# GENOMIC AND IMMUNOINFORMATIC CHARACTERIZATION OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 VARIANTS IN NORTH-RIFT KENYA AND THEIR OVERLAP WITH ENDEMIC HUMAN CORONAVIRUSES

**BY**

**ELIUS MURIUNGI MBOGORI**

**A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY DEGREE IN MEDICAL IMMUNOLOGY TO THE COLLEGE OF HEALTH SCIENCES, SCHOOL OF MEDICINE, DEPARTMENT OF PATHOLOGY, MOI UNIVERSITY**

## DECLARATION

I declare that the work presented herein is my original work and has not been presented for any award in any other university or institution.

**Elius Muriungi Mbogori**

**PHD/IM/4291/20**

Signature:…… Date:……..

**Declaration by the Supervisors**

This thesis has been submitted for examination with our approval as university supervisors.

**Prof. Kirtika Patel (Ph.D.)**

Department of Pathology,

Moi University.

Signature:…… Date:……..

**Dr. Richard Biegon (Ph.D.)**

Department of Pathology,

Moi University.

Signature:…… Date:……..

## DEDICATION

Dedicated to my wife, Caren and my three lovely sons Lee, Leonel and Lennox, who gave me great support to complete this study

## ABSTRACT

**Background:** Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the virus responsible for COVID-19, continuously mutates, posing new challenges in pandemic management. It is necessary to study these mutations across all geographical regions, including North Rift Kenya, as they are influenced by environmental and host factors among other factors. Such mutations can undermine vaccine efficacy and diagnostic accuracy by altering epitopes. Therefore, identifying conserved epitopes is crucial for developing long-term immunity strategies. Additionally, understanding cross-immunity with less harmful coronaviruses could help use endemic strains in combating the pandemic.

**Objectives:** To investigate the genetic diversity and immunogenicity of SARS-CoV-2 and their overlap with endemic human coronaviruses in North Rift Kenya

**Methods**: A cross-sectional study was conducted to identify the various SARS-CoV-2 mutations circulating in North Rift, Kenya. Purposive sampling method was used to select 44 samples from COVID-19 patients tested at Moi Teaching and Referral Hospital. Samples these patients, which had a cycle threshold (Ct) value of <30 as determined by Polymerase Chain Reaction (PCR), were sequenced using the Illumina MiSeq platform. Bioinformatic tools were employed to process sequences, identify clades and annotate mutations in the sequences. Basic Local Alignment Search Tool (BLAST) was used to identify conserved epitopes by aligning the annotated proteins with epitopes from the initial Wuhan strain deposited in the Immune Epitope Database (IEDB). The study also examined the immune cross-reactivity of SARS-CoV-2 with endemic coronaviruses by aligning its annotated proteins with epitopes from hCoV-OC43, hCoV-HKU1, hCoV-NL63, and hCoV-229E, also found in the IEDB.

**Results:** Out of the analyzed samples, 2 (5%) samples were Delta of clade AY.46 and 42 samples were Omicron which represents clades BA.1 (2%), BA.1.1(86%), BA.1.1.1(5%), and BA.1.14(2%). There were 65 overlapping mutations of which 38 of them occurred in 88% of the patients, 2 occurred in 63% of the patients 18 in 38% of the patients and 7 mutations in 13% of the patients. Two new mutations (S:R214R and NSP12:A555A) were discovered in the region. Other notable mutations included S:D614G; S:D796Y; S:E484A; and S:N501Y. A total of 5154 (358 T-Cell MHC Class1; 661 T-Cell MHC Class2; and 4135 B-Cell) conserved epitopes were described from a pool of 12285 Wuhan genome epitopes. Various endemic coronaviruses showed cross-reactivity with SARS-CoV-2: hCoV-OC43 having 11 cross-reacting epitopes; hCoV-HKU1 had 13; hCoV-NL63 had 1; and hCoV-229E had 3. Collectively, these cross-reacting epitopes had similarities with 7 immunodominant epitopes.

**Conclusion:** The study identified Delta and Omicron variants, with Omicron being predominant. Among the 65 mutations, most notable ones included S:D614G; S:D796Y; S:E484A; and S:N501Y, with two new mutations, S:R214R and NSP12: A555A. Conserved epitopes from the Wuhan genome were described, and cross-reactivity with endemic coronaviruses was observed, identifying 28 cross-reacting epitopes that shared similarities with 7 immunodominant epitopes.

**Recommendations**: The study recommend continuous monitoring of new SARS-CoV-2 mutations and investigation of the benefits of endemic coronaviruses for insight cross-immunity. An in-depth analysis of endemic coronaviruses is also recommend and its data documented for future researches. Further investigations should focus on mechanisms of cross-reactivity, the diagnostic and therapeutic potential of specific epitopes, and immune responses to coronaviruses to enhance outbreak management.

# TABLE OF CONTENTS

**LIST OF TABLES**

**LIST OF FIGURES**

**LIST OF ABBREVIATIONS AND ACRONYMS**

| | |
|---|---|
| +ve ssRNA | Positive-sense single strand RNA |
| ACE2 | angiotensin-converting enzyme2 |
| ARDS | acute respiratory distress syndrome |
| BLAST | Basic Local Alignment Search Tool |
| BLASTp | Basic Local Alignment Search Tool for proteins |
| BtCoV | bat coronavirus |
| CFR | Case Fatality Rate |
| COVID-19 | Coronavirus disease 2019 |
| DPP4 | dipeptidyl peptidase 4 |
| ERGIC | Endoplasmic Reticulum – Golgi Intermediate Compartment |
| GIT | Gastrointestinal tract |
| HAT | human airway trypsin |
| hCoV | human coronavirus |
| ICTV | International Committee on Taxonomy of Viruses |
| ICU | Intensive Care Unit |
| IFN | interferon |
| INDELs | Insertion-deletions |
| ISGs | IFN-stimulated genes |
| KEMRI | Kenya Medical Research Institute |
| MAVS | Mitochondrial antiviral signaling protein |
| MERS-CoV | Middle East Respiratory Syndrome Coronavirus |
| MTRH | Moi Teaching and Referral Hospital |
| NAAT | Nucleic acid tests |
| NK cells | Natural Killer cells |
| NSPS | Non-structural proteins |
| ORFs | open reading frames |
| PRRs | Pathogen recognition receptors |

| | |
|---|---|
| RAAS | renin angiotensin aldosterone system |
| RBD | Receptor binding domains |
| RNA | Ribonucleic acid |
| SARS-CoV | Severe Acute Respiratory Syndrome Coronavirus |
| SARS-CoV-2 | Severe Acute Respiratory Syndrome Coronavirus 2 |
| sgmRNAs | subgenomic messenger RNAs |
| STING | Stimulator of interferon genes |
| TMPRSS2 | type 2 transmembrane serine protease |
| TLR | Toll-like receptors |
| USA | United States of America |
| VOC | Variants of Concern |
| VTM | viral transport medium |
| WHO | World Health Organization |

## DEFINITIONS OF TERMS

**Adaptive immune response**: The specific immune response mediated by T cells and B cells, which recognize and remember specific pathogens to provide long-term immunity.

**Amplicon-based sequencing**: A sequencing technique that involves amplifying specific DNA regions (amplicons) and then sequencing them to analyze genetic material.

**Cellular immunity**: A component of the immune system that involves the activation of immune cells, such as T cells, to combat infections, particularly intracellular pathogens.

**Cytokine**: Small proteins released by cells, especially those of the immune system, that mediate and regulate immunity, inflammation, and hematopoiesis.

**Efficacy**: The ability of a treatment or vaccine to produce a desired effect, such as preventing infection or disease.

**Epitope**: The specific part of an antigen that is recognized and bound by an antibody or T-cell receptor.

**Gene**: A unit of heredity that is transferred from parent to offspring and determines some characteristic of the offspring.

**Glycoproteins**: Proteins that have carbohydrate groups attached, often found on the surface of cells and viruses, playing roles in cell signaling and immune recognition.

**Immune evasion**: Strategies used by pathogens to avoid detection and destruction by the host immune system.

**Immunodominant**: Refers to the part of an antigen (epitope) that elicits the strongest immune response.

**Nanopore sequencing**: A next-generation sequencing technology that detects DNA or RNA sequences by measuring changes in electrical current as molecules pass through a nanopore.

**Next-generation sequencing (NGS)**: High-throughput sequencing technologies that allow for rapid and large-scale sequencing of DNA or RNA.

**Pandemic**: An outbreak of a disease that occurs over a wide geographic area and affects an exceptionally high proportion of the population.

**Pathogenesis**: The process by which a disease develops, including the mechanisms and cellular events involved.

**Polymerase Chain Reaction (PCR)**: A technique used to amplify specific DNA sequences, making millions of copies of a particular DNA segment.

**Pro-inflammatory**: Refers to substances or responses that promote inflammation, often as part of the immune response.

**Protein**: Large, complex molecules made up of amino acids that perform a vast array of functions in the body, including structural support, enzymatic activity, and signaling.

**Proteases**: Enzymes that break down proteins by hydrolyzing peptide bonds.

**Quarantine**: The isolation of individuals who may have been exposed to a contagious disease to prevent potential spread.

**Subgenomic**: Refers to smaller RNA molecules derived from a larger viral genome, often encoding specific viral proteins.

**Variants of concern**: Viral strains that have mutations leading to increased transmissibility, severity, or immune evasion, raising public health concerns.

**Wild-type virus**: The original, non-mutated form of a virus as it exists in nature, before any genetic changes or adaptations.

## ACKNOWLEDGEMENT

I am deeply grateful to God Almighty for His grace and guidance throughout this journey.

My heartfelt thanks go to my supervisors, Prof. Kirtika Patel and Dr. Richard Biegon, for their unwavering support, insightful guidance, and mentorship. Their expertise and encouragement were pivotal to the success of this study and my growth as a researcher.

I extend my sincere appreciation to Prof. Elijah Songok for facilitating the sequencing of my samples at KEMRI, and to the KEMRI team, especially Dr. Damaris and Kelvin, for their technical assistance. I am also grateful to Prof. Binhua Liang for his mentorship in bioinformatics and to Dr. Stanslaus Musyoki for his invaluable advice and support throughout this work.

Special thanks to Ms. Caroline Gikunyu for her companionship and collaboration during the study, and to Dr. Winfrida Cheriro for her significant contributions.

I also acknowledge the CEO and Management of Moi Teaching and Referral Hospital, the Department of Laboratory and Pathology, and Moi University, College of Health Sciences, for their institutional support and resources.

Finally, to my family, I owe my greatest gratitude. My late Dad's unwavering belief in me continues to inspire me, and my Mum's love and sacrifices have been my foundation. My siblings—Edith, Patrick, Faith, Freda, Lucyline, and Eric—have been my pillars of strength, and I am eternally thankful for their encouragement and love. This achievement is as much theirs as it is mine.

This journey would not have been possible without the collective support of all those mentioned above. I am truly blessed to have had such a remarkable network of encouragement and guidance.

# CHAPTER ONE

## 1.0 INTRODUCTION

### 1.1 Study Background

The global impact of COVID-19 has been magnified by the continual emergence of genetically distinct SARS-CoV-2 variants. These variants, which harbor mutations in key genomic regions, have significantly altered transmission dynamics, vaccine effectiveness, clinical manifestations, and diagnostic accuracy (Planas et al., 2021; P. Wang et al., 2021). While much of the global surveillance has focused on variant characterization in high-income countries, limited data exists on the genetic landscape of SARS-CoV-2 in sub-Saharan Africa, and particularly in Kenya's North Rift region. This creates a significant gap in understanding local viral evolution and its implications for public health responses.

Studies have highlighted the role of SARS-CoV-2 variants in driving changes in transmission dynamics. For example, Buss *et al*. (2021) reported a three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic, indicating the heightened transmissibility of certain variants. Similarly, Chinazzi *et al*. (2020) demonstrated the impact of travel restrictions on the spread of COVID-19, underscoring the importance of monitoring variants to inform targeted interventions aimed at reducing transmission.

Concerns have also been raised regarding the impact of SARS-CoV-2 variants on vaccine efficacy. Planas *et al*. (2021) documented reduced sensitivity of the Delta variant to antibody neutralization, suggesting potential challenges for vaccination efforts. Wang *et al*. (2021) further investigated the antibody resistance of variants such as B.1.351 and B.1.1.7,

highlighting the need for ongoing surveillance to assess the effectiveness of vaccines against emerging strains.

In addition to affecting transmission dynamics and vaccine efficacy, SARS-CoV-2 variants may also influence disease severity. Davies *et al*. (2021) estimated the transmissibility and severity of the Alpha variant in England, indicating an increased risk of hospitalization compared to earlier strains. These findings underscore the importance of monitoring variants to identify factors that contribute to disease severity and inform clinical management strategies.

Understanding the genetic makeup of SARS-CoV-2 variants is crucial for developing therapeutic interventions. Liu *et al*. (2021) investigated the neutralizing activity of serum elicited by vaccines, highlighting the need to assess the efficacy of existing treatments against emerging variants. By studying variants, researchers can identify potential drug targets and develop alternative treatment strategies to combat evolving strains.

The emergence of SARS-CoV-2 variants has also posed challenges for diagnostic testing. Cele *et al*. (2022) documented the Omicron variant's escape from neutralization by the Pfizer BNT162b2 vaccine, emphasizing the importance of adapting diagnostic assays to detect infections caused by emerging strains. By studying variants and their genetic signatures, researchers can ensure the accuracy of diagnostic tests and refine testing protocols to effectively identify infections.

Effective public health measures rely on timely interventions informed by scientific evidence. By monitoring SARS-CoV-2 variants, policymakers can implement targeted

interventions to limit the spread of specific strains and prevent outbreaks. The findings of studies such as those by Buss *et al*. (2021) and Chinazzi *et al*. (2020) provide valuable insights into the effectiveness of measures such as travel restrictions and vaccination campaigns in controlling the spread of COVID-19.

Although several international studies have detailed the mutation profiles of globally circulating SARS-CoV-2 variants (e.g., Alpha, Beta, Delta, Omicron), regional genomic data from rural and underserved parts of Kenya remain scarce. The North Rift region, with its unique demographic, ecological, and epidemiological features, may harbor distinct viral variants that have not been fully characterized. Without detailed genomic surveillance, it is impossible to ascertain how these regional strains differ from global prototypes, what novel mutations may be present, or how these mutations might influence infectivity and pathogenesis. This study seeks to bridge this gap by employing next-generation sequencing and comparative genomics to map the diversity of SARS-CoV-2 in the North Rift.

A critical step in vaccine development and immune monitoring is the identification of conserved B-cell and T-cell epitopes. While global efforts have led to the prediction and validation of such epitopes in SARS-CoV-2, regional variations in viral sequences may significantly influence epitope architecture and immunogenicity. In regions like Kenya's North Rift, where viral genomic surveillance is limited, the absence of localized epitope mapping hinders the development of region-specific diagnostics and vaccines. Moreover, immune escape mutations arising in local strains may compromise the effectiveness of globally designed immunological tools. This study therefore seeks to address this gap by performing in silico epitope prediction based on SARS-CoV-2 variants sequenced from the

North Rift, with the aim of identifying candidate targets for vaccines and diagnostic tools that are locally relevant.

The significance of such epitope prediction becomes even more evident when examined within the broader context of SARS-CoV-2 immunogenicity. Several studies have underscored the diverse nature of immune responses to the virus. For instance, Guðbjartsson et al. (2020) observed that individuals who had recovered from SARS-CoV-2 infection exhibited a robust humoral immune response, characterized by high seropositivity and stable antiviral antibody titers over time. These findings emphasize the importance of characterizing the specific antigenic determinants that elicit such durable immune responses.

Conversely, not all viral components elicit protective immunity. Flower and colleagues highlighted the role of the SARS-CoV-2 ORF8 protein in immune evasion, demonstrating how certain structural elements of the virus may undermine host defenses. Understanding such mechanisms is essential for designing vaccines and therapeutics that can overcome viral strategies for immune escape (Flower et al., 2020).

Epitope prediction, particularly through bioinformatic and structural approaches, offers a valuable pathway for addressing these challenges. Grifoni et al. (2020) employed computational tools to predict B-cell and T-cell epitopes in SARS-CoV-2, identifying immunodominant regions crucial for eliciting protective immune responses. This method not only accelerates vaccine design but also enables tailored responses to specific viral strains, an approach particularly useful in regions with evolving variant profiles. Mir et al. (2022) further emphasized the utility of such predictive strategies in accounting for viral variability and enhancing vaccine efficacy.

Given the urgent global demand for effective therapeutics and vaccines during the COVID-19 pandemic, the role of epitope prediction in vaccine development cannot be overstated. As Doytchinova and Tchorbanov (2020) noted, identifying immunodominant epitopes forms the foundation of rational vaccine design. Additionally, the integration of immunoinformatics with structural biology, as demonstrated by Mukherjee et al. (2020), supports the development of multi-epitope vaccines capable of inducing broad and robust immune responses. By focusing on both B and T cell epitopes—particularly those conserved across local and global SARS-CoV-2 strains—researchers can design vaccines that are not only effective against circulating variants but also resilient to future mutations.

By January 2021, Africa had reported approximately 2 million confirmed SARS-CoV-2 infections and 34,000 deaths (WHOa, 2021). In contrast, Europe had recorded over 30 million cases and 660,000 deaths, while Kenya alone accounted for 98,000 infections and 1,723 deaths (WHOa, 2021; WHOb, 2021). These figures highlight a perplexing disparity: despite Africa's high population density, relatively low hygiene standards, and under-resourced healthcare systems, the continent has experienced a disproportionately lower burden of COVID-19 compared to Europe, Asia, and the Americas. The reasons behind this phenomenon remain incompletely understood, but one leading hypothesis points to immune system priming by previously circulating, closely related viruses, particularly endemic human coronaviruses (HCoVs).

Among these, HCoV-OC43, HCoV-NL63, HCoV-HKU1, and HCoV-229E are the most prevalent and are commonly associated with mild respiratory illnesses such as the common cold. Interactions between these endemic coronaviruses and SARS-CoV-2 may modulate immune responses, potentially affecting both the short-term and long-term dynamics of

COVID-19 through mechanisms such as cross-protection or antibody-dependent enhancement (Borchering, 2019; D. He, 2015; Peiris, 2003). However, the precise nature of these immunological interactions remains largely speculative and underexplored.

Emerging evidence suggests that prior exposure to related coronaviruses may elicit immune responses capable of recognizing SARS-CoV-2. For example, Zhu et al. (2020) demonstrated that serum antibodies from individuals previously infected with SARS-CoV-1 exhibited cross-reactive neutralization of SARS-CoV-2, underscoring the potential significance of pre-existing immunity. Additionally, several studies have detected SARS-CoV-2-neutralizing antibodies and SARS-CoV-2-reactive CD4 T cells in individuals with no known history of COVID-19, suggesting possible cross-reactivity with antigens from common cold HCoVs (Crawford et al., 2020; Dan et al., 2021; Jianzhong Shi et al., 2020; Kai Wang et al., 2020). Despite these findings, it remains unclear whether such pre-existing immune responses offer protection, influence disease severity, or impact vaccine efficacy.

In this context, Kenya, like many African countries, presents a unique immunological landscape. The circulation of endemic HCoVs may play a role in shaping local immune responses to SARS-CoV-2, potentially explaining the relatively lower rates of severe disease. Yet, the extent of molecular and immunogenic overlap between SARS-CoV-2 and endemic coronaviruses within the Kenyan population remains uncharacterized. Critical questions persist regarding whether this cross-reactivity targets conserved epitopes and whether it confers beneficial or detrimental effects on immune protection.

This study seeks to address these important knowledge gaps by investigating the molecular similarity and epitope-level immunogenicity overlap between SARS-CoV-2 variants and

endemic HCoVs circulating in Kenya's North Rift region. Through this work, we aim to shed light on the potential role of prior coronavirus exposures in modulating COVID-19 outcomes and inform region-specific public health strategies, vaccine designs, and therapeutic approaches.

## 1.2 Statement of the problem

SARS-CoV-2 has continued to evolve, giving rise to variants that impact transmission, vaccine effectiveness, and disease severity. While global studies have documented these effects, there is limited data on how these variants behave in specific local settings such as North Rift Kenya, where factors like population dynamics, healthcare access, and prior viral exposure may influence viral evolution and immune response.

The North Rift region, served by Moi Teaching and Referral Hospital (MTRH), played a central role in COVID-19 testing and treatment. However, genomic surveillance in this region has been minimal, creating a gap in understanding the genetic diversity and mutation patterns of circulating SARS-CoV-2 variants. Without this information, regional public health efforts and vaccine strategies may be misaligned with local viral dynamics.

In addition, while epitope mapping has guided global vaccine design, there is insufficient data on immune targets (epitopes) present in local variants. Predicting these epitopes is essential to determine whether current vaccines remain effective and to guide development of more regionally appropriate immunogens.

Furthermore, SARS-CoV-2 shares structural similarities with endemic human coronaviruses (HCoV-OC43, HCoV-HKU1, HCoV-229E, and HCoV-NL63). Emerging evidence suggests

that prior exposure to these viruses may generate cross-reactive immune responses. However, the extent of this cross-reactivity in the North Rift population remains unexplored.

This study therefore aims to characterize SARS-CoV-2 genetic variations in North Rift Kenya, predict epitopes in circulating variants, and investigate immune overlap with endemic coronaviruses. The findings will support context-specific public health interventions, vaccine optimization, and preparedness for future coronavirus threats.

### 1.3 Justification of the study

The COVID-19 pandemic has had a profound impact on global health and economies. To mitigate these effects, the proposed study aims to characterize the genomic diversity of SARS-CoV-2 variants circulating in North Rift Kenya and predict their immunoinformatic epitope profiles. By investigating these aspects, the study will provide important insights into the genetic variations of the virus and how these influence immune responses, disease transmission, and potential immune escape mechanisms.

Understanding the genetic diversity of SARS-CoV-2 variants in North Rift Kenya is essential for tracking the evolution of the virus and its potential for widespread transmission. This study will focus on identifying specific genetic variations in the circulating variants, aligning with the objective of determining the genetic variations of SARS-CoV-2 in the region. Variants of concern, such as Delta and Beta, have shown altered transmission dynamics and varying vaccine efficacy (Planas et al., 2021; P. Wang et al., 2021). By studying the genetic makeup of these variants, we can gain a better understanding of their behavior and predict their impact on public health strategies. This knowledge is vital for optimizing current vaccination campaigns and ensuring their effectiveness against circulating strains.

The study will also predict epitopes within the SARS-CoV-2 variants circulating in North Rift Kenya, in line with the objective to predict viral epitopes. Epitopes are critical in shaping immune responses and understanding these variations will contribute to the development of vaccines that are more effective against the circulating variants. By focusing on immunoinformatic tools, the study will predict potential epitopes that may provoke immune responses, which could help refine vaccine formulations and improve their targeting of specific viral proteins.

In addition, this study will explore the immunological overlap between the circulating SARS-CoV-2 variants and endemic human coronaviruses in the region. This will help identify any common immunogenic features that could provide cross-protection or inform strategies for more effective vaccine and therapeutic development. Investigating this overlap aligns with the objective of examining the immunological relationship between SARS-CoV-2 variants and endemic coronaviruses.

The outcomes of this study will provide valuable data for public health authorities, enhancing our understanding of the genetic diversity and immunogenic potential of SARS-CoV-2 in North Rift Kenya. This information is critical for the development of targeted public health interventions, including vaccination strategies and therapeutic approaches, that can address the unique challenges posed by the regional variants. Additionally, understanding the immune overlap with endemic coronaviruses could inform future approaches to vaccine design and cross-species immunity.

This study is focused on the genetic characterization of SARS-CoV-2 variants, the prediction of epitopes, and the analysis of immunological overlap with endemic human coronaviruses.

It does not extend to drug target identification or assessment of therapeutic interventions, which are outside the scope of the current investigation. By strictly aligning the study's objectives with these aims, the research will contribute to an improved understanding of the viral variants in North Rift Kenya, and ultimately assist in the development of more effective vaccines and public health interventions tailored to the local context.

## 1.4 Research question

1. What are the variants of SARS-CoV-2 in North Rift Kenya?
2. Which are the probable epitopes of SARS-CoV-2 in the variants circulating in North Rift Kenya?
3. What is the extent of epitope overlap between SARS-CoV-2 variants present in North Rift Kenya and endemic human coronaviruses?

## 1.5 Objectives

### 1.5.1 Broad objective

To characterize the genomic diversity and immunoinformatic epitope profiles of Severe Acute Respiratory Syndrome Coronavirus 2 variants in North Rift Kenya and examine their immunological overlap with endemic human coronaviruses.

### 1.5.2 Specific Objectives

1. To determine genetic variations of SARS-CoV-2 in North Rift Kenya
2. To predict epitopes of SARS-CoV-2 in the variants circulating in North Rift Kenya
3. To investigate the immune overlap among the SARS-CoV-2 variants circulating in North Rift Kenya and the endemic human coronaviruses.

# CHAPTER TWO

## 2.0 LITERATURE REVIEW

### 2.1 SARS-COV-2: disease, genome and epidemiology

#### 2.1.1 COVID-19

The COVID-19 pandemic, the virus cause of which is SARS-CoV-2 has expanded significantly across the world thereby presenting a major challenge to public health, economies, and societies (Buss et al., 2021). It falls under the family of coronaviruses that are known to cause respiratory illnesses in people. The International Committee on Taxonomy of Viruses (ICTV) classified it as a Betacoronavirus within the Coronaviridae family (*International Committee on Taxonomy of Viruses: Http://Www.Ictvonline.Org/*, 2019). This makes it one among other famous coronaviruses including SARS-CoV and MERS-CoV.

#### 2.1.2 Classes of Coronaviruses

Coronaviruses belong to the subfamily (*Coronavirinae*) in the family *Coronaviridae* of the *Nidovirales* order (Cavanagh, 1997). They were further classified into three groups: groups I and II (infective viruses for humans and mammals) and group III (poultry). Group I contains human coronavirus (hCoV- 229E and hCoV- NL63) (Van Der Hoek *et al*., 2004). Human pathogenic group II viruses consist of hCoV-HKU1 (Woo *et al*., 2005) and hCoV-OC43 (McIntosh *et al*., 1967), and the severe acute respiratory syndrome (SARS) coronavirus (Snijder *et al*., 2003). The International Committee on Taxonomy of Viruses (ICTV) later documented four genera within the *Coronavirinae* subfamily: *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*, (that were formerly referred to as coronavirus

groups I, II, and III,) and *Deltacoronavirus* (King *et al*., 2012). *Deltacoronavirus* lineages include viruses detected mainly in bats, such as *Pipistrellus* bat coronavirus HKU5 (BtCoV-HKU5), *Tylonycteris* bat coronavirus HKU4 (BtCoV-HKU4), and *Rousettus* bat coronavirus HKU9 (BtCoV-HKU9) (*Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, 2012). Of these coronaviruses, HCoV NL63, HCoV 229E, HCoV OC43 and HCoV HKU1 are the endemic coronaviruses known to cause disease in human, mostly mild flu and common cold.

### 2.1.3 Coronavirus Replication

All coronaviruses translate proteins for spikes, viral replication and nucleocapsid via their explicit genes in ORF1 and other downstream sections of their genome (van Boheemen *et al*., 2012). The glycoprotein spikes are found on the viral envelope of coronaviruses and play a huge role in the attachment and entry of the virus into host cells. The receptor-binding domain (RBD) is part of the spike that is lightly attached to the virus, giving the virus ability to infect several hosts (Perlman & Netland, 2009; Raj *et al*., 2013). Most of the coronaviruses typically recognize aminopeptidases or carbohydrates as the main receptor for entry to human cells whereas SARS-CoV and MERS-CoV recognize exopeptidases (N. Wang *et al*., 2013). The entry mechanism of a coronavirus is dependent on cellular proteases which include, transmembrane protease serine 2 (TMPRSS2), human airway trypsin-like protease (HAT) and cathepsins that split the spike protein and launch further penetration changes (Bertram *et al*., 2011). MERS-coronavirus uses dipeptidyl peptidase 4 (DPP4), while HcoV-NL63 and SARS-coronavirus require angiotensin-converting enzyme 2 (ACE2) as the main receptor (Perlman & Netland, 2009; N. Wang *et al*., 2013).

Coronaviruses initiate entry into the target cell by binding their S glycoprotein found on their surface to the host receptor on the target cell. The domain of S protein known as receptor binding (RBD) is present on the S1 subunit. In some coronaviruses, the RBD is found at the N-terminus region of S1, while in SARS-CoV, it is located at the C-terminus (P. K. C. Cheng *et al*., 2004; F. Taguchi *et al*., 1995). The fusogenic feature of the virus to cell membrane is regulated by two tandem domains known as heptad repeats (HR1,2) which are found on the S2 subunit of S protein (Xia *et al*., 2020; N. Zhang *et al*., 2014).

The Spike glycoprotein of the SARS-CoV-2 has also been reported to be modified via homologous recombination. It is the combination of bat SARS-CoV and an unknown Betacoronavirus (B. Li *et al*., 2020). It has a 3-D structure in the RBD region maintained by the van der Waals forces (Xu *et al*., 2020). The critical lysine 31 residue on the human ACE2 receptor recognizes the 394 glutamine residue in the RBD region of SARS-CoV-2 (Wan *et al*., 2020). Wan *et al* in their fluorescent study further have shown that a single N501T mutation in SARS-CoV-2's Spike protein can significantly increase its binding affinity for ACE2 (Wan *et al*., 2020). M protein shapes the virions and binds nucleocapsid. The E-protein is involved in viral pathogenesis through virion assembly and release. N protein packages the encapsulated genome into virions (P. Zhou *et al*., 2020).

ACE2 is a membrane-bound protein which is important in the renin-angiotensin aldosterone system (RAAS). It is expressed in different organs in humans, including the kidneys, cardiovascular system, gastrointestinal tract, adipose tissues, the lungs, and the central nervous system. Inhibition of ACE2 can decrease the clearance of harmful molecules like angiotensin II as well as the creation of anti-inflammatory and pro-regenerative molecules such as angiotensin 1–7 (A. Wu *et al*., 2020). Earlier studies in animal models showed the

protective effect of ACE2 against ARDS. It has been identified as the receptor molecule for SARS-CoV-2 cellular entry. Consequently, ACE2-positive cells are the target cells and are vulnerable to SARS-CoV-2 infection. In human tissues, the highest expression of ACE2 is in the heart, small intestine, kidneys, testis, adipose tissue and thyroid. Its lowest expression in muscle, the blood, bone marrow, spleen, blood vessels and brain.

There are other viral elements that take part in entry and interaction between viruses and their hosts apart from spike protein. For example, ORF3a protein has been reported to enhance viral infectivity as well as modulation of host immune responses (D.-Y. Chen *et al*., 2022). Changes in the interaction between variants and host immune pathways due to mutated ORF3a can alter disease progression, and hence vaccine efficacy.

Host immune responses play a key role in controlling viral infections while SARS-CoV-2 has evolved mechanisms for evading them or suppressing them. Viral proteins involved in immune evasion that undergo variation can both affect disease severity and vaccine efficacy greatly. Take for instance Epsilon variant (B.1.427/B.1.429), which has a mutation at the ORF1a gene that likely impairs its ability to suppress host antiviral responses (Damavandi *et al*., 2022).

SARS-CoV-2 penetrates the lung parenchyma and begins proliferation. The virus follows the naso-oral cavity path to reach the lungs. Once inhaled, the virus is, crosses into the epithelial cells of the nasal cavity by linking its viral RBD to the ACE2 receptor of the cell and initiates classical coronavirus replication (Markus Hoffmann *et al*., 2020; Sungnak *et al*., 2020; Walls *et al*., 2020). TMPRSS2 and ACE2 play crucial roles in the cellular entry of SARS-CoV-2, like in that of SARS-CoV (Markus Hoffmann *et al*., 2020). The S1 domain of the spike

protein attaches to the host cell surface via the cellular ligand ACE2. Priming by TMPRSS2, a cellular protease, enables cleavage of spike protein at the S1/S2 subunit. The S2′ subunit then allows the fusing of the virus to the host cell through its membranes (Markus Hoffmann *et al*., 2020). Coronavirus genome transcription and replication take place at cytosolic membranes upon viral entry into the cell. The replicase complex enhances the continuous and discontinuous synthesis of RNA, where 16 viral subunits and several of cellular proteins are then generated. Budding out from the host cell happens when replicated RNA is added into virions (Mousavizadeh & Ghasemi, 2020). At this point, zero symptoms to endemic constitutional symptoms and initial immune response by innate system are experienced.

Once inside the host cell, the coronavirus through membrane fusion releases its +ve ssRNA genome into the cytosol and begins the translation of ORF-1a and ORF-1b. The resultant of this translation is formation of two large polyproteins (pp1a and pp1ab). These polyproteins are then cleaved into 16 Non-structural proteins (NSP1-16) by three functional proteases, eventually creating accessory proteins for virus assembly among them, the viral RNA polymerase (Brockway *et al*., 2003; de Wilde *et al*., 2018; Sawicki & Sawicki, 2005). Consequently, there is an uninterrupted replication-transcription which results in the establishment of various nested sets of subgenomic messenger RNAs (sgmRNAs) that eventually translate into several structural and nonstructural proteins (S. Hussain *et al*., 2005). After synthesis, the E glycoproteins are fused into the Golgi membrane of rough endoplasmic reticulum. The +ve ssRNA becomes encapsulated by the capsid protein to form the nucleocapsid, followed by bourgeoning of assembled virus particles in the Endoplasmic Reticulum – Golgi Intermediate Compartment (ERGIC) (Perrier *et al*., 2019). Finally, the vesicles loaded with virus particles are amalgamated with the cell membrane and shed out

(De Wit *et al*., 2016). These new highly contagious virions not only infect the bordering healthy cells but also are released into the nearby environment through respiratory droplets to potentially spread the disease to a healthy population.

### 2.1.4 Pathogenesis of COVID-19

The pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) involves a complex interplay between direct viral cytopathic effects and host immune responses. This dynamic interaction has been a subject of extensive research due to its critical role in disease progression and clinical outcomes (Lamers & Haagmans, 2022).

Following exposure, SARS-CoV-2 initially replicates in the upper respiratory tract, particularly in the nasal and pharyngeal mucosa, during an early asymptomatic phase that typically lasts 1–2 days. During this phase, viral replication proceeds largely unchecked by innate immune defenses. Symptomatic illness generally develops within 2–14 days post-exposure, presenting with signs similar to other coronaviruses such as SARS-CoV and MERS-CoV. These include pharyngitis, fever, dry cough, myalgia, fatigue, and dyspnea (Centre for Disease Control and Prevention, 2020).

As the virus spreads into the lower respiratory tract, a more vigorous innate immune response is triggered. This is often accompanied by a surge in pro-inflammatory cytokines, sometimes culminating in a cytokine storm. Such immune dysregulation can lead to viral sepsis, acute respiratory distress syndrome (ARDS), pulmonary edema, multi-organ failure, and in severe cases, death (Shereen *et al*., 2020).

Although respiratory manifestations are most prominent, gastrointestinal involvement has also been documented. This is attributed to the high expression of angiotensin-converting enzyme 2 (ACE2) receptors in gastrointestinal epithelial cells, which serve as the entry point for the virus. Gastrointestinal symptoms such as diarrhea, though less frequent, do occur and complicate the clinical presentation (S. J. Lee *et al*., 2022). Similarly, organs such as the liver and pancreas, also rich in ACE2 receptors, may sustain viral injury, potentially leading to hepatic and pancreatic dysfunction (S. J. Lee *et al*., 2022).

The clinical spectrum of COVID-19 ranges from asymptomatic or mild disease to critical illness. Common symptoms include fever, cough, shortness of breath, fatigue, myalgia, and anosmia or ageusia (Dawei Wang *et al*., 2020). Severe disease is more common in older adults and individuals with underlying comorbidities such as cardiovascular, pulmonary, or renal conditions (Guan *et al*., 2020). In critical cases, an exaggerated immune response induces a hyperinflammatory state, marked by excessive cytokine release and uncontrolled infiltration of immune cells into lung tissue. This cascade often results in ARDS, a life-threatening condition, and may be associated with systemic microvascular injury affecting multiple organs (Khreefa et al., 2023; Sherif et al., 2023).

Understanding the multi-organ involvement and immunopathogenesis of SARS-CoV-2 is crucial for developing targeted therapeutic and preventive strategies. As research continues, greater insight into these mechanisms may guide interventions to mitigate the clinical severity and reduce mortality associated with COVID-19 (S. J. Lee *et al*., 2022).

### 2.1.5 Epidemiology

The historical origins of SARS-CoV-2 are linked to repeated zoonotic spillover events, whereby animal viruses cross species barriers to infect human populations, often resulting in outbreaks. It is widely accepted that SARS-CoV-2 originated in animals, with bats considered the most probable natural reservoir. The initial human cases were associated with the Huanan Seafood Wholesale Market in Wuhan, China, suggesting a potential site of interspecies transmission (Pagani *et al*., 2023). Genomic analyses of SARS-CoV-2 have revealed evidence of multiple recombination events among coronaviruses across different species and geographic regions. These recombination events have played a critical role in shaping the virus's genetic diversity and enhancing its capacity to infect both humans and other mammals (Meega & Kumar, 2022). As the virus spread rapidly across the globe and adapted efficiently within human hosts, the accumulation of mutations increased. This ongoing viral evolution raises the likelihood of further recombination events, thereby sustaining the potential emergence of novel variants with enhanced transmissibility or immune evasion (Farahat *et al*., 2022).

#### 2.1.5.1 Transmission and spread of SARS-CoV-2

SARS-CoV-2 spreads through respiratory tract transmission, as well as other routes such as fecal-oral, vertical transmission, and aerosol-eye transmission also exist (Rodrigues, 2024). Primarily, the mode of transmission for SARS-CoV-2 is through respiratory droplets. The droplets are discharged when infected individuals cough, sneeze, or talk. Additionally, close contact with infected individuals (Zhang *et al*., 2020). Apart from person-to-person spreading, environmental pollution is another way to transmit the virus. COVID-19 can easily be spread indirectly when a non-infected individual touches own mouth, eyes, and

nose after touching objects soiled with infectious droplets. Studies have shown that SARS-CoV-2 can remain stable and communicable in aerosols for hours. On some surfaces made of stainless steel or plastic, the virus can remain stable for days (van Doremalen *et al*., 2020).This high transmissibility has contributed to the rapid global spread of the virus. Once inside the body, the virus enters human tissue by attaching its Spike (S) protein to the angiotensin converting enzyme 2 (ACE2) receptor on the surface of epithelial cells (Higgins *et al*., 2023). The clinical presentation of the consequent disease can range from having no symptoms to severe disease. The most common symptoms include, dry cough, fever, fatigue (Yiting Zhang *et al*., 2023) and loss of taste or smell (Dawei Wang *et al*., 2020).

After its identification in Wuhan in late December 2019, COVID-19 started spreading rapidly to other provinces as well as other countries including Japan, Korea, and Thailand. Thereafter it spread throughout Asia. Through various transmission routes, including international transport aero plane and cruise ship, local and community transmission, other continents, like Europe and the United States of America (USA) followed (Tsang *et al*., 2020). By March 2020, WHO had recorded the total number of confirmed deaths in Italy at 3,407 surpassing those 3,253 in China (WHO, 2021), and the total number of confirmed cases of the infection in the USA was 85,228 had exceeded China's 82,213 making USA the country with the majority number of confirmed cases (*Situation Report-68 HIGHLIGHTS*, 2020). In May 2020, WHO confirmed 3,525,087 cases and 248,913 deaths worldwide. A month later, the confirmed COVID-19 cases surged to 6,535,354 and 387,155 deaths reported (WHO Headquarters & WHO Worldwide, 2020). The global trend of confirmed COVID-19, cases and deaths keep increasing significantly, By October 2020, WHO reported the numbers of confirmed cases and deaths as 33,842,281 and 1,010,634, respectively (WHO, 2020).

The prevalence of SARS-CoV-2 varies greatly across countries — global seroprevalence estimates range from 3% to 15% (Islam *et al*., 2021). The Eastern Mediterranean region has the highest prevalence at 15%; followed by America at 8%, Europe at 5% and the Western Pacific at 3% (Donohue, 2022). However, these estimates are based on seroprevalence studies, which may not give a complete picture of infection. Understanding SARS-CoV-2 prevalence also requires considering genomic surveillance data, which provides valuable insights into how common different variants are. Globally, Delta is the most common variant overall, accounting for proportions ranging from 67.58% to 98.31% in various areas (Azami *et al*., 2022). But sequencing coverage and availability of genomic data vary among regions. Globally, less than 3.5% of all confirmed cases have been sequenced (V. Patel et al., 2020). To accurately monitor SARS-CoV-2's spread and measure its impact on public health globally, we need improved sequencing capacity and more sharing of genomic data (Z. Chen *et al*., 2021).

Across England, one of the nations that suffered highly from this disease outbreak, there has been a significant fluctuation in the proportion of people testing positive for SARS-CoV-2 between April and November 2020. The fluctuation entailed as number of positive cases initially reduced then remained relatively low, and subsequently rose very high by end of August 2020. These findings reveal that transmission dynamics of the virus within a country are not straightforward (Pouwels *et al*., 2021).

In Ontario, Canada, similar seroprevalence estimates designed to determine how many people have antibodies against SARS-CoV-2 ranged from 0.4% to 1.4% between March and August 2020. For example, some studies observed that antibody level reduced over time indicating fading immunity among individuals. Thus, this points out why it is important to

monitor long-term immunity against this pathogen considering its impact on public health (Bolotin *et al*., 2021).

In the United States, a nation that was hugely affected by coronavirus pandemic; most places had seroprevalence rates below 10%. However, certain areas like New York City had higher rates reaching up to 23.2% compared to Florida which recorded 13.3% of the population. The observed regional disparities in seroprevalence were consistent with the peaks in reported cases, thereby implying that there is a close relationship between the two variables. On top of this it is important to note that the number of infections estimated cumulatively exceeded the number of reported cases across all sites, suggesting underreporting and possible underestimation of disease burden (T. Lim *et al*., 2021).

The prevalence of SARS-CoV-2, the virus that causes COVID-19, varies greatly across African countries. For example, an analysis of blood samples collected during the second wave of the pandemic in South Africa found a seroprevalence of 45.2%. That was a considerable increase from the 26.9% seroprevalence observed in December 2020 and rose to 47.1% by April 2021 (Wolter *et al*., 2022). It demonstrates how quickly the virus can spread through a population and highlights the need for ongoing monitoring and interventions.

To get an idea of the overall seroprevalence patterns on the continent, researchers conducted a systematic review and meta-analysis that estimated an average rate of about 16% (Hajissa *et al*., 2022). Subpopulations like healthcare workers had higher rates compared to groups such as blood donors, pregnant women, or general populations.

In Burkina Faso, a West African country, researchers found a much lower seroprevalence rate: just 3.2%. They also found evidence that seven malaria patients were co-infected with SARS-CoV-2 (López-Farfán *et al*., 2022), which says more research is needed into how different diseases interact with each other.

In Mozambique, located on Africa's southeastern coast, researchers found that among members of its general population who were tested before being vaccinated against COVID-19, there were seroprevalences as low as 0.9% and as high as 3.0%. But among specific occupational groups like healthcare workers and people working jobs involving close contact with others (like taxi drivers), it ranged from about 3% to about 7% (Arnaldo *et al*., 2022).

Researchers also conducted testing on Mozambican pregnant women living with HIV who are participating in an unrelated study to see if their antiretroviral therapy interacts with COVID-19 vaccinations. The team found a seroprevalence rate of 7.1% at the time of enrollment and 8.5% at the time of delivery (González *et al*., 2022). In total, they estimated that about 11.3% of the pregnant women living with HIV and their babies were infected between September 2020 and July 2021.

Therefore it is evident that SARS-CoV-2 was widely prevalent in Africa. Differences exist in seroprevalence between countries and populations studied in South Africa, Burkina Faso, Mozambique, healthcare workers and pregnant women living with HIV. These differences highlight how responses to COVID-19 need to be tailored to regions, populations and other contextual factors. Ongoing monitoring research and interventions are necessary to understand what's happening on the ground on a continent that has been dealing with the

pandemic for nearly as long as any other place in the world but has comparatively fewer resources

SARS-CoV-2's prevalence varied throughout Kenya's different regions and time periods, showing that its an evolving and dynamic situation. To shine more light on how common anti-SARS-CoV-2 IgG is in three representative populations of unvaccinated people in Kisumu, Nairobi, and Kilifi, a comprehensive study was done from December 2020 to May 2021. The spread was wild with the lowest being 14.5% and highest at 50.2% (Kagucia *et al*., 2022). Another study conducted between March 2021 and July 2021 looked into how much multiple variants of concern (VOC) of SARS-CoV-2 were circulating in Nairobi and counties close by. The Alpha variant (B.1.1.7), Delta variant (B.1.617.2), and Beta variant (B.1.351), which are highly contagious strains, were all detected (G. Wilson, 2022). Then to take it even further, a new study conducted between February to June 2022 estimated seropositivity rates for anti-S IgG in Kilifi and Nairobi; which is just a blood test that shows if you have the virus or not through your antibodies. The study concluded that about 69.1% of Kilifi residents had already gotten sick while in Nairobi it was around 88%. These results show how much SARS-CoV-2's presence fluctuates in Kenya proving that constant testing needs to be done along with strong public health interventions to lower this viruses impacts on the country (S. A. Khamadi *et al*., 2022).

According to Lauer *et al*, in a study of 181 confirmed cases, the mean incubation period of COVID-19 is approximated to be 5.1 days with 97.5% of patients developing symptoms in 11.5 days after the infection. Around 1% of infected individuals in quarantine, developed symptoms after 14 days (Lauer *et al*., 2020). It has also been documented that as much as the people with symptoms are a source of infection, the asymptomatic people can also act as the

hidden COVID-19 sources (Rothe *et al*., 2020). The estimated basic reproduction number ($R_0$) of SARS-CoV-2, gotten by different studies, varied from 1.4 to 6.49, with the mean being 3.28 (Ying Liu *et al*., 2020), which is a little higher than that of SARS-CoV ($R_0$ at 2 to 5) in 2002. When $R_0$ is more than 1, then the infected number is expected to increase exponentially which can lead to an epidemic or pandemic. On the other hand, $R_0$ is less than 1, which means that the transmissibility of that virus is minimized and not invasive (Sheng *et al*., 2020).

The fatality rate of COVID-19 worldwide was found to be 3.4%, which is higher than that of seasonal influenza (Gebhard *et al*., 2020). In a study carried out in Italy and China, the fatality rate was shown to be increasing as age advanced, ranging from 0–0.2% for those aged 0–19 years old to 8.0–12.8% for those 70–79 years old and 14.8%-20.2% for 80 years old and above (Onder *et al*., 2020). Mainly, deaths are caused by acute respiratory distress syndrome (ARDS), coagulopathy, metabolic acidosis, septic shock, acute respiratory failure, and cardiovascular complications (Gebhard *et al*., 2020). In Europe, France showed the highest mortality rate of 13.38%, among different countries which exceeded 50,000 confirmed cases, and Germany showed the least mortality rate of 1.94% in this category (WHO Headquarters & WHO Worldwide, 2020).

The collective mortality contrasts may be due to hygiene cognizance and overload of the health care system and/or state of personal health. Personal health status comprises of age and underlying conditions which will most likely, lead to an abnormal mortality rate. While an exhaustive analysis of the underlying causes is deficient, studies from Europe and China displayed sexual dimorphism with respect to detected COVID-19 cases and their case fatality rate. The median age was documented as 49 years old and 73% of infected persons were

male. Also observed was comorbidity, as 32% of the people were having an underlying disease, such as cardiovascular diseases, hypertension, and diabetes (C. Huang *et al*., 2020). Further, reports from Italy similarly suggested that persons above 60 years, particularly those aged 80 years and above, and the presence of preexisting health conditions are possible risk factors for severe COVID-19 (COVID-19 Surveillance Group, 2020). The U.S. states report from February 12 to March 28, 2020, also showed an association between preexisting health conditions and severe disease outcomes. Patients with at least one preexisting health condition who required intensive care unit (ICU) admission was 78% and patients necessitating hospitalization without ICU admission was 71%. The most frequently reported conditions were diabetes mellitus, chronic lung diseases and cardiovascular diseases (N. Chow *et al*., 2020).

### 2.1.6 Factors affecting the global spread of COVID-19

The prevalence of SARS-CoV-2 in different countries is affected by multiple aspects. These include age and age mixing, testing rates, sampling strategies, sequencing proportions, environmental factors like temperature, sunlight and humidity, public health parameters as well as vaccination rate. It is important to analyze the effects of these factors in order to gain a comprehensive understanding of how the virus spreads and its impact.

Temperature conditions as well as other environmental related issues affects the incidence of SARS-CoV-2. Temperature and sunlight have been shown to have strong negative correlation with infectivity particularly during winter months (Tripathi *et al*., 2022). Contrariwise the link between humidity level and infectivity is not uniform for SARS-CoV-

2 demanding further studies on this aspect. Understanding these environmental factors will guide public health interventions.

Apart from this few other variables are independently correlated with the Case Fatality Rate (CFR) which include elderly population rate male ratio nurses/hospital beds physicians/hospital beds public health spending smoking rate unvaccinated population rate (Papadopoulos *et al*., 2022). There has been much debate about the vulnerability of different age groups to infection. A study showed that contacts aged 0 - 11 and 12 - 17 years had lower susceptibility to infection compared with those aged 65 years or older (Bistaraki *et al*., 2021). This discovery demonstrates that when determining how common SARS-CoV-2 is in a population one should also consider its age structure.

Probable causes leading to the differences in incidence among sexes include variances in the regulation of hormone-regulated expression of genes encoding for the host type 2 transmembrane serine protease (TMPRSS2) and angiotensin-converting enzyme (ACE) 2 (M. Fischer *et al*., 2002), which are needed for SARS-CoV-2 to enter target cells; and different behaviours. Some reports have shown that there are higher circulating levels of ACE2 in males than those in females. Besides, a higher tissue expression of ACE2 has correspondingly been detected in Asian males as compared to their counterparts (Gebhard *et al*., 2020). The variance in the levels of expression of sex hormones between females and males may perhaps explain the difference in ACE2 expression as swelling evidence show that sex hormones, such as estradiol, are involved in the regulation of the renin-angiotensin-aldosterone system (RAAS), as well as ACE2 (M. Fischer *et al*., 2002; Regitz-Zagrosek & Seeland, 2011). Stelzig, *et al*. reported that estrogen can regulate the expression of ACE2 in differentiated airway epithelial cells (Stelzig *et al*., 2020).

Another factor that could affect detection and estimation of variant prevalence significantly is testing rates and spatiotemporal biases. Low testing rates coupled with spatial and temporal biases concerning tests can result in delays in detecting new variants besides unreliable estimates on their prevalence (Han *et al*., 2022). Therefore it's essential to have enough testing capabilities with a balanced distribution across regions over time periods. Considering these factors will help policy makers and healthcare providers understand some contributing factors to severe cases relating COVID-19 hence they can be tailored accordingly.

The COVID-19 severity is worse in aged persons and people with underlying poor health history, like those immune-compromised by chemotherapy for cancer or by HIV infection. Patients with diabetes and asthma, as well as individuals with obesity, hypertension, or cardiac, renal, or liver disorders, are too at greater risk if they contract the disease (Centers for Disease Control and Prevention, 2020; C. Huang *et al*., 2020). Only a very small number of children, as contrasting to adults, has been infected with SARS-CoV-2 (Lu *et al*., 2020).

However various social behavioral aspects related to social behavior and collective immunity can affect the dynamics of COVID-19 epidemic. Factors such as social responsibility, travel regulation and collective immunity to new mutation variants of the virus can shape wavelike epidemic process of COVID-19 (Nechaev *et al*., 2022). Understanding these social and behavioral factors is crucial for designing effective strategies to control the spread of the virus and mitigate its impact.

### 2.1.7 SARS-COV-2 Genome

SARS-COV-2 genome is roughly 30 kilobases in size. It encodes various pivotal structural and non-structural proteins indispensable for the life cycle of the virus. Its genes can be

divided into regions, that include, the 5' untranslated region (UTR), ORF1ab gene, spike (S) gene, envelope (E) gene, membrane (M) gene, nucleocapsid (N) gene, and the 3' UTR. The 5' UTR and 3' UTR are non-coding sections that are of utmost significance for the replication and packaging of viral RNA (Zarandi *et al*., 2021).

The significance of the ORF1ab gene in SARS-CoV-2 cannot be overstated as regards virus reproduction and pathogenesis. The gene, which is situated at the 5'-end of the viral genetic material, encode non-structural proteins (nsps) that play a central role in viral replication and transcription. These nsps play a crucial role in the constitution of the replication-transcription complex (RTC) of the virus, which carries out the replication and transcription of the viral RNA. Extensive research has identified various potential targets for pharmaceutical intervention within these replicase genes, thereby offering promising avenues for the development of antiviral drugs (Piccoli *et al*., 2020). Additionally, it harbors accessory proteins which contribute greatly to the ability of the virus to cause disease (Jawad et al., 2022). It should be noted that ORF1ab gene has conserved regions that could elicit powerful T-cell responses. This quality makes it an appealing candidate for booster vaccines development (Nguyen *et al*., 2023). Moreover, certain sequences in the ORF1ab gene are recognized by γδT cells suggesting their potential use as targets for therapeutic strategies (Du *et al*., 2022). Also of importance are mutations within the ORF1ab gene where several such changes can impact on severe hepatitis caused by these viruses and autoimmune T-cell responses. Overall, SARS-CoV-2's ORF1ab gene is pivotal in viral replication, pathogenesis, and immune response induction.

The function of the spike protein of SARS-CoV-2 in viral entry into host cells is very important and vital as it also determines the organs that the virus can infect and cause damage

to, apart from influencing the course of disease itself (D.-Y. Chen *et al*., 2022). In order to enter a cell through viral infection, this spike protein binds efficiently with a specific receptor located on the host cell surface called angiotensin-converting enzyme 2 (ACE2) which is considered as the most fundamental step in infection process (Banerjee *et al*., 2022). This is worth mentioning that spike protein gets a precise enzymatic cleavage at a particular position named S1/S2 by cellular proteases such as furin, and this step has great importance for determining ability of fusion with host cell membrane, facilitating virus entrance into infected cells, therefore increasing infectiousness thereof (Williams & Vethan, 2023; M. Yamamoto *et al*., 2022). Furthermore, it should be noted that one of the most important residues in ACE2-spikes interaction point C488 locates close to where spike proteins directly interact with ACE2 receptor. It's also worthy to note that any mutation occurring within spike protein, for example those found in Omicron variant can potentially enable virus evade from immune responses stimulated by vaccination particularly humoral immunity generated by vaccines. Moreover, earlier studies have shown that exposure to S1 protein of SARS-CoV-2 substantially modifies phenotypic characteristics of neurons during their early growth stages leading to altered patterns in electric bursts exhibited by neurons thereby revealing possible involvement of spike protein in development and functioning of central nervous system during infections.

The envelope (E) protein of SARS-CoV-2 performs a multitude of functions throughout both the viral life cycle and pathogenesis. One of its crucial roles is the facilitation of assembly and budding of new virions at the ER-Golgi intermediate compartment (ERGIC) (Mehregan *et al*., 2022). Remarkably, the E protein demonstrates the ability to engage in ion channel activity, particularly pertaining to the alteration of the ionic balance of calcium ions (Ca2+)

within the host cells (Antonides *et al*., 2022). By binding to calcium ions, the E protein effectively initiates the activation of the channel, which subsequently has the potential to release Ca2+ from the ER (Nahálka, 2022). Furthermore, the E protein interacts with diverse host proteins and pathways, thereby inducing stress responses, translational shutoff, and the activation of autophagy (Waisner *et al*., 2022). Additionally, the E protein exerts influence over genes involved in the SARS-CoV-2 life cycle and pathogenesis, either promoting or repressing their expression. These genes encompass those related to adaptive immunity, pulmonary tissue function, and even diseases such as Alzheimer's and Parkinson's (H. Huang *et al*., 2023). Collectively, the impact of the E protein of SARS-CoV-2 extends significantly beyond its primary role in viral replication, profoundly affecting host cell physiology and eliciting intricate immune responses.

The M gene in SARS-CoV-2 serves a multitude of functions within the viral system, encompassing initial attachment to host cells, the assembly of viral proteins, as well as the facilitation of glucose transport (L. Shen *et al*., 2021). It holds the distinction of being the most abundant structural protein within the viral makeup, and is heavily implicated in the process of viral assembly and morphogenesis (Z. Zhang *et al*., 2022). Through its formation as a dimer, the M protein further proceeds to assemble into higher-order oligomers (Dolan *et al*., 2022). Furthermore, it engages in interactions with various other viral components, such as the nucleocapsid protein (N) and RNA, thereby serving as a mediator for their recruitment (Y. Ren *et al*., 2021). Moreover, it has been discovered that the M protein possesses the capability to induce apoptosis through a caspase-dependent pathway, ultimately leading to cellular demise. This effect is achieved with the assistance of the nucleocapsid protein N. An additional noteworthy aspect is the electropositive nature of the M protein's cytosolic surface,

which may prove pivotal in facilitating interactions with other viral proteins and RNA. Taken together, the M gene emerges as an indispensable element in the overall process of viral assembly, providing essential structural integrity and potentially contributing to the pathogenicity of the virus.

The ORF3 gene, present in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), exhibits a multitude of functions. One of its primary roles is that of a viroporin, which actively disrupts the activities of ion channels in both the plasma membrane and endomembranes of the host cell. This disruption has significant implications for various stages of the viral life cycle, including viral entry, transcription, replication, and release (J. Zhang *et al*., 2022). Furthermore, the ORF3a protein, which is encoded by the ORF3 gene, plays a pivotal role in triggering cellular innate immune responses as well as pro-inflammatory immune responses. This subsequently leads to the initiation of a cytokine storm, a phenomenon characterized by an excessive release of cytokines, and ultimately culminates in cell death through various mechanisms such as apoptosis, necrosis, and pyroptosis. Consequently, this detrimental cascade of events results in significant tissue damage (Landherr *et al*., 2023). Additionally, ORF3a is known to interact with STING (stimulator of interferon genes) and disrupt the interaction between STING and LC3 (microtubule-associated protein 1A/1B-light chain 3), thereby impeding the process of cGAS-STING-induced autophagy. This disruption ultimately enhances the replication of the virus within the host cell (J. Su *et al*., 2022). Another accessory protein, known as ORF3c, is encoded within the ORF3a sgmRNA and localizes specifically to the mitochondria. Upon reaching the mitochondria, ORF3c exerts its inhibitory effects on the host cell's innate immune response by restricting the production of interferon-beta (IFN-β). This hindrance of

IFN-β production ultimately impairs the host cell's ability to mount an effective antiviral response (Stewart *et al*., 2022). Furthermore, ORF3c also exerts an influence on mitochondrial metabolism, thereby altering the intricate balance of cellular energy production. This alteration consequently leads to an increase in the production of reactive oxygen species (ROS), which can be detrimental to the host cell. Additionally, ORF3c has been found to impede the normal process of autophagic flux, which is crucial for maintaining cellular homeostasis. These multifaceted effects of ORF3c suggest that it may play a crucial role in the progression of the disease (Mozzi *et al*., 2022).

The subject of research regarding the ORF10 gene in SARS-CoV-2 has been undertaken in order to gain a comprehensive understanding of its various functions and mechanisms. Numerous studies have been conducted, yielding valuable insights into the role that ORF10 plays within the viral genome. Initially, there was speculation that the ORF10 gene encoded a hypothetical protein; however, recent findings have presented a growing body of evidence that challenges this notion, suggesting that ORF10 may not be a coding region (Pancer *et al*., 2020). In light of these developments, it has been discovered that ORF10 actually plays a significant role in the inhibition of the innate immune response. Through its actions, ORF10 effectively suppresses the expression of essential type I interferon genes as well as interferon-stimulated genes, thereby leading to the degradation of the crucial mitochondrial antiviral signaling protein (MAVS) via the autophagy pathway (X. Li *et al*., 2021). Despite its observed interaction with an E3 ubiquitin ligase known as CRL2ZYG11B, there is currently no substantial evidence to support the idea that ORF10 functions to actively inhibit or hijack this particular ligase (Mena *et al*., 2021). Moreover, it is worth noting that the ORF10 gene has also been found to harbor multiple mutations and co-mutations, which may have

significant implications for the severity of the virus and its associated consequences (Hassan *et al*., 2021). Consequently, in order to obtain a complete and thorough understanding of the biological relevance of ORF10 and its putative protein, further investigations and research endeavors are necessary (Schuster, 2020).

SARS-CoV-2's N gene performs several functions, all with different roles in the virus' life cycle. To begin with, it actively enhances NLRP3 inflammasome activation, which is a key component of the host immune response that ultimately leads to excessive inflammation and severe COVID-19 symptoms (P. Pan *et al*., 2021). Secondly, it efficiently suppresses interferon-beta (INF-β) production as well as interrupts RIG-I signaling pathways involved in the recognition of viral infections needed for activating an effective immune response (Savellini *et al*., 2021). Therefore, the interaction between the immunity system of the host and N gene reveals its crucial role during disease progression.

Moreover, besides viral reproduction process, transcription and assembly, there are other important functions that the N gene plays. In combination with viral genomic RNA; it forms a complex referred to as viral RNA-protein complex (VRPC), essential for replication and transcription (Lutomski *et al*., 2021). This is necessary for efficient production of more viruses particles as well as spread of infection. Additionally, these events demonstrate versatility of N protein which bind to specific GGG motifs in RNA enabling them to contact cyclophilin A; a cellular protein implicated in various viral infections (W. Wang *et al*., 2022). Many potential therapeutic options result from such interactions.

Lastly, another unique feature possessed by the N protein is its association with G3BP1 and G3BP2 proteins involved in stress granule formation. By getting involved in such proteins

too; it prevents creation of stress granules which are significant structures within cells responding under stressful conditions. Thus this inhibition eventually contributes to SARS-CoV-2 pathogenesis by increasing susceptibility to infection and facilitating higher levels of replication (H. Liu *et al*., 2022). Therefore numerous diverse functions carried out by the N gene indicate its major involvement into intricate relationship between the virus and a host; enlightening possible areas for intervention development or/and therapy while providing some insights into how COVID-19 occurs.

## 2.2 SARS-COV-2 mutations and variants

### 2.2.1 SARS-COV-2 mutations

The early genome analysis showed two lineages, A and B, which exhibit dissimilarities in numerous aspects. The two lineages demonstrated distinct patterns of dominance and evolution, as elucidated by various studies (Yoon *et al*., 2023). Over time, lineage B domineered over lineage A. This can be attributed to the acquisition of the Spike D614G substitution, thereby enhancing its capacity for replication and transmission among humans (Aoki *et al*., 2021). Conversely, lineage A variants eventually underwent mutations, including Spike D614G and Spike L452R, which augmented the ability for in vitro replication within human nasal epithelial cells (L. Tan *et al*., 2013). In the case of Marseilleviridae viruses, lineage B viruses have been observed to induce host Acanthamoeba cells to form aggregations known as "bunches," thereby amplifying their potential for infecting Acanthamoeba cells (Yujiang Zhang *et al*., 2007). In stark contrast, lineage A viruses do not exhibit the propensity to form such "bunches" (Pyo *et al*., 2010). These noteworthy findings provide compelling evidence that lineage A and lineage B viruses possess distinct characteristics and employ diverse strategies for replication and infection.

Various mutations that influence how SARS-CoV-2 develops include those in the spike gene and other accessory genes (Johnson *et al*., 2022; McGrath *et al*., 2022). Multiple changes exist within the spike gene, which are represented by R203K+G204R, N501Y, E484K, L452R, and E484Q that have a prospective in escalating viral replication, fitness and immune evasion (Dang et al., 2022; Hoter & Naim, 2022). These particular mutations in this part of the spike gene may intensify virus- ACE2 receptor interplay which is mandatory for this virus to enter host cells (Dang et al., 2022; Hoter & Naim, 2022). Moreover, it has been discovered that some mutations happening within SARS-CoV-2 accessory genes may interrupt with its intracellular innate response against the virus. These non-spike ones may influence disease severity as well as viral propagation processes; hence they might cause emergence of less virulent phenotypes among circulating strains. Therefore, it is apparent that both spike and non-spike genetic alterations play important roles in the pathogenesis of SARS-CoV-2.

Their influence on the pathogenesis of the virus is brought about by mutations in SARS-CoV-2 through a number of different mechanisms. One such important mechanism is how mutations in spike gene are capable of having profound impacts, hence alteration with potential to viral replication, enhances fitness levels as well as supporting overall pathogenesis (Johnson *et al*., 2022). There is also evidence that mutations within the accessory genes of SARS-CoV-2 can help cause some diseases by interfering with the host's innate immune response (McGrath *et al*., 2022). Also, it should be noted that such specific mutations occurring within accessory genes can have implications for both viral replication and disease manifestation in mice (Dang et al., 2022). Additionally, it is worth noting that certain mutations happening in the viral genome of SARS-CoV-2 could result into antigenic drift or shift which makes it possible for the virus to sidestep human immunity and change

its transmissibility and virulence aspects in terms of antigenicity (Dang *et al*., 2022). Thus, considering all these factors, one may assume that there are complex effects caused by SARS-CoV-2 on its pathogenesis due to these mutations making them not only enhancing but also breaking up host immunity leading to changes in transmissibility and virulence.

Jia *et al* (2005) conducted one of first studies that suggested the significance of mutations in SARS-CoV. During human airway epithelia differentiation, they investigated ACE2 receptor expression that serves as an entry point for this virus into human cells. The researchers found out that ACE2 receptor expression was increased with differentiation suggesting cellular differentiation controlled host cell susceptibility to the virus besides being influenced by other factors such as changes in the virus's genetic material as a result of mutation (Jia *et al*., 2005). This finding demonstrates how mutations can affect viral tropism and infectivity.

The E484K mutation in SARS-CoV-2 affects pathogenesis through several mechanisms. The mutation enhances the binding affinity between the receptor-binding domain (RBD) of the virus and the human angiotensin-converting enzyme 2 (hACE2) receptor, leading to improved transmissibility (W. B. Wang *et al*., 2021). Additionally, the E484K mutation causes conformational rearrangements in the RBD, resulting in a tighter binding interface with hACE2 and the formation of new hydrogen bonds (McGrath *et al*., 2022). These changes contribute to the increased binding affinity between the virus and its receptor (L. Wu *et al*., 2021). Furthermore, the E484K mutation reduces the binding affinities of neutralizing antibodies and nanobodies to the RBD, potentially weakening the effectiveness of these antibodies in neutralizing the virus (W.-T. Yang *et al*., 2022). Overall, the E484K mutation enhances viral binding to hACE2 and reduces the effectiveness of neutralizing antibodies, which may contribute to increased transmissibility and immune evasion ("SARS-CoV-2

E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations," 2022).

The pathogenesis of SARS-CoV-2 is affected by the D614G mutation in its spike protein, which operates through multiple mechanisms. Firstly, the D614G mutation has the ability to enhance the trafficking of the spike protein to lysosomes, ultimately resulting in the reprogramming of lysosomes by Spike and a subsequent decrease in the expression of Spike on the cell surface (Khatri *et al*., 2022). This particular mutation also provides stabilization to the association between the S1 and S2 regions of the spike protein, thus facilitating the selection of additional mutations that promote an increase in the cleavage of S1/S2. As a consequence, variants with spikes that possess a highly fusogenic nature emerge (Yurkovetskiy *et al*., 2023). Furthermore, the D614G mutation contributes to an elevation in the expression of the spike protein on the cell surface and simultaneously restricts the shedding of the S1 subunit of the spike protein, thereby intensifying the infectivity of the virus (Guo *et al*., 2022). Moreover, the D614G mutation is the predominant mutation found in the Variants of Concern (VoCs) and has provided the virus with an improved ability to survive and evade the host's defense mechanisms (Noaman, 2022). In summary, the D614G mutation plays an indispensable role in the function of the spike protein, the infectivity of the virus, and the emergence of highly transmissible variants of SARS-CoV-2.

Another mutation that has been observed in the spike protein of the SARS-CoV-2 virus is N501Y, and it has several mechanisms through which it affected pathogenesis. This mutation firstly enhances how well this virus sticks to human angiotensin converting enzyme 2 (ACE2) receptors (W. S. Ahmed *et al*., 2022; X. Hou *et al*., 2021). Furthermore, it was found out that N501Y mutation gives a moderate level of resistance to some monoclonal antibodies (mAbs)

(Yichao Zhu *et al*., 2022). Molecular dynamics simulations studies have shown that the spike protein with N501Y mutation binds more strongly and remains bound for longer time than spike protein without the mutation interaction with ACE2 (Yichao Zhu *et al*., 2022). Another important consequence of this mutation is its effect on local interactions with binding partners, which results in a rise in local hydrophobicity and increases its interactions with neighboring residues (Hoter & Naim, 2022). Also, it was noticed that binding by neutralizing antibodies is influenced by N501Y mutation. For certain antibodies, there will be reduced affinity while others will be more likely to bind them. In conclusion, these scientific results are significant for understanding molecular mechanisms underlying viral infectivity as well as antibody recognition against N501Y mutant strains (Hoter & Naim, 2022).

The spike protein's K417N mutation in SARS-CoV-2 affects the pathogenesis of the virus via several mechanisms. For instance, a K417N mutation has been observed to have an impressive ability to significantly increase affinity between the receptor-binding domain (RBD) of spike protein and ACE2 receptor, which is necessary for virus entry into host cells (Fratev, 2020). This heightened affinity may be associated with escalated infectivity and spread of the virus (Sun et al., 2021). Additionally, K417N mutation abolishes the hidden interfacial salt bridge that connects the RBD and neutralizing antibodies, thus reducing their binding energy and enhancing escape from antibody mediated immune responses (Luan & Huynh, 2021). Such a collateral effect might affect efficacy of therapeutic antibodies as well as question viability of currently available vaccines (McGrath *et al*., 2022). If we take an overview on all this impact about K417N mutation it becomes apparent that this has a pivotal role that cannot be dispensed with when it comes to adaptation by SARS-CoV-2. Moreover,

one can suggest plausibly that this alteration may indeed cause enhanced virulence or evasion from immunity known in relation to this pathogen (Das *et al*., 2021).

Research and analysis have extensively been done on the study of non-spike mutations in SARS-CoV-2. Many studies have sought to determine the most common non-spike mutations, especially in proteins such as Nsp4, Nsp3, ORF9b and Nsp6-10. The most prevalent non-spike mutations observed in the SARS-CoV-2 virus are NSP3_P1469S, NSP3_A488S, NSP3_P1228L, NSP4_V167L, NSP4_T492I, NSP6_T77A, NSP14_A394V, NSP12_G671S, and NSP13_P77L, as reported in a published study (Anwar *et al*., 2022). These specific mutations were identified in the genome sequences obtained from the Punjab province of Pakistan during the fourth wave of infections in September 2021. Furthermore, there are numerous distinct mutations that have been detected in the papain-like protease (NSP3), main protease (NSP5), and RNA-dependent RNA polymerase (NSP12) of the virus (Anwar *et al*., 2022). The presence of these genetic alterations could potentially impact the efficacy of therapeutic interventions and vaccines targeting SARS-CoV-2.

### 2.2.2 Variants of concern

Mutations are among primary sources of genetic variation in viruses. They are spontaneous changes taking place in the viral genome during replication. As it goes from one person to another person through replication, SARS-CoV-2 gathers mutations in its genome structure. Many mutations do not cause any harm or have an insignificant impact on virus properties. However, some can allow selective advantages like being more transmissible; evading immunity or changing virulence by mutation on their gene sequence. The occurrence of such

mutated viral lineages will increase infection rates and create a challenge to public health interventions.

It is important to note that during the pandemic of COVID-19 caused by SARS-CoV-2, infectious disease dynamics is characterized by emergence of viral lineages and variants of concern. Viruses like SARS-CoV-2, which belong to the family of coronaviruses, constantly change their genetic makeup as a result of mutations, deletions, insertions and recombination. These are all mechanisms responsible for genetic variation in viruses that are significant for the epidemiology and clinical outcome of viral infections.

Since its emergence in late 2019, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has undergone numerous genetic mutations, resulting in the development of multiple variants of concern. These variants have raised global alarm due to their potential effects on transmissibility, disease severity, and the efficacy of existing vaccines. The significance of these variants is primarily attributed to mutations occurring within the spike protein (Planas et al., 2021; WHO, 2021).

#### 2.2.2.1 Alpha Variant (B.1.1.7):

The variant strains have been reported associated with more transmissibility with altered severity for disease and reduced vaccine effectiveness. A good example is B1.1.7 (Alpha variant) first reported in the UK which has mutation at the spike protein; N501Y, that increases its binding to ACE2 receptor. Its high transmissibility popularized it globally (S. Khamadi *et al*., 2022). One study on this variant was conducted by Harvey *et al* (2021) who found some mutations on its spike proteins including N501Y that enhance interaction between spike protein with ACE2 receptors thereby promoting greater virus entry into host

cells (Harvey *et al*., 2021). Moreover, other changes exist within the receptor-binding domain (RBD) region within their spikes thus increasing infectivity while enabling immune evasion. This strain has been linked to higher viral loads, increased hospitalization rates, and more fatalities compared with earlier strains (Winger & Caspari, 2021).

#### 2.2.2.2 Beta Variant (B.1.351):

On the same note, another variant of virus concern, Beta variant (B.1.351) was discovered in South Africa and it has some changes on the spike protein which includes E484K and K417N that have been linked with reduced susceptibility to neutralizing antibodies (Choi & Smith, 2021).. The spike protein may also be mutated in B.1.427/B.1.429 called Epsilon variant initially found in California These variants have raised alarms due to their potential impact on vaccine efficacy and the course of the pandemic (Kimita *et al*., 2022b). The RBD structure may be altered by these mutations, enabling the virus to escape host immunity and evade reinfection or breakthrough vaccine-associated infections.

#### 2.2.2.3 Gamma Variant (P.1):

Gamma variant originated from Brazil with similar mutations to Beta variant including E484K alongside other spike proteins mutations such as K417T and N501Y (Pondé, 2022). In combination these mutations may contribute to increasing transmissibility via immune escape mechanism. Studies have revealed that the Gamma variant has been linked to increased reinfection rates and partial resistance against neutralizing antibodies (Neerukonda *et al*., 2021).

#### 2.2.2.4 Delta Variant (B.1.617.2):

Delta variant first identified in India became super dominant in many countries within a short period of time. It has several mutations on the spike protein including L452R and T478K (Pondé, 2022) . Transmissibility is raised by these mutations and immune escape occurs as well. The Delta variant has been associated with higher viral loads and severe disease as characterized by increased hospitalization and death rates observed in some parts of the world (Laine *et al*., 2022). The B.1.617.2 (Delta) had several amino acids deleted from its spike protein's N-terminus region thus making it more transmissible. This strain quickly dominated several countries and caused increased hospitalization rates (Githinji *et al*., 2022). The emergence and subsequent predominance of delta variant (B.1.617.2) in multiple countries across Africa, Asia, Europe, North America and Oceania have generated grave concerns worldwide as regards outbreaks associated with SARS-CoV-2. Since its increased transmissibility, has been a cause for concern leading to significant attention being directed on this variant because of its high potential to spread quickly and compromise control efforts. Nevertheless, remained P1 Gamma variant largely prevalent within South America, thereby indicating unique geographical aspects exhibited by variants. These observations underscore the need for international monitoring and response efforts aimed at controlling changing dynamics related to SARS-CoV-2 variants (Donohue, 2022).

#### 2.2.2.5 Omicron Variant (B.1.1.529):

The first appearance of the Omicron variant was recorded in South Africa where it gained worldwide recognition because of its large number of changes on spike protein especially receptor-binding domain (RBD). Some key mutations include N501Y, E484A, P681H among others (M Hoffmann *et al*., 2023). Consequently, cases exploded due to high

transmissibility across different nations. Omicon might have an element of immunity evasion from previous infection or vaccination although severity remains uncertain as per Hoffmann *et al*.'s research. It is important that retrospective studies be conducted to unravel the origins of SARS-CoV-2, thereby enabling us to take proactive measures to prevent the occurrence of future pandemics (Crone *et al*., 2023).

It must be remembered that vaccines still remain critical tools for controlling the spread and impact of SARS-CoV-2 variants. Knowing how viruses enter cells and interact with host proteins will aid in designing vaccines that target conserved parts of the viral genome thus minimizing immune escape risk. The fight against Omicron variant (B.1.1.529) which has many spike mutations; some are associated with enhanced infection while others may help it evade immune responses requires more data on whether available strategies are effective (D.-Y. Chen *et al*., 2022).

There are implications for vaccine efficacy based on spike protein changes seen in SARS-CoV-2 variants of concern. Vaccines produced from earlier strains might be less effective against specific variants because they were able to escape immunological response. Studies have also established that neutralizing antibodies generated by vaccines are partially ineffective against certain strains like Beta and Gamma (Pondé, 2022). Therefore, additional shots and improved vaccines specifically targeting given strains may be necessary for continued protection from these emerging varieties.

Increase in infectivity due to spike mutations has led to outbreaks of covid-19 in the regions affected. In this regard, the quantity of viruses has increased by some variants, for instance delta and omicron, which might be a possible cause of fast and prevalent infection. This

would require such public measures as wearing masks, maintaining social distance and engaging in vaccination campaigns so as to protect people against new variants thus reducing their spread.

Some variants of concern have been associated with different types of diseases. Alpha is more likely than other forms to result in hospitalization or even death (Winger & Caspari, 2021). Likewise, delta is said to be more dangerous than other strains (Laine *et al*., 2022). However, despite being highly transmissible Omicron appears less severe based on certain studies. Nonetheless, medical facilities are bothered about that since the sheer number of cases (M Hoffmann *et al*., 2023).

Global Public Health has had significant difficulties since SARS-CoV-2 Variants of Concern showed up rapidly. For these variants the role played by spike mutations cannot be overemphasized since they enhance viral transmissibility while allowing them escape from immunity thus altering severity levels (Donohue, 2022). As a result, it is necessary to monitor and comprehend epidemiology and characteristics inherent in these mutated strains, thereby tailoring our public health strategies and vaccine development towards effectively managing this pandemic. Collaborations between researchers, healthcare providers and policy makers would go a long way in dealing with these strains for a better future.

At global stage, emergence and increase in prevalence of new mutations and variants of SARS-CoV-2 have attracted considerable attention worldwide including fear among populations due to their perception about this issue. Transmission dynamics and vaccine efficacy are being questioned due to these variants including UK variant (B.1.1.7), South Africa variant (B.1.351) and Brazil's P.1 and P.2 variants among others. Their spread has

been very fast making them become dominant forms thus requiring continuous surveillance as well as adaptable public health interventions which can provide relief against possible consequences in future (Weber *et al*., 2021).

