE2 primers used in PCR amplification and the specific binding positions
1498 on the cDNA product of the six ACE2 isoforms

Specific binding positions of primers

	ACE2 Variant	ACE2 Variant	ACE2 Variant	ACE2 Variant	ACE2 Variant	ACE2 Variant
Primer name	1	2	3	4	5	6
ACE2_4F	727	984	727	727	NA	856
ACE2_4R	-1075	-1332	-1075	-1075	NA	-1204
ACE2_6F	1330	1587	1330	1330	351	1459
ACE2_6R	-1691	-1948	-1691	-1691	-712	-1820
ACE2_7F	1648	1905	1648	1648	669	1777
ACE2_7R	-1997	-2254	-1997	NA	-1018	NA
ACE2_8F	1950	2207	1950	NA	971	NA
ACE2_8R	-2309	-2566	-2309	-1976	-1330	-2105

1499

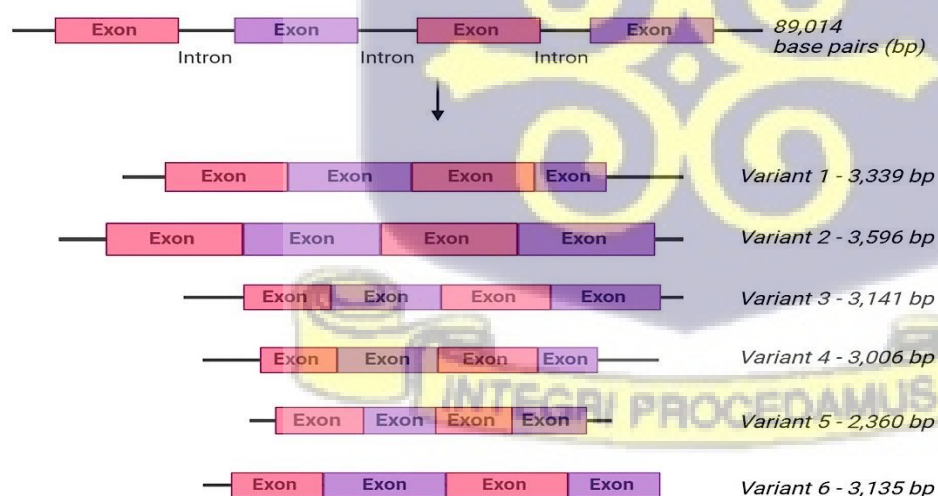
1500

1501

Appendix 3: schematic of ACE2 gene and transcript variants produced after splicing