### 2.2.3 SARS-CoV-2 Viral Lineages and Variants of Concern in Kenya

During the COVID-19 pandemic, Kenya has experienced the emergence of VOCs and SARS-CoV-2 viral lineages like other countries before it. Therefore, continuous genomic monitoring and surveillance are important to stop the spread of imported variants and circulation of multiple variants in the country. The findings from relevant studies and guidelines have provided valuable insights into the dynamics of these variants, contributing to the development of targeted interventions, prevention strategies, and vaccination efforts. With ongoing research, global collaboration, robust public health measures we can effectively manage and control spread of viral variants in Kenya during this pandemic.

Variations of SARS-CoV-2 have caused global concerns. In Kenya, second year pandemic imported variants have been documented (Nasimiyu *et al*., n.d.). This indicates that this country is prone to introduction and establishment of novel viral lineages thus requiring constant monitoring and surveillance.

Kimita *et.al* (2022)'s important study conducted in Kenya looked into temporal lineage replacements as well as prevalence of the imported variants of concern during the COVID-19 pandemic. Their findings demonstrate how dynamic are the viral lineages circulating within the nation while stressing on significance genomic surveillance for identification and response towards emerging varieties (Kimita *et al*., 2022a).

The World Health Organisation (WHO) has given crucial instructions for preventing, caring for COVID-19 infection including those in Kenya. These guidelines have directed the nation's reaction to the pandemic with regards to emergence of viral variants.

On top of these imports, multiple SARS-CoV-2 strains were seen circulating across different regions within Kenya. Khamadi *et al*., 2020 conducted whole-genome sequence analysis from samples collected from Nairobi County as well as its environs which portrayed multiple variant types at some point in time. This therefore emphasizes how essential genomic surveillance can be used to understand virological lineage movement within a country.

The importance goes beyond just numbers related to these viral variants. The researchers discussed the clinical epidemiology of COVID-19 in hospitalized children in rural Western Kenya (Tsegaye *et al*., 2023). Moreover, this study offered insights into its impact among vulnerable populations and called for targeted public health measures for children.

Another concern is whether vaccines will remain effective against these new viral variants. Song *et al*., (2020) tackled cross-reactive serum and memory B cell responses to spike protein among SARS-CoV-2 and endemic corona virus infections. In Kenya, understanding immune responses to these changes would be helpful in developing effective vaccination strategies and vaccines.

Kenya therefore needs to strengthen its public health infrastructure and enhance genomic surveillance and testing capacity as it grapples with the current pandemic. A study by Nasimiyu et.al highlights the importance of monitoring imported variants that may exacerbate infection rates within the country (Nasimiyu, Matoke-Muhia, Rono, Osoro, Obado, *et al*., 2022).

Also, since dealing with these new forms is far from straightforward, it is imperative that we effectively communicate risks associated with them towards our people so as they can understand exactly what they mean by following preventive measures including vaccination progress. WHO guidelines have been instrumental in shaping the response of Kenya to this outbreak (WHO, 2015) thus should always be revised due to changing virus landscape.

### 2.2.4 SARS-COV-2 waves and dominant Variants

The first cases of SARS-CoV-2 in Kenya were due to the presence of the wild-type virus. This was followed by a remarkable emergence and eventual dominance of some major VOCs namely Beta, Alpha, Delta and Omicron as is documented by credible sources (Githinji *et al*., 2022). A closer look showed that entry and subsequent spread of these VOCs corresponded with specific regions exhibiting high disease transmission rates, such as Coast, Nairobi, and Western Kenya regions. Additionally, it should be noted that these areas enjoy convenient proximity to established road and air transport routes which may have facilitated rapid disease spread (Barr *et al*., 2022).

The demographic composition also played a role in its overall spread as distinct age groups had different VOCs detected more often than others (Lambisia *et al*., 2022). Furthermore, clinical symptoms depicted by these VOCs differed significantly where Omicron being associated with mild upper respiratory tract illness while Delta was most commonly linked moderate to severe lower respiratory tract illnesses (S. A. Khamadi *et al*., 2022). It must be emphasized that despite implementing public health measures multiple importations occurred raising questions on effectiveness of these measures in preventing virus spread. Besides, it is important to note that secondary attack rate of SARS-CoV-2 among household

contacts in rural Kenya is lower compared to global average and interestingly majority of the infections were asymptomatic. These findings serve as a timely reminder on the need for good knowledge about local epidemiology before implementing evidence-based control strategies aimed at curbing the spread of this virus and protecting public health welfare.

The pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Kenya was characterized by the occurrence of multiple waves, each possessing unique and distinguishable features. It is noteworthy to mention that the initial two waves happened prior to the emergence of VoCs, and their attack rates were comparatively lower, constituting 65.4 and 118.2 per 100,000 individuals respectively (Nasimiyu, Matoke-Muhia, Rono, Osoro, Ouso, *et al*., 2022). Subsequently, waves 3, 4, and 5 unfolded during the second year, and each of these waves was predominantly driven by the Alpha, Delta, and Omicron VoCs, respectively.

In contrast to the earlier waves, these subsequent waves exhibited a higher AR, ranging from 74.9% to 98.7% (Njenga, 2022). Delving further into the specific characteristics of each VoC, it is noteworthy that the Alpha VoC was primarily associated with mild symptoms affecting the upper respiratory tract. Conversely, the Delta VoC was found to be linked with symptoms ranging from moderate to severe, impacting the lower respiratory tract and also causing fever. Notably, some cases involving the Delta VoC also exhibited post-acute phase neurological complications (Kimita *et al*., 2022b). Despite the implementation of government-imposed restrictions during these waves, the intended objective of mitigating the progression of the pandemic was not achieved, leading to a higher incidence of infections spreading extensively across the entire country.

The 6th wave of the COVID-19 pandemic exhibited a decrease in the number of hospitalizations, admissions to intensive care units, and fatalities in comparison to the preceding waves. This decline can be attributed to the implementation of early detection methods and the wide-scale administration of vaccines (Lobo-Rodriguez *et al*., 2023). Furthermore, the study conducted on the subject discovered that there was a significant rise in the incidence of COVID-19 cases within the initial 16 months of the pandemic in Kenya. This finding serves to emphasize the urgent requirement for the implementation of supplementary mitigation strategies, specifically the distribution of vaccines, particularly in densely populated areas characterized by low socioeconomic status (Lobo-Rodríguez *et al*., 2023).

### 2.2.5 Testing SARS-COV-2

Although there are other testing methods such as antibody testing can be used, WHO adopted Nucleic acid tests (NAAT) using reverse transcription-polymerase chain reaction as the main confirmatory tool for COVID-19. Although no particular medication yet discovered, preventive measures such as vaccines, specialized facial masks, and social distancing have demonstrated their effectiveness (Pagani *et al*., 2023). The rapid release of drugs to the market has been facilitated by the innovations in drug discovery and development, including regulation.

The testing methods employed for the detection of SARS-CoV-2, the virus responsible for the COVID-19 pandemic, encompass a wide range of techniques that include reverse transcriptase polymerase chain reaction (RT-PCR), immunoassays, and molecular assays. RT-PCR, a highly prevalent approach, is extensively utilized for the early identification of

SARS-CoV-2 RNA in clinical samples, thereby aiding in the prompt diagnosis of infected individuals (Biswal *et al*., 2021). Furthermore, the rRT-PCR assay serves as a valuable tool in the identification of SARS-CoV-2 RNA in diverse clinical specimens, including but not limited to blood, sputum, bronchoalveolar lavage fluid, biopsies, pharyngeal and nasal swabs, feces, and fibrobronchoscope brush biopsies, thereby enabling comprehensive analysis and assessment of the disease (Tahmasebi *et al*., 2020). In addition to these techniques, fully integrated molecular assays like Idylla™ SARS-CoV-2 have revolutionized the field of diagnostic testing as they facilitate the simultaneous detection of not only SARS-CoV-2 but also other respiratory viruses such as influenza A and B, and RSV, thereby enhancing the efficiency and accuracy of the diagnostic process (Farfour *et al*., 2022).

The significance of the primers employed in the Polymerase Chain Reaction (PCR) for the detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) cannot be overstated, as they play a pivotal role in ensuring the accuracy and specificity of the test results. The specificity of most Reverse Transcription PCR (RT-PCR) tests hinges upon the meticulous design of primers that are tailored to the unique genomic sequence of SARS-CoV-2, thereby enabling the identification of the virus with utmost precision and reliability (Sethuraman *et al*., 2020). In the realm of molecular diagnostic techniques for SARS-CoV-2 detection, Real-time PCR has emerged as a commonly employed methodology, thereby highlighting the paramount importance of primer design in facilitating successful detection of the virus (Gogoi *et al*., 2021). A recent study conducted to compare different primer sets has revealed that regions such as ORF1ab, Nucleocapsid, and RdRp exhibit commendable primer design, thereby enabling the recognition of SARS-CoV-2 RNA. This finding serves to underscore the crucial role played by region-specific primer selection in facilitating

accurate detection of the virus (Mollaie *et al*., 2020). Furthermore, the aforementioned study has also emphasized the development of particular primer sets alongside standard positive controls, which serve to ascertain the authenticity of negative results, thereby bolstering the overall accuracy in the diagnosis of SARS-CoV-2, as false positives can potentially lead to misguided conclusions (Cho, 2021).

The most specific primers for the detection of SARS-CoV-2, possess the ability to amplify segments of the SARS-CoV-2 genome (Kandel *et al*., 2023). A primer sequence that has been utilized for the purpose of detecting SARS-CoV-2 is ACCGTAGCTGGTGTCTCTAT (forward primer) and GTGCCAACCACCATAGAATTTG (reverse primer (Merdekawati & Nurhayati, 2023). Moreover, it has been observed that the primers which target the N (nucleocapsid phosphoprotein) and S (surface glycoprotein) genes exhibit a remarkably high analytical sensitivity and are capable of providing results within a mere duration of approximately 20 minutes (Shaka *et al*., 2023). Furthermore, the RT-MIPLAMP method, which involves the incorporation of a pair of inner primers, has also been developed for the purpose of detecting SARS-CoV-2 and has demonstrated enhanced sensitivity and a shorter reaction time in comparison to the conventional RT-LAMP method (AlHamid *et al*., 2022; X. He *et al*., 2022)

It is worth noting that the utilization of multiple genes in the PCR-based detection of SARS-CoV-2 can indeed result in heightened sensitivity. In support of this notion, a duplex RT-PCR method has been devised, focusing on conserved genes of both SARS-CoV and Middle East Respiratory Syndrome Coronavirus (MERS-CoV), thereby suggesting the potential for increased sensitivity through the simultaneous detection of multiple genes (N. Zhang *et al*., 2020). Furthermore, the integration of serologic testing in conjunction with reverse

transcription polymerase chain reaction testing has been shown to significantly enhance the sensitivity of SARS-CoV-2 detection, thereby implying that a combination of diverse testing methods can synergistically amplify the sensitivity threshold for viral detection (Espejo *et al*., 2020). Moreover, a separate study has shed light on the development of three distinct sets of primers, each targeting different genes of SARS-CoV-2, and has demonstrated that these primer sets exhibit enhanced sensitivity and specificity, thereby offering further evidence to support the notion that the utilization of multiple primer sets can contribute to heightened sensitivity in viral detection (Kalita *et al*., 2021). Additionally, the importance of employing multiple genes for detection purposes has been underscored within the context of loop-mediated isothermal amplification (LAMP) assays, wherein the specificity and sensitivity of the primers designed for different genes have been shown to be critical in achieving accurate detection of SARS-CoV-2 (W. Yang *et al*., 2020). Furthermore, a study has provided evidence that the concentration of cellular material present in nasopharyngeal swabs can impact the clinical sensitivity of SARS-CoV-2 RT-PCR, thereby suggesting that the quantity and quality of the genetic material being tested can significantly influence the overall sensitivity of the test (Kannian *et al*., 2020). In summary, it is evident that the utilization of multiple genes, alongside careful considerations pertaining to primer specificity, testing methods, and sample quality, collectively contribute to heightened sensitivity in the detection of SARS-CoV-2 through PCR.

Another significant category of testing methods comprises immunoassays, which comprise techniques such as Enzyme Linked Immunosorbent Assay and receptor-binding domain-based immunoassays, primarily employed for serosurveillance and differentiation between individuals who have been exposed to SARS-CoV-2 and those who have not, thus aiding in

the identification and monitoring of the spread of the virus within a population (Pulido *et al*., 2021; S. Wang *et al*., 2021). Furthermore, the impact of SARS-CoV-2 infection on mortality and hospitalization rates is closely monitored through the calculation of infection rates, a crucial parameter for gauging the severity and magnitude of the disease outbreak (Bulgaresi *et al*., 2023). Additionally, it is imperative to recognize the critical role played by animal surveillance and routine testing in outbreak response and prevention, as they provide valuable insights into the prevalence and transmission dynamics of SARS-CoV-2, thereby enabling the implementation of timely and effective control measures (K. Deng *et al*., 2022).

The potential utility of lung ultrasound as a diagnostic tool for SARS-CoV-2 infection has also been explored, with promising results indicating its effectiveness in the detection and assessment of lung abnormalities associated with the disease (Bosso *et al*., 2020). Moreover, recent studies have reported the detection of SARS-CoV-2 RNA in various clinical specimens, including pancreatic pseudocyst fluid samples, thereby highlighting the diverse manifestations of the virus and the need for comprehensive testing strategies (Schepis *et al*., 2020).

It is crucial to note that the co-circulation of influenza and SARS-CoV-2 can present diagnostic and management challenges due to the overlapping clinical presentations of these two viral infections, necessitating the implementation of integrated and tailored diagnostic approaches for accurate and efficient identification (Laris-González *et al*., 2021). Furthermore, the exploration of inflammatory biomarkers such as TNF-α in the context of SARS-CoV-2 infection has shed light on the intricate host-virus interactions and the underlying immunopathological mechanisms, thus providing potential targets for therapeutic interventions (C. Y. Lee *et al*., 2021). Lastly, the elucidation of the interaction between the

SARS-CoV-2 spike protein and the human ACE2 receptor through sophisticated modeling, docking, and molecular dynamics simulation techniques has broadened our understanding of the viral entry process and has the potential to inform the development of effective antiviral strategies (R. Li *et al*., 2020).

The sequencing techniques available for SARS-CoV-2 encompass a wide array of methods, including next-generation sequencing, amplicon-based sequencing, metagenomic methods, sequence capture or enrichment methods, and hybridization-based capture (Alpert *et al*., 2021; Clark *et al*., 2022; Park *et al*., 2022; Simonetti *et al*., 2021; K. L. Tan *et al*., 2021). These techniques have been instrumental in generating a substantial number of SARS-CoV-2 genome sequences, which are crucial for understanding the epidemiology and evolution of the virus (Chiara *et al*., 2020; Crits-Christoph *et al*., 2020). The diversity of approaches and methods applied to SARS-CoV-2 genomes underscores the significance of genomic sequencing in tracking and monitoring genetic changes in the virus (S. H. Zhan *et al*., 2021).

Furthermore, the availability of SARS-CoV-2 genome sequences has provided valuable resources for understanding viral evolution and diversity (Lambert *et al*., 2020). The use of chaos game representation (CGR) and recurrence quantification analysis (RQA) has been proposed as an effective process to extract valuable features from genomic sequences of SARS-CoV-2, demonstrating the application of advanced analytical techniques in genomic analysis (Olyaee *et al*., 2020). Additionally, the sequencing of viral concentrates and RNA extracted directly from wastewater has been shown to identify multiple SARS-CoV-2 genotypes, highlighting the utility of sequencing in environmental surveillance and public health efforts (Crits-Christoph *et al*., 2020). The significance of SARS-CoV-2 sequencing is further emphasized by its role in identifying genetically distinct virus strains, tracking the

COVID-19 pandemic, and studying the evolution of the virus (V. Gupta *et al*., 2020; Seemann *et al*., 2020). The availability of over 130,000 complete or nearly complete genome sequences of SARS-CoV-2 in public databases underscores the extensive use of sequencing in characterizing the virus (Cotten *et al*., 2021).

Nanopore sequencing, which has emerged as an invaluable tool in the realm of scientific research, has revolutionized the field of rapid and accurate detection and characterization of the SARS-CoV-2 virus. Researchers have effectively utilized the MinION nanopore sequencing technology to meticulously delineate the intricate transcriptome of the SARS-CoV-2 virus, thereby providing compelling evidence to support the efficacy of nanopore sequencing in the comprehensive analysis of viral genomes (D.-W. Kim *et al*., 2020).

Another study has underscored the tremendous utility of nanopore sequencing in the prompt detection and thorough characterization of the SARS-CoV-2 virus, specifically highlighting the multitude of advantages it offers over conventional sequencing protocols (Arévalo *et al*., 2022). Moreover, researchers have also presented a remarkable extension of the Nanopore toolset known as minoTour, which enables real-time monitoring and comprehensive analysis of SARS-CoV-2 nanopore sequencing data. This significant advancement in the field serves as a testament to the continuous progress being made in the realm of nanopore sequencing (Munro *et al*., 2021). Additionally, a study has demonstrated the remarkable versatility of nanopore sequencing technology in various applications related to viral detection and characterization, specifically highlighting its profound impact in the identification of aptamers targeting the SARS-CoV-2 spike protein (Khrenova *et al*., 2022). Furthermore, researchers have provided groundbreaking insights into the infrequent occurrence of host-virus chimeric events in SARS-CoV-2-infected cells, thereby shedding invaluable light on

the complex dynamics and intricate interactions that take place within host cells (B. Yan *et al*., 2021). This groundbreaking research has significantly contributed to our understanding of the viral genome and its interactions with host cells. Furthermore, researchers have successfully employed nanopore sequencing in combination with poly(T) adaptors and an R9.4 flowcell on a GridION platform to perform direct RNA sequencing and conduct early evolution analysis of the SARS-CoV-2 virus. This pioneering approach has revealed the immense potential of nanopore sequencing in unraveling the genomic characteristics and intricate evolutionary patterns of the virus (Taiaroa *et al*., 2020). Moreover, another study has adeptly utilized amplicon-based MinION sequencing to rapidly identify the SARS-CoV-2 virus and comprehensively characterize the metagenomic makeup of nasopharyngeal swabs obtained from COVID-19 patients. This study has successfully highlighted the incredible speed and efficiency of nanopore sequencing in clinical settings, thus reaffirming its status as an invaluable tool in the field of viral detection and characterization (S. C. Moore *et al*., 2020). All in all, the remarkable capabilities of nanopore sequencing in the detection, characterization, and real-time monitoring of the SARS-CoV-2 virus cannot be overstated. This cutting-edge technology has consistently provided researchers with rapid and accurate insights into the intricate details of the viral genome and its intricate interactions with host cells.

The utilization of the Illumina MiSeq platform for the purpose of whole-genome sequencing of SARS-CoV-2 has been extensively documented and explored, showcasing its remarkable efficacy in the characterization of viral strains as well as the facilitation of outbreak investigations. In one study, researchers conducted next-generation sequencing (NGS) on the Illumina MiSeq instrument, enabling the comprehensive analysis and understanding of the

full-length genome and the subsequent phylogenetic analysis of various SARS-CoV-2 strains (Gunadi *et al*., 2020). Similarly, another investigation sought to compare the genome sequences of SARS-CoV-2 that were obtained using both the Illumina NovaSeq and MiSeq instruments, subsequently validating the modified protocol that was employed for the purpose of high-throughput sequencing (Mze *et al*., 2023). Additionally, scientists decided to employ the Illumina MiSeq instrument in order to conduct quasispecies analysis of SARS-CoV-2, effectively highlighting and shedding light upon its pivotal role in the exploration and understanding of viral diversity (Bader *et al*., 2022). Furthermore, another study set out to assess the effectiveness and utility of the Illumina MiSeq ARTIC protocol, specifically in the context of whole-genome sequencing of SARS-CoV-2, effectively emphasizing its immense potential and value in the field of clinical microbiology (Plitnick *et al*., 2021). Moreover, researchers have successfully adapted and optimized the existing Illumina protocols, allowing for quick and accurate outbreak investigations of SARS-CoV-2, thus effectively showcasing the platform's inherent adaptability and suitability in the context of pandemic response (Pillay *et al*., 2020). Collectively, these studies serve to underscore and highlight the pivotal and indispensable role that the Illumina MiSeq platform plays in enabling the comprehensive and rapid whole-genome sequencing of SARS-CoV-2, thereby making significant contributions to the understanding of viral diversity, phylogenetic analysis, as well as outbreak management.

A key step in SARS-COV-2 sequencing is library preparation. Numerous methodologies have been elucidated for the preparation of SARS-CoV-2 libraries with the intent of sequencing. In one instance, improvements were delineated in the ARTIC multiplex PCR technique for SARS-CoV-2 genome sequencing employing nanopore technology, which has

garnered popularity owing to its simplicity and cost-effectiveness (Tyson *et al*., 2020). Another investigation devised a swift library preparation protocol predicated on tiled amplification of the S gene, followed by a PCR barcoding step and subsequent sequencing utilizing Nanopore platforms (Salazar *et al*., 2023). Likewise, the utilization of tiled amplification ensued by library preparation for sequencing on either Oxford Nanopore Technologies or Illumina platforms was underscored, encompassing the incorporation of sequencing adapters and barcodes to enable the simultaneous sequencing of multiple samples (Child *et al*., 2023). Additionally, an assessment was conducted on miniaturized Illumina DNA preparation protocols for SARS-CoV-2 whole-genome sequencing, which successfully demonstrated the preservation of sequence quality whilst concurrently reducing the volumes of library preparation reagents by a factor of eight (Pillay *et al*., 2023). Furthermore, an adaptation of the NovaSeq library was employed to sequence SARS-CoV-2 genomes from a staggering 384 samples utilizing the MiSeq instrument, thus maximizing the sequencing capacity in a singular run (Mze *et al*., 2023). In a separate study, the performance of sequencing SARS-CoV-2 genomes on both nanopore and Illumina platforms was exhibited and subsequently juxtaposed with the results obtained from the ARTIC LoCost nanopore method (Baker *et al*., 2021).

A rapid and cost-effective tailed amplicon method for sequencing SARS-CoV-2 was also described, entailing the preparation of libraries with sample-specific barcodes and sequencing via short-read (Illumina) or long-read (Oxford Nanopore, PacBio) technologies (Gohl *et al*., 2020). Furthermore, the creation of individual sequencing libraries from each sample was accomplished employing a commercial kit that is compatible with SARS-CoV-2 whole-genome sequencing (Simonetti *et al*., 2021). Collectively, these studies collectively

showcase the diverse array of approaches for SARS-CoV-2 library preparation for sequencing, encompassing methodologies for both nanopore and Illumina platforms, while simultaneously emphasizing the significance of cost-effectiveness, high throughput, and the preservation of sequence quality.

## 2.3 Immunogenicity of SARS-COV-2

The definition of immunogenicity may be stated as “the ability of a molecule or substance to elicit an immune response” or “the intensity of an immune response” (Murphy & Casey Weaver, 2017) In this definition, the term “immune response” refers to “an combined systemic response to an antigen, especially one arbitrated by lymphocytes and consist of recognition of antigens by specific antibodies or formerly sensitized lymphocytes”.

### 2.3.1 Immunity to SARS-COV-2

Once the SARS-CoV-2 virus enters into the host cell, the host immune system may recognize the complete virus or its surface epitopes, provoking innate and adaptive immunity. Pathogen recognition receptors (PRRs) existing on immune cells, primarily Toll-like receptors (TLR) 3, 7, and 8, are the first to recognize the virus. This leads to heightened interferon (IFN) synthesis (W. Hu *et al*., 2012; Le Bon & Tough, 2002; Thiel & Weber, 2008). The production of type I interferons (IFN) is initiated after a viral attack, promoting downstream JAK-STAT signal pathway for enhanced expression of IFN-stimulated genes (ISGs). Cross infection of the virus is controlled by NK cells and macrophages (G. Li *et al*., 2020). For its survival, the SARS-CoV virus blocks the production of IFN using the N protein (Mason, 2020). Theoretically, the lengthy incubation period of the virus successfully helps it to escape the innate immune response at the initial stage of infection (Prompetchara *et al*., 2020). The

innate immunity by type I IFN and monocyte-macrophages is not clear and need to be further studied in COVID-19 patients, as the dysregulated innate immune response has been seen to cause deadly pneumonia in youthful SARS-CoV-2-infected patients (Prompetchara *et al*., 2020).

In adaptive immunity, helper T cells trigger T-dependent B cells for SARS-COV-2 specific antibodies production. Cytotoxic T cells directly kill virus-infected host cells. Majorly, helper T cells enhances NF-Kb signaling pathway for pro-inflammatory cytokines production (G. Li *et al*., 2020). The humoral response towards SARS-CoV-2 has been shown to be similar to that towards other coronavirus infections, including the characteristic IgG and IgM production. At the beginning of SARS-CoV infection, B cells evoke an initial response against the N protein. Antibodies against S protein can also be detected in 4–8 days from the experience of first symptoms (Y. J. Tan *et al*., 2004; H. S. Wu *et al*., 2004). While N protein is smaller than S protein, it is extremely antigenic. The absence of glycosylation domains on it, results in production of N-specific neutralizing antibody at an early phase of severe infection (Meyer *et al*., 2014).

### 2.3.2 Effects of mutations on immunity

The rapid and extensive spread of the virus has been accompanied by the emergence of multiple viral variants through continuous genomic mutations (Lippi & Henry, 2021). These genetic variations in the SARS-CoV-2 genome have implications for viral infectivity, immune evasion, and vaccine efficacy (Meng *et al*., 2022).

One critical aspect of SARS-CoV-2 genome variation that affects immunity is its impact on viral entry into host cells. The spike (S) protein of SARS-CoV-2 plays a central role in viral

entry by binding to the angiotensin-converting enzyme 2 (ACE2) receptor on host cells (Iwata-Yoshikawa *et al*., 2022). Several studies have identified mutations in the spike protein that enhance viral infectivity and fusogenicity, leading to increased viral replication and transmission (Ao *et al*., 2021). For instance, the Omicron variant has been associated with an increased number of mutations in the spike protein, including H655Y and P681R, which enhance the endosomal entry pathway and reduce cell surface entry pathways, making the virus more efficient at infecting lung cells (Y. Hou *et al*., 2022; Y. Yamamoto *et al*., 2022). The alteration of entry preferences can impact the virus's ability to evade the host immune response and potentially alter disease severity (Strobelt *et al*., 2023).

Furthermore, some SARS-CoV-2 variants, such as Delta and Omicron, exhibit increased kinetic stability of open spike conformations, facilitating more efficient receptor binding and membrane fusion during viral entry (Z.-J. Yang *et al*., 2022). The altered kinetics of spike protein conformation can impact the host immune response by evading recognition by neutralizing antibodies and enhancing viral infectivity (Z.-J. Yang *et al*., 2021)

Another critical aspect of SARS-CoV-2 genome variation is its effect on antigenic escape. Continuous genomic mutations in the virus can lead to changes in viral epitopes, affecting the recognition and binding of neutralizing antibodies produced by the host immune system (Kuzmina, Wattad, *et al*., 2021). Studies have shown that certain variants, such as the Delta variant, exhibit a minor decrease in neutralization sensitivity to convalescent or post-vaccination sera, contributing to breakthrough infections in vaccinated individuals and potentially influencing the duration of vaccine-induced protection (Carabelli *et al*., 2023; Kuzmina, Wattad, *et al*., 2021).

Moreover, the emergence of SARS-CoV-2 variants with mutations in the spike protein's S1/S2 cleavage site has been shown to promote infectious potential and impact viral transmission (G. Liu & Gack, 2022). Mutations in this region can affect the efficiency of viral entry and the processing of the spike protein by host cell proteases, such as TMPRSS2 (Iwata-Yoshikawa *et al*., 2022). Consequently, these changes can alter viral fitness and contribute to the adaptation of the virus to its human host (Meng *et al*., 2022).

The SARS-CoV-2 genome variation also affects the viral interferon response, influencing the interaction between the virus and the host's immune system (Y. Hou *et al*., 2022). Some variants have been found to manipulate the antiviral interferon response, leading to enhanced viral replication and evasion of immune surveillance (G. Liu & Gack, 2022). For example, the Delta variant has been associated with increased infectivity and a decrease in interferon sensitivity compared to earlier variants (G. Liu & Gack, 2022). This alteration in the viral-host interaction can influence disease severity and the overall trajectory of the pandemic.

Importantly, SARS-CoV-2 genome variation impacts vaccine development and efficacy. The spike protein is the primary target for most COVID-19 vaccines, and changes in its sequence due to genomic mutations can affect vaccine-induced immune responses (Carabelli *et al*., 2023). Studies have shown that certain variants, like Omicron, exhibit a high number of mutations in the spike protein, leading to reduced vaccine efficacy against infection and transmission (Kuzmina *et al*., 2023). The frequent emergence of new variants poses challenges for vaccine manufacturers to keep pace with the evolving virus and highlights the need for ongoing surveillance and updated vaccine strategies (Carabelli *et al*., 2023).

In conclusion, the genetic variations in the SARS-Cov-2 genome have significant effects on immunity, viral infectivity, and vaccine response. Mutations in the spike protein can alter viral entry, promote antigenic escape, and impact the host immune response (Carabelli *et al*., 2023; G. Liu & Gack, 2022). Variants with enhanced infectivity and altered receptor binding kinetics can facilitate viral transmission and contribute to breakthrough infections (Y. Hou *et al*., 2022). Additionally, some variants exhibit increased resistance to neutralizing antibodies, potentially reducing vaccine efficacy (Kuzmina, Khalaila, *et al*., 2021). Understanding the effects of SARS-CoV-2 genome variation on immunity is crucial for developing effective public health strategies and improving vaccination efforts to control the pandemic. Continuous surveillance of viral evolution and adaptation is essential to stay ahead of the virus and safeguard global health (Carabelli *et al*., 2023).

### 2.3.3 Mechanisms Host immunity

Host immunity to SARS-CoV-2 is primarily mediated by both the humoral and cellular arms of the immune system. The spike protein, which plays a central role in viral entry into host cells, is the primary target for neutralizing antibodies generated during infection or vaccination (Jackson *et al*., 2021).

Studies have shown that infection or vaccination with the original Wuhan strain elicits a robust humoral immune response, leading to the production of neutralizing antibodies (Greaney, Starr, Eguia, *et al*., 2021). However, the emergence of SARS-CoV-2 variants with mutations in the spike protein has raised concerns about the effectiveness of these antibodies against newer strains (Choi & Smith, 2021) .

The Alpha variant, characterized by the N501Y mutation in the receptor-binding domain (RBD), has shown some escape from neutralization by convalescent sera and vaccine-induced antibodies (Hirabara *et al*., 2022). Similarly, the Beta variant, with the E484K mutation in the RBD, has been associated with reduced recognition by neutralizing antibodies (K. Chen *et al*., 2022). These findings highlight the importance of monitoring viral mutations and their impact on humoral immunity.

In response to the ongoing evolution of the virus, efforts have been made to develop vaccines targeting the spike protein. mRNA-based vaccines, such as Pfizer-BioNTech and Moderna, have been effective in inducing strong humoral immune responses (Mclean *et al*., 2022). However, studies have shown that some variants, particularly those with mutations in the RBD like E484K, may partially escape neutralization by vaccine-induced antibodies (Mittal *et al*., 2022). As a result, booster doses and updated vaccines targeting specific variants have been considered to enhance protection (Torbati *et al*., 2021).

In addition to humoral immunity, cellular immunity plays a crucial role in defending against SARS-CoV-2 infection. T cells, particularly CD4+ and CD8+ T cells, recognize and target infected cells, contributing to viral clearance and providing immune memory (Golota, 2022).

Studies have shown that T cells generated from natural infection or vaccination can recognize a broad range of SARS-CoV-2 variants, indicating that cellular immunity remains effective against different viral strains (Pecetta *et al*., 2022). This cross-reactive T cell response is attributed to the presence of conserved regions in the spike protein and other viral components (Mistry *et al*., 2022).

Moreover, studies have demonstrated that T cell responses in individuals with a history of prior infection or vaccination are associated with milder disease outcomes upon re-infection with SARS-CoV-2 variants (Errico *et al*., 2022). This suggests that T cell immunity may confer a level of protection against severe disease, even in the presence of viral variants.

The interplay between humoral and cellular immunity is essential for optimal protection against SARS-CoV-2 variants. Neutralizing antibodies can prevent viral entry into host cells, while T cells can recognize and clear infected cells. This dual approach is critical for controlling viral replication and limiting disease severity (Corti *et al*., 2021).

In response to the ongoing evolution of the virus and the potential for immune escape, the development of new therapeutic strategies has been explored. Monoclonal antibodies targeting the spike protein have been used as a treatment for COVID-19, and their use has expanded to include specific variants (Mclean *et al*., 2022).

Additionally, efforts are being made to develop pan-coronavirus vaccines that target conserved regions in the spike protein to provide broad protection against multiple coronaviruses, including SARS-CoV-2 variants (Pecetta *et al*., 2022).

In conclusion, the ongoing evolution of SARS-Cov-2 and the emergence of variants of concern have underscored the importance of understanding host immunity to combat the pandemic effectively. Humoral immunity, mediated by neutralizing antibodies, plays a crucial role in preventing viral entry, while cellular immunity, mediated by T cells, contributes to viral clearance and immune memory. The interplay between these arms of the immune system is critical for controlling viral replication and limiting disease severity.

Continued research and surveillance are essential to develop effective vaccines and therapeutic strategies to combat the evolving landscape of SARS-CoV-2 variants.

### 2.3.4 Role of Antibodies in Neutralizing the Virus

Antibodies are essential components of the adaptive immune response and play a central role in neutralizing viral infections. These specialized proteins are produced by B cells in response to the presence of foreign antigens, such as those displayed on the surface of the SARS-CoV-2 virus. The spike protein of the virus is the primary target of neutralizing antibodies, as it mediates viral entry into host cells. Several studies have investigated the impact of SARS-CoV-2 variants on the binding and neutralization potential of antibodies.

A study by Harvey *et al*. (2021) highlighted that some SARS-CoV-2 variants carry mutations in the spike protein, leading to reduced antibody recognition and neutralization (Harvey *et al*., 2021). For instance, the Beta variant, which harbors the E484K mutation in the receptor-binding domain of the spike protein, has been associated with a lower susceptibility to neutralization by convalescent plasma and vaccine-induced antibodies. Similarly, the Gamma variant, also known as the P.1 variant, possesses multiple mutations, such as K417T, E484K, and N501Y, which confer resistance to neutralization by some monoclonal antibodies (Choi & Smith, 2021).

The impact of SARS-CoV-2 mutations on vaccine efficacy has been a subject of intense investigation. Damavandi *et al*. (2022) reported that the spike protein mutations found in different variants can reduce the effectiveness of vaccines (Damavandi *et al*., 2022). The Omicron variant, in particular, has garnered significant attention due to its extensive spike protein mutations, including K417N, E484A, and N501Y, among others. Studies have shown

that vaccine-elicited antibodies may have reduced neutralizing activity against the Omicron variant compared to earlier variants, potentially affecting vaccine effectiveness (Mclean *et al*., 2022).

The humoral immune response to SARS-CoV-2 infection and vaccination goes beyond neutralizing antibodies. Fryer *et al*. (2022) discussed the role of B-cell memory, which includes memory B cells and plasma cells that secrete antibodies, in providing long-term protection against reinfection. While the Omicron variant may reduce neutralization by vaccine-induced antibodies, existing memory B cells can rapidly respond to the variant upon reinfection, offering some degree of protection (Fryer *et al*., 2022).

In addition to neutralizing antibodies, cellular immunity also plays a crucial role in controlling SARS-CoV-2 infections. Golota (2022) highlighted the significance of cellular immune responses mediated by T cells in recognizing and eliminating virus-infected cells (Golota, 2022). Even when neutralizing antibodies are less effective against certain variants, T cells can still recognize conserved viral epitopes and mount a targeted immune response. This dual response of both neutralizing antibodies and T cells provides a robust defense against the virus.

Efforts to tackle COVID-19 have extended to the development of neutralizing monoclonal antibodies. Corti *et al*. (2021) discussed the use of monoclonal antibodies as therapeutic agents to treat infected individuals or provide passive immunity. However, the emergence of variants has raised concerns about the potential loss of therapeutic efficacy. Continuous monitoring of viral evolution and antibody interactions is critical to adjust treatment strategies accordingly.

Despite the challenges posed by SARS-CoV-2 variants, vaccination remains a crucial tool in controlling the pandemic. Vaccines induce a diverse repertoire of antibodies, some of which may retain neutralizing activity against different variants (Mistry *et al*., 2022). Ongoing research and development of new vaccines or booster shots targeting specific variant antigens aim to enhance neutralizing activity and broaden immune protection.

In conclusion, the emergence of SARS-CoV-2 variants of concern has highlighted the importance of understanding the role of antibodies in neutralizing the virus. While some variants carry spike protein mutations that reduce neutralization by antibodies, existing memory B cells and T-cell responses can still contribute to protective immunity. Efforts to develop effective vaccines and therapeutic antibodies must consider the evolving viral landscape and adapt strategies to ensure continued success in combating the pandemic.

### 2.3.5 T Cell Responses and Their Significance in Viral Clearance

The ongoing COVID-19 pandemic caused by the SARS-CoV-2 with the emergence of various variants of concern, have raised significant global health concerns. Understanding the role of T cell responses in combating these variants is crucial for devising effective strategies to control the spread of the virus and prevent severe disease outcomes.

T cells are a crucial component of the adaptive immune response, playing a central role in recognizing and eliminating virus-infected cells. Two main types of T cells involved in the immune response against viral infections are CD8+ cytotoxic T cells and CD4+ helper T cells. CD8+ T cells recognize viral peptides presented on the surface of infected cells through major histocompatibility complex (MHC) class I molecules. Once activated, they can directly kill the infected cells, preventing further viral replication (Hirabara *et al*., 2022).

Studies have shown that T cell responses play a critical role in controlling SARS-CoV-2 infections and are not solely dependent on neutralizing antibodies. The role of T cells in viral clearance has been demonstrated in various studies. For instance, Lou *et al*. (2021) reported that patients with mild or asymptomatic COVID-19 had a robust and functional T cell response, suggesting that T cells play a crucial role in controlling viral replication and preventing disease progression (Lou *et al*., 2021).

As SARS-CoV-2 variants have emerged, questions have arisen regarding the impact of these mutations on T cell responses. Khamadi *et al*. (2022) conducted whole-genome sequence analysis in Kenya and found the circulation of multiple SARS-CoV-2 variants of concern (S. Khamadi et al., 2022). Despite the variant diversity, it is essential to assess whether T cell responses can recognize and target these variants to ensure effective viral clearance.

One of the major concerns with SARS-CoV-2 variants is their ability to evade the immune response, including T cell recognition. A study investigated the molecular basis of immune evasion by the Delta and Kappa variants. They reported that the Delta variant exhibited reduced sensitivity to neutralizing antibodies and altered recognition by CD8+ T cells. Similarly, the Kappa variant displayed escape from neutralization by convalescent plasma and vaccine-induced antibodies, leading to concerns about potential immune evasion by T cells as well (McCallum *et al*., 2021).

While some variants may partially evade T cell responses, other studies have shown that T cell immunity remains broad and can recognize multiple viral epitopes, including those found in different variants. Hirabara *et al*. discussed the potential of immune evasion by SARS-CoV-2 variants but emphasized that T cell responses to conserved epitopes can still offer

protection against different viral strains (Hirabara *et al*., 2022). This finding highlights the significance of T cell responses in maintaining immune surveillance against a diverse range of variants.

In addition to natural infection, T cell responses elicited by vaccination have shown promise in providing protection against SARS-CoV-2 variants. Chen *et al*. (2022) discussed the evolution of SARS-CoV-2 variants and the impact on vaccine development. Despite concerns about reduced neutralization by vaccine-induced antibodies against certain variants, T cell responses induced by vaccination can still recognize and target multiple epitopes on the virus, potentially providing cross-variant protection (D.-Y. Chen *et al*., 2022).

Furthermore, T cell responses may be more durable than antibody responses, as they can persist for an extended period, contributing to long-term immune memory. Understanding the longevity of T cell responses to variants will be critical in assessing vaccine efficacy and the need for booster doses.

However, it is essential to continue monitoring T cell responses and their impact on viral clearance in the context of SARS-CoV-2 variants. New variants may arise with additional mutations that could further impact T cell recognition and immunity. Therefore, ongoing research is crucial in identifying potential escape mutants and developing strategies to counteract their immune evasion capabilities.

In conclusion, T cell responses play a significant role in combating SARS-CoV-2 infections and are critical in viral clearance. As SARS-CoV-2 variants of concern continue to emerge, understanding the impact of these mutations on T cell recognition and immunity is essential. T cells have the capacity to recognize conserved epitopes, providing cross-variant protection.

Still, continuous monitoring and research are necessary to ensure the effectiveness of T cell responses against the evolving virus and to inform the development of vaccines and therapeutic interventions.

### 2.3.6 Effects of SARS-CoV-2 Genome Variation on Innate Immunity

SARS-CoV-2 genomic variations, particularly in the spike protein, can influence viral recognition by pattern recognition receptors (PRRs) of the innate immune system. PRRs detect viral components, triggering the release of pro-inflammatory cytokines and type I interferons (IFNs). Variants like B.1.351 and P.1 have shown the ability to escape detection by PRRs, leading to reduced IFN production and delayed immune response (M Hoffmann *et al*., 2021). These alterations in viral recognition can facilitate viral replication and dissemination, potentially contributing to the increased transmissibility observed in certain variants (J. Davies *et al*., 2020).

IFNs are essential antiviral effectors that establish an antiviral state within infected cells and neighboring uninfected cells. Some SARS-CoV-2 variants, such as B.1.1.7 and B.1.351, have demonstrated the ability to suppress the IFN response (Tegally *et al*., 2021). This evasion of the IFN response allows the virus to replicate more efficiently and spread within the host. Additionally, genomic variations in the viral non-structural proteins can influence their ability to antagonize host proteins involved in the IFN pathway (Bushman *et al*., 2021). By evading the IFN response, SARS-CoV-2 variants can evade the early antiviral defenses of the host's immune system, potentially leading to more severe and prolonged infections.

Innate immune activation triggers the release of pro-inflammatory cytokines to recruit and activate immune cells. Some SARS-CoV-2 variants, like B.1.1.7, have been associated with

increased inflammation and higher viral loads in the upper respiratory tract (Bushman *et al*., 2021). Excessive cytokine production can lead to an uncontrolled inflammatory response, resulting in a cytokine storm. These storms are associated with severe COVID-19 cases and can cause extensive tissue damage, particularly in the lungs. Additionally, the presence of certain SARS-CoV-2 variants has been linked to increased expression of pro-inflammatory chemokines, further exacerbating inflammation and tissue damage (M Hoffmann *et al*., 2021).

SARS-CoV-2 variants can exhibit differential recognition by innate immune sensors, resulting in distinct immune responses. For instance, certain variants may have altered structures or modifications in their genomic material that affect their detection by immune sensors (Bushman *et al*., 2021). This differential recognition can lead to variation in the magnitude and nature of the immune response, potentially influencing disease severity and clinical outcomes.

The impact of SARS-CoV-2 genome variation on innate immunity is complex and multifaceted. Genomic changes in the virus can alter its recognition by PRRs, evade the IFN response, and modulate the activity of innate immune cells. These alterations can lead to differences in viral replication, immune activation, and inflammatory responses. Understanding the interplay between SARS-CoV-2 variants and innate immunity is critical for developing effective strategies to combat the ongoing pandemic. Continuous surveillance and research on the effects of genomic variation on innate immunity are essential to guide public health measures and vaccine development efforts to address the evolving challenges of the COVID-19 pandemic.

### 2.3.7 Epitopes and Their Relevance in SARS-CoV-2 Immune Recognition

The immune system's capacity to identify and respond to pathogens such as the SARS-CoV-2 relies on epitopes – specific antigenic determinants that stimulate immune reactions. Epitopes, or antigenic determinants, are distinct molecular segments present on antigens – the surface proteins of pathogens. These epitopes serve as recognition sites for immune cells, allowing them to discriminate between self and foreign molecules and to initiate targeted immune responses. Epitopes essentially function as markers guiding the immune system in identifying and neutralizing invading pathogens.

Epitopes, which are specific regions of antigens, engage in molecular interactions with receptors in the immune system (Peters *et al*., 2005). These epitopes hold immense significance in the context of immune responses, as they facilitate the identification of pathogens and the creation of vaccines. Linear and conformational epitopes represent two distinct types, with the former being linear sequences of amino acids and the latter being formed by amino acids that assemble within the folded structure of the antigen (Albukhari *et al*., 2002). The immune system possesses the ability to recognize these epitopes, thereby initiating a cascade of events that activate immune cells, including T cells and B cells.

Multiple studies have shed light on the critical role of epitopes across various scenarios, ranging from comprehending antibody responses to allergens (Mueller *et al*., 2019) to mapping T cell responses during viral infections such as dengue (Chaudhury *et al*., 2017; Grifoni *et al*., 2017) and identifying epitopes for developing vaccines against pathogens like HIV (J. Kim *et al*., 2018; Richmond *et al*., 2011). Notably, in the case of influenza, it has been observed that while current vaccines predominantly trigger responses to variable

epitopes in the hemagglutinin head region, they fail to enhance responses to the conserved stem region (Ellebedy *et al*., 2014). This observation underscores the significance of strategically targeting specific epitopes to elicit broad and cross-reactive immune responses. Moreover, the diversity of epitopes within pathogens such as Ebola and measles carries substantial implications for the design of immunotherapies and vaccines (Saphire & Aman, 2016). Gaining a comprehensive understanding of the epitopes recognized by neutralizing antibodies is of utmost importance in the development of effective vaccines that can withstand viral evolution and immune evasion (Greaney, Starr, Barnes, *et al*., 2021). Furthermore, the identification of epitopes that trigger polyfunctional responses in diseases like HIV underscores the need for a thorough evaluation of vaccine candidates (Richmond *et al*., 2011). In conclusion, epitopes represent fundamental constituents of the immune response, playing a pivotal role in the recognition of antigens, the development of vaccines, and the comprehension of immune reactions against diverse pathogens. Ongoing research on epitopes continues to yield invaluable insights into the formulation of effective vaccines and immunotherapies against infectious diseases.

### 2.3.8 Linear and conformational epitopes

The epitopes can be classified as linear, consisting of continuous amino acid sequences, or conformational, formed by amino acids brought together in the folded antigen structure (Malonis *et al*., 2019). Linear epitopes are recognized by antibodies even without a specific conformation but often adopt local secondary structures upon binding (Gottardo *et al*., 2013). The immune system's recognition of these epitopes leads to the activation of immune cells like T cells and B cells. Studies have shown that conformation-independent epitopes are typically linear stretches of residues, with some adopting secondary structures upon antibody

binding (Malonis *et al*., 2019). Linear epitopes in specific regions of antigens have been associated with reduced infection risk in vaccine trials (Gottardo *et al*., 2013). Peptide-based vaccines are under development to induce specific humoral responses by promoting peptide secondary structures (W. Li *et al*., 2014). The prediction of continuous B-cell epitopes in antigens using neural networks has demonstrated moderate sensitivity, specificity, and positive prediction values (Saha & Raghava, 2006). Research on HIV and SARS-CoV-2 has emphasized the significance of linear epitopes in inducing virus neutralization and protection (Korber *et al*., 2017; Lü *et al*., 2021). Antibodies targeting epitopes can facilitate functions like virus neutralization and antibody-dependent cellular cytotoxicity (Pollara *et al*., 2013).

Computational tools are increasingly utilized to predict epitopes and design vaccines against various pathogens (A. Gupta *et al*., 2020; Prasasty *et al*., 2019). Additionally, investigations into epitope-based DNA vaccines have shown promising results in focusing antibody responses (Mitchell *et al*., 2017). Terefore, epitopes, whether linear or conformational, are vital for immune recognition and response. Understanding these epitopes is crucial for vaccine development and combating infectious diseases. Ongoing research aims to explore epitope mapping, prediction, and their role in designing effective vaccines against a range of pathogens.

Antibodies targeting linear epitopes are less influenced by the protein's conformation, making them more robust in various conditions (J. P. Moore & Ho, 1993). These antibodies can bind to small peptides or unfolded allergens, indicating their versatility in antigen recognition (Mueller *et al*., 2019). Linear epitope-specific antibodies have been found to be broadly cross-reactive against diverse strains of viruses like hepatitis C (Ahsan *et al*., 2021). Predicting linear B cell epitopes is particularly valuable as they can be recognized by

antibodies even when isolated from the protein context as synthetic peptides (Ras-Carmona *et al*., 2021). While antibodies recognizing linear epitopes do not require a specific conformation for binding, they still bind to a specific conformation of the linear epitope, indicating a conformational component in their recognition (Forsström *et al*., 2015). Linear epitope antibodies may be preferred when the target antigen is denatured during sample preparation, highlighting their practical utility in certain applications (Tuong *et al*., 2021).

In contrast to linear epitopes, antibodies targeting conformational epitopes, which are more prevalent in infected individuals, are highly sensitive to protein conformation (Robert-Guroff *et al*., 1994). Conformational epitopes, also referred to as discontinuous epitopes, are comprised of residues that are not sequentially adjacent but are spatially close in three-dimensional space, highlighting the importance of protein folding in antigenic reactivity (Ras-Carmona *et al*., 2021).

Research suggests that the majority of antibodies generated against allergens target conformational epitopes, emphasizing the influence of the three-dimensional structure on antibody specificity (Madsen *et al*., 2014). Moreover, the most antigenic epitopes are intricate structures made up of various protein segments in a three-dimensional conformation, underscoring the necessity of considering the overall protein structure in epitope prediction (Y. Chen *et al*., 2023; Raoufi *et al*., 2019). Overall, the distinction between linear and conformational epitopes is crucial in understanding antibody-antigen interactions. Linear epitopes offer advantages in terms of stability and ease of recognition, making them valuable targets for various research and diagnostic applications.

### 2.3.9 B and T cell epitopes

The identification and prediction of B cell epitopes hold great significance in the field of vaccine design and the comprehension of immune responses. To this end, various computational tools and methods have been developed, encompassing neural networks and rule mining systems (Yajie Cao et al., 2016; Xie et al., 2020). The objective of these tools is to enhance the accuracy of epitope prediction and increase the success rate of experimental identification (Yajie Cao *et al*., 2016). However, it is worth mentioning that the experimental determination of B cell epitopes has traditionally focused more on linear epitopes, despite the belief that conformational epitopes account for the majority of B cell epitopes (EL-Manzalawy & Honavar, 2010). Conformational epitopes are considered more pertinent in terms of antibody recognition; however, isolating them from their protein context can prove to be a challenging endeavor (Ras-Carmona *et al*., 2021). In conclusion, B cell epitopes play a pivotal role in immune responses, and it is of utmost importance to comprehend their nature, categorization, and prediction methodologies for the advancement of vaccine development and immunological research. While the study of linear epitopes is more straightforward from an experimental standpoint, it is widely believed that conformational epitopes are more prevalent and crucial for antibody recognition. The continuous development of computational tools continues to revolutionize the field of B cell epitope prediction, thereby facilitating the design of effective vaccines and enhancing our understanding of immune responses.

T cells epitopes, which are particular parts of antigens recognized by T cells, which is a vital part of immune system, have an important function in the immunodominance and memory formation in immune responses (Kwok *et al*., 2012). The presence of these T cell epitopes

within various other antigens like ESAT-6, CFP-10 as well as viral proteins i.e. SARS-CoV-2 influences the recognition and response of the immune systems significantly (Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, *et al*., 2020). Furthermore, there is great need to determine and foretell the presence of t-cell epitopes for numerous applications such as vaccine development and immunotherapy. An example is when T-cell epitope-specific peptide vaccines against pathogens like Bacillus cereus or Echinococcus multilocularis are designed so that protective immune responses can be induced (Pang *et al*., 2022; Rasheed *et al*., 2021). Besides this fact it is also useful to know about different proteins such as lectins or outer membrane proteins that derive T cell epitopes, because they act as the vaccine candidate too (Ranjbarian *et al*., 2023). The discipline of immunoinformatics has had a key role to play which has helped predict T cell epitopes that can lead to stronger immunity response (Zaheer *et al*., 2018). These tools are useful in identifying some peptides bearing sequence resemblance with human sequences or having weak antigenic properties thus giving helpful clues about potential tolerance induction (J. Li *et al*., 2011). Another proof of the significance of an epitope presentation for a proper immune activation is that it has been shown that joining together b cells and t cell epitopes at one molecular platform improves T-cell help and enhances antibody responses (Cañas-Arranz *et al*., 2021).

HLA type significantly influences epitope specificity. Research has shown that HLA-DR and -DQ alleles can impact autoantibody recognition of distinct epitopes within specific autoantigens (Richardson *et al*., 2015). Certain HLA alleles can lead to differences in antibody binding and affect the recognition of particular epitopes (Richardson *et al*., 2015).

Additionally, HLA class I genotype has been identified as a determinant of epitope selection, with mutations in immunodominant epitopes altering the pattern of cytotoxic T lymphocyte responses (Goulder *et al*., 1997).

Studies have demonstrated that modifications in MHC contact residues can enhance the immunogenicity of specific epitopes (Vertuani *et al*., 2004). The binding affinity of epitope peptides to HLA molecules can be influenced by specific residues, thereby affecting the presentation of epitopes to T cells (Borghan *et al*., 2005). Furthermore, individuals' HLA supertypes have been found to strongly predict epitope-specific responses, although responses to epitopes outside the predicted HLA type have also been observed (C. C. Wilson *et al*., 2003). Moreover, the diversity of HLA alleles can impact the repertoire, specificity, and binding characteristics of epitopes presented to T cells (Matthews *et al*., 2012). The interaction between HLA shared epitope alleles and environmental factors like smoking can influence autoimmunity and the development of autoantibodies (Willemze *et al*., 2011). These findings underscore the complex relationship between HLA type and epitope recognition, underscoring the necessity of considering HLA diversity in understanding immune responses.

The investigation of SARS-CoV-2 epitopes is of utmost importance in comprehending the intricacies of immune responses and the creation of effective vaccines. Utilizing computational prediction algorithms, researchers have successfully employed this approach to predict the presence of potent B and T-cell epitopes in the structural proteins of both SARS-CoV-2 and related viruses (Pinto *et al*., 2020). These epitopes serve a pivotal role in the humoral immune response directed against SARS-CoV-2, as emphasized by Pinto *et al*. in their research. Moreover, the profound serological profiling undertaken has unveiled more

than 800 epitopes within the SARS-CoV-2 proteome, with a subset likely to be acknowledged by neutralizing antibodies (Shrock *et al*., 2020). The identification of these epitopes is of utmost importance in the field of vaccine design, as they possess the ability to elicit protective T cell immunity against SARS-CoV-2 and its various strains (Lü *et al*., 2021). Numerous studies have shed light on the conservative nature of T cell epitopes, thereby suggesting their potential inclusion in vaccines as a means to induce a state of protective immunity, as underscored by Lu *et al*. in their research. In addition, Chouchane *et al*. have successfully illustrated the presence of neutralizing antibodies that specifically target certain epitopes found in the SARS-CoV-2 spike protein (Chouchane *et al*., 2021). Remarkably, these antibodies are capable of targeting both conformational and linear epitopes, thus signifying the crucial role played by epitope diversity in eliciting immune responses, as outlined by Chouchane *et al* (2021) in their study. Ongoing efforts are currently underway to identify SARS-CoV-2 epitopes that prompt the development of protective T cell responses, thereby underscoring the necessity of delineating T cell receptor repertoires that are specifically attuned to these epitopes (D. Wu *et al*., 2021). The immune target epitopes of SARS-CoV-2 constitute a critical aspect of comprehending the host's immune responses and developing effective therapeutic and preventive measures (Kiyotani *et al*., 2020). The existence of certain epitopes that are conserved among different coronaviruses, including SARS-CoV and bat coronaviruses, underscores the potential for a broad-spectrum immunity targeting these shared epitopes (Lü *et al*., 2021).

In summary, T-cell epitopes compromise key parts of antigens causing immune responses and are important in vaccination creation, immunotherapy creation as well as our understanding of immunity recognition. The detection, prediction and use of t-cell epitopes

therefore assist to formulate better ways to combat infections diseases or other conditions related with the immune system. The in-depth exploration of SARS-CoV-2 epitopes is an indispensable component in comprehending immune responses, designing effective vaccines, and developing therapeutic interventions aimed at combating COVID-19. The identification of conserved epitopes, cross-reactive epitopes, and neutralizing epitopes provides invaluable insights into the immune targeting of SARS-CoV-2, while concurrently shedding light on the potential for broad protection against closely related coronaviruses.

### 2.3.10 Structural and non-structural protein epitopes

The SARS-COV-2 epitopes are seen on both structural and non-structural proteins of the virus. The primary structural proteins of SARS-CoV-2 are the spike (S) glycoprotein, nucleocapsid (N) protein, membrane (M) glycoprotein, and small envelope (E) glycoprotein (B. Wang *et al*., 2021). The structural proteins contribute significantly to the overall architecture of the virus. The spike protein is crucial for viral entry into host cells and is a key target for vaccine development due to its antigenic properties (Dongliang Wang *et al*., 2020). The arrangement of these structural proteins forms the virion's supramolecular organization, with the proteins forming specific patterns within the viral envelope (Dongliang Wang *et al*., 2020). Understanding the structural proteins of SARS-CoV-2 is vital for developing therapeutic interventions. Targeting these proteins, especially the spike protein, can lead to the development of vaccines and antiviral drugs that disrupt viral entry and replication. The interactions and functions of these structural proteins provide insights into the virus's life cycle and pathogenesis, offering potential targets for therapeutic strategies aimed at combating COVID-19.

The spike protein is highlighted as the main structural protein in both SARS-CoV-2 and SARS-CoV (Shekhawat & Chowdhury, 2022). The spike protein plays a crucial role in viral infection by interacting with the human Angiotensin-Converting Enzyme 2 (ACE2) receptor (Xie et al., 2020). It is an indispensable constituent situated on the external surface of the virus, comprises a diverse array of epitopes that are acknowledged by the immune system. These epitopes are specifically targeted for the purpose of generating vaccines and therapeutic interventions. The spike protein assumes a pivotal role in various indispensable functions including viral attachment, fusion, and entry into the host cells (Tai *et al*., 2020).

Numerous T cell and B cell epitopes have been identified within the spike glycoprotein through rigorous studies, and it is these epitopes that are crucial for the process of immune recognition and response (Baruah & Bose, 2020). Furthermore, there is a highly conserved epitope present within the receptor-binding domain of the spike protein, which facilitates cross-reactive binding between SARS-CoV-2 and SARS-CoV, thereby underscoring the significance of these specific epitopes in eliciting immune responses (Yuan *et al*., 2020). The spike protein is often described as a pivotal component for the purpose of achieving effective immunization against SARS-CoV-2, as it has the ability to induce the production of neutralizing antibodies and precisely target certain structural domains within the protein (Vernet *et al*., 2021). Additionally, the spike glycoprotein is a primary target for vaccine development due to its critical involvement in the process of viral entry into the host cells (Vishweshwaraiah *et al*., 2022).

Extensive research has also been dedicated to the identification of novel antibody epitopes within the spike protein, and these epitopes hold the potential to serve as potent targets for antibody-based therapies and vaccines (Zheng & Lun, 2020). Several studies have

demonstrated that the SARS-CoV-2 spike protein exhibits similarities with other viral proteins, such as paramyxovirus surface proteins, thereby suggesting the possibility of cross-reactivity and shared epitopes (Ahmadi *et al*., 2021).

The identification of conserved epitopes within the spike protein is of utmost importance as it aids in the comprehension of immune responses and the development of efficacious vaccines (Forcelloni *et al*., 2021). Furthermore, the structural dynamics of the spike protein, including the variations observed in the epitopes, play a significant role in determining the antigenicity of the virus and its recognition by the immune system (Guerin *et al*., 2022). In conclusion, the epitopes that are present within the SARS-CoV-2 spike protein hold crucial significance in terms of immune recognition and response. Gaining a comprehensive understanding of these epitopes, their degree of conservation, as well as the variations they may exhibit, is absolutely vital for the development of effective vaccines, therapeutics, and diagnostic tools that can successfully combat COVID-19.

The receptor-binding domain (RBD) of the SARS-CoV-2 spike protein plays a pivotal role in the process of viral attachment and entry into host cells. This is achieved through a direct interaction with the host cell receptor, known as angiotensin-converting enzyme 2 (ACE2). Numerous studies, including those conducted by Shang *et al*. (2020), Tai *et al*. (2020), and Ma *et al*. (2021), have provided evidence demonstrating the RBD's specific recognition of ACE2 as its receptor, thereby facilitating the entry of the virus into host cells (Ma *et al*., 2021; Shang *et al*., 2020; Tai *et al*., 2020). It is important to note that the interaction between the RBD of the viral spike protein and ACE2 is not only essential for viral transmission but also crucial for its infectivity (Borkotoky *et al*., 2022). Extensive research has shown that the

binding affinity between the RBD domain and ACE2 is significantly higher in SARS-CoV-2 compared to its predecessor, SARS-CoV.

This finding, highlighted by Angeli *et al*. and He *et al*. underscores the immense importance of this interaction in the viral pathogenesis of SARS-CoV-2 (Angeli *et al*., 2021; J. He *et al*., 2020). Furthermore, investigations into the impact of mutations on the binding efficacy of SARS-CoV-2 variants with the human ACE2 receptor have placed a strong emphasis on the need to comprehend the structural alterations occurring within the RBD. This knowledge is critical for understanding viral infectivity and the potential development of antibody resistance (Costa et al., 2022; A. Khan et al., 2022). To combat the ongoing threat of COVID-19, substantial efforts have been dedicated to the development of inhibitors and vaccines that specifically target the RBD (Hampton *et al*., 2022; Hanke *et al*., 2020). Additionally, computational models have played a vital role in predicting the binding affinity of new spike variants for ACE2. These models have further highlighted the significance of mutations occurring within the RBD in determining viral infectivity (Tragni *et al*., 2021, 2022). In conclusion, the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein is an indispensable factor in the viral pathogenesis of COVID-19. Understanding the intricate interactions between the RBD and ACE2 is of paramount importance in the development of effective therapeutic interventions and vaccines against this highly contagious disease.

The nucleocapsid (N) protein is essential for virus replication and assembly, as it has been found to have the most interactions among the structural and accessory proteins of SARS-CoV-2 (M. Chen *et al*., 2021). It aids in viral replication and assembly, maintaining the virus's structure (M. Chen *et al*., 2021). It assumes the form of a ribonucleoprotein complex, thereby effectively safeguarding the viral RNA from any potential interference by host factors,

thereby greatly assisting in the processes of viral replication and assembly (Mao *et al*., 2023). A noteworthy fact is that this protein is expressed at exceedingly high levels within infected cells, thereby serving as a catalyst for enhancing the efficiency of viral RNA transcription, and accordingly, it stands as an indispensable player in the realm of viral replication (Savastano *et al*., 2020).

Research endeavors have successfully managed to identify a number of B and T cell epitopes within the nucleocapsid protein, thereby rendering it a matter of utmost significance in the context of immune responses and the development of vaccines (H. Chen *et al*., 2020; Oliveira *et al*., 2020). In addition to these roles, the nucleocapsid protein itself has emerged as a major diagnostic marker for SARS-CoV-2 infection, as well as the field of immune protection (Terry *et al*., 2021). It is pertinent to note that the phosphorylation sites of the nucleocapsid protein do indeed exert an influence on the intricate interactions that take place between the host and the pathogen, thereby influencing viral genome processes in a profound manner (Tung & Limtung, 2020). The protein in question undergoes a substantial degree of phosphorylation and it is this feature that enables it to establish interactions with viral RNA, ultimately culminating in the formation of the viral core and nucleocapsid (Pk *et al*., 2020). Furthermore, it is essential to underscore the fact that the protein has been the subject of extensive exploration due to its potential as a target for the development of vaccines, given its proven ability to incite cytotoxic T lymphocyte responses (H. Chen & Wu, 2020).

Epitope mapping studies have consistently managed to shed light on the existence of conserved epitopes within the nucleocapsid protein, thereby contributing to a deeper understanding of immune responses and also serving as a potential blueprint for the design of vaccines (Forcelloni *et al*., 2020, 2021). Moreover, it is of utmost importance to

acknowledge the fact that any mutations that may arise within the nucleocapsid protein have the potential to exert a significant impact on its antigenicity, consequently influencing the extent to which the immune system acknowledges and responds to this protein (Sushant Kumar *et al*., 2022).

The identification and subsequent detection of this protein within infected individuals has been employed for diagnostic purposes, thereby effectively highlighting the immunogenic nature of this protein (Lau *et al*., 2004). While it is true that antibodies that are specific to the nucleocapsid protein may not possess the ability to neutralize cell-free SARS-CoV-2, it is worth mentioning that they have demonstrated a certain degree of potential in mediating antibody-dependent cell-mediated cytotoxicity (Dangi, Sanchez, Class, *et al*., 2022). In conclusion, the nucleocapsid protein of SARS-CoV-2 assumes an absolutely vital role in the life cycle of the virus, immune responses, and diagnostic strategies. The epitopes, phosphorylation sites, and mutations that are associated with this protein each occupy a significant position in the broader understanding of viral pathogenesis, immunity, and the development of vaccines.

The membrane and envelope proteins are involved in virus assembly and budding (Dongliang Wang *et al*., 2020). The identification of epitopes that can be recognized by the immune system has been attributed to the membrane proteins of SARS-CoV-2, which encompass the membrane glycoprotein (M) and non-structural protein-13 (NSP13) (K. Pan *et al*., 2021). Although these epitopes may exhibit a lower degree of immunogenicity when compared to the spike and nucleocapsid proteins, they still fulfill a significant role in the induction of a T cell response (K. Pan *et al*., 2021). Furthermore, the host immune system has been found to

target the structural proteins of SARS-CoV-2, including the spike, membrane, nucleocapsid, and envelope proteins (J. Deng *et al*., 2021).

It has been noted that unique epitopes can be found in non-structural proteins (NSP) 1 and 3, as well as the surface glycoprotein of SARS-CoV-2 (Guevarra & Ulanday, 2020). Additionally, studies have successfully identified T-cell epitopes from the surface proteins of SARS-CoV-2, which signifies the presence of immunogenic regions within these specific proteins (Kibria *et al*., 2022). Moreover, the existence of epitope similarities between SARS-CoV-2 and other viruses, such as Bacillus Calmette–Guérin, hints at the potential for cross-reactive adaptive immunity to be established (Urbán *et al*., 2020).

The incorporation of epitopes from endogenous proteins into recombinant antigens could potentially offer protection against newly evolved genomic variants of SARS-CoV-2 (Barman *et al*., 2022). Furthermore, the glycosylated structural protein M of SARS-CoV-2 has been discovered to be indispensable for viral particle assembly (Fu *et al*., 2020). Additionally, the investigation into the landscape of antibody binding in SARS-CoV-2 infection has shed light on the immunogenicity of the coronavirus membrane (M), ORF3a, and ORF8 proteins (Heffron *et al*., 2021). Various research studies have successfully predicted immunodominant linear B-cell epitopes on the trimeric S protein of SARS-CoV-2, thereby highlighting regions that possess a high degree of antigenicity (Dongliang Wang *et al*., 2020). To conclude, it is evident that while the spike and nucleocapsid proteins of SARS-CoV-2 serve as the primary targets for the immune system, the membrane proteins also harbor epitopes that have the capability to elicit immune responses. A comprehensive understanding of the immunogenicity exhibited by different viral proteins is of paramount

importance in the development of effective vaccines and therapeutic strategies against COVID-19.

The envelope proteins of SARS-CoV-2, although they may exhibit a potentially lower level of immunogenicity when compared to the spike and nucleocapsid proteins, still possess specific regions, known as epitopes, that are capable of being recognized by the complex and intricate immune system (S. F. Ahmed *et al*., 2020). These envelope proteins, which are integral components of the SARS-CoV-2 virions, alongside the nucleocapsid, spike, and membrane proteins, collectively contribute to the overall structure and composition of the virus (B. Wang *et al*., 2021).

Through extensive investigation, it has been revealed that the envelope protein of SARS-CoV-2 can be subjected to immunoinformatics analysis to effectively evaluate the potential for cross-protection against COVID-19 (Tilocca *et al*., 2020). Furthermore, compelling evidence has emerged indicating that the envelope protein shares certain epitopes with other members of the betacoronavirus family, thereby suggesting the possibility of cross-recognition by the intricate immune system (Tajuelo *et al*., 2022). The identification and comprehensive understanding of these epitopes within the envelope protein play a pivotal role in the development and design of effective vaccines, as well as in the elucidation of the immune response to SARS-CoV-2 (Lü *et al*., 2021). Extensive research has highlighted the utmost significance of epitope profiling in order to ascertain the fundamental antigenic determinants within SARS-CoV-2, including those that are primarily located in the receptor binding motif (RBM), which is of critical importance (Jiang *et al*., 2021). Moreover, the envelope protein, alongside the spike, membrane, and nucleocapsid proteins, assumes a

substantial role in the overall structural framework of SARS-CoV-2, thereby underscoring its significance (Jain *et al*., 2020).

Numerous scientific investigations have concentrated their efforts on the prediction of epitopes housed within the envelope protein of SARS-CoV-2, employing sophisticated immunoinformatics methodologies and approaches (Ghafouri *et al*., 2020; Jakhar & Gakhar, 2020). These studies strive to harness the power of computational techniques in order to identify potential targets for vaccine development, as well as to gain a comprehensive understanding of the intricate immunopathogenesis underlying diseases such as Guillain-Barré syndrome (Morsy, 2020). Furthermore, the preservation and conservation of epitopes between SARS-CoV and its counterpart SARS-CoV-2 present intriguing possibilities regarding potential similarities in immune responses elicited by these closely related viruses (Dongliang Wang *et al*., 2020). In conclusion, although the envelope proteins of SARS-CoV-2 may not exhibit the same level of immunogenicity as other structural proteins, they nevertheless encompass specific epitopes that are capable of being recognized by the multifaceted immune system. Consequently, acquiring a comprehensive understanding of these epitopes assumes paramount importance in the realms of vaccine design, cross-protection assessment, and the elucidation of the intricate immune response to COVID-19.

The non-structural proteins of SARS-CoV-2 contain epitopes that are recognized by the immune system. Although much has been talked about regarding structural proteins such as spike (S), envelope (E), membrane (M) and nucleocapsid (N), recent studies have also looked into epitopes within non-structural proteins (G. Yang *et al*., 2023). The specific epitopes came to light through the use of immunoinformatics analyses, making it possible to predict

and map T-cell epitopes based on known immune responses to these two viruses' infections, SARS-CoV, and SARS-CoV-2 (Devi *et al*., 2021).

It is crucial that these epitopes be identified since they function as key mediators of antiviral immunity responsible for sustaining T cell immunity even in individuals lacking detectable antibody responses to an infection (Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, *et al*., 2020). Further research indicates that the non-structural proteins of severe acute respiratory syndrome coronavirus 2 share conserved epitopes with those found in the non-structural proteins of severe acute respiratory syndrome coronavirus (Grifoni *et al*., 2020a). This observation then implies a possibility for cross-reactivity and immune recognition between these two viruses. Thus, understanding cross-reactivity is very important in appreciating the induced immune responses by both viruses and designing vaccines against them targeting these shared epitopes (S. F. Ahmed *et al*., 2020).

Presence of such epitopes among other proteins highlights the fact that multi-target vaccines should be considered for designing vaccines against COVID-19. This will help include both structural and non-structural proteins so that a stronger immune response can be raised against it (Gauttier *et al*., 2020). To summarize briefly, it seems clear that both structural protein and Nsps contains several protective targets being recognized by host reactions for SARS-CoV-2. This understanding is vital in the development of vaccines and therapeutics against the virus, as they are the major factors that induce immune responses and confer protection after infection.

### 2.3.11 Immunodominant Epitopes

Immunodominant epitopes, are crucial constituents in the immune response directed against pathogens, which play a significant and pivotal role in the development of vaccines and diagnostics. These epitopes, which are localized and specific regions on a protein of the pathogen, elicit a robust and vigorous immune response, particularly through the recognition by antibodies that are produced by the immune system (Gruta *et al*., 2006). Studies have demonstrated that the recruitment of specific T cells into a response that is epitope-specific is influenced by various factors such as the dose of the antigen and the frequencies of precursor cells, ultimately shaping the composition of the repertoire of T cell receptors that are specific to the epitope (Gruta *et al*., 2006). Moreover, the existence of an immunodominant epitope on a single viral protein can have an impact on the immune responses that are induced by epitopes located on other viral proteins, thereby indicating a complex and intricate interplay that exists among different epitopes (Rodríguez *et al*., 2002).

Research has also emphasized the utmost importance of identifying immunodominant epitopes in diverse pathogens. For instance, in the case of SARS-CoV-2, regions that are immunodominant have been precisely defined and delineated based on criteria that are pertinent to B-cell and T-cell epitopes, mapping, antigens that are predicted to be protective, and the similarity that they bear to experimentally validated epitopes of SARS-CoV (Mukherjee *et al*., 2020). Similarly, in the context of Mycobacterium tuberculosis, specific immunodominant epitopes have been identified and linked to the responses of CD8+ T cells, thereby offering valuable and insightful information regarding potential targets for the development of vaccines against tuberculosis (Woodworth *et al*., 2011). Furthermore, the manipulation and alteration of immunodominant and subdominant epitopes has been

extensively explored and investigated in order to enhance and augment immune responses. Studies have presented evidence that flanking sequences that are portable and encompass immunodominant epitopes can regulate the processing of epitopes, ultimately leading to an increase in the production and antigenicity of subdominant epitopes (Gall *et al*., 2007).

The removal and deletion of an immunodominant epitope in the case of respiratory syncytial virus has been thoroughly examined and scrutinized, with the purpose of boosting the response of CD8+ T cells to a subdominant epitope (Mok *et al*., 2008). To conclude, it is absolutely imperative and crucial to have a comprehensive understanding of immunodominant epitopes and their interactions with the immune system, as this knowledge is indispensable for the advancement and progress of vaccine design and immunotherapies. Through the elucidation of the characteristics and roles that these epitopes possess, researchers are able to optimize and enhance immune responses directed against pathogens, and potentially develop and create diagnostic tools and vaccines that are more efficacious and effective.

The development of diagnostic tools, vaccine candidates, and neutralizing antibodies against SARS-CoV-2 relies heavily on the identification and validation of immunodominant epitopes. Various studies have explored different approaches to achieve this goal. For instance, demonstrated the safety and immunogenicity of RNA-based COVID-19 vaccine candidates, which elicited dose-dependent SARS-CoV-2–neutralizing titers (Walsh *et al*., 2020). Similarly, Gao *et al*. highlighted the development of an inactivated vaccine candidate for SARS-CoV-2, emphasizing the importance of various vaccine types under development, including DNA- and RNA-based formulations, recombinant subunits, adenovirus-based vectors, and purified inactivated virus (Z. Liu *et al*., 2020).

The study by Yu and his colleagues  focused on an RBD-Fc-based COVID-19 vaccine candidate that induced potent neutralizing antibody responses by targeting multiple dominant neutralizing epitopes (Yong *et al*., 2022). This approach underscores the significance of identifying key epitopes for vaccine design. Additionally, investigated DNA vaccine protection against SARS-CoV-2 in rhesus macaques, demonstrating the immunogenicity and protective efficacy of prototype DNA vaccines expressing SARS-CoV-2 spike proteins (Yong *et al*., 2022). This approach aligns with the concept of designing multi-epitope vaccines, as discussed in a study by , which demonstrated the preclinical efficacy of gamma-irradiated inactivated SARS-CoV-2 vaccine candidates in producing neutralizing antibodies (Karakus *et al*., 2021). In conclusion, the synthesis of these studies underscores the diverse strategies employed to identify immunodominant epitopes for the design of effective diagnostic tools, vaccine candidates, and neutralizing antibodies against SARS-CoV-2. By leveraging various vaccine platforms, epitope mapping techniques, and preclinical evaluations, researchers aim to develop robust solutions to combat the ongoing COVID-19.

Identification of immunodominant epitopes on the nucleocapsid (NP) and spike (S) proteins of SARS-CoV-2 has been investigated in several studies. In Iranian COVID-19 patients Maghsood *et al*. found that the IgG response to NP is dominated by linear epitopes, with a strong reactivity towards a single peptide (Maghsood *et al*., 2021). Musicò *et al*. identified immunodominant regions on the S and NP proteins, with an epitope in the N protein  showing 92% sensitivity and 100% specificity for IgG detection in COVID-19 samples (Musicò *et al*., 2020). Vandervaart *et al*. used structural biology and epitope mapping to define antigenic regions of the NP protein, identifying six public and four private epitope regions across NP, some of which are unique to their study (Vandervaart *et al*., 2023). Amrun *et al*. identified

four immunodominant epitopes on the S and N viral proteins, with N4P5 achieving the highest level of specificity (100%) and sensitivity (>96%) against SARS-CoV-2 (Amrun *et al*., 2020). Zhang *et al*. analyzed linear B cell immunodominant (ID) sites on the S protein and identified nine linear ID sites, including four in the RBD fragment (B. Zhang *et al*., 2020). Some immunodominant epitopes that have been isolated from these studies as listed in table 1 below (Maghsood et al., 2021).

**Table 1: Immunodominant epitopes previously identified** *(Maghsood et al., 2021).*

| SPIKE | NUCLEOCAPSID | |
|---|---|---|
| RVQPTESIVRFPNITNLCPF | SDNGPQNQRNAPRITFGGPS | LNQLESKMSGKGQQQQGQTV |
| NLCPFGEVFNATRFASVYAW | FGGPSDSTGSNQNGERSGAR | QGQTVTKKSAAEASKKPRQK |
| SVYAWNRKRISNCVADYSV | RSGARSKQRRPQGLPNNTAS | KPRQKRTATKAYNVTQAFGR |
| DYSVLYNSASFSTFKCYGVS | NNTASWFTALTQHGKEDLKF | QAFGRRGPEQTQGNFGDQEL |
| CYGVSPTKLNDLCFTNVYAD | EDLKFPRGQGVPINTNSSPD | GDQELIRQGTDYKHWPQIAQ |
| NVYADSFVIRGDEVRQIAPG | NSSPDDQIGYYRRATRRIRG | PQIAQFAPSASAFFGMSRIG |
| QIAPGQTGKIADYNYKLPDD | RRIRGGDGKMKDLSPRWYFY | MSRIGMEVTPSGTWLTYTAA |
| KLPDDFTGCVIAWNSNNLDS | RWYFYYLGTGPEAGLPYGAN | TYTAAIKLDDKDPNFKDQVI |
| NNLDSKVGGNYNYLYRLFRK | PYGANKDGIIWVATEGALNT | KDQVILLNKHIDAYKTFPPT |
| RLFRKSNLKPFERDISTEIY | GALNTPKDHIGTRNANNAA | TFPPTEPKKDKKKKADETQA |
| STEIYQAGSTPCNGVEGFNC | ANNAAIVLQLPQGTTLPKGF | DETQALPQRQKKQQTVTLLP |
| EGFNCYFPLQSYGFQPTNGV | LPKGFYAEGSRGGSQASSRS | VTLLPAADLDDFSKQLQQSM |
| PTNGVGYQPYRVVVLSFELL | ASSRSSSRSRNSSRNSTPGS | LDDFSKQLQQSMSSADSTQA |
| SFELLHAPATVCGPKKSTNL | GNGGDAALALLLLDRLNQLE | STPGSSRGTSPARMAGNGGD |
| ATVCGPKKSTNLVKNKCVNF | | |

### 2.3.12 Viral protein synthesis

SARS-CoV-2, the virus responsible for the COVID-19 pandemic, utilizes various proteins encoded in its genome to facilitate its replication, virulence, and evasion of the host immune response. The virus also manipulates host biosynthetic machinery to produce essential proteins for its propagation (O'Keefe *et al*., 2021). Studies have shown that SARS-CoV-2

can invade host cells through unique routes, such as the CD147-spike protein pathway (Ke Wang *et al*., 2020). Furthermore, the virus can antagonize the host's type I interferon response, contributing to its pathogenesis (Q. Zhang *et al*., 2021).

The virus's interaction with host cells can lead to various complications. For instance, SARS-CoV-2 infection has been linked to endothelial dysfunction, which may contribute to conditions like erectile dysfunction (Kresch *et al*., 2021). The virus can also infect kidney tissues, leading to tissue damage and contributing to conditions like acute tubular necrosis (Radovic *et al*., 2023). Furthermore, severe COVID-19 patients may develop anti-cardiac autoantibodies, highlighting the systemic impact of the virus on various organs (Fagyas *et al*., 2022).

The intricate process by which SARS-CoV-2 synthesizes its proteins within host cells plays a critical role in its pathogenesis and the development of COVID-19. The initial stage of protein synthesis by SARS-CoV-2 encompasses the process of translating its genomic RNA into two considerably large polyproteins, namely pp1a and pp1ab, through the utilization of the ribosomes found within the host cell.

Once synthesized, these polyproteins undergo cleavage by viral proteases, specifically the primary protease (Mpro) and the papain-like protease (PLpro), resulting in the formation of distinct non-structural proteins (nsps) (X. Shen *et al*., 2021). The main protease (Mpro) of SARS-CoV-2 plays an exceedingly vital role in the cleavage of the viral polyproteins into functional units, thus significantly contributing to viral replication (J. Chen et al., 2022). In addition to this, the papain-like protease (PLpro) actively participates in the cleavage of viral polyproteins and has been specifically identified as an ideal target for inhibition in order to

impede SARS-CoV-2 replication (Swaim *et al*., 2020). The genomic RNA of SARS-CoV-2 comprises open reading frames that are responsible for encoding both non-structural and structural proteins, with the former being expressed through subgenomic mRNAs (Narwal *et al*., 2022). The production of polyproteins resulting from the translation of viral RNA is absolutely essential for the process of replication and the establishment of pathogenesis in relation to SARS-CoV-2 (Parvez & Sakina, 2021). Furthermore, it is important to note that the polyprotein substrates of SARS-CoV-2 effectively regulate the stepwise cleavage reaction mediated by the main protease (Mpro) (Narwal *et al*., 2022).

The interaction between SARS-CoV-2 and host cell proteins is undeniably crucial for the successful replication of the virus, with the main protease (Mpro) being responsible for the cleavage of cellular targets, which ultimately contributes to the pathogenicity of the virus (Miczi *et al*., 2020). The genome of SARS-CoV-2 is composed of distinctive features, such as an insertion of four amino acids at the S1/S2 junction of the spike protein, thereby setting it apart from other coronaviruses (Jaimes *et al*., 2020). Furthermore, an extensive analysis of the SARS-CoV-2 genome has been conducted in order to identify mutations occurring in major target proteins, thus highlighting variations that may potentially impact the characteristics of the virus (M. I. Khan *et al*., 2020). In conclusion, the translation of SARS-CoV-2 genomic RNA into polyproteins, followed by the subsequent cleavage of these polyproteins into individual non-structural proteins by viral proteases, represents an absolutely critical stage within the life cycle of the virus.

SARS-CoV-2 encodes several nonstructural proteins (nsps) crucial for its replication. Among these nsps, RNA-dependent RNA polymerase (RdRp or nsp12) plays a central role in replicating the viral genome, while helicase (nsp13) unwinds viral RNA duplexes to aid in

replication (L. Yan *et al*., 2020). Additionally, nsps such as nsp3, nsp4, and nsp6 contribute to the formation of the replication-transcription complex (RTC), a specialized structure where viral RNA synthesis occurs (Angelini *et al*., 2013). Studies have shown that nsp6 is involved in the formation of double-membrane vesicles (DMVs) during SARS-CoV-2 infection (D. Y. Chen *et al*., 2023). The architecture of the SARS-CoV-2 mini RTC has been elucidated, revealing the assembly of nsp12, nsp7, nsp8, and two nsp13 helicase molecules 175]. Nsp13 has been found to possess NTPase and RNA helicase activities critical for viral replication, making it a potential target for antiviral therapies (Shu *et al*., 2020). Furthermore, the interaction between nsp12 and nsp13 enhances the unwinding of nucleic acids, aiding in viral replication (Adedeji *et al*., 2012).

Mutational hotspots have been identified in the SARS-CoV-2 genome, impacting the virus's evolution and potential clinical outcomes (Zekri *et al*., 2021). Studies have also highlighted the importance of nsps like nsp3, nsp4, and nsp6 in the formation of DMVs, essential for viral replication (D. Y. Chen *et al*., 2023). Moreover, the conservation of nsps across different SARS-CoV-2 strains underscores their significance in viral biology and potential as therapeutic targets (Nawaz *et al*., 2023). In conclusion, the nsps of SARS-CoV-2, including RdRp, helicase, and other key proteins like nsp3, nsp4, and nsp6, play vital roles in viral replication and the formation of replication complexes. Understanding the functions and interactions of these nsps is crucial for developing targeted antiviral strategies and combating COVID-19.

The structural proteins essential for the assembly of new SARS-CoV-2 viral particles include the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. These proteins are synthesized on the endoplasmic reticulum (ER) membrane, where they are inserted and

then transported to the site of viral assembly (Jain *et al*., 2020). The nucleocapsid (N) protein of SARS-CoV-2 interacts with human ribonucleoproteins, modulating protein/RNA granule formation to enhance viral replication (Savastano *et al*., 2020). Additionally, the spike (S) protein of SARS-CoV-2 is crucial for viral entry into host cells, with the spike protein having a higher binding affinity to ACE2 compared to SARS-CoV-1, potentially leading to a more robust cytokine response (Ouyang *et al*., 2021). Furthermore, the SARS-CoV-2 spike protein has fusogenic activity, promoting membrane fusion and viral entry (Ouyang *et al*., 2021). The spike protein also elicits cell signaling in human cells, which has implications for the consequences of COVID-19 vaccines (Suzuki & Gychka, 2021). Moreover, the SARS-CoV-2 spike protein has been linked to fatal diseases like pulmonary arterial hypertension (PAH) by promoting cell growth signaling in human lung vascular cells (Suresh & Suzuki, 2020). In terms of vaccine development, combining spike- and nucleocapsid-based vaccines has been shown to improve the control of SARS-CoV-2 infection (Dangi *et al*., 2021). Additionally, nucleocapsid-specific humoral responses have been found to enhance the control of SARS-CoV-2, indicating the importance of targeting multiple viral antigens for effective immune responses (Dangi, Sanchez, Park, *et al*., 2022). In summary, the structural proteins of SARS-CoV-2 play critical roles in viral assembly, entry into host cells, modulation of immune responses, and vaccine development. Understanding the functions and interactions of these proteins is essential for developing effective therapeutic strategies against COVID-19.

Post-translational modifications are essential for the maturation and functionality of viral proteins. Modifications such as glycosylation, phosphorylation, and proteolytic cleavage play a crucial role in ensuring proper protein function and stability (Reeves, 2020). Glycosylation,

for example, is important for the correct folding and quality control of newly synthesized proteins at the ER-Golgi intermediate compartment (ERGIC) (Appenzeller-Herzog & Hauri, 2006). Post-translational modifications like glycosylation and phosphorylation are common in viral proteins and impact various functions of these proteins (T. Zhou *et al*., 2022). These modifications are vital for the stability of viral proteins and also influence their interactions with host factors and cellular membranes (Z. Zhu *et al*., 2019). Following post-translational modifications, viral proteins gather at specific cellular membranes, such as the ERGIC, to form new viral particles (Pearson *et al*., 2021). The ERGIC is crucial for concentrating, folding, and quality control of these modified proteins, facilitating viral particle assembly (Appenzeller-Herzog & Hauri, 2006). Additionally, the Golgi region aids in transporting viral nucleocapsids for particle formation, where viral glycoproteins accumulate (Stertz *et al*., 2007). Accumulation of viral envelope proteins at the ERGIC is necessary for virus assembly (Lontok *et al*., 2004; McBride *et al*., 2007).

The ERGIC plays a significant role in the biogenesis of autophagosomes, with membrane remodeling at ER-exit sites initiating a pathway for autophagosome formation (Ge *et al*., 2017). This process involves contributions from the endomembrane system and various membrane and lipid interactions (S. Li *et al*., 2021). Autophagosomes, double-membrane vesicles generated during autophagy, are crucial for cellular homeostasis and the degradation of cellular components (Ge *et al*., 2017). In conclusion, post-translational modifications of viral proteins, their assembly at specific cellular membranes like the ERGIC, and their involvement in processes such as autophagosome biogenesis underscore the intricate interplay between viral proteins and cellular machinery during viral replication and pathogenesis.

### 2.3.13 Viral protein prediction

The ORF (Open Reading Frame) Finder is an invaluable bioinformatics tool that is employed to identify possible protein-coding regions within DNA sequences. This particular tool plays an extremely important and pivotal role in the exploration and investigation of genes, such as those that are associated with the SARS-CoV-2 virus, by precisely locating open reading frames that have the potential to encode proteins. The true significance and importance of the ORF Finder lies in its remarkable and highly accurate ability to predict genes and translation initiation sites within DNA sequences (Hyatt *et al*., 2010). By conducting a thorough and comprehensive analysis of the %G+C content of the predicted open reading frames, as well as evaluating dinucleotide bias and determining the presence of mobility genes, the ORF Finder greatly aids in the identification of pathogenicity islands that are present in bacterial genomes (Che *et al*., 2014). Furthermore, this tool has also been effectively utilized in the evaluation and assessment of the ORFs of Acinetobacter baumannii genomes, thereby clearly demonstrating its immensely practical and applicable nature when it comes to the study of complex carbohydrate biosynthesis loci (Kenyon & Hall, 2013).

The utility of the ORF Finder extends beyond solely focusing on bacterial genomics since it has also been extensively employed in the annotation of protein-coding genes in a wide variety of different organisms, including the green-tide forming alga Ulva compressa (Meredith *et al*., 2012). Moreover, the incredible versatility of this tool can be further exemplified by its successful application in the prediction of intergenic small ORFs within the Caenorhabditis elegans genome, which effectively showcases its remarkable adaptability and flexibility when it comes to dealing with a wide range of diverse genetic contexts (Casimiro-Soriguer *et al*., 2020). The highly effective and efficient nature of this tool, in

terms of accurately identifying ORFs, has been widely and extensively acknowledged and recognized in a multitude of different studies, thereby emphasizing and underscoring its pivotal role in gene prediction and analysis (S. Wu *et al*., 2011). Overall, the ORF Finder undoubtedly emerges as an absolutely fundamental and essential tool within the field of bioinformatics, as it successfully empowers and enables researchers to accurately and precisely identify potential protein-coding regions. The remarkable and highly commendable ability of this tool to predict genes, evaluate pathogenicity islands, and annotate genomes effectively highlights and underscores its immense significance and relevance within the realm of genomic research, while also highlighting its immensely practical applications within a wide range of different biological contexts.

To understand the proteins and subsequently the epitopes made by a virion inside a cell, it is important to employ bioinformatic tools that include ORF finder to predict the viral genes, NCBI blast to annotate proteins and IEDB to find saved epitopes and check for homology with the experimental genome sequence.

When employing the ORF Finder tool to make predictions about genes associated with the SARS-CoV-2 virus, it is of utmost importance to possess a thorough understanding of the significance of open reading frames (ORFs) in the realm of genetics. ORFs essentially serve as DNA sequences that hold the potential to encode proteins, which are typically characterized by the presence of a start codon (often AUG), a series of codons, and ultimately a stop codon (UAA, UAG, or UGA) (Sardi *et al*., 2020).

The identification of ORFs within a given viral genome is absolutely crucial, as it can provide valuable insights into the fundamental genes responsible for various viral functions,

including but not limited to replication, transcription, and interactions with host organisms. In the realm of current prediction algorithms, the utilization of conservation signatures plays a pivotal role in effectively distinguishing functional ORFs (Brunet *et al*., 2018). For example, in the study conducted by [insert author name], the prediction of ORFs was carried out through the utilization of the NCBI ORF Finder tool, wherein a minimum ORF length of 100 nucleotides was defined (Sardi *et al*., 2020). This particular example underscores the immense importance of employing such tools to accurately identify regions within the viral genome that hold the potential to code for proteins. Furthermore, the study conducted by Taguchi shed light on the significant impact that specific ORFs can have on viral functions.

In their research involving *Pseudomonas syringae* pv. tabaci, it was observed that the deletion of orf3 did not have any effect on flagellin glycosylation, but it did significantly reduce the virulence of the mutant on tobacco plants (Fumiko Taguchi *et al*., 2010). This particular finding serves as a stark reminder of how a thorough understanding of individual ORFs can provide valuable insights into the functionality of viral genes. Therefore, it is absolutely imperative to possess a comprehensive understanding of ORFs and their relevance within the field of genetics when utilizing tools such as the ORF Finder for the purpose of predicting genes associated with the SARS-CoV-2 virus. By effectively recognizing and analyzing ORFs within viral genomes, researchers have the potential to uncover essential genes that play a critical role in various viral activities, thereby aiding in the understanding of the virus's behavior and identifying potential therapeutic targets.

The process of annotation and analysis of predicted open reading frames (ORFs) in the genome of SARS-CoV-2 encompasses a multi-step computational and bioinformatics endeavor, which entails various methodologies and tools. In the initial phase, the utilization

of the ORF Finder software allows for the prediction of potential ORFs by thoroughly scanning the entire genome and applying specific criteria, such as the identification of start and stop codons, thereby aiding in the identification of potential protein-coding regions (Finkel *et al*., 2020) (Finkel *et al*., 2020). To identify the genes of SARS-CoV-2 by means of the ORF Finder, researchers typically acquire the comprehensive sequence of the virus's genome through the utilization of various sequencing techniques, such as next-generation sequencing (NGS) or the polymerase chain reaction (PCR), which is then followed by sequencing processes (J. Zhan *et al*., 2022).

Once the genome sequence has been obtained, it can subsequently be inputted into the software known as ORF Finder, which is accessible online or can be utilized as standalone applications, and serves the purpose of identifying the open reading frames (ORFs) within the viral genome (D.-W. Kim *et al*., 2020). The genome of SARS-CoV-2, being a single-stranded RNA virus, possesses a remarkably large non-segmented genome that measures approximately 30 kilobases in length (Ye *et al*., 2022). The annotation of the SARS-CoV-2 genome has revealed that there are 10 ORFs present within it, with the ORF10 being situated downstream of the N gene (Pancer *et al*., 2020). Furthermore, it has been ascertained that the genetic makeup of the SARS-CoV-2 genome is significantly more stable when compared to other coronaviruses, such as SARS-CoV and MERS-CoV (Y. C. F. Su *et al*., 2020).

Research studies have demonstrated that the genome of SARS-CoV-2 consists of potential G4-forming sequences, which are known as PQSs, and have been subject to extensive investigation (Zhao *et al*., 2020). Moreover, the genome of SARS-CoV-2 has been meticulously examined and characterized in relation to its accessory proteins, which serve a significant role in the pathogenicity of the virus itself (Silvas *et al*., 2021). Areas that pertain

to comprehending the pathogenesis of SARS-CoV-2 and the influence that viral proteins have on the outcomes of the disease are still shrouded in mystery and necessitate further investigation (Silvas *et al*., 2021). Consequently, it is crucial to attain an understanding of the genetic composition of SARS-CoV-2 through the process of genome sequencing and the thorough analysis of its open reading frames (ORFs) and accessory proteins, as this knowledge is of paramount importance when it comes to unraveling the pathogenicity of the virus and formulating potential therapeutic targets.

The identification and annotation of open reading frames in the SARS-CoV-2 genome have emerged as crucial steps in the process of decoding the abundant genetic information carried by this particular virus. Numerous studies have made significant contributions to revealing the coding capacity and functional aspects of these ORFs, thereby enlightening our understanding of the intricacies of this viral genome. In their comprehensive study, Finkel and his colleagues underscored the paramount importance of unearthing the complete coding potential harbored within the SARS-CoV-2 genome, an endeavor that undoubtedly lays the foundation for subsequent and essential functional investigations (Finkel *et al*., 2020). Similarly, Kim and colleagues aptly stressed the weighty significance of extensively exploring the transcriptomic architecture of SARS-CoV-2, as this endeavor is vital for us to acquire a more profound comprehension of the virus (D.-W. Kim *et al*., 2020). Additionally, researchers have delved deep into the functional roles played by specific ORFs, unearthing valuable insights into the multifaceted nature of this virus. For instance, Kopecky-Bromberg *et al*. meticulously demonstrated that certain ORFs within the SARS-CoV genome, such as ORF 3b, ORF 6, and nucleocapsid proteins, effectively function as interferon antagonists, hence demonstrating the virus's remarkable ability to skillfully evade immune responses

mounted by the host (Kopecky-Bromberg *et al*., 2007). Likewise, Silvas *et al*. conducted an in-depth investigation into the contribution made by SARS-CoV-2 accessory proteins to viral pathogenicity, an inquiry that crucially shed light on the intricate and complex viral-host interactions that underpin the pathogenicity of this particular virus (Silvas *et al*., 2021).

Studies have also successfully unraveled the profound impact that individual ORFs can exert upon host cells, thereby providing us with invaluable insights into the intricate and multifaceted nature of this viral genome. For example, Shi *et al*. astutely revealed the manner in which SARS-Coronavirus ORF-9b can successfully repress innate immunity, accomplishing this by specifically targeting mitochondria, thereby effectively illustrating the disruptive capabilities that small ORFs possess (C. S. Shi *et al*., 2014). In a similar vein, Count meticulously explored group-specific ORFs within the SARS-CoV genome, conclusively indicating the fact that such ORFs do not play essential roles in viral replication within cell cultures and mice (Yount *et al*., 2005). Moreover, the comprehensive characterization of accessory genes and non-canonical subgenomic RNAs within the SARS-CoV-2 genome has remarkably enriched our current comprehension of the viral genome's coding potential (Michel *et al*., 2020; Nomburg *et al*., 2020).

The identification of overlapping ORFs and the subsequent exploration of their implications have emerged as topics of immense interest, effectively underscoring the critical necessity for precise genome annotations (Jungreis *et al*., 2021; Pavesi, 2021). In conclusion, it is undoubtedly evident that the annotation and thorough analysis of ORFs within the SARS-CoV-2 genome represent vital pursuits that are instrumental in unraveling the virus's extensive genetic information, gaining a comprehensive understanding of the multifaceted functional roles played by specific ORFs, and accurately assessing the profound impact that

these ORFs can exert upon host cells. Consequently, it is clear that the cumulative findings of these detailed studies represent a significant advancement in our current knowledge of SARS-CoV-2, and moreover, these findings effectively facilitate the development of targeted interventions and therapeutics that can successfully combat this viral entity.

Following the preliminary step by ORF finder, sequence alignment tools such as the Basic Local Alignment Search Tool (BLAST) are employed to compare the predicted ORFs with previously known protein sequences, thus facilitating the identification of similarities and homologies among them (Grifoni *et al*., 2020b). This comparative analysis plays a crucial role in providing insights into the functional aspects of the identified ORFs.

In order to infer the functions of the predicted ORFs, the process of functional annotation is performed, which involves leveraging information from homologous proteins in other organisms (Michel *et al*., 2020). By drawing upon the wealth of knowledge accumulated regarding the functions of proteins in other organisms, this annotation process contributes to the understanding of the potential roles and characteristics of the identified ORFs. Furthermore, orthology analysis is conducted to discern orthologous genes across different viral strains, thereby shedding light on the evolutionary relationships and gene conservation patterns within the SARS-CoV-2 genome (Almansour & Boudellioua, 2022). This analysis allows for the identification of genes that have been conserved throughout the evolutionary history of the virus, providing valuable insights into the genetic makeup and evolution of SARS-CoV-2.

In cases where the predicted ORFs exhibit a lack of similarity to known proteins, alternative approaches such as protein structure modeling and prediction methods are employed to

predict their three-dimensional structures, thereby offering valuable insights into their potential functions and interactions (Michel *et al*., 2020). These computational methods enable researchers to explore the structural features of these novel ORFs, providing a foundation for further investigations into their functional roles. However, it is important to note that experimental validation is of paramount significance in confirming the accuracy and reliability of the computational predictions. Techniques such as gene and protein expression analysis are employed in this validation process, allowing for the assessment of the actual expression and functionality of the predicted ORFs (Finkel *et al*., 2020). By integrating computational predictions with experimental validation, this comprehensive approach ensures a thorough understanding of the protein-coding regions within the SARS-CoV-2 genome.

The annotated ORFs, along with their predicted functions, are then stored in databases like GenBank, serving as invaluable resources for ongoing research efforts (Grifoni *et al*., 2020b). These databases provide a centralized repository of information, enabling researchers from diverse fields to access and utilize the knowledge derived from the annotation and analysis of the ORFs in the SARS-CoV-2 genome, fostering collaboration and facilitating further investigations into the virus and its implications. In conclusion, the process of ORF annotation and analysis in the SARS-CoV-2 genome encompasses a comprehensive and interdisciplinary approach that combines computational predictions, comparative analyses, experimental validation, and database storage. This integrative methodology ensures a robust understanding of the protein-coding regions within the viral genome, contributing to our knowledge of the virus and informing ongoing research efforts.

## 2.4 Other coronaviruses

Within the Coronaviridae family, four major genera can be distinguished: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus. These genera represent different branches on the evolutionary tree of coronaviruses with Alphacoronavirus and Betacoronavirus predominantly infecting mammals including humans and Gammacoronavirus and Deltacoronavirus being mainly bird pathogens (Curran *et al*., 2020). Classification of coronaviruses into these genera is not random but based on a careful analysis of their genomic structure, antigenic properties as well as host range. This scientific approach has made it possible for researchers to better understand the diversity and characteristics of coronaviruses leading to development of efficient diagnostic tests and therapeutic interventions (Galina *et al*., 2020). The two subfamilies within the Coronaviridae family are Torovirinae and Coronavirinae which include these groups respectively. The latter subfamily, Coronavirinae is highly compelling because it contains human diseases caused by coronaviruses. Among the Coronavirinae subfamily mentioned earlier which includes those four genera that have vital role in epidemiology and pathogenicity of coronavirus infections. The categorizations into these genera are not mere classification but rather scientific deductions supported by key genomic features such as genetic make-up, antigen presentation as well as species specificity (Hemmati-Dinarvand *et al*., 2020).

HCoV's genome sequence reveals that it falls under four main groups that correspond to four different genera: Alphacoronavirus (α), Betacoronavirus (β), Gammacoronavirus (γ), and Deltacoronavirus (δ). It should be pointed out that this division does not occur randomly; instead, it reflects how far back a strain goes in time and its genetic diversity. Based on categorizing HCoVs into these groups, similarities and variances in their genetic composition

can be identified and this would help to explain their biological features and pathogenicity. This information is important for the development of effective diagnostic methods and therapeutic options against HCoV infections (A. Singh *et al*., 2020).

An expansive research has been done with a view of exploring the genomic traits of coronaviruses in depth. In particular, bat coronaviruses have been identified as the major source for Alphacoronavirus and Betacoronavirus gene whereas avian coronaviruses are considered to be responsible for Gammacoronavirus and Deltacoronavirus genes (Kai Wang *et al*., 2015). Importantly, this viral group is widely distributed among many species including animals, birds, and humans thereby implying its diversity (Marchenko *et al*., 2022). Moreover, Gammacoronaviruses and Deltacoronaviruses have a well-documented genetic diversity and they are especially common in both wild and domesticated birds. Additionally, these viruses have shown their ability to infect mammalian reservoirs indicating that they might cross over into them easily as a consequence of their adaptability which enables them to exploit a wide range of hosts.

The coronaviruses that have the ability to infect humans Alphacoronaviruses and Betacoronaviruses. The Alphacoronaviruses group encompasses HCoV-229E and HCoV-NL63, while the Betacoronaviruses group consists of HCoV-OC43, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2 (B. Huang *et al*., 2020). These two genera are of particular significance due to the fact that all known human coronaviruses fall under their classification, thereby emphasizing their potential to give rise to future pandemics (Ghai *et al*., 2021). In the case of Alphacoronaviruses such as HCoV-229E, feline coronavirus, and canine coronavirus, their recognition and attachment to aminopeptidase N (APN) as their receptor play a crucial role.

Conversely, Betacoronaviruses such as SARS-CoV, MERS-CoV, and SARS-CoV-2 are associated with severe human morbidity and high case fatality rates (F. Li, 2016; Llanes *et al*., 2020). The alternating biennial winter incidence peaks observed in both alphacoronaviruses and betacoronaviruses suggest immune-mediated interactions that impact the epidemiology of these viruses (Dyrdak *et al*., 2021). Understanding the biology, transmission dynamics, and potential threats posed by alphacoronaviruses is crucial in light of past outbreaks and the ongoing COVID-19 pandemic (Algaissi *et al*., 2020). The genetic characterization of alphacoronaviruses and their prevalence in various animal species emphasize the need for continued research to mitigate the risks associated with these viruses (Orłowska *et al*., 2022).

There are various reasons why viruses show specificity in terms of the hosts they infect. To start with, viruses depend on specific receptor molecules on the surface of host cells to gain entry, and variations in these receptors across species can be traced back to inherent differences in cellular structure and physiology (Bhella, 2015). Even when a virus can bind itself on the surface of a cell, it must hijack the complex machinery of the host cell completely for replication. This process however varies among different host species hence hindering its efficient replication (Mitnaul *et al*., 2000).

Host species have immune systems that are equipped with defenses that are specifically meant for their respective species making it easier for viruses to either overcome or dodge such defenses in one species while being ineffective against the others (J. Chow *et al*., 2015). Over time viruses co-evolve with their hosts leading to evolutionary changes that enhance their ability to infect some hosts but reduce infectivity to non-hosts (Dennehy, 2016). Additionally, viral transmission is influenced by ecological factors such as close contact and

interactions between individual hosts which influence both frequency and likelihood of transmission (D. Zhang *et al*., 2022). Also, there is an intrinsic link between viral evolution rates and the geography of its host range thus showing viral evolution is associated with its own host species (Streicker *et al*., 2012). Conversely, phylogenetic relationships within families highlight vast diversity among viruses as well as potential risks from spill-over into human population(s) (Tirera *et al*., 2021). Moreover, socio-ecological environments interacting with climate can greatly affect outbreaks brought about by viruses like temperature among other socio-ecological issues influencing dynamics related to viral transmission (Chenlu Li *et al*., 2021).

To fully comprehend how viral communities work intricately we must develop a broad knowledge about variation existing within several populations of different hosts, as this is likely to affect viral communities within host organisms (Bergner *et al*., 2019). Summing up, the specificity of viruses in relation to certain host species results from a complex interplay between receptor interactions, host immunity factors, co-evolution and viral diversity. The ability of a virus to infect, replicate and transmit effectively in specific host populations is collectively determined by these factors hence demonstrating that viruses have intricate relationships with their respective hosts.

### 2.4.1 Alphacoronaviruses

Important focus has been on alphacoronaviruses which are a subgroup of coronaviruses and are highly significant in both animals and humans. These coronaviruses are commonly known to infect a variety of animals, including bats, pigs and humans. HCoV-229E is one of the leading examples within this group as well as HCoV-NL63 which often leads to mild

respiratory illnesses in humans (Mei *et al*., 2022). Additionally, bats have been implicated as important reservoirs for diverse coronaviruses encompassing alphacoronaviruses (Caraballo *et al*., 2022). In America, there are seven different lineages of alphacoronaviruses that occur in bats and these alphaCoVs have gone through several recent cases of host switching (Korneenko *et al*., 2022). Bat populations act as hosts for various types of coronaviruses with SARS-CoV and SARS-CoV-2 being no exceptions (Saxena *et al*., 2023). Moreover, an emerging alphacoronavirus identified in swine called swine acute diarrhea syndrome coronavirus (SADS-CoV), causing diarrhea in newborn piglets has been found in southern China (Scheffer *et al*., 2023).

Scientific inquiries have successfully detected previously unidentified alphacoronaviruses in meso-carnivores and rodents, thereby indicating the potential for interspecies transmission and, consequently, the imperative necessity for comprehensive surveillance encompassing diverse animal populations (Olarte-Castillo, 2023). The extensive array of coronaviruses that exist within wildlife, particularly those harbored by bats and rodents, unequivocally underscores the utmost importance of conducting thorough investigations into these specific populations as potential repositories for emerging human pathogens (Arteaga *et al*., 2023). Thus, it is evident from these findings that bats remain key hosts and reservoirs of alphacoronaviruses capable of crossing species barriers into other animals including humans.

Alphacoronaviruses are known for their distinctive composition, encompassing spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, which contribute to their structural integrity. Of particular significance in the entry of the virus into the host cells is the spike protein, as it binds to specific receptors and initiates the infection process (F. Li,

2016). The spike protein exhibits two distinct conformations, namely prefusion and postfusion, each playing a crucial role in the viral life cycle (Chunhua Li *et al*., 2017).

Several investigations have shed light on the structural characteristics of the spike protein in alphacoronaviruses, such as the Porcine Epidemic Diarrhea Virus, revealing unique attributes that may aid in immune evasion, including concealed receptor-binding sites and shielded critical epitopes (Shang *et al*., 2018; Wrapp & McLellan, 2019). Furthermore, coronaviruses' spike protein induces conformational alterations in the three-dimensional structure of the ACE2 protein, a pivotal factor for viral entry into host cells (Basu *et al*., 2020). Moreover, the genome of alphacoronaviruses encodes non-structural and accessory proteins that play vital roles in viral replication, immune evasion, and pathogenesis (Forni *et al*., 2022).

A comprehensive comprehension of the molecular interactions between viral proteins, such as the spike protein, and host cell receptors is indispensable for the development of effective antiviral strategies. Consequently, researchers have devoted their efforts to investigating the interplay between the spike protein of coronaviruses and host cell receptors like ACE2, providing valuable insights into potential targets for therapeutic interventions (M. Hussain *et al*., 2020; Ibrahim *et al*., 2020). Moreover, advancements in structural biology, exemplified by cryo-electron microscopy, have revolutionized our ability to visualize the spike protein in its prefusion conformation, thereby unraveling novel insights into its intricate structure and multifaceted functionality (Shang *et al*., 2018; Wrapp & McLellan, 2019). These structural studies have not only expanded our understanding of the spike protein but have also paved the way for the identification of neutralizing epitopes, which hold promise as targets for the development of innovative vaccines and therapeutics (Chunhua Li *et al*., 2017; Thavorasak *et al*., 2022).

#### 2.4.1.1 HCOV-229E

Human Coronavirus 229E (HCoV-229E), a viral microorganism classified under the Coronaviridae family, is renowned for its propensity to induce respiratory ailments in both human beings and animals. First identified during the 1960s, HCoV-229E stands as one of the four commonly occurring coronaviruses that afflict the human population, typically giving rise to mild respiratory symptoms (Van Der Hoek *et al*., 2004). Belonging to a genus of alpha coronaviruses, HCoV-229E manifests weak neutralizing activity while exhibiting substantial binding reactivity when juxtaposed with other coronaviruses (Kubota-Koketsu *et al*., 2020).

HCoV-229E has attracted attention due to its association with upper respiratory diseases. Different investigations have shown that this virus in particular can lead to severe respiratory infections including among children, elderly people and people with chronic medical conditions (Paules *et al*., 2020). However, it must be noted that while HCoV-229E is typically associated with upper respiratory infection like the common cold, there have been cases where it has caused severe lower respiratory tract illness in old people and those whose immunity is compromised (D. Liu *et al*., 2023). Therefore, although usually HCoV-229E results into mild/moderate symptoms of upper airway infections; still it might result into more severe illnesses for individuals who are at risk (Fausto, 2023). For instance, in immunocompromised patients, presence of HCoV represents an important factor causing upper respiratory tract infections (Lepiller *et al*., 2013). In addition, several surveys indicated that such infection including blocked ciliary movement leading to replication within outlying parts of nasal passage could be responsible for upper airway pathologies such as colds (Prill

*et al*., 2012). HCoV-229E has been found to infect polarized airway epithelia from the apical surface, indicating its ability to target respiratory cells efficiently (G. Wang *et al*., 2000).

Despite mainly being linked to URTI's involving nasopharyngeal area; rarely do normal adults experience acute proximal pneumonia during bouts of infection by this virus (Paul *et al*., 2022). This finding underlines how different levels of sickness can arise from infection via HCoV-229E which happens based on immune status or other underlying health conditions of the victims. It is therefore clear that HCoV-229E is a significant viral pathogen causing upper respiratory illnesses and can cause severe respiratory tract infections in specific populations of vulnerable individuals. Hence, it is important to understand the nature of diseases caused by HCoV-229E fully so as to design specific strategies for preventing and treating them particularly among those who are most at risk.

HCoV-229E is a positive-sense, single-stranded RNA virus with a genome size ranging from 27 to 32 kilobases. This virus encodes structural proteins like spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as non-structural proteins crucial for its replication and transcription. The spike protein of HCoV-229E plays a key role in viral attachment and entry into host cells by interacting with cellular receptors, particularly angiotensin-converting enzyme 2 (ACE2) (Breslin *et al*., 2003). Studies have shown that HCoV-229E preferentially uses the cell surface protease TMPRSS2 for entry into host cells rather than endosomal cathepsin L, allowing it to bypass the endosome and potentially evade the host's innate immune response triggered by Toll-like receptors (Shirato *et al*., 2017).

The evolutionary history of HCoV-229E suggests an ancestral association with bats, similar to the Middle East respiratory syndrome coronavirus (MERS-CoV), indicating a zoonotic

origin (Corman *et al*., 2015). The virus has been extensively studied for its fusion core, which could be a potential target for drug development to inhibit membrane fusion, a critical step in viral infection (W. Zhang *et al*., 2018). In conclusion, HCoV-229E's genomic structure, protein composition, cellular entry mechanisms, evolutionary origins, and pathogenicity have been well-documented in various studies, shedding light on its characteristics and behavior within the host.

The virus has undergone meticulous scrutiny, with in-depth analyses of its genetic drift over time revealing limited variation within specific gene sequences (Chibo & Birch, 2006). Moreover, HCoV-229E has been utilized as a surrogate agent for investigating the inactivation of alternative coronaviruses, such as SARS-CoV-2, on surfaces (Vaze *et al*., 2022). Furthermore, since the 1960s, HCoV-229E, in conjunction with HCoV-OC43, has been recognized as one of the two human pathogenic members within the coronavirus family (Geller *et al*., 2009). In terms of its mode of operation, the virus capitalizes on aminopeptidase N (APN) glycoproteins as receptors, a phenomenon akin to other coronaviruses including transmissible gastroenteritis virus and feline coronavirus (Jeffers *et al*., 2004). Additionally, scientific investigations have explored the potential antiviral effects exhibited by various compounds, such as chloroquine, on HCoV-229E infections, thereby shedding light on potential treatment avenues (Kono *et al*., 2008).

Studies conducted on the human coronavirus HCoV-229E have provided evidence supporting its prevalence in both temperate and tropical regions. In the context of temperate climates, such as North America, Europe, and certain parts of Asia, it has been observed that infections tend to reach their peak during the colder months, aligning with the typical seasonal pattern exhibited by respiratory viruses. Conversely, in tropical regions

characterized by relatively consistent climatic conditions throughout the year, instances of HCoV-229E infections may occur without a distinct seasonal peak, thereby manifesting as a year-round occurrence (Gaunt *et al*., 2010; Lau *et al*., 2006; S. Zhang *et al*., 2018). Notably, investigations have revealed that within tropical regions such as Hong Kong, HCoV-229E infections may emerge during the early summer months, reaching their peak during the autumn period, while remaining absent during winter (Lau *et al*., 2006). This stands in stark contrast to the observations made within temperate regions like Shanghai, where HCoV infections exhibit a higher prevalence during the winter and spring seasons (S. Zhang *et al*., 2018). The epidemiological characteristics attributed to HCoV-229E have been associated with significant genomic alterations, including deletions within the viral spike glycoprotein gene, thus indicating a potential ancestral association with bats (Corman *et al*., 2015). Furthermore, seroepidemiological studies have demonstrated a substantial prevalence of antibodies specific to HCoV-229E within nasal secretions, thereby suggesting frequent reinfections despite the presence of preexisting serum antibodies (Gorse *et al*., 2010). In summary, the prevalence and seasonal patterns exhibited by HCoV-229E infections showcase notable variations between temperate and tropical regions, with discernible peaks evident within temperate climates during colder months, while tropical areas exhibit more consistent year-round occurrences. Consequently, these findings emphasize the utmost importance of comprehending the regional disparities in HCoV-229E epidemiology to develop and implement effective prevention and control strategies.

HCoV-229E, a human coronavirus, is prevalent throughout Africa, particularly causing respiratory infections during cooler and drier months. Although comprehensive surveillance data may be lacking in many African countries, various studies and reports indicate that

HCoV-229E significantly contributes to the burden of respiratory illnesses, especially affecting children and vulnerable populations (Faye *et al*., 2022; Tso *et al*., 2021). The presence of cross-reactive antibodies against SARS-CoV-2 in sub-Saharan Africa suggests exposure to other coronaviruses, potentially including HCoV-229E (Tso *et al*., 2021). Studies have shown that African bats harbor viruses closely related to HCoV-229E, indicating a possible origin of this coronavirus (Tao *et al*., 2017). Additionally, the recombination history of bat coronaviruses in Kenya further supports the presence of coronaviruses related to HCoV-229E in Africa (Tao *et al*., 2017). Furthermore, the sero-epidemiological data in rural communities in Ghana also suggests the circulation of human coronaviruses, including HCoV-229E, in Africa (Owusu *et al*., 2021). Respiratory infections, including those caused by HCoV-229E, are a significant public health concern in Africa, with studies highlighting the burden of viral pneumonia and other respiratory illnesses in the region (Mchomvu *et al*., 2019).

The role of co-infections, such as RSV co-infection with rhinovirus, in modulating disease severity in southern Africa remains an area requiring further research (Ihling *et al*., 2021; Loevinsohn *et al*., 2021). Moreover, the potential impact of influenza vaccine roll-out on antibiotic use in Africa underscores the importance of addressing respiratory infections comprehensively in the region (Knight *et al*., 2018). In conclusion, HCoV-229E is present in Africa, contributing to respiratory infections, especially in children and vulnerable populations. Further research and surveillance are crucial to better understand the epidemiology of HCoV-229E and other respiratory viruses in Africa to implement effective public health measures.

Given that Kenya has a significant international link and diverse population, it is arguable that HCoV-229E may be present in the country, as these respiratory viruses are known to globally circulate like coronaviruses (Y. Li *et al*., 2022). Studies conducted in Uganda and Ghana indicate an occurrence of HCoV-229E in East Africa thereby largely implying that Kenya, a densely populated country within the region, also harbors this particular coronavirus (Mulabbi *et al*., 2021; Owusu *et al*., 2021). In-depth molecular characterization of human coronaviruses in Kenya between 2009 and 2012 during which acute respiratory infections among children were initiated by these pathogens (Sipulwa *et al*., 2016). Also, several surveillance studies on pneumonia associated endemic human coronaviruses have been done Kilifi, Kenya confirming the need for monitoring of these viruses in the area (Otieno et al., 2020). Besides Africa possesses rich fauna and biodiversity hence geographic dispersal and evolutionary dynamics by African Bats beta coronaviruses make it a potential hotbed for viral diseases outbreak (Motayo *et al*., 2020). This ecological factor greatly supports possible existence of different types of CoVs including HCoV-229E inside the Kenyan boundaries.

Therefore, with regard to its unique traits alongside confirmed cases of human coronavirus propagation to neighboring countries as well as other parts within Africa at large makes it quite reasonable to hypothesize that HCoV-229E is indeed prevalent in Kenya. Hence, there is a need for extensive surveillance, research undertakings and monitoring programs to attain deep understanding of the pervasiveness and consequences of HCoV-229E in the country.

Research HCoV-229E, in Africa has historically faced limitations due to inadequate infrastructure and funding (J. Singh *et al*., 2021). The insufficiency of resources and financial support has posed significant challenges to the advancement of scientific investigations pertaining to respiratory viruses, with a specific focus on HCoV-229E, in the African

continent. Consequently, the dearth of research has impeded the attainment of a thorough comprehension regarding the prevalence, transmission dynamics, and the consequential public health implications associated with HCoV-229E in African countries. These limitations have underscored the necessity for concerted efforts aimed at fortifying the existing surveillance systems and carrying out comprehensive epidemiological studies. By doing so, it is envisaged that there will be a substantial increase in awareness regarding the presence and the significant impact of the virus within the African context (Tao *et al*., 2017). It is crucial to recognize that a multitude of studies have unequivocally indicated that bats serve as the ultimate reservoir hosts for various coronaviruses, including the ancestors of HCoV-229E (Dijkman *et al*., 2008).

The evolutionary trajectory of HCoV-229E has been inextricably linked to bats, thus suggesting a plausible zoonotic origin (Corman *et al*., 2015). Furthermore, scientific investigations have brought attention to the inherent importance of gaining a comprehensive understanding of the structure and the receptor binding properties of the HCoV-229E spike glycoprotein as it relates to immune evasion strategies and the potential for cross-species transmission (Z. Li *et al*., 2019). Notably, endeavors aimed at augmenting our epidemiological knowledge pertaining to human coronaviruses, including HCoV-229E, have yielded valuable insights into interseasonal variations in terms of detection frequencies as well as the clinical manifestations exhibited by infected patients (Cabeça *et al*., 2013). Additionally, studies have underscored the urgent need to incorporate HCoV-229E within routine respiratory virus detection panels in order to avoid underestimating its prevalence and the associated significance (L. Wang et al., 2018). In conclusion, although research pertaining to HCoV-229E in Africa has been encumbered by various limitations, recent

global studies have played a pivotal role in illuminating the evolutionary history, the transmission dynamics, and the clinical implications associated with this particular virus.

It is imperative to underscore the criticality of bolstering surveillance systems, conducting comprehensive epidemiological studies, and integrating HCoV-229E within routine testing protocols as fundamental steps towards enhancing our understanding of the virus and its profound implications on public health.

Epitopes are of utmost importance in comprehending the immune responses elicited against viruses, particularly those of the HCoV-229E classification. It has been demonstrated through research that the presence of antibodies to HCoV-229E and related coronaviruses can result in cross-reactivity, owing to the existence of conserved viral antigenic epitopes (Gorse *et al*., 2020). Specifically, the N-terminal domain of the HCoV-229E spike protein has been identified as a region that exerts dominance in terms of antigenicity, thereby potentially serving as a pan-α-HCoVs epitope (Xiang *et al*., 2022).

It is crucial However to acknowledge that neutralizing epitopes located on the receptor-binding domain (RBD) of HCoV-229E have the propensity to undergo evolutionary changes over time (Jiale Shi *et al*., 2022). In light of this, studies have been conducted to establish the existence of a relatively high degree of amino acid similarity between recognized SARS-CoV-2 epitopes and those found in seasonal HCoVs such as HCoV-OC43, HCoV-HKU1, HCoV-229E, and HCoV-NL63 (Karlsson *et al*., 2020). Moreover, it has been suggested by research that the presence of pre-existing T-cell reactivity in healthy individuals could potentially be attributed to their previous encounters with common coronaviruses like HCoV-229E, HCoV-OC43, HCoV-NL63, or HCoV-HKU1, thereby indicating the possibility of

cross-reactive T-cell responses (Noronha *et al*., 2020). Furthermore, it is worth noting that a pan-coronavirus vaccine has been designed with the intention of achieving comprehensive coverage, and this vaccine incorporates conserved cytotoxic T lymphocyte (CTL) epitopes sourced from various HCoVs, including HCoV-229E (M. Li *et al*., 2021).

Studies have successfully identified potential epitopes that exhibit cross-reactivity between SARS-CoV-2 and seasonal human coronaviruses like HCoV-229E (M. H. Cheng *et al*., 2020; Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, *et al*., 2020). The significance of these findings cannot be overstated, as they highlight the critical nature of comprehending the role epitopes play in the development of vaccines and the subsequent immune responses triggered against coronaviruses.

In summation, Human Coronavirus 229E (HCoV-229E) retains its status as a consequential pathogen within the Coronaviridae family, imparting respiratory illnesses upon the human population. Consequently, comprehending its distinctive characteristics, transmission dynamics, and intricate interplay with host cells assumes paramount importance in the development and implementation of effective prevention and treatment strategies, particularly in the context of emerging coronavirus infections.

#### 2.4.1.2 HCOV NL63

Human coronavirus NL63 (HCoV-NL63), which belongs to the Coronaviridae family and is specifically classified within the genus Alphacoronavirus, was first identified in the year 2004 (L. Zhang et al., 2020). This particular virus is known to primarily induce mild to moderate respiratory illness in individuals who are in good health, yet it can lead to severe

complications in populations that are vulnerable, including infants, elderly individuals, and those with weakened immune systems (Gaunt *et al*., 2010). HCoV-NL63 has been linked to respiratory conditions such as croup (Gaunt *et al*., 2010), bronchiolitis, and pneumonia (Weng *et al*., 2019).

Although infections caused by HCoV-NL63 typically manifest as mild and exhibit symptoms resembling the common cold, they have the potential to progress into more severe lower respiratory tract diseases in specific populations (D. Liu *et al*., 2023). Studies have demonstrated that HCoV-NL63 primarily infects children and individuals with compromised immune systems, resulting in symptoms like coughing, fever, rhinorrhea, and croup (Gao *et al*., 2021). Research has also indicated that HCoV-NL63 shares recognized epitopes with SARS-CoV-2, suggesting certain similarities in their humoral responses (Simula *et al*., 2020). Moreover, the virus has been found to exhibit a higher level of efficiency in infecting a small subpopulation of ACE2+ human respiratory epithelial cells compared to SARS-CoV-2 (Castillo *et al*., 2023). Investigations have additionally been conducted to explore the antiviral properties against HCoV-NL63, with studies revealing the potential antiviral activity of specific plant extracts and phenolic acid constituents against this coronavirus (Tsai *et al*., 2020; Weng *et al*., 2019).

Understanding the epidemiology of HCoV-NL63 is of utmost importance for global surveillance, as evidenced by studies carried out in Taiwan (S. Huang *et al*., 2017). Furthermore, the virus has been associated with severe lower respiratory tract infections in certain subgenotypes, underscoring the significance of monitoring its genetic variations (Y. Wang *et al*., 2020). In summary, HCoV-NL63, a member of the Alphacoronavirus genus, presents a risk of causing severe respiratory illnesses in vulnerable populations, despite

typically resulting in mild symptoms in healthy individuals. Ongoing research endeavors continue to investigate various aspects of this virus, encompassing its clinical manifestations, potential antiviral treatments, and genetic variations, with the aim of gaining a better understanding of HCoV-NL63 infections and effectively managing them.

HCoV-NL63 has the ability to infect individuals of all age groups, not only limited to children and immunocompromised individuals. Infection by HCoV-NL63 can manifest in various symptoms, including cough, fever, rhinorrhea, and croup (K Wang *et al*., 2021). Extensive research has indicated that HCoV-NL63 shares recognized epitopes with SARS-CoV-2, suggesting potential similarities in their humoral responses (Simula *et al*., 2020). Furthermore, scientific investigations have revealed that HCoV-NL63 exhibits a higher level of efficiency in infecting a specific subset of ACE2 human respiratory epithelial cells compared to SARS-CoV-2 (Castillo *et al*., 2022) . Notably, HCoV-NL63 infections are strongly associated with both upper and lower respiratory tract infections across all age groups, with the possibility of severe complications particularly in the elderly and immunocompromised individuals (Kesheh *et al*., 2021).

In affected individuals, this virus has been observed to cause mild upper respiratory symptoms as well as serious lower respiratory tract complications such as bronchiolitis, croup, and pneumonia (P. Li *et al*., 2021). Moreover, it is worth noting that HCoV-NL63 infections can result in severe respiratory illnesses among individuals spanning different age groups (Moës *et al*., 2005). Furthermore, it is important to recognize that HCoV-NL63 infections are frequent and can be acquired at any age, with potent neutralizing activity detected in the sera of patients aged 8 years or older (Hofmann *et al*., 2005).

This virus, HCoV-NL63, has a global presence and has been linked to severe lower respiratory tract disease in individuals across all age groups (Arden *et al*., 2005). Although the majority of HCoV-NL63 infections are associated with common cold symptoms, it is essential to acknowledge that severe respiratory illnesses can occur in individuals from various age groups (L. Zhang et al., 2020). In conclusion, HCoV-NL63 is a highly significant pathogen capable of affecting individuals across all age groups, leading to a wide range of respiratory symptoms. The shared humoral responses between HCoV-NL63 and SARS-CoV-2, as well as the high infectivity of HCoV-NL63 in respiratory epithelial cells, underscore the utmost importance of comprehensively understanding and effectively managing infections caused by HCoV-NL63.

HCoV-NL63, being a constituent of the Alphacoronavirus genus, is distinguished by its enveloped, positive-sense, single-stranded RNA genome of approximately 27–32 kilobases (Ravindra *et al*., 2020). This particular virus encodes a variety of proteins, including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, with particular emphasis on the spike protein, which plays a crucial role in viral entry by binding to angiotensin-converting enzyme 2 (ACE2) receptors (Brielle *et al*., 2020). Notably, HCoV-NL63, akin to SARS-CoV-2, leverages the ACE2 receptor for cellular entry (Gard *et al*., 2021). The spike protein of HCoV-NL63 assumes a significant function in its interaction with the ACE2 receptor, effectively facilitating the virus's entry into host cells (Brielle *et al*., 2020).

Extensive research has confirmed that HCoV-NL63 has the capacity to infect a broad range of host cells, including tertiary monkey kidney cells and the monkey kidney LLC-MK2 cell line (Van Der Hoek *et al*., 2004). Moreover, studies have indicated that HCoV-NL63 might share epitopes with SARS-CoV-2, thereby leading to the development of highly neutralizing

antibodies against SARS-CoV-2 in individuals recovering from COVID-19 (Woldemeskel *et al*., 2021). Furthermore, HCoV-NL63 has been identified as the etiological agent responsible for mild to severe respiratory tract infections, with repeated infections observed throughout an individual's lifetime (Kiyuka *et al*., 2018). Furthermore, HCoV-NL63 has been proposed as a safe model for investigating SARS-CoV-2 infection in a BSL-2 laboratory environment, given its resemblance in terms of utilizing the ACE2 receptor pathway for entry (Gard *et al*., 2021).

The virus has been discovered to interact with heparan sulfate proteoglycans to attach itself to target cells, thereby providing further insights into its mechanisms of infection (Milewska *et al*., 2014). Additionally, HCoV-NL63 has been associated with a unique strategy for interacting with the ACE2 human receptor, thereby highlighting its distinctive characteristics within the realm of coronaviruses (Brielle *et al*., 2020). In conclusion, as an integral member of the Alphacoronavirus genus, HCoV-NL63 manifests specific genomic and protein characteristics that enable its interaction with host cells via the ACE2 receptor, akin to other coronaviruses. Comprehending the molecular mechanisms underlying HCoV-NL63 is of utmost importance when it comes to unraveling its pathogenesis and exploring its potential implications for therapeutic interventions, vaccine development and herd immunity.

It has been evidenced through extensive research that HCoV-NL63 has a wide distribution across the globe and is frequently associated with respiratory tract infections (RTIs) in children (Hofmann *et al*., 2005). In comparison to other human coronaviruses, the prevalence of infections caused by HCoV-NL63 and HCoV-OC43 is higher during early childhood (Dijkman *et al*., 2012).

Numerous studies have unequivocally indicated the global prevalence of HCoV-NL63, with detections documented in numerous countries spanning across all continents (Z. Zeng *et al*., 2017). The initial identification of this virus dates back to the year 2004 in the Netherlands, and since then, it has been reported to be the cause of infections on a global scale (F. Wang *et al*., 2016). Furthermore, it has been established that HCoV-NL63 is associated with severe lower respiratory tract diseases in various regions, including Australia and Kenya (Arden *et al*., 2005; Kiyuka *et al*., 2018). The virus has been found in nasopharyngeal aspirates of patients, with a higher frequency observed among young adults who are experiencing acute respiratory infections (Arden *et al*., 2005). Moreover, comprehensive phylogenetic analyses have provided insights into the evolutionary patterns and circulation dynamics of HCoV-NL63 in diverse regions, such as China and Taiwan (S. Huang *et al*., 2017; Weng *et al*., 2019). Additionally, studies have highlighted the stability of infectious HCoV-NL63 virions, thereby suggesting the potential of the virus to persist in the environment (Florek *et al*., 2014). In conclusion, research conducted in various countries has provided substantial evidence supporting the global distribution of HCoV-NL63, thereby emphasizing its ability to circulate widely and cause respiratory infections in diverse populations worldwide.

The HCoV-NL63 infections have been widely reported in various parts of Africa, among which include countries such as South Africa, Nigeria, Egypt, Morocco, Kenya and Ghana. This therefore demonstrates the virus's great circulation across the whole continent (Otieno et al., 2020). Several researches have definitely shown that there is no limited gender, age and geography for HCoV-NL63 infections hence it affects people irrespective of their demographics and location on the earth. Also, these can account for about 5 percent of all

cases of acute respiratory tract diseases especially during winter season though they affect all age groups (Cabeça & Bellei, 2012).

It is also important to note that in Africa HCoV-NL63 occurs in similar patterns with global ones where this virus has been shown to be responsible for respiratory infections whose severity ranges from mild cold-like symptoms to severe lower respiratory tract infections mainly among populations at risk like infants, elderly people and those individuals whose immune systems are compromised (Dijkman *et al*., 2012; Rochars *et al*., 2016). Amazingly, different scientific studies have consistently demonstrated that there are seasonal fluctuations in the occurrence rates of HCoV-NL63 with spikes reported from summer through fall to winter hence showing a clear temporal relationship for these infections (L. J. Cui *et al*., 2011). Furthermore it must be noted that many investigations have identified upper respiratory tract symptoms such as rhinitis, coughs, fever and sore throat associated with HCoV-NL63 infection whereas lower respiratory tracts infections manifesting bronchitis bronchiolitis and pneumonia may occur indicating how diverse the presentations of the disease can be (Florek *et al*., 2014). Moreover it must be taken into account that there is a high degree of variability in terms of severity when it comes to HCoV-NL63 so that in some cases, pneumonia or CNS diseases may develop in those who are prone to such complications (Y. Wang *et al*., 2020).

Besides, it is of significance to mention that the prevalence of HCoV-NL63 infection has been extensively studied in various settings including hospitals and rural communities thus stressing the necessity for sustained and rigorous surveillance as well as a profound knowledge of this strain's epidemiology (Owusu *et al*., 2021). In essence, the detection of HCoV-NL63 presence in different parts of the world points out that there is an urgent need for continuous monitoring as well as active involvement in research programs aimed at

improving our understanding on this specific strain of coronavirus impact and transmission dynamics.

Covertly, the HCoV-NL63 that is a variant of human corona virus may fluctuate with seasons in Kenya. These variations can occur during winter months which are normally peak seasons for respiratory diseases. Prevalence rates and epidemiological patterns of HCoV-NL63 in Kenya; these are influenced by various factors like location on geographical basis, climatic conditions, population density and individual health-seeking behavior (Kiyuka *et al*., 2018). Earlier researches have also shown that HCoV-NL63 was most detected in winter-spring and not as described seasonality of summer-to-autumn period in tropical countries (Gaunt *et al*., 2010). Furthermore, HCoV-NL63 has been observed to exhibit diverse geospatial patterns (Lau *et al*., 2006). Studies have indicated that during the winter disease season CoVs such as HCoV-229E and HCoV-OC43 cause illness in temperate regions whereas the seasonality of HCoV-NL63 differs across different regions (Lau *et al*., 2006). Similarly, there is a difference between prevalence patterns for various strains of HCoVs with majority having higher frequencies during spring and autumn periods in some areas (Z. Zeng *et al*., 2017). Worldwide activity of this pathogen appears to follow sporadic or lack seasonal nature because it does not show any peaks and valleys throughout the year (Sipulwa *et al*., 2016).

In Kenya specifically, where surveys on this virus have been done, it has been associated with childhood pneumonia cases while measures were put up to monitor its occurrence rate. The molecular epidemiology and evolutionary patterns of NL63-like viruses have been extensively studied among rural coastal communities in Kenya where similarities were found with other endemic coronaviruses such as OC43-like viruses and 229E-like viruses (Kiyuka *et al*., 2018). In summary, the prevalence and seasonality of HCoV-NL63 in Kenya may be

similar to what has been experienced in other parts of the world but can also be influenced by local factors. Therefore, it is necessary to have a constant surveillance system for HCoV-NL63 as well as new investigations about its effects on respiratory illnesses in Kenya.

Research studies on respiratory viruses, particularly HCoV-NL63, are crucial for enhancing the understanding of the epidemiology, clinical characteristics, and genetic diversity of these viruses in Kenya. Collaborative efforts with local and international partners can significantly contribute to the development of evidence-based strategies for the prevention, diagnosis, and management of respiratory infections (Gaunt *et al*., 2010). HCoV-NL63, along with other common human coronaviruses, has been associated with a range of respiratory outcomes, including bronchiolitis and pneumonia (Gaunt *et al*., 2010). Studies have shown that HCoV-NL63 infections can peak during specific seasons, such as in autumn, providing valuable insights for surveillance and preparedness efforts (Lau *et al*., 2006). Furthermore, the prevalence of pre-existing serological cross-reactivity against SARS-CoV-2 in sub-Saharan Africa has been documented, highlighting the importance of understanding the interactions between different coronaviruses, including HCoV-NL63 (Tso *et al*., 2021). Additionally, research has indicated that HCoV-NL63 and SARS-CoV-2 share recognized epitopes, raising questions about potential cross-immunity and the need for further investigation into the implications of previous exposure to SARS-CoV-2 on immunity against HCoV-NL63 (Karami *et al*., 2023).

Collaborative surveillance studies, such as those conducted in Belgium and Japan, have utilized multiplex quantitative RT-PCRs to monitor human coronaviruses, including HCoV-NL63, in primary care settings and hospitals, providing valuable data on the prevalence and circulation of these viruses (N. Fischer *et al*., 2021; Matoba *et al*., 2015). Understanding the

epidemiology and clinical characteristics of HCoV-NL63, as demonstrated in studies from Taiwan and China, can aid in the identification and management of respiratory infections, particularly in hospitalized patients (S. Huang *et al*., 2017; Z. Zeng *et al*., 2017). The collaborative research efforts focusing on HCoV-NL63 and other respiratory viruses therefore are essential for advancing knowledge on these pathogens, which can ultimately inform public health strategies and interventions to combat respiratory infections effectively.

Several studies have indicated that HCoV-NL63 shares epitopes with SARS-CoV-2, thereby suggesting the possibility of cross-reactivity between these two viruses. This cross-reactivity has been observed in both humoral and cellular immune responses, thereby indicating that prior exposure to HCoV-NL63 may have an impact on the immune response to SARS-CoV-2 (Denninger *et al*., 2022; Sayama *et al*., 2023; Simula *et al*., 2020). In addition, the existence of common epitopes between HCoV-NL63 and SARS-CoV-2 has important implications for the development of vaccines and the comprehension of immune memory (Denninger *et al*., 2022; Simula *et al*., 2020). In fact, research has identified conserved epitopes in the spike proteins of various human coronaviruses, including HCoV-NL63, which could potentially serve as targets for the development of pan-coronavirus vaccines (M. Li *et al*., 2021; Xiang *et al*., 2022). Furthermore, numerous studies have underscored the significance of memory immunity in the defense against SARS-CoV-2 and other coronaviruses, such as HCoV-NL63. The reactivation of memory and the occurrence of cross-reactivity play a prominent role in the immune response, wherein antibodies specific to HCoV-NL63 and HCoV-HKU1 are derived from memory immunity rather than as a response to SARS-CoV-2 (Denninger *et al*., 2021, 2022). Taken together, the body of research suggests a intricate interplay between HCoV-NL63 and other coronaviruses, particularly SARS-CoV-2, with regards to shared

epitopes, immune responses, and the development of vaccines. Gaining a comprehensive understanding of these interactions is of utmost importance in advancing our knowledge of coronavirus immunity and devising effective strategies to combat both current and future coronavirus outbreaks.

### 2.4.2 BETACORONAVIRUSES

Betacoronaviruses, a classification that encompasses highly pathogenic viruses such as SARS-CoV, MERS-CoV, and SARS-CoV-2, which are responsible for the ongoing COVID-19 pandemic, demonstrate a wider spectrum of potential host species compared to other coronaviruses. Moreover, it has been observed that alphacoronaviruses exhibit a wider host range and greater genetic diversity in bats when compared to betacoronaviruses (B. He *et al*., 2014). These viruses are closely associated with severe respiratory ailments and pose significant threats to global health security. Their ability to infect a wide range of natural reservoir hosts, which encompass humans, bats, rodents, pigs, and camels, further highlights the magnitude of the danger they present (G. Lee *et al*., 2021).

The expansive range of intermediary hosts that betacoronaviruses and other coronaviruses can infect is exemplified by instances where animals come into contact with HCoV-OC43 or BCoV due to previous close proximity with cattle (Schindell *et al*., 2022). Moreover, it is important to note that genetic recombination within coronaviruses can lead to the expansion of host range or alterations in tissue tropism, further emphasizing the significance of comprehending the ecology and diversity of these viruses (So *et al*., 2019). In summary, betacoronaviruses exhibit a broader host range, infecting mammals such as bats, rodents, and humans, in contrast to other coronaviruses. The knowledge of the host range of these viruses

is of utmost importance in predicting and preventing future zoonotic events and outbreaks. Thus, efforts to investigate the genetic diversity, ecology, and evolution of coronaviruses are imperative for upholding global health security and ensuring pandemic preparedness.

#### 2.4.2.1 HCoV Hku1

Human Coronavirus HKU1 (HCoV-HKU1) is a distinct variety of coronavirus, first identified in 2005 at the prestigious University of Hong Kong, which is a leading institution in scientific research and medical advances. This virus belongs to the renowned family Coronaviridae and it has an affinity towards attacking mainly the complex respiratory system of its hapless hosts thus causing myriad distressing symptoms with severity ranging from relatively mild cold-like conditions to severe debilitating forms of respiratory distress experienced by patients affected by this infection (Dominguez *et al*., 2013). According to virus classification, HCoV-HKU1 was placed appropriately into the genus Betacoronavirus as a positive-strand RNA virus under viral taxonomy (Sechan *et al*., 2022). It is interesting to note that all coronaviruses including HKU1 are commonly viewed as viral pathogens that cause relatively mild diseases mimicking common colds especially among those who have not been infected by any disease for long time (Paules *et al*., 2020). However, new studies have shown a different aspect about these viruses where their impact on respiratory health in diverse populations particularly those living under poor sanitary conditions or are more vulnerable due underlying pathologies like asthma or HIV – AIDS among others (Georgakopoulou *et al*., 2021).

In clinical practice focused on diagnosing and treating various types of respiratory infections, HCoV-HKU1 has frequently been detected among patients admitted for acute respiratory

tract infection; however, it has been intriguingly observed that this particular virus can coexist with other viruses such as human respiratory syncytial virus (HRSV) (Gaunt *et al*., 2010). Epidemiology is a discipline that studies patterns, factors and distribution of diseases in populations. In the case of HCoV-HKU1, it has been closely studied and analyzed by epidemiologists who have documented its widespread distribution among different regions throughout the world including China and Australia (Shao *et al*., 2022; Sloots *et al*., 2006).

Molecular evolution studies on viruses help to show how they change with time. On this basis, research has shown that HCoV-HKU1 as well as other human coronaviruses have globally spread beyond borders and are involved with varying degrees of respiratory disease severities exhibited clinically (Shao *et al*., 2022). Furthermore, it should be noted that HCoV-HKU1 has been detected in various clinical settings leading to cases of encephalitis as well as severe lower respiratory tract infections caused directly by this virus (P. Cao *et al*., 2020; Dijkman *et al*., 2013). Human Coronavirus HKU1 (HCoV-HKU1) therefore is a virus which has undoubtfully been established as a major pathogen causing viral upper respiratory tract infection in humans. Formerly considered to be less severe compared to pathogens causing mild illnesses such as common cold, recent scientific researches clearly show otherwise that this illness can cause serious consequences if not properly managed. Therefore, understanding the genetic makeup, complex epidemiology and diverse clinical manifestations associated with HCoV-HKU1 is crucial towards effective management control and prevention of respiratory infections caused by this viral pathogen.

A change in the concept of human coronaviruses, specifically HCoV-HKU1, has been reported by recent studies. Those viruses were previously perceived as agents causing mild diseases similar to normal colds. It is now recognized that they could cause severe respiratory

syndromes (Gaunt *et al*., 2010; Wevers & Hoek, 2010). These viruses, including HCoV-HKU1, are regularly found in patients diagnosed with acute respiratory tract infections and commonly coexist with other respiratory viruses such as human respiratory syncytial virus (HRSV) (Anirudhan *et al*., 2021). They were initially seen as being quite safe but it turned out that this was not true since human coronaviruses have been found to be associated with far more serious clinical complications especially within susceptible groups (Shao *et al*., 2022). Several cases of bronchiolitis and pneumonia have been linked to four types of human coronaviruses: HCoV-229E, HCoV-HKU1, HCoV-NL63, and HCoV-OC43 among others (Kesheh *et al*., 2021). These cause common colds and limited infections of the respiratory tract which account for a substantial number of all common cold cases (Kesheh *et al*., 2021). They have shown that it could result in either mild common-cold or severe disease symptoms (L. Ren *et al*., 2010). The burden imposed by human coronaviruses, e.g. HCoV-229E, HCoV-OC43, HCoV-HKU1 and HCoV-NL63 in causing acute respiratory tract infections has been well-documented (L. Ren *et al*., 2010).

HCoV-HKU1 is a positive-sense, single-stranded RNA virus with an enveloped structure. Its genome is approximately 29,900 nucleotides long and encodes for various structural proteins like spike (S), envelope (E), membrane (M), and nucleocapsid (N), along with non-structural proteins crucial for viral replication and transcription (L. Zhang et al., 2020), (X. Huang *et al*., 2015).. The virus is closely related to bat-derived coronaviruses, bat-SL-CoVZC45, and bat-SL-CoVZXC21, with genetic identities of 88% and is distinct from other coronaviruses like SARS-CoV and MERS-CoV (L. Zhang et al., 2020). HCoV-HKU1 is known to cause acute respiratory tract infections in children and is prevalent globally (X. Chen *et al*., 2022). It is one of the four endemic coronaviruses and is associated with a broad spectrum of

respiratory diseases of varying severity (Shao *et al*., 2022). Epidemiologically, HCoV-HKU1 is detected across all age groups and is likely to cause repeated infections throughout life (Gaunt *et al*., 2010). The virus uses O-acetylated sialic acid residues as attachment receptor determinants and employs hemagglutinin-esterase protein as a receptor-destroying enzyme during infection (X. Huang *et al*., 2015).

Despite being associated with respiratory diseases, HCoV-HKU1 remains poorly understood due to the lack of an in vitro model for viral replication and the inability to culture the virus in vitro (Dominguez *et al*., 2013; Milewska *et al*., 2020). Clinical presentations of HCoV infections, including HCoV-HKU1, do not show species-specific differences except for a trend towards higher gastrointestinal symptoms severity in HCoV-HKU1 infections (Bouvier *et al*., 2018). The virus has been detected in patients with acute lower respiratory tract infections, bronchiolitis, and asthmatic exacerbations, with seasonal peaks observed during winter and spring (Yu *et al*., 2012). HCoV-HKU1 infections have been reported in various regions globally, emphasizing the importance of continued surveillance and research to better understand and manage infections caused by this coronavirus (Y. Cui *et al*., 2021).

The global distribution of HCoV-HKU1 extends to various countries, such as Australia (Sloots *et al*., 2006), Hong Kong (Lau *et al*., 2006), and Greece (Georgakopoulou *et al*., 2021), which serves as indicative evidence of its widespread prevalence. Through meticulous investigations and comprehensive studies, it has been ascertained that HCoV-HKU1 is closely associated with respiratory diseases of varying degrees of severity (Shao *et al*., 2022). Furthermore, the virus has been detected in patients suffering from viral pneumonia, as astutely determined by Joyjinda *et al*. (2019) , and it is noteworthy that it is globally pervasive, causing acute respiratory infections primarily amongst children (Joyjinda *et al*.,

2019; Sipulwa *et al*., 2016). It is imperative to acknowledge that HCoV-HKU1 effortlessly invades human ciliated airway epithelial cell cultures, permitting its replication and subsequent production of progeny virions (Pyrc *et al*., 2010). In light of this, it is crucial to emphasize that the virus exhibits a universal prevalence and is significantly linked to both upper and lower respiratory tract diseases, which can manifest in individuals of all age groups, encompassing both children and adults (Dominguez *et al*., 2013).

The information on the occurrence and prevalence of HCoV-HKU1 in Africa may be relatively few comparing to other areas, but various studies have confirmed the presence of this virus in many countries across the continent (Gaunt *et al*., 2010). They are known to cause such respiratory diseases as bronchiolitis as well as pneumonia (Kesheh *et al*., 2021). The virus invades host cells through a mechanism that is well characterized and involves a highly glycosylated spike protein (Y. Cui *et al*., 2021). It has been shown that this virus can be found in different parts of the world such as Australia, Hong Kong etc. (Lau *et al*., 2006; Sloots *et al*., 2006). As a result it is considered to be an endemic respiratory pathogen introduced zoonotically (Hulswit *et al*., 2019). Besides, infection with HCoV-HKU1 has been documented among various mammalian species indicating its interspecies transmission potential (Woo *et al*., 2009).

Among other things formerly among hospitalized children in Guangzhou China where it was identified as a common causative agent for pneumonia together with other types of lower respiratory tract infections including asthma or bronchitis (Z. Zeng *et al*., 2017), extensive research has revealed that epidemiology of HCOv-HKU1 is wide-spread. Moreover there are seasonal fluctuations in the prevalence of viral respiratory diseases during different years in Israel's population making them arise every year due to influenza viruses and coronaviruses

especially during late autumn through to early spring time frames showing that seasonality is observed among these diseases. In Greece, an immunocompromised patient suffering from severe lower respiratory tract infection was found to be positive for HCoV-HKU1 hence showing that such viruses have an impact on vulnerable populations (Georgakopoulou *et al*., 2021). In conclusion, comprehensive data on the distribution and prevalence of HCoV-HKU1 in Africa are limited. Nonetheless, global studies provide evidence of this virus' presence and effects on respiratory infections across different populations. More investigations should therefore be carried out to understand better HCoV-HKU1's extent in African public health.

HCoV-HKU1 has been recognized as one of the respiratory viruses that are circulating globally, including in Kenya, a nation located in East Africa. Although there may be a scarcity of specific information regarding the prevalence and distribution of HCoV-HKU1 in Kenya, diligent surveillance endeavors and extensive research studies have successfully identified the presence of this virus within the Kenyan population (Sipulwa *et al*., 2016). HCoV-HKU1 belongs to the group 2 classification of coronaviruses, alongside its counterpart HCoV-OC43, and is prominently linked to various respiratory ailments that afflict individuals worldwide (Lau *et al*., 2006). It is of utmost importance to note that HCoV-HKU1 is categorized as one of the four endemic coronaviruses, which denotes its continuous existence within specific populations over a prolonged period of time (Sechan *et al*., 2022). In order to effectively invade and infect host cells, the virus employs a highly glycosylated spike protein, as elucidated by (Y. Cui *et al*., 2021), and relies upon O-acetylated sialic acid as a determinant for attachment to its receptor (X. Huang *et al*., 2015).

The epitopes of HCoV-HKU1 have been a subject of interest due to their potential cross-reactivity with other coronaviruses. Studies have shown that memory CD4+ T cells can cross-react with SARS-CoV-2 and common cold coronaviruses like HCoV-HKU1 (Mateus *et al*., 2020). Additionally, research has identified potential cross-reactive epitopes between SARS-CoV-2 and common cold coronaviruses, including HCoV-HKU1, through sequence alignments (Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, *et al*., 2020). Furthermore, monoclonal antibodies against HCoV-HKU1 spike proteins have been found to recognize specific immunodominant epitopes, shedding light on the regions that are targeted by the immune system (Guruprasad, 2021). Moreover, the specificity of antibodies towards HCoV-NL63 and HCoV-HKU1 epitopes suggests that these antibodies may be derived from memory immunity rather than a response to SARS-CoV-2 (Denninger *et al*., 2022). This highlights the role of memory reactivation and cross-reactivity in the defense against coronaviruses. The conserved epitopes on the S2 subunit of various human coronaviruses, including HCoV-HKU1, contribute to cross-reactivity and immune responses (Klose *et al*., 2023). The presence of antibodies specific to HCoV-HKU1 spike proteins in sera indicates immune responses to this specific coronavirus (Bangaru *et al*., 2022). Additionally, the high amino acid similarity between recognized SARS-CoV-2 epitopes and seasonal HCoVs like HCoV-HKU1 suggests potential cross-reactivity and shared immune responses (Karlsson *et al*., 2020). The identification of CD8+ T cell epitopes with homology to HCoV-HKU1 and other coronaviruses further supports the concept of conserved antigen recognition (Pushpakumara *et al*., 2021). In conclusion, the research on HCoV-HKU1 epitopes underscores the interconnectedness of immune responses to different coronaviruses and the potential for cross-reactivity. Understanding the epitopes

of HCoV-HKU1 and their interactions with the immune system is crucial for elucidating broader immune responses to coronaviruses and also informing vaccine development strategies.

In conclusion, HCoV-HKU1 stands as an influential respiratory virus that continuously circulates worldwide, even extending its reach to the nation of Kenya. Although specific data pertaining to its prevalence within the Kenyan population may be limited in nature, research conducted in other countries unequivocally accentuates its correlation with respiratory diseases and its capability to infiltrate human cells. Henceforth, it is incumbent upon the scientific community to undertake enhanced surveillance and research endeavors to gain a more comprehensive understanding of the prevalence and impact of HCoV-HKU1 within the population of Kenya.

#### 2.4.2.2 HCOV OC43

Human Coronavirus OC43 (HCoV-OC43) is classified as a member of the Coronaviridae family, characterized as an enveloped, positive-sense single-stranded RNA virus. Initially identified in 1967 from patients exhibiting symptoms of the common cold, HCoV-OC43 is acknowledged as one of the common human coronaviruses alongside HCoV-229E, NL63, and HKU1 (Lau *et al*., 2006). These coronaviruses are known to cause various respiratory illnesses, including pneumonia, with HCoV-OC43 being part of group 2 coronaviruses (Z. Zeng *et al*., 2017).

Despite typically causing mild respiratory infections, HCoV-OC43 has raised concerns due to its potential to induce severe diseases in vulnerable populations and its genetic variability, which presents challenges in vaccine development and therapeutic strategies (Lau *et al*.,

2006). Research has shown that HCoV-OC43 is associated with mild upper respiratory infections globally and is the coronavirus most commonly linked to "common colds" (G. Lee *et al*., 2021). Additionally, studies have indicated that HCoV-OC43 is prevalent among infants, causing a significant proportion of upper and lower respiratory tract illnesses (Dijkman *et al*., 2012). Furthermore, investigations into the evolutionary history of closely related group 2 coronaviruses, including HCoV-OC43, have revealed relatively recent common ancestors for these species-specific coronaviruses (Moës *et al*., 2005).

The presence of preexisting antibodies against HCoV-OC43 has been suggested to influence the durability and cross-reactivity of the humoral response to SARS-CoV-2 vaccination, indicating a potential interplay between different coronaviruses in the immune response (B. Hu *et al*., 2022). This cross-reactivity has also been observed in studies where higher antibody levels against seasonal human coronaviruses, including HCoV-OC43, were associated with a more robust humoral immune response following SARS-CoV-2 vaccination (Asamoah-Boaheng *et al*., 2022). Thus, HCoV-OC43, as a member of the Coronaviridae family, plays a significant role in respiratory infections, ranging from mild illnesses to severe diseases. Its genetic variability and interactions with other coronaviruses highlight the importance of understanding its impact on human health and the development of effective preventive and therapeutic measures.

As HCoV-OC43 elicits mild upper respiratory infections, frequently presents itself in the form of common colds (G. Lee *et al*., 2021), extensive research has demonstrated it as highly prevalent among infants and significantly contributes to both upper and lower respiratory tract illnesses (Dijkman *et al*., 2012). Investigations focusing on the evolutionary history of group 2 coronaviruses, including HCoV-OC43, have successfully identified recent common

ancestors for these species-specific coronaviruses (N. Fischer *et al*., 2021). In fact, HCoV-OC43, along with other human coronaviruses, has been estimated to be the second most common cause of common colds on a global scale (N. Fischer *et al*., 2021). Infections caused by HCoV-OC43 can exhibit a wide range of manifestations, varying from asymptomatic to mild to moderate upper respiratory tract illnesses, thereby establishing it as a frequent etiological agent of mild respiratory disease (Kissler *et al*., 2020) . Furthermore, compelling investigations have indicated that human coronaviruses, including HCoV-OC43, are typically associated with mild upper respiratory tract infections; however, severe cases have indeed been documented (Masse *et al*., 2020) . Notably, the epidemiology of human coronaviruses, such as HCoV-OC43, has been subjected to extensive scrutiny across diverse geographic regions worldwide, thus emphasizing their broad distribution and profound impact on respiratory health (Faye *et al*., 2022; S. Zhang *et al*., 2018). These meticulously conducted studies underscore the utmost importance of monitoring and comprehending the intricate circulation dynamics of human coronaviruses in order to augment the management of respiratory infections (Al-Khannaq *et al*., 2016; Sipulwa *et al*., 2016). Furthermore, the prevalence of human coronaviruses, including HCoV-OC43, has been meticulously documented in diverse patient groups, thereby providing invaluable insights into the clinical characteristics and epidemiology of these infections (Cabeça *et al*., 2013; Yip *et al*., 2016; Z. Zeng *et al*., 2017). To summarize, HCoV-OC43 undoubtedly serves as a substantial contributor to mild respiratory infections, specifically common colds. Thus, it is of utmost importance to thoroughly comprehend its prevalence, evolutionary background, and clinical implications in order to effectively manage and surveil public health measures.