1502

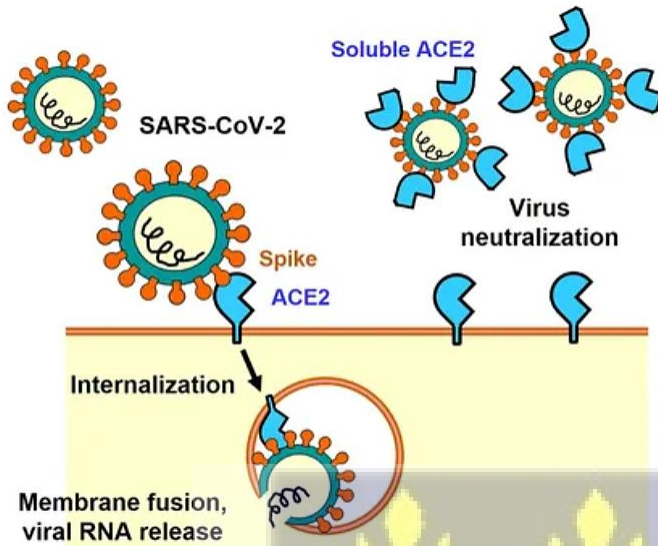
1503



1511

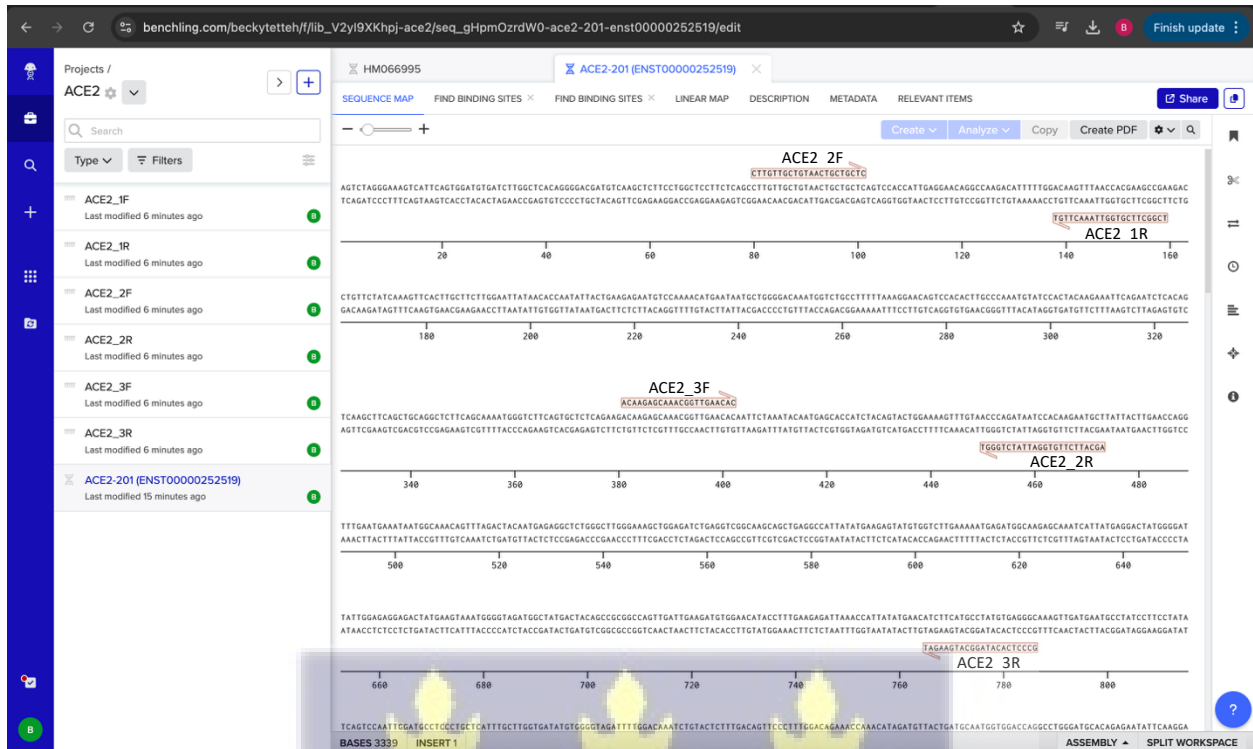
1512

1513 Appendix 4: ACE2 mediated cell entry of SARS-CoV-2 and inhibition of virus infection by
1514 recombinant soluble ACE2 protein



1515
1516 Source: (Kuba et al., 2021)

1517
1518



Appendix 5: Benchling software showing the attachment of ACE2 primers to ACE2 gene sequence (Ensembl: ENSG00000130234).

The full length ACE2 template (Ensembl: ENSG00000130234) was uploaded on the Benchling software showing the attachment of five ACE2 primers used in this project. Although primer ACE_1F was included in the primers attached, the software shows no binding between the ACE2 template and the ACE2_1F primer. Thus, no amplification was observed for ACE2 primer pair 1.

1529 Appendix 6: Total number of samples which were processed for sequencing and total number of
1530 samples which passed for sequencing after quality control from each centre

ACE Partner	Country	Number of samples				Samples successfully sequenced
		VTM	cDNA	RNA	Sputum	
CEA-PCMT	Guinea Conakry	0	40	0	0	21
ACEGID	Nigeria	40	0	0	0	21
ACEPHAP	Nigeria	30*	5*	10	0	9
ACENTDFB	Nigeria	40	0	0	0	29
ACEMFS	Nigeria	40	0	0	0	12
IRCB/CERPAGE	Benin	40*	0	40	0	4
CEA-ITECH/MTV	Burkina Faso	0	0	0	0	NA
WACCBIP	Ghana	82*	0	0	8*	23

1531
1532
1533
1534
1535
1536
1537
1538