HCoV-OC43, exhibits both structural and genetic similarities to SARS-CoV-2, thereby establishing a connection between these two coronaviruses. The genome of HCoV-OC43, which spans approximately 30,000 nucleotides, encodes a diverse array of proteins, including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, all of which play pivotal roles in the viral life cycle and pathogenesis (C. Wang et al., 2021) . Of particular importance is the spike protein of HCoV-OC43, which acts as a key determinant for viral entry by binding to specific receptors on the host cell surface, facilitating membrane fusion, and initiating the infectious process (Cerutti *et al*., 2021).

The extent of relatedness between HCoV-OC43 and SARS-CoV-2 is further underscored by studies demonstrating a sequence identity of approximately 30% between their respective spike proteins (Woldemeskel *et al*., 2021). Consequently, it becomes evident that HCoV-OC43 assumes a prominent position among the endemic human coronaviruses, predominantly giving rise to respiratory illnesses that range from mild manifestations reminiscent of the common cold to more severe respiratory infections, particularly in individuals who are immunocompromised or harbor underlying health conditions (Zhao *et al*., 2014). Moreover, the association between HCoV-OC43 and COVID-19 disease severity has been established, with the presence of cross-reactive antibodies against the spike protein of HCoV-OC43 significantly correlating with the severity of the infection (Cerutti *et al*., 2021). Consequently, the investigation of antibody responses to HCoV-OC43 has attracted considerable scientific interest, yielding valuable insights into various contexts. For instance, studies have revealed that higher levels of antibodies against HCoV-OC43 are indicative of a more robust humoral immune response subsequent to SARS-CoV-2 vaccination, thereby

highlighting the potential interplay between these coronaviruses in terms of immunological memory (Becerra-Artiles et al., 2023).

Therefore it becomes increasingly apparent that HCoV-OC43, much like its counterpart SARS-CoV-2, occupies a central position in the realm of respiratory infections. Consequently, an in-depth understanding of its genetic composition, protein structure, as well as its intricate interactions with the immune system, becomes of paramount importance in order to fully comprehend its pathogenic potential and the potential implications it carries for the development of vaccines and the management of associated diseases.

The seasonal circulation patterns of HCoV-OC43, similar to other coronaviruses, demonstrate fluctuations contingent upon the prevailing climatic conditions. Its epidemiology has been extensively investigated in various parts of the world. In Guangzhou China where it was first recognized over forty years ago , HCOv – OC43 has been found to cause common colds and mild respiratory tracts infections (S. Zhang *et al*., 2018). Similarly, in Belgium, HCoV-OC43 and HCoV-229E were first identified in the 1960s, shedding light on their presence and circulation within the human population (Moës *et al*., 2005). Additionally, these genotypes have emerged over time suggesting that Yamagata has experienced genetic diversity of HCoV-OC43 due to continuous evolution processes (Matoba et al., 2015).

In regions classified as temperate, such as France and the United States, the incidence of HCoV-OC43 is more pronounced during the winter months, mirroring the behavior of influenza and other respiratory viruses. This observed seasonal variation can be attributed to various factors including diminished humidity levels, heightened indoor crowding, and a

compromised immune response during colder weather. Conversely, in tropical regions like Thailand and Malaysia, where the climate remains relatively stable throughout the year, the manifestation of seasonal patterns pertaining to HCoV-OC43 may be less conspicuous (Al-Khannaq *et al*., 2016; Owusu *et al*., 2021).

Extensive investigations have divulged that HCoV-OC43 tends to reach its zenith during the autumn and winter months, albeit with some variability in peak activity witnessed across different geographical locations. For instance, in Hong Kong, HCoV-OC43 predominated during the fall and winter, while HCoV-HKU1 exhibited its peak activity during the winter months, albeit with a handful of cases occurring in the spring and summer seasons as well (Yip *et al*., 2016). Similarly, in Korea, HCoV-OC43 reigned supreme, with its highest prevalence observed consistently during the winter months (Choi & Smith, 2021). The seasonal circulation of HCoV-OC43 is by no means restricted to specific regions; rather, it represents a global phenomenon. Studies conducted in Iran and Senegal have also reported analogous findings pertaining to the seasonality of HCoV-OC43 (Faye *et al*., 2022; Karami *et al*., 2023). Furthermore, the presence of preexisting antibodies against HCoV-OC43 has been postulated to exert an influence on the durability and cross-reactivity of the humoral response to SARS-CoV-2, thereby suggesting a plausible connection between seasonal coronaviruses and immunity (B. Hu *et al*., 2022).

In conclusion, the seasonal circulation of HCoV-OC43 adheres to a pattern that is characteristic of respiratory viruses, with an augmented prevalence during the winter months in temperate regions. Nonetheless, the seasonal patterns may exhibit variations in tropical regions characterized by more stable climates. A comprehensive understanding of these seasonal fluctuations assumes paramount importance in the realm of public health

preparedness and response strategies, particularly when confronted with the advent of emerging respiratory infections, such as COVID-19.

There is also considerable research examining potential treatment or interventions targeting OC43. This includes some studies looking into the antiviral properties of chloroquine against 'HCoV-OC43' as well as others that focus on how green tea polyphenols can inhibit the replication of this virus (C. Hu et al., 2022). As such, these efforts provide a great deal of information regarding the features, dynamics of transmission and targeting possibilities linked to HCov – OC43. Thus, knowledge about HCOV-OC43 can be achieved by conducting thorough investigations.

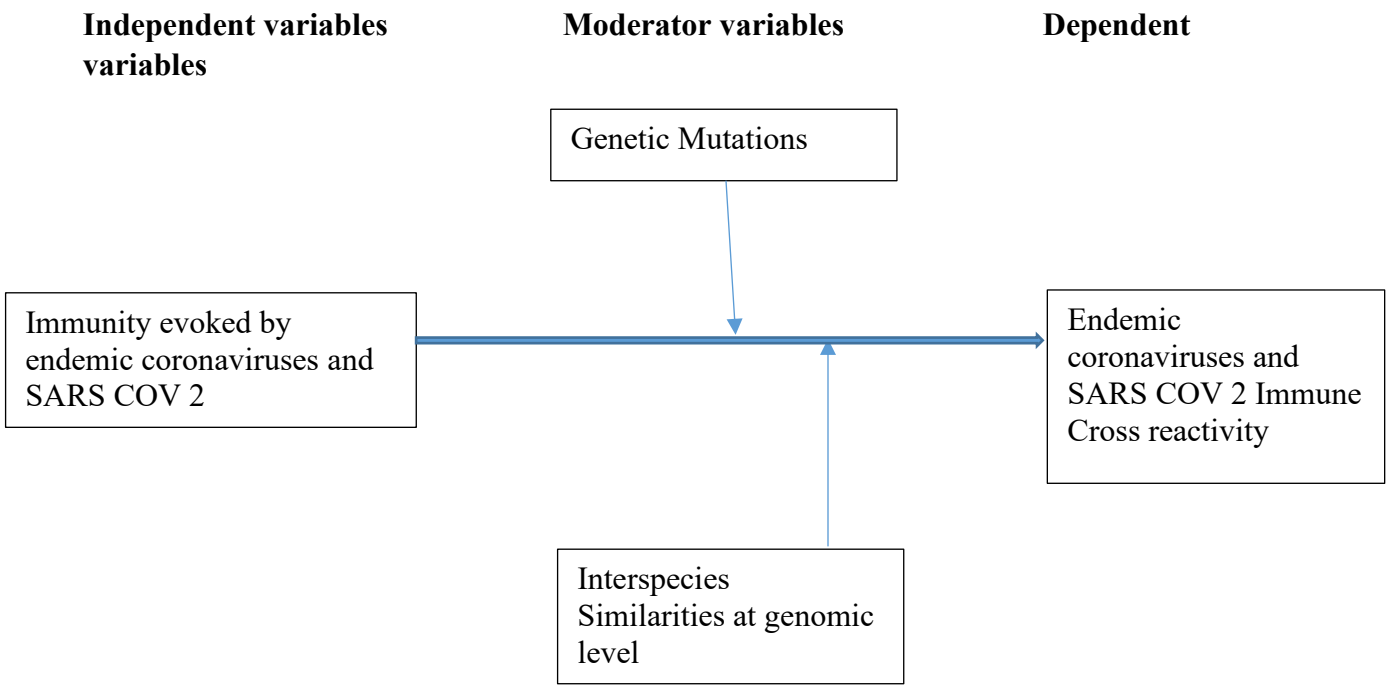
The molecular characteristic aspects and circulation dynamics of human coronavirus have been extensively studied in Kenya where respiratory infections particularly those caused by HCoV-OC43 virus remain substantial areas of research focus (Sipulwa *et al*., 2016). However, it should be noted that comprehensive surveillance networks for respiratory viruses including coronaviruses may not be uniformly implemented throughout the country thereby resulting in lack of specific information regarding prevalence and distribution of HCoV-OC43 in Kenya. In view of the fact that COVID-19 cases are distributed across continents while taking into account that this virus has become endemic to Kenya, it becomes important to enhance surveillance so as to understand better how common this virus is within Kenya. This would help manage respiratory diseases effectively besides contributing towards global efforts aimed at controlling transmission routes for corona viruses.

Epitope mapping is essential for understanding immune responses to viruses like HCoV-OC43. Studies have identified immunodominant epitopes on the spike protein of related

coronaviruses like SARS-CoV, with Site IV being a major epitope (Y. He *et al*., 2004). Cross-reactive antibody responses have been observed between SARS-CoV-2 and seasonal coronaviruses, including unique epitopes of HCoV-OC43 (Klompus *et al*., 2020). Conservation analysis has shown candidate CD8+ T cell epitopes of SARS-CoV-2 with homology to HCoVs like OC43 (Pushpakumara *et al*., 2021). Additionally, T cell responses to SARS-CoV-2 have been associated with cross-reactions with HCoVs, including OC43, suggesting potential immune memory involvement (Gallais *et al*., 2021). Furthermore, studies have highlighted the presence of pre-existing T-cell reactivity in healthy individuals, likely due to past exposure to common cold coronaviruses such as HCoV-OC43 (Noronha *et al*., 2020).

The design of a pan-coronavirus vaccine has incorporated conserved CTL epitopes from various HCoVs, including OC43, to provide broad coverage (M. Li *et al*., 2021). Moreover, shared B and T cell epitopes have been identified among relevant coronaviruses, including HCoV-OC43, indicating potential cross-protection (Pacheco-Olvera *et al*., 2020). Research has also shown that antibodies specific to HCoVs like OC43 and HKU1 are present in individuals, with potential implications for immune responses to SARS-CoV-2 (Bangaru *et al*., 2022). Additionally, cross-reactivity of antibodies against N proteins of coronaviruses like HCoV-OC43 has been observed, suggesting some level of immune recognition overlap (Tajuelo *et al*., 2022). In conclusion, the studies provide valuable insights into the epitopes, immune responses, and potential cross-reactivity involving HCoV-OC43, shedding light on the complex interplay between different coronaviruses and the immune system.

**Conceptual framework**



# CHAPTER THREE

## 3.0 RESEARCH DESIGN AND METHODOLOGY

### 3.1 Study area

The study was conducted at Moi Teaching and Referral Hospital (MTRH), an urban tertiary healthcare facility located in North Rift, Kenya. MTRH serves as the primary referral center for the North Rift Region and hosts a dedicated COVID-19 isolation unit that admits patients from surrounding counties. The hospital's diagnostic infrastructure includes a well-equipped molecular laboratory, which functions as a regional hub for COVID-19 testing. This laboratory receives and processes SARS-CoV-2 samples from across the North Rift using real-time polymerase chain reaction (PCR) technology.

### 3.2 Study design

This is a cross-sectional study that was aimed at analyzing samples from patients visiting Moi Teaching and Referral Hospital in December 2021.

### 3.3 Study Population

The study population comprised individuals who tested positive for COVID-19. Their SARS-COV-2 variants was analysed and their epitopes determined.

### 3.4 Inclusion and Exclusion criteria

#### 3.4.1 Inclusion criteria

a. Consented adult patients diagnosed with COVID-19

### 3.4.2 Exclusion criteria

a. Patients that did not grant informed written consent

b. Samples with SARS-CoV-2 RT-PCR Cycle threshold (Ct) > 30

## 3.5 Sample size Determination

Yamane 1967 formula as cited by Israel (Israel, 1992)

$$n = \frac{N}{1 + N(e)^2}$$

- Where
  - n is the sample size
  - N is the population previous recently estimate 90
  - e is the error margin (0.05)

Thus sample size was 74.

## 3.6 Laboratory procedures

### 3.6.1 Recruitment and Sample collection

Nasopharyngeal and oropharyngeal swab samples were collected in December 2021 from patients scheduled for routine sample collection at Moi Teaching and Referral Hospital. All participants provided written informed consent prior to sampling and consecutive sampling was done. Some individuals exhibited symptoms consistent with COVID-19, while others, who were attending the hospital for unrelated medical conditions, were asymptomatic. Sterile swabs were used to collect the specimens, which were then placed in viral transport medium

(VTM) and transported to the laboratory within six hours under cold chain conditions. Upon arrival, samples were stored at −20°C. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was isolated from swabs that tested positive and subsequently used for sequencing. Demographic information was recorded for all participants. Patients were classified as symptomatic or asymptomatic based on the World Health Organization COVID-19 case definition guidelines.

### 3.6.2 Molecular variations of SARS-CoV-2 in North Rift Kenya

#### 3.6.2.1 RNA extraction, genome amplification, and sequencing

Viral RNA was extracted manually from nasopharyngeal and oropharyngeal swab specimens using the DaAn Gene Viral RNA Extraction Kit (DaAn Gene Co., Ltd., Guangzhou, China). Clinical specimens, previously collected in viral transport medium (VTM), were first equilibrated to room temperature and gently mixed to ensure homogeneity.

For each extraction, an aliquot of 200 µL of the clinical sample was transferred into a sterile 1.5 mL microcentrifuge tube. Subsequently, 400 µL of lysis buffer (Buffer R) was added to the sample to facilitate disruption of viral particles and release of RNA. The mixture was vortexed briefly and incubated at room temperature for 10 minutes to allow for complete lysis and viral inactivation.

Following the lysis step, 400 µL of binding buffer containing isopropanol was added to the lysate to promote binding of RNA to the magnetic beads provided in the kit. Approximately 20–40 µL of magnetic bead suspension was introduced into the mixture, followed by gentle mixing and incubation for 5–10 minutes. The tube was then placed on a magnetic stand until the beads separated, and the supernatant was carefully discarded without disturbing the pellet.

The bound RNA was subjected to two successive washes to remove impurities. First, 500 µL of Wash Buffer 1 was added, followed by gentle mixing, magnetic separation, and supernatant removal. This was followed by a second wash using 500 µL of Wash Buffer 2 (ethanol-based), after which the beads were again magnetically separated, and the supernatant discarded. To remove residual ethanol, the magnetic bead pellet was air-dried at room temperature for 5–10 minutes, taking care not to overdry the beads to preserve RNA integrity.

Elution of purified RNA was carried out by resuspending the dried pellet in 50–100 µL of RNase-free elution buffer. The mixture was incubated at room temperature for 5 minutes to facilitate release of RNA from the beads. Finally, the tube was placed on the magnetic stand, and the eluate containing purified viral RNA was transferred into a fresh RNase-free microcentrifuge tube.

The extracted RNA was either processed immediately for downstream applications such as reverse transcription quantitative polymerase chain reaction (RT-qPCR) or stored at −80 °C until further use.

Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was carried out using the DaAn Gene Detection Kit for 2019-nCoV (DaAnGene, China) on the Rotor-Gene Q real-time PCR platform. RNA extraction and real-time reverse transcription polymerase chain reaction (RT-PCR) were performed at the Moi Teaching and Referral Hospital (MTRH) Molecular Laboratory. Samples yielding positive results with cycle threshold (Ct) values below 30 for combinations of target genes, E & RdRp, S & RdRp, or

ORF1ab & N were selected for sequencing in Kenya Medical Research Institute (KEMRI) Nairobi as illustrated in figure 1 below

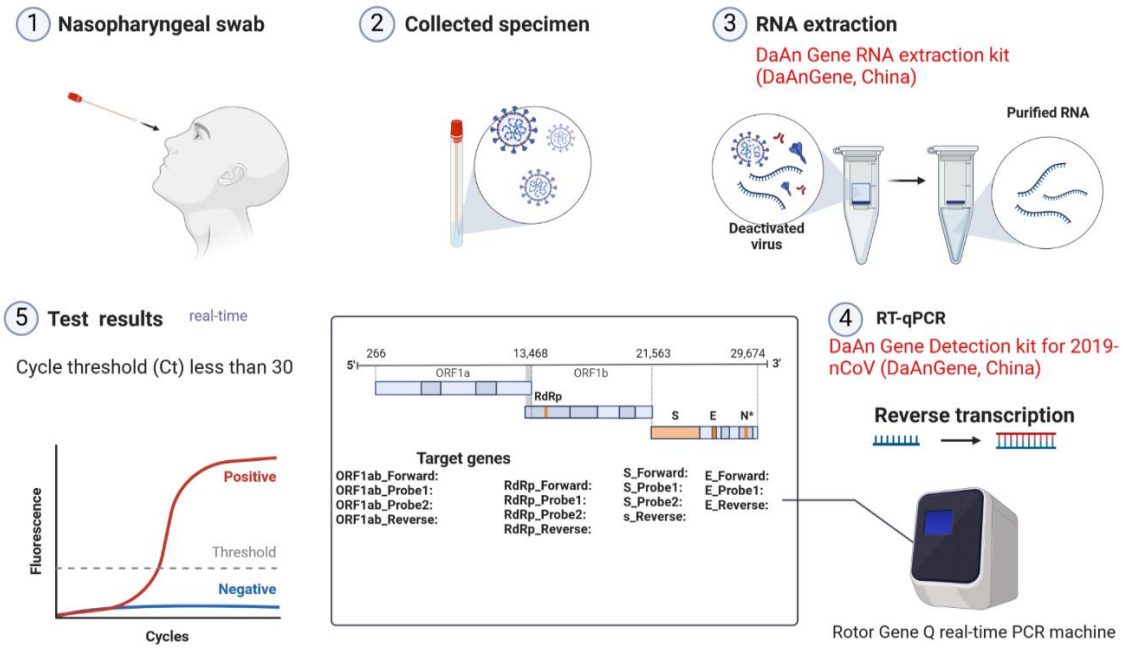



*Figure 1: Lab procedures: Sample collection and amplification illustrated*

Prior to sequencing, sample preparation was done through the following steps. Firstly, reverse transcription (RT), a procedure in which RNA is transformed into complementary DNA (cDNA) through SuperScript™ IV Reverse Transcriptase kit from Illumina was done. The cDNA then underwent PCR amplification with NEBNext® Ultra™ II DNA Library Prep Kit directed at SARS-CoV-2 genes. During library preparation, adapters for Illumina sequencing were added to the amplified fragments followed by cleanup with Agencourt AMPure XP system and size selection to remove contaminants. To ensure optimal loading concentration, the library's concentration was quantified using qubit. Finally, following the

manufacturer's instructions, the sequencing was done using the MiSeq sequencing platform (Illumina, California USA).

### 3.6.2.2 Bioinformatics

The sequence data underwent processing through bioinformatic tools within UseGalaxy. Initial quality checks were performed using FastQC for FastQ data. Subsequent trimming was carried out using the trim sequence tool. Alignment was executed, referencing the SARS-CoV-2 genome with NCBI accession number MN908947.2 (Miller et al., 2020), utilizing bowtie2. This reference genome is the first genome to be sequenced and subsequent mutations are documented in reference to it. To eliminate duplicates, the Markduplicates tool was used. Variant calling was conducted using FreeBayes, followed by VCF filtering using VCFfilter to achieve specific criteria, including DP >30, AF >0.2, and QUAL >30. Ultimately, the consensus tool was employed to generate a fastA file consolidating the variant information as shown in figure 2.

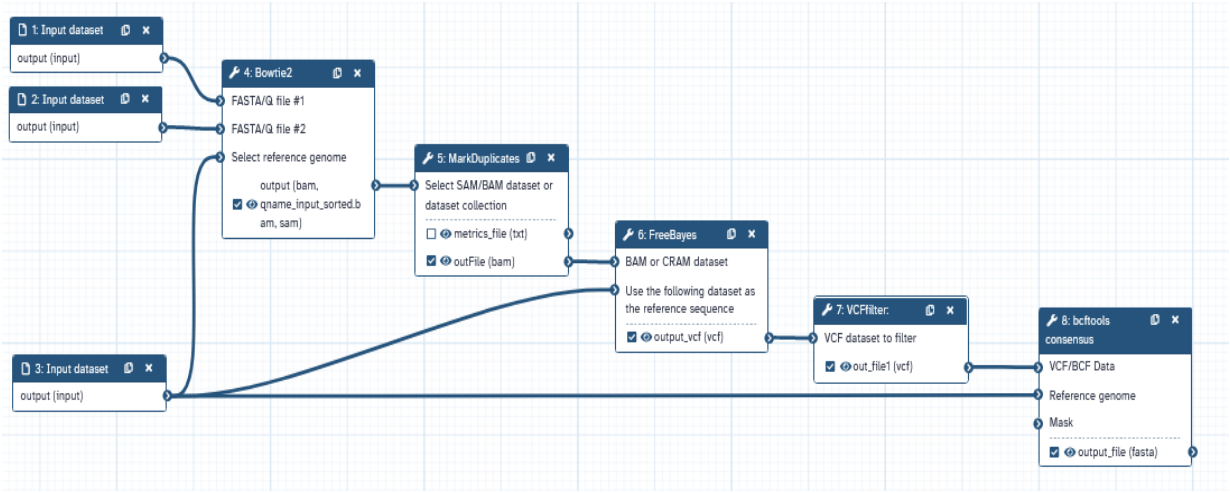



*Figure 2: Workflow in GALAXY: Data in FASTQ format was entered after trimming through box1 (for forward) and box 2 (for reverse). The reference genome was put through box 3 and was used for bowrie2, freebays and consensus.*

The SARS-CoV-2 genome sequences underwent categorization into genetic lineages using the Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) tool. This platform is widely adopted for the dynamic classification of SARS-CoV-2 lineages, allowing for standardized global tracking of viral evolution and transmission patterns. Its frequent updates and alignment with global genomic surveillance initiatives make it suitable for identifying and comparing emerging variants in real time. Annotation of the sequences was performed using the COVID-19 Annotator (Mercatelli et al., 2021), which enables high-throughput detection of mutations and facilitates detailed genomic characterization essential for understanding variant-specific molecular signatures. This pipeline was selected for its robustness, reproducibility, and relevance in global SARS-CoV-2 genomic epidemiology.

To find the new mutations, sequences from earlier mutations were retrieved from GISAID database. These mutations were annotated using the COVID-19 Annotation tool (http://giorgilab.unibo.it/coronannotator/) (Mercatelli et al., 2021) for comparison with mutations identified in the study.

#### 3.6.2.3 Phylogenetic tree

The phylogenetic tree, representing the most likely evolutionary history, was performed in MEGA11. The evolutionary relationships among the sequences were reconstructed using the Neighbor-Joining method. Evolutionary distances were calculated using the Maximum Composite Likelihood method, expressed as the number of base substitutions per site. To account for rate variation across sites, a gamma distribution was applied (shape parameter = 1). All ambiguous positions were removed for each sequence pair (pairwise deletion).

### 3.6.3 Predicting the epitopes

The T and B cell epitopes were predicted using the Immune Epitope Database (IEDB) bioinformatics tools. The first stage comprised of applying ORF Finder to experimental SARS-CoV-2 genomes in order to identify Open Reading Frames (ORFs). The ORFs were then carefully annotated to determine their possible roles or features. The next step after annotation was attempted was the translation of protein coding sequences found by EXPASY into all six frames. This comprehensive translation has led to numerous potential protein sequences which ensures that no probable coding regions are ignored. Then, a BlastP analysis was performed where these translated SARS-CoV-2 proteins were compared against a database of known proteins obtained from the National Center for Biotechnology Information (NCBI). The aim of this comparison was identifying most likely matches for the translated sequences and providing insights into the functions that may be attributed to the identified proteins. Also, epitope specific information on SARS-CoV-2 was obtained from Immune Epitope Database (IEDB), which is an all-inclusive source of epitope data. These well-known epitopes served as useful basis for further study. After obtaining the epitopes, we matched them with those predicted by IEDB for every such protein using BlastP. This careful cross-examination made it possible to discover any meaningful hits amid forecasted proteins and identified epitopes, thereby highlighting potential immunogenic sites within the experimental of SARS-CoV-2 genome.

### 3.6.4 Molecular similarities with other Coronaviruses

The epitopes were compared to the published epitope genes of the other Coronaviruses using local alignment methods to identify the similarities and conserved regions among the coronaviruses.

#### 3.6.4.1 Comparing SARS-CoV-2 genome proteins with epitopes from endemic coronaviruses

Known epitopes for each endemic coronavirus strain, including HCOV HKU1, HCOV OC43, HCOV NL63, and HCOV 229E, were collected from IEDB, establishing essential reference points for subsequent alignment with SARS-CoV-2 proteins. Using BlastP, an alignment analysis was conducted, wherein the annotated proteins of SARS-CoV-2 were aligned with the epitopes obtained from endemic coronaviruses. This intricate comparative analysis aimed to elucidate regions of similarity and potential cross-reactivity between the viral strains. The BlastP results were then subjected to thorough scrutiny, with a focus on identifying regions of significant alignment between the epitopes of endemic coronaviruses and the annotated proteins of SARS-CoV-2. The evaluation process involved meticulous scrutiny to prioritize potential cross-reactive epitopes for further investigation. Regions demonstrating notable alignment were singled out for detailed examination, with an emphasis on their potential implications for immune recognition and cross-reactivity. Finally, the aligned epitopes were meticulously documented alongside the corresponding SARS-CoV-2 annotated proteins.

#### 3.6.4.2 Comparing epitopes from SARS-COV-2 and endemic coronaviruses

In order to understand potential cross-reactivity and immune responses across different coronaviruses, epitopes from endemic coronaviruses (HCOV HKU1, HCOV OC43, HCOV NL63, and HCOV 229E) were systematically aligned with those from SARS-CoV-2 annotated proteins using BlastP.

Firstly, the known epitopes for every endemic coronavirus were collected painstakingly from Immune Epitope Database (IEDB). Using BlastP, epitopes of endemic coronaviruses were aligned with the epitopes of annotated proteins from SARS-CoV-2. This alignment aimed at revealing similarities in regions and possible cross-reactivities between the epitopes of other types of corona viruses

A subsequent analysis was then done on BlastP results by carefully examining them to discover areas of significant similarity between the epitopes of endemic coronaviruses and those of annotated proteins for SARS-COV-2. This is an important analysis because it helps identify characteristic features that are conserved among various SARS-COV-2 annotated proteins.

The resultant set of blast results was analyzed thoroughly to identify significantly aligning parts in HCoVs' epitopes and SARS-CoV-2's annotated proteins. From this result some common antigenic determinants could be retained within these two groups.

The aligned epitope sequences were subsequently recorded with their corresponding SARS-CoV-2 annotated protein names.

Additionally, the overlapping epitopes were aligned with SARS-CoV-2 immunodominant epitopes that were downloaded from a previous article (Maghsood *et al*., 2021) to estimate the impact of those epitopes.

### 3.7 Data Analysis

Descriptive statistics were used to summarize the data, including frequencies. Inferential statistics were also performed. The Chi-Square Test of Independence was used to assess

associations between categorical variables such as clade and symptom status, and clade and gender. Fisher's Exact Test was applied in cases where expected cell frequencies were less than five, due to small sample sizes. A Z-Test for Two Proportions was conducted to compare the proportion of symptomatic patients between males and females. Relative Risk and Odds Ratio were calculated to quantify the strength of associations. The Cochran-Mantel-Haenszel Test was used to evaluate the association between clade and symptom status while controlling for gender. All statistical analyses were conducted using R version 4.2.1, with statistical significance set at a threshold of 0.05.

### 3.8 Ethical Considerations

Ethical approval for the study was obtained from the Moi Teaching and Referral Hospital/Moi University Institutional Research and Ethics Committee (Ref. IREC/036/2021 and ELD/MTRH/R&P/10/2/V.2/2010). To maintain confidentiality, participants were assigned unique code identifiers in place of their personal names. Participation was entirely voluntary, and no coercion was involved. All participants were fully informed about the purpose, procedures, potential risks, and benefits of the study prior to enrollment, and written informed consent was obtained. Participants retained the right to withdraw from the study at any point without penalty or loss of access to healthcare services.

The study excluded children in order to minimize ethical complexity and avoid involving vulnerable populations without strong scientific justification. Recruitment was conducted fairly, ensuring that no specific group was overburdened or unjustly excluded.

All participant data and laboratory results were securely stored in password-protected digital databases and locked physical cabinets. These data will be retained for five years after the

conclusion of the study, after which all personally identifiable information will be securely destroyed—electronic records will be permanently deleted, and physical documents will be shredded.

Research findings are being disseminated to participants, the local community, and relevant stakeholders through summary reports, institutional meetings, and community engagement forums. Broader dissemination includes presentations at scientific conferences and submission to peer-reviewed journals. Efforts have been made to ensure that findings are communicated in an accessible and comprehensible manner to both scientific and non-specialist audiences.

# CHAPTER FOUR

## 4.0 RESULTS

### 4.1 Genetic Diversity of SARS-COV-2 in North Rift Kenya

#### 4.1.1 Sequencing data

The sequencing data obtained from the collection of 44 SARS-CoV-2 samples exhibited an average base coverage of 95%. On average, each base had a coverage depth of 282.73, with the maximum coverage depth soaring to 1031.16. The sequences have been submitted to GISAID and are published under the accession numbers EPI_ISL_19004000 to EPI_ISL_19004016 and EPI_ISL_19004981 to EPI_ISL_19005007.

#### 4.1.2 Variants of concern

The genomic evaluation indicated the detection of two variants of concern: Delta (2 samples) and Omicron (42 samples). Within the Omicron variant, a complex genetic diversity was identified, featuring four distinct clades labeled as BA.1, BA.1.1, BA.1.1.1, and BA.1.14. It is particularly noteworthy that the BA.1.1 clade is predominant, representing a significant

majority at 86.4% of all sequenced samples. These genetic findings are visually illustrated in Figure 2.

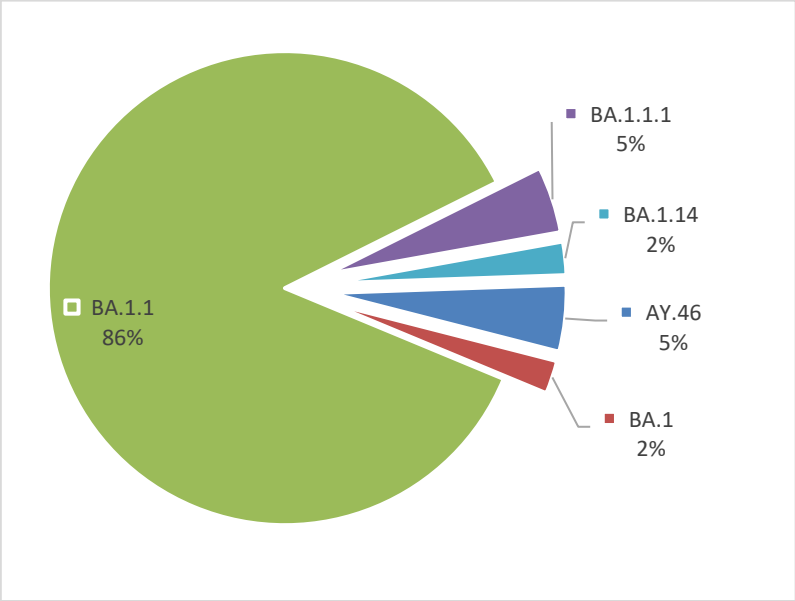



*Figure 3: SARS-CoV-2 Clade frequency:* ***The nomenclature for clades follows the PANGO system. BA.1 signifies the initial Omicron variant. Each following number indicates a subvariant of the one prior. AY.46 is derived from the Delta variant.***

Among symptomatic patients, 16 (42%) were male and 22 (58%) were female. For symptomatic males, the BA.1.1 variant was detected in 14 (88%) cases, while one case (6%) each was reported for the BA.1 and AY.46 variants. No symptomatic males were found to have the BA.1.1.1 or BA.1.14 variants. Among symptomatic females, 18 (82%) cases were attributed to the BA.1.1 variant, with 2 (10%) cases of the BA.1.1.1 variant and one case (5%) each for the BA.1.14 and AY.46 variants. No symptomatic female cases were reported for the BA.1 variant.

In asymptomatic patients, 2 (33%) were male and 4 (67%) were female. All asymptomatic individuals, regardless of gender, were infected with the BA.1.1 variant. No asymptomatic

cases were identified for the BA.1.1.1, BA.1, BA.1.14, or AY.46 variants in either gender as shown in Table 2.

**Table 2: Distribution of the clades in relation to gender and symptoms**

| Clade | Symptoms | Gender | | |
|---|---|---|---|---|
| | | Male | Female | Total |
| **BA.1.1** | Symptomatic | 14 | 18 | 32 |
| | Asymptomatic | 2 | 4 | 6 |
| **BA.1.1.1** | Symptomatic | 0 | 2 | 2 |
| | Asymptomatic | 0 | 0 | 0 |
| **BA.1** | Symptomatic | 1 | 0 | 1 |
| | Asymptomatic | 0 | 0 | 0 |
| **BA.1.14** | Symptomatic | 0 | 1 | 1 |
| | Asymptomatic | 0 | 0 | 0 |
| **AY.46** | Symptomatic | 1 | 1 | 2 |
| | Asymptomatic | 0 | 0 | 0 |
| **Total** | Symptomatic | 16 | 22 | 38 |
| | Asymptomatic | 2 | 4 | 6 |
| | All | 18 | 26 | 44 |

### 4.1.3 Relationship between SARS-CoV-2 Mutation Clades, Symptoms, and Gender

In this study, a dataset of patients infected with various clades of SARS-CoV-2 was analyzed to investigate whether specific clades are associated with symptomatic disease and whether these associations differ by gender. The aim was

1. To determine if there is an association between clade and symptom status.
2. To assess whether gender influences the likelihood of being symptomatic.
3. To quantify the strength of these associations using odds ratios and relative risks.

To address these objectives, a combination of inferential statistical methods were employed including chi-square tests, logistic regression, and proportion comparisons.

The dataset consisted of 44 patients categorized by clade (BA.1.1, BA.1.1.1, BA.1, BA.1.14, AY.46), symptom status (symptomatic or asymptomatic), and gender (male or female). The distribution of patients across these categories is summarized in Table 2.

#### 4.1.3.1 Association Between Clade and Symptom Status

A chi-square test of independence was conducted to examine the association between clade and symptom status. The results showed a significant association ($\chi 2(4)=12.45$, $p=0.014$), indicating that the distribution of symptomatic and asymptomatic patients differed across clades.

For smaller clades (e.g., BA.1.1.1, BA.1.14), Fisher's exact test was used due to low sample sizes. No significant associations were found for these clades as shown in Appendix 1.

#### 4.1.3.2 Comparison of Symptomatic Proportions by Gender

A Z-test for two proportions was conducted to compare the proportion of symptomatic males ($p1=16/18=0.89$) and symptomatic females ($p2=22/26=0.85$). The test yielded a non-significant result ($z=0.45$, $p=0.653$), indicating no difference in symptomatic rates between genders as shown in Appendix 1.

The relative risk (RR) of being symptomatic for males compared to females was calculated as:

$RR=0.85/0.89=1.05$

The odds ratio (OR) was:

$OR=8/5.5 =1.45$

These values suggest that there was 45% more likelihood of being symptomatic for males, compared to females

#### 4.1.3.3 Association between clade and symptom status while controlling for gender

The CMH test was used to assess the association between clade and symptom status while controlling for gender. The test yielded a significant result (CMH=11.23, p=0.024), indicating that the association between clade and symptom status remained after adjusting for gender as shown in Appendix 1.

### 4.1.4 Mutation frequencies

A total of 65 mutations were identified across various samples with differing frequencies. The "5'UTR" region exhibited a variant at position 241 with a frequency of 31.8%. In the NSP2 protein, the A555A mutation at position 2470 was observed in 18.2% of individuals. Numerous mutations were detected in the NSP3 protein, including A889A, I388L, F106F, S1265, and A1892T, with frequencies ranging from 11.4% to 97.7%.

The T492I mutation at position 10029 in the NSP4 protein, was identified in 97.7% of individuals. NSP6 exhibited variants such as L105 and I189V, each occurring at a frequency of 95.5%. High-frequency mutations were also noted in NSP10 and NSP12b, including V57V and P314L, the latter of which was present in all individuals (100%). NSP12b also features N591N at a lower frequency of 11.4%. In NSP13, the G170S mutation was found in 63.6% of individuals. The NSP14 protein contained the I42V mutation, observed in 95.5% of individuals. In the Spike (S) protein, multiple mutations such as A67, T95I, I210, N211K, and others were found, with frequencies ranging from 6.8% to 100%. Notably, the D614G variant in the Spike protein was present in all individuals (100%).

Other notable mutations included G339D, R346K, and S371L, with frequencies of 84.1%, 81.8%, and 22.7%, respectively. The ORF3a protein exhibited a T64T mutation in 95.5% of

individuals. Mutations in the Envelope (E) and Membrane (M) proteins, such as T9I and Q19E, showed frequencies of 88.6% and 95.5%, respectively. The ORF6 and ORF7b regions included variants like M19M and L17L with frequencies of 95.5% and 61.4%, respectively as shown in table 3.

**Table 3: Mutation frequency for each mutated position in Percentage**

| *Ref pos* | *Protein* | *Ref var* | *Q var* | *Variant* | *Var name* | *No. of Subjects* | *Mutation frequency in Percentage* |
|---|---|---|---|---|---|---|---|
| 241 | 5'UTR | C | T | 241 | 5'UTR:241 | 14 | 31.8 |
| 2470 | NSP2 | C | T | A555A | NSP2:A555A | 8 | 18.2 |
| 3037 | NSP3 | C | T | F106F | NSP3:F106F | 43 | 97.7 |
| 3881 | NSP3 | A | C | I388L | NSP3:I388L | 5 | 11.4 |
| 5386 | NSP3 | T | G | A889A | NSP3:A889A | 42 | 95.5 |
| 6513 | NSP3 | GTT | . | S1265 | NSP3:S1265 | 42 | 95.5 |
| 8393 | NSP3 | G | A | A1892T | NSP3:A1892T | 42 | 95.5 |
| 10029 | NSP4 | C | T | T492I | NSP4:T492I | 43 | 97.7 |
| 11286 | NSP6 | TGTCTGGTT | . | L105 | NSP6:L105 | 42 | 95.5 |
| 11537 | NSP6 | A | G | I189V | NSP6:I189V | 42 | 95.5 |
| 13195 | NSP10 | T | C | V57V | NSP10:V57V | 42 | 95.5 |
| 14408 | NSP12b | C | T | P314L | NSP12b:P314L | 44 | 100.0 |
| 15240 | NSP12b | C | T | N591N | NSP12b:N591N | 5 | 11.4 |
| 16744 | NSP13 | G | A | G170S | NSP13:G170S | 28 | 63.6 |
| 18163 | NSP14 | A | G | I42V | NSP14:I42V | 42 | 95.5 |
| 21762 | S | C | . | A67 | S:A67 | 42 | 95.5 |
| 21764 | S | A | . | A67 | S:A67 | 42 | 95.5 |
| 21767 | S | CATG | . | I68 | S:I68 | 42 | 95.5 |
| 21846 | S | C | T | T95I | S:T95I | 42 | 95.5 |
| 21987 | S | GTGTTTATT | . | G142 | S:G142 | 42 | 95.5 |
| 22193 | S | . | T | I210 | S:I210 | 19 | 43.2 |
| 22195 | S | T | G | N211K | S:N211K | 19 | 43.2 |
| 22197 | S | TA | GC | L212C | S:L212C | 19 | 43.2 |
| 22201 | S | . | AGC | S214 | S:S214 | 19 | 43.2 |
| 22202 | S | . | A | V213 | S:V213 | 19 | 43.2 |
| 22203 | S | . | A | R214 | S:R214 | 19 | 43.2 |
| 22204 | S | T | A | R214R | S:R214R | 19 | 43.2 |
| 22578 | S | G | A | G339D | S:G339D | 37 | 84.1 |
| 22599 | S | G | A | R346K | S:R346K | 36 | 81.8 |
| 22673 | S | TC | CT | S371L | S:S371L | 10 | 22.7 |
| 22679 | S | T | C | S373P | S:S373P | 10 | 22.7 |
| 22686 | S | C | T | S375F | S:S375F | 10 | 22.7 |
| 22813 | S | G | T | K417N | S:K417N | 3 | 6.8 |
| 22882 | S | T | G | N440K | S:N440K | 18 | 40.9 |
| 22898 | S | G | A | G446S | S:G446S | 18 | 40.9 |
| 22992 | S | G | A | S477N | S:S477N | 16 | 36.4 |
| 22995 | S | C | A | T478K | S:T478K | 18 | 40.9 |
| 23013 | S | A | C | E484A | S:E484A | 18 | 40.9 |
| 23040 | S | A | G | Q493R | S:Q493R | 17 | 38.6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 23048 | S | G | A | G496S | S:G496S | 17 | 38.6 |
| 23055 | S | A | G | Q498R | S:Q498R | 17 | 38.6 |
| 23063 | S | A | T | N501Y | S:N501Y | 16 | 36.4 |
| 23075 | S | T | C | Y505H | S:Y505H | 16 | 36.4 |
| 23202 | S | C | A | T547K | S:T547K | 42 | 95.5 |
| 23403 | S | A | G | D614G | S:D614G | 44 | 100.0 |
| 23525 | S | C | T | H655Y | S:H655Y | 42 | 95.5 |
| 23599 | S | T | G | N679K | S:N679K | 42 | 95.5 |
| 23604 | S | C | A | P681H | S:P681H | 42 | 95.5 |
| 23854 | S | C | A | N764K | S:N764K | 38 | 86.4 |
| 23948 | S | G | T | D796Y | S:D796Y | 40 | 90.9 |
| 24130 | S | C | A | N856K | S:N856K | 42 | 95.5 |
| 24424 | S | A | T | Q954H | S:Q954H | 42 | 95.5 |
| 24469 | S | T | A | N969K | S:N969K | 42 | 95.5 |
| 24503 | S | C | T | L981F | S:L981F | 42 | 95.5 |
| 25000 | S | C | T | D1146D | S:D1146D | 42 | 95.5 |
| 25584 | ORF3a | C | T | T64T | ORF3a:T64T | 42 | 95.5 |
| 26270 | E | C | T | T9I | E:T9I | 39 | 88.6 |
| 26577 | M | C | G | Q19E | M:Q19E | 42 | 95.5 |
| 26709 | M | G | A | A63T | M:A63T | 42 | 95.5 |
| 27259 | ORF6 | A | C | M19M | ORF6:M19M | 42 | 95.5 |
| 27807 | ORF7b | C | T | L17L | ORF7b:L17L | 27 | 61.4 |
| 28271 | 3'UTR | A | T | 28271 | 3'UTR:28271 | 42 | 95.5 |
| 28311 | N | C | T | P13L | N:P13L | 42 | 95.5 |
| 28362 | N | GAGAACGCA | . | E31 | N:E31 | 42 | 95.5 |
| 28881 | N | GGG | AAC | RG203KR | N:RG203KR | 42 | 95.5 |

### 4.1.5 Relationship between SARS-CoV-2 Mutation Frequencies and Proteins Types

The dataset of mutations across various proteins of the SARS-CoV-2 genome was analyzed to investigate the following:

Which mutations are most frequent and whether certain proteins exhibit higher mutation rates.

Whether there is an association between protein type and mutation frequency.

Patterns of co-occurrence among mutations.

To address these questions, biostatistical methods, including chi-square tests, logistic regression, and network analysis were employed.

#### 4.1.5.1 Association between protein type and mutation frequency categories

A chi-square test was conducted to examine the association between SARS-CoV-2 protein type and mutation frequency categories. The results showed a significant association ($\chi 2(6)=18.45$, $p=0.005$), suggesting that mutation frequencies vary significantly across proteins as shown in Appendix 1.

#### 4.1.5.2 Predicting the probability of a mutation being frequent (>50%) based on protein type and position

A logistic regression model was fitted to predict the probability of a SARS-CoV-2 mutation being frequent (>50%) based on protein type and position. Mutations in the Spike protein were significantly more likely to be frequent compared to other proteins (OR = 3.42, 95% CI: 1.12–10.34, p=0.031) as shown in Appendix 1.

#### 4.1.5.3 Co-occurrence patterns among mutations in the Spike protein

Network analysis revealed strong co-occurrence patterns among mutations in the Spike protein, particularly D614G, P681H, and N501Y. These mutations formed a densely connected cluster, indicating potential functional dependencies as shown in Appendix 1.

### 4.1.6 Synonymous and Non-Synonymous mutations

Among the 65 SARS-CoV-2 mutations identified in the study, 42 were classified as non-synonymous, 10 as synonymous, 7 as deletions, 4 as insertions, and 2 as extragenic mutations located in the 3'UTR and 5'UTR areas, as shown in figure 3.

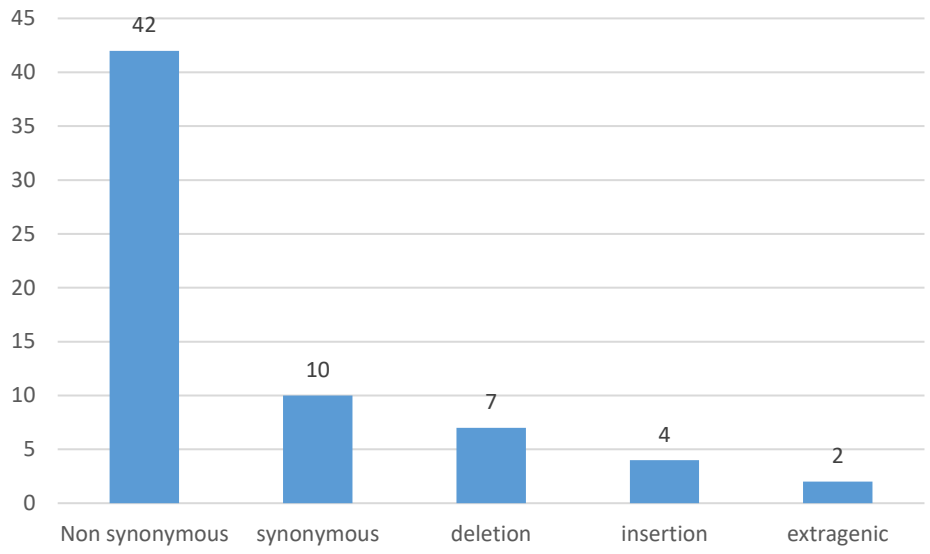



***Figure 4: Distribution of various mutations of SARS-CoV-2 in respect to their type***

Among the 42 non-synonymous mutations, seven mutations: NSP3:I388L; S:E484A; S:G446S; S:G496S; S:N211K; S:N440K; and S:N501Y were uniquely identified in symptomatic patients. All four insertions: S:I210; S:R214; S:S214; and S:V213, were present only in symptomatic patients. Likewise, one synonymous mutation; S:R214R out of ten and one extragenic mutation: 5'UTR:241 out of two were found solely in symptomatic patients. The remaining mutations were observed in both symptomatic and asymptomatic individuals.

### 4.1.7 Phylogenetic Analysis of SARS-CoV-2 Sequences and Their Evolutionary Relationships

The evolutionally relationships of the sequences were displayed in a phylogenetic tree (figure 4). The analysis included 45 nucleotide sequences. The final dataset comprised a total of 29,888 positions. The reference genome was the initial Wuhan strain very close relationship with SME26_MTRH_S37_SARS-CoV-2 Genome while other genomes such as

SME112_MTRH_S17_SARS-CoV-2 and SME124_MTRH_S18_SARS-CoV-2 had very distanced relationship with the initial genome

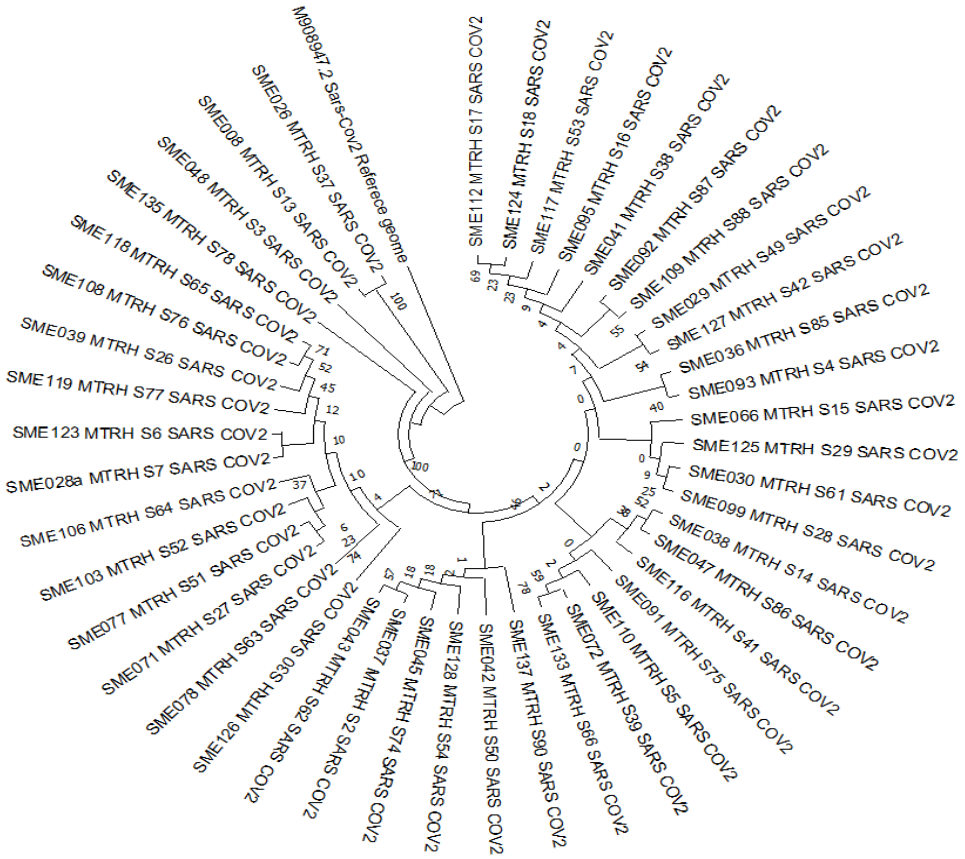



*Figure 5: Evolutionary relationships among SARS-CoV-2 sequences. Each branch represents SARS-CoV-2 genomes found in the study samples.*

### 4.1.8 Emerging mutations

The initial mutation, designated as S:R214R, stands out due to its significant occurrence, representing a substantial 43.2% of the total cases analyzed within the scope of the research study conducted. Yet, it is particularly fascinating to note that this specific mutation has not been detected in any of the previously examined waves, leading to the intriguing conclusion that its appearance might be uniquely associated with the particular cohort being studied in this instance. On the other hand, the second mutation, referred to as NSP2:A555A, has been

found in 18.2% of the cases included in the study group. This mutation, much like its predecessor, was also notably absent from the earlier waves of analysis, which raises questions about its timing and potential implications as shown in table 4.

**Table 4: Newly discovered mutations**

| Name | Cases in study | Symptomatic |
|---|---|---|
| S:R214R | 19 (43.2%) | 19 (100%) |
| NSP2:A555A | 8 (18.2%) | 7 (88%) |

## 4.2 Immunoinformatic Profiling of SARS-CoV-2 Variants in North Rift Kenya

### 4.2.1 B cell and T cell epitopes

A total of 12,285 SARS-CoV-2 epitopes derived from the reference Wuhan genome were retrieved from the Immune Epitope Database (IEDB). Comparative analysis revealed that 5,154 of these epitopes were conserved within the genomes sequenced from clinical isolates collected at Moi Teaching and Referral Hospital (MTRH), Kenya, in December 2021. The epitope repertoire was further classified into functional immune categories, including 8,597 linear B-cell epitopes, of which 4,087 (47.5%) were retained in the MTRH genomes; 2,326 T-cell MHC class I epitopes, of which 374 (16.1%) were identified; and 1,362 T-cell MHC class II epitopes, with 693 (50.9%) being present in the MTRH-derived sequences.

The detected epitopes were distributed across multiple coding regions of the SARS-CoV-2 genome. However, a subset of genes—namely envelope (E), ORF7b, and ORF9b—did not exhibit any matching epitopes from the Wuhan reference in the MTRH sequences, suggesting possible sequence divergence or deletions in these regions. A gene-by-gene breakdown of

epitope conservation between the Wuhan reference strain and MTRH isolates is provided in Table 5.

**Table 5: Number of downloaded for various immune cells Epitope across each protein of Wuhan SARS-CoV-2 genome and those of them that were found o MTRH genome**

| | Bcell | | Tcell-MHC1 | | Tcell-MHC2 | | **Total** | |
|---|---|---|---|---|---|---|---|---|
| | Wuhan | MTRH | Wuhan | MTRH | Wuhan | MTRH | **Wuhan** | **MTRH** |
| **Envelope** | 41 | 0 | 33 | 0 | 20 | 0 | **94** | **0** |
| **Membrane** | 198 | 165 | 134 | 33 | 101 | 77 | **433** | **275** |
| **Nucleoprotein** | 784 | 405 | 189 | 45 | 198 | 134 | **1171** | **584** |
| **ORF1ab** | 4598 | 2487 | 1122 | 177 | 327 | 249 | **6047** | **2913** |
| **ORF3a** | 181 | 81 | 85 | 8 | 26 | 8 | **292** | **97** |
| **ORF6** | 33 | 33 | 15 | 5 | 10 | 8 | **58** | **46** |
| **ORF7a** | 59 | 20 | 30 | 2 | 12 | 0 | **101** | **22** |
| **ORF7b** | 2 | 0 | 8 | 0 | 1 | 0 | **11** | **0** |
| **ORF8** | 97 | 52 | 28 | 6 | 26 | 9 | **151** | **67** |
| **ORF9b** | 1 | 0 | 10 | 0 | 1 | 0 | **12** | **0** |
| **Spike** | 2603 | 844 | 672 | 98 | 640 | 208 | **3915** | **1150** |
| Total | **8597** | **4087** | **2326** | **374** | **1362** | **693** | **12285** | **5154** |

## 4.3 Immune overlap of the SARS-CoV-2 variants circulating in North Rift Kenya and the endemic human coronaviruses

### 4.3.1 Overlapping epitopes

The genome of SARS-CoV-2 isolated in North Rift Kenya contains epitopes from endemic coronaviruses – HCoV-OC43, HCoV-HKU1, HCoV-NL63 and HCoV-229E.

The Membrane gene of SARS-CoV-2 has 3 epitopes that are similar to some one's found in HCoV-OC43 thus suggesting their cross-reactivity potential. However, there is no matching in terms of epitopes between SARS-CoV-2 and HCoV-HKU1.

With respect to the Nucleoprotein gene, only two epitopic regions from SARS-CoV-2 were found to be similar to those present in HCoV-OC43 while one is similar to that for HCoC-HKU1 with no matches identified for either HCoC-NL63 or HCoC229E.

Similarly, there are a significant number of common epitopes between SARS-CoV and all four endemic coronaviruses within the ORF1ab gene hence making them distinct. In particular, the total number of such areas in SARS-CoV is composed of 306 which conform with those seen with respect to OC43; 192 resemble those associated with MKU1; 135 correspond similarly like NL63 while as many as 179 fully match with 229E virus.

In case of the Spike gene, eleven epitopes on the genome of SARS– Co-Virus -2 show similarity with those found on OC43 and thirteen .with HKU

Furthermore, one spike-like protein displays similarity to NL63 while three spike-like proteins share limited resemblance towards. Therefore for this study our focus will be on those genes that have cross reactivity between SARS-CoV-2 and other viruses circulating in this region.

In general, from the genomes of SARS-CoV-2 virus isolated in North Rift Kenya, 322 epitopes demonstrated cross-reactivity towards epitopes of endemic coronaviruses. Namely, there are 136 epitopes from HcoV-HKU1; 206 from HCoV-OC43; 136 from HCoV-NL63 and finally 182 from HCoV-229E which suggests that there is a possibility for immune interactions occurring between these viruses and SARSCoV2. Table 6 provides the detailed breakdown of these epitopes.

**Table 6: Number of epitopes for endemic coronaviruses found on SARS-COV-2 genome isolated in North rift Kenya**

| Gene | HCoV-OC43 | HCoV-HKU1 | HCoV-NL63 | HCoV-229E |
|---|---|---|---|---|
| Membrane | 3 | 0 | 0 | 0 |
| Nucleoprotein | 2 | 1 | 0 | 0 |
| ORF1ab | 306 | 192 | 135 | 179 |
| Spike | 11 | 13 | 1 | 3 |
| Total | 322 | 206 | 136 | 182 |

A total of 27 spike protein epitopes from endemic human coronaviruses were identified as potentially cross-reactive with SARS-CoV-2. These epitopes were retrieved from the Immune Epitope Database (IEDB) and represent linear B-cell and T-cell epitopes derived from HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E.

The majority of the epitopes originated from HCoV-OC43 and HCoV-HKU1, both of which are betacoronaviruses, like SARS-CoV-2. Most of the identified epitopes were B-cell epitopes, suggesting potential antibody-mediated cross-reactivity. A smaller number of epitopes were T-cell epitopes presented by MHC class II molecules, such as the sequences YEMYVKWPWYVWLLI from HCoV-HKU1 and SFFEDLLFDKVKLSDVGFVE, also from HCoV-HKU1.

Several epitopes such as YYVKWPWYVWLL (IEDB ID 84110), LKDIGTYEYYVKWPWY (85650), and YEMYVKWPWYVWLLIS (1708895) contain the conserved motif "VKWPWYVWLL", which appears in both HCoV-HKU1 and HCoV-OC43. This motif is structurally similar to immunodominant regions found in the SARS-CoV-2 spike protein, suggesting a possible shared immunogenic region that could trigger memory responses in previously exposed individuals.

Other recurrent sequences, such as SRLDALEAEAQIDRLI and its variants, also appear across multiple epitopes, further supporting the presence of conserved immunological hotspots among seasonal coronaviruses as shown in table 7.

**Table 7: The cross-reacting spike epitope from endemic coronaviruses**

| IEDB ID | Virus | Immune cell | Epitope Sequence |
|---|---|---|---|
| **77848** | HCoV-OC43 | Bcell | ECSKASSRSAIEDLLFDKVKLSDV |
| **84110** | HCoV-OC43 | Bcell | YYVKWPWYYVWLL |
| **77858** | HCoV-OC43 | Bcell | TSIPNLPDFKEELDQWFKNQTSVA |
| **84513** | HCoV-OC43 | Bcell | DLICVQSYKGIKVLPP |
| **85237** | HCoV-OC43 | Bcell | ILSRLDALEAEAQIDR |
| **85634** | HCoV-OC43 | Bcell | LICVQSYKGIKVLPPL |
| **85650** | HCoV-OC43 | Bcell | LKDIGTYEYYVKWPWY |
| **85739** | HCoV-OC43 | Bcell | LSRLDALEAEAQIDRL |
| **86254** | HCoV-OC43 | Bcell | RDLICVQSYKGIKVLP |
| **86533** | HCoV-OC43 | Bcell | SRLDALEAEAQIDRLI |
| **87237** | HCoV-OC43 | Bcell | TSIPNLPDFKEELDQWFKNQTSVAP |
| **1656676** | HCoV-NL63 | Bcell | FENYIKWPWWVWLIIS |
| **1310951** | HCoV-HKU1 | Tcell-MHC2 | YEMYVKWPWYVWLLI |
| **1386658** | HCoV-HKU1 | Bcell | PHCGSSSRSFFEDLLFDKVKLSDVGF |
| **1644870** | HCoV-HKU1 | Bcell | ALEAQVQIDRLINGRL |
| **1649596** | HCoV-HKU1 | Bcell | DALEAQVQIDRLINGR |
| **i1667205** | HCoV-HKU1 | Bcell | ILSRLDALEAQVQIDR |
| **1673904** | HCoV-HKU1 | Bcell | LEAQVQIDRLINGRLT |
| **1675295** | HCoV-HKU1 | Bcell | LKDIGTYEMYVKWPWY |
| **1677933** | HCoV-HKU1 | Bcell | LSRLDALEAQVQIDRL |
| **1691286** | HCoV-HKU1 | Bcell | RSFFEDLLFDKVKLSD |
| **1695319** | HCoV-HKU1 | Bcell | SRLDALEAQVQIDRLI |
| **1708895** | HCoV-HKU1 | Bcell | YEMYVKWPWYVWLLIS |
| **1869240** | HCoV-HKU1 | Tcell-MHC2 | SFFEDLLFDKVKLSDVGFVE |
| **2180202** | HCoV-HKU1 | Bcell | SSRSLLEDLLFNKVKLSDV |
| **1655637** | HCoV-229E | Bcell | ETYIKWPWWVWLCISV |
| **1683464** | HCoV-229E | Bcell | NRVETYIKWPWWVWLC |
| **1702665** | HCoV-229E | Bcell | VETYIKWPWWVWLCIS |

### 4.3.2 Cross-Reactivity Between HCoV- OC43 and SARS-CoV-2 Spike Epitopes Comparison of SARS-CoV-2 Epitopes

An analysis of the cross-reactivity between SARS-CoV-2 epitopes and particular spike epitopes of the human coronavirus OC43 (HCoV OC43) was done. For each spike epitope from HCoV OC43, it has its unique iedbID and Table 8 counts the number of SARS-CoV-2 epitopes having similarity with every HCoV OC43 spike epitope.

From iedbID77848, a huge cross-reactivity is observed where 392 SARS-CoV-2 epitopes share similarity with this specific HCoV OC43 spike epitope. Epitope iedbID77858, had 278 SARS-CoV-2 epitopes similar to this HCoV OC43 spike site. Additionally, iedbID84513 also points out relatively smaller although remarkable overlap where 140 SARS-CoV-2 epitopes showed relatedness Similarly, as few as 204 different types of SARS-CoV-2 epitopes can cross-react with iedbID85237, pointing to another potential immunological similarity between HCov OC43 and SARSCo V2

When it comes to iedbID85634, the evaluation recognizes that 141 SARS-CoV-2 epitopes are comparable, indicating a moderate level of cross-reactivity at this specific epitope. Further along the same line of thought is iedbID85650, which corresponds with 185 SARS-CoV-2 Nsp3 proteins suggesting another potential point at which these two coronaviruses might interact.

However, iedbID85739 has been identified as a significant cross-reactivity where 229 SARS-CoV-2 epitopes show similarity indicating large overlapping antigenic properties of spike epitopes from HCoV OC43 and SARS-CoV-2. Conversely, there are only 124 SARS-CoV-

2 epitopes showing similarities with the one found in iedbID86254. They however give us an insight into possible immunological interactions.

Thus, on this specific site alone in fact illustrated by IEDB ID86533 (iedbID86533), the latter had similarities with 220 different types of SARS-CoV-2. Again IEDB ID87237 shows 278 of them have similar structures to that of SARS-COV-2.

**Table 8: Number of SARS-COV-2 epitopes similar to each cross-reacting HCOV OC43 spike epitope**

| HCOV OC43 | Number of similar SARS-COV-2 epitopes |
|---|---|
| iedbID77848 | 392 |
| iedbID77858 | 278 |
| iedbID84513 | 140 |
| iedbID85237 | 204 |
| iedbID85634 | 141 |
| iedbID85650 | 185 |
| iedbID85739 | 229 |
| iedbID86254 | 124 |
| iedbID86533 | 220 |
| iedbID87237 | 278 |
| Total | 2191 |

### 4.3.3 Cross-Reactivity Between HCoV-NL63 and SARS-CoV-2 Spike Epitopes

#### Comparison of SARS-CoV-2 Epitopes

In the case of iedbID1656676, there is a noticeable cross-reactivity, where 171 SARS-CoV-2 epitopes show similarity to this specific HCoV NL63 spike epitope. This observation could imply that these two coronaviruses have an overlap in their immunogenicity profiles at this particular epitope site.

Table 9 explores the potential for cross-reactivity between SARS-CoV-2 epitopes and some human coronavirus NL63 spike epitopes. It also shows that each unique iedbID of a spike epitope from HCoV NL63 is matched by specific search results (SARS-CoV-2) for recognition similar to HCoV NL63 spike epitopes.

**Table 9: Number of SARS-CoV-2 epitopes similar to each crossreacting HCOV NL63 spike epitopes**

| HCOV NL63 | Number of similar SARS-COV-2 epitopes |
|---|---|
| iedbID1656676 | 171 |

### 4.3.4 Cross-Reactivity Between HCoV-HKU1 and SARS-CoV-2 Spike Epitopes

#### Comparison of SARS-CoV-2 Epitopes

A detailed analysis was conducted to assess potential cross-reactivity between spike protein epitopes of HCoV-HKU1 and those of SARS-CoV-2. The comparison was based on shared sequence similarity and immunogenic overlap, as summarized in Table 10. Each entry in the table corresponds to a unique IEDB ID representing an HCoV-HKU1 spike epitope, along with the number of SARS-CoV-2 spike epitopes exhibiting significant similarity.

The analysis revealed varying degrees of epitope cross-reactivity. Notably, IEDB ID 1386658 demonstrated high cross-reactivity, sharing similarity with 427 SARS-CoV-2 spike epitopes, suggesting extensive overlap in immunogenic properties between the two viruses. Similarly, IEDB ID 1869240 and IEDB ID 2180202 exhibited similarity with 334 and 352 SARS-CoV-2 epitopes, respectively, indicating robust shared antigenic determinants.

Moderate levels of cross-reactivity were observed in epitopes such as IEDB ID 1644870 (125 matches), IEDB ID 1673904 (127 matches), IEDB ID 1675295 (154 matches), IEDB ID

1677933 (228 matches), and IEDB ID 1695319 (221 matches). These findings support the presence of partial antigenic overlap that may influence immune recognition or memory responses.

Additionally, IEDB ID 1667205 and IEDB ID 1691286 exhibited 209 and 203 matching SARS-CoV-2 epitopes, respectively, reinforcing the possibility of conserved antigenic sites. IEDB ID 1649596 and IEDB ID 1675114 also showed modest cross-reactivity, with 149 and 146 shared epitopes, respectively.

**Table 10: Number of SARS-CoV-2 epitopes similar to each crossreacting HCOV HKU1 spike epitope**

| HCOV HKU1 | Number of similar SARS-COV-2 epitopes |
|---|---|
| iedbID1310951 | 230 |
| iedbID1386658 | 427 |
| iedbID1644870 | 125 |
| iedbID1649596 | 149 |
| iedbID1667205 | 209 |
| iedbID1673904 | 127 |
| iedbID1675295 | 154 |
| iedbID1677933 | 228 |
| iedbID1691286 | 203 |
| iedbID1695319 | 221 |
| iedbID1708895 | 243 |
| iedbID1869240 | 334 |
| iedbID2180202 | 352 |
| Total | 3002 |

### 4.3.5 Cross-Reactivity Between HCoV-229E and SARS-CoV-2 Spike Epitopes

An in-depth comparative analysis was conducted to evaluate potential cross-reactivity between spike protein epitopes of Human Coronavirus 229E (HCoV-229E) and SARS-CoV-2. The analysis focused on identifying SARS-CoV-2 epitopes that exhibit sequence similarity with known HCoV-229E spike epitopes, using epitope data obtained from the Immune Epitope Database (IEDB). The results are summarized in Table 11, which presents each HCoV-229E spike epitope by its IEDB ID and the corresponding number of SARS-CoV-2 epitopes exhibiting significant similarity.

The analysis revealed multiple instances of potential cross-reactivity. Notably, IEDB ID 1655637 demonstrated the highest degree of similarity, with 117 SARS-CoV-2 spike epitopes sharing conserved features, indicating a substantial overlap in antigenic determinants at this epitope site. Likewise, IEDB ID 1683464 and IEDB ID 1702665 exhibited high levels of similarity, with 124 and 120 SARS-CoV-2 epitopes, respectively, suggesting the presence of shared immunogenic regions that may be relevant to cross-reactive immune responses.

In total, 361 cross-reactive instances were identified, wherein various spike epitopes of HCoV-229E aligned with SARS-CoV-2 epitopes. These findings suggest a considerable degree of antigenic similarity between the spike proteins of these two alphacoronaviruses.

**Table 11: Number of SARS-CoV-2 epitopes similar to each crossreacting HCOV 229E spike epitope**

| HCOV 229E | Number of similar SARS-COV-2 epitopes |
|---|---|
| iedbID1655637 | 117 |
| iedbID1683464 | 124 |
| iedbID1702665 | 120 |

| | |
|---|---|
| Total | 361 |

### 4.3.6 Immunodominant epitopes

To assess the potential for immune cross-reactivity, a targeted analysis was performed comparing select immunodominant epitopes of SARS-CoV-2 against epitope datasets from four endemic human coronaviruses (HCoVs): HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E. The results, summarized in Table 12, outline the number of cross-reactive matches for each SARS-CoV-2 peptide and their corresponding HCoV counterparts.

The peptide "NNLDSKVGGNYNYLYRLFRK" was found to share similarity with one epitope from HCoV-OC43, suggesting a potential conserved antigenic site between the two viruses. The peptide "PTNGVGYQPYRVVVLSFELL" exhibited overlap with two epitopes from HCoV-HKU1, indicating a possible shared immunogenic region that may elicit cross-reactive responses.

For "LPKGFYAEGSRGGSQASSRS", cross-reactivity was observed with one HCoV-OC43 epitope and two from HCoV-HKU1, suggesting immunological convergence across these viruses. Similarly, "ASSRSSSRSRNSSRNSTPGS" demonstrated broader overlap, aligning with two epitopes from HCoV-OC43 and four from HCoV-HKU1, pointing to a significant potential for shared antigenicity in this region.

The peptide "GNGGDAALALLLLDRLNQLE" showed the highest level of overlap, with six matching epitopes from HCoV-OC43 and seven from HCoV-HKU1, indicating a substantial degree of conservation in this antigenic region across betacoronaviruses.

In contrast, the peptides "RRIRGGDGKMKDLSPRWYFY" and "GDQELIRQGTDYKHWPQIAQ" demonstrated overlap with one epitope each from

HCoV-NL63, highlighting isolated but notable instances of cross-reactivity with alphacoronaviruses.

**Table 12: Number of epitopes from endemic coronaviruses overlapping with each SARS-CoV-2 Immunodominant Epitope**

| Immunodominant epitope peptide | HCOV oc43 | HCOV nl63 | HCOV Hku1 | HCOV 229e |
|---|---|---|---|---|
| NNLDSKVGGNYNYLYRLFRK | 1 | 0 | 0 | 0 |
| PTNGVGYQPYRVVVLSFELL | 0 | 0 | 2 | 0 |
| LPKGFYAEGSRGGSQASSRS | 1 | 0 | 2 | 0 |
| ASSRSSSRSRNSSRNSTPGS | 2 | 0 | 4 | 0 |
| GNGGDAALALLLLDRLNQLE | 6 | 0 | 7 | 0 |
| RRIRGGDGKMKDLSPRWYFY | 1 | 0 | 0 | 0 |
| GDQELIRQGTDYKHWPQIAQ | 0 | 1 | 0 | 0 |

# CHAPTER FIVE

## 5.0 DISCUSSION

### 5.1 Genetic Diversity of SARS-CoV-2 in North Rift Kenya

#### 5.1.1 Variants of concern

The genomic analysis of 44 SARS-CoV-2 samples sequenced in this study demonstrated a clear predominance of the Omicron variant over other circulating lineages during the period under investigation. These findings are consistent with those reported in Mexico City by Cedro-Tanda et al. (2022), who noted that the Omicron variant was a major driver of increased COVID-19 case numbers at the time. The predominance of Omicron can be attributed to a combination of factors, including enhanced transmissibility, immune evasion capabilities, and decreased sensitivity to neutralizing antibodies, which collectively contribute to reduced vaccine efficacy (Cedro-Tanda et al., 2022). Among the Omicron sub-lineages detected, BA.1.1 was the most prevalent, accounting for 86.4% of the total cases. The widespread distribution of this sub-lineage is consistent with its reported selective advantage and enhanced fitness, as supported by previous studies (Cloete et al., 2022; Mohanty et al., 2022; Onishchenko et al., 2022).

A comparative evaluation of clinical and genomic data revealed notable patterns in the distribution of SARS-CoV-2 variants across symptomatic and asymptomatic individuals, disaggregated by gender. Overall, symptomatic infections were more frequently observed in female patients than in males. The BA.1.1 lineage was the predominant subvariant among symptomatic cases in both groups, representing 88% of symptomatic males and 82% of symptomatic females, suggesting its central role in driving symptomatic disease during the

period. This observation suggests that BA.1.1's has increased transmissibility and capacity to evade host immune responses.

Interestingly, further sub-lineage stratification revealed gender-specific differences in variant distribution. Symptomatic males demonstrated lower genomic diversity, with BA.1 and AY.46 each accounting for 6% of cases. Conversely, symptomatic females harbored a more diverse set of Omicron sub-lineages, including BA.1.1.1, BA.1.14, and AY.46. Notably, BA.1 was absent among symptomatic females, while BA.1.1.1 and BA.1.14 were undetected among symptomatic males. These findings suggest potential sex-based differences in susceptibility or immune responsiveness to specific SARS-CoV-2 variants, potentially modulating the distribution of circulating sub-lineages among symptomatic individuals. This contrasts with the conclusions drawn by Pongou et al. (2023), who reported no statistically significant gender-based differences in immune responses or clinical severity of COVID-19.

Statistical analysis provided additional insight into the relationships between viral clades, disease manifestation, and demographic variables. Notably, the BA.1.1 clade was significantly associated with symptomatic disease, indicating that mutations characteristic of this lineage may contribute to increased pathogenicity. Gender-based comparisons revealed a slightly higher odds ratio for symptomatic infection among males; however, this association did not reach statistical significance, suggesting a limited role of sex in determining disease expression.

To validate these associations and account for potential confounding factors, the Cochran-Mantel-Haenszel (CMH) test was employed. The results affirmed that the association

between BA.1.1 and symptomatic disease remained significant after adjusting for gender, underscoring the independent effect of viral genetic factors on clinical outcomes. These findings provide valuable insights into the interplay between viral genetics and host clinical presentation, reinforcing the critical role of genomics in informing evidence-based public health strategies during the COVID-19 pandemic.

The analysis of mutations across a wide array of viral genes from five SARS-CoV-2 clades further highlighted the genetic plasticity and evolutionary adaptability of the virus. The current study, conducted during the fifth wave, identified clade BA.1.1 as the dominant variant in circulation. Previous studies have demonstrated that these clades accumulated lineage-defining mutations early in the pandemic and subsequently disseminated across various geographical regions, thus contributing to the evolving epidemiology of COVID-19 (Patro et al., 2021). In the initial pandemic wave, the B.1 global parental lineage predominated, accounting for 94% of sequenced genomes. During the second wave, its prevalence declined to 71%, coinciding with the emergence of diverse lineages, including Variants of Concern (VoCs) such as B.1.351 (Beta) and B.1.1.7 (Alpha), as well as non-VoCs such as B.1.3x, B.1.5x, B.1.525 (Eta), and A.23x. By the mid-point of the third wave, Alpha had become the predominant lineage, representing 74.9% of sequenced genomes. The Delta variant (B.1.617.2) and its AY sub-lineages were first reported in March 2021 and rapidly displaced previous strains, achieving dominance during the fourth wave with a prevalence of 99.3% (Nasimiyu, Matoke-Muhia, Rono, Osoro, Ouso, et al., 2022).

The comprehensive genomic analysis included a wide range of structural and non-structural genes, such as 3'UTR, 5'UTR, E, M, N, NSP1-NSP16, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and S. Through the interrogation of these genomic regions, this study elucidated the

roles of viral proteins in modulating host-pathogen interactions. Coronaviruses, including SARS-CoV-2, are known to exploit host cellular mechanisms such as autophagy to facilitate replication and propagation (Shroff & Nazarko, 2021). The analysis also revealed evidence of epistasis, wherein specific combinations of mutations interact to influence viral fitness and pathogenicity (H. Zeng et al., 2020). Furthermore, lineage-specific mutational signatures were observed across different clades, underscoring the dynamic and multifactorial nature of SARS-CoV-2 evolution. These findings underscore the importance of continued genomic surveillance and mutational profiling in understanding viral adaptation and its implications for diagnostics, therapeutics, and public health. Detailed knowledge of these mutations not only informs the design of variant-specific interventions but also enhances our preparedness for future waves of infection driven by novel variants.

### 5.1.2 Mutated bases per gene for each Clade

A notable mutation observed in this study was the cytosine-to-thymine (C>T) transition at position 241 in the 5′ untranslated region (UTR) of the SARS-CoV-2 genome. This mutation was detected in the BA.1.1 variant at a frequency of 34.2%, a finding that is consistent with observations reported by Yunlong Cao et al. (2022). The 5′ UTR, although non-coding, plays a critical regulatory role in the viral life cycle, including processes such as RNA replication, transcription, translation, and genome packaging. Variations within this region, particularly at position 241, have been documented in several genomic studies and have been proposed to contribute to the phenotypic differences observed among SARS-CoV-2 lineages (Ullah et al., 2021).

Experimental evidence has shown that mutations in the 5′ UTR can negatively affect viral RNA synthesis by disrupting RNA structural elements or altering interactions with host and viral proteins (Luo *et al*., 2021). Despite its functional significance, the evolutionary dynamics and host-interacting mechanisms of the 5′ UTR remain incompletely understood (Mohammadi-Dehcheshmeh *et al*., 2021). Nevertheless, studies have emphasized its role in RNA–protein interactions and viral replication fidelity, suggesting that the region could serve as a potential target for antiviral interventions (Verma *et al*., 2021).

Furthermore, the 5′ UTR is implicated in modulating the translation efficiency and stability of viral RNAs, processes that are essential for efficient viral replication and protein synthesis (Raheja et al., 2023; M. Shi et al., 2020). Mutations in this region have been consistently linked to changes in viral fitness, adaptability, and virulence, highlighting the 5′ UTR as a hotspot for evolutionary adaptation (Alam *et al*., 2020). Therefore, the detection of the C>T mutation at position 241 in the BA.1.1 lineage reinforces the significance of this regulatory region in SARS-CoV-2 pathobiology.

The genomic investigation of the SARS-CoV-2 Omicron subvariant BA.1.1 revealed consistent mutations at several positions within the non-structural protein 3 (NSP3) coding region, specifically at NSP3:F106F, NSP3:A889A, NSP3:S1265, and NSP3:A1892T. These sites exhibited a mutation prevalence of 100.0%, indicating fixed or near-fixed substitutions within the BA.1.1 population. This high conservation of mutations is consistent with findings by Singh et al. (2022), who reported similar patterns of nucleotide variation in Omicron lineages, emphasizing the dynamic nature of the viral genome and its continuous evolution.

In addition to NSP3, the non-structural protein 4 (NSP4) region demonstrated a notable mutation at reference position 10029, resulting in the T492I substitution. This mutation, present in 100.0% of the BA.1.1 sequences analyzed, is consistent with the findings of Lin et al. (2023), who proposed that such high-frequency mutations may play a pivotal role in viral evolution and represent attractive targets for the development of broad-spectrum antiviral agents.

The NSP6 region also exhibited significant alterations, including deletions at positions L105 and I189V, both with a 100.0% mutation rate in BA.1.1 genomes. These deletions have been previously associated with modifications in the formation of autophagosomal vesicles, potentially aiding the virus in evading host immune surveillance mechanisms (Lin *et al*., 2023). Together, the mutations in NSP4 and NSP6 reflect selective pressures that may favor viral phenotypes with enhanced replication or immune modulation capabilities.

In conclusion, the recurrent and conserved mutations observed in the NSP3, NSP4, and NSP6 regions of the SARS-CoV-2 Omicron variant BA.1.1 highlight the virus's ongoing adaptive evolution. These non-structural proteins play critical roles in replication, pathogenesis, and immune evasion. As such, their mutation profiles not only reflect evolutionary pressures but also present potential targets for next-generation antiviral development. Future research should prioritize functional characterization of these mutations to elucidate their mechanistic contributions to viral fitness and disease outcomes, ultimately guiding the development of effective therapeutic and preventive measures.

### 5.1.3 Significance of various mutations

The present study identified a total of 65 distinct mutations in the SARS-CoV-2 BA.1.1 variant, spanning both structural and non-structural proteins, with frequencies ranging from 6.8% to 100.0%. Among the 42 non-synonymous mutations, seven mutations: NSP3:I388L; S:E484A; S:G446S; S:G496S; S:N211K; S:N440K; and S:N501Y were uniquely identified in symptomatic patients. All four insertions: S:I210; S:R214; S:S214; and S:V213, were present only in symptomatic patients. A majority of these mutations were localized within the spike (S) glycoprotein, echoing patterns reported in previous studies (Suresh Kumar et al., 2022; Viana et al., 2022), thereby reinforcing the critical role of this protein in viral adaptation. However, unlike earlier studies that predominantly focused on a subset of spike mutations, the current investigation extends its scope to incorporate mutations across NSP2, NSP3, NSP4, NSP6, NSP12b, and NSP14, offering a broader genomic context.

The mutation S:D614G was found at a fixed frequency of 100%, consistent with global findings that have established it as a dominant mutation since early 2020 (Ogawa et al., 2020; Verkhivker, Agajanian, Kassab, et al., 2021; Verkhivker, Agajanian, Oztas, et al., 2021). The extensive literature agrees that D614G enhances infectivity through increased spike stability, furin cleavage, and higher virion spike density. This concordance strengthens the confidence in its evolutionary advantage and provides a solid benchmark for comparing newer spike mutations.

Similarly, high-frequency mutations such as S:P681H (95.5%) and S:N501Y mirror findings from previous studies where these mutations were linked to enhanced furin cleavage potential

and increased ACE2 receptor binding, respectively (Gu et al., 2020; Lubinski et al., 2022). The consistency across studies suggests robust selective pressure favoring these mutations.

In contrast, the mutation S:E484A was observed in 42.1% of cases, which is notably lower than frequencies reported in some earlier Omicron datasets (e.g., >70%) (Saranya et al., 2022). This may reflect geographic or temporal differences in variant evolution or founder effects specific to the sampled population. Moreover, while S:E484A has been consistently implicated in antibody escape, its frequency variability may affect population-level susceptibility to neutralizing antibodies.

Particularly, the study identified S:R214R, a synonymous mutation present in 43.2% of sequences, which evolved from an earlier insertion mutation (S:R214). While synonymous mutations were historically considered neutral, recent research supports their role in altering RNA structure and translation kinetics (Renn, 2023). This study adds to that body of knowledge by suggesting that synonymous mutations may undergo positive selection, although functional validation remains a gap.

Unlike many prior studies focused primarily on spike protein evolution, this study extensively profiled mutations in NSPs. Notably, NSP3 mutations such as F106F, A889A, S1265, and A1892T were present at 100%, in agreement with findings by Singh et al. (2022). These mutations are located in regions essential for viral replication and immune modulation, including the Mac1 domain. Taha et al. (2023) demonstrated that alterations in Mac1 significantly affect viral replication and pathogenesis, supporting the biological relevance of these fixed NSP3 mutations.

In NSP4, the T492I mutation also reached fixation (100%), consistent with Lin et al. (2023), who proposed its involvement in membrane rearrangements necessary for replication. Deletions in NSP6 at positions L105 and I189V also mirrored those found in early Omicron studies, suggesting a conserved adaptation mechanism across sub-lineages (Lin *et al*., 2023).

This broader genomic scope is a strength of the present study, allowing for a more holistic view of viral evolution. However, a key limitation is the lack of functional validation experiments, which constrains the ability to definitively assign phenotypic outcomes to observed mutations.

The chi-square test revealed a significant association between mutation frequency categories and protein type ($\chi^2(6) = 18.45$, $p = 0.005$), indicating that some proteins—particularly Spike and NSP3—harbor disproportionately high-frequency mutations. This result aligns with their functional roles in host interaction and immune evasion, confirming earlier assertions of intense positive selection acting on these regions.

A logistic regression model further supported this observation, demonstrating that mutations in the Spike protein were 3.42 times more likely to occur at high frequency than those in other proteins (95% CI: 1.12–10.34, $p = 0.031$). The non-significant association between mutation frequency and genomic position ($p = 0.182$) reinforces that protein function, rather than location, governs evolutionary constraint.

These statistical findings echo previous large-scale analyses of SARS-CoV-2 diversity (Bloom & Neher, 2023), but add new granularity by integrating non-structural protein data. However, one statistical limitation is the modest sample size (n=44), which may limit generalizability and the power of post-hoc comparisons.

The co-occurrence network revealed tightly linked clusters of spike mutations, particularly D614G, P681H, and N501Y. These interactions suggest either compensatory mechanisms or shared selective pressures. Previous studies have proposed that co-evolving mutations may collectively enhance viral fitness more than single substitutions (Gellenoncourt *et al*., 2022). This study substantiates those hypotheses by demonstrating that these mutations frequently occur together in the same variant genomes.

The statistical strength of these associations is reflected in the dense connectivity of the network, although further computational modeling (e.g., Bayesian epistasis analysis) would be needed to quantify their interdependence. Nonetheless, the network's structural coherence supports the biological plausibility of synergistic interactions, particularly those affecting viral entry and immune evasion.

## 5.2 Immunoinformatic Profiling of SARS-CoV-2 Variants in North Rift Kenya

The identification and characterization of SARS-CoV-2 epitopes are essential for understanding immune responses to COVID-19 and for advancing vaccine and diagnostic development. In this study, 12,285 epitopes from the Wuhan SARS-CoV-2 reference genome were examined and 5,154 (42%) conserved epitopes in viral genomes isolated from Moi Teaching and Referral Hospital (MTRH) in Kenya. These conserved epitopes were predominantly found in the spike (S), nucleocapsid (N), and membrane (M) proteins, which are critical for viral pathogenesis and immune recognition.

The results of this study align with prior research emphasizing the significance of conserved epitopes in these structural proteins. For example, Patel et al. (2023) showed that lipopeptides from conserved S and N protein epitopes can induce cross-protective immunity in mice,

highlighting their potential for vaccine design (R. K. Patel & Agrawal, 2023). Similarly, Wang et al. (2023) identified conserved epitopes in the S1 and S2 subunits of the spike protein, reinforcing their relevance for universal vaccine strategies (C. Y. Wang et al., 2023).

Recent studies have also highlighted the conservation of epitopes across major SARS-CoV-2 variants, such as Delta and Omicron. Rodrigues-da-Silva et al. (2023) identified stable B-cell epitopes in the nucleocapsid protein, which remain consistent despite genetic variability (Rodrigues-da-Silva et al., 2023). Tarigan et al. (2023) characterized linear epitopes in the spike protein conserved across multiple variants, suggesting their utility in diagnostic applications (Tarigan et al., 2023).

A key finding of this study is the conservation of epitopes in structural proteins, particularly the spike (S) and nucleocapsid (N) proteins. The spike protein, a primary target for neutralizing antibodies, contained the majority of conserved B-cell epitopes (844 out of 2,603). This aligns with earlier research emphasizing the spike protein's importance in vaccine development due to its role in viral entry and immunodominant properties (Grifoni et al., 2020b; Walls et al., 2020). Similarly, the nucleocapsid protein, involved in viral replication and assembly, also exhibited significant epitope conservation (405 out of 784 B-cell epitopes). These findings are consistent with studies showing that conserved epitopes in these proteins can trigger strong immune responses, even against emerging variants (Tarke et al., 2021).

Recent advancements in phage-based vaccines, such as those targeting the spike protein, offer promising opportunities to leverage conserved epitopes. Research by Markosian et al. (2022) and Staquicini et al. (2021) has demonstrated the effectiveness of phage-displayed

epitopes in eliciting robust immune responses against SARS-CoV-2 variants (Markosian et al., 2022; Staquicini et al., 2021). The findings of this research support these efforts by identifying conserved regions suitable for incorporation into phage-based vaccine platforms.

The conservation of T-cell epitopes, mostly MHC class II epitopes (693 out of 1,362), further highlights the potential for cross-reactive immunity. T-cell responses are vital for viral clearance and long-term immunity, and the identification of conserved T-cell epitopes aids in developing vaccines that induce broad and durable immune responses (Sette & Crotty, 2021). In contrast, MHC class I epitopes were less conserved, with a notable reduction in the MTRH genomes (374 out of 2,326). This may reflect selective pressures from host immune responses or genetic drift, as MHC class I epitopes are crucial for cytotoxic T-cell activity (Sette & Crotty, 2021). The lower conservation of these epitopes could impact cross-protective immunity, as mutations in these regions may facilitate immune evasion.

While this study provides valuable insights into epitope conservation, it has limitations. First, the analysis relies on in silico predictions and database-derived epitopes, which may not fully capture their immunogenicity in vivo. Experimental validation, including peptide binding assays and functional studies, is necessary to confirm the immunogenicity and protective potential of the identified epitopes. Second, the study is based on a limited dataset of 44 genomes from Kenya, collected in December 2021, which may restrict the generalizability of the findings. SARS-CoV-2 exhibits regional variability due to differences in transmission dynamics, host immunity, and viral evolution (Rambaut et al., 2021). Future studies should expand the geographic and temporal scope to capture a broader spectrum of viral diversity.

Despite these limitations, the findings have significant implications for vaccine development and immune surveillance. The identification of conserved epitopes in the spike and nucleocapsid proteins supports the design of multiepitope vaccines capable of providing broad protection against multiple variants (Tregoning et al., 2021). The use of bioinformatics tools, such as machine learning, to predict epitope reactivity across variants further enhances the potential for tailored vaccine strategies (P. Wang et al., 2021). Additionally, the conservation of key epitopes in the MTRH genomes suggests that vaccines targeting these regions may be effective in Kenya and other regions with similar viral lineages.

## 5.3 Immune overlap of the SARS-CoV-2 variants circulating in North Rift Kenya and the endemic human coronaviruses

### 5.3.1 Overlapping epitopes

The North Rift Kenya isolated SARS-CoV-2 virus genomes displayed cross-reactivity with the epitopes originating from endemic coronaviruses, that is HCoV-HKU1, HCoV-OC43, HCoV-NL63 and HCoV-229E. This observed phenomenon is indicative of the possibility of immune interactions between these endemic coronavirus strains and SARS-CoV-2 (Sagar *et al*., 2021). There have been several studies to show that pre-existing T-cell responses to SARS-CoV-2 virus exist in individuals who had not been previously exposed to it which suggests that there may be some degree of cross-reactivity between common cold coronaviruses and SARS-CoV-2 virus (Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, et al., 2020). The shared sequence homology seen between the endemic coronaviruses and the novel coronavirus might elicit immune responses with cross-reacting antigens against the antigenic determinants present in the SARS-CoV-2

virus (Nelde, Bilich, Heitmann, Maringer, Salih, Roerden, Lübke, Bauer, Rieth, Wacker, Peter, Hörber, Traenkle, Kaiser, Rothbauer, Becker, Junker, Krause, Strengert, Schneiderhan-Marra, et al., 2020). Furthermore, conserved epitopes found in both HCoV-OC43and HCoV-HKU1 are another piece of evidence for possible cross-reaction with the novel corona virus (Piccaluga *et al*., 2021). Research has identified potential T-cell epitopes which have shared reactivity by structural proteins from endemic coronaviruses specific to antibodies produced against them thus forming a platform where immune interactions take place (Johansson *et al*., 2021). It also worth noting that cross reactive serum antibodies were detected in individuals donating before COVID-19 suggesting pre-existent cross reactivity between these endemic corona viruses and SARS-Cov 2 (Song *et al*., 2021). Notably, conserved epitopes within spike region among SARS-CoV-2 virus, bat SL-CoV and SARS-CoV signify potential of cross-reactivity and ensuing immune interactions (Baloch *et al*., 2022). The identification of immunodominant and cross-reactive epitopes is important in terms of guiding the development of vaccines that can be used to control emerging SARS-CoV-2 variants (Low *et al*., 2021). Some studies have shown that individuals who were previously infected with or vaccinated against SARS-CoV-2 virus produce antibodies that show cross reactivity with related betacoronaviruses, indicating possible immune interactions (Geanes et al., 2022). Moreover, highly conserved T-cell epitopes shared by the SARS-CoV-2 virus and other human coronaviruses indicate the likelihood for cross-reactivity and subsequent immune responses. To sum up, shared epitopes, as well as sequence homology between endemic coronavirus strains, such as SARS-CoV-2 virus may result in cross-reacting immune interactions. In order to develop effective immunization strategies and comprehend host responses to SARS-CoV-2 fully it is very necessary to understand these interplays.

Epitopes obtained from different structural proteins of SARS-CoV-2, including those derived from the spike (S) and nucleocapsid (N), consist of various B cell epitopes that are recognized by AY.46 clade of SARS-CoV-2 (S. F. Ahmed et al., 2020). Also, it has been found that some of these B-cell epitopes were highly conserved among those identified in SARS-CoV-1 implying potential cross-reactivity as well as shared immunogenicity between these viruses (H. X. Lim et al., 2021).

These predicted epitopes are non-allergenic and have high proteasomal cleavage and TAP transport efficiency with 100% conservancy among different clades of COVID-19 virus within the AY.46 clade (Khairkhah et al., 2020). This is important for vaccines to target many strains of the virus simultaneously. It is also significant to predict bioinformatically using structural sequences of SARS-CoV-2 since it helps understand antigen–antibody relationship (Jiang et al., 2021).

Besides, analysis of T cell epitopes is critical because it aids in understanding cellular immunity and designing peptide-based diagnostics, therapeutics, and vaccines (Jensen et al., 2018). In addition, B-cell and T-cell epitope mapping on SARSCoV-2 including those found in unannotated ORF antigens can be exploited to design polyvalent epitope-based vaccines against this virus (Uttamrao et al., 2021). The most critical aspect about the knowledge regarding b-cell epitopes and MHC-I binding regions in structural proteins of coronavirus is because they will aid in developing effective diagnostics as well as therapies for SARS-CoV-2 infections (Tu et al., 2017).

To sum up, comprehensive identification and characterization of B cell and T cell epitopes within the AY.46 clade of SARS-CoV-2 offers valuable insights into vaccine design,

immunotherapy development, and immune response to the virus. Therefore, these epitopes are instrumental in fighting COVID-19 using targeted and effective strategies.

# CHAPTER SIX

## 6.0 CONCLUSION AND RECOMMENDATION

### 6.1 Conclusion

From this study, the BA.1.1 clade was identified as the most prevalent in the region, with other clades such as BA.1, BA.1.1.1, BA.1.14, and AY.46 also detected. Outstandingly, gender-based differences in the distribution of variants among symptomatic and asymptomatic cases were observed, highlighting the need for further research to uncover the biological, immunological, and sociological factors that may contribute to these disparities.

A total of 65 mutations, both synonymous and non-synonymous, were identified in symptomatic and asymptomatic patients. Among these, the non-synonymous mutations G446S, E484A, G496S, N440K, and N501Y in the spike protein were found exclusively in symptomatic patients.

Additionally, two new synonymous mutations, S:R214R and NSP2:A555A , were detected. Although these mutations do not alter the amino acid sequence, they were predominantly observed in symptomatic patients.

This study further identified 5,154 conserved epitopes from the Wuhan SARS-CoV-2 genome in the genomes isolated from Kenya. These epitopes, predominantly located in the M, N, and S proteins, represent promising targets for the development of vaccines and diagnostic tools. However, the study has certain limitations, such as the absence of experimental validation and the restricted geographic scope of the dataset, which emphasize the need for further investigation.

Through bioinformatics tools and laboratory techniques, this study provides essential findings on genetic mutations and immune cross reactivity, necessary for vaccine development as well as understanding of cross-immunity among coronaviruses.

The findings presented highlight significant evidence of cross-reactivity between SARS-CoV-2 and endemic human coronaviruses such as HCoV-HKU1, HCoV-OC43, HCoV-NL63, and HCoV-229E. Sequence homology and epitope conservation across these viruses suggest a potential for pre-existing immune responses in individuals previously exposed to common cold coronaviruses, which may contribute to variability in SARS-CoV-2 susceptibility and disease outcomes. The detection of conserved B-cell and T-cell epitopes, particularly within the spike and nucleocapsid proteins, supports the notion of shared immunogenic regions that could mediate cross-protective immunity.

## 6.2 Recommendations

**Objective 1: Determine Genetic Variations of SARS-CoV-2 in North Rift Kenya**

1. Expand genomic surveillance across counties and over time to track emerging variants.
2. Prioritize functional studies of mutations (e.g., G446S, E484A, N501Y) linked to symptomatic cases.
3. Investigate potential regulatory effects of synonymous mutations like S:R214R and NSP2:A555A.

**Objective 2: Predict Epitopes of SARS-CoV-2 in Circulating Variants**

1. Experimentally validate predicted B- and T-cell epitopes using immunological assays.
2. Incorporate conserved epitopes into multivalent or pan-variant vaccine designs.

**Objective 3: Investigate Immune Overlap with Endemic Human Coronaviruses**

1. Confirm bioinformatic cross-reactivity using local pre- and post-pandemic serum samples.
2. Focus on conserved spike and nucleocapsid epitopes shared with HCoVs for cross-protective vaccine targets.
3. Support longitudinal studies on the role of pre-existing immunity in SARS-CoV-2 susceptibility and response.

# REFERENCES

## 7.0 References

Adedeji, A. O., Marchand, B., Velthuis, A. J. t., Snijder, E. J., Weiss, S. R., Eoff, R. L., Singh, K., & Sarafianos, S. G. (2012). Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. *Plos One*. https://doi.org/10.1371/journal.pone.0036521

Ahmadi, E., Zabihi, M. R., Hosseinzadeh, R., Khosroshahi, L. M., & Noorbakhsh, F. (2021). SARS-CoV-2 Spike Protein Displays Sequence Similarities With Paramyxovirus Surface Proteins; A Bioinformatics Study. *Plos One*. https://doi.org/10.1371/journal.pone.0260360

Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses*. https://doi.org/10.3390/v12030254

Ahmed, W. S., Philip, A. M., & Biswas, K. H. (2022). Decreased Interfacial Dynamics Caused by the N501Y Mutation in the SARS-CoV-2 S1 Spike:ACE2 Complex. *Frontiers in Molecular Biosciences*, *9*. https://doi.org/10.3389/fmolb.2022.846996

Ahsan, A., Dar, S., Hassan, F., Ghafoor, F., Yousuf, M., & Shahzad-ul-Hussan, S. (2021). Characterization of Linear Epitope Specificity of Antibodies Potentially Contributing to Spontaneous Clearance of Hepatitis C Virus. *Plos One*. https://doi.org/10.1371/journal.pone.0256816

Al-Khannaq, M. N., Ng, K. T., Oong, X. Y., Pang, Y. K., Takebe, Y., Chook, J. B., Hanafi, N. S., Kamarulzaman, A., & Tee, K. K. (2016). Diversity and Evolutionary Histories of Human Coronaviruses NL63 and 229E Associated With Acute Upper Respiratory Tract Symptoms in Kuala Lumpur, Malaysia. *American Journal of Tropical Medicine and Hygiene*. https://doi.org/10.4269/ajtmh.15-0810

Alam, A., Islam, O. K., Hasan, M. S., Al-Emran, H. M., Jahid, I. K., & Hossain, M. A. (2020). *Evolving Infection Paradox of SARS-CoV-2: Fitness Costs Virulence?* https://doi.org/10.31219/osf.io/y36ve

Albukhari, T. A., Radford, P. M., Bouras, G., Davenport, C., Trigwell, S., Bottazzo, G. F., Lai, M., Schwartz, H. L., Tighe, P. J., & Todd, I. (2002). Distinct Antigenic Features

of Linear Epitopes at the N-Terminus and C-Terminus of 65 kDa Glutamic Acid Decarboxylase (GAD65): Implications for Autoantigen Modification During Pathogenesis. *Clinical & Experimental Immunology*. https://doi.org/10.1046/j.1365-2249.2002.01960.x

Algaissi, A., Alharbi, N. S., Hassanain, M., & Hashem, A. M. (2020). *Preparedness and Response to COVID-19 in Saudi Arabia: Lessons Learned From MERS-CoV ≪!--EndFragment--≫* https://doi.org/10.20944/preprints202004.0018.v1

AlHamid, G., Tombuloglu, H., Motabagani, D., Motabagani, D., Rabaan, A. A., Unver, K., Dorado, G., Al-Suhaimi, E. A., & Unver, T. (2022). Colorimetric and fluorometric reverse transcription loop-mediated isothermal amplification (RT-LAMP) assay for diagnosis of SARS-CoV-2. *Functional & Integrative Genomics*, *22*(6), 1391–1401. https://doi.org/10.1007/s10142-022-00900-5

Almansour, I., & Boudellioua, I. (2022). hCoronavirusesDB: An Integrated Bioinformatics Resource for Human Coronaviruses. *Database*. https://doi.org/10.1093/database/baac017

Alpert, T., Vogels, C. B., Breban, M. I., Petrone, M. E., Wyllie, A. L., Grubaugh, N. D., & Fauver, J. R. (2021). *Sequencing SARS-CoV-2 Genomes From Saliva*. https://doi.org/10.1101/2021.06.21.21259289

Amrun, S. N., Lee, C. Y.-P., Lee, B., Fong, S.-W., Fong, S.-W., Young, B. E., Chee, R. S.-L., Yeo, N. K.-W., Torres-Ruesta, A., Carissimo, G., Poh, C. M., Chang, Z. W., Tay, M. Z., Chan, Y.-H., Chen, M. I.-C., Tambyah, P. A., Kalimuddin, S., Kalimuddin, S., Pada, S., … Ng, L. F. P. (2020). Linear B-Cell Epitopes in the Spike and Nucleocapsid Proteins as Markers of SARS-CoV-2 Exposure and Disease Severity. *Social Science Research Network*. https://doi.org/10.2139/SSRN.3605114

Angeli, F., Reboldi, G., & Verdecchia, P. (2021). SARS-CoV-2 Infection and ACE2 Inhibition. *Journal of Hypertension*. https://doi.org/10.1097/hjh.0000000000002859

Angelini, M. M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. (2013). Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *Mbio*. https://doi.org/10.1128/mbio.00524-13

Anirudhan, V., Lee, H., Cheng, H., Cooper, L., & Rong, L. (2021). Targeting SARS-CoV-2 Viral Proteases as a Therapeutic Strategy to Treat COVID-19. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.26814

Antonides, L. H., Hurst, Q. W., Ives, C. M., Ramberg, K. O., Ostrovitsa, N., Scanlan, E. M., Caffrey, M., Pitt, S. J., & Zachariae, U. (2022). The SARS-CoV-2 envelope (E) protein forms a calcium- and voltage-activated calcium channel. *BioRxiv*. https://doi.org/10.1101/2022.10.11.511775

Anwar, M. Z., Lodhi, M. S., Khan, M. T., Khan, M. I., & Sharif, S. (2022). Coronavirus Genomes and Unique Mutations in Structural and Non-Structural Proteins in Pakistani SARS-CoV-2 Delta Variants during the Fourth Wave of the Pandemic. *Genes*, *13*(3), 552. https://doi.org/10.3390/genes13030552

Ao, Z., Ouyang, M. J., Olukitibi, T., & Yao, X. (2021). SARS-CoV-2 Delta spike protein enhances the viral fusogenicity and inflammatory cytokine production. *IScience*, *25*, null. https://doi.org/10.1016/j.isci.2022.104759

Aoki, K., Fukaya, S., Takahashi, H., Kobayashi, M., Sasaki, K., & Takemura, M. (2021). Marseilleviridae Lineage B Diversity and Bunch Formation Inhibited by Galactose. *Microbes and Environments*, *36*(1), 1–6. https://doi.org/10.1264/JSME2.ME20139

Appenzeller-Herzog, C., & Hauri, H. (2006). The ER-Golgi Intermediate Compartment (ERGIC): In Search of Its Identity and Function. *Journal of Cell Science*. https://doi.org/10.1242/jcs.03019

Arden, K. E., Nissen, M. D., Sloots, T. P., & Mackay, I. M. (2005). New Human Coronavirus, HCoV-NL63, Associated With Severe Lower Respiratory Tract Disease in Australia. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.20288

Arévalo, M. T., Karavis, M. A., Katoski, S. E., Harris, J. V, Hill, J. M., Deshpande, S. V, Roth, P. A., Liem, A. T., & Bernhards, R. C. (2022). A Rapid, Whole Genome Sequencing Assay for Detection and Characterization of Novel Coronavirus (SARS-CoV-2) Clinical Specimens Using Nanopore Sequencing. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2022.910955

Arnaldo, P., Mabunda, N., Young, P., Tran, T. D., Sitoe, N., Chelene, I., Nhanombe, A.,

Isamael, N., Júnior, A., Cubula, B., Inlamea, O. F., Gudo, E. S., & Jani, I. V. (2022). Prevalence of SARS-CoV-2 antibodies in the Mozambican population: a cross-sectional Serologic study in three cities, July-August 2020. *Clinical Infectious Diseases*, *75*(Supplement_2), S285–S293. https://doi.org/10.1093/cid/ciac516

Arteaga, F. L., Miragaya, M., Molina, N., Mondino, M., Bracamonte, C., Capitelli, G., Mundo, S. L., Torres, C., & Bratanich, A. (2023). Identification of Coronaviruses in Bats and Rodents in Northern and Central Argentina. *Archives of Virology*. https://doi.org/10.1007/s00705-023-05703-y

Azami, M., Moradi, Y., Moradkhani, A., & Aghaei, A. (2022). SARS-CoV-2 seroprevalence around the world: an updated systematic review and meta-analysis. *European Journal of Medical Research*, *27*(1). https://doi.org/10.1186/s40001-022-00710-2

Bader, W., Delerce, J., Aherfi, S., Scola, B. L., & Colson, P. (2022). Quasispecies Analysis of SARS-CoV-2 of 15 Different Lineages During the First Year of the Pandemic Prompts Scratching Under the Surface of Consensus Genome Sequences. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms232415658

Baker, D., Aydin, A., Le-Viet, T., Kay, G. L., Rudder, S., Martins, L. d. O., Tedim, A. P., Kolyva, A., Díaz, M., Alikhan, N. F., Meadows, L., Bell, A., Gutiérrez, A. V, Trotter, A. J., Thomson, N., Gilroy, R., Griffith, L., Adriaenssens, E. M., Stanley, R., … O'Grady, J. (2021). CoronaHiT: High-Throughput Sequencing of SARS-CoV-2 Genomes. *Genome Medicine*. https://doi.org/10.1186/s13073-021-00839-5

Baloch, Z., Ikram, A., Shamim, S., Obaid, A., Awan, F. M., Naz, A., Rauff, B., Gilani, K., & Qureshi, J. A. (2022). Human Coronavirus Spike Protein Based Multi-Epitope Vaccine Against COVID-19 and Potential Future Zoonotic Coronaviruses by Using Immunoinformatic Approaches. *Vaccines*. https://doi.org/10.3390/vaccines10071150

Banerjee, S., Wang, X., Du, S., Zhu, C., Jia, Y., Wang, Y., & Cai, Q. (2022). Comprehensive role of SARS-CoV-2 spike glycoprotein in regulating host signaling pathway. *Journal of Medical Virology*, *94*(9), 4071–4087. https://doi.org/10.1002/jmv.27820

Bangaru, S., Antanasijevic, A., Kose, N., Sewall, L. M., Jackson, A. M., Suryadevara, N.

C., Zhan, X., Torres, J. L., Copps, J., Peña, A. T. de la, Crowe, J. E., & Ward, A. B. (2022). Structural Mapping of Antibody Landscapes to Human Betacoronavirus Spike Proteins. *Science Advances*. https://doi.org/10.1126/sciadv.abn2911

Barman, N., De, A., Paul, J., Haldar, S., Bhattacharya, A. R., & Pal, K. (2022). Strategy to Configure Multi-Epitope Recombinant Immunogens With Weightage on Proinflamatory Response Using SARS-CoV-2 Spike Glycoprotein (S-Protein) and RNA-dependent RNA Polymerase (RdRp) as Model Targets. *Journal of Pure and Applied Microbiology*. https://doi.org/10.22207/jpam.16.1.17

Barr, B. A. T., Herman-Roloff, A., Mburu, M., Murnane, P. M., Sang, N., Bukusi, E. A., Oele, E. A., Odhiambo, A. O., Lewis-Kulzer, J., Onyango, C., Hunsperger, E., Odhiambo, F., Joseph, R. H., Munyua, P., Othieno, K., Mulwa, E., Akelo, V., Muok, E. M. O., Bulterys, M., … Cohen, C. R. (2022). First cases of SARS-CoV-2 infection and secondary transmission in Kisumu, Kenya. *PLOS Global Public Health*, *2*(9), e0000951–e0000951. https://doi.org/10.1371/journal.pgph.0000951

Baruah, V., & Bose, S. (2020). Immunoinformatics-aided Identification of T Cell and B Cell Epitopes in the Surface Glycoprotein of 2019-nCoV. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25698

Basu, A., Sarkar, A., & Maulik, U. (2020). Molecular Docking Study of Potential Phytochemicals and Their Effects on the Complex of SARS-CoV2 Spike Protein and Human ACE2. *Scientific Reports*. https://doi.org/10.1038/s41598-020-74715-4

Becerra-Artiles, A., Nanaware, P. P., Muneeruddin, K., Weaver, G. C., Shaffer, S. A., Calvo-Calle, J. M., & Stern, L. J. (2023). Immunopeptidome profiling of human coronavirus OC43-infected cells identifies CD4 T-cell epitopes specific to seasonal coronaviruses or cross-reactive with SARS-CoV-2. *PLoS Pathogens*, *19*(7 July). https://doi.org/10.1371/journal.ppat.1011032

Bergner, L. M., Orton, R., Benavides, J. A., Becker, D. J., Tello, C., Biek, R., & Streicker, D. G. (2019). Demographic and Environmental Drivers of Metagenomic Viral Diversity in Vampire Bats. *Molecular Ecology*. https://doi.org/10.1111/mec.15250

Bertram, S., Glowacka, I., Muller, M. A., Lavender, H., Gnirss, K., Nehlmeier, I., Niemeyer, D., He, Y., Simmons, G., Drosten, C., Soilleux, E. J., Jahn, O., Steffen, I.,

& Pohlmann, S. (2011). Cleavage and Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by Human Airway Trypsin-Like Protease. *Journal of Virology*. https://doi.org/10.1128/jvi.05300-11

Bhella, D. (2015). The Role of Cellular Adhesion Molecules in Virus Attachment and Entry. *Philosophical Transactions of the Royal Society B Biological Sciences*. https://doi.org/10.1098/rstb.2014.0035

Bistaraki, A., Roussos, S., Tsiodras, S., & Sypsa, V. (2021). Age-dependent effects on infectivity and susceptibility to SARS-CoV-2 infection: results from nationwide contact tracing data in Greece. *Infectious Diseases*, 1–10. https://doi.org/10.1080/23744235.2021.1995627

Biswal, J. K., Ranjan, R., Dahiya, S. S., Mallick, S., & Mohapatra, J. K. (2021). Regenerated Silica-Based RNA Purification Columns to Address the Short Supply of RNA Purification Kits for COVID-19 Diagnosis. *Molecular Biology Reports*. https://doi.org/10.1007/s11033-021-06688-0

Bloom, J. D., & Neher, R. A. (2023). Fitness effects of mutations to SARS-CoV-2 proteins. *BioRxiv*. https://doi.org/10.1101/2023.01.30.526314

Bolotin, S., Tran, V., Osman, S., Brown, K. A., Buchan, S. A., Joh, E., Deeks, S. L., & Allen, V. (2021). SARS-CoV-2 seroprevalence survey estimates are affected by anti-nucleocapsid antibody decline. *The Journal of Infectious Diseases*, *223*(8), 1334–1338. https://doi.org/10.1093/INFDIS/JIAA796

Borchering, R. (2019). Impacts of Zika emergence in Latin America on endemic dengue transmission. *Nat. Commun.*, *10*.

Borghan, M. A., Oka, S., & Takiguchi, M. (2005). Identification of HLA-A*3101-restricted Cytotoxic T-lymphocyte Response to Human Immunodeficiency Virus Type 1 (HIV-1) in Patients With Chronic HIV-1 Infection. *Tissue Antigens*. https://doi.org/10.1111/j.1399-0039.2005.00489.x

Borkotoky, S., Dey, D., & Hazarika, Z. (2022). *Structural Perspective of the Interactions of ACE2 and SARS CoV-2 Spike Protein RBD*. https://doi.org/10.21203/rs.3.rs-2053408/v1

Bosso, G., Allegorico, E., Pagano, A., Porta, G., Serra, C., Minerva, V., Mercurio, V., Russo, T., Altruda, C., Arbo, P., Sio, C. d., Vicario, F. D., & Numis, F. G. (2020). Lung Ultrasound as Diagnostic Tool for SARS-CoV-2 Infection. *Internal and Emergency Medicine*. https://doi.org/10.1007/s11739-020-02512-y

Bouvier, M., Chen, W. J., Arnold, J., Fairchok, M., Danaher, P., Lalani, T., Malone, L., Mor, D., Ridore, M., Burgess, T., & Millar, E. V. (2018). Species-specific Clinical Characteristics of Human Coronavirus Infection Among Otherwise Healthy Adolescents and Adults. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.12538

Breslin, J. J., Mørk, I., Smith, M. J., Vogel, L. K., Hemmila, E. M., Bonavia, A., Talbot, P. J., Sjöström, H., Norén, O., & Holmes, K. V. (2003). Human Coronavirus 229E: Receptor Binding Domain and Neutralization by Soluble Receptor at 37°C. *Journal of Virology*. https://doi.org/10.1128/jvi.77.7.4435-4438.2003

Brielle, E. S., Stroopinsky, D., & Linial, M. (2020). The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting With the ACE2 Human Receptor. *Viruses*. https://doi.org/10.3390/v12050497

Brockway, S. M., Clay, C. T., Lu, X. T., & Denison, M. R. (2003). Characterization of the Expression, Intracellular Localization, and Replication Complex Association of the Putative Mouse Hepatitis Virus RNA-Dependent RNA Polymerase. *Journal of Virology*, *77*(19), 10515–10527. https://doi.org/10.1128/jvi.77.19.10515-10527.2003

Brunet, M.-C., Levesque, S., Hunting, D. J., Cohen, A. A., & Roucou, X. (2018). Recognition of the Polycistronic Nature of Human Genes Is Critical to Understanding the Genotype-Phenotype Relationship. *Genome Research*. https://doi.org/10.1101/gr.230938.117

Bulgaresi, M., Rivasi, G., Tarantini, F., Tofalos, S. E., Re, L. M. D., Salucci, C., Turrin, G., Barucci, R., Bandinelli, C., Fattorini, L., Borchi, D., Betti, M., Checchi, S., Baggiani, L., Collini, F., Lorini, C., Bonaccorsi, G., Ungar, A., Mossello, E., & Benvenuti, E. (2023). Impact of SARS-CoV2 Infection on Mortality and Hospitalization in Nursing Home Residents During the "Omicron Era." *Aging Clinical and Experimental Research*. https://doi.org/10.1007/s40520-023-02415-w

Bushman, M., Kahn, R., Taylor, B. P., Lipsitch, M., & Hanage, W. (2021). Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell*, *184*, 6229-6242.e18. https://doi.org/10.1016/j.cell.2021.11.026

Buss, L. F., Prete, C. A., Abrahim, C. M. M., Mendrone, A., Salomon, T., De Almeida-Neto, C., França, R. F. O., Belotti, M. C., Carvalho, M. P. S. S., Costa, A. G., Crispim, M. A. E., Ferreira, S. C., Fraiji, N. A., Gurzenda, S., Whittaker, C., Kamaura, L. T., Takecian, P. L., Da Silva Peixoto, P., Oikawa, M. K., … Sabino, E. C. (2021). Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*, *371*(6526), 288–292. https://doi.org/10.1126/science.abe9728

Cabeça, T. K., & Bellei, N. (2012). Human Coronavirus NL-63 Infection in a Brazilian Patient Suspected of H1N1 2009 Influenza Infection: Description of a Fatal Case. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2011.09.006

Cabeça, T. K., Granato, C. F. H., & Bellei, N. (2013). Epidemiological and Clinical Features of Human Coronavirus Infections Among Different Subsets of Patients. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.12101

Cañas-Arranz, R., León, P. d., Defaus, S., Torres, E., Forner, M., Bustos, M. J., Andreu, D., Blanco, E., & Sánchez-Madrid, F. (2021). Immunogenicity of Foot-and-Mouth Disease Virus Dendrimer Peptides: Need for a T-Cell Epitope and Ability to Elicit Heterotypic Responses. *Molecules*. https://doi.org/10.3390/molecules26164714

Cao, P., Huang, Y., Wu, Z., Sun, H., Ma, W., & Fang, T. (2020). A Case of Coronavirus HKU1 Encephalitis. *Acta Virologica*. https://doi.org/10.4149/av_2020_214

Cao, Yajie, Liu, J., Liu, T., Liu, D., & Zhang, Z. (2016). Theoretical Study of Continuous B-Cell Epitopes With Developed BP Neural Network. *Computational Chemistry*. https://doi.org/10.4236/cc.2016.43008

Cao, Yunlong, Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., Du, S., Wang, J., Li, Q., Chen, X., Yu, Y., Wang, P., Zhang, Z., Liu, P., An, R., Hao, X., Wang, Y., Wang, J., Feng, R., … Xie, X. S. (2022). BA.2.12.1, BA.4 and BA.5 Escape Antibodies Elicited by Omicron Infection. *Nature*. https://doi.org/10.1038/s41586-022-04980-y

Caraballo, D. A., Sabio, M. S. S., Colombo, V. C., Piccirilli, M. G., Vico, L., Riade, S. M. H., Campos, J., Martínez, G., Beltrán, F. J., Baumeister, E., & Cisterna, D. (2022). The Role of Molossidae and Vespertilionidae in Shaping the Diversity of Alphacoronaviruses in the Americas. *Microbiology Spectrum*, *10*(6). https://doi.org/10.1128/spectrum.03143-22

Carabelli, A., Peacock, T., Thorne, L., Harvey, W. T., Hughes, J., de Silva, T. D., Peacock, S., Barclay, W., Towers, G., & Robertson, D. L. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature Reviews. Microbiology*, *21*, 162–177. https://doi.org/10.1038/s41579-022-00841-7

Casimiro-Soriguer, C. S., Rigual, M. M., Brokate-Llanos, A. M., Muñoz, M. J., Garzón, A., Pérez-Pulido, A. J., & Jiménez, J. (2020). Using AnABlast for Intergenic sORF Prediction in the <i>Caenorhabditis Elegans</I> Genome. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btaa608

Castillo, G., Mora-Díaz, J. C., Nelli, R. K., & Giménez-Lirola, L. (2022). Human Air-Liquid-Interface Organotypic Airway Cultures Express Significantly More ACE2 Receptor Protein and Are More Susceptible to HCoV-NL63 Infection Than Monolayer Cultures of Primary Respiratory Epithelial Cells. *Microbiology Spectrum*. https://doi.org/10.1128/spectrum.01639-22

Castillo, G., Nelli, R. K., Phadke, K. S., Bravo-Parra, M., Mora-Díaz, J. C., Bellaire, B. H., & Giménez-Lirola, L. (2023). SARS-CoV-2 Is More Efficient Than HCoV-NL63 in Infecting a Small Subpopulation of ACE2+ Human Respiratory Epithelial Cells. *Viruses*. https://doi.org/10.3390/v15030736

Cavanagh, D. (1997). Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Archives of Virology*.

Cedro-Tanda, A., Gómez-Romero, L., de Anda-Jauregui, G., Garnica-López, D., Alfaro-Mora, Y., Sánchez-Xochipa, S., García-García, E. F., Mendoza-Vargas, A., Frías-Jiménez, E. J., Moreno, B., Campos-Romero, A., Moreno-Camacho, J. L., Alcantar-Fernández, J., Ortíz-Ramírez, J., Benitez-González, M., Trejo-González, R., Aguirre-Chavarría, D., Núñez-Martínez, M. E., Uribe-Figueroa, L., … Herrera, L. A. (2022). Early Genomic, Epidemiological, and Clinical Description of the SARS-CoV-2

Omicron Variant in Mexico City. *Viruses*, *14*(3). https://doi.org/10.3390/v14030545

Cele, S., Jackson, L., Khoury, D. S., Khan, K., Moyo-Gwete, T., Tegally, H., San, J. E., Cromer, D., Scheepers, C., Amoako, D. G., Karim, F., Bernstein, M., Lustig, G., Archary, D., Smith, M., Ganga, Y., Jule, Z., Reedoy, K., Hwa, S. H., … Sigal, A. (2022). Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*, *602*(7898), 654–656. https://doi.org/10.1038/s41586-021-04387-1

Centers for Disease Control and Prevention. (2020). People who are at higher risk for severe illness | CDC. *Centers for Disease Control and Prevention*.

Centre for Disease Control and Prevention. (2020). *Information for Healthcare Professionals: COVID-19 and Underlying Conditions*. CDC.

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E., Yu, J., Bahna, F., Bimela, J. S., Huang, Y., Katsamba, P., Liu, L., Nair, M., Rawi, R., Olia, A., Wang, P., Chuang, G., Ho, D., … Shapiro, L. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host & Microbe*, *29*, 819-833.e7. https://doi.org/10.1016/j.chom.2021.03.005

Chaudhury, S., Gromowski, G. D., Ripoll, D. R., Khavrutskii, I. V, Desai, V., & Wallqvist, A. (2017). Dengue Virus Antibody Database: Systematically Linking Serotype-Specificity With Epitope Mapping in Dengue Virus. *Plos Neglected Tropical Diseases*. https://doi.org/10.1371/journal.pntd.0005395

Che, D., Hasan, M. S., & Chen, B. (2014). Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens*. https://doi.org/10.3390/pathogens3010036

Chen, D.-Y., Kenney, D., Chin, C. V., Tavares, A. H., Khan, N., Conway, H. L., Liu, G., Choudhary, M. C., Gertje, H., O'Connell, A., Kotton, D. N., Herrmann, A., Ensser, A., Connor, J. H., Bosmann, M., Li, J. Z., Gack, M. U., Baker, S. C., Kirchdoerfer, R. N., … Saeed, M. (2022). Role of spike in the pathogenic and antigenic behavior of SARS-CoV-2 BA.1 Omicron. *BioRxiv*. https://doi.org/10.1101/2022.10.13.512134

Chen, D. Y., Chin, C. V, Kenney, D., Tavares, A. H., Khan, N., Conway, H. L., Liu, G., Choudhary, M., Gertje, H., O'Connell, A. K., Adams, S. R., Kotton, D. N., Herrmann, A., Ensser, A., Connor, J. H., Bosmann, M., Li, J. Z., Gack, M. U., Baker, S. C., … Saeed, M. (2023). Spike and Nsp6 Are Key Determinants of SARS-CoV-2 Omicron BA.1 Attenuation. *Nature*. https://doi.org/10.1038/s41586-023-05697-2

Chen, H., Tang, L., Yu, X., Zhou, J., Chang, Y., & Xiang, W. (2020). *Bioinformatics Analysis of Epitope-Based Vaccine Design Against the Novel SARS-CoV-2*. https://doi.org/10.21203/rs.3.rs-24655/v1

Chen, H., & Wu, X. (2020). *Bioinformatic Analysis of Epitope-Based Vaccine Design Against the Novel SARS-COV-2*. https://doi.org/10.21203/rs.3.rs-16262/v1

Chen, J., Li, Z., Guo, J., Xu, S., Zhou, J., Chen, Q., Tong, X., Wang, D., Peng, G., Fang, L., & Xiao, S. (2022). SARS-CoV-2 Nsp5 Exhibits Stronger Catalytic Activity and Interferon Antagonism Than Its SARS-CoV Ortholog. *Journal of Virology*. https://doi.org/10.1128/jvi.00037-22

Chen, K., Huang, D. T.-N., & Huang, L.-M. (2022). SARS-CoV-2 variants – Evolution, spike protein, and vaccines. *Biomedical Journal*, *45*, 573–579. https://doi.org/10.1016/j.bj.2022.04.006

Chen, M., Chen, Y., Qin, F., Zheng, L., & Zhang, X.-E. (2021). The Intraviral Protein-Protein Interaction of SARS-CoV-2 Reveals the Key Role of N Protein in Virus-Like Particle Assembly. *International Journal of Biological Sciences*. https://doi.org/10.7150/ijbs.64977

Chen, X., Zhu, Y., Li, Q., Liu, G., Li, C., Jin, R., Li, L., Xu, B., Gao, L., Yin, J., & Xie, Z. (2022). Genetic Characteristics of Human Coronavirus HKU1 in Mainland China During 2018. *Archives of Virology*. https://doi.org/10.1007/s00705-022-05541-4

Chen, Y., Jin, T., Li, M.-S., Xiao, Y., Huan, F., Liu, Q., Hu, M., Wei, X., Zheng, P., & Liu, G. M. (2023). Crystal Structure Analysis of Sarcoplasmic-Calcium-Binding Protein: An Allergen in <i>Scylla Paramamosain</I>. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.2c07267

Chen, Z., Azman, A. S., Azman, A. S., Chen, X., Zou, J., Tian, Y., Sun, R., Xu, X., Wu, Y.,

Lu, W., Ge, S., Zhao, Z., Yang, J., Leung, D. T., Domman, D., & Yu, H. (2021). Landscape of SARS-CoV-2 genomic surveillance, public availability extent of genomic data, and epidemic shaped by variants: a global descriptive study. *MedRxiv*. https://doi.org/10.1101/2021.09.06.21263152

Cheng, M. H., Zhang, S., Porritt, R. A., Rivas, M. N., Paschold, L., Willscher, E., Binder, M., Arditi, M., & Bahar, İ. (2020). Superantigenic Character of an Insert Unique to SARS-CoV-2 Spike Supported by Skewed TCR Repertoire in Patients With Hyperinflammation. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2010722117

Cheng, P. K. C., Wong, D. A., Tong, L. K. L., Ip, S. M., Lo, A. C. T., Lau, C. S., Yeung, E. Y. H., & Lim, W. W. L. (2004). Viral shedding patterns of coronavirus in patients with probable severe acute respiratory syndrome. *Lancet*, *363*(9422), 1699–1700. https://doi.org/10.1016/S0140-6736(04)16255-7

Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Horner, D. S., & Pesole, G. (2020). Next Generation Sequencing of SARS-CoV-2 Genomes: Challenges, Applications and Opportunities. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbaa297

Chibo, D., & Birch, C. (2006). Analysis of Human Coronavirus 229E Spike and Nucleoprotein Genes Demonstrates Genetic Drift Between Chronologically Distinct Strains. *Journal of General Virology*. https://doi.org/10.1099/vir.0.81662-0

Child, H. T., O'Neill, P., Moore, K., Rowe, W., Denise, H., Bass, D., Wade, M. J., Loose, M., Paterson, S., Aerle, R. v., & Jeffries, A. R. (2023). Optimised Protocol for Monitoring SARS-CoV-2 in Wastewater Using Reverse Complement PCR-based Whole-Genome Sequencing. *Plos One*. https://doi.org/10.1371/journal.pone.0284211

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Elizabeth Halloran, M., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400. https://doi.org/10.1126/science.aba9757

Cho, K. B. (2021). Development of Reverse Transcriptase Polymerase Chain Reaction

Primer Sets and Standard Positive Control Capable of Verifying False Positive for the Detection of <i>Severe Acute Respiratory Syndrome Coronavirus</I> 2. *Biomedical Science Letters*. https://doi.org/10.15616/bsl.2021.27.4.283

Choi, J., & Smith, D. M. (2021). SARS-CoV-2 Variants of Concern. *Yonsei Medical Journal*, *62*, 961–968. https://doi.org/10.3349/ymj.2021.62.11.961

Chouchane, L., Grivel, J.-C., Farag, E., Pavlovski, I., Maacha, S., Sathappan, A., Al-Romaihi, H. E., Abuaqel, S. W., Ata, M., Chouchane, A. I., Remadi, S., Halabi, N., Rafii, A., Al-Thani, M., Marr, N., Mahadevan, S., & Shan, J. (2021). Dromedary Camels as a Natural Source of Neutralizing Nanobodies Against SARS-CoV-2. *Jci Insight*. https://doi.org/10.1172/jci.insight.145785

Chow, J., Franz, K. M., & Kagan, J. C. (2015). PRRs Are Watching You: Localization of Innate Sensing and Signaling Regulators. *Virology*. https://doi.org/10.1016/j.virol.2015.02.051

Chow, N., Fleming-Dutra, K., Gierke, R., Hall, A., Hughes, M., Pilishvili, T., Ritchey, M., Roguski, K., Skoff, T., & Ussery, E. (2020). Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 — United States, February 12–March 28, 2020. *MMWR. Morbidity and Mortality Weekly Report*. https://doi.org/10.15585/mmwr.mm6913e2

Clark, C., Hardison, M., Houdeshell, H., Vest, A., Whitlock, D. A., Skola, D., Koble, J., Oberholzer, M., & Schroth, G. P. (2022). *Evaluation of an Optimized Protocol and Illumina ARTIC V4 Primer Pool for Sequencing of SARS-CoV-2 Using COVIDSeq™ and DRAGEN™ COVID Lineage App Workflow*. https://doi.org/10.1101/2022.01.07.475443

Cloete, J., Kruger, A., Masha, M., du Plessis, N. M., Mawela, D., Tshukudu, M., Manyane, T., Komane, L., Venter, M., Jassat, W., Goga, A., & Feucht, U. (2022). Paediatric hospitalisations due to COVID-19 during the first SARS-CoV-2 omicron (B.1.1.529) variant wave in South Africa: a multicentre observational study. *The Lancet Child and Adolescent Health*, *6*(5), 294–302. https://doi.org/10.1016/S2352-4642(22)00027-X

Corman, V. M., Baldwin, H., Tateno, A. F., Zerbinati, R. M., Annan, A., Owusu, M., Nkrumah, E. E., Maganga, G. D., Oppong, S., Adu-Sarkodie, Y., Vallo, P., Filho, L.

V. R. F. da S., Leroy, E. M., Thiel, V., Hoek, L. v. d., Poon, L. L. M., Tschapka, M., Drosten, C., & Drexler, J. F. (2015). Evidence for an Ancestral Association of Human Coronavirus 229E With Bats. *Journal of Virology*. https://doi.org/10.1128/jvi.01755-15

Corti, D., Purcell, L., Snell, G., & Veesler, D. (2021). Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell*, *184*, 3086–3108. https://doi.org/10.1016/j.cell.2021.05.005

Costa, C. H. S. da, Freitas, C. A. B. de, Alves, C. N., & Lameira, J. (2022). *Assessment of Mutations on RBD in the Spike Protein of SARS-CoV-2 Alpha, Delta and Omicron Variants*. https://doi.org/10.21203/rs.3.rs-1401835/v1

Cotten, M., Bugembe, D. L., Kaleebu, P., & Phan, M. V. T. (2021). Alternate Primers for Whole-Genome SARS-CoV-2 Sequencing. *Virus Evolution*. https://doi.org/10.1093/ve/veab006

COVID-19 Surveillance Group. (2020). Characteristics of COVID-19 patients dying in Italy Report based on available data on March 20 th , 2020. *COVID-19 Surveillance Group*.

Crawford, K. H. D., Dingens, A. S., Eguia, R., Wolf, C. R., Wilcox, N., Logue, J. K., Shuey, K., Casto, A. M., Fiala, B., Wrenn, S., Pettie, D., King, N. P., Greninger, A. L., Chu, H. Y., & Bloom, J. D. (2020). Dynamics of neutralizing antibody titers in the months after SARS-CoV-2 infection. *The Journal of Infectious Diseases*, *223*(2), 197–205. https://doi.org/10.1093/infdis/jiaa618

Crits-Christoph, A., Kantor, R., Olm, M. R., Whitney, O. N., Al-Shayeb, B., Lou, Y. C., Flamholz, A., Kennedy, L. C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F. W., Schroth, G. P., Kuersten, S., Banfield, J. F., & Nelson, K. L. (2020). *Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants*. https://doi.org/10.1101/2020.09.13.20193805

Crone, M. A., Hakki, S., Zhou, J., Oliveira, C. R. de, Madon, K., Koycheva, A., Badhan, A., Jonnerby, J., Fenn, J., Kundu, R., Barnett, J., Nevin, S. T., Conibear, E., Derqui-Fernández, N., Pillay, T., Varro, R., Luca, C., Quinn, V., Shazaad, A., … Lalvani, A. (2023). Rapid emergence of transmissible SARS-CoV-2 variants in mild community

cases. *MedRxiv*. https://doi.org/10.1101/2023.02.15.23285923

Cui, L. J., Zhang, C., Zhang, T., Lu, R., Zheng, X., Zhang, L. L., Liu, C. Y., Zhou, W., Ruan, L., Xue, J., & Tan, W. (2011). Human Coronaviruses HCoV-NL63 and HCoV-HKU1 in Hospitalized Children With Acute Respiratory Infections in Beijing, China. *Advances in Virology*. https://doi.org/10.1155/2011/129134

Cui, Y., Dong, X., Zhang, X., Chen, C., Fu, D., Li, X., & Liang, X. (2021). Deciphering the O-Glycosylation of HKU1 Spike Protein With the Dual-Functional Hydrophilic Interaction Chromatography Materials. *Frontiers in Chemistry*. https://doi.org/10.3389/fchem.2021.707235

Curran, C. S., Rivera, D. R., & Kopp, J. B. (2020). COVID-19 Usurps Host Regulatory Networks. *Frontiers in Pharmacology*. https://doi.org/10.3389/fphar.2020.01278

Damavandi, A. R., Dowran, R., Sharif, S. Al, Kashanchi, F., & Jafari, R. (2022). Molecular variants of SARS-CoV-2: antigenic properties and current vaccine efficacy. *Medical Microbiology and Immunology*, *211*, 79–103. https://doi.org/10.1007/s00430-022-00729-6

Dan, J. M., Mateus, J., Kato, Y., Hastie, K. M., Yu, E. D., Faliti, C. E., Grifoni, A., Ramirez, S. I., Haupt, S., Frazier, A., Nakao, C., Rayaprolu, V., Rawlings, S. A., Peters, B., Krammer, F., Simon, V., Saphire, E. O., Smith, D. M., Weiskopf, D., … Crotty, S. (2021). Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science*. https://doi.org/10.1126/science.abf4063

Dang, S., Ren, L., & Wang, J. (2022). Functional mutations of SARS-CoV-2: implications to viral transmission, pathogenicity and immune escape. *Chinese Medical Journal*, *135*, 1213–1222. https://doi.org/10.1097/CM9.0000000000002158

Dangi, T., Class, J., Palacio, N., Richner, J. M., & MacMaster, P. P. (2021). Combining Spike- And Nucleocapsid-Based Vaccines Improves Distal Control of SARS-CoV-2. *Cell Reports*. https://doi.org/10.1016/j.celrep.2021.109664

Dangi, T., Sanchez, S., Class, J., Richner, M., Visvabharathy, L., Chung, Y. R., Bentley, K., Stanton, R. J., Koralnik, I. J., Richner, J. M., & Penaloza-MacMaster, P. (2022). Improved Control of SARS-CoV-2 by Treatment With a Nucleocapsid-Specific

Monoclonal Antibody. *Journal of Clinical Investigation*. https://doi.org/10.1172/jci162282

Dangi, T., Sanchez, S., Park, M.-C., Class, J., Richner, M., Richner, J. M., & Penaloza-MacMaster, P. (2022). *Nucleocapsid-Specific Humoral Responses Improve the Control of SARS-CoV-2*. https://doi.org/10.1101/2022.03.09.483635

Das, J. K., Thakuri, B., MohanKumar, K., Roy, S., Sljoka, A., Sun, G.-Q., Sun, G.-Q., & Chakraborty, A. (2021). *Mutation-Induced Long-Range Allosteric Interactions in the Spike Protein Determine the Infectivity of SARS-CoV-2 Emerging Variants. 6*(46), 31312–31327. https://doi.org/10.1021/ACSOMEGA.1C05155

Davies, J., Almasy, K. M., McDonald, E. F., & Plate, L. (2020). Comparative Multiplexed Interactomics of SARS-CoV-2 and Homologous Coronavirus Nonstructural Proteins Identifies Unique and Shared Host-Cell Dependencies. *Acs Infectious Diseases*. https://doi.org/10.1021/acsinfecdis.0c00500

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., Diaz-Ordaz, K., Keogh, R., Eggo, R. M., … Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, *372*(6538). https://doi.org/10.1126/science.abg3055

de Wilde, A. H., Snijder, E. J., Kikkert, M., & van Hemert, M. J. (2018). Host factors in coronavirus replication. In *Current Topics in Microbiology and Immunology* (Vol. 419, pp. 1–42). Springer Verlag. https://doi.org/10.1007/82_2017_25

De Wit, E., Van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. In *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro.2016.81

Deng, J., Jiang, P., Qiu, M., Mao, L., Wang, Z., Zhu, G., Gao, L., Su, J., Hu, Y., Luo, O. J., Chen, G., & Wang, P. (2021). Identification of HLA-A2 Restricted CD8+ T Cell Epitopes in SARS-CoV-2 Structural Proteins. *Journal of Leukocyte Biology*. https://doi.org/10.1002/jlb.4ma0621-020r

Deng, K., Uhlig, S., Goodman, L. B., Ip, H. S., Killian, M. L., Nemser, S. M., Ulaszek, J., Kiener, S., Kmet, M., Frost, K., Hettwer, K., Colson, B., Nichani, K., Schlierf, A., Mlalazi-Oyinloye, M., Scott, A., Reddy, R., & Tkachenko, A. (2022). Second Round of an Interlaboratory Comparison of SARS-CoV2 Molecular Detection Assays Used by 45 Veterinary Diagnostic Laboratories in the United States. *Journal of Veterinary Diagnostic Investigation*. https://doi.org/10.1177/10406387221115702

Dennehy, J. J. (2016). Evolutionary Ecology of Virus Emergence. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.13304

Denninger, V., Xu, C. K., Meisl, G., Morgunov, A. S., Fiedler, S., Ilsley, A., Emmenegger, M., Malik, A. Y., Piziorska, M. A., Schneider, M. M., Devenish, S. R. A., Kosmoliaptsis, V., Aguzzi, A., Fiegler, H., & Knowles, T. P. J. (2021). *Understanding the Role of Memory Re-Activation and Cross-Reactivity in the Defense Against SARS-CoV-2*. https://doi.org/10.1101/2021.07.23.453352

Denninger, V., Xu, C. K., Meisl, G., Morgunov, A. S., Fiedler, S., Ilsley, A., Emmenegger, M., Malik, A. Y., Piziorska, M. A., Schneider, M. M., Devenish, S. R. A., Kosmoliaptsis, V., Aguzzi, A., Fiegler, H., & Knowles, T. P. J. (2022). Microfluidic Antibody Affinity Profiling Reveals the Role of Memory Reactivation and Cross-Reactivity in the Defense Against SARS-CoV-2. *Acs Infectious Diseases*. https://doi.org/10.1021/acsinfecdis.1c00486

Devi, Y. D., Goswami, H. B., Konwar, S., Doley, C., Dolley, A., Devi, A., Chongtham, C., Dowerah, D., Biswa, V., Jamir, L., Kumar, A., Satapathy, S. S., Ray, S. K., Deka, R. C., Doley, R., Mandal, M., Das, S., Singh, C. S., Borah, P. P., … Namsa, N. D. (2021). Immunoinformatics Mapping of Potential Epitopes in SARS-CoV-2 Structural Proteins. *Plos One*. https://doi.org/10.1371/journal.pone.0258645

Dijkman, R., Jebbink, M. F., Gaunt, E., Rossen, J. W. A., Templeton, K., Kuijpers, T. W., & Hoek, L. v. d. (2012). The Dominance of Human Coronavirus OC43 and NL63 Infections in Infants. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2011.11.011

Dijkman, R., Jebbink, M. F., Idrissi, N. B. E., Pyrć, K., Müller, M. A., Kuijpers, T. W., Zaaijer, H. L., & Hoek, L. v. d. (2008). Human Coronavirus NL63 and 229E

Seroconversion in Children. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.00533-08

Dijkman, R., Jebbink, M. F., Koekkoek, S. M., Deijs, M., Jonsdottir, H. R., Molenkamp, R., Ieven, M., Goossens, H., Thiel, V., & Hoek, L. v. d. (2013). Isolation and Characterization of Current Human Coronavirus Strains in Primary Human Epithelial Cell Cultures Reveal Differences in Target Cell Tropism. *Journal of Virology*. https://doi.org/10.1128/jvi.03368-12

Doddapaneni, H., Cregeen, S. J., Sucgang, R., Meng, Q., Qin, X., Avadhanula, V., Chao, H., Menon, V., Nicholson, E., Henke, D., Piedra, F.-A., Rajan, A., Momin, Z., Kottapalli, K., Hoffman, K. L., Sedlazeck, F. J., Metcalf, G., Piedra, P. A., Petrosino, J. F., & Gibbs, R. A. (2020). *Oligonucleotide Capture Sequencing of the SARS-CoV-2 Genome and Subgenomic Fragments From COVID-19 Individuals*. https://doi.org/10.1101/2020.07.27.223495

Dolan, K. A., Dutta, M., Kern, D. M., Kotecha, A., Voth, G. A., & Brohawn, S. G. (2022). Structure of SARS-CoV-2 M protein in lipid nanodiscs. *BioRxiv*, *11*. https://doi.org/10.1101/2022.06.12.495841

Dominguez, S. R., Travanty, E. A., Qian, Z., & Mason, R. J. (2013). Human Coronavirus HKU1 Infection of Primary Human Type II Alveolar Epithelial Cells: Cytopathic Effects and Innate Immune Response. *Plos One*. https://doi.org/10.1371/journal.pone.0070129

Donohue, W. A. (2022). Geographical prevalence of SARS-CoV-2 variants, August 2020 to July 2021. *Dental Science Reports*, *12*(1). https://doi.org/10.1038/s41598-022-08684-1

Doytchinova, I., & Tchorbanov, A. (2020). Design of Multi-Epitope Vaccine Against SARS-CoV-2. *Cybernetics and Information Technologies*. https://doi.org/10.2478/cait-2020-0072

Du, B., Guo, Y., Li, G., Zhu, Y., Wang, Y., & Xi, X. (2022). Non-structure protein ORF1ab (NSP8) in SARS-CoV-2 contains potential γδT cell epitopes. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.936272

Dyrdak, R., Hodcroft, E. B., Wahlund, M., Neher, R. A., & Albert, J. (2021). Interactions Between Seasonal Human Coronaviruses and Implications for the SARS-CoV-2 Pandemic: A Retrospective Study in Stockholm, Sweden, 2009-2020. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2021.104754

EL-Manzalawy, Y., & Honavar, V. (2010). Recent Advances in B-Cell Epitope Prediction Methods. *Immunome Research*. https://doi.org/10.1186/1745-7580-6-s2-s2

Ellebedy, A. H., Krammer, F., Li, G., Miller, M. S., Chiu, C., Wrammert, J., Chang, C., Davis, C. W., McCausland, M., Elbein, R., Edupuganti, S., Spearman, P., Andrews, S. F., Wilson, P. C., García-Sastre, A., Mulligan, M. J., Mehta, A. K., Palese, P., & Ahmed, R. (2014). Induction of Broadly Cross-Reactive Antibody Responses to the Influenza HA Stem Region Following H5N1 Vaccination in Humans. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1414070111

Errico, J., Adams, L., & Fremont, D. (2022). Antibody-mediated immunity to SARS-CoV-2 spike. *Advances in Immunology*, *154*, 1–69. https://doi.org/10.1016/bs.ai.2022.07.001

Espejo, A. P., Akgun, Y., Mana, A. F. A., Tjendra, Y., Millan, N., Gomez-Fernandez, C., & Cray, C. (2020). Review of Current Advances in Serologic Testing for COVID-19. *American Journal of Clinical Pathology*. https://doi.org/10.1093/ajcp/aqaa112

Fagyas, M., Nagy, B., Ráduly, A. P., Mányiné, I. S., Mártha, L., Erdősi, G., Sipka, S., Enyedi, E. E., Szabó, A. Á., Pólik, Z., Kappelmayer, J., Papp, Z., Borbély, A., Szabó, T., Balla, J., Balla, G., Bai, P., Bacsi, A., & Tóth, A. (2022). The Majority of Severe COVID-19 Patients Develop Anti-Cardiac Autoantibodies. *Geroscience*. https://doi.org/10.1007/s11357-022-00649-6

Farahat, R., Abdelaal, A., Umar, T. P., El-Sakka, A. A., Benmelouka, A. Y., Albakri, K., Ali, I. H., Al-Ahdal, T. M. A., Abdelazeem, B., Sah, R., & Rodriguez-Morales, A. J. (2022). *The emergence of SARS-CoV-2 Omicron subvariants: current situation and future trends. 30 4*(4), 480–494. https://doi.org/10.53854/liim-3004-2

Farfour, E., Yung, T. M., Baudoin, R., & Vasse, M. (2022). Evaluation of Four Fully Integrated Molecular Assays for the Detection of Respiratory Viruses During the Co-Circulation of SARS-CoV-2, Influenza and RSV. *Journal of Clinical Medicine*.

https://doi.org/10.3390/jcm11143942

Fausto, A. (2023). Improved Culture Methods for Human Coronaviruses HCoV-OC43, HCoV-229E, and HCoV-NL63. *Current Protocols*. https://doi.org/10.1002/cpz1.914

Faye, M. N., Barry, M. A., Jallow, M. M., Wade, S. F., Mendy, M. P., Sy, S., Fall, A., Kiori, D. E., Ndiaye, N. K., Goudiaby, D., Diamanka, A., Niang, M. N., & Dia, N. (2022). Epidemiology of Non-Sars-CoV2 Human Coronaviruses (HCoVs) in People Presenting With Influenza-Like Illness (ILI) or Severe Acute Respiratory Infections (SARI) in Senegal From 2012 to 2020. *Viruses*. https://doi.org/10.3390/v15010020

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2020). *The Coding Capacity of SARS-CoV-2*. https://doi.org/10.1101/2020.05.07.082909

Fischer, M., Baessler, A., & Schunkert, H. (2002). Renin angiotensin system and gender differences in the cardiovascular system. In *Cardiovascular Research*. https://doi.org/10.1016/S0008-6363(01)00479-5

Fischer, N., Dauby, N., Bossuyt, N., Reynders, M., Gerard, M., Lacor, P., Daelemans, S., Lissoir, B., Holemans, X., Magerman, K., Jouck, D., Bourgeois, M., Delaere, B., Quoilin, S., Gucht, S. V, Thomas, I., Barbezange, C., & Subissi, L. (2021). Monitoring of Human Coronaviruses in Belgian Primary Care and Hospitals, 2015–20: A Surveillance Study. *The Lancet Microbe*. https://doi.org/10.1016/s2666-5247(20)30221-4

Florek, D., Burmistrz, M., Potempa, J., & Pyrc, K. (2014). Stability of Infectious Human Coronavirus NL63. *Journal of Virological Methods*. https://doi.org/10.1016/j.jviromet.2014.04.001

Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., & Hurley, J. H. (2020). Structure of SARS-CoV-2 ORF8, a Rapidly Evolving Immune Evasion Protein. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2021785118

Forcelloni, S., Benedetti, A., Dilucca, M., & Giansanti, A. (2020). *Identification of*

*Conserved Epitopes in SARS-CoV-2 Spike and Nucleocapsid Protein*. https://doi.org/10.1101/2020.05.14.095133

Forcelloni, S., Benedetti, A., Dilucca, M., & Giansanti, A. (2021). Identification of Conserved Epitopes in SARS-CoV-2 Spike and Nucleocapsid Protein. *Current Genomics*. https://doi.org/10.2174/1389202923666211216162605

Forni, D., Cagliani, R., Molteni, C., Arrigoni, F., Mozzi, A., Clerici, M., Gioia, L. D., & Sironi, M. (2022). Homology-based Classification of Accessory Proteins in Coronavirus Genomes Uncovers Extremely Dynamic Evolution of Gene Content. *Molecular Ecology*. https://doi.org/10.1111/mec.16531

Forsström, B., Axnäs, B. B., Rockberg, J., Danielsson, H., Anna, B., & Uhlén, M. (2015). Dissecting Antibodies With Regards to Linear and Conformational Epitopes. *Plos One*. https://doi.org/10.1371/journal.pone.0121673

Fratev, F. (2020). The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *BioRxiv*. https://doi.org/10.1101/2020.12.23.424283

Fryer, H. A., Hartley, G. E., Edwards, E., O'Hehir, R., & van Zelm, M. V. (2022). Humoral immunity and B-cell memory in response to SARS-CoV-2 infection and vaccination. *Biochemical Society Transactions*, *50*, 1643–1658. https://doi.org/10.1042/BST20220415

Fu, Y.-Z., Wang, S., Zheng, Z.-Q., Huang, Y., Li, W., Xu, Z., & Wang, Y.-Y. (2020). SARS-CoV-2 Membrane Glycoprotein M Antagonizes the MAVS-mediated Innate Antiviral Response. *Cellular and Molecular Immunology*. https://doi.org/10.1038/s41423-020-00571-x

Galina, A. C., Sarzi, D. S., Medeiros, L. C. d., Sampaio, A. L. F., & Leta, J. (2020). *Scientific Output on Coronavirus and WHO's Solidarity Project: A Science-Based Choice?* https://doi.org/10.1101/2020.11.16.20232488

Gall, S. L., Stamegna, P., & Walker, B. D. (2007). Portable Flanking Sequences Modulate CTL Epitope Processing. *Journal of Clinical Investigation*. https://doi.org/10.1172/jci32047

Gallais, F., Velay, A., Nazon, C., Wendling, M.-J., Partisani, M., Sibilia, J., Candon, S., & Fafi-Kremer, S. (2021). Intrafamilial Exposure to SARS-CoV-2 Associated With Cellular Immune Response Without Seroconversion, France. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid2701.203611

Gao, H., Jiang, H., Hu, X., Zhang, Y., Zhou, X., Zhong, F., Lin, C. W., Li, J., Luo, J., & Zhang, J. (2021). Crystal Structures of Human Coronavirus NL63 Main Protease at Different pH Values. *Acta Crystallographica Section F Structural Biology Communications*. https://doi.org/10.1107/s2053230x21009523

Gard, A. L., Luu, R. J., Miller, C., Maloney, R., Cain, B. P., Marr, E. E., Burns, D., Gaibler, R., Mulhern, T. J., Wong, C., Alladina, J., Coppeta, J., Liu, P., Wang, J. P., Azizgolshani, H., Fezzie, R. F., Balestrini, J. L., Isenberg, B. C., Medoff, B. D., … Borenstein, J. T. (2021). High-Throughput Human Primary Cell-Based Airway Model for Evaluating Influenza, Coronavirus, or Other Respiratory Viruses in Vitro. *Scientific Reports*. https://doi.org/10.1038/s41598-021-94095-7

Gaunt, E., Hardie, A., Claas, E. C. J., Simmonds, P., & Templeton, K. (2010). Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected Over 3 Years Using a Novel Multiplex Real-Time PCR Method. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.00636-10

Gauttier, V., Morello, A., Girault, I., Mary, C., Belarif, L., Desselle, A., Wilhelm, E., Bourquard, T., Pengam, S., Teppaz, G., Thepenier, V., Biteau, K., Barbeyrac, E. D., Kiepferlé, D., Vasseur, B., Flem, F. L., Debieuvre, D., Costantini, D., & Poirier, N. (2020). *Tissue-Resident Memory CD8 T-Cell Responses Elicited by a Single Injection of a Multi-Target COVID-19 Vaccine*. https://doi.org/10.1101/2020.08.14.240093

Ge, L., Zhang, M., Kenny, S. J., Liu, D., Maeda, M., Saito, K., Mathur, A., Xu, K., & Schekman, R. (2017). Remodeling Of <scp>ER</Scp> -exit Sites Initiates a Membrane Supply Pathway for Autophagosome Biogenesis. *Embo Reports*. https://doi.org/10.15252/embr.201744559

Geanes, E. S., LeMaster, C., Fraley, E., Khanal, S., McLennan, R., Grundberg, E., Selvarangan, R., & Bradley, T. (2022). Cross-Reactive Antibodies Elicited to Conserved Epitopes on SARS-CoV-2 Spike Protein After Infection and Vaccination.

*Scientific Reports*. https://doi.org/10.1038/s41598-022-10230-y

Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H. K., Morgan, R., & Klein, S. L. (2020). Impact of sex and gender on COVID-19 outcomes in Europe. In *Biology of Sex Differences*. https://doi.org/10.1186/s13293-020-00304-9

Gellenoncourt, S., Saunders, N., Robinot, R., Auguste, L., Rajah, M. M., Kervevan, J., Jeger-Madiot, R., Staropoli, I., Planchais, C., Mouquet, H., Buchrieser, J., Schwartz, O., & Chakrabarti, L. (2022). The Spike-Stabilizing D614G Mutation Interacts with S1/S2 Cleavage Site Mutations To Promote the Infectious Potential of SARS-CoV-2 Variants. *Journal of Virology*, *96*, null. https://doi.org/10.1128/jvi.01301-22

Geller, C., Fontanay, S., Finance, C., & Duval, R. E. (2009). A New Sephadex$^{TM}$-Based Method for Removing Microbicidal and Cytotoxic Residues When Testing Antiseptics Against Viruses: Experiments With a Human Coronavirus as a Model. *Journal of Virological Methods*. https://doi.org/10.1016/j.jviromet.2009.03.023

Georgakopoulou, V., Petsinis, G., Mantzouranis, K., Damaskos, C., Melemeni, D., Gkoufa, A., Chlapoutakis, S., Garmpis, N., Sklapani, P., Trakas, N., & Tsiafaki, X. (2021). First Detection of Human Coronavirus HKU1 in Greece, in an Immunocompromised Patient With Severe Lower Respiratory Tract Infection. *Acta Medica Lituanica*. https://doi.org/10.15388/amed.2021.28.1.21

Ghafouri, F., Cohan, R. A., Noorbakhsh, F., Samimi, H., & Haghpanah, V. (2020). *An in-Silico Approach to Develop of a Multi-Epitope Vaccine Candidate Against SARS-CoV-2 Envelope (E) Protein*. https://doi.org/10.21203/rs.3.rs-30374/v1

Ghai, R. R., Carpenter, A., Liew, A., Martin, K. B., Herring, M. K., Gerber, S. I., Hall, A. J., Sleeman, J. M., VonDobschuetz, S., & Behravesh, C. B. (2021). Animal Reservoirs and Hosts for Emerging Alphacoronaviruses and Betacoronaviruses. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid2704.203945

Githinji, G., Lambisia, A. W., Omah, I. F., O'Toole, A., Mohamed, K., Mwanga, M. J., Mburu, M. W., Makori, T., Morobe, J. M., Ong'era, E. M., Ndwiga, L., Laurent, Z. R. de, Gathii, K., Thiongo, K., Omuoyo, D. O., Chepkorir, E., Musyoki, J. N., Kingwara, L., Matoke, D., … Agoti, C. N. (2022). The genomic epidemiology of SARS-CoV-2 variants of concern in Kenya. *MedRxiv*. https://doi.org/10.1101/2022.10.26.22281446

Gogoi, S., Bora, I., Debnath, E., Sarkar, S., Jais, M., & Sharma, A. (2021). Perplexity vs Clarity in Choosing the Right Molecular Diagnostic Techniques for SARS-COV2 Detection in Indian Setup. *Journal of Family Medicine and Primary Care*. https://doi.org/10.4103/jfmpc.jfmpc_1793_20

Gohl, D. M., Garbe, J., Grady, P. G. S., Daniel, J., Watson, R. H. B., Auch, B., Nelson, A. C., Yohe, S., & Beckman, K. B. (2020). A Rapid, Cost-Effective Tailed Amplicon Method for Sequencing SARS-CoV-2. *BMC Genomics*. https://doi.org/10.1186/s12864-020-07283-6

Golota, A. S. (2022). Cellular Immunity in the Patients with COVID-19: Molecular Biology, Pathophysiology, and Clinical Implications. *Journal of Clinical Practice*, *null*, null. https://doi.org/10.17816/clinpract106239

González, R., Nhampossa, T., Figueroa-Romero, A., Mendes, AMazuze, M., García-Otero, L., E, S., Piqueras, M., Egri, N., Bedini, J., Saúte, F., & Menendez, C. (2022). SARS-CoV-2 Seropositivity and HIV Viral Load Among Mozambican Pregnant Women. *Journal of Acquired Immune Deficiency Syndromes*, *92*(2), 115–121. https://doi.org/10.1097/qai.0000000000003120

Gorse, G. J., Donovan, M. M., & Patel, G. B. (2020). Antibodies to Coronaviruses Are Higher in Older Compared With Younger Adults and Binding Antibodies Are More Sensitive Than Neutralizing Antibodies in Identifying Coronavirus-associated Illnesses. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25715

Gorse, G. J., Patel, G. B., Vitale, J. N., & O'Connor, T. (2010). Prevalence of Antibodies to Four Human Coronaviruses Is Lower in Nasal Secretions Than in Serum. *Clinical and Vaccine Immunology*. https://doi.org/10.1128/cvi.00278-10

Gottardo, R., Bailer, R. T., Korber, B. T., Gnanakaran, S., Phillips, J., Shen, X., Tomaras, G. D., Turk, E., Imholte, G., Eckler, L., Wenschuh, H., Zerweck, J., Greene, K., Gao, H., Berman, P. W., Francis, D. P., Sinangil, F., Lee, C., Nitayaphan, S., … Montefiori, D. C. (2013). Plasma IgG to Linear Epitopes in the V2 and V3 Regions of HIV-1 Gp120 Correlate With a Reduced Risk of Infection in the RV144 Vaccine Efficacy Trial. *Plos One*. https://doi.org/10.1371/journal.pone.0075665

Goulder, P., Sewell, A. K., Lalloo, D. G., Price, D. A., Whelan, J. A., Evans, J. W., Taylor,

G. P., Luzzi, G., Giangrande, P., Phillips, R. E., & McMichael, A. J. (1997). Patterns of Immunodominance in HIV-1–specific Cytotoxic T Lymphocyte Responses in Two Human Histocompatibility Leukocyte Antigens (HLA)-identical Siblings With HLA-A*0201 Are Influenced by Epitope Mutation. *The Journal of Experimental Medicine*. https://doi.org/10.1084/jem.185.8.1423

Greaney, A. J., Starr, T. N., Barnes, C., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., Wang, Z., Poston, D., Muecksch, F., Hatziioannou, T., Bieniasz, P., Robbiani, D., Nussenzweig, M., Bjorkman, P., & Bloom, J. (2021). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nature Communications*, *12*, null. https://doi.org/10.1038/s41467-021-24435-8

Greaney, A. J., Starr, T. N., Eguia, R. T., Loes, A., Khan, K., Karim, F., Cele, S., Bowen, J. E., Logue, J., Corti, D., Veesler, D., Chu, H., Sigal, A., & Bloom, J. (2021). A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. *BioRxiv*, *null*, null. https://doi.org/10.1101/2021.10.12.464114

Grifoni, A., Angelo, M. A., López, B., O'Rourke, P., Sidney, J., Cerpas, C., Balmaseda, Á., Silveira, C. G. T., Maestri, A., Costa, P. R., Durbin, A. P., Diehl, S. A., Phillips, E., Mallal, S., Silva, A. D. D., Nchinda, G., Nkenfou, C. N., Collins, M. H., Silva, A. M. d., … Weiskopf, D. (2017). Global Assessment of Dengue Virus-Specific CD4+ T Cell Responses in Dengue-Endemic Areas. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2017.01309

Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020a). A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host & Microbe*. https://doi.org/10.1016/j.chom.2020.03.002

Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020b). A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host and Microbe*, *27*(4), 671-680.e2. https://doi.org/10.1016/j.chom.2020.03.002

Gruta, N. L. L., Kedzierska, K., Pang, K. C., Webby, R. J., Davenport, M. P., Chen, W.,

Turner, S. J., & Doherty, P. C. (2006). A Virus-Specific $CD8^{+}$T Cell Immunodominance Hierarchy Determined by Antigen Dose and Precursor Frequencies. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0510429103

Gu, H., Chen, Q., Guan, Y., He, L., Fan, H., Deng, Y.-Q., Wang, Y., Teng, Y., Zhao, Z., Cui, Y., Li, Y., Li, X.-F., Li, J., Zhang, N.-N., Yang, X., Chen, S., Guo, Y., Zhao, G., Wang, X., … Zhou, Y. (2020). Adaptation of SARS-CoV-2 in BALB/c Mice for Testing Vaccine Efficacy. *Science*, *369*(6511), 1603–1607. https://doi.org/10.1126/science.abc4730

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., … Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, *382*(18), 1708–1720. https://doi.org/10.1056/nejmoa2002032

Guðbjartsson, D. F., Norddahl, G. L., Melsted, P., Gunnarsdóttir, K., Hólm, H., Eyþórsson, E., Arnthórsson, A. Ö., Helgason, D., Bjarnadóttir, K., Ingvarsson, R. F., Thorsteinsdottir, B., Kristjánsdóttir, S., Birgisdottir, K., Kristinsdottir, A. M., Sigurðsson, M. I., Arnadottir, G. A., Ívarsdóttir, E. V, Andrésdóttir, M. B., Jónsson, F., … Stefánsson, K. (2020). Humoral Immune Response to SARS-CoV-2 in Iceland. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa2026116

Guerin, P. M., Yahi, N., Azzaz, F., Chahinian, H., Sabatier, J.-M., & Fantini, J. (2022). Structural Dynamics of the SARS-CoV-2 Spike Protein: A 2-Year Retrospective Analysis of SARS-CoV-2 Variants (From Alpha to Omicron) Reveals an Early Divergence Between Conserved and Variable Epitopes. *Molecules*. https://doi.org/10.3390/molecules27123851

Guevarra, L. A., & Ulanday, G. E. L. (2020). *Immune Epitope Map of the Reported Protein Sequences of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*. https://doi.org/10.21203/rs.3.rs-18689/v1

Gunadi, Wibawa, H., Marcellus, Hakim, M. S., Daniwijaya, E. W., Rizki, L. P., Supriyati, E., Nugrahaningsih, D. A. A., Afiahayati, Siswanto, S., Iskandar, K., Anggorowati, N.,

Kalim, A. S., Puspitarani, D. A., Athollah, K., Arguni, E., Nuryastuti, T., & Wibawa, T. (2020). Full-Length Genome Characterization and Phylogenetic Analysis of SARS-CoV-2 Virus Strains From Yogyakarta and Central Java, Indonesia. *Peerj*. https://doi.org/10.7717/peerj.10575

Guo, C., Tsai, S. J., Ai, Y., Li, M., Anaya, E. F., Pekosz, A., Cox, A. L., & Gould, S. J. (2022). The D614G mutation redirects SARS-CoV-2 spike to lysosomes and suppresses deleterious traits of the furin cleavage site insertion mutation. *Science Advances*, *8*(51). https://doi.org/10.1126/sciadv.ade5085

Gupta, A., Rosato, A. J., & Cui, F. (2020). Vaccine Candidate Designed Against Carcinoembryonic Antigen-Related Cell Adhesion Molecules Using Immunoinformatics Tools. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1797539

Gupta, V., Bhoyar, R. C., Jain, A., Srivastava, S., Upadhayay, R., Imran, M., Jolly, B., Divakar, M. K., Sharma, D., Sehgal, P., Ranjan, G., Gupta, R. K., Scaria, V., & Sivasubbu, S. (2020). *Asymptomatic Reinfection in Two Healthcare Workers From India With Genetically Distinct SARS-CoV-2*. https://doi.org/10.31219/osf.io/4fmrg

Guruprasad, L. (2021). Human Coronavirus Spike Protein-Host Receptor Recognition. *Progress in Biophysics and Molecular Biology*. https://doi.org/10.1016/j.pbiomolbio.2020.10.006

Hajissa, K., Islam, M. A., Hassan, S. A., Zaidah, A. R., Ismail, N., & Mohamed, Z. (2022). Seroprevalence of SARS-CoV-2 Antibodies in Africa: A Systematic Review and Meta-Analysis. *International Journal of Environmental Research and Public Health*, *19*(12), 7257. https://doi.org/10.3390/ijerph19127257

Hampton, J. T., Lalonde, T., Tharp, J. M., Alugubelli, Y. R., Roundy, C. M., Hamer, G. L., Xu, S., Liu, W. R., & Kurra, Y. (2022). *A Novel Regioselective Approach to Cyclize Phage-Displayed Peptides in Combination With Epitope-Directed Selection to Identify a Potent Neutralizing Macrocyclic Peptide for SARS-CoV-2*. https://doi.org/10.1101/2022.07.06.498864

Han, A. X., Toporowski, A., Sacks, J. A., Perkins, M. D., Briand, S., Kerkhove, M. D. Van, Hannay, E. J., Carmona, S., Rodriguez, B., Parker, E., Nichols, B. E., & Russell, C. A.

(2022). SARS-CoV-2 diagnostic testing rates determine the sensitivity of genomic surveillance programs. *MedRxiv*. https://doi.org/10.1101/2022.05.20.22275319

Hanke, L., Perez, L. V, Sheward, D. J., Das, H., Schulte, T., Morro, A. M., Corcoran, M., Achour, A., Hedestam, G. K., Hallberg, B., Murrell, B., & McInerney, G. M. (2020). *An Alpaca Nanobody Neutralizes SARS-CoV-2 by Blocking Receptor Interaction*. https://doi.org/10.1101/2020.06.02.130161

Harvey, W. T., Carabelli, A., Jackson, B., Gupta, R. K., Thomson, E., Harrison, E., Ludden, C., Reeve, R., Rambaut, A., Peacock, S., & Robertson, D. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews. Microbiology*, *19*, 409–424. https://doi.org/10.1038/s41579-021-00573-0

Hassan, S. S., Lundstrom, K., Serrano-Aroca, Á., Adadi, P., Aljabali, A. A. A., Redwan, E. M., Lal, A., Kandimalla, R., El-Aziz, T. M. A., Choudhury, P. P., Azad, G. K., Sherchan, S. P., Tambuwala, M. M., Chauhan, G., Takayama, K., Barh, D., Palù, G., Basu, P., & Uversky, V. N. (2021). *Emergence of Unique SARS-CoV-2 ORF10 Variants and Their Impact on Protein Structure and Function*. https://doi.org/10.20944/PREPRINTS202107.0554.V1

He, B., Zhang, Y., Xu, L., Yang, W., Yang, F., Feng, Y., Xia, L., Zhou, J., Wei-bin, Z., Feng, Y., Guo, H., Zhang, H., & Tu, C. (2014). Identification of Diverse Alphacoronaviruses and Genomic Characterization of a Novel Severe Acute Respiratory Syndrome-Like Coronavirus From Bats in China. *Journal of Virology*. https://doi.org/10.1128/jvi.00631-14

He, D. (2015). Global spatio-temporal patterns of influenza in the post-pandemic era. *Sci. Rep.*, *5*.

He, J., Tao, H., Yan, Y., Huang, S.-Y., & Xiao, Y. (2020). Molecular Mechanism of Evolution and Human Infection With SARS-CoV-2. *Viruses*. https://doi.org/10.3390/v12040428

He, X., Su, F., Chen, Y., & Li, Z. (2022). Novel reverse transcription-multiple inner primer loop-mediated isothermal amplification (RT-MIPLAMP) for visual and sensitive detection of SARS-CoV-2. *Analytical Methods*, *14*(47), 5012–5018. https://doi.org/10.1039/d2ay01330d

He, Y., Zhou, Y., Wu, H., Luo, B., Chen, J., Li, W., & Jiang, S. (2004). Identification of Immunodominant Sites on the Spike Protein of Severe Acute Respiratory Syndrome (SARS) Coronavirus: Implication for Developing SARS Diagnostics and Vaccines. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.173.6.4050

Heffron, A. S., McIlwain, S. J., Amjadi, M. F., Baker, D., Khullar, S., Armbrust, T., Halfmann, P., Kawaoka, Y., Sethi, A. K., Palmenberg, A. C., Shelef, M. A., O'Connor, D. H., & Ong, I. M. (2021). The Landscape of Antibody Binding in SARS-CoV-2 Infection. *Plos Biology*. https://doi.org/10.1371/journal.pbio.3001265

Hemmati-Dinarvand, M., Saedi, S., Zarei, M., & Seghatoleslam, A. (2020). Mysterious Virus: A Review on Behavior and Treatment Approaches of the Novel Coronavirus, 2019-nCoV. *Archives of Medical Research*. https://doi.org/10.1016/j.arcmed.2020.04.022

Higgins, A., C., H, B. E. N.-P., Bonaventure, B., Kurland, A. P., Ye, C., Yaron, M., T., Johnson, J. L., Prithy Adhikary, I. G., Pa, M., & Johnson, J. R. (2023). SARS-CoV-2 hijacks p38β/MAPK11 to promote virus replication. *Mbio*. https://doi.org/10.1128/mbio.01007-23

Hirabara, S., Serdan, T. D., Gorjão, R., Masi, L. N., Pithon-Curi, T., Covas, D., Curi, R., & Durigon, E. (2022). SARS-COV-2 Variants: Differences and Potential of Immune Evasion. *Frontiers in Cellular and Infection Microbiology*, *11*, null. https://doi.org/10.3389/fcimb.2021.781429

Hoffmann, M, Wong, L., Arora, P., Zhang, L., Rocha, C., Odle, A. E., Nehlmeier, I., Kempf, A., Richter, A., Halwe, N., Schön, J., Ulrich, L., Hoffmann, D., Beer, M., Drosten, C., Perlman, S., & Pöhlmann, S. (2023). Omicron subvariant BA.5 efficiently infects lung cells. *Nature Communications*, *14*, null. https://doi.org/10.1038/s41467-023-39147-4

Hoffmann, M, Zhang, L., Krüger, N., Graichen, L., Kleine-Weber, H., Hofmann-Winkler, H., Kempf, A., Nessler, S., Riggert, J., Winkler, M., Schulz, S. R., Jäck, H., & Pöhlmann, S. (2021). SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization. *Cell Reports*, *35*, 109017. https://doi.org/10.1016/j.celrep.2021.109017

Hoffmann, Markus, Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. https://doi.org/10.1016/j.cell.2020.02.052

Hofmann, H., Pyrć, K., Hoek, L. v. d., Geier, M., Berkhout, B., & Pöhlmann, S. (2005). Human Coronavirus NL63 Employs the Severe Acute Respiratory Syndrome Coronavirus Receptor for Cellular Entry. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0409465102

Hoter, A., & Naim, H. Y. (2022). Biochemical Characterization of SARS-CoV-2 Spike RBD Mutations and Their Impact on ACE2 Receptor Binding. *Frontiers in Molecular Biosciences*, *9*, 893843. https://doi.org/10.3389/fmolb.2022.893843

Hou, X., Zhang, Z., Gao, J., Gao, J., & Wang, Y. (2021). *SARS-CoV-2 spike protein N501Y mutation causes differential species transmissibility and antibody sensitivity: a molecular dynamics and alchemical free energy study*. *6*(11), 11. https://doi.org/10.1039/D1ME00086A

Hou, Y., Zhao, S., Liu, Q., Zhang, X., Sha, T., Su, Y., Zhao, W., Bào, Y., Xue, Y., & Chen, H. (2022). Ongoing Positive Selection Drives the Evolution of SARS-CoV-2 Genomes. *Genomics, Proteomics & Bioinformatics*, *20*, 1214–1223. https://doi.org/10.1016/j.gpb.2022.05.009

Hu, B., Chan, J., Liu, H., Liu, Y., Chai, Y., Shi, J., Shuai, H., Hou, Y., Huang, X., Yuen, T. T., Yoon, C., Zhu, T., Zhang, J., Li, W., Zhang, A. J., Zhou, J., Yuan, S., Zhang, B.-Z., Yuen, K., & Chu, H. (2022). Spike mutations contributing to the altered entry preference of SARS-CoV-2 omicron BA.1 and BA.2. *Emerging Microbes & Infections*, *11*, 2275–2287. https://doi.org/10.1080/22221751.2022.2117098

Hu, C., Wang, Z., Ren, L., Hao, Y., Zhu, M., Jiang, H., Wang, S., Li, D., & Shao, Y. (2022). Pre-existing anti-HCoV-OC43 immunity influences the durability and cross-reactivity of humoral response to SARS-CoV-2 vaccination. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.978440

Hu, W., Yen, Y. T., Singh, S., Kao, C. L., & Wu-Hsieh, B. A. (2012). SARS-CoV regulates

immune function-related gene expression in human monocytic cells. *Viral Immunology*. https://doi.org/10.1089/vim.2011.0099

Huang, B., Ling, R., Cheng, Y., Wen, J., Dai, Y., Huang, W., Zhang, S., Lu, X., Luo, Y., & Jiang, Y. (2020). Characteristics of the Coronavirus Disease 2019 and Related Therapeutic Options. *Molecular Therapy — Methods & Clinical Development*. https://doi.org/10.1016/j.omtm.2020.06.013

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, H., Li, X., Zha, D., Lin, H., Yang, L., Wang, Y., Xu, L., Wang, L., Lei, T., Zhou, Z., Xiao, Y.-F., Xin, H., Fu, M., & Qian, Y. (2023). SARS-CoV-2 E protein-induced THP-1 pyroptosis is reversed by Ruscogenin. *Biochemistry and Cell Biology*. https://doi.org/10.1139/bcb-2022-0359

Huang, S., Su, M. C., Tien, N., Huang, C. J., Lan, Y. C., Lin, C. S., Chen, C. H., & Lin, C. W. (2017). Epidemiology of Human Coronavirus NL63 Infection Among Hospitalized Patients With Pneumonia in Taiwan. *Journal of Microbiology Immunology and Infection*. https://doi.org/10.1016/j.jmii.2015.10.008

Huang, X., Dong, W., Milewska, A., Gołda, A., Qi, Y., Zhu, Q., Marasco, W. A., Baric, R. S., Sims, A. C., Pyrć, K., Li, W., & Sui, J. (2015). Human Coronavirus HKU1 Spike Protein Uses <i>O</I> -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *Journal of Virology*. https://doi.org/10.1128/jvi.00854-15

Hulswit, R. J., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., Ophorst, B., Kuppeveld, F. J. M. van, Boons, G., Bosch, B. J., Huizinga, E. G., & Groot, R. J. d. (2019). Human Coronaviruses OC43 and HKU1 Bind to 9- <i>O</I> -Acetylated Sialic Acids via a Conserved Receptor-Binding Site in Spike Protein Domain A. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1809667116

Hussain, M., Jabeen, N., Amanullah, A., Baig, A., Aziz, B., Shabbir, S., Raza, F., & Uddin,

N. (2020). Molecular Docking Between Human TMPRSS2 and SARS-CoV-2 Spike Protein: Conformation and Intermolecular Interactions. *Aims Microbiology*. https://doi.org/10.3934/microbiol.2020021

Hussain, S., Pan, J., Chen, Y., Yang, Y., Xu, J., Peng, Y., Wu, Y., Li, Z., Zhu, Y., Tien, P., & Guo, D. (2005). Identification of Novel Subgenomic RNAs and Noncanonical Transcription Initiation Signals of Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology*, *79*(9), 5288–5295. https://doi.org/10.1128/jvi.79.9.5288-5295.2005

Hyatt, D. R., Chen, G. L., LoCascio, P. F., Land, M., Larimer, F. W., & Hauser, L. (2010). Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-11-119

Ibrahim, I., Abdelmalek, D. H., Elshahat, M., & Elfiky, A. A. (2020). *COVID-19 Spike-Host Cell Receptor GRP78 Binding Site Prediction*. https://doi.org/10.21203/rs.2.24599/v1

Ihling, C. M., Schnitzler, P., Heinrich, N., Mangu, C., Sudi, L., Souares, A., Gies, S., Sié, A., Coulibaly, B., Ouédraogo, A. T., Mordmüller, B., Held, J., Adegnika, A. A., Fernandes, J. F., Eckerle, I., May, J., Hogan, B., Eibach, D., & Tabatabai, J. (2021). Molecular Epidemiology of Respiratory Syncytial Virus in Children in Sub-Saharan Africa. *Tropical Medicine & International Health*. https://doi.org/10.1111/tmi.13582

*International Committee on Taxonomy of Viruses: http://www.ictvonline.org/*. (2019). https://scholar.google.com/scholar_lookup?title=International+Committee+on+Taxonomy+of+Viruses+(ICTV).+Virus+Taxonomy:+2018b+Release&publication_year=2019&

Islam, A., Sayeed, M. A., Kalam, M. A., Ferdous, J., Rahman, M. K., Abedin, J., Islam, S., Shano, S., Saha, O., Shirin, T., & Hassan, M. M. (2021). *Molecular epidemiology of sars-cov-2 in diverse environmental samples globally*. *9*(8), 1696. https://doi.org/10.3390/MICROORGANISMS9081696

Israel, G. D. (1992). *Determining Sample Size 1 Strategies for Determining Sample Size*. *Figure 1*, 1–5. http://edis.ifas.ufl.edu.

Iwata-Yoshikawa, N., Kakizaki, M., Shiwa-Sudo, N., Okura, T., Tahara, M., Fukushi, S., Maeda, K., Kawase, M., Asanuma, H., Tomita, Y., Takayama, I., Matsuyama, S., Shirato, K., Suzuki, T., Nagata, N., & Takeda, M. (2022). Essential role of TMPRSS2 in SARS-CoV-2 infection in murine airways. *Nature Communications*, *13*, null. https://doi.org/10.1038/s41467-022-33911-8

Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2021). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews. Molecular Cell Biology*, *23*, 3–20. https://doi.org/10.1038/s41580-021-00418-x

Jaimes, J. A., Millet, J. K., & Whittaker, G. R. (2020). Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3581359

Jain, K., Gadhave, A., Florance, A., & Jain, K. (2020). Development of Genlisa$^{TM}$ SARS-CoV-2 ELISA (KBVH015-2 and KBVH015-8) for Estimation of Novel Coronavirus (nCoV/CoV-2) IgG Antibodies in Serum and Plasma. *International Journal of Research and Studies Publishing*. https://doi.org/10.29322/ijsrp.10.06.2020.p10238

Jakhar, R., & Gakhar, S. K. (2020). An Immunoinformatics Study to Predict Epitopes in the Envelope Protein of SARS-CoV-2. *Canadian Journal of Infectious Diseases and Medical Microbiology*. https://doi.org/10.1155/2020/7079356

Jawad, N. A., Al-Khafaji, Z. A., & Al-Tamimi, B. J. (2022). Molecular detection and Phylogenetic study role of ORF1ab in SARS-COV2 on patients in Hilla city. *International Journal of Health Sciences (IJHS)*, 4112–4120. https://doi.org/10.53730/ijhs.v6ns3.6693

Jeffers, S. A., Tusell, S. M., Gillim-Ross, L., Hemmila, E. M., Achenbach, J. E., Babcock, G. J., Thomas, W. D., Thackray, L. B., Young, M. J., Mason, R. J., Ambrosino, D. M., Wentworth, D. E., DeMartini, J. C., & Holmes, K. V. (2004). CD209L (L-Sign) Is a Receptor for Severe Acute Respiratory Syndrome Coronavirus. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0403812101

Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, *154*(3), 394–406.

https://doi.org/10.1111/imm.12889

Jia, H. P., Look, D. C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., & McCray, P. B. (2005). ACE2 Receptor Expression and Severe Acute Respiratory Syndrome Coronavirus Infection Depend on Differentiation of Human Airway Epithelia. *Journal of Virology*, *79*(23), 14614–14621. https://doi.org/10.1128/jvi.79.23.14614-14621.2005

Jiang, M., Zhang, G., Liu, H., Ding, P., Liu, Y., Tian, Y., Wang, Y., & Wang, A. (2021). Epitope Profiling Reveals the Critical Antigenic Determinants in SARS-CoV-2 RBD-Based Antigen. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.707977

Johansson, A. M., Malhotra, U., Kim, Y. G., Gomez, R., Krist, M. P., Wald, A., Koelle, D. M., & Kwok, W. W. (2021). Cross-Reactive and Mono-Reactive SARS-CoV-2 CD4+ T Cells in Prepandemic and COVID-19 Convalescent Individuals. *Plos Pathogens*. https://doi.org/10.1371/journal.ppat.1010203

Johnson, B. A., Zhou, Y., Lokugamage, K. G., Vu, M. N., Bopp, N. E., Crocquet-Valdes, P. A., Kalveram, B., Schindewolf, C., Liu, Y., Scharton, D., Plante, J. A., Xie, X., Aguilar, P. V, Weaver, S. C., Shi, P. Y., Walker, D. H., Routh, A., Plante, K. S., & Menachery, V. D. (2022). Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. *PLOS Pathogens*, *18*(6), e1010627–e1010627. https://doi.org/10.1371/journal.ppat.1010627

Joyjinda, Y., Rodpan, A., Chartpituck, P., Suthum, K., Yaemsakul, S., Cheun-Arom, T., Bunprakob, S., Olival, K. J., Stokes, M. M., Hemachudha, T., & Wacharapluesadee, S. (2019). First Complete Genome Sequence of Human Coronavirus HKU1 From a Nonill Bat Guano Miner in Thailand. *Microbiology Resource Announcements*. https://doi.org/10.1128/mra.01457-18

Jungreis, I., Nelson, C. W., Ardern, Z., Finkel, Y., Krogan, N. J., Sato, K., Ziebuhr, J., Stern-Ginossar, N., Pavesi, A., Firth, A. E., Gorbalenya, A. E., & Kellis, M. (2021). Conflicting and Ambiguous Names of Overlapping ORFs in the SARS-CoV-2 Genome: A Homology-Based Resolution. *Virology*. https://doi.org/10.1016/j.virol.2021.02.013

Kagucia, E. W., Ziraba, A., Nyagwange, J., Kutima, B., Kimani, M., Akech, D., Ng'oda, M., Sigilai, A., Mugo, D., Karanja, H. K., Karani, A., Toroitich, M. J., Karia, B., Otiende, M., Njeri, A., Aman, R., Amoth, P., Mwangangi, M., Kasera, K., … Agweyu, A. (2022). SARS-CoV-2 seroprevalence and implications for population immunity: Evidence from two Health and Demographic Surveillance System sites in Kenya, February-June 2022. *MedRxiv*. https://doi.org/10.1101/2022.10.10.22280824

Kalita, M. J., Dutta, K., Hazarika, G., Dutta, R., Kalita, S., Das, P. P., Sarma, M. P., Banu, S., Idris, M. G., Talukdar, A. K., Dutta, S., Sharma, A., & Medhi, S. (2021). *In-House Reverse Transcriptase Polymerase Chain Reaction for Detection of SARS-CoV2 With Increased Sensitivity*. https://doi.org/10.21203/rs.3.rs-408343/v1

Kandel, S., Hartzell, S., Ingold, A., Kennedy, J. L., & Ussery, D. W. (2023). Genomic Surveillance of SARS-CoV-2 Using Long-Range PCR Primers. *BioRxiv*. https://doi.org/10.1101/2023.07.10.548464

Kannian, P., Mahanathi, P., & Ashwini. (2020). *Concentration of the Cellular Material in the Nasopharyngeal Swabs Increases the Clinical Sensitivity of SARS-CoV2 RT-PCR*. https://doi.org/10.1101/2020.10.31.20218958

Karakus, G. S., Tastan, C., Kancagi, D. D., Yurtsever, B., Tumentemur, G., Demir, S., Turan, R. D., Abanuz, S., Cakirsoy, D., Seyis, U., Ozer, S., Elibol, O., Elek, M., Ertop, G., Arbak, S., Elmas, M. A., Hemsinlioglu, C., Kocagoz, A. S., Ng, O. H., … Ovali, E. (2021). Preclinical Efficacy and Safety Analysis of Gamma-Irradiated Inactivated SARS-CoV-2 Vaccine Candidates. *Scientific Reports*. https://doi.org/10.1038/s41598-021-83930-6

Karami, H., Sadeghi, K., Zadheidar, S., Saadatmand, F., Mirsalehi, N., Ardestani, N. H., Kalantari, S., Farahmand, M., Yavarian, J., & Mokhtari-Azad, T. (2023). Surveillance of Endemic Coronaviruses During the COVID-19 Pandemic in Iran, 2021–2022. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.13128

Karlsson, A. C., Humbert, M., & Buggert, M. (2020). The Known Unknowns of T Cell Immunity to COVID-19. *Science Immunology*. https://doi.org/10.1126/sciimmunol.abe8063

Kenyon, J. J., & Hall, R. M. (2013). Variation in the Complex Carbohydrate Biosynthesis

Loci of Acinetobacter Baumannii Genomes. *Plos One*. https://doi.org/10.1371/journal.pone.0062160

Kesheh, M. M., Hosseini, P., Soltani, S., & Zandi, M. (2021). An Overview on the Seven Pathogenic Human Coronaviruses. *Reviews in Medical Virology*. https://doi.org/10.1002/rmv.2282

Khairkhah, N., Aghasadeghi, M. R., Namvar, A., & Bolhassani, A. (2020). Design of Novel Multiepitope Constructs-Based Peptide Vaccine Against the Structural S, N and M Proteins of Human COVID-19 Using Immunoinformatics Analysis. *Plos One*. https://doi.org/10.1371/journal.pone.0240577

Khamadi, S. A., Opanda, S., Symekher, S. L., Limbaso, S. K., Langat, S., Cirindi, J. K., Mwangi, M., Mwikwabe, N., Okeyo, S., Koskei, E., Mutisya, J., Owaka, S., Nyunja, A., Koka, H., Wadegu, M., Chitechi, E., Achilla, R., Majanja, J. M., Kanyara, L., … Bulimo, W. (2022). Whole-genome sequence analysis reveals the circulation of multiple SARS-CoV-2 variants of concern in Nairobi and neighboring counties, Kenya between March and July 2021. *Virology Journal*, *19*(1). https://doi.org/10.1186/s12985-022-01895-y

Khamadi, S., Opanda, S., Symekher, S., Limbaso, S., Langat, S., Cirindi, J., Mwangi, M., Mwikwabe, N. M., Okeyo, S., Koskei, E., Mutisya, J., Owaka, S., Nyunja, A., Koka, H., Wadegu, M., Chitechi, E., Achilla, R., Majanja, J. M., Kanyara, L., … Bulimo, W. D. (2022). Whole-genome sequence analysis reveals the circulation of multiple SARS-CoV-2 variants of concern in Nairobi and neighboring counties, Kenya between March and July 2021. *Virology Journal*, *19*(1). https://doi.org/10.1186/s12985-022-01895-y

Khan, A., Khan, S. A., Zia, K., Altowyan, M. S., Barakat, A., & Ul-Haq, Z. (2022). Deciphering the Impact of Mutations on the Binding Efficacy of SARS-CoV-2 Omicron and Delta Variants With Human ACE2 Receptor. *Frontiers in Chemistry*. https://doi.org/10.3389/fchem.2022.892093

Khan, M. I., Khan, Z. A., Baig, M. H., Ahmad, I., Farouk, A.-E., Song, Y. G., & Dong, J.-J. (2020). Comparative Genome Analysis of Novel Coronavirus (SARS-CoV-2) From Different Geographical Locations and the Effect of Mutations on Major Target

Proteins: An in Silico Insight. *Plos One*. https://doi.org/10.1371/journal.pone.0238344

Khatri, R., Siddqui, G., Sadhu, S., Maithil, V., Vishwakarma, P., Lohiya, B., Goswami, A., Ahmed, S., Awasthi, A., & Samal, S. (2022). Intrinsic D614G and P681R/H mutations in SARS-CoV-2 VoCs Alpha, Delta, Omicron and viruses with D614G plus key signature mutations in spike protein alters fusogenicity and infectivity. *Medical Microbiology and Immunology*, *212*(1), 103–122. https://doi.org/10.1007/s00430-022-00760-7

Khreefa, Z., Barbier, M. T., Koksal, A. R., Love, G., & Valle, L. Del. (2023). Pathogenesis and Mechanisms of SARS-CoV-2 Infection in the Intestine, Liver, and Pancreas. *Cells*, *12*(2), 262. https://doi.org/10.3390/cells12020262

Khrenova, M. G., Nikiforova, L. A., Grabovenko, F. I., Orlova, N. A., Sinegubova, M., Kolesov, D., Zavyalova, E., Spiridonova, V. A., Zatsepin, T. S., & Zvereva, M. I. (2022). *In Vitro Selection of an Aptamer Targeting SARS-CoV-2 Spike Protein With Nanopore Sequence Identification Reveals Discrimination Between the Authentic Strain and Omicron*. https://doi.org/10.26434/chemrxiv-2022-d9gcs

Kibria, K. M. K., Faruque, M. O., Islam, M. N., Ullah, H., Mahmud, S., Miah, M., & Saleh, A. S. A. (2022). A Conserved Subunit Vaccine Designed Against SARS-CoV-2 Variants Showed Evidence in Neutralizing the Virus. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-022-11988-x

Kim, D.-W., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*. https://doi.org/10.1016/j.cell.2020.04.011

Kim, J., Cruz, J. d. l., Lam, K., Ng, H. L., Daar, E. S., Balamurugan, A., & Yang, O. O. (2018). $CD8^{+}$Cytotoxic T Lymphocyte Responses and Viral Epitope Escape in Acute HIV-1 Infection. *Viral Immunology*. https://doi.org/10.1089/vim.2018.0040

Kimita, G., Nyataya, J., Omuseni, E., Sigei, F., Lemtudo, A., Muthanje, E., Andika, B., Liyai, R., Githii, R., Masakwe, C., Ochola, S., Awinda, G., Kifude, C. M., Mutai, B., Gatata, R. M., & Waitumbi, J. N. (2022a). Temporal lineage replacements and dominance of imported variants of concern during the COVID-19 pandemic in Kenya. *Communications Medicine*, *2*(1). https://doi.org/10.1038/s43856-022-00167-8

Kimita, G., Nyataya, J., Omuseni, E., Sigei, F., Lemtudo, A., Muthanje, E., Andika, B., Liyai, R., Githii, R., Masakwe, C., Ochola, S., Awinda, G., Kifude, C., Mutai, B., Gatata, R. M., & Waitumbi, J. (2022b). Temporal lineage replacements and dominance of imported variants of concern during the COVID-19 pandemic in Kenya. *Communications Medicine*, *2*(1). https://doi.org/10.1038/s43856-022-00167-8

King, A. M. Q., Lefkowitz, E., Adams, M. J., & Carstens, E. B. (2012). Virus Taxonomy. In *Virus Taxonomy*. https://doi.org/10.1016/b978-012465330-6/50022-1

Kiyotani, K., Toyoshima, Y., Nemoto, K., & Nakamura, Y. (2020). Bioinformatic Prediction of Potential T Cell Epitopes for SARS-Cov-2. *Journal of Human Genetics*. https://doi.org/10.1038/s10038-020-0771-5

Kiyuka, P. K., Agoti, C. N., Munywoki, P. K., Njeru, R., Bett, A., Otieno, J. R., Otieno, G. P., Kamau, E., Clark, T. G., Hoek, L. V, Kellam, P., Nokes, D. J., & Cotten, M. (2018). Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya. *The Journal of Infectious Diseases*. https://doi.org/10.1093/infdis/jiy098

Klompus, S., Leviatan, S., Vogl, T., Kalka, I., Godneva, A., Shinar, E., Weinberger, A., & Segal, E. (2020). *Cross-Reactive Antibody Responses Against SARS-CoV-2 and Seasonal Common Cold Coronaviruses*. https://doi.org/10.1101/2020.09.01.20182220

Klose, A., Kosoy, G., & Miller, B. L. (2023). Arrayed Imaging Reflectometry Monitoring of Anti-Viral Antibody Production Throughout Vaccination and Breakthrough Covid-19. *Plos One*. https://doi.org/10.1371/journal.pone.0277846

Knight, G. M., Clarkson, M., & Silva, T. I. d. (2018). Potential Impact of Influenza Vaccine Roll-Out on Antibiotic Use in Africa. *Journal of Antimicrobial Chemotherapy*. https://doi.org/10.1093/jac/dky172

Kono, M., Tatsumi, K., Imai, A. M., Saito, K., Kuriyama, T., & Shirasawa, H. (2008). Inhibition of Human Coronavirus 229E Infection in Human Epithelial Lung Cells (L132) by Chloroquine: Involvement of P38 MAPK and ERK. *Antiviral Research*. https://doi.org/10.1016/j.antiviral.2007.10.011

Kopecky-Bromberg, S. A., Martínez-Sobrido, L., Frieman, M. B., Baric, R., & Palese, P.

(2007). Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *Journal of Virology*. https://doi.org/10.1128/jvi.01782-06

Korber, B., Hraber, P., Wagh, K., & Hahn, B. H. (2017). Polyvalent Vaccine Approaches to Combat <scp>HIV</Scp>-1 Diversity. *Immunological Reviews*. https://doi.org/10.1111/imr.12516

Korneenko, E. V, Samoilov, A., Artyushin, I. V, Yusefovich, A. P., Dolotova, S. M., Klyuchnikova, E. O., Sbarzaglia, V., Gladkikh, A., Dedkov, V. G., & Speranskaya, A. S. (2022). Alphacoronaviruses detected in fecal samples of bats captured in Moscow and Rostov-on-Don in 2021. *Medicinskij Akademičeskij Žurnal*, *2*(2), 203–208. https://doi.org/10.17816/maj108718

Kresch, E., Achua, J. K., Saltzman, R. G., Khodamoradi, K., Arora, H., Ibrahim, E., Kryvenko, O. N., Almeida, V. W., Firdaus, F., Hare, J. M., & Ramasamy, R. (2021). COVID-19 Endothelial Dysfunction Can Cause Erectile Dysfunction: Histopathological, Immunohistochemical, and Ultrastructural Study of the Human Penis. *The World Journal of Men S Health*. https://doi.org/10.5534/wjmh.210055

Kubota-Koketsu, R., Terada, Y., Yunoki, M., Sasaki, T., Nakayama, E. E., Kamitani, W., & Shioda, T. (2020). Neutralizing and Binding Activities Against <scp>SARS-CoV</Scp>-1/2, <scp>MERS-CoV</Scp>, and Human Coronaviruses <scp>229E</Scp> and <scp>OC43</Scp> by Normal Human Intravenous Immunoglobulin Derived From Healthy Donors in Japan. *Transfusion*. https://doi.org/10.1111/trf.16161

Kumar, Suresh, Karuppanan, K., & Subramaniam, G. (2022). Omicron (BA.1) and Sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 Spike Infectivity and Pathogenicity: A Comparative Sequence and Structural-based Computational Assessment. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.27927

Kumar, Sushant, Kumari, K., & Azad, G. K. (2022). Immunoinformatics Study of SARS-CoV-2 Nucleocapsid Phosphoprotein Identifies Promising Epitopes With Mutational Implications. *Moscow University Biological Sciences Bulletin*. https://doi.org/10.3103/s0096392522040125

Kuzmina, A., Khalaila, Y., Voloshin, O., Keren-Naus, A., Boehm-Cohen, L., Raviv, Y., Shemer-Avni, Y., Rosenberg, E., & Taube, R. (2021). SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell Host & Microbe*, *29*, 522-528.e2. https://doi.org/10.1016/j.chom.2021.03.008

Kuzmina, A., Korovin, D., lass, I. C., Atari, N., Ottolenghi, A., Hu, P., Mandelboim, M., Rosental, B., Rosenberg, E., Diaz-Griffero, F., & Taube, R. (2023). Changes within the P681 residue of spike dictate cell fusion and syncytia formation of Delta and Omicron variants of SARS-CoV-2 with no effects on neutralization or infectivity. *Heliyon*, *9*, e16750–e16750. https://doi.org/10.1016/j.heliyon.2023.e16750

Kuzmina, A., Wattad, S., Khalaila, Y., Ottolenghi, A., Rosental, B., Engel, S., Rosenberg, E., & Taube, R. (2021). SARS CoV-2 Delta variant exhibits enhanced infectivity and a minor decrease in neutralization sensitivity to convalescent or post-vaccination sera. *IScience*, *24*, null. https://doi.org/10.1016/j.isci.2021.103467

Kwok, W. W., Tan, V., Gillette, L., Littell, C. T., Soltis, M. A., LaFond, R. E., Yang, J., James, E. A., & DeLong, J. H. (2012). Frequency of Epitope-Specific Naive CD4+ T Cells Correlates With Immunodominance in the Human Memory Repertoire. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.1102190

Laine, L., Skön, M., Väisänen, E., Julkunen, I., & Österlund, P. (2022). SARS-CoV-2 variants Alpha, Beta, Delta and Omicron show a slower host cell interferon response compared to an early pandemic variant. *Frontiers in Immunology*, *13*, null. https://doi.org/10.3389/fimmu.2022.1016108

Lambert, N., Ndishimye, P., & Zakham, F. (2020). *Genomic Diversity Analysis of SARS-CoV-2 Genomes in Rwanda*. https://doi.org/10.1101/2020.12.14.422793

Lambisia, A. W., Mudhune, G. H., Morobe, J. M., Mohammed, K. S., Makori, T. O., Ndwiga, L., Mburu, M. W., Moraa, E. O., Musyoki, J., Murunga, N., Waliaula, I. N., Mumelo, A. K., Bejon, P., Ochola-Oyier, L. I., Githinji, G., Nokes, J., & Agoti, C. N. (2022). Temporal distribution and clinical characteristics of the Alpha, Delta and Omicron SARS-CoV-2 variants of concern in Laikipia, Kenya: institutional and community-based genomic surveillance. *Wellcome Open Research*, *7*, 235.

https://doi.org/10.12688/wellcomeopenres.18306.1

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology*, *20*(5), 270–284. https://doi.org/10.1038/s41579-022-00713-0

Landherr, M., O-Uchi, J., Cypress, M. W., Chaput, I., Jhun, B. S., & Polina, I. (2023). Impact of ORF3a-Q57H variant of SARS-CoV-2 on apoptotic signaling cascades. *Physiology*, *38*(S1). https://doi.org/10.1152/physiol.2023.38.s1.5732631

Laris-González, A., Avilés-Robles, M., Domínguez-Barrera, C., Parra-Ortega, I., Sánchez-Huerta, J. L., Ojeda-Diezbarroso, K., Bonilla-Pellegrini, S., Olivar-López, V., Chávez-López, A., & Jiménez-Juárez, R. N. (2021). Influenza vs. COVID-19: Comparison of Clinical Characteristics and Outcomes in Pediatric Patients in Mexico City. *Frontiers in Pediatrics*. https://doi.org/10.3389/fped.2021.676611

Lau, S. K. P., Woo, P. C. Y., Wong, B. H., Tsoi, H.-W., Woo, G. K. S., Poon, R. W., Chan, K.-H., Wei, W. I., Peiris, M., & Yuen, K.-Y. (2004). Detection of Severe Acute Respiratory Syndrome (SARS) Coronavirus Nucleocapsid Protein in SARS Patients by Enzyme-Linked Immunosorbent Assay. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.42.7.2884-2889.2004

Lau, S. K. P., Woo, P. C. Y., Yip, C. C., Tse, H., Tsoi, H.-W., Cheng, V. C., Lee, P., Tang, B. S., Cheung, C. H. Y., Lee, R. A., So, L., Lau, Y., Chan, K., & Yuen, K. (2006). Coronavirus HKU1 and Other Coronavirus Infections in Hong Kong. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.02614-05

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*. https://doi.org/10.7326/M20-0504

Le Bon, A., & Tough, D. F. (2002). Links between innate and adaptive immunity via type I interferon. In *Current Opinion in Immunology*. https://doi.org/10.1016/S0952-7915(02)00354-0

Lee, C. Y., Huang, C. H., Rastegari, E., Rengganaten, V., Liu, P. C., Tsai, P.-H., Chin, Y.-F., Wu, J., Chiou, S. H., Teng, Y., Lee, C. W., Liang, Y., Chen, A. Y., Hsu, S. C.,

Hung, Y. J., Sun, J.-R., Chien, C.-S., & Chien, Y. (2021). Tumor Necrosis Factor-Alpha Exacerbates Viral Entry in SARS-CoV2-Infected iPSC-Derived Cardiomyocytes. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms22189869

Lee, G., Budhathoki, S., Lee, G. Y., Oh, K. J., Ham, Y. K., Kim, Y. J., Lim, Y. R., Hoang, P. T., Lee, Y.-J., Lim, S. W., Kim, J. M., Cho, S., Kim, T. H., Song, J. W., Lee, S., & Kim, W. K. (2021). Broad-Spectrum Antiviral Activity of 3D8, a Nucleic Acid-Hydrolyzing Single-Chain Variable Fragment (scFv), Targeting SARS-CoV-2 and Multiple Coronaviruses in Vitro. *Viruses*. https://doi.org/10.3390/v13040650

Lee, S. J., Kim, Y. J., & Ahn, D. G. (2022). Distinct Molecular Mechanisms Characterizing Pathogenesis of SARS-CoV-2. *Journal of Microbiology and Biotechnology*, *32*(9), 1073–1085. https://doi.org/10.4014/jmb.2206.06064

Lepiller, Q., Barth, H., Lefèbvre, F., Herbrecht, R., Lutz, P., Kessler, R., Fafi-Kremer, S., & Stoll-Keller, F. (2013). High Incidence but Low Burden of Coronaviruses and Preferential Associations Between Respiratory Viruses. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.01078-13

Li, B., Si, H.-R., Zhu, Y., Yang, X.-L., Anderson, D. E., Shi, Z.-L., Wang, L.-F., & Zhou, P. (2020). Discovery of Bat Coronaviruses through Surveillance and Probe Capture-Based Next-Generation Sequencing. *MSphere*. https://doi.org/10.1128/msphere.00807-19

Li, Chenlu, Wu, X., Sheridan, S. C., Lee, J., Yin, J., & Han, J. (2021). Interaction of Climate and Socio-Ecological Environment Drives the Dengue Outbreak in Epidemic Region of China. *Plos Neglected Tropical Diseases*. https://doi.org/10.1371/journal.pntd.0009761

Li, Chunhua, Li, W., Esesarte, E. L. d., Guo, H., Elzen, P. v. d., Aarts, E., Born, E. v. d., Rottier, P. J. M., & Bosch, B.-J. (2017). Cell Attachment Domains of the Porcine Epidemic Diarrhea Virus Spike Protein Are Key Targets of Neutralizing Antibodies. *Journal of Virology*. https://doi.org/10.1128/jvi.00273-17

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*. https://doi.org/10.1146/annurev-virology-110615-042301

Li, G., Fan, Y., Lai, Y., Han, T., Li, Z., Zhou, P., Pan, P., Wang, W., Hu, D., Liu, X., Zhang, Q., & Wu, J. (2020). Coronavirus infections and immune responses. In *Journal of Medical Virology* (Vol. 92, Issue 4, pp. 424–432). John Wiley and Sons Inc. https://doi.org/10.1002/jmv.25685

Li, J., Zhang, S., Tan, S., Zheng, B., & Gao, G. F. (2011). Revival of the Identification of Cytotoxic T-Lymphocyte Epitopes for Immunological Diagnosis, Therapy and Vaccine Development. *Experimental Biology and Medicine*. https://doi.org/10.1258/ebm.2010.010278

Li, M., Zeng, J., Li, R., Wen, Z., Cai, Y., Wallin, J. J., Shu, Y., Du, X., & Sun, C. (2021). Rational Design of a Pan-Coronavirus Vaccine Based on Conserved CTL Epitopes. *Viruses*. https://doi.org/10.3390/v13020333

Li, P., Rajpoot, S., Saqib, U., Yu, P., Li, Y., Li, Y., Ma, Z., Baig, M. S., & Pan, Q. (2021). Comparative Assessment of Favipiravir and Remdesivir Against Human Coronavirus NL63 in Molecular Docking and Cell Culture Models. *Scientific Reports*. https://doi.org/10.1038/s41598-021-02972-y

Li, R., Li, H., & Chen, W. (2020). Silico Analysis of Interaction Between Full-Length SARS-CoV2 S Protein With Human Ace2 Receptor: Modelling, Docking, MD Simulation. *Biophysical Chemistry*. https://doi.org/10.1016/j.bpc.2020.106472

Li, S., Yan, R., Xu, J., Zhao, S., Ma, X., Sun, Q., Zhang, M., Liu, Y., Liu, J. J., Chen, L., Li, S., Xu, K., & Ge, L. (2021). *A New Type of ERGIC-ERES Membrane Contact Mediated by TMED9 and SEC12 Is Required for Autophagosome Biogenesis*. https://doi.org/10.1101/2021.05.19.444771

Li, W., Joshi, M., Singhania, S., Ramsey, K. H., & Murthy, A. K. (2014). Peptide Vaccine: Progress and Challenges. *Vaccines*. https://doi.org/10.3390/vaccines2030515

Li, X., Hou, P., Ma, W., Wang, X., Wang, H., Yu, Z., Chang, H., Wang, T., Jin, S., Wang, X., Wang, W., Zhao, Y., Zhao, Y., Xu, C., Ma, X., Gao, Y., & He, H. (2021). SARS-CoV-2 ORF10 suppresses the antiviral innate immune response by degrading MAVS through mitophagy. *Cellular & Molecular Immunology*, *29*, 29. https://doi.org/10.1038/S41423-021-00807-4

Li, Y., Merbah, M., Wollen-Roberts, S. E., Beckman, B., Mdluli, T., Swafford, I., Mayer, S. V, King, J., Corbitt, C., Currier, J. R., Liu, H., Esber, A., Pinyakorn, S., Parikh, A., Francisco, L. V, Phanuphak, N., Maswai, J., Owuoth, J., Kibuuka, H., … Rolland, M. (2022). Coronavirus Antibody Responses Before COVID-19 Pandemic, Africa and Thailand. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid2811.221041

Li, Z., Tomlinson, A. C. A., Wong, A. H. M., Zhou, D., Desforges, M., Talbot, P. J., Benlekbir, S., Rubinstein, J. L., & Rini, J. M. (2019). The Human Coronavirus HCoV-229E S-Protein Structure and Receptor Binding. *Elife*. https://doi.org/10.7554/elife.51230

Lim, H. X., Lim, J., Jazayeri, S. D., Poppema, S., & Poh, C. L. (2021). Development of Multi-Epitope Peptide-Based Vaccines Against SARS-CoV-2. *Biomedical Journal*. https://doi.org/10.1016/j.bj.2020.09.005

Lim, T., Delorey, M. J., Bestul, N., Johannson, M. A., Reed, C., Hall, A. J., Fry, A. M., Edens, C., Semenova, V. A., Li, H., Browning, P., Desai, R., Epperson, M., Jia, T., Thornburg, N. J., Schiffer, J., & Havers, F. (2021). Changes in Severe Acute Respiratory Syndrome Coronavirus 2 Seroprevalence Over Time in 10 Sites in the United States, March-August, 2020. *Clinical Infectious Diseases*, *73*(10), 1831–1839. https://doi.org/10.1093/CID/CIAB185

Lin, X., Sha, Z., Trimpert, J., Kunec, D., Jiang, C., Xiong, Y., Xu, B., Zhu, Z., Xue, W., & Wu, H. (2023). *T492I Mutation Alters SARS-CoV-2 Properties via Modulating Viral Non-Structural Proteins*. https://doi.org/10.1101/2023.01.15.524090

Lippi, G., & Henry, B. (2021). The landscape of SARS-CoV-2 genomic mutations. *Journal of Laboratory and Precision Medicine*, *null*, null. https://doi.org/10.21203/rs.3.rs-1343942/v1

Liu, D., Chen, C., Chen, D., Zhu, A., Li, F., Zhuang, Z., Mok, C. K. P., Dai, J., Li, X., Jin, Y., Chen, Z., Sun, J., Wang, Y., Li, Y., Zhang, Y., Wen, L., Zhang, Z., Zhuo, J., Wang, J., … Zhao, J. (2023). Mouse Models Susceptible to HCoV-229E and HCoV-NL63 and Cross Protection From Challenge With SARS-CoV-2. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2202820120

Liu, G., & Gack, M. U. (2022). Insights into pandemic respiratory viruses: manipulation of

the antiviral interferon response by SARS-CoV-2 and influenza A virus. *Current Opinion in Immunology*, *78*, 102252. https://doi.org/10.1016/j.coi.2022.102252

Liu, H., Bai, Y., Zhang, X., Gao, T., Liu, Y., Li, E., Wang, X., Cao, Z., Zhu, L., Dong, Q., Hu, Y., Wang, G. F., Song, C., Niu, X., Zheng, T., Wang, D., Liu, Z., Jin, Y., Li, P., … Liu, X. (2022). SARS-CoV-2 N Protein Antagonizes Stress Granule Assembly and IFN Production by Interacting with G3BPs to Facilitate Viral Replication. *Journal of Virology*, *96*(12). https://doi.org/10.1128/jvi.00412-22

Liu, Yang, Liu, J., Xia, H., Zhang, X., Fontes-Garfias, C. R., Swanson, K. A., Cai, H., Sarkar, R., Chen, W., Cutler, M., Cooper, D., Weaver, S. C., Muik, A., Sahin, U., Jansen, K. U., Xie, X., Dormitzer, P. R., & Shi, P.-Y. (2021). Neutralizing Activity of BNT162b2-Elicited Serum. *New England Journal of Medicine*, *384*(15), 1466–1468. https://doi.org/10.1056/nejmc2102017

Liu, Ying, Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*. https://doi.org/10.1093/jtm/taaa021

Liu, Z., Xu, W., Xia, S., Gu, C., Wang, X., Wang, Q., Zhou, J., Wu, Y., Cai, X., Qu, D., Ying, T., Xie, Y., Lu, L., Yuan, Z., & Jiang, S. (2020). RBD-Fc-based COVID-19 Vaccine Candidate Induces Highly Potent SARS-CoV-2 Neutralizing Antibody Response. *Signal Transduction and Targeted Therapy*. https://doi.org/10.1038/s41392-020-00402-5

Llanes, A., Restrepo, C. M., Caballero, Z., Rajeev, S., Kennedy, M., & Lleonart, R. (2020). Betacoronavirus Genomes: How Genomic Information Has Been Used to Deal With Past Outbreaks and the COVID-19 Pandemic. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms21124546

Lobo-Rodriguez, C., López-Calleja, A. M., Luna, P. M. de los R., & López-Esteban, A. (2023). Comparison of the first and sixth waves of the SARS-COV-2 coronavirus pandemic in patients with immune-mediated disease. *Reumatología Clínica*. https://doi.org/10.1016/j.reumae.2023.02.011

Lobo-Rodríguez, C., López-Calleja, A. M., Sánchez-Vázquez, S., & López-Esteban, A. (2023). [Comparison of the first and sixth waves of the SARS-COV-2 coronavirus

pandemic in patients with immune-mediated disease.]. *Reumatología Clínica*. https://doi.org/10.1016/j.reuma.2023.02.008

Loevinsohn, G., Hardick, J., Sinywimaanzi, P., Fenstermacher, K., Shaw-Saliba, K., Monze, M., Gaydos, C. A., Rothman, R. E., Pekosz, A., Thuma, P. E., & Sutcliffe, C. G. (2021). Respiratory Pathogen Diversity and Co-Infections in Rural Zambia. *International Journal of Infectious Diseases*. https://doi.org/10.1016/j.ijid.2020.10.054

Lontok, E., Corse, E., & Machamer, C. E. (2004). Intracellular Targeting Signals Contribute to Localization of Coronavirus Spike Proteins Near the Virus Assembly Site. *Journal of Virology*. https://doi.org/10.1128/jvi.78.11.5913-5922.2004

López-Farfán, D., Yerbanga, R. S., Parres-Mercader, M., Torres-Puente, M., Gómez-Navarro, I., Sanou, D. M. S., Yao, A., Ouédraogo, J. B., Comas, I., Irigoyen, N., & Gómez-Díaz, E. (2022). Prevalence of SARS-CoV-2 and co-infection with malaria during the first wave of the pandemic (the Burkina Faso case). *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1048404

Lou, F., Li, M., Pang, Z., Jiang, L., Guan, L., Tian, L., Hu, J., Fan, J., & Fan, H. (2021). Understanding the Secret of SARS-CoV-2 Variants of Concern/Interest and Immune Escape. *Frontiers in Immunology*, *12*, null. https://doi.org/10.3389/fimmu.2021.744242

Lü, S., Xie, X. x., Zhao, L., Wang, B., Zhu, J., Yang, T., Yang, G., Ji, M., Lv, C. p., Xue, J., Dai, E. H., Fu, X. m., Liu, D. q., Zhang, L., Hou, S. j., Yu, X., Wang, Y. l., Gao, H. x., Shi, X. h., … Liu, R. t. (2021). The Immunodominant and Neutralization Linear Epitopes for SARS-CoV-2. *Cell Reports*. https://doi.org/10.1016/j.celrep.2020.108666

Lu, X., Zhang, L., Du, H., Zhang, J., Li, Y. Y., Qu, J., Zhang, W., Wang, Y., Bao, S., Li, Y., Wu, C., Liu, H., Liu, D., Shao, J., Peng, X., Yang, Y., Liu, Z., Xiang, Y., Zhang, F., … Wong, G. W. K. (2020). SARS-CoV-2 Infection in Children. *New England Journal of Medicine*. https://doi.org/10.1056/nejmc2005073

Luan, B., & Huynh, T. (2021). Insights into SARS-CoV-2's Mutations for Evading Human Antibodies: Sacrifice and Survival. *Journal of Medicinal Chemistry*. https://doi.org/10.1021/ACS.JMEDCHEM.1C00311

Lubinski, B., Fernandes, M., Frazier, L., Tang, T., Daniel, S., Diel, D. G., Jaimes, J. A., & Whittaker, G. R. (2022). Functional Evaluation of the P681H Mutation on the Proteolytic Activation of the SARS-CoV-2 Variant B.1.1.7 (Alpha) Spike. *Iscience*. https://doi.org/10.1016/j.isci.2021.103589

Luo, Y., Yu, F., Zhou, M., Liu, Y., Xia, B., Zhang, X., Li, J., Zhang, J., Du, Y., Li, R., Wu, L., Zhang, X., Pan, T., Guo, D., Peng, T., & Zhang, H. (2021). Engineering a Reliable and Convenient SARS-CoV-2 Replicon System for Analysis of Viral RNA Synthesis and Screening of Antiviral Inhibitors. *Mbio*. https://doi.org/10.1128/mbio.02754-20

Lutomski, C. A., El-Baba, T. J., Bolla, J. R., & Robinson, C. V. (2021). *Multiple Roles of SARS-CoV-2 N Protein Facilitated by Proteoform-Specific Interactions with RNA, Host Proteins, and Convalescent Antibodies. 1*(8), 1147–1157. https://doi.org/10.1021/JACSAU.1C00139

Ma, H., Zeng, W., Meng, X., Huang, X., Yu, Y., Zhao, D., Zhou, P., Wang, X., Zhao, C., Sun, Y., Wang, P., Ou, H., Hu, X., Xiang, Y., & Jin, T. (2021). Potent Neutralization of SARS-CoV-2 by Hetero-Bivalent Alpaca Nanobodies Targeting the Spike Receptor-Binding Domain. *Journal of Virology*. https://doi.org/10.1128/jvi.02438-20

Madsen, J. L., Kroghsbo, S., Madsen, C. B., Pozdnyakova, I., Barkholt, V., & Bøgh, K. L. (2014). The Impact of Structural Integrity and Route of Administration on the Antibody Specificity Against Three Cow's Milk Allergens - a Study in Brown Norway Rats. *Clinical and Translational Allergy*. https://doi.org/10.1186/2045-7022-4-25

Maghsood, F., Shokri, M.-R., Jeddi-Tehrani, M., Mt, R., Salimi, Kardar, G. A., Zarnani, A.-H., Amiri, M. M., & Shokri, F. (2021). *Identification of immunodominant epitopes on nucleocapsid and spike proteins of the SARS-CoV-2 in Iranian COVID-19 patients*. https://doi.org/10.22541/AU.162223663.36566742/V1

Malonis, R. J., Lai, J. R., & Vergnolle, O. (2019). Peptide-Based Vaccines: Current Progress and Future Challenges. *Chemical Reviews*. https://doi.org/10.1021/acs.chemrev.9b00472

Mao, S., Cai, X.-F., Niu, S., Wei, J., Jiang, N., Deng, H., Wang, W., Zhang, J., Shen, S., Ma, Y., Wu, X., Peng, Q., Huang, A., & Wang, D. (2023). TRIM21 Promotes

Ubiquitination of SARS-CoV-2 Nucleocapsid Protein to Regulate Innate Immunity. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.28719

Marchenko, V. Y., Danilenko, A. V, Kolosova, N. G., Bragina, M., Molchanova, M. S., Bulanovich, Y., Gorodov, V., Leonov, S., Gudymo, A. S., Onkhonova, G. S., Svyatchenko, S. V, & Ryzhikov, A. B. (2022). *Diversity of Gammacoronaviruses and Deltacoronaviruses in Wild Birds and Poultry in Russia*. https://doi.org/10.21203/rs.3.rs-1827146/v1

Markosian, C., Staquicini, D. I., Dogra, P., Dodero-Rojas, E., Lubin, J. H., Tang, F. H. F., Smith, T. L., Contessoto, V. G., Libutti, S. K., Wang, Z., Cristini, V., Khare, S. D., Whitford, P. C., Burley, S. K., Onuchic, J. N., Pasqualini, R., & Arap, W. (2022). Genetic and Structural Analysis of SARS-CoV-2 Spike Protein for Universal Epitope Selection. *Molecular Biology and Evolution*, *39*(5). https://doi.org/10.1093/molbev/msac091

Mason, R. J. (2020). Pathogenesis of COVID-19 from a cell biology perspective. In *European Respiratory Journal* (Vol. 55, Issue 4). European Respiratory Society. https://doi.org/10.1183/13993003.00607-2020

Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S. I., Dan, J. M., Burger, Z. C., Rawlings, S. A., Smith, D. M., Phillips, E., Mallal, S., Lammers, M., Rubiro, P., Quiambao, L., Sutherland, A., Yu, E. D., Antunes, R. d. S., Greenbaum, J. A., Frazier, A., … Weiskopf, D. (2020). Selective and Cross-Reactive SARS-CoV-2 T Cell Epitopes in Unexposed Humans. *Science*. https://doi.org/10.1126/science.abd3871

Matoba, Y., Abiko, C., Ikeda, T., Aoki, Y., Suzuki, Y., Yahagi, K., Matsuzaki, Y., Itagaki, T., Katsushima, F., Katsushima, Y., & Mizuta, K. (2015). Detection of the Human Coronavirus 229E, HKU1, NL63, and OC43 Between 2010 and 2013 in Yamagata, Japan. *Japanese Journal of Infectious Diseases*. https://doi.org/10.7883/yoken.jjid.2014.266

Matthews, P. C., Koyanagi, M., Kløverpris, H. N., Harndahl, M., Stryhn, A., Akahoshi, T., Gatanaga, H., Oka, S., Molina, C. J., Ponce, H. V, Ríos, S., Cole, D. K., Carlson, J. M., Payne, R., Ogwu, A., Bere, A., Ndung'u, T., Gounder, K., Chen, F., … Goulder, P. (2012). Differential Clade-Specific HLA-B*3501 Association With HIV-1 Disease

Outcome Is Linked to Immunogenicity of a Single Gag Epitope. *Journal of Virology*. https://doi.org/10.1128/jvi.01381-12

McBride, C. E., Li, J., & Machamer, C. E. (2007). The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction With Membrane Protein. *Journal of Virology*. https://doi.org/10.1128/jvi.02146-06

McCallum, M., Walls, A., Sprouse, K. R., Bowen, J. E., Rosen, L. E., Dang, H., Marco, A. De, Franko, N. M., Tilles, S., Logue, J., Miranda, M. C., Ahlrichs, M., Carter, L., Snell, G., Pizzuto, M., Chu, H., Voorhis, W. V. Van, Corti, D., & Veesler, D. (2021). Molecular basis of immune evasion by the delta and kappa SARS-CoV-2 variants. *BioRxiv*, *null*, eabl8506. https://doi.org/10.1126/science.abl8506

McGrath, M., Xue, Y., Dillen, C. A., Oldfield, L. M., Assad-Garcia, N., Zaveri, J., Singh, N., Baracco, L., Taylor, L., Vashee, S., & Frieman, M. B. (2022). SARS-CoV-2 variant spike and accessory gene mutations alter pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(37). https://doi.org/10.1073/pnas.2204717119

Mchomvu, E., Mbunda, G., Simon, N. R., Kitila, F., Temba, Y., Msumba, I., Namamba, J., Kilindimo, S. S., Mgubike, H., Gingo, W., Hatz, C., Paris, D., Weisser, M., & Rohacek, M. (2019). Diagnoses Made in an Emergency Department in Rural Sub-Saharan Africa. *Swiss Medical Weekly*. https://doi.org/10.4414/smw.2019.20018

McIntosh, K., Becker, W. B., & Chanock, R. M. (1967). Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.58.6.2268

Mclean, G., Kamil, J., Lee, B., Moore, P., Schulz, T., Muik, A., Şahin, U., Türeci, Ö., & Pather, S. (2022). The Impact of Evolving SARS-CoV-2 Mutations and Variants on COVID-19 Vaccines. *MBio*, *13*, null. https://doi.org/10.1128/mbio.02979-21

Meega, R., & Kumar, R. (2022). *Emergence of SARS-CoV-2 Variant of Concern Omicron: Biological Features and Genomic Concern*. https://doi.org/10.56950/itrx2370

Mehregan, A., Pérez-Conesa, S., Zhuang, Y., Elbahnsi, A., Pasini, D., Lindahl, E., Howard, R., Ulens, C., & Delemotte, L. (2022). Probing effects of the SARS-CoV-2 E protein on membrane curvature and intracellular calcium. *Biochimica Et Biophysica Acta - Biomembranes*, *1864*(10), 183994. https://doi.org/10.1016/j.bbamem.2022.183994

Mei, X. Q., Qin, P., Yang, Y.-L., Liao, M., Liang, Q., Zhao, Z., Shi, F.-S., Wang, B., & Huang, Y.-W. (2022). First evidence that an emerging mammalian alphacoronavirus is able to infect an avian species. *Transboundary and Emerging Diseases*, *69*(5). https://doi.org/10.1111/tbed.14535

Mena, E. L., Donahue, C. J., Vaites, L. P., Li, J., Róna, G., Róna, G., O'Leary, C., Lignitto, L., Miwatani-Minter, B., Paulo, J. A., Dhabaria, A., Ueberheide, B., Gygi, S. P., Pagano, M., Pagano, M., Harper, J. W., Davey, R. A., & Elledge, S. J. (2021). ORF10-Cullin-2-ZYG11B complex is not required for SARS-CoV-2 infection. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(17). https://doi.org/10.1073/PNAS.2023157118

Meng, B., Datir, R., Choi, J., Bradley, J., Smith, K. G. C., Lee, J., & Gupta, R. K. (2022). SARS-CoV-2 spike N-terminal domain modulates TMPRSS2-dependent viral entry and fusogenicity. *Cell Reports*, *40*, null. https://doi.org/10.1101/2022.05.07.491004

Mercatelli, D., Holding, A. N., & Giorgi, F. M. (2021). Web tools to fight pandemics: The COVID-19 experience. In *Briefings in Bioinformatics* (Vol. 22, Issue 2, pp. 690–700). https://doi.org/10.1093/bib/bbaa261

Merdekawati, F., & Nurhayati, B. (2023). DESAIN PRIMER GEN PENGKODE RNA DEPENDENT RNA POLIMERASE (RdRp) UNTUK DETEKSI SARS COV2 DENGAN MENGGUNAKAN REAL TIME POLYMERASE CHAIN REACTION. *Jurnal Riset Kesehatan Poltekkes Depkes Bandung*. https://doi.org/10.34011/juriskesbdg.v15i1.2179

Meredith, M. M., Liu, K., Darrasse-Jeze, G., Kamphorst, A. O., Schreiber, H. A., Guermonprez, P., Idoyaga, J., Cheong, C., Yao, K. H., Niec, R. E., & Nussenzweig, M. C. (2012). Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *Journal of Experimental Medicine*, *209*(6), 1153–1165. https://doi.org/10.1084/jem.20112675

Meyer, B., Drosten, C., & Müller, M. A. (2014). Serological assays for emerging coronaviruses: Challenges and pitfalls. *Virus Research*. https://doi.org/10.1016/j.virusres.2014.03.018

Michel, C., Mayer, C., Poch, O., & Thompson, J. (2020). *Characterization of Accessory Genes in Coronavirus Genomes*. https://doi.org/10.1101/2020.05.26.118208

Miczi, M., Golda, M., Kunkli, B., Nagy, T., Tőzsér, J., & Mótyán, J. A. (2020). Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms21249523

Milewska, A., Falkowski, K., Kulczycka, M., Bielecka, E., Naskalska, A., Mak, P., Lesner, A., Ochman, M., Urlik, M., Diamandis, E. P., Prassas, I., Potempa, J., Kantyka, T., & Pyrć, K. (2020). Kallikrein 13 Serves as a Priming Protease During Infection by the Human Coronavirus HKU1. *Science Signaling*. https://doi.org/10.1126/scisignal.aba9902

Milewska, A., Zarębski, M., Nowak, P., Stożek, K., Potempa, J., & Pyrć, K. (2014). Human Coronavirus NL63 Utilizes Heparan Sulfate Proteoglycans for Attachment to Target Cells. *Journal of Virology*. https://doi.org/10.1128/jvi.02078-14

Miller, D., Martin, M. A., Harel, N., Tirosh, O., Kustin, T., Meir, M., Sorek, N., Gefen-Halevi, S., Amit, S., Vorontsov, O., Shaag, A., Wolf, D., Peretz, A., Shemer-Avni, Y., Roif-Kaminsky, D., Kopelman, N. M., Huppert, A., Koelle, K., & Stern, A. (2020). Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nature Communications 2020 11:1*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-19248-0

Mir, S. A., Alaidarous, M., Alshehri, B., Dukhyil, A. A. B., Banawas, S., Madkhali, Y., Alsagaby, S. A., & Othaim, A. A. (2022). Immunoinformatics-Based Identification of B and T Cell Epitopes in RNA-Dependent RNA Polymerase of SARS-CoV-2. *Vaccines*. https://doi.org/10.3390/vaccines10101660

Mistry, P., Barmania, F., Mellet, J., Peta, K., Strydom, A., Viljoen, I., James, W., Gordon, S., & Pepper, M. S. (2022). SARS-CoV-2 Variants, Vaccines, and Host Immunity. *Frontiers in Immunology*, *12*, null. https://doi.org/10.3389/fimmu.2021.809244

Mitchell, D. A., Dupuy, L. C., Sanchez-Lockhart, M., Palacios, G., Back, J. W., Shimanovskaya, K., Chaudhury, S., Ripoll, D. R., Wallqvist, A., & Schmaljohn, C. S. (2017). Epitope Mapping of Ebola Virus Dominant and Subdominant Glycoprotein Epitopes Facilitates Construction of an Epitope-Based DNA Vaccine Able to Focus the Antibody Response in Mice. *Human Vaccines & Immunotherapeutics*. https://doi.org/10.1080/21645515.2017.1347740

Mitnaul, L. J., Matrosovich, M., Castrucci, M. R., Tuzikov, A., Бовин, H. B., Kobasa, D., & Kawaoka, Y. (2000). Balanced Hemagglutinin and Neuraminidase Activities Are Critical for Efficient Replication of Influenza a Virus. *Journal of Virology*. https://doi.org/10.1128/jvi.74.13.6015-6020.2000

Mittal, A., Khattri, A., & Verma, V. (2022). Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathogens*, *18*, null. https://doi.org/10.1371/journal.ppat.1010260

Moës, E., Vijgen, L., Keyaerts, E., Zlateva, K. T., Li, S., Maes, P., Pyrć, K., Berkhout, B., Hoek, L. v. d., & Ranst, M. V. (2005). A Novel Pancoronavirus RT-PCR Assay: Frequent Detection of Human Coronavirus NL63 in Children Hospitalized With Respiratory Tract Infections in Belgium. *BMC Infectious Diseases*. https://doi.org/10.1186/1471-2334-5-6

Mohammadi-Dehcheshmeh, M., Moghbeli, S. M., Rahimirad, S., Alanazi, I. O., Shehri, Z. S. A., & Ebrahimie, E. (2021). A Transcription Regulatory Sequence in the 5′ Untranslated Region of SARS-CoV-2 Is Vital for Virus Replication With an Altered Evolutionary Pattern Against Human Inhibitory MicroRNAs. *Cells*. https://doi.org/10.3390/cells10020319

Mohanty, M., Mishra, B., Singh, A. K., Mohapatra, P. R., Gupta, K., Patro, B. K., Sahu, D. P., Kar, P., Purushotham, P., Saha, S., Das, S., Mamidi, P., Panda, S., Mandal, M. C., & Bhuniya, S. (2022). Comparison of Clinical Presentation and Vaccine Effectiveness Among Omicron and Non-omicron SARS Coronavirus-2 Patients. *Cureus*, *14*(12), e32354. https://doi.org/10.7759/cureus.32354

Mok, H., Lee, S., Wright, D. W., & Crowe, J. E. (2008). Enhancement of the CD8+ T Cell Response to a Subdominant Epitope of Respiratory Syncytial Virus by Deletion of an

Immunodominant Epitope. *Vaccine*. https://doi.org/10.1016/j.vaccine.2008.07.012

Mollaie, H., Afshar, A. A., Kalantar-Neyestanaki, D., & Fazlalipour, M. (2020). Comparison Five Primer Sets From Different Genome Region of COVID-19 for Detection of Virus Infection by Conventional RT-PCR. *Iranian Journal of Microbiology*. https://doi.org/10.18502/ijm.v12i3.3234

Moore, J. P., & Ho, D. D. (1993). Antibodies to Discontinuous or Conformationally Sensitive Epitopes on the Gp120 Glycoprotein of Human Immunodeficiency Virus Type 1 Are Highly Prevalent in Sera of Infected Humans. *Journal of Virology*. https://doi.org/10.1128/jvi.67.2.863-875.1993

Moore, S. C., Penrice-Randal, R., Alruwaili, M., Dong, X., Pullan, S. T., Carter, D., Bewley, K. R., Zhao, Q., Sun, Y., Hartley, C., Zhou, E.-M., Solomon, T., Beadsworth, M., Cruise, J., Bogaert, D., Crook, D. W., Matthews, D. A., Davidson, A. D., Mahmood, Z. H., … Hiscox, J. A. (2020). *Amplicon Based MinION Sequencing of SARS-CoV-2 and Metagenomic Characterisation of Nasopharyngeal Swabs From Patients With COVID-19*. https://doi.org/10.1101/2020.03.05.20032011

Morsy, S. (2020). NCAM Protein and SARS-COV-2 Surface Proteins: In-Silico Hypothetical Evidence for the Immunopathogenesis of Guillain-Barré Syndrome. *Medical Hypotheses*. https://doi.org/10.1016/j.mehy.2020.110342

Motayo, B. O., Oluwasemowo, O. O., & Akinduti, P. A. (2020). Evolutionary Dynamics and Geographic Dispersal of Beta Coronaviruses in African Bats. *Peerj*. https://doi.org/10.7717/peerj.10434

Mousavizadeh, L., & Ghasemi, S. (2020). Genotype and phenotype of COVID-19: Their roles in pathogenesis. In *Journal of Microbiology, Immunology and Infection*. https://doi.org/10.1016/j.jmii.2020.03.022

Mozzi, A., Oldani, M., Forcella, M., Vantaggiato, C., Cappelletti, G., Pontremoli, C., Valenti, F., Forni, D., Biasin, M., Sironi, M., Fusi, P., & Cagliani, R. (2022). SARS-CoV-2 ORF3c impairs mitochondrial respiratory metabolism, oxidative stress and autophagic flow. *BioRxiv*. https://doi.org/10.1101/2022.10.04.510754

Mueller, G. A., Min, J.-Y., Foo, A. C. Y., Pomés, A., & Pedersen, L. C. (2019). Structural

Analysis of Recent Allergen-Antibody Complexes and Future Directions. *Current Allergy and Asthma Reports*. https://doi.org/10.1007/s11882-019-0848-4

Mukherjee, S., Tworowski, D., Detroja, R., Mukherjee, S. B., & Frenkel-Morgenstern, M. (2020). Immunoinformatics and Structural Analysis for Identification of Immunodominant Epitopes in SARS-CoV-2 as Potential Vaccine Targets. *Vaccines*. https://doi.org/10.3390/vaccines8020290

Mulabbi, E. N., Tweyongyere, R., Wabwire-Mangen, F., Mworozi, E., Koehlerb, J., Kibuuka, H., Millard, M., Erima, B., Tugume, T., Aquino, U. Q., & Byarugaba, D. K. (2021). Seroprevalence of Human Coronaviruses Among Patients Visiting Hospital-Based Sentinel Sites in Uganda. *BMC Infectious Diseases*. https://doi.org/10.1186/s12879-021-06258-6

Munro, R., Holmes, N., Moore, C. J., Carlile, M., Payne, A., Santos, R., & Loose, M. (2021). *Real-Time Monitoring and Analysis of SARS-CoV-2 Nanopore Sequencing With minoTour*. https://doi.org/10.1101/2021.09.13.459777

Murphy, K. M., & Casey Weaver. (2017). *Janeway's Immunobiology - Kenneth Murphy, Casey Weaver - Google Books*. Garland Science. https://books.google.co.ke/books?hl=en&lr=&id=GmPLCwAAQBAJ&oi=fnd&pg=PP1&ots=6bri0bx_ri&sig=Gzy4tks7Ovx1UtHw2SU-bGdQUdM&redir_esc=y#v=onepage&q&f=false

Musicò, A., Frigerio, R., Mussida, A., Barzon, L., Sinigaglia, A., Riccetti, S., Gobbi, F., Piubelli, C., Bergamaschi, G., Chiari, M., Gori, A., & Cretich, M. (2020). SARS-CoV-2 epitope mapping on microarrays highlights strong immune-response to N protein region. *BioRxiv*. https://doi.org/10.1101/2020.11.09.374082

Mze, N. P., Kacel, I., Beye, M., Tola, R., Sarr, M., Basco, L. K., Bogreau, H., Colson, P., & Fournier, P.-E. (2023). High Throughput SARS-CoV-2 Genome Sequencing From 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. *Genes*. https://doi.org/10.3390/genes14030681

Nahálka, J. (2022). Transcription of the Envelope Protein by 1-L Protein–RNA Recognition Code Leads to Genes/Proteins That Are Relevant to the SARS-CoV-2 Life Cycle and Pathogenesis. *Current Issues in Molecular Biology*, *44*(2), 791–816.

https://doi.org/10.3390/cimb44020055

Narwal, M., Armache, J. P., Edwards, T., & Murakami, K. (2022). *SARS-CoV-2 Polyprotein Substrate Regulates the Stepwise Mpro Cleavage Reaction*. https://doi.org/10.1101/2022.09.09.507133

Nasimiyu, C., Matoke-Muhia, D., Rono, G. K., Osoro, E., Obado, D. O., Mwangi, J. M., Mwikwabe, N. M., O, K. T., Dawa, J., Ngere, I., Gachohi, J., Kariuki, S., Amukoye, E., Mureithi, M. W., Ngere, P., Amoth, P., Were, I., Makayotto, L., Nene, V., … Oyola, S. O. (2022). Imported SARS-COV-2 Variants of Concern Drove Spread of Infections Across Kenya During the Second Year of the Pandemic. *MedRxiv*. https://doi.org/10.1101/2022.02.28.22271467

Nasimiyu, C., Matoke-Muhia, D., Rono, G. K., Osoro, E., Ouso, D. O., Mwangi, J. M., Mwikwabe, N., Thiong'o, K., Dawa, J., Ngere, I., Gachohi, J., Kariuki, S., Amukoye, E., Mureithi, M., Ngere, P., Amoth, P., Were, I., Makayotto, L., Nene, V., … Oyola, S. O. (2022). Imported SARS-CoV-2 Variants of Concern Drove Spread of Infections across Kenya during the Second Year of the Pandemic. *COVID*, *2*(5), 586–598. https://doi.org/10.3390/covid2050044

Nasir, J. A., Kozak, R., Aftanas, P., Raphenya, A. R., Smith, K. M., Maguire, F., Maan, H., Alruwaili, M., Banerjee, A., Mbareche, H., Alcock, B., Knox, N., Mossman, K., Wang, B., Hiscox, J. A., McArthur, A. G., & Mubareka, S. (2020). A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture. *Viruses*. https://doi.org/10.3390/v12080895

Nawaz, R., Arif, M. A., Ahmad, Z., Ahad, A., Shahid, M., Hassan, Z., Husnain, A., Aslam, A., Mehmood, U., & Khan, I. (2023). *An Immunoinformatic Approach to Design siRNA Molecules Against SARS-CoV-2 Double Membrane Vesicle Formation and Accessory Genes*. https://doi.org/10.21203/rs.3.rs-3141999/v1

Nechaev, V. V, Lobzin, Y. V, Gusev, D. A., Kholin, A. V, Vinogradova, N. V, Shmeleva, O. V, Boburova, L. E., Gorziy, E. S., Levina, M. S., & Darina, M. G. (2022). Epidemiological patterns and features of SARS-CoV-2 on a regional, territorial scale and at the level of megacities. *Sanitarnyj Vrač*, *8*, 524–539. https://doi.org/10.33920/med-08-2208-01

Neerukonda, S., Vassell, R., Lusvarghi, S., Wang, R., Echegaray, F., Bentley, L., Eakin, A., Erlandson, K. J., Katzelnick, L., Weiss, C., & Wang, W. (2021). SARS-COV-2 Delta variant displays moderate resistance to neutralizing antibodies and spike protein properties of higher soluble ACE2 sensitivity, enhanced cleavage and fusogenic activity. *BioRxiv*, *13*, null. https://doi.org/10.1101/2021.11.05.467523

Nelde, A., Bilich, T., Heitmann, J. S., Maringer, Y., Salih, H. R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., Peter, A., Hörber, S., Traenkle, B., Kaiser, P., Rothbauer, U., Becker, M., Junker, D., Krause, G., Strengert, M., … Walz, J. S. (2020). SARS-CoV-2-derived Peptides Define Heterologous and COVID-19-induced T Cell Recognition. *Nature Immunology*. https://doi.org/10.1038/s41590-020-00808-x

Nelde, A., Bilich, T., Heitmann, J. S., Maringer, Y., Salih, H. R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., Peter, A., Hörber, S., Traenkle, B., Kaiser, P., Rothbauer, U., Becker, M., Junker, D., Krause, G., Strengert, M., … Walz, J. S. (2020). *SARS-CoV-2 T-Cell Epitopes Define Heterologous and COVID-19-induced T-Cell Recognition*. https://doi.org/10.21203/rs.3.rs-35331/v1

Nguyen, C. M., Luong, B. A., & Tran, L. S. (2023). Design and generation of mRNAs encoding conserved regions of SARS-CoV-2 ORF1ab for T cell-mediated immune activation. *Future Virology*. https://doi.org/10.2217/fvl-2023-0066

Njenga, M. K. (2022). Imported SARS-CoV-2 Variants of Concern Drove Spread of Infections across Kenya during the Second Year of the Pandemic. *COVID*, *2*(5), 586–598. https://doi.org/10.3390/covid2050044

Noaman, V. (2022). *The spike-stabilizing D614G mutation interacts with S1/S2 cleavage site mutations to promote the infectious potential of SARS-CoV-2 variants*. https://doi.org/10.1101/2022.05.20.492832

Nomburg, J., Meyerson, M., & DeCaprio, J. A. (2020). *Pervasive Generation of Non-Canonical Subgenomic RNAs by SARS-CoV-2*. https://doi.org/10.1101/2020.04.28.066951

Noronha, P., Paraschoudi, G., Sousa, E. d., Kamiki, J., António, P., Condenço, C., Silva, I., Maia, A., Castillo-Martin, M., Beltran, A., Carvalho, C., Lérias, J. R., Zumla, A., & Maeurer, M. (2020). *326 sars-CoV-2 Specific T-Cells in TIL From Patients With*

*Epithelial Cancer*. https://doi.org/10.1136/jitc-2020-sitc2020.0326

O'Keefe, S., Roboti, P., Duah, K. B., Zong, G., Schneider, H., Shi, W., & High, S. (2021). Ipomoeassin-F Inhibits the <i>in Vitro</I> Biogenesis of the SARS-CoV-2 Spike Protein and Its Host Cell Membrane Receptor. *Journal of Cell Science*. https://doi.org/10.1242/jcs.257758

Ogawa, J., Zhu, W., Tonnu, N., Singer, O., Hunter, T., Ryan, A. L., & Pao, G. M. (2020). *The D614G Mutation in the SARS-CoV2 Spike Protein Increases Infectivity in an ACE2 Receptor Dependent Manner*. https://doi.org/10.1101/2020.07.21.214932

Olarte-Castillo, X. A. (2023). Detection and Characterization of Novel Luchacoviruses, Genus <i>Alphacoronavirus</I> , in Saliva and Feces of Meso-Carnivores in the Northeastern United States. *Journal of Virology*. https://doi.org/10.1128/jvi.00829-23

Oliveira, S. C., Magalhães, M. T. d., & Homan, E. J. (2020). Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2020.587615

Olyaee, M. H., Pirgazi, J., Khalifeh, K., & Khanteymoori, A. (2020). RCOVID19: Recurrence-Based SARS-CoV-2 Features Using Chaos Game Representation. *Data in Brief*. https://doi.org/10.1016/j.dib.2020.106144

Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. In *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.2020.4683

Onishchenko, G. G., Sizikova, T. E., Lebedev, V. N., & Borisevich, S. V. (2022). The Omicron Variant of the Sars-Cov-2 Virus As the Dominant Agent of a New Risk of Disease amid the COVID-19 Pandemic. *Herald of the Russian Academy of Sciences*, *92*(4), 381–391. https://doi.org/10.1134/S1019331622040074

Orłowska, A., Smreczak, M., Thor, K., Niedbalska, M., Pawelec, D., Trębas, P., & Rola, J. (2022). The Genetic Characterization of the First Detected Bat Coronaviruses in Poland Revealed SARS-Related Types and Alphacoronaviruses. *Viruses*. https://doi.org/10.3390/v14091914

Otieno, G. P., Murunga, N., Agoti, C. N., Gallagher, K. E., Awori, J. O., & Nokes, D. J.

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

Ouyang, W., Xie, T., Fang, H., Gao, C.-Z., Stantchev, T. S., Clouse, K. A., yuan, kun, Ju, T., & Frucht, D. M. (2021). Variable Induction of Pro-Inflammatory Cytokines by Commercial SARS CoV-2 Spike Protein Reagents: Potential Impacts of LPS on in Vitro Modeling and Pathogenic Mechanisms in Vivo. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms22147540

Owusu, M., Sylverken, A. A., El-Duah, P., Acheampong, G., Mutocheluh, M., & Adu-Sarkodie, Y. (2021). Sero-Epidemiology of Human Coronaviruses in Three Rural Communities in Ghana. *Pan African Medical Journal*. https://doi.org/10.11604/pamj.2021.38.244.26110

Pacheco-Olvera, D. L., Remy-Hernández, S. S., Acevedo-Ochoa, E., Arriaga-Pizano, L., Cérbulo-Vázquez, A., Ferat-Osorio, E., Rivera-Hernandez, T., & López-Macías, C. (2020). *Bioinformatic Analysis of Shared B and T Cell Epitopes Amongst Relevant Coronaviruses to Human Health: Is There Cross-Protection?* https://doi.org/10.1101/2020.07.14.202887

Pagani, I., Ghezzi, S., Alberti, S., Poli, G., & Vicenzi, E. (2023). Origin and evolution of SARS-CoV-2. *European Physical Journal Plus*, *138*(2). https://doi.org/10.1140/epjp/s13360-023-03719-6

Pan, K., Chiu, Y., Huang, E. L., Chen, M., Wang, J., Lai, I., Singh, S., Shaw, R., MacCoss, M. J., & Yee, C. (2021). *Immunogenic SARS-CoV2 Epitopes Defined by Mass Spectrometry*. https://doi.org/10.1101/2021.07.20.453160

Pan, P., Shen, M., Yu, Z., Ge, W., Chen, K., Tian, M., Tian, M., Xiao, F., Wang, Z., Wang, J., Jia, Y., Wang, W., Wan, P., Zhang, J., Chen, W., Lei, Z., Chen, X., Luo, Z., Zhang, Q., … Wu, J. (2021). SARS-CoV-2 N protein promotes NLRP3 inflammasome activation to induce hyperinflammation. *Nature Communications*, *12*(1), 4664. https://doi.org/10.1038/S41467-021-25015-6

Pancer, K., Milewska, A., Owczarek, K., Dabrowska, A., Kowalski, M., Łabaj, P. P., Branicki, W., Sanak, M., & Pyrc, K. (2020). The SARS-CoV-2 ORF10 is not essential

in vitro or in vivo in humans. *PLOS Pathogens*, *16*(12). https://doi.org/10.1371/JOURNAL.PPAT.1008959

Pang, M., Lu, Y., Tang, F., Wang, H., Zhou, Y., Ren, L., Li, R., Zhou, H., Wan, C.-F., Liu, C.-C., Luosang, D., Cairang, Y., & Fan, H. (2022). Prediction and Identification of Epitopes in the Echinococcus Multilocularis Thrombospondin 3 Antigen. *Technology and Health Care*. https://doi.org/10.3233/thc-212983

Papadopoulos, V., Emmanouilidou, A., Yerou, M., Panagaris, S., Souleiman, C., Varela, D., Avramidou, P., Melissopoulou, E., Pappas, C., Iliadou, Z., Piperopoulos, I., Somadis, V., Partsalidis, A., Metaxa, E., Feresiadis, I., & Filippou, D. (2022). SARS-CoV-2 Vaccination Coverage and Key Public Health Indicators May Explain Disparities in COVID-19 Country-Specific Case Fatality Rate Within European Economic Area. *Cureus*, *14*. https://doi.org/10.7759/cureus.22989

Park, C., Kim, K. W., Park, D., Hassan, Z. U., Park, E. C., Lee, C.-S., Rahman, M. D. T., Yi, H., & Kim, S. (2022). Rapid and Sensitive Amplicon-Based Genome Sequencing of SARS-CoV-2. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2022.876085

Parvez, M. K., & Sakina, N. (2021). The Genomic and Structural Organization of SARS-CoV-2: A Mutational Perspective. *Scimedicine Journal*. https://doi.org/10.28991/scimedj-2021-0301-8

Patel, R. K., & Agrawal, B. (2023). Mucosal immunization with lipopeptides derived from conserved regions of SARS-CoV-2 antigens induce robust cellular and cross-variant humoral immune responses in mice. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/fimmu.2023.1178523

Patel, V., McCarthy, C., Taylor, R. A., Moir, R., Kelly, L. A., & Snary, E. (2020). An improved methodology for estimating the prevalence of SARS-CoV-2. *MedRxiv*. https://doi.org/10.1101/2020.08.04.20168187

Patro, L. P. P., Sathyaseelan, C., Uttamrao, P. P., & Rathinavelan, T. (2021). Global Variation in SARS-CoV-2 Proteome and Its Implication in Pre-Lockdown Emergence and Dissemination of 5 Dominant SARS-CoV-2 Clades. *Infection Genetics and Evolution*. https://doi.org/10.1016/j.meegid.2021.104973

Paul, D., Gupta, A., Bhatia, V., & Gupta, E. (2022). The Neglected Pathogen: Case Reports of Severe Lower Respiratory Tract Infection by Human Coronavirus 229E. *Access Microbiology*. https://doi.org/10.1099/acmi.0.000311

Paules, C. I., Marston, H. D., & Fauci, A. S. (2020). Coronavirus Infections—More Than Just the Common Cold. *Jama*. https://doi.org/10.1001/jama.2020.0757

Pavesi, A. (2021). Prediction of Two Novel Overlapping ORFs in the Genome of SARS-CoV-2. *Virology*. https://doi.org/10.1016/j.virol.2021.07.011

Pearson, G., Broncel, M., Snijders, A. P., & Carlton, J. G. (2021). *ER-export and ARFRP1/AP-1-dependent Delivery of SARS-CoV-2 Envelope to Lysosomes Controls Late Stages of Viral Replication*. https://doi.org/10.1101/2021.06.30.450614

Pecetta, S., Kratochvil, S., Kato, Y., Vadivelu, K., & Rappuoli, R. (2022). Immunology and Technology of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Vaccines. *Pharmacological Reviews*, *74*, 313–339. https://doi.org/10.1124/pharmrev.120.000285

Peiris, J. (2003). Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*, *361*, 1767–1772.

Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. In *Nature Reviews Microbiology* (Vol. 7, Issue 6, pp. 439–450). https://doi.org/10.1038/nrmicro2147

Perrier, A., Bonnin, A., Desmarets, L., Danneels, A., Goffard, A., Rouillé, Y., Dubuisson, J., & Belouzard, S. (2019). The C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal Running title: MERS-CoV M protein TGN localization Downloaded from. *Journal of Biological Chemistry*.

Peters, B., Sidney, J., Bourne, P., Bui, H.-H., Buus, S., Doh, G., Fleri, W., Kronenberg, M., Kubo, R. T., Lund, O., Nemazee, D., Ponomarenko, J., Sathiamurthy, M., Schoenberger, S. P., Stewart, S. J., Surko, P., Way, S., Wilson, S., & Sette, A. (2005). The Design and Implementation of the Immune Epitope Database and Analysis Resource. *Immunogenetics*. https://doi.org/10.1007/s00251-005-0803-5

Piccaluga, P. P., Malerba, G., Navari, M., Diani, E., Concia, E., & Gibellini, D. (2021). Cross-Immunization Against Respiratory Coronaviruses May Protect Children From SARS-CoV2: More Than a Simple Hypothesis? *Frontiers in Pediatrics*. https://doi.org/10.3389/fped.2020.595539

Piccoli, L., Park, Y.-J., Tortorici, M. A., Czudnochowski, N., Walls, A., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L. E., Bowen, J. E., Acton, O., Jaconi, S., Guarino, B., Minola, A., Zatta, F., Sprugasci, N., Bassi, J., Peter, A., Marco, A. De, … Veesler, D. (2020). Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell*, *183*, 1024-1042.e21. https://doi.org/10.1016/j.cell.2020.09.037

Pillay, S., Giandhari, J., Tegally, H., Wilkinson, E., Chimukangara, B., Lessells, R., Moosa, M.-Y. S., Mattison, S. A., Gazy, I., Fish, M., Singh, L., Khanyile, K. S., San, J. E., Fonseca, V., Giovanetti, M., Alcântara, L. C. J., & Oliveira, T. d. (2020). Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation During a Pandemic. *Genes*. https://doi.org/10.3390/genes11080949

Pillay, S., San, J. E., Tshiabuila, D., Naidoo, Y., Pillay, Y., Maharaj, A., Anyaneji, U. J., Wilkinson, E., Tegally, H., Lessells, R., Baxter, C., Oliveira, T. d., & Giandhari, J. (2023). Evaluation of Miniaturized Illumina DNA Preparation Protocols for SARS-CoV-2 Whole Genome Sequencing. *Plos One*. https://doi.org/10.1371/journal.pone.0283219

Pinto, D., Park, Y.-J., Beltramello, M., Walls, A., Tortorici, M. A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., Marco, A. De, Peter, A., Guarino, B., Spreafico, R., Cameroni, E., Case, J., Chen, R. E., Havenar-Daughton, C., Snell, G., Telenti, A., … Corti, D. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, *583*, 290–295. https://doi.org/10.1038/s41586-020-2349-y

Pk, D., Ampajwala, M., Chappel, C., Ab, G., Hoppers, M., Wang, M., Rosen, R. L., Young, S., Zissman, E., & Montaño, M. A. (2020). *A Rapid, High-Sensitivity SARS-CoV-2 Nucleocapsid Immunoassay to Aid Diagnosis of Acute COVID-19 at the Point of Care*. https://doi.org/10.1101/2020.12.11.20238410

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., … Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, *596*(7871), 276–280. https://doi.org/10.1038/s41586-021-03777-9

Plitnick, J., Griesemer, S. B., Lasek-Nesselquist, E., Singh, N., Lamson, D. M., & George, K. S. (2021). Whole-Genome Sequencing of SARS-CoV-2: Assessment of the Ion Torrent AmpliSeq Panel and Comparison With the Illumina MiSeq ARTIC Protocol. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.00649-21

Pollara, J., Bonsignori, M., Moody, M. A., Pazgier, M., Haynes, B. F., & Ferrari, G. (2013). Epitope Specificity of Human Immunodeficiency Virus-1 Antibody Dependent Cellular Cytotoxicity [ADCC] Responses. *Current Hiv Research*. https://doi.org/10.2174/1570162x113116660059

Pondé, R. (2022). Physicochemical effect of the N501Y, E484K/Q, K417N/T, L452R and T478K mutations on the SARS-CoV-2 spike protein RBD and its influence on agent fitness and on attributes developed by emerging variants of concern. *Virology*, *572*, 44–54. https://doi.org/10.1016/j.virol.2022.05.003

Pongou, R., Ahinkorah, B. O., Mabeu, M. C., Agarwal, A., Maltais, S., Moumouni, A. B., & Yaya, S. (2023). Identity and COVID-19 in Canada: Gender, ethnicity, and minority status. *PLOS Global Public Health*, *3*(5). https://doi.org/10.1371/journal.pgph.0001156

Pouwels, K. B., Pouwels, K. B., House, T., House, T., Pritchard, E., Pritchard, E., Robotham, J. V, Birrell, P. J., Birrell, P. J., Gelman, A., K-D., V., K-D., V., Bowers, N., Boreham, I., Thomas, H., Lewis, J., Bell, I., Bell, J., Newton, J. N., … Walker, A.-S. (2021). Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey. *The Lancet. Public Health*, *6*(1). https://doi.org/10.1016/S2468-2667(20)30282-6

Prasasty, V. D., Grazzolie, K., Rosmalena, R., Yazid, F., Ivan, F. X., & Sinaga, E. (2019). Peptide-Based Subunit Vaccine Design of T- And B-Cells Multi-Epitopes Against

Zika Virus Using Immunoinformatics Approaches. *Microorganisms*. https://doi.org/10.3390/microorganisms7080226

Prill, M. M., Iwane, M. K., Edwards, K. M., Williams, J. V, Weinberg, G. A., Staat, M. A., Willby, M. J., Talbot, H. K., Hall, C. B., Szilâgyi, P. G., Griffin, M. R., Curns, A. T., & Erdman, D. D. (2012). Human Coronavirus in Young Children Hospitalized for Acute Respiratory Illness and Asymptomatic Controls. *The Pediatric Infectious Disease Journal*. https://doi.org/10.1097/inf.0b013e31823e07fe

Prompetchara, E., Ketloy, C., & Palaga, T. (2020). Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. In *Asian Pacific Journal of Allergy and Immunology* (Vol. 38, Issue 1, pp. 1–9). Allergy and Immunology Society of Thailand. https://doi.org/10.12932/AP-200220-0772

Pulido, J. C., García-Durán, M., Fernández-Antonio, R., Galán, C. R., López, L., Vela, C., Venteo, Á., Rueda, P., & Rivas, L. A. (2021). Receptor-Binding Domain–based Immunoassays for Serosurveillance Differentiate Efficiently Between SARS-CoV2–exposed and Non-Exposed Farmed Mink. *Journal of Veterinary Diagnostic Investigation*. https://doi.org/10.1177/10406387211057859

Pushpakumara, P. D., Madusanka, D., Danasekara, S., Jeewandara, C., Ogg, G., & Malavige, G. N. (2021). *Identification of Novel Candidate CD8+ T Cell Epitopes of the SARS-CoV2 With Homology to Other Seasonal Coronaviruses*. https://doi.org/10.21203/rs.3.rs-228306/v1

Pyo, C. W., Guethlein, L. A., Vu, Q., Wang, R., Abi-Rached, L., Norman, P., Marsh, S. G. E., Miller, J. S., Parham, P., & Geraghty, D. E. (2010). Different Patterns of Evolution in the Centromeric and Telomeric Regions of Group A and B Haplotypes of the Human Killer Cell Ig-Like Receptor Locus. *PLOS ONE*, *5*(12). https://doi.org/10.1371/JOURNAL.PONE.0015115

Pyrc, K., Sims, A. C., Dijkman, R., Jebbink, M. F., Long, C., Deming, D., Donaldson, E. F., Vabret, A., Baric, R. S., Hoek, L. v. d., & Pickles, R. J. (2010). Culturing the Unculturable: Human Coronavirus HKU1 Infects, Replicates, and Produces Progeny Virions in Human Ciliated Airway Epithelial Cell Cultures. *Journal of Virology*. https://doi.org/10.1128/jvi.00947-10

Radovic, S., Meng, W., Chen, L., Mondolfi, A. E. P., Bryce, C., Grimes, Z., Sordillo, E. M., Cordon-Cardo, C., Guo, H., Huang, Y., & Gao, S.-J. (2023). SARS-CoV-2 Infection of Kidney Tissues From Severe COVID-19 Patients. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.28566

Raheja, H., Sahu, R., Ghosh, T., Rani, P., George, B., & Tripathi, S. (2023). *Human Antigen R Differentially Modulates the Translation of SARS-CoV-2 Genomic and Sub-Genomic RNAs*. https://doi.org/10.1101/2023.03.15.532790

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. A., Dijkman, R., Muth, D., Demmers, J. A. A., Zaki, A., Fouchier, R. A. M., Thiel, V., Drosten, C., Rottier, P. J. M., Osterhaus, A. D. M. E., Bosch, B. J., & Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. https://doi.org/10.1038/nature12005

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., & Volz, E. (2021). on behalf of COVID-19 Genomics Consortium UK (CoG-UK). 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Report on Virological. Org Website: Https://Virological. Org/t/Preliminary-Genomic-Characterisation-of-an-Emergent-Sars-Cov-2-Lineage-in-the-Uk-Defined-by-a-Novel-Set-of-Spike-Mutations/563*, 2019. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

Ranjbarian, P., Goudarzi, F., Akya, A., Heidarinia, H., Farasat, A., & Rostamian, M. (2023). Finding Epitopes of Klebsiella Pneumoniae Outer Membrane Protein-K17 (OMPK17) and Introducing a 25-Mer Peptide of It as a Vaccine Candidate. *Biologia*. https://doi.org/10.1007/s11756-023-01371-0

Raoufi, E., Hemmati, M., Eftekhari, S., Khaksaran, K., Mahmodi, Z., Farajollahi, M. M., & Mohsenzadegan, M. (2019). Epitope Prediction by Novel Immunoinformatics Approach: A State-of-the-Art Review. *International Journal of Peptide Research and Therapeutics*. https://doi.org/10.1007/s10989-019-09918-z

Ras-Carmona, A., Pelaez-Prestel, H. F., Lafuente, E. M., & Reche, P. A. (2021). BCEPS: A

Web Server to Predict Linear B Cell Epitopes With Enhanced Immunogenicity and Cross-Reactivity. *Cells*. https://doi.org/10.3390/cells10102744

Rasheed, M., Awais, M., Aldhahrani, A., Althobaiti, F., Alhazmi, A., Sattar, S., Afzal, U., Baeshen, H. A., Enshasy, H. A. E., Dailin, D. J., AL-surhanee, A. A., & Kabir, F. (2021). Designing a Highly Immunogenic Multi Epitope Based Subunit Vaccine Against Bacillus Cereus. *Saudi Journal of Biological Sciences*. https://doi.org/10.1016/j.sjbs.2021.06.082

Ravindra, N. G., Alfajaro, M. M., Gasque, V., Habet, V., Jin, W., Filler, R., Huston, N. C., Wan, H., Szigeti-Buck, K., Bao, W., Wang, G., Montgomery, R. R., Eisenbarth, S. C., Williams, A., Pyle, A. M., Iwasaki, A., Horváth, T. L., Foxman, E. F., Pierce, R. W., … Wilen, C. B. (2020). *Single-Cell Longitudinal Analysis of SARS-CoV-2 Infection in Human Airway Epithelium*. https://doi.org/10.1101/2020.05.06.081695

Reeves, M. B. (2020). Cell Signaling and Cytomegalovirus Reactivation: What Do SRC Family Kinases Have to Do With It? *Biochemical Society Transactions*. https://doi.org/10.1042/bst20191110

Regitz-Zagrosek, V., & Seeland, U. (2011). Sex and gender differences in myocardial hypertrophy and heart failure. *Wiener Medizinische Wochenschrift*. https://doi.org/10.1007/s10354-011-0892-8

Ren, L., Gonzalez, R. R., Xu, J., Xiao, Y., Li, Y., Zhou, H., Li, J., Yang, Q., Zhang, J., Chen, L., Wang, W., Vernet, G., Paranhos-Baccalà, G., Wang, Z., & Wang, J. (2010). Prevalence of Human Coronaviruses in Adults With Acute Respiratory Tract Infections in Beijing, China. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.21956

Ren, Y., Wang, A., Fang, Y., Shu, T., Wu, D., Wang, C., Huang, M., Min, J., Jin, L., Zhou, W., Qiu, Y., & Xi, Z. (2021). SARS-CoV-2 Membrane Glycoprotein M Triggers Apoptosis With the Assistance of Nucleocapsid Protein N in Cells. *Frontiers in Cellular and Infection Microbiology*, *11*, 706252. https://doi.org/10.3389/FCIMB.2021.706252

Renn, O. (2023). *Synonymous base substitution in SARS-CoV-2 EE.2 lineage furin arginine CGG–CGG codons*. https://doi.org/10.21203/rs.3.rs-3139964/v1

Richardson, C. C., McLaughlin, K. A., Morgan, D., Feltbower, R., & Christie, M. R. (2015). Influence of HLA-DR and -DQ Alleles on Autoantibody Recognition of Distinct Epitopes Within the Juxtamembrane Domain of the IA-2 Autoantigen in Type 1 Diabetes. *Diabetologia*. https://doi.org/10.1007/s00125-015-3803-5

Richmond, M., McKinnon, L. R., Kiazyk, S., Wachihi, C., Kimani, M., Kimani, J., Plummer, F. A., & Ball, T. B. (2011). Epitope Mapping of HIV-Specific CD8 <sup>+</Sup> T Cell Responses by Multiple Immunological Readouts Reveals Distinct Specificities Defined by Function. *Journal of Virology*. https://doi.org/10.1128/jvi.01707-10

Robert-Guroff, M., Louie, A. T., Myagkikh, M., Michaels, F. H., Kieny, M. P., White-Scharf, M. E., Potts, B., Grogg, D., & Reitz, M. S. (1994). Alteration of V3 Loop Context Within the Envelope of Human Immunodeficiency Virus Type 1 Enhances Neutralization. *Journal of Virology*. https://doi.org/10.1128/jvi.68.6.3459-3466.1994

Rochars, V. M. B. De, Lednicky, J. A., White, S. K., Loeb, J. C., Elbadry, M. A., Telisma, T., Chavannes, S., Anilis, M. G., Cella, E., Ciccozzi, M., Okech, B. A., Salemi, M., & Morris, J. G. (2016). Isolation of Coronavirus NL63 From Blood From Children in Rural Haiti: Phylogenetic Similarities With Recent Isolates From Malaysia. *American Journal of Tropical Medicine and Hygiene*. https://doi.org/10.4269/ajtmh.16-0585

Rodrigues-da-Silva, R. N., Conte, F. de P., Silva, G. D. P. da, Kuriyama, S. N., Neto, A. A., & Lima-Junior, J. da C. (2023). Identification of B-Cell Linear Epitopes in the Nucleocapsid (N) Protein B-Cell Linear Epitopes Conserved among the Main SARS-CoV-2 Variants. *Viruses*, *15*(4), 923. https://doi.org/10.3390/v15040923

Rodrigues, T. S. F. (2024). *SARS-CoV (COVID-19 and SARS)*. 812–818. https://doi.org/10.1016/b978-0-323-80932-0.00146-4

Rodríguez, F., Harkins, S., & Whitton, J. L. (2002). Immunodominance in Virus-Induced $CD8^{+}$T-Cell Responses Is Dramatically Modified by DNA Immunization and Is Regulated by Gamma Interferon. *Journal of Virology*. https://doi.org/10.1128/jvi.76.9.4251-4259.2002

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P.,

Zwirglmaier, K., Zange, S., Wölfel, R., & Hoelscher, M. (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine*. https://doi.org/10.1056/nejmc2001468

Sagar, M., Reifler, K., Rossi, M. R., Miller, N. S., Sinha, P., White, L. F., & Mizgerd, J. P. (2021). Recent Endemic Coronavirus Infection Is Associated With Less-Severe COVID-19. *Journal of Clinical Investigation*. https://doi.org/10.1172/jci143380

Saha, S., & Raghava, G. P. S. (2006). Prediction of Continuous B-cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins Structure Function and Bioinformatics*. https://doi.org/10.1002/prot.21078

Salazar, C., Ferrés, I., Paz, M., Costábile, A., Moratorio, G., Moreno, P., & Iraola, G. (2023). Fast and Cost-Effective SARS-CoV-2 Variant Detection Using Oxford Nanopore Full-Length Spike Gene Sequencing. *Microbial Genomics*. https://doi.org/10.1099/mgen.0.001013

Saphire, E. O., & Aman, M. J. (2016). Feverish Quest for Ebola Immunotherapy: Straight or Cocktail? *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2016.05.008

Saranya, S., Sangavai, C., B, R., & Chellapandi, P. (2022). *Studying Folding Kinetics of Omicron to Understand Its Hijack Mechanism in Human Host Cells*. https://doi.org/10.21203/rs.3.rs-2392479/v1

Sardi, S. I., Carvalho, R., Pacheco, L. G., Almeida, J. F. L. de, Belitardo, E. M. M. D. A., Pinheiro, C. S., Campos, G. S., & Aguiar, E. R. G. R. (2020). High-Quality Resolution of the Outbreak-Related Zika Virus Genome and Discovery of New Viruses Using Ion Torrent-Based Metatranscriptomics. *Viruses*. https://doi.org/10.3390/v12070782

SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations. (2022). *Infection and Drug Resistance*, *Volume 15*, 373–385. https://doi.org/10.2147/idr.s344099

Savastano, A., Opakua, A. I. d., Ranković, M., & Zweckstetter, M. (2020). Nucleocapsid Protein of SARS-CoV-2 Phase Separates Into RNA-rich Polymerase-Containing Condensates. *Nature Communications*. https://doi.org/10.1038/s41467-020-19843-1

Savellini, G. G., Anichini, G., Gandolfo, C., & Cusi, M. G. (2021). SARS-CoV-2 N Protein

Targets TRIM25-Mediated RIG-I Activation to Suppress Innate Immunity. *Viruses*, *13*(8), 1439. https://doi.org/10.3390/V13081439

Sawicki, S. G., & Sawicki, D. L. (2005). Coronavirus transcription: A perspective. In *Current Topics in Microbiology and Immunology* (Vol. 287, pp. 31–55). Springer Verlag. https://doi.org/10.1007/3-540-26765-4_2

Saxena, D., Batra, L., & Verma, S. (2023). Broad-Spectrum Antivirals against Multiple Human and Animal Coronaviruses Infection. *Pathogens*, *12*(6), 823. https://doi.org/10.3390/pathogens12060823

Sayama, Y., Okamoto, M., Saito, M., Saito-Obata, M., Tamaki, R., Joboco, C. D., Lupisan, S., & Oshitani, H. (2023). Seroprevalence of Four Endemic Human Coronaviruses And, Reactivity and Neutralization Capability Against SARS-CoV-2 Among Children in the Philippines. *Scientific Reports*. https://doi.org/10.1038/s41598-023-29072-3

Scheffer, K. C., Neto, R. dos S. C., Fahl, W. de O., Zamudio, R. M., Zamudio, M. M., Asano, K. M., Chierato, M. E. R., Pedrozo, D. F. P., Enio Mori, K. I., Cocchi, M., Chaves, L. B., Cássia, A. de, Silva, R. da, & Batista, H. B. de C. R. (2023). *Bat Coronaviruses in the World*. 1–34. https://doi.org/10.2174/9789815123371123030003

Schepis, T., Larghi, A., Papa, A., Miele, L., Panzuto, F., Biase, L. D., Annibale, B., Cattani, P., & Rapaccini, G. L. (2020). SARS-CoV2 RNA Detection in a Pancreatic Pseudocyst Sample. *Pancreatology*. https://doi.org/10.1016/j.pan.2020.05.016

Schindell, B., Allardice, M., McBride, J. A., Dennehy, B., & Kindrachuk, J. (2022). SARS-CoV-2 and the Missing Link of Intermediate Hosts in Viral Emergence - What We Can Learn From Other Betacoronaviruses. *Frontiers in Virology*. https://doi.org/10.3389/fviro.2022.875213

Schuster, N. A. (2020). A theoretical analysis of the putative ORF10 protein in SARS-CoV-2. *BioRxiv*. https://doi.org/10.1101/2020.10.26.355784

Sechan, F., Grobben, M., Edridge, A. W. D., Jebbink, M. F., Loens, K., Ieven, M., Goossens, H., Hemert-Glaubitz, S. v., Gils, M. J. v., & Hoek, L. v. d. (2022). Atypical Antibody Dynamics During Human Coronavirus HKU1 Infections. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2022.853410

Seemann, T., Lane, C. R., Sherry, N. L., Duchêne, S., Silva, A. G. d., Caly, L., Sait, M., Ballard, S. A., Horan, K., Schultz, M. B., Hoang, T., Easton, M., Dougall, S., Stinear, T. P., Druce, J., Catton, M., Sutton, B., Diemen, A. v., Alpren, C., … Howden, B. P. (2020). *Tracking the COVID-19 Pandemic in Australia Using Genomics*. https://doi.org/10.1101/2020.05.12.20099929

Sethuraman, N., Jeremiah, S. S., & Ryo, A. (2020). Interpreting Diagnostic Tests for SARS-CoV-2. *Jama*. https://doi.org/10.1001/jama.2020.8259

Sette, A., & Crotty, S. (2021). Adaptive immunity to SARS-CoV-2 and COVID-19. In *Cell* (Vol. 184, Issue 4, pp. 861–880). Elsevier B.V. https://doi.org/10.1016/j.cell.2021.01.007

Shaka, K., Jones, K., Putzke, A. P., & Measor, P. (2023). *A 3D printed microfluidic PCR device towards detecting SARS-CoV-2*. *12387*, 123870C-123870C. https://doi.org/10.1117/12.2648519

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature*. https://doi.org/10.1038/s41586-020-2179-y

Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Tai, W., Du, L., Zhou, Y., Zhang, W., & Li, F. (2018). Cryo-Electron Microscopy Structure of Porcine Deltacoronavirus Spike Protein in the Prefusion State. *Journal of Virology*. https://doi.org/10.1128/jvi.01556-17

Shao, N., Zhang, C., Dong, J., Sun, L., Chen, X., Xie, Z., Xu, B., An, S., Zhang, T., & Yang, F. (2022). Molecular Evolution of Human Coronavirus-Nl63, -229e, -HKU1 and -OC43 in Hospitalized Children in China. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2022.1023847

Shekhawat, U., & Chowdhury, A. R. (2022). Computational and Comparative Investigation of Hydrophobic Profile of Spike Protein of SARS-CoV-2 and SARS-CoV. *Journal of Biological Physics*. https://doi.org/10.1007/s10867-022-09615-x

Shen, L., Bard, J. D., Triche, T. J., Judkins, A. R., Biegel, J. A., & Gai, X. (2021). Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved

target with potential pathological and therapeutic implications. *Emerging Microbes & Infections*, *10*(1), 885–893. https://doi.org/10.1080/22221751.2021.1922097

Shen, X., Tang, H., McDanal, C., Wagh, K., Fischer, W., Theiler, J., Yoon, H., Li, D., Haynes, B., Sanders, K. O., Gnanakaran, S., Hengartner, N., Pajon, R., Smith, G., Dubovsky, F., Glenn, G., Korber, B., & Montefiori, D. (2021). SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral Spike vaccines. *BioRxiv*, *29*, 529-539.e3. https://doi.org/10.1101/2021.01.27.428516

Sheng, Z., Diao, M. Y., Yu, W., Pei, L., Lin, Z., & Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International Journal of Infectious Diseases*. https://doi.org/10.1016/j.ijid.2020.02.033

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. In *Journal of Advanced Research* (Vol. 24, pp. 91–98). Elsevier B.V. https://doi.org/10.1016/j.jare.2020.03.005

Sherif, Z. A., Gomez, C. R., Connors, T. J., Henrich, T. J., & Reeves, W. B. (2023). Pathogenic mechanisms of post-acute sequelae of SARS-CoV-2 infection (PASC). *ELife*, *12*. https://doi.org/10.7554/eLife.86002

Shi, C. S., Qi, H., Boularan, C., Huang, N., Abu-Asab, M., Shelhamer, J. H., & Kehrl, J. H. (2014). SARS-Coronavirus Open Reading Frame-9b Suppresses Innate Immunity by Targeting Mitochondria and the MAVS/TRAF3/TRAF6 Signalosome. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.1303196

Shi, Jiale, Shi, Y., Xiu, R., Gang, W., Liang, R., Jiao, Y., Shen, Z., Zhu, C., & Peng, G. (2022). Identification of a Novel Neutralizing Epitope on the N-Terminal Domain of the Human Coronavirus 229E Spike Protein. *Journal of Virology*. https://doi.org/10.1128/jvi.01955-21

Shi, Jianzhong, Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., Liu, R., He, X., Shuai, L., Sun, Z., Zhao, Y., Liu, P., Liang, L., Cui, P., Wang, J., Zhang, X., Guan, Y., Tan, W., Wu, G., … Bu, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*.

https://doi.org/10.1126/science.abb7015

Shi, M., Wang, L., Fontana, P., Vora, S. M., Zhang, Y., Fu, T.-M., Lieberman, J., & Wu, H. (2020). *SARS-CoV-2 Nsp1 Suppresses Host but Not Viral Translation Through a Bipartite Mechanism*. https://doi.org/10.1101/2020.09.18.302901

Shirato, K., Kanou, K., Kawase, M., & Matsuyama, S. (2017). Clinical Isolates of Human Coronavirus 229E Bypass the Endosome for Cell Entry. *Journal of Virology*. https://doi.org/10.1128/jvi.01387-16

Shrock, E., Fujimura, E., Kula, T., Timms, R. T., Lee, I.-H., Leng, Y., Robinson, M. L., Sie, B. M., Li, M. Z., Chen, Y., Logue, J. K., Zuiani, A., McCulloch, D. J., Lelis, F., Henson, S. N., Monaco, D. R., Travers, M., Habibi, S., Clarke, W., … Wong, C. J. (2020). Viral Epitope Profiling of COVID-19 Patients Reveals Cross-Reactivity and Correlates of Severity. *Science*. https://doi.org/10.1126/science.abd4250

Shroff, A., & Nazarko, T. Y. (2021). The Molecular Interplay Between Human Coronaviruses and Autophagy. *Cells*. https://doi.org/10.3390/cells10082022

Shu, T., Huang, M., Wu, D., Ren, Y., Zhang, X., Han, Y., Mu, J., Wang, R., Qiu, Y., Zhang, D.-Y., & Zhou, X. (2020). SARS-Coronavirus-2 Nsp13 Possesses NTPase and RNA Helicase Activities That Can Be Inhibited by Bismuth Salts. *Virologica Sinica*. https://doi.org/10.1007/s12250-020-00242-1

Silvas, J. A., Vasquez, D. M., Park, J.-G., Chiem, K., Allué-Guardia, A., Garcia-Vilanova, A., Platt, R. N., Miorin, L., Kehrer, T., Cupic, A., Gonzalez-Reiche, A. S., Bakel, H. v., García-Sastre, A., Anderson, T., Torrelles, J. B., Ye, C., & Martínez-Sobrido, L. (2021). Contribution of SARS-CoV-2 Accessory Proteins to Viral Pathogenicity in K18 Human ACE2 Transgenic Mice. *Journal of Virology*. https://doi.org/10.1128/jvi.00402-21

Simonetti, M., Zhang, N., Harbers, L., Milia, M. G., Brossa, S., Nguyen, T. T. H., Cerruti, F., Berrino, E., Sapino, A., Bienko, M., Sottile, A., Ghisetti, V., & Crosetto, N. (2021). *COVseq Is a Cost-Effective Workflow for Mass-Scale SARS-CoV-2 Genomic Surveillance*. https://doi.org/10.21203/rs.3.pex-1338/v3

Simula, E. R., Manca, M. A., Jasemi, S., Uzzau, S., Rubino, S., Manchia, P., Bitti, A.,

Palermo, M., & Sechi, L. A. (2020). HCoV-NL63 and SARS-CoV-2 Share Recognized Epitopes by the Humoral Response in Sera of People Collected Pre- And During CoV-2 Pandemic. *Microorganisms*. https://doi.org/10.3390/microorganisms8121993

Singh, A., Kumar, M., & Dubey, A. K. (2020). Effect of Pre-existing Diseases on COVID-19 Infection and Role of New Sensors and Biomaterials for Its Detection and Treatment. *Medical Devices & Sensors*. https://doi.org/10.1002/mds3.10140

Singh, J., Pandit, P., McArthur, A. G., Banerjee, A., & Mossman, K. (2021). Evolutionary Trajectory of SARS-CoV-2 and Emerging Variants. *Virology Journal*. https://doi.org/10.1186/s12985-021-01633-w

Singh, U., Deb, S., Chaudhry, R., Bala, K., Rani, L., Gupta, R., Kumari, L., Ahmed, J., Gaurav, S., Perumalla, S., Mishra, A., Stephenraj, J., Shukla, J., Bhardwaj, D., Jamshed, N., Aggarwal, P., Kabra, M., Ahuja, V., Sinha, S., & Guleria, R. (2022). *SARS-CoV-2 Omicron Variant Wave in India: Advent, Phylogeny and Evolution*. https://doi.org/10.1101/2022.05.14.491911

Sipulwa, L. A., Ongus, J. R., Coldren, R. L., & Bulimo, W. (2016). Molecular Characterization of Human Coronaviruses and Their Circulation Dynamics in Kenya, 2009–2012. *Virology Journal*. https://doi.org/10.1186/s12985-016-0474-x

*Situation Report-68 HIGHLIGHTS*. (2020).

Sloots, T. P., McErlean, P., Speicher, D., Arden, K. E., Nissen, M. D., & Mackay, I. M. (2006). Evidence of Human Coronavirus HKU1 and Human Bocavirus in Australian Children. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2005.09.008

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M., & Gorbalenya, A. E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(03)00865-9

So, R. T. Y., Chu, D. K., Miguel, E., Perera, R. A., Oladipo, J. O., Fassi-Fihri, O., Aylet, G., Ko, R. C., Zhou, Z., Cheng, M.-S., Kuranga, S. A., Roger, F., Chevalier, V.,

Webby, R. J., Woo, P. C. Y., Poon, L. L. M., & Peiris, M. (2019). Diversity of Dromedary Camel Coronavirus HKU23 in African Camels Revealed Multiple Recombination Events Among Closely Related Betacoronaviruses of the Subgenus Embecovirus. *Journal of Virology*. https://doi.org/10.1128/jvi.01236-19

Song, G., He, W., Callaghan, S., Anzanello, F., Huang, D., Ricketts, J., Torres, J., Beutler, N., Peng, L., Vargas, S., Cassell, J., Parren, M., Yang, L., Ignacio, C., Smith, D. M., Voss, J., Nemazee, D., Ward, A., Rogers, T., … Andrabi, R. (2021). Cross-reactive serum and memory B-cell responses to spike protein in SARS-CoV-2 and endemic coronavirus infection. *Nature Communications*, *12*, null. https://doi.org/10.1038/s41467-021-23074-3

Staquicini, D. I., Tang, F. H. F., Markosian, C., Yao, V. J., Staquicini, F. I., Dodero-Rojas, E., Contessoto, V. G., Davis, D., O'Brien, P., Habib, N., Smith, T. L., Bruiners, N., Sidman, R. L., Gennaro, M. L., Lattime, E. C., Libutti, S. K., Whitford, P. C., Burley, S. K., Onuchic, J. N., … Pasqualini, R. (2021). Design and proof of concept for targeted phage-based COVID-19 vaccination strategies with a streamlined cold-free supply chain. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(30), e2105739118. https://doi.org/10.1073/pnas.2105739118

Stelzig, K. E., Canepa-Escaro, F., Schiliro, M., Berdnikovs, S., Prakash, Y. S., & Chiarella, S. E. (2020). Estrogen regulates the expression of SARS-CoV-2 receptor ACE2 in differentiated airway epithelial cells. *American Journal of Physiology - Lung Cellular and Molecular Physiology*. https://doi.org/10.1152/AJPLUNG.00153.2020

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martínez-Sobrido, L., García-Sastre, A., Weber, F., & Kochs, G. (2007). The Intracellular Sites of Early Replication and Budding of SARS-coronavirus. *Virology*. https://doi.org/10.1016/j.virol.2006.11.027

Stewart, H., Lu, Y., O'Keefe, S., Valpadashi, A., Cruz-Zaragoza, L. D., Michel, H. A., Nguyen, S. K., Carnell, G., Lukhovitskaya, N. I., Milligan, R., Jungreis, I., Lulla, V., Davidson, A. D., Matthews, D. A., High, S., Rehling, P., Emmott, E., Heeney, J. L., Edgar, J. R., … Firth, A. E. (2022). The SARS-CoV-2 protein ORF3c is a mitochondrial modulator of innate immunity. *BioRxiv*. https://doi.org/10.1101/2022.11.15.516323

Streicker, D. G., Lemey, P., Velasco-Villa, A., & Rupprecht, C. E. (2012). Rates of Viral Evolution Are Linked to Host Geography in Bat Rabies. *Plos Pathogens*. https://doi.org/10.1371/journal.ppat.1002720

Strobelt, R., Broennimann, K., Adler, J., & Shaul, Y. (2023). SARS-CoV-2 Omicron Specific Mutations Affecting Infectivity, Fusogenicity, and Partial TMPRSS2-Independency. *Viruses*, *15*, null. https://doi.org/10.3390/v15051129

Su, J., Shen, S., Hu, Y., Chen, S., Cheng, L., Cai, Y., Wei, W., Wang, Y., Rui, Y., & Yu, X. (2022). SARS-CoV-2 ORF3a inhibits cGAS-STING-mediated autophagy flux and antiviral function. *Journal of Medical Virology*, *95*(1). https://doi.org/10.1002/jmv.28175

Su, Y. C. F., Anderson, D. E., Young, B. E., Zhu, F., Linster, M., Kalimuddin, S., Low, J., Yan, Z., Jayakumar, J., Sun, L., Yan, G., Mendenhall, I. H., Leo, Y., Lye, D. C., Wang, L., & Smith, G. (2020). *Discovery of a 382-Nt Deletion During the Early Evolution of SARS-CoV-2*. https://doi.org/10.1101/2020.03.11.987222

Sun, S., Gu, H., Cao, L., Chen, Q., Ye, Q., Yang, G., Li, R., Fan, H., Deng, Y.-Q., Song, X., Qi, Y., Li, M., Lan, J., Feng, R., Guo, Y., Zhu, N., Qin, S., Wang, L., Zhang, Y., … Qin, C.-F. (2021). Characterization and structural basis of a lethal mouse-adapted SARS-CoV-2. *Nature Communications*, *12*(1), 5654. https://doi.org/10.1038/S41467-021-25903-X

Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-López, C., Maatz, H., Reichart, D., Sampaziotis, F., Worlock, K. B., Yoshida, M., Barnes, J. L., Banovich, N. E., Barbry, P., Brazma, A., Collin, J., Desai, T. J., Duong, T. E., … Figueiredo, F. (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*. https://doi.org/10.1038/s41591-020-0868-6

Suresh, S. J., & Suzuki, Y. (2020). SARS-CoV-2 Spike Protein and Lung Vascular Cells. *Journal of Respiration*. https://doi.org/10.3390/jor1010004

Suzuki, Y., & Gychka, S. G. (2021). SARS-CoV-2 Spike Protein Elicits Cell Signaling in Human Host Cells: Implications for Possible Consequences of COVID-19 Vaccines. *Vaccines*. https://doi.org/10.3390/vaccines9010036

Swaim, C. D., Perng, Y.-C., Zhao, X., Canadeo, L. A., Harastani, H. H., Boon, A. C. M., Lenschow, D. J., & Huibregtse, J. M. (2020). *6-Thioguanine Blocks SARS-CoV-2 Replication by Inhibition of PLpro Protease Activities*. https://doi.org/10.1101/2020.07.01.183020

Taguchi, F., Kubo, H., Suzuki, H., & Yamada, Y. K. (1995). Localization of neutralizing epitopes and receptor-binding site in murine coronavirus spike protein. *Advances in Experimental Medicine and Biology*, *380*, 359–365. https://doi.org/10.1007/978-1-4615-1899-0_58

Taguchi, Fumiko, Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., & Ichinose, Y. (2010). The Siderophore Pyoverdine Of <i>Pseudomonas Syringae</I> Pv. Tabaci 6605 Is an Intrinsic Virulence Factor in Host Tobacco Infection. *Journal of Bacteriology*. https://doi.org/10.1128/jb.00689-09

Tahmasebi, S., Khosh, E., & Esmaeilzadeh, A. (2020). The Outlook for Diagnostic Purposes of the 2019-novel Coronavirus Disease. *Journal of Cellular Physiology*. https://doi.org/10.1002/jcp.29804

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., & Du, L. (2020). Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cellular and Molecular Immunology*. https://doi.org/10.1038/s41423-020-0400-4

Taiaroa, G., Rawlinson, D., Featherstone, L., Pitt, M. E., Caly, L., Druce, J., Purcell, D., Harty, L., Tran, T., Roberts, J. A., Scott, N. E., Catton, M., Williamson, D. A., Coin, L., & Duchêne, S. (2020). *Direct RNA Sequencing and Early Evolution of SARS-CoV-2*. https://doi.org/10.1101/2020.03.05.976167

Tajuelo, A., Lopez-Siles, M., Mas, V., Pérez-Romero, P., Aguado, J. M., Briz, V., McConnell, M. J., Martín-Galiano, A. J., & López, D. (2022). Cross-Recognition of SARS-CoV-2 B-Cell Epitopes With Other Betacoronavirus Nucleoproteins. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms23062977

Tan, K. L., Tiong, V., Tan, J., Wong, J.-E., Teoh, B.-T., Abd-Jamil, J., Johari, J., Nor'e, S. S., Khor, C.-S., Yaacob, C.-N., Zulkifli, M.-M.-S., CheMatSeri, A., Mahfodz, N. H.,

Azizan, N. S., & AbuBakar, S. (2021). Multiplex Sequencing of SARS-Cov-2 Genome Directly From Clinical Samples Using the Ion Personal Genome Machine (PGM). *Tropical Biomedicine*. https://doi.org/10.47665/tb.38.3.069

Tan, L., Coenjaerts, F. E. J., Houspie, L., Viveen, M. C., Bleek, G. M. van, Wiertz, E. J. H. J., Martin, D. P., & Lemey, P. (2013). The Comparative Genomics of Human Respiratory Syncytial Virus Subgroups A and B: Genetic Variability and Molecular Evolutionary Dynamics. *Journal of Virology*, *87*(14), 8213–8226. https://doi.org/10.1128/JVI.03278-12

Tan, Y. J., Goh, P. Y., Fielding, B. C., Shen, S., Chou, C. F., Fu, J. L., Leong, H. N., Leo, Y. S., Ooi, E. E., Ling, A. E., Lim, S. G., & Hong, W. (2004). Profiles of Antibody Responses against Severe Acute Respiratory Syndrome Coronavirus Recombinant Proteins and Their Potential Use as Diagnostic Markers. *Clinical and Diagnostic Laboratory Immunology*. https://doi.org/10.1128/CDLI.11.2.362-371.2004

Tao, Y., Shī, M., Chommanard, C., Queen, K., Zhang, J., Markotter, W., Kuzmin, I. V, Holmes, E. C., & Tong, S. (2017). Surveillance of Bat Coronaviruses in Kenya Identifies Relatives of Human Coronaviruses NL63 and 229E and Their Recombination History. *Journal of Virology*. https://doi.org/10.1128/jvi.01953-16

Tarigan, S., Dharmayanti, N. I., Sugiartanti, D., Putri, R., Andriani, Nuradji, H., Robinson, M., Wiendayanthi, N., & Djufri, F. (2023). Characterization of two linear epitopes SARS CoV-2 spike protein formulated in tandem repeat. *PLOS ONE*, *18*(1), e0280627–e0280627. https://doi.org/10.1371/journal.pone.0280627

Tarke, A., Sidney, J., Methot, N., Zhang, Y., Dan, J. M., Goodwin, B., Rubiro, P., Sutherland, A., da Silva Antunes, R., Frazier, A., Rawlings, S. A., Smith, D. M., Peters, B., Scheuermann, R. H., Weiskopf, D., Crotty, S., Grifoni, A., & Sette, A. (2021). Negligible impact of SARS-CoV-2 variants on CD4 + and CD8 + T cell reactivity in COVID-19 exposed donors and vaccinees. *BioRxiv : The Preprint Server for Biology*, 2021.02.27.433180. https://doi.org/10.1101/2021.02.27.433180

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A., Engelbrecht, S., van Zyl, G. V, … de

Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, *592*, 438–443. https://doi.org/10.1038/s41586-021-03402-9

Terry, J. S., Anderson, L. B. R., Scherman, M. S., McAlister, C., Perera, R., Schountz, T., & Geiss, B. J. (2021). Development of a SARS-CoV-2 Nucleocapsid Specific Monoclonal Antibody. *Virology*. https://doi.org/10.1016/j.virol.2021.01.003

Thavorasak, T., Chulanetra, M., Glab-ampai, K., Teeranitayatarn, K., Songserm, T., Yodsheewan, R., Saelim, N., Lekcharoensuk, P., Sookrung, N., & Chaicumpa, W. (2022). Novel Neutralizing Epitope of PEDV S1 Protein Identified by IgM Monoclonal Antibody. *Viruses*. https://doi.org/10.3390/v14010125

Thiel, V., & Weber, F. (2008). Interferon and cytokine responses to SARS-coronavirus infection. In *Cytokine and Growth Factor Reviews*. https://doi.org/10.1016/j.cytogfr.2008.01.001

Tilocca, B., Soggiu, A., Sanguinetti, M., Babini, G., Maio, F. D., Britti, D., Zecconi, A., Bonizzi, L., Urbani, A., & Roncada, P. (2020). Immunoinformatic Analysis of the SARS-CoV-2 Envelope Protein as a Strategy to Assess Cross-Protection Against COVID-19. *Microbes and Infection*. https://doi.org/10.1016/j.micinf.2020.05.013

Tirera, S., Thoisy, B. d., Donato, D., Bouchier, C., Lacoste, V., Franc, A., & Lee, A. (2021). The Influence of Habitat on Viral Diversity in Neotropical Rodent Hosts. *Viruses*. https://doi.org/10.3390/v13091690

Torbati, E., Krause, K., & Ussher, J. E. (2021). The Immune Response to SARS-CoV-2 and Variants of Concern. *Viruses*, *13*, null. https://doi.org/10.3390/v13101911

Tragni, V., F, P., Laera, L., Onofrio, A., Mercurio, I., Todisco, S., Volpicella, M., Grassi, A., & Pierri, C. L. (2022). Modeling SARS-CoV-2 Spike/Ace2 Protein–protein Interactions for Predicting the Binding Affinity of New Spike Variants for ACE2, and Novel ACE2 Structurally Related Human Protein Targets, for COVID-19 Handling in the 3PM Context. *The Epma Journal*. https://doi.org/10.1007/s13167-021-00267-w

Tragni, V., F, P., Laera, L., Onofrio, A., Todisco, S., Volpicella, M., Grassi, A., & Pierri, C. L. (2021). *A Modular Molecular Framework for Quickly Estimating the Binding Affinity of the Spike Protein of SARS-CoV-2 Variants for ACE2, in Presence of*

*Mutations at the Spike Receptor Binding Domain*. https://doi.org/10.1101/2021.05.26.445422

Tregoning, J. S., Flight, K. E., Higham, S. L., Wang, Z., & Pierce, B. F. (2021). Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape. In *Nature Reviews Immunology* (Vol. 21, Issue 10, pp. 626–636). Nature Publishing Group. https://doi.org/10.1038/s41577-021-00592-1

Tripathi, V., Bundel, R., & Mandal, C. C. (2022). Effect of environmental factors on SARS-CoV-2 infectivity in northern hemisphere countries: a 2-year data analysis. *Public Health*, *208*, 105–110. https://doi.org/10.1016/j.puhe.2022.04.011

Tsai, Y., Lee, C.-L., Yen, H.-R., Chang, Y. S., Lin, Y. P., Huang, S., & Lin, C. W. (2020). Antiviral Action of Tryptanthrin Isolated From Strobilanthes Cusia Leaf Against Human Coronavirus NL63. *Biomolecules*. https://doi.org/10.3390/biom10030366

Tsang, H. F., Chan, L. W. C., Cho, W. C. S., Yu, A. C. S., Yim, A. K. Y., Chan, A. K. C., Ng, L. P. W., Wong, Y. K. E., Pei, X. M., Li, M. J. W., & Wong, S.-C. C. (2020). An update on COVID-19 pandemic: the epidemiology, pathogenesis, prevention and treatment strategies. *Expert Review of Anti-Infective Therapy*, 1–12. https://doi.org/10.1080/14787210.2021.1863146

Tsegaye, A. T., Sherry, C., Oduol, C., Otieno, J., Rwigi, D., Masheti, M., Machura, I., Liru, M., Akuka, J., Omedo, D., Symekher, S., Khamadi, S. A., Isaaka, L., Ogero, M., Mumelo, L., Berkley, J. A., Agweyu, A., Walson, J. L., Singa, B. O., & Tickell, K. D. (2023). Clinical epidemiology of COVID-19 among hospitalized children in rural western Kenya. *PLOS Global Public Health*, *3*(6), e0002011. https://doi.org/10.1371/journal.pgph.0002011

Tso, F. Y., Lidenge, S. J., Peña, P. B., Clegg, A. A., Ngowi, J. R., Mwaiselage, J., Ngalamika, O., Julius, P., West, J. T., & Wood, C. (2021). High Prevalence of Pre-Existing Serological Cross-Reactivity Against Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) in Sub-Saharan Africa. *International Journal of Infectious Diseases*. https://doi.org/10.1016/j.ijid.2020.10.104

Tu, T., Budzinska, M. A., Shackel, N. A., & Urban, S. (2017). HBV DNA integration: Molecular mechanisms and clinical implications. *Viruses*, *9*(4).

https://doi.org/10.3390/v9040075

Tung, H. Y. L., & Limtung, P. (2020). Mutations in the Phosphorylation Sites of SARS-CoV-2 Encoded Nucleocapsid Protein and Structure Model of Sequestration by Protein 14-3-3. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2020.08.024

Tuong, H. T., Jeong, J. H., Choi, Y.-K., Park, H., Baek, Y. H., & Yeo, S. J. (2021). Development of a Rapid Fluorescent Diagnostic System to Detect Subtype H9 Influenza a Virus in Chicken Feces. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms22168823

Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. R., Lapointe, H. R., Kamelian, K., Smith, A. D., Prystajecky, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020). *Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore*. https://doi.org/10.1101/2020.09.04.283077

Ullah, M. F., Alnour, T. M., Elssaig, E. H., & Ahmed-Abakur, E. H. (2021). Characterization of Altered Genomic Landscape of SARS-CoV-2 Variants Isolated in Saudi Arabia in a Comparative in Silico Study. *Saudi Journal of Biological Sciences*. https://doi.org/10.1016/j.sjbs.2021.07.054

Urbán, S., Paragi, G., Burián, K., McLean, G. R., & Virók, D. (2020). Identification of Similar Epitopes Between Severe Acute Respiratory Syndrome Coronavirus-2 and Bacillus Calmette–Guérin: Potential for Cross-reactive Adaptive Immunity. *Clinical & Translational Immunology*. https://doi.org/10.1002/cti2.1227

Uttamrao, P. P., Sathyaseelan, C., Patro, L. P. P., & Rathinavelan, T. (2021). Revelation of Potent Epitopes Present in Unannotated ORF Antigens of SARS-CoV-2 for Epitope-Based Polyvalent Vaccine Design Using Immunoinformatics Approach. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2021.692937

van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, A. M., Osterhaus, A. D. M. E., Haagmans, B. L., Gorbalenya, A. E., Snijder, E. J., & Fouchier, R. A. M. (2012). Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*.

https://doi.org/10.1128/mBio.00473-12

Van Der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., Wertheim-Van Dillen, P. M. E., Kaandorp, J., Spaargaren, J., & Berkhout, B. (2004). Identification of a new human coronavirus. *Nature Medicine*, *10*(4), 368–373. https://doi.org/10.1038/nm1024

van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., & Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*. https://doi.org/10.1056/nejmc2004973

Vandervaart, J. P., Inniss, N. L., Ling-Hu, T., Minasov, G., Wiersum, G., Rosas-Lemus, M., Shuvalova, L., Achenbach, C. J., Hultquist, J. F., Satchell, K. J. F., & Bachta, K. E. R. (2023). Serodominant SARS-CoV-2 Nucleocapsid Peptides Map to Unstructured Protein Regions. *Microbiology Spectrum*, *11*(3). https://doi.org/10.1128/spectrum.00324-23

Vaze, N., Soorneedi, A., Moore, M. D., & Demokritou, P. (2022). Inactivating SARS-CoV-2 Surrogates on Surfaces Using Engineered Water Nanostructures Incorporated With Nature Derived Antimicrobials. *Nanomaterials*. https://doi.org/10.3390/nano12101735

Verkhivker, G. M., Agajanian, S., Kassab, R., & Krishnan, K. (2021). *The Landscape-Based Protein Stability Analysis and Network Modeling of Multiple Conformational States of the SARS-CoV-2 Spike D614 Mutant: Conformational Plasticity and Frustration-Driven Allostery as Energetic Drivers of Highly Transmissible Spike Varian*. https://doi.org/10.1101/2021.12.09.471953

Verkhivker, G. M., Agajanian, S., Oztas, D., & Gupta, G. (2021). Computational Analysis of Protein Stability and Allosteric Interaction Networks in Distinct Conformational Forms of the SARS-CoV-2 Spike D614G Mutant: Reconciling Functional Mechanisms through Allosteric Model of Spike Regulation. *BioRxiv*, *null*, null. https://doi.org/10.1080/07391102.2021.1933594

Verma, R., Saha, S., Kumar, S., Mani, S., Maiti, T. K., & Surjit, M. (2021). *RNA-protein Interaction Analysis of SARS-CoV-2 5'- And 3'-Untranslated Regions Identifies an*

*Antiviral Role of Lysosome-Associated Membrane Protein-2*. https://doi.org/10.1101/2021.01.05.425516

Vernet, R., Charrier, E., Grogg, J., & Mach, N. (2021). A Quantitative ELISA Protocol for Detection of Specific Human IgG Against the SARS-CoV-2 Spike Protein. *Vaccines*. https://doi.org/10.3390/vaccines9070770

Vertuani, S., Sette, A., Sidney, J., Southwood, S., Fikes, J., Keogh, E., Lindencrona, J. A., Ishioka, G., Levitskaya, J., & Kiessling, R. (2004). Improved Immunogenicity of an Immunodominant Epitope of the Her-2/Neu Protooncogene by Alterations of MHC Contact Residues. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.172.6.3501

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji, U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., Plessis, L. d., Engelbrecht, S., Everatt, J., Giandhari, J., … Oliveira, T. d. (2022). Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa. *Nature*, *603*(7902), 679–686. https://doi.org/10.1038/s41586-022-04411-y

*Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. (2012). Elsevier. http://books.google.com/books?hl=en&lr=&id=KXRCYay3pH4C&pgis=1

Vishweshwaraiah, Y., Hnath, B., Rackley, B., Wang, J., Gontu, A., Chandler, M., Afonin, K. A., Kuchipudi, S. V, Christensen, N. D., Yennawar, N. H., & Dokholyan, N. V. (2022). Adaptation-Proof SARS-CoV-2 Vaccine Design. *Advanced Functional Materials*. https://doi.org/10.1002/adfm.202206055

Waisner, H., Grieshaber, B., Saud, R., Henke, W., Stephens, E. B., & Kalamvoki, M. (2022). SARS-CoV-2 Harnesses Host Translational Shutoff and Autophagy To Optimize Virus Yields: the Role of the Envelope (E) Protein. *Microbiology Spectrum*, *11*(1). https://doi.org/10.1128/spectrum.03707-22

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. https://doi.org/10.1016/j.cell.2020.02.058

Walsh, E. E., Frenck, R. W., Falsey, A. R., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K. M., Mulligan, M. J., Bailey, R., Swanson, K. A., Li, P., Koury, K., Kalina, W. V, Cooper, D. A., Fontes-Garfias, C. R., Shi, P. Y., Türeci, Ö., Tompkins, K., … Gruber, W. C. (2020). Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa2027906

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology*. https://doi.org/10.1128/jvi.00127-20

Wang, B., Zhong, C., & Tieleman, D. P. (2021). *The Supramolecular Organization of SARS-CoV and SARS-CoV-2 Virions Revealed by Coarse-Grained Models of Intact Virus Envelopes*. https://doi.org/10.1101/2021.09.16.460716

Wang, C., van Haperen, R., Gutiérrez-Álvarez, J., Li, W., Okba, N. M. A., Albulescu, I., Widjaja, I., van Dieren, B., Fernandez-Delgado, R., Sola, I., Hurdiss, D. L., Daramola, O., Grosveld, F., van Kuppeveld, F. J. M., Haagmans, B. L., Enjuanes, L., Drabek, D., & Bosch, B. J. (2021). A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies. *Nature Communications*, *12*(1), null. https://doi.org/10.1038/s41467-021-21968-w

Wang, C. Y., Peng, W.-J., Kuo, B., Ho, Y., Wang, M.-S., Yang, Y.-T., Chang, P.-Y., Shen, Y.-H., & Hwang, K.-P. (2023). Toward a pan-SARS-CoV-2 vaccine targeting conserved epitopes on spike and non-spike proteins for potent, broad and durable immune responses. *PLOS Pathogens*, *19*(4), e1010870–e1010870. https://doi.org/10.1371/journal.ppat.1010870

Wang, Dawei, Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - Journal of the American Medical Association*, *323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585

Wang, Dongliang, Mai, J., Zhou, W., Yu, W., Zhan, Y., Wang, N., Epstein, N. D., & Yang,

Y. (2020). Immunoinformatic Analysis of T- And B-Cell Epitopes for SARS-CoV-2 Vaccine Design. *Vaccines*. https://doi.org/10.3390/vaccines8030355

Wang, F., Chen, C., Tan, W., Yang, K., & Yang, H. (2016). Structure of Main Protease From Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Scientific Reports*. https://doi.org/10.1038/srep22677

Wang, G., Deering, C., Macke, M., Shao, J., Burns, R., Blau, D. M., Holmes, K. V, Davidson, B. L., Perlman, S., & McCray, P. B. (2000). Human Coronavirus 229E Infects Polarized Airway Epithelia From the Apical Surface. *Journal of Virology*. https://doi.org/10.1128/jvi.74.19.9234-9239.2000

Wang, K, Cao, Y., Zhou, Y., Wu, J., Jia, Z., Hu, Y., Yisimayi, A., Fu, W., Wang, L., Liu, P., Fan, K., Chen, R., Jing, L., Wang, Y., Ge, X., Zhang, Q., Wang, N., Wu, W., Gao, Y., … Wang, X. (2021). *A third dose of inactivated vaccine augments the potency, breadth, and duration of anamnestic responses against SARS-CoV-2*. https://doi.org/10.1101/2021.09.02.21261735

Wang, Kai, Long, Q. X., Deng, H. J., Hu, J., Gao, Q. Z., Zhang, G. J., He, C. L., Huang, L. Y., Hu, J. L., Chen, J., Tang, N., & Huang, A. L. (2020). Longitudinal dynamics of the neutralizing antibody response to SARS-CoV-2 infection. In *medRxiv*. https://doi.org/10.1101/2020.07.14.20151159

Wang, Kai, Ping, X., Yu, W., Zhou, Q., Xiong, S., & 孙冰. (2015). The ns12.9 Accessory Protein of Human Coronavirus OC43 Is a Viroporin Involved in Virion Morphogenesis and Pathogenesis. *Journal of Virology*. https://doi.org/10.1128/jvi.01986-15

Wang, Ke, Chen, W., Zhou, Y., Lian, J.-Q., Zhang, Z., Du, P., Gong, L., Zhang, Y., Cui, H., Geng, J.-J., Wang, B., Sun, X., Wang, C., Xu, Y., Peng, L., Deng, Y.-Q., Ding, W., Yang, X., Zhu, Y., … Chen, Z. (2020). *SARS-CoV-2 Invades Host Cells via a Novel Route: CD147-spike Protein*. https://doi.org/10.1101/2020.03.14.988345

Wang, L., Su, S., Bi, Y., Wong, G., & Gao, G. F. (2018). Bat-Origin Coronaviruses Expand Their Host Range to Pigs. *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2018.03.001

Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K. C., Chen, Y. H., Zhang, L., & Wang, X. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Research*, *23*(8), 986–993. https://doi.org/10.1038/cr.2013.92

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P. D., Graham, B. S., Mascola, J. R., Chang, J. Y., Yin, M. T., Sobieszczyk, M., Kyratsous, C. A., Shapiro, L., Sheng, Z., Huang, Y., & Ho, D. D. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*, *593*(7857), 130–135. https://doi.org/10.1038/s41586-021-03398-2

Wang, S., Wu, D., Xiong, H., Wang, J., Tang, Z., Chen, Z., Wang, Y.-Z., Zhang, Y., Ying, D., Lin, X., Liu, C., Guo, S., Tian, W., Lin, Y., Zhang, X., Yuan, Q., Yu, H., Zhang, T., Zheng, Z., & Xia, N. (2021). Quantatitive Analysis of Conserved Sites on the SARS-CoV-2 Receptor-Binding Domain to Promote Development of Universal SARS-Like Coronavirus Vaccines. *BioRxiv*, *null*, null. https://doi.org/10.1101/2021.04.10.439161

Wang, W. B., Liang, Y., Jin, Y. Q., Zhang, J., Su, J. G., & Li, Q. M. (2021). E484K mutation in SARS-CoV-2 RBD enhances binding affinity with hACE2 but reduces interactions with neutralizing antibodies and nanobodies: Binding free energy calculation studies. *Journal of Molecular Graphics & Modelling*, *109*, 108035. https://doi.org/10.1016/J.JMGM.2021.108035

Wang, W., Chen, J., Yu, X., & Lan, H. Y. (2022). Signaling mechanisms of SARS-CoV-2 Nucleocapsid protein in viral infection, cell death and inflammation. *International Journal of Biological Sciences*, *18*(12), 4704–4713. https://doi.org/10.7150/ijbs.72663

Wang, Y., Li, X., Liu, W., Gan, M., Zhang, L., Wang, J., Zhang, Z., Zhu, A., Li, F., Sun, J., Zhang, G., Zhuang, Z., Luo, J., Chen, D., Qiu, S., Zhang, L., Xu, D., Mok, C. K. P., Zhang, F., … Zhao, J. (2020). Discovery of a Subgenotype of Human Coronavirus NL63 Associated With Severe Lower Respiratory Tract Infection in China, 2018. *Emerging Microbes & Infections*. https://doi.org/10.1080/22221751.2020.1717999

Weber, S., Ramirez, C. M., Weiser, B., Burger, H., Doerfler, W., & Doerfler, W. (2021). SARS-CoV-2 worldwide replication drives rapid rise and selection of mutations

across the viral genome: a time-course study - potential challenge for vaccines and therapies. *Embo Molecular Medicine*, *13*(6). https://doi.org/10.15252/EMMM.202114062

Weng, J. R., Lin, C. S., Lai, H. C., Lin, Y. P., Wang, C. Y., Tsai, Y., Wu, K. C., Huang, S., & Lin, C. W. (2019). Antiviral Activity of Sambucus FormosanaNakai Ethanol Extract and Related Phenolic Acid Constituents Against Human Coronavirus NL63. *Virus Research*. https://doi.org/10.1016/j.virusres.2019.197767

Wevers, B. A., & Hoek, L. v. d. (2010). Renin–angiotensin System in Human Coronavirus Pathogenesis. *Future Virology*. https://doi.org/10.2217/fvl.10.4

WHO. (2020). Weekly Operational Update on COVID-19. *Emergency Situational Updates*, *77*, 1–10. https://www.who.int/publications/m/item/weekly-update-on-covid-19---2-october-2020

WHO. (2021). *WHO Coronavirus Disease Dashboard*. WHO.Int. https://covid19.who.int/%0Ahttps://covid19.who.int/%0Ahttps://covid19.who.int/%0Ahttps://covid19.who.int/%0Ahttps://covid19.who.int/region/searo/country/bd%0Ahttps://covid19.who.int/%0Ahttps://covid19.who.int/?gclid=CjwKCAjwnK36BRBVEiwAsMT8WJ3y00_BUzvrLsv

WHO Headquarters, & WHO Worldwide. (2020). *Situation Report-80 HIGHLIGHTS*. https://apps.who.int/iris/handle/10665/331778

WHOa. (2021). *WHO AFRICA REPORT 2021*. https://doi.org/www.afro.who.int/health-topics/coronavirus-covid-19

WHOb. (2021). *WHO World COVID 19 Situation in 2021*. https://doi.org/covid19.who.int/

Willemze, A., Woude, D. v. d., Ghidey, W., Levarht, E. W. N., Stoeken-Rijsbergen, G., Verduyn, W., Vries, R. R. P. De, Houwing-Duistermaat, J. J., Huizinga, T. W. J., Trouw, L. A., & Toes, R. E. M. (2011). The Interaction Between HLA Shared Epitope Alleles and Smoking and Its Contribution to Autoimmunity Against Several Citrullinated Antigens. *Arthritis & Rheumatism*. https://doi.org/10.1002/art.30409

Williams, G., & Vethan, A. (2023). SARS-CoV-2 spike protein reduces burst activities in neurons measured by micro-electrode arrays. *BioRxiv*, *85*, 3469–3476.

https://doi.org/10.1097/MS9.0000000000000950

Wilson, C. C., McKinney, D. M., Anders, M., MaWhinney, S., Forster, J. E., Crimi, C., Southwood, S., Sette, A., Chesnut, R. W., Newman, M. J., & Livingston, B. (2003). Development of a DNA Vaccine Designed to Induce Cytotoxic T Lymphocyte Responses to Multiple Conserved Epitopes in HIV-1. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.171.10.5611

Wilson, G. (2022). *SARS-CoV-2 Seroprevalence in three Kenyan Health and Demographic Surveillance Sites, December 2020-May 2021*. https://doi.org/10.1101/2022.02.07.22270012

Winger, A., & Caspari, T. (2021). The Spike of Concern—The Novel Variants of SARS-CoV-2. *Viruses*, *13*, null. https://doi.org/10.3390/v13061002

Woldemeskel, B. A., Garliss, C. C., & Blankson, J. N. (2021). SARS-CoV-2 mRNA Vaccines Induce Broad CD4+ T Cell Responses That Recognize SARS-CoV-2 Variants and HCoV-NL63. *Journal of Clinical Investigation*. https://doi.org/10.1172/jci149335

Wolter, N., Tempia, S., Gottberg, A. von, Bhiman, J. N., Walaza, S., Kleynhans, J., Moyes, J., Buys, A., McMorrow, M., Aitken, S., Magni, S., Yun, J. A., Fellows, T., Maakamedi, T., Weiner, R., Cawood, C., Martinson, N. A., Lebina, L., Jassat, W., … Cohen, C. (2022). Seroprevalence of SARS-CoV-2 after the second wave in South Africa in HIV-infected and uninfected persons: a cross-sectional household survey. *Clinical Infectious Diseases*, *75*(1), e57–e68. https://doi.org/10.1093/cid/ciac198

Woo, P. C. Y., Lau, S. K. P., Chu, C., Chan, K., Tsoi, H., Huang, Y., Wong, B. H. L., Poon, R. W. S., Cai, J. J., Luk, W., Poon, L. L. M., Wong, S. S. Y., Guan, Y., Peiris, J. S. M., & Yuen, K. (2005). Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *Journal of Virology*. https://doi.org/10.1128/jvi.79.2.884-895.2005

Woo, P. C. Y., Lau, S. K. P., Huang, Y., & Yuen, K. Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. In *Experimental Biology and Medicine* (Vol. 234, Issue 10, pp. 1117–1127). https://doi.org/10.3181/0903-MR-94

Woodworth, J. S., Shin, D. S., Volman, M., Nunes-Alves, C., Fortune, S. M., & Behar, S. M. (2011). <i>Mycobacterium Tuberculosis</I>Directs Immunofocusing of CD8+ T Cell Responses Despite Vaccination. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.1002911

Wrapp, D., & McLellan, J. S. (2019). The 3.1-Angstrom Cryo-Electron Microscopy Structure of the Porcine Epidemic Diarrhea Virus Spike Protein in the Prefusion Conformation. *Journal of Virology*. https://doi.org/10.1128/jvi.00923-19

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., & Jiang, T. (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host and Microbe*. https://doi.org/10.1016/j.chom.2020.02.001

Wu, D., Колесников, A. C., Yin, R., Gowthaman, R., Shmelev, A., Serdyuk, Y., Efimov, G. A., Pierce, B. G., & Mariuzza, R. A. (2021). *Structural Basis for Recognition of Two HLA-A2-restricted SARS-CoV-2 Spike Epitopes by Public and Private T Cell Receptors*. https://doi.org/10.1101/2021.07.28.454232

Wu, H. S., Hsieh, Y. C., Su, I. J., Lin, T. H., Chiu, S. C., Hsu, Y. F., Lin, J. H., Wang, M. C., Chen, J. Y., Hsiao, P. W., Chang, G. D., Wang, A. H. J., Ting, H. W., Chou, C. M., & Huang, C. J. (2004). Early detection of antibodies against various structural proteins of the SARS-associated coronavirus in SARS patients. *Journal of Biomedical Science*. https://doi.org/10.1159/000075294

Wu, L., Peng, C., Yang, Y., Shi, Y., Zhou, L., Xu, Z., & Zhu, W. (2021). Exploring the immune evasion of SARS-CoV-2 variant harboring E484K by molecular dynamics simulations. *Briefings in Bioinformatics*. https://doi.org/10.1093/BIB/BBAB383

Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: A Customizable Web Server for Fast Metagenomic Sequence Analysis. *BMC Genomics*. https://doi.org/10.1186/1471-2164-12-444

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S., & Lu, L. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. In *Cellular and Molecular Immunology* (Vol. 17, Issue 7, pp.

765–767). Springer Nature. https://doi.org/10.1038/s41423-020-0374-2

Xiang, J., Su, J., Lan, Q., Zhao, W., Zhou, Y., Xu, Y., Ni, J., Xia, S., Qi, Q., Sidhu, S. S., Lu, L., Miersch, S., & Yang, B. (2022). Antigenic Mapping Reveals Sites of Vulnerability on A-HCoV Spike Protein. *Communications Biology*. https://doi.org/10.1038/s42003-022-04160-8

Xie, Y., Karki, C. B., Du, D., Li, H., Wang, J., Sobitan, A., Teng, S., Tang, Q., & Li, L. (2020). Spike Proteins of SARS-CoV and SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. *Frontiers in Molecular Biosciences*. https://doi.org/10.3389/fmolb.2020.591873

Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. In *Science China Life Sciences*. https://doi.org/10.1007/s11427-020-1637-5

Yamamoto, M., Tomita, K., Hirayama, Y., Inoue, J., Kawaguchi, Y., & Gohda, J. (2022). SARS-CoV-2 Omicron spike H655Y mutation is responsible for enhancement of the endosomal entry pathway and reduction of cell surface entry pathways. *BioRxiv*, *null*, null. https://doi.org/10.1101/2022.03.21.485084

Yamamoto, Y., Inoue, T., Inoue, M., Murae, M., Fukasawa, M., Kaneko, M., Kato, Y., & Noguchi, K. (2022). SARS-Cov-2 Spike Protein Mutation at Cysteine-488 Impairs Its Golgi Localization And Intracellular S1/S2 Processing. *International Journal of Molecular Sciences*, *23*(24). https://doi.org/10.20944/preprints202210.0245.v1

Yan, B., Chakravorty, S., Mirabelli, C., Wang, L., Trujillo-Ochoa, J. L., Chauss, D., Kumar, D., Lionakis, M. S., Olson, M. R., Wobus, C. E., Afzali, B., & Kazemian, M. (2021). Host-Virus Chimeric Events in SARS-CoV-2-Infected Cells Are Infrequent and Artifactual. *Journal of Virology*. https://doi.org/10.1128/jvi.00294-21

Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M., Wang, Q., Rao, Z., & Lou, Z. (2020). Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex. *Nature Communications*. https://doi.org/10.1038/s41467-020-19770-1

Yang, G., Wang, J., Sun, P., Qin, J., Yang, X., Chen, D., Zhang, Y., Zhong, N., & Wang, Z. (2023). SARS-CoV-2 Epitope-Specific T Cells: Immunity Response Feature, TCR Repertoire Characteristics and Cross-Reactivity. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2023.1146196

Yang, W.-T., Huang, W. H., Liao, T.-L., Hsiao, T. H., Chuang, H.-N., & Liu, P. Y. (2022). SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations. *Infection and Drug Resistance*, *15*, 373–385. https://doi.org/10.2147/IDR.S344099

Yang, W., Dang, X., Wang, Q., Xu, M., Zhao, Q., Yang, Z., Zhao, H., Wang, L., Xu, Y. X., Wang, J., Han, S., Wang, M., Li, H., Pei, F., & Wang, Y. (2020). *Rapid and Accurate Detection of Three Genes in SARS-CoV-2 Using RT-LAMP Method*. https://doi.org/10.21203/rs.3.rs-28070/v1

Yang, Z.-J., Han, Y., Ding, S., Finzi, A., Mothes, W., & Lu, M. (2021). SARS-CoV-2 variants exhibit increased kinetic stability of open spike conformations as an evolutionary strategy. *BioRxiv*, *null*, null. https://doi.org/10.1101/2021.10.11.463956

Yang, Z.-J., Han, Y., Ding, S., Shi, W., Zhou, T., Finzi, A., Kwong, P., Mothes, W., & Lu, M. (2022). SARS-CoV-2 Variants Increase Kinetic Stability of Open Spike Conformations as an Evolutionary Strategy. *MBio*, *13*, null. https://doi.org/10.1128/mbio.03227-21

Ye, C., Park, J.-G., Chiem, K., Dravid, P., Guardia, A. A., Garcia-Vilanova, A., Kapoor, A., Walter, M. R., Kobie, J. J., Plemper, R. K., Torrelles, J. B., & Martínez-Sobrido, L. (2022). *Immunization With Recombinant Accessory Protein-Deficient SARS-CoV-2 Protects Against Lethal Challenge and Viral Transmission*. https://doi.org/10.1101/2022.03.13.484172

Yip, C. C. Y., Lam, C. S. F., Luk, H. K., Wong, E. Y., Lee, R. A., So, L.-Y., Chan, K.-H., Cheng, V. C., Yuen, K.-Y., Woo, P. C. Y., & Lau, S. K. (2016). A Six-Year Descriptive Epidemiological Study of Human Coronavirus Infections in Hospitalized Patients in Hong Kong. *Virologica Sinica*. https://doi.org/10.1007/s12250-016-3714-8

Yong, C. Y., Liew, W.-P.-P., Ong, H. K., & Poh, C. L. (2022). Development of Virus-like Particles-based Vaccines Against Coronaviruses. *Biotechnology Progress*.

https://doi.org/10.1002/btpr.3292

Yoon, S., Anaya, E., Sachithanandham, J., Pinsky, B. A., Sullivan, D., Mostafa, H. H., & Pekosz, A. (2023). Convergent Evolution of A-Lineage (Clade 19B) SARS-CoV-2 Spike Sequences with B-Lineage Variants of Concern Affects Virus Replication in a Temperature-Dependent Manner on Human Nasal Epithelial Cell Cultures. *BioRxiv*. https://doi.org/10.1101/2023.03.03.531067

Yount, B., Roberts, R., Sims, A. C., Deming, D. J., Frieman, M. B., Sparks, J. S., Denison, M. R., Davis, N., & Baric, R. S. (2005). Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *Journal of Virology*. https://doi.org/10.1128/jvi.79.23.14909-14922.2005

Yu, J., Zhang, R.-F., Xie, Z., Yan, K., Gao, H., Song, J., Yuan, X., Cheng, W., Hou, Y., & Duan, Z. (2012). Newly Identified Respiratory Viruses Associated With Acute Lower Respiratory Tract Infections in Children in Lanzou, China, From 2006 to 2009. *Clinical Microbiology and Infection*. https://doi.org/10.1111/j.1469-0691.2011.03541.x

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T., Lv, H., Mok, C. K. P., & Wilson, I. A. (2020). A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. *Science*. https://doi.org/10.1126/science.abb7269

Yurkovetskiy, L., Egri, S. B., Kurhade, C., Díaz-Salinas, M. A., Jaimes, J. A., Nyalile, T., Xie, X., Choudhary, M. C., Dauphin, A., Li, J. Z., Munro, J. B., Shi, P. Y., Shen, K., & Luban, J. (2023). S:D614G and S:H655Y are gateway mutations that act epistatically to promote SARS-CoV-2 variant fitness. *BioRxiv*. https://doi.org/10.1101/2023.03.30.535005

Zaheer, T., Khan, M. T. S., Dar, H. A., Ashraf, S. T., Paracha, R. Z., & Ali, A. (2018). *Prediction of CD8 and CD4 T Cell Epitopes in the Polyprotein of Zika Virus; An Immunoinformatics Approach*. https://doi.org/10.7287/peerj.preprints.26499v1

Zarandi, P. K., Zinatizadeh, M. R., Zinatizadeh, M., Yousefi, M., & Rezaei, N. (2021). SARS-CoV-2: From the pathogenesis to potential anti-viral treatments. *Biomedicine & Pharmacotherapy*, *137*, 111352. https://doi.org/10.1016/j.biopha.2021.111352

Zekri, A.-R. N., Mohanad, M., Hafez, M. M., Soliman, H., Hassan, Z. K., Abouelhoda, M., Amer, K., Seadawy, M. G., & Ahmed, O. (2021). Genome Sequencing of SARS-CoV-2 in a Cohort of Egyptian Patients Revealed Mutation Hotspots That Are Related to Clinical Outcomes. *Biochimica Et Biophysica Acta (Bba) - Molecular Basis of Disease*. https://doi.org/10.1016/j.bbadis.2021.166154

Zeng, H., Dichio, V., Horta, E. R., Thorell, K., & Aurell, E. (2020). Global Analysis of More Than 50,000 SARS-CoV-2 Genomes Reveals Epistasis Between Eight Viral Genes. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.2012331117

Zeng, Z., Chen, D., Tan, W., Qiu, S., Xu, D., Liang, H.-X., Chen, M., Li, X., Lin, Z., Liu, W., & Zhou, Z. (2017). Epidemiology and Clinical Characteristics of Human Coronaviruses OC43, 229E, NL63, and HKU1: A Study of Hospitalized Children With Acute Respiratory Tract Infection in Guangzhou, China. *European Journal of Clinical Microbiology & Infectious Diseases*. https://doi.org/10.1007/s10096-017-3144-z

Zhan, J., Xie, Y., Wang, J., & Hang, X. (2022). *Investigating the Optimum Sample Type and Target Genes for SARS-CoV-2 Detection*. https://doi.org/10.1101/2022.05.25.22275564

Zhan, S. H., Alamouti, S. M., Kwok, B., Lee, M.-H., Khattra, J., Daneshpajouh, H., Houck, H., & Rand, K. H. (2021). *Target Capture Sequencing of SARS-CoV-2 Genomes Using the ONETest Coronaviruses Plus*. https://doi.org/10.1101/2021.03.25.437083

Zhang, B., Zhang, B., Hu, Y. F., Hu, Y. F., Chen, L., Yau, T., Tong, Y., Hu, J., Cai, J.-P., Chan, K.-H., Dou, Y., Deng, J., Wang, X., Hung, I., To, K. K. W., Yuen, K.-Y., Huang, J.-D., & Huang, J.-D. (2020). Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Research*, *30*(8), 702–704. https://doi.org/10.1038/S41422-020-0366-X

Zhang, D., Yang, Y., Li, M., Lu, Y., Liu, Y., Jiang, J., Liu, R., Liu, J., Huang, X., Li, G., & Qu, J. (2022). Ecological Barrier Deterioration Driven by Human Activities Poses Fatal Threats to Public Health Due to Emerging Infectious Diseases. *Engineering*. https://doi.org/10.1016/j.eng.2020.11.002

Zhang, J., Ejikemeuwa, A., Gerzanich, V., Nasr, M., Tang, Q., Simard, J. M., & Zhao, R. Y. (2022). Understanding the Role of SARS-CoV-2 ORF3a in Viral Pathogenesis and COVID-19. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.854567

Zhang, L., Gan, M., Zhang, Z., Li, X., Liu, W., Zhu, A., Sun, J., Li, F., Wang, Y., Zhang, F., Zhao, J., Zhou, R., & Zhao, J. (2020). Complete Genome Sequences of Five Human Coronavirus NL63 Strains Causing Respiratory Illness in Hospitalized Children in China. *Microbiology Resource Announcements*. https://doi.org/10.1128/mra.01597-19

Zhang, N., Jiang, S., & Du, L. (2014). Current advancements and potential strategies in the development of MERS-CoV vaccines. In *Expert Review of Vaccines* (Vol. 13, Issue 6, pp. 761–774). Expert Reviews Ltd. https://doi.org/10.1586/14760584.2014.912134

Zhang, N., Wang, L., Deng, X., Liang, R., Su, M., He, C., Hu, L., Su, Y., Ren, J., Yu, F., Du, L., & Jiang, S. (2020). Recent Advances in the Detection of Respiratory Virus Infection in Humans. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25674

Zhang, Q., Chen, Z., Huang, C., Sun, J., Xue, M., Feng, T., Pan, W., Wang, K., & Dai, J. (2021). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Membrane (M) and Spike (S) Proteins Antagonize Host Type I Interferon Response. *Frontiers in Cellular and Infection Microbiology*. https://doi.org/10.3389/fcimb.2021.766922

Zhang, S., Tuo, J., Huang, X., Zhu, X., Zhang, D., Zhou, K., Yuan, L., Luo, H., Zheng, B., Yuen, K.-Y., Li, M., Chen, K., & Xu, L. (2018). Epidemiology Characteristics of Human Coronaviruses in Patients With Respiratory Infection Symptoms and Phylogenetic Analysis of HCoV-OC43 During 2010-2015 in Guangzhou. *Plos One*. https://doi.org/10.1371/journal.pone.0191789

Zhang, W., Qin, Z., Yan, M., Chen, X., Yang, H., Zhou, W., & Rao, Z. (2018). Structural Characterization of the HCoV-229E Fusion Core. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2018.02.136

Zhang, Yiting, Wu, H., Li, W., Zhou, C., Lin, K., Wang, J., Wang, X., & Wang, L. (2023). A comprehensive overview on the transmission, pathogenesis, diagnosis, treatment, and prevention of SARS-CoV-2. *Journal of Medical Virology*, *95*(5). https://doi.org/10.1002/jmv.28776

Zhang, Yujiang, Xu, W., Shen, K., Xie, Z., Sun, L., Lu, Q., Liu, C., Liang, G., Beeler, J. A., & Anderson, L. J. (2007). Genetic variability of group A and B human respiratory syncytial viruses isolated from 3 provinces in China. *Archives of Virology*, *152*(8), 1425–1434. https://doi.org/10.1007/S00705-007-0984-3

Zhang, Z., Nomura, N., Muramoto, Y., Ekimoto, T., Uemura, T., Liu, K., Yui, M., Kono, N., Aoki, J., Ikeguchi, M., Noda, T., Iwata, S., Ohto, U., & Shimizu, T. (2022). Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-32019-3

Zhao, C., Geng, Q., Niu, J., Wang, Z., Wang, C., Ren, J., & Qu, X. (2020). Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angewandte Chemie*. https://doi.org/10.1002/anie.202011419

Zheng, M., & Lun, S. (2020). Novel Antibody Epitopes Dominate the Antigenicity of Spike Glycoprotein in SARS-CoV-2 Compared to SARS-CoV. *Cellular and Molecular Immunology*. https://doi.org/10.1038/s41423-020-0385-z

Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. Di, Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., … Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

Zhou, T., Wang, M., Cheng, A., Yang, Q., Tian, B., Wu, Y., Jia, R., Chen, S., Liu, M., Zhao, X., Ou, X., Mao, S., Sun, D., Zhang, S., Zhu, D., Huang, J., Gao, Q., Yu, Y., & Zhang, L. (2022). Regulation of Alphaherpesvirus Protein via Post-Translational Phosphorylation. *Veterinary Research*. https://doi.org/10.1186/s13567-022-01115-z

Zhu, Yichao, Zhou, W., Niu, Z., Sun, J., Zhang, Z., Li, Q., Zheng, Y.-X., Wang, C., Gao, L., & Sun, Q. (2022). Long-range enhancement of N501Y-endowed mouse infectivity of SARS-CoV-2 by the non-RBD mutations of Ins215KLRS and H655Y. *Biology Direct*, *17*(1). https://doi.org/10.1186/s13062-022-00325-x

Zhu, Yuanmei, Yu, D., Han, Y., Huang, Y., Chong, H., Ren, L., Wang, J., Li, T., & He, Y. (2020). Cross-Reactive Neutralization of SARS-CoV-2 by Serum Antibodies From Recovered SARS Patients and Immunized Animals. *Science Advances*.

https://doi.org/10.1126/sciadv.abc9999

Zhu, Z., Chu, H., Wen, L., Yuan, S., Chik, K. K., Yuen, T. T., Yip, C. C., Wang, D., Zhou, J., Yin, F., Jin, D.-Y., Kok, K.-H., Yuen, K.-Y., & Chan, J. F. (2019). Targeting SUMO Modification of the Non-Structural Protein 5 of Zika Virus as a Host-Targeting Antiviral Strategy. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20020392

# APPENDICIES

## App: 1 APPENDIX 1: STATISTICAL ANALYSES

### App: 1.1 Statistical Analysis of SARS-CoV-2 Mutation Clades, Symptoms, and Gender

### App: 1.1.1 Association Between Clade and Symptom Status Variants of concern

#### App: 1.1.1.1 Chi-Square Test of Independence

**Purpose:**
To determine if there is an association between categorical variables (clade and symptom status).

**Questions Addressed:**
- Is there an association between clade and symptom status?

**Hypotheses:**
- Null Hypothesis ($H_0$): There is no association between the variables.
- Alternative Hypothesis ($H_1$): There is an association between the variables.

**Contingency Table for Clade vs. Symptom Status:**

| CLADE | SYMPTOMATIC | ASYMPTOMATIC | TOTAL |
|---|---|---|---|
| BA.1.1 | 32 | 6 | 38 |
| BA.1.1.1 | 2 | 0 | 2 |
| BA.1 | 1 | 0 | 1 |
| BA.1.14 | 1 | 0 | 1 |
| AY.46 | 2 | 0 | 2 |
| Total | 38 | 6 | 44 |

**Steps:**

1. Calculate the expected frequencies for each cell under the null hypothesis (no association between clade and symptom status).
2. Use the formula for the chi-square statistic:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

3. Determine the degrees of freedom: $df = (\text{rows} - 1) \times (\text{columns} - 1) = (5 - 1) \times (2 - 1) = 4$.
4. Use a chi-square distribution table or software to find the p-value corresponding to the calculated $\chi^2$ value and degrees of freedom.

The chi-square test evaluates whether there is an association between clade and symptom status:. $\chi 2=11.65$; $p=0.02$; $df=4$

Since the p-value ($p=0.02$) is less than 0.05, we reject the null hypothesis. There is a significant association between clade and symptom status.

### App: 1.1.1.2 Fisher's Exact Test

**Purpose**:
To test for associations in small sample sizes or when the assumptions of the chi-square test are violated (e.g., expected frequencies < 5).
**Questions Addressed**:

- Is there an association between clade and symptom status for smaller clades like BA.1.1.1 or BA.1.14?

Fisher's exact test is used when sample sizes are small or expected frequencies are less than 5. For example:

- BA.1.1.1: Symptomatic = 2, Asymptomatic = 0.
- BA.1.14: Symptomatic = 1, Asymptomatic = 0.

Use software (e.g., R, Python) or online calculators to compute the p-value for these small tables.

**Example for BA.1.1.1:**

Contingency table:

| SYMPTOMATIC | ASYMPTOMATIC |
|---|---|
| 2 | 0 |
| 36 | 6 |

Use Fisher's exact test formula or software to calculate the p-value.

### Fisher's Exact Test for BA.1.1.1

The Fisher's exact test evaluates whether there is an association between symptom status and the BA.1.1.1 clade.

- **Odds Ratio** : Undefined (division by zero because there are no asymptomatic cases)
- **P-Value** : p=1.0

The p-value is not significant, so we fail to reject the null hypothesis. There is no evidence of an association between symptom status and the BA.1.1.1 clade.

### Fisher's Exact Test for BA.1.14

The Fisher's exact test evaluates whether there is an association between symptom status and the BA.1.14 clade.

- **Odds Ratio** : Undefined (division by zero because there are no asymptomatic cases)
- **P-Value** : p=1.0

The p-value is not significant, so we fail to reject the null hypothesis. There is no evidence of an association between symptom status and the BA.1.14 clade.

### App: 1.1.2 Comparison of Symptomatic Proportions by Gender

#### App: 1.1.2.1 Proportion Comparison (Z-Test)

**Purpose**:
To compare proportions of symptomatic patients between two groups (e.g., males vs. females).

**Questions Addressed:**
Are males more likely to be symptomatic than females?

**Hypotheses**:
Null Hypothesis ($H_0$): The proportions are equal.
Alternative Hypothesis ($H_1$): The proportions differ.

**Steps:**

1. Calculate proportions:
   - Proportion of symptomatic males: $p_1 = \frac{16}{18}$
   - Proportion of symptomatic females: $p_2 = \frac{22}{26}$
2. Compute pooled proportion:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{16 + 22}{18 + 26} = 0.8636$$

3. Compute standard error:

$$SE = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

4. Compute z-statistic:

$$z = \frac{p_1 - p_2}{SE}$$

5. Compare z-statistic to critical value from the standard normal distribution.

**Interpretation**:
If the p-value (z=0.45, p=0.653) is less than the significance level, reject the null hypothesis.

### App: 1.1.3 The relative risk (RR) of being symptomatic for males compared to females

#### App: 1.1.3.1 Relative Risk (RR) and Odds Ratio (OR)

**Purpose**:
To quantify the strength of association between clades, gender, and symptom status.

**Questions Addressed**:
How much more likely are patients with a specific clade to be symptomatic compared to others?
How much more likely are males to be symptomatic compared to females?

**Relative Risk:**

$$RR = \frac{\text{Proportion Symptomatic in Males}}{\text{Proportion Symptomatic in Females}}$$

$$RR = \frac{\frac{16}{18}}{\frac{22}{26}} = 1.15$$

**Odds Ratio:**

$$OR = \frac{\text{Odds Symptomatic in Males}}{\text{Odds Symptomatic in Females}}$$

$$OR = \frac{\frac{16}{2}}{\frac{22}{4}} = 1.45$$

**Interpretation**:
RR > 1 indicates higher risk in males.
OR > 1 indicates higher odds in males.
These values suggest a slight increase in the likelihood of being symptomatic for males, though not statistically significant.

### App: 1.1.4 The association between clade and symptom status while controlling for gender

#### App: 1.1.4.1 Cochran-Mantel-Haenszel Test

**Purpose**:
To assess the association between two variables while controlling for a third variable (e.g., clade and symptom status while controlling for gender).

**Questions Addressed:**
Is there an association between clade and symptom status after adjusting for gender?

**Hypotheses:**
Null Hypothesis ($H_0$): No association exists between clade and symptom status after controlling for gender.
Alternative Hypothesis ($H_1$): An association exists between clade and symptom status after controlling for gender.

**Steps:**

1. Stratify the data by gender.
2. Compute the CMH statistic:

$$CMH = \frac{(\sum[A_i - E(A_i)])^2}{\sum V(A_i)}$$

where $A_i$ is the observed count in stratum $i$, $E(A_i)$ is the expected count, and $V(A_i)$ is the variance.

3. Use software (e.g., R, Python) to compute the test.

**Interpretation**
Test yielded a significant result (CMH=11.23, p=0.024), indicating that the association between clade and symptom status remained significant after adjusting for gender.

## App: 1.2 Statistical Analysis of SARS-CoV-2 Mutation Frequencies Across Proteins

### App: 1.2.1 Association between protein type and mutation frequency

#### App: 1.2.1.1 Chi-Square Test of Independence

**Purpose:**
To test for associations between categorical variables (e.g., protein and mutation frequency).

**Questions Addressed:**
Is there an association between the protein and the frequency of mutations?
Do certain proteins have significantly higher mutation rates than others?

**Hypotheses:**
Null Hypothesis ($H_0$): Mutation frequency is independent of the protein.
Alternative Hypothesis ($H_1$): Mutation frequency depends on the protein.

**Steps:**

1. **Create Contingency Table :**

| PROTEIN | LOW FREQUENCY (<50%) | HIGH FREQUENCY (>50%) | TOTAL |
|---|---|---|---|
| NSP3 | 1 | 10 | 11 |
| Spike | 5 | 15 | 20 |
| ORF3a | 0 | 1 | 1 |
| Others | 3 | 8 | 11 |
| Total | 9 | 34 | 43 |

2. **Calculate Expected Frequencies :**
   For each cell:

$$E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

   Example for NSP3 (Low Frequency):

$$E = \frac{11 \times 9}{43} = 2.28$$

3. **Calculate Chi-Square Statistic :**

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

   Example for NSP3 (Low Frequency):

$$\chi^2 = \frac{(1 - 2.28)^2}{2.28} = 0.52$$

   Repeat for all cells.

4. **Degrees of Freedom :**

$$df = (\text{Rows} - 1) \times (\text{Columns} - 1)$$

   Here, $df = (4 - 1) \times (2 - 1) = 3$.

5. **Compare to Critical Value :**
   Use a chi-square distribution table with $df = 3$ and $\alpha = 0.05$. If $\chi^2 >$ critical value, reject the null hypothesis.

The results showed a significant association ($\chi 2(6)=18.45$, $p=0.005$), suggesting that mutation frequencies vary significantly across proteins.

### App: 1.2.2 To predict the probability of a mutation being frequent (>50%) based on protein type and position.

#### App: 1.2.2.1 Logistic Regression Model

**Purpose:**
To model the probability of a mutation being frequent (e.g., >50%) based on predictors such as protein type or position.
**Questions Addressed:**
What factors (e.g., protein, position) predict whether a mutation will be frequent?
Are mutations in certain proteins more likely to reach high frequencies?

**Model:**

$$\text{logit}(P(\text{Frequent})) = \beta_0 + \beta_1 \cdot \text{Protein} + \beta_2 \cdot \text{Position}$$

**Interpretation:**
The coefficients (β1,β2) indicate the effect of protein and position on the log-odds of a mutation being frequent. Exponentiating these coefficients gives odds ratios. A logistic regression model was Mutations in the Spike protein were significantly more likely to be frequent compared to other proteins (OR = 3.42, 95% CI: 1.12–10.34, p=0.031).

### App: 1.2.3 To explore interactions between mutations and identify co-occurring mutations

#### App: 1.2.3.1 Network Analysis

**Purpose:**
Identify co-occurring mutations.
**Questions Addressed:**
Are certain mutations more likely to co-occur in the same subjects?
Do co-occurring mutations suggest compensatory effects or functional dependencies?

Network analysis revealed strong co-occurrence patterns among mutations in the Spike protein, particularly D614G, P681H, and N501Y. These mutations formed a densely connected cluster, indicating potential functional dependencies

# App: 2 APPENDIX 2: CONSERVED SARS-COV-2 EPITOPES

## B Cell epitopes

AACCHLAKALNDFSNS; AACRKVQHMVVKAAL; AAEASKKPRQKRTAT; AAEASKKPRQKRTATKAYNV; AAEIRASANLAATKM; AAFATAQEAYEQAVAN; AAISDYDYYRYNLPTM; AAIVLQLPQGTTLPKG; AAKAYKDYLASGGQPI; AAKKNNLPFKLTCATT; AALCTFLLNKEMYLK; AALLADKFPVLH; AALQIPFAMQMAYRF; AALQIPFAMQMAYRFN; AALQIPFAMQMAYRFNGIGVTQNVL; AALTNNVAFQTVKPGN; AARQGFVDSDVETKD; AARVVRSIFSRTLET; AASGNLLLDKRTTC; AATRGATVVIGTSKF; AATRGATVVIGTSKFY; AATVRLQAGNATEVPA; AAVNSVPWDTIANYA; AAVNSVPWDTIANYAK; AAYSRYRIGNYKLNTD; ACCHLAKALNDFSNSG; ACHNSEVGPEHSLAE; ACIDCSARHINAQVA; ACLVGLMWLSYFIAS; ACVLAAECTIFKDAS; ADDLNQLTGYKKPAS; ADDLNQLTGYKKPASR; ADIVEEAKKVKPTVVV; ADIVVFDEISMATNY; ADIVVFDEISMATNYD; ADKYVRNLQHRLYEC; ADLKSFDLGDELGTD; ADNNCYLATALLTLQ; ADPAMHAASGNLLLD; ADQLTPTWRVYS; ADQLTPTWRVYSTGS; ADQLTPTWRVYSTGSN; ADSIIIGGAKLKALN; ADSKIVQLSEISMDN; ADSTLIGDCATVHTAN; ADTTDAVRDPQTLEIL; ADVEWKFYDAQPCSD; ADVEWKFYDAQPCSDK; ADVTKIKPHNSHEGKT; AEAELAKNVSLDNVLS; AEASKKPRQKRTATK; AEIPKEEVKPFITES; AEIPKEEVKPFITESK; AEIRASANLAATKMSE; AEIRASANLAATKMSECVLG; AELAKNVSLDNVLST; AELAKNVSLDNVLSTF; AELEGIQYGRSGETL; AENVTGLFKDCSKVI; AESHVDTDLTKPYIKW; AEVAVKMFDAYVNTFS; AEVQIDRLITGR; AEVQIDRLITGRLQS; AFATAQEAYEQAVAN; AFATAQEAYEQAVANG; AFAVDAAKAYKDYLAS; AFEYYHTTDPSFLGRY; AFFGMSRIGMEVTPS; AFFGMSRIGMEVTPSG; AFGRRGPEQTQGNFGD; AFGRRGPEQTQGNFGDQELI; AFKQIVESCGNFKVT; AFLPFAMGIIAMSAF; AFPSGKVEGCMVQVTC; AFQTVKPGNFNKDFY; AFVETVKGLDYKAFK; AFVETVKGLDYKAFKQ; AFVTNVNASSSEAFL; AFVTNVNASSSEAFLI; AFYILPSIISNEKQE; AFYILPSIISNEKQEI; AGALNKATNNAMQVE; AGALNKATNNAMQVES; AGDSGFAAYSRYRIGN; AGFPFNKWGKARLYY; AGGHSYGADLKSFDLG; AGHHLGRCDIKDLPKE; AGHHLGRCDIKDLPKEITVATSRTLS; AGIVGVLTLDNQDLNG; AGLPYGANKDGIIWVA; AGNATEVPANSTVLS; AGSDKGVAPGTAVLRQ; AGSKSPIQYIDI; AGSKSPIQYIDIGNYT; AGSYKDWSYSGQSTQ; AGSYKDWSYSGQSTQL; AGTDLEGNFYGPFVD; AGTDLEGNFYGPFVDR; AHFPREGVFVSNGTH; AHIQWMVMFTPLVPF; AICHDGKAHFPR; AICHDGKAHFPREGVF; AIDYKHYTPSFKKGA; AIDYKHYTPSFKKGAK; AIGKIQDSLSSTASAL; AIGLALYYPSARIVY; AIHADQLTPTWR; AIHADQLTPTWRVYST; AIKCVPQADVEWKFY; AIKCVPQADVEWKFYD; AIKITEHSWNADLYKL; AIKLDDKDPNFKDQVI; AIRHVRAWIGFDVEG; AISDYDYYRYNLPTMC; AISMWALIISVTSNY; AITVTPEANMDQESFG; AIVLQLPQGTTLPKG; AIVLQLPQGTTLPKGF; AIVLQLPQGTTLPKGFYAEGS; AKAYKDYLASGGQPIT; AKCWTETDLTKGPHEF; AKHDFFKFRIDGDMV; AKHDFFKFRIDGDMVP; AKHYVYIGDPAQLPA; AKKFDTFNGECPNF; AKKFDTFNGECPNFVF; AKKGAWNIGEQKSILS; AKKNNLPFKLTCATTR; AKLMVVIPDYNTYKNT; AKPFLNKVVSTTTNIV; AKVGILCIMSDRDLY; ALDQAISMWALIISV; ALGGSVAIKITEHSW; ALLADKFPVLHDIGNP; ALLSTDGNKIADKYV; ALLSTDGNKIADKYVR; ALLTKSSEYKGPITDV; ALLTLQQIELKFNPPA; ALNDFSNSGSDVLYQP; ALNKATNNAMQVESD; ALNKATNNAMQVESDD; ALNTPKDHIGTRNPAN; ALPETTADIVVFDEI; ALQDAYYRARAGEAA; ALQDAYYRARAGEAAN; ALQIPFAMQMAYRFN; ALQIPFAMQMAYRFNG; ALRHFDEGNCDTLKE; ALRKVPTDNYITTYPG; ALTAESHVDTDLTKP; ALTAESHVDTDLTKPY; ALTNNVAFQTVKPGNF; ALWEIQQVVDADSKI; ALYNKYKYFSGAMDTT; AMQMAYRFNGIG; AMQMAYRFNGIGVTQN; AMTQMYKQARSEDKR; AMTQMYKQARSEDKRA; ANDPVGFTLKNTVCTV; ANKDGIIWVATE; ANKDGIIWVATEGALN; ANKWDLIISDMYDPKT; ANNAAIVLQLPQGTTL; ANNAAIVLQLPQGTTLPKG; ANNAAIVLQLPQGTTLPKGF; ANNAAIVLQLPQGTTLPKGFY; ANQFNSAIGKIQDSLS; ANQVIVNNLDKSAGFP; ANRNRFLYIIKLIFL; ANSVFNICQAVTANV; ANSVFNICQAVTANVN; ANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNR; ANVNALLSTDGNKIAD; ANYAKPFLNKVVSTTT; ANYQKVGMQKYSTLQG; APAICHDGKAHF; APAICHDGKAHFPRE; APAICHDGKAHFPREG; APATVCGPKKSTN; APATVCGPKKSTNLV; APATVCGPKKSTNLVK; APATVCGPKKSTNLVKNKCVN; APHGVVFLHVTYVPA; APHGVVFLHVTYVPAQEKNF; APRTLLTKGTLEPEYF; APSASAFFGMSRIGM; APTHLSVDTKFKTEG; APTLVPQEHYVRITGL; AQAAGTDTTITVNVL; AQCVQLHNDILLAKDT; AQDGNAAISDYDYYREY; AQEAYEQAVANGDSEV; AQGMDNLACEDL; AQLPAPRTLLTKGTLE; AQPCSDKAYKIEELF; AQPCSDKAYKIEELFY; AQYELKHGTFTCASE; AQYTSALLAGTITSG; ARAGKASCTLSEQLD; ARARVECFDKFKVNS; ARARVECFDKFKVNST; ARDLSLQFKRPINPT; ARDLSLQFKRPINPTD; ARFPKSDGTGTIYTE; ARHINAQVAKSHNIA; ARIVYTACSHAAVDAL; ARKHTTCCSLSHRFY; ARLRAKHYVYIGDPAQ; ARLYYDSMSYEDQDA; ARQHLKDGTCGLVEV; ARSKQRRPQGLPNNTA; ARSKQRRPQGLPNNTASWF; ARTAPHGHVMVELVA; ARTVAGVSICSTMTN; ARVECFDKFKVNSTLE; ASAFFGMSRIGMEVTP; ASAFFGMSRIGMEVTPSGTW; ASALWEIQQVVDADSK; ASCCLYCRCHIDHPN; ASFDNFKFVCDNIKFA; ASFSASTSAFVETVK; ASFYYVWKSYVHVVD; ASKILGLPTQTVDSSQ; ASLPFGWLIVGVALL; ASQRVAGDSGFAAYS; ASQRVAGDSGFAAYSR; ASTSAFVETVKGLDY; ASTSAFVETVKGLDYK; ASVYYSQLMCQPILLL; ASWFTALTQHGKED; ATAEAELAKNVSLDN; ATAEAELAKNVSLDNV; ATAQEAYEQAVANG; ATAQEAYEQAVANGD; ATAQEAYEQAVANGDS; ATATIPIQASLPFGW; ATEGALNTPKDHIGTR; ATHSDKFTDGVCLFW; ATIPIQASLPFGWLI; ATKAYNVTQAFGRRG; ATKAYNVTQAFGRRGP; ATKYLVQQESPFVMM; ATKYLVQQESPFVMMS; ATLVSDIDITFLKKD; ATLVSDIDITFLKKDA; ATNKATYKPNTWCIR; ATNKATYKPNTWCIRC; ATNVVIKVCEFQ; ATRGATVVIGTSKFY; ATRRIRGGDGKMKDLS; ATRVECTTIVNGVRR; ATVCGPKKSTNLVKN; ATVCGPKKSTNLVKNK; ATVCGPKKSTNLVKNKCVNF; ATVCGPKKSTNLVKNKCVNFNFNG; ATVHTANKWDLIIS; ATVHTANKWDLIISD; ATVHTANKWDLIISDM; ATVRLQAGNATEVPA; ATVRLQAGNATEVPAN; ATVSVSSPDAVTAYNG; ATYKPNTWCIRCLWS; AVAKHDFFKFRIDG; AVAKHDFFKFRIDGDM; AVASKILGLPTQTVDS; AVDAAKAYKDYLASGG; AVDALCEKALKYLP; AVDINKLCEEMLDNR; AVDINKLCEEMLDNRA; AVFISPYNSQNAVASK; AVGNICYTPSKLIEYT; AVGTNLPLQLGFSTGV; AVILRGHLRIAGHHL; AVMYMGTLSYEQFKKG; AVNSVPWDTIANYAKP; AVPTGYVDTPNNTDFS; AVRDPQTLEILD; AVRDPQTLEILDITP; AVRDPQTLEILDITPC; AVRDPQTLEILDITPCSFGGVS; AVTANVNALLSTDGNK; AVTAYNGYLTSSSKT; AVTAYNGYLTSSSKTP; AVVCFNSTYASQGLVA; AVVLLILMTARTVYD; AWIGFDVEGCHATRE; AWWTAFVTNVNASSS; AWWTAFVTNVNASSSE; AYANSVFNICQAVTAN; AYCNKTVGELGDVRE; AYESLRPDTRYVLMD; AYESLRPDTRYVLMDG; AYFNMVYMPASWVMR; AYKDYLASGGQPITNC; AYLRKHFSMMILSDDA; AYNGYLTSSSKTPEEH; AYNVTQAFGRRGPEQT; AYRFNGIGVTQN; AYRFNGIGVTQNVLY; AYRKVLLRKNGNKGAG; AYSRYRIGNYKLNTD; AYSRYRIGNYKLNTDH; AYTNSFTRGVYYPDKV; AYTNSFTRGVYYPDKVFRSS; AYTRYVDNNFCGPDGY; AYVNTFSSTFNVPMEK; CAAGTTQTACTDDN; CANDPVGFTLKNTVCT; CATTRQVVNVVTTKIA; CCHLAKALNDFSNSGS; CCKFDEDDSEPVLKG; CCKFDEDDSEPVLKGV; CCLYCRCHIDHPNPK; CCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT; CCREHEHEIAWYTER; CCSCGSCCKFDEDDSEPVLKGV; CCSCGSCCKFDEDDSEPVLKGVKL; CCSCGSCCKFDEDDSEPVLKGVKLHYT; CDDDYFNKKDWYDFV; CDGTTFTYASALWEI; CDHCGETSWQTGDF; CDIKDLPKEITVATS; CDIKDLPKEITVATSR; CDIPIGAGICASYQTQ; CDIRQLLFVVEVVDK; CDLQNYGDSATLPKGI; CDNIKFADDLNQLTGY; CDRAMPNMLRIMASL; CDVVIGIVNNTVYDP; CEEMLDNRATLQAIAS; CEFQFCNDPFLG; CFDKFKVNSTLEQYVF; CFVKRVDWTIEYPII; CFVLAAVYRINWITG; CGFIQQKLALGGSVA; CGKGYHLMSFPQ; CGKQATKYLVQQESPF; CGNFKVTKGKAKKGA; CGPKKSTNLVKNK; CGPKKSTNLVKNKCV; CGPKKSTNLVKNKCVN; CGQQQTTLKGVEAVMY; CGSVGFNIDYDCVSF; CHDGKAHFPREG; CHDGKAHFPREGVFVSNGTHWFVTQRNFYE; CHIDHPNPKGFCDLKG; CHLAKALNDFSNSGSD; CIDCSARHINAQVAKS; CIDGALLTKSSEYKGP; CIKVATVQSKMSDVK; CIKVATVQSKMSDVKC; CIMSDRDLYDKLQFTS; CINANQVIVNNLDKSA; CIRCLWSTKPVETSNS; CKDGHVETFYPKLQSS; CKFDEDDSEPVLKGVK; CKTCGQQQTTLKGVE; CKTCGQQQTTLKGVEA; CLGDIAARDLICAQKF; CLKLSHQSDIEVTGD; CLLQFAYANRNRFLY; CLNRVCTNYMPYFFT; CLPFTINCQEPKLGSL; CLSTLMKCDHCGETS; CLWSTKPVETSNSFD; CLWSTKPVETSNSFDV; CLYRNRDVDTFDVNE; CNQMCLSTLMKCDHC; CNYLGKPREQIDGYVM; CPACHNSEVGPEHSL; CPAVAKHDFFKFRIDG; CQAVTANVNALLSTDG; CQEPKLGSLVVRCSF; CQPILLLDQALVSDV; CRCHIDHPNPKGFCDL; CRFQEKDEDDNLIDSY; CRKVQHMVVKAALLA; CRMNSRNYIAQV; CRMNSRNYIAQVDVVN; CRRCPAEIVDTVSALV; CSCGSCCKFDEDDSEPVLKGVKLH; CSFGGVSVITPGTNTS; CSFYEDFLEYHDVRV; CSLSHRFYRLANECA; CSQHTMLVKQGDDYVY; CSRIIPARARVECFD; CSSGTYEGNSPFHPLA; CSVIDLLLDDFVEII; CTCGKQATKYLVQQES; CTCYFGLFCLLNRYF; CTEIDPKLDNYYKKDN; CTEVPVAIHADQ; CTEVPVAIHADQLTP; CTEVPVAIHADQLTPTWRVYST; CTEVPVAIHADQLTPTWRVYSTGSN; CVCSVIDLLLDDFVE; CVDEAGSKSPIQYID; CVDEAGSKSPIQYIDI; CVLAAECTIFKDASGK; CVPQADVEWKFYDAQP; CVRGTTVLLKEPCSSG; CWTETDLTKGPHEFCS; CYDGGCINANQVIVN; CYDGGCINANQVIVNN; CYFGLFCLLNRYFRL; CYLATALLTLQQIEL; CYTPSKLIEYTDFATS; DAAKAYKDYLASGGQP; DAAMQRKLEKMADQAM; DADSKIVQLSEISMDN; DADSTLIGDCATVHTA; DALFAYTKRNVIPTI; DALFAYTKRNVIPTIT; DAMRNAGIVGVLTLD; DAPYIVGDVVQEGVL; DARTAPHGHVMVELVA; DAVRDPQTLEIL; DAVRDPQTLEILDIT; DAVRDPQTLEILDITP; DAVRDPQTLEILDITPCSFG; DAVTAYNGYLTSSSK; DAVTAYNGYLTSSSKT; DAYYRARAGEAANFCA; DCATVHTANKWDLII; DCLGDIAARDLICAQ; DCPAVAKHDFFKFRID; DCYDGGCINANQVIVN; DDDYFNKKDWYDFVEN; DDFVEIIKSQDLSVV; DDKDPNFKDQVILLN; DDKDPNFKDQVILLNK; DDLNQLTGYKKPASR; DDLNQLTGYKKPASRE; DDQIGYYRRATRRIRGGDGK; DDSEPVLKGVKLH; DDSEPVLKGVKLHYT; DDTLRVEAFEYYHTT; DDVVYCPRHVICTSE; DDYFNKKDWYDFVEN; DDYFNKKDWYDFVENP; DDYVYLPYPDPSRIL; DEAGSKSPIQYI; DEAGSKSPIQYIDIGN; DEDDNLIDSYFVVKR; DEDDNLIDSYFVVKRH; DEDDSEPVLKGVKLH; DEDDSEPVLKGVKLHY; DEFSSNVANYQKVGMQ; DEISMATNYDLSVVN; DEISMATNYDLSVVNA; DELGTDPYEDFQEN; DEMIAQYTSALLAGT; DFCGKGYHLMSF; DFCGKGYHLMSFPQS; DFCGKGYHLMSFPQSA; DFFKFRIDGDMVPHIS; DFIDTKRGVYCCREH; DFIDTKRGVYCCREHE; DFKLNEEIAIILASF; DFQENWNTKHSS; DFQENWNTKHSSGVTR; DFSNSGSDVLYQPPQ; DFSNSGSDVLYQPPQT; DFSRVSAKPPPGDQFK; DFSRVSAKPPPGDQFKH; DFVSDADSTLIGDCA; DFVSDADSTLIGDCAT; DFYDFAVSKGFFKEGS; DGADVTKIKPHNSHE; DGALLTKSSEYKGPIT; DGCNSSTCMMCYKRN; DGEVITFDNLKTLLSL; DGGCINANQVIVNNLD; DGHVETFYPKLQSSQA; DGIIWVVATEGALNTP; DGIIWVVATEGALNTPK; DGKAHFPREGVFVSN; DGKMKDLSPRWYFYY; DGKMKDLSPRWYFYYL; DGNAAISDYDYYRYNL; DGTCGLVEVEKGVLP; DGTGTIYTELEPPCR; DGTTSPISEHDYQIG; DGTTSPISEHDYQIGG; DGVPFVVSTGYHFRE; DGYPLECIKDLLARA; DGYVMHANYIFWRNT; DHCGETSWQTGDFVKA; DHIGTRNPANNAAIVL; DHSSSSDNIALLVQ; DIAARDLICAQK; DIADTTDAVRDPQTL; DIADTTDAVRDPQTLE; DIADTTDAVRDPQTLEILDITPCSFG; DIDITFLKKDAPYIVG; DIGNYTVSCLPF; DIKDLPKEITVATSR; DIKDLPKEITVATSRT; DINKLCEEMLDNRAT; DINKLCEEMLDNRATL; DIPIGAGICASYQTQT; DITFLKKDAPYIVGDV; DITPCSFGGVSVITP; DKCSRIIPARARVEC; DKDPNFKDQVILLNK; DKDPNFKDQVILLNKH; DKDPNFKDQVILLNKHIDAY; DKDPNFKDQVILLNKHIDAYKTFPPTEPKKDKKKKADET; DKFKVNSTLEQYVFCT; DKRAKVTSAMQTMLF; DKRTTCFSVAALTNN; DKSVYYTSNPTTFHL; DKVEAEVQIDRLITG; DKVFRSSVLHSTQDL; DKYFDCYDGGCINANQ; DKYFKNHTSPDVDL; DKYFKNHTSPDVDLG; DKYFKNHTSPDVDLGD; DKYVRNLQHRLYECL; DLEGNFYGPFVDRQTA; DLGDELGTDPYEDFQ; DLGDELGTDPYEDFQE; DLIISDMYDPKTKNVT; DLKFPRGQGVPINT; DLKFPRGQGVPINTNS; DLKGKYVQIPTTCAND; DLKSFDLGDELGTDPY; DLLKYDFTEERLKLFD; DLLLDDFVEIIKSQD; DLLLDDFVEIIKSQDL; DLNDFVSDADSTLIGD; DLNGDVVAIDYKHYTP; DLNQLTGYKKPASRE; DLNQLTGYKKPASREL; DLPKEITVATSRTLS; DLPKEITVATSRTLSY; DLQDLKWARFPKSDGT; DLQNYGDSATLPKGIM; DLSPRWYFYYLGTGP; DLSVVNARLRAKHYVY; DLVPNQPYPNASFDNF; DLVYALRHFDEGNCD; DLYDKLQFTSLEIPR; DLYDKLQFTSLEIPRR; DMLNPNYEDLLIRKSN; DMYDPKTKNVTKENDS; DNFKFVCDNIKFADD; DNIKFADDLNQLTGYK; DNLLEILQKEKVNIN; DNLLEILQKEKVNINI; DNLRANNTKGSLPINV; DNQDLNGNWYDFGDF; DNSYFTEQPIDLVPNQ; DNTFVSGNCDVV; DNTSRYWEPEFYEAM; DNTSRYWEPEFYEAMY; DNVYIKNADIVEEAK; DNVYIKNADIVEEAKK; DNYITTYPGQGLNGYT; DNYYKKDNSYFTEQPI; DPAQLPAPRTLLTK; DPAQLPAPRTLLTKGT; DPKLDNYYKKDNSYFT; DPLQPELDSFKEELD; DPLQPELDSFKEELDK; DPLQPELDSFKEELDKYFKNHTSPDVDLGDIS; DPNFKDQVILLNKHI; DPNFKDQVILLNKHID; DPSFLGRYMSALNHTK; DPYEDFQENWNT; DPYEDFQENWNTKHS; DPYEDFQENWNTKHSS; DQAISMWALIISVTS; DQAMTQMYKQARSEDK; DQDALFAYTKRNVIPT; DQELIRQGTDYKHWP; DQFKHLIPLMYKGLP; DQIGYYRRATRRIRG; DQLTPTWRVYST; DQLTPTWRVYSTGSNV; DQTYHPNCVNCLDDRC; DRDLYDKLQFTSLEI; DRLITGRLQSLQTYV; DRLITGRLQSLQTYVTQQLI; DRQTAQAAGTDTTIT; DRWFLNRFTTTLNDFN; DRYFKYWDQTYHPNCV; DSAEVAVKMFDAYVN; DSATLPKGIMMNVAKY; DSATLVSDIDITFLK; DSATLVSDIDITFLKK; DSCNNYMLTYNKVEN; DSCNNYMLTYNKVENM; DSDLNDFVSDADSTLI; DSEPVLKGVKLHYT; DSEYCRHGTCERSEA; DSFKEELDKYFKNHT; DSFKEELDKYFKNHTS; DSFKEELDKYFKNHTSPDVD; DSGFAAYSRYRIGNY; DSGFAAYSRYRIGNYK; DSKTQSLLIVNNATN; DSKTQSLLIVNNATNV; DSMSYEDQDALFAYTK; DSSQGSEYDYVIFTQ; DSTLIGDCATVHTAN; DSTLIGDCATVHTANK; DSVEEVLSEARQHLKD; DSVTVKNGSIHLYFD; DTDLTKPYIKWDLLK; DTDLTKPYIKWDLLKY; DTFCAGSTFISDEVA; DTFNGECPNFVFPLN; DTFNGECPNFVFPLNS; DTIANYAKPFLNKVVS; DTKRGVYCCREHEHEI; DTPNNTDFSRVSAKPP; DTRVLSNLNLPGCDGG; DTTDAVRDPQTL; DTTDAVRDPQTLEIL; DTTDAVRDPQTLEILD; DTTDAVRDPQTLEILDITPCSF; DVENPHLMGWDYPKC; DVENPHLMGWDYPKCD; DVEWKFYDAQPCSDK; DVEWKFYDAQPCSDKA; DVFYKENSYTTTIKPV; DVGDSAEVAVKMFDAY; DVLECNVKTTEVVGD; DVLKSEDAQGMDNLA; DVLKSEDAQGMDNLAC; DVLVRGFGDSVEEVLS; DVNCTEVPVAIHADQL; DVNCTEVPVAIHADQLTPTW; DVNCTEVPVAIHADQLTPTWRVYSTGS; DVSSAINRPQIGVVRE; DVTKIKPHNSHEGKTF; DVVAIDYKHYTPSFKK; DVVECLKLSHQSDIE; DVVIGIVNNTVYDPL; DWTIEYPIIGDELKI; DWTIEYPIIGDELKIN; DYDYYRYNLPTMCDI; DYDYYRYNLPTMCDIR; DYFNKKDWYDFVENPD; DYFVLTSHTVMPLSAP; DYIINLIIKNLSKSL; DYKHWPQIAQFAPSAS; DYNTYKNTCDGT; DYNTYKNTCDGTTFTY;

DYQIGGYTEKWESGVK; DYVIFTQTTETAHSCN; EAACCHLAKALNDFS; EAACCHLAKALNDFSN; EAELAKNVSLDNVLST; EAEVQIDRLITGRLQ; EAEVQIDRLITGRLQS; EAFEYYHTTDPSFLG; EAGLPYGANKDG; EAGLPYGANKDGIIW; EAGLPYGANKDGIIWV; EAGSKSPIQYID; EAGSKSPIQYIDIGN; EAGSKSPIQYIDIGNY; EAKCWTETDLTKGPH; EAKCWTETDLTKGPHE; EAKKVKPTVVVNAAN; EAKKVKPTVVVNAANV; EASFNYLKSPNFSKL; EASFNYLKSPNFSKLI; EASKKPRQKRTATKA; EAVGTNLPLQLGFST; EAVMYMGTLSYEQFK; EAVMYMGTLSYEQFKK; ECDIPIGAGIC; ECDIPIGAGICA; ECDIPIGAGICASYQ; ECDIPIGAGICASYQT; ECFDKFKVNSTLEQYV; ECPNFVFPLNSI; ECTTIVNGVRRSFYV; ECVRGTTVLLKEPCS; EDAQGMDNLACEDLK; EDDSEPVLKGVKLHY; EDDSEPVLKGVKLHYT; EDFLEYHDVRVVLDF; EDFQENWNTKHS; EDFQENWNTKHSSGVT; EDLLIRKSNHNFLVQ; EDMLNPNYEDLLIRK; EDMLNPNYEDLLIRKS; EDQDALFAYTKRNVI; EDQDALFAYTKRNVIP; EEAALCTFLLNKEMY; EEAIRHVRAWIGFDV; EEAKKVKPTVVVNAAN; EEAKTVLKKCKSAFY; EEHFIETISLAGSYK; EELDKYFKNHTSPDV; EELDKYFKNHTSPDVD; EELFYSYATHSDKFTD; EEMLDNRATLQAIAS; EEVGHTDLMAAYVDN; EEVKPFITESKPSVEQ; EEVLSEARQHLKDGT; EEVLSEARQHLKDGTC; EEVVENPTIQKDVLEC; EFCGTENLTKEGATT; EFCSQHTMLVKQGDDY; EFDRDAAMQRKLEKM; EFLKRGDKSVYYTSN; EFLKRGDKSVYYTSNP; EFSSLPSYAAFATAQE; EFSSNVANYQKVGMQK; EFYAYLRKHFSMMIL; EGALNTPKDHIGTRN; EGALNTPKDHIGTRNP; EGATTCGYLPQNAVVK; EGCHATREAVGTNLPL; EGKTFYVLPNDDTLR; EGKTFYVLPNDDTLRV; EGLNDNLLEILQKEK; EGLNDNLLEILQKEKV; EGNFYGPFVDRQTAQA; EGSVAYESLRPDTRY; EGSVAYESLRPDTRYV; EGSVRVVTTFDSEYC; EGVFVSNGTHWFVTQRNFYE; EHDYQIGGYTEKWESG; EHDYQIGGYTEKWESGV; EHDYQIGGYTEKWESGVKDCVVL; EHEHEIAWYTER; EHSWNADLYKLMGHFA; EHYVRITGLYPTLNI; EIDPKLDNYYKKDNSY; EIIKSQDLSVVSKVV; EIKLAKKFDTFNGEC; EIKLAKKFDTFNGECP; EILDITPCSFGG; EILDITPCSFGGVSV; EILQKEKVNINIVGD; EILQKEKVNINIVGDF; EIPKEEVKPFITESKP; EIPRRNVATLQAENVT; EIRASANLAATKMSE; EIRASANLAATKMSEC; EISFMLWCKDGHVET; EISMATNYDLSVVNAR; EITVATSRTLSYYKL; EKCDLQNYGDSATLPK; EKDEDDNLIDSYFVV; EKDEDDNLIDSYFVVK; EKGDYGDAVVYRGTTT; EKGVLPQLEQPYVFIK; EKLKTLVATAEAELAK; EKMADQAMTQMYKQA; EKNFTTAPAICHDGK; EKQVEQKIAEIPKEE; EKQVEQKIAEIPKEEV; EKSNIIRGWIFGTTL; EKSYELQTPFEIKLA; EKSYELQTPFEIKLAK; EKVNINIVGDFKLNEE; EKWESGVKDCVVLHS; ELAKNVSLDNVLSTFI; ELCVDEAGSKSPIQYI; ELDKYFKNHTSP; ELDKYFKNHTSPDVD; ELDKYFKNHTSPDVDL; ELDSFKEELDKYFKN; ELDSFKEELDKYFKNH; ELDSFKEELDKYFKNHTS; ELDSFKEELDKYFKNHTSPD; ELDSFKEELDKYFKNHTSPDV; ELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDR; ELEPPCRFVTDTPKG; ELEPPCRFVTDTPKGP; ELFYSYATHSDKFTD; ELFYSYATHSDKFTDG; ELGDVRETMSYLFQH; ELGDVRETMSYLFQHA; ELGTDPYEDFQENWN; ELGTDPYEDFQENWNT; ELGVVHNQDVNLHSS; ELIRQGTDYKHWPQI; ELIRQGTDYKHWPQIA; ELIRQGTDYKHWPQIAQFAP; ELKFNPPALQDAYYRA; ELKHFFFAQDGNAAI; ELKVTFFPDLNGDVV; ELKVTFFPDLNGDVVA; ELLHAPATVCGPKKSTNLVK; ELLHAPATVCGPKKSTNLVKN; ELLVYAADPAMHAAS; ELPTGVHAGTDLEGNF; ELQTPFEIKLAKKFD; ELQTPFEIKLAKKFDT; ELSRVLGLKTLATHG; ELTGHMLDMYSVMLT; ELTPVVQTIEVNSFSG; ELTSMKYFVKIGPER; ELTSMKYFVKIGPERT; ELVIGAVILRGHLRI; ELYHYQECVRGTTVL; EMIAQYTSALLAGTI; EMKSEKQVEQKIAEI; EMKSEKQVEQKIAEIP; EMLDNRATLQAIASEF; ENAFLPFAMGIIAMS; ENKYSQLDEEQPMEI; ENKYSQLDEEQPMEID; ENLTKEGATTCGYLPQ; ENPDILRVYANLGERV; ENPHLMGWDYPKCDR; ENPHLMGWDYPKCDRA; ENPTIQKDVLECNVKT; ENQKLIANQFNSAIGK; ENSYTTTIKPVTYKL; ENSYTTTIKPVTYKLD; EPEYFNSVCRLMKTI; EPKLGSLVVRCSFYE; EPPCRFVTDTPKGPKV; EPQIITTDNTFVSGN; EPQIITTDNTFVSGNC; EPVLKGVKLHYT; EQIDGYVMHANYIFWR; EQKIAEIPKEEVKPF; EQKIAEIPKEEVKPFI; EQKSILSPLYAFASEA; EQPIDLVPNQPYPNAS; EQRKQDDKKIKACVEE; EQTQGNFGDQELIRQ; EQTQGNFGDQELIRQG; ERHSLSHFVNLDNLR; ERLKLFDRYFKYWDQ; ERLKLFDRYFKYWDQT; ERSEAGVCVSTSGRW; ERSEKSYELQTPFEIK; ERVRQALLKTVQFCD; ERVRQALLKTVQFCDA; ESCGNFKVTKGKAKK; ESELVIGAVILRGHL; ESGVKDCVVLHSYFTS; ESHVDTDLTKPYIKW; ESHVDTDLTKPYIKWD; ESKPSVEQRKQDDKKI; ESLRPDTRYVLMDGS; ESLRPDTRYVLMDGSI; ESNKKFLPFQQFGRDI; ESNKKFLPFQQFGRDIA; ESNKKFLPFQQFGRDIADTT; ESPFVMMSAPPAQYEL; ESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHI; ESSSKLWAQCVQLHN; ESSSKLWAQCVQLHND; ETAQNSVRVLQKAAIT; ETDLTKGPHEFCSQH; ETFYPKLQSSQAWQPG; ETKDVVECLKLSHQSD; ETLGVLVPHVGEIPV; ETMSYLFQHANLDSC; ETNILLNVPLHGTIL; ETSNSFDVLKSEDAQ; ETSNSFDVLKSEDAQG; ETSWQTGDFVKATCE; ETTADIVVFDEISMAT; ETVKGLDYKAFKQIVE; EVARDLSLQFKRPIN; EVEKGVLPQLEQPYVF; EVITFDNLKTLLSLRE; EVKPFITESKPSVEQ; EVKPFITESKPSVEQR; EVNSFSGYLKLTDNV; EVNSFSGYLKLTDNVY; EVPVAIHADQLT; EVPVAIHADQLTPTW; EVPVAIHADQLTPTWR; EVPVAIHADQLTPTWRVYSTGSNVFQTRA; EVQIDRLITGRL; EVTGDSCNNYMLTYN; EVTGDSCNNYMLTYNK; EVTPSGTWLTYT; EVVDKYFDCYDGGCI; EVVENPTIQKDVLEC; EVVENPTIQKDVLECN; EWKFYDAQPCSDKAY; EWKFYDAQPCSDKAYK; EYCPIFFITGNTLQC; EYCRHGTCERSEAGV; EYDYVIFTQTTETAH; EYDYVIFTQTTETAHS; EYKGPITDVFYKENS; EYKGPITDVFYKENSY; EYPIIGDELKINAACR; EYTDFATSACVLAAE; EYTFEKGDYGDAVVYR; EYTGNYQCGHYKHITS; EYYHTTDPSFLGRYM; EYYHTTDPSFLGRYMS; FAAYSRYRIGNYKLN; FAAYSRYRIGNYKLNT; FADDLNQLTGYKKPAS; FAMQMAYRFNGI; FAMQMAYRFNGIGVT; FAMQMAYRFNGIGVTQ; FAPSASAFFGMSRIG; FAQDGNAAISDYDYY; FAQDGNAAISDYDYYR; FARTRSMWSFNPETN; FATAQEAYEQAVANGD; FATSACVLAAECTIF; FAVDAAKAYKDYLASG; FAVSKGFFKEGSSVE; FAYANRNRFLYIIKL; FAYTKRNVIPTITQM; FCAFAVDAAKAYKDY; FCGKGYHLMSFP; FCGKGYHLMSFPQSA; FCGKGYHLMSFPQSAP; FCGKGYHLMSFPQSAPHGVVFLHVT; FCGPDGYPLECIKDL; FCKLHNWNCVNCDTF; FCSQHTMLVKQGDDY; FCSQHTMLVKQGDDYV; FDCYDGGCINANQVIV; FDEDDSEPVLKG; FDEDDSEPVLKGVKLH; FDEDDSEPVLKGVKLHY; FDEISMATNYDLSVVN; FDKFKVNSTLEQYVFC; FDLGDELGTDPYEDFQ; FDNPVLPFNDGV; FDNPVLPFNDGVYFA; FDRDAAMQRKLEKM; FDSEYCRHGTCERSE; FDSEYCRHGTCERSEA; FDTFNGECPNFVFPLN; FDTRVLSNLNLPGCD; FDVLKSEDAQGMDNLA; FEIKLAKKFDTFNGEC; FEKGDYGDAVVYRGTT; FELLHAPATVCGPKKSTNLVK; FEYYHTTDPSFLGRY; FEYYHTTDPSFLGRYM; FFFAQDGNAAISDYD; FFGMSRIGMEVTPSGT; FFKEGSSVELKHFFF; FFKFRIDGDMVPHIS; FFPDLNGDVVAIDYKH; FGAGAALQIPFAMQM; FGAGAALQIPFAMQMAYRFNGI; FGDQELIRQGTDYKH; FGDQELIRQGTDYKHW; FGDSVEEVLSEARQHL; FGGPSDSTGSNQNGER; FGGVSVITPGTNTSN; FGGVSVITPGTNTSNQ; FGMSRIGMEVTPSGTW; FGMSRIGMEVTPSGTWLTYTGAIKL; FGPLVRKIFVDGVPF; FGRDIADTTDAVR; FGRDIADTTDAVRDP; FGRDIADTTDAVRDPQ; FGRDIADTTDAVRDPQTLEI; FGRDIADTTDAVRDPQTLEILDITPCSFGGVSVI; FGRRGPEQTQGNFGD; FGRRGPEQTQGNFGDQ; FGTTLDSKTQSLLIV; FHLDGEVITFDNLKT; FHLDGEVITFDNLKTL; FHLVDFQVTIAEILL; FHQKLLKSIAATRGAT; FIDTKRGVYCCREHEH; FIKGLNNNLNRGMVLG; FIKGLNNNLNRGMVLGS; FIKRSDARTAPHGHV; FIQQKLALGGSVAIK; FISDEVARDLSLQFK; FISPYNSQNAVASKIL; FITESKPSVEQRKQD; FITESKPSVEQRKQDD; FITREEAIRHVRAWI; FKDQVILLNKHIDAY; FKDQVILLNKHIDAYK; FKEELDKYFKNH; FKEELDKYFKNHTSPD; FKELLVYAADPAMHAA; FKFRIDGDMVPHISR; FKFVCDNIKFADDLN; FKFVCDNIKFADDLNQ; FKKGAKLLHKPIVWH; FKKGVQIPCTCGKQA; FKLTCATTRQVVNVV; FKMNYQVNGYPNMFIT; FKNHTSPDVDLGDIS; FKNHTSPDVDLGDISG; FKNHTSPDVDLGDISGINAS; FKNHTSPDVDLGDISGINASVVNIQKEI; FKSVLYYQNNVFMSE; FKYWDQTYHPNCVNCL; FLAHIQWMVMFTPLV; FLEMKSEKQVEQKIA; FLEMKSEKQVEQKIAE; FLFDMSKFPLKLRGT; FLIGCNYLGKPREQID; FLKKDAPYIVGDVVQ; FLKRGDKSVYYTSNPT; FLLNKEMYLKLRSDVL; FLNGSCGSVGFNIDYD; FLNKVVSTTTNIVTRC; FLNRFTTTLNDFNLV; FLNRFTTTLNDFNLVA; FLPFQQFGRDIA; FLPFQQFGRDIADTT; FLPFQQFGRDIADTTD; FLPGVYSVIYLYLTF; FLTWICLLQFAYANR; FLYIIKLIFLWLLWP; FMGRIRSVYPVASPN; FMGRIRSVYPVASPNE; FMSEAKCWTETDLTKG; FMSLSEQLRKQIRSAA; FNGECPNFVFPLNSI; FNGIGVTQNVLYENQK; FNKKDWYDFVENPDIL; FNMVYMPASWVVMRIM; FNPETNILLNVPLHG; FNPETNILLNVPLHGT; FNSAIGKIQDSLSST; FNSAIGKIQDSLSSTA; FNSVCRLMKTIGPDM; FPDLNGDVVAIDYK; FPDLNGDVVAIDYKH; FPDLNGDVVAIDYKHY; FPKSDGTGTIYTELE; FPLKLRGTAVMSLKE; FPLNSIIKTIQPRVEK; FPQSAPHGVVFL; FPREGVFVSNGTHWF; FPRGQGVPINTNSSPD; FPRGQGVPINTNSSPDS; FQEKDEDDNLIDSYF; FQEKDEDDNLIDSYFV; FQENWNTKHSSGVTRE; FQQFGRDIADTTDAV; FQQFGRDIADTTDAVR; FQTRAGCLIGAE; FQTVKPGNFNKDFYDF; FRELGVVHNQDVNLH; FRIDGDMVPHISRQR; FRLFARTRSMWSFNP; FRSSVLHSTQDLFLP; FRSSVLHSTQDLFLPFFSNVTW; FSASTSAFVETVKGLD; FSAVGNICYTPSKLI; FSGYLKLTDNVYIKNA; FSNYQHEETIYNLLKD; FSRTLETAQNSVRVL; FSRVSAKPPPGDQFKH; FSSLPSYAAFATAQE; FSSLPSYAAFATAQEA; FSSTFNVPMEKLKT; FSTFEEAALCTFLLN; FSTGVNLVAVPTGYV; FTCASEYTGNYQCGH; FTCASEYTGNYQCGHY; FTDGVCLFWNCNVDR; FTEERLKLFDRYFK; FTEQPIDLVPNQPYPN; FTGYRVTKNSKVQIGE; FTIGTVTLKQGEIKDA; FTINCQEPKLGSLVV; FTSLEIPRRNVATLQA; FTTAPAICHDGKAHF; FTTTLNDFNLVAMKY; FTTTLNDFNLVAMKYN; FVCDNIKFADDLNQLT; FVDGVPFVVSTGYHF; FVDRQTAQAAGTDTTI; FVENPDILRVYANLGE; FVKATCEFCGTENLT; FVMMSAPPAQYELKH; FVNEFYAYLRKHFSMM; FVNLDNLRANNTKGSL; FVSDADSTLIGDCATV; FVSGNCDVVIGIVNN; FVSGNCDVVIGIVNNTVYDPLQ; FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNH; FVSNGTHWFVTQ; FVSNGTHWFVTQRNF; FVTDTPKGPKVKYLYF; FVTHSKGLYRKCVKS; FVTQRNFYEPQI; FVTQRNFYEPQIITTD; FVVKRHTFSNYQHEE; FWITIAYIICISTKH; FYEDFLEYHDVRVVL; FYEPQIITTDNTFVSG; FYGPFVDRQTAQAAGT; FYKENSYTTTIKPVTY; FYLTNDVSFLAHIQW; FYPKLQSSQAWQPGVA; FYRLANECAQVLSEM; FYSYATHSDKFTDGV; FYSYATHSDKFTDGVC; GAALQIPFAMQMAYR; GACIDCSARHINAQVA; GADVTKIKPHNSHEGK; GAGAALQIPFAMQMA; GAGAALQIPFAMQMAY; GAGGHSYGADLKSFDL; GAGICASYQTQT; GAIKLDDKDPNFKDQV; GAKLLHKPIVWHVNN; GALLTKSSEYKGPITD; GALNKATNNAMQVESD; GAMDTTSYREAACCHL; GANKDGIIWVATEGA; GARSKQRRPQGLPN; GARSKQRRPQGLPNN; GARSKQRRPQGLPNNT; GASQRVAGDSGFAAYS; GAVDINKLCEEMLDNR; GAVILRGHLRIAGHH; GAYTRYVDNNFCGPD; GCHATREAVGTNLPLQ; GCINANQVIVNNLDKS; GCNYLGKPREQIDGY; GCNYLGKPREQIDGYV; GDCATVHTANKWDLI; GDCLGDIAARDLICA; GDELGTDPYEDFQEN; GDELGTDPYEDFQENW; GDFIQTTPGSGVPV; GDFVKATCEFCGTEN; GDGKMKDLSPRW; GDGKMKDLSPRWYFY; GDGKMKDLSPRWYFYY; GDGTTSPISEHDYQIG; GDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHS; GDIAARDLICAQ; GDPAQLPAPRTLLTKG; GDQELIRQGTDYKHWP; GDRWFLNRFTTTLND; GDSCNNYMLTYNKVEN; GDSGFAAYSRYRIGN; GDSGFAAYSRYRIGNY; GDSVEEVLSEARQHLK; GDVRETMSYLFQHANL; GDVVAIDYKHYTPSFK; GDVVQEGVLTAVVIPT; GDYFVLTSHTVMPLS; GDYGDAVVYRGTTTYK; GEGSEGLNDNLLEIL; GELGDVRETMSYLFQH; GEVITFDNLKTLLS; GEYSHVVAFNTLLFL; GFAAYSRYRIGNYKLN; GFCDLKGKYVQIPTTC; GFFKEGSSVELKHF; GFFTYICGFIQQKLA; GFGDSVEEVLSEARQ; GFGDSVEEVLSEARQH; GFKMNYQVNGYPNMF; GFKMNYQVNGYPNMFI; GFLFLTWICLLQFAY; GFVDSDVETKDVVEC; GGAKLKALNLGETFVT; GGASCCLYCRCHIDH; GGCINANQVIVNNLDK; GGDGKMKDLSPRWYF; GGDGKMKDLSPRWYFY; GGDGKMKDLSPRWYFYY; GGGVAGALNKATNNA; GGGVAGALNKATNNAM; GGHSYGADLKSFDLGD; GGKPCIKVATVQSKMS; GGPSDSTGSNQNGERS; GGSVAIKITEHSWNA; GGTSSGDATTAYANS; GGTSSGDATTAYANSV; GGVAGALNKATNNAMQ; GGVSVITPGTNTSNQ; GGVSVITPGTNTSNQV; GHHLGRCDIKDLPKE; GHHLGRCDIKDLPKEI; GHLRIAGHHLGRCDI; GHSMQNCVLKLKVDTA; GHSYGADLKSFDLGDE; GHTDLMAAYVDNSSLT; GHVETFYPKLQSSQAW; GHVMVELVAELEGIQ; GIEFLKRGDKSVYYTS; GIGVTQNVLYEN; GIGVTQNVLYENQKLI; GIIWVATEGALNTPKD; GILCIMSDRDLYDKLQ; GIQYGRSGETLGVLV; GIQYGRSGETLGVLVP; GIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISG; GKAHFPREGVFV; GKAKKGAWNIGEQKSI; GKARLYYDSMSYEDQ; GKGFCKLHNWNCVNC; GKGFCKLHNWNCVNCD; GKGYHLMSFPQSAPHG; GKIQDSLSSTASALGK; GKMKDLSPRWYFYYLGTGPEAGLPYGANK; GKPCIKVATVQSKMSD; GKPREQIDGYVMHANY; GKQATKYLVQQESPF; GKQATKYLVQQESPFV; GKSHFAIGLALYYPS; GKTFYVLPNDDTLRVE; GKYVQIPTTCANDPVG; GLAAVNSVPWDTIA; GLAAVNSVPWDTIAN; GLAAVNSVPWDTIANY; GLDYKAFKQIVESCGN; GLHPTQAPTHLSVDT; GLKTLATHGLAAVNSV; GLNDNLLEILQKEKVN; GLNNNLNRGMVLGSLAA; GLPNNTASWFTALTQHGKED; GLPTQTVDSSQGSEYD; GLPYGANKDGII; GLPYGANKDGIIWVA; GLPYGANKDGIIWVAT; GLTSIKWADNNCYLAT; GLTVLPPLLTDEMIA; GLVEVEKGVLPQLEQ; GLVEVEKGVLPQLEQP; GLYPTLNISDEFSSNV; GMEVTPSGTWLTYT; GMEVTPSGTWLTYTG; GMEVTPSGTWLTYTGAIKLD; GMSRIGMEVTPSGTW; GMSRIGMEVTPSGTWL; GNAAISDYDYYRYNL; GNAAISDYDYYRYNLP; GNFGDQELIRQGTDY; GNFGDQELIRQGTDYK; GNFNKDFYDFAVSKGF; GNFYGPFVDRQTAQAA; GNICYTPSKLIEYTDF; GNKGAGGHSYGADLKSFDL; GNLLLDKRTTCFSVAA; GNPKAIKCVPQADVE; GNVQLRVIGHSMQNC; GNYKLNTDHSSSSDN; GNYKLNTDHSSSSDNI; GPDGYPLECIKDLLA; GPEAGLPYGANKDGI; GPEAGLPYGANKDGII; GPEHSLAEYHNESGLK; GPEQTQGNFGDQELI; GPEQTQGNFGDQELIR; GPFVDRQTAQAAGTDT; GPITDVFYKENSYTTT; GPKKSTNLVKNKC; GPKKSTNLVKNKCVN; GPKKSTNLVKNKCVNF; GPSDSTGSNQNGERS; GPSDSTGSNQNGERSG; GQAITVTPEANMDQES; GQGLNGYTVEEAKTVL; GQGVPINTNSSP; GQGVPINTNSSPDDQI; GQQQTTLKGVEAVMY; GQQQTTLKGVEAVMYM; GQSTQLGIEFLKRGDK; GRCDIKDLPKEITVA; GRCDIKDLPKEITVAT; GRDIADTTDAVRDPQ; GRDIADTTDAVRDPQT; GRDIADTTDAVRDPQTLEILDI; GRDIADTTDAVRDPQTLEILDITPCSFGGVSV; GRIRSVYPVASPNEC; GRIRSVYPVASPNECN; GRRGPEQTQGNFGDQE; GRYMSALNHTKKWKY; GRYMSALNHTKKWKYPQV; GSCCKFDEDDSEPVLK; GSCGSVGFNIDYDCV; GSCGSVGFNIDYDCVS; GSDKGVAPGTAVLRQW; GSEYDYVIFTQTTET; GSFLNGSCGSVGFNID; GSKSPIQYIDIGNYT; GSKSPIQYIDIGNYTV; GSKSPIQYIDIGNYTVSC; GSKSPIQYIDIGNYTVSCLPF; GSLAATVRLQAGNAT; GSLAATVRLQAGNATE; GSLVVRCSFYEDFLE; GSLYVKPGGTSSGDAT; GSNVFQTRAGCL; GSSKCVCSVIDLLLD; GSVAYESLRPDTRYVL; GSVRVVTTFDSEYCRH; GSYKDWSYSGQSTQLG; GTCRRCPAEIVDTVSA; GTDLEGNFYGPFVDRQ; GTDPYEDFQENWNTK; GTDPYEDFQENWNTKH; GTDYKHWPQIAQFAP; GTDYKHWPQIAQFAPS; GTGPEAGLPYGANKD; GTGPEAGLPYGANKDG; GTGQAITVTPEANMDQ; GTGTIYTELEPPCRFV; GTGVLTESNKKFLPFQ; GTGVLTESNKKFLPFQQFGR; GTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILD; GTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITP; GTHWFVTQRNFY; GTHWFVTQRNFYEPQ; GTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQF; GTIYTELEPPCRFVTD; GTKRFDNPVLPFNDGV; GTLLVDSDLNDFVSD; GTLLVDSDLNDFVSDA; GTLSYEQFKKGVQIPC; GTNGTKRFDNPV; GTRNPANNAAIVLQLP; GTTLDSKTQSLLIVN; GTTLPKGFYAEGSRGG; GTTSPISEHDYQ; GTTSPISEHDYQIGGY; GTTTLNGLWLDDVVY;

GTVSWNLREMLAHAEE; GTVTLKQGEIKDATP; GTVTLKQGEIKDATPS; GTWLTYTGAIKLDDK; GTWLTYTGAIKLDDKD; GTYEGNSPFHPLADN; GTYEGNSPFHPLADNK; GVAGALNKATNNAMQ; GVAGALNKATNNAMQV; GVAMPNLYKMQRMLL; GVEAVMYMGTLSYEQF; GVFVSNGTHWFV; GVFVSNGTHWFVTQR; GVFVSNGTHWFVTQRNFYEP; GVFVSNGTHWFVTQRNFYEPQI; GVGGKPCIKVATVQS; GVGGKPCIKVATVQSK; GVHAGTDLEGNFYGPF; GVITHDVSSAINRPQI; GVKDCVVLHSYFTSDY; GVLTESNKKFLPFQ; GVLTESNKKFLPFQQ; GVLTESNKKFLPFQQF; GVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILD; GVLTLDNQDLNGNWYD; GVLVPHVGEIPVAYR; GVPINTNSSPDDQIGY; GVPINTNSSPDDQIGYYRRA; GVPVVDSYYSLLMPI; GVQIPCTCGKQATKYL; GVSVITPGTNTSNQVA; GVSVITPGTNTSNQVAVLYQ; GVTQNVLYENQKLIA; GVTQNVLYENQKLIAN; GVVCTEIDPKLDNYY; GVVCTEIDPKLDNYYK; GVVREFLTRNPAWRKA; GVYCCREHEHEIAW; GVYYHKNNKSWMESE; GVYYHKNNKSWMESEF; GVYYHKNNKSWMESEFRVY; GWDYPKCDRAMPNML; GWDYPKCDRAMPNMLR; GWIFGTTLDSKT; GWIFGTTLDSKTQSL; GWTFGAGAALQIPFA; GYHLMSFPQSAP; GYKKPASRELKVTFF; GYLKLTDNVYIKNADI; GYLTSSSKTPEEHFI; GYLTSSSKTPEEHFIE; GYRVTKNSKVQIGEYT; GYTVEEAKTVLKKCK; GYTVEEAKTVLKKCKS; GYVDTPNNTDFSRVSA; GYYRRATRRIRGGDG; GYYRRATRRIRGGDGK; HADQLTPTWRVY; HADQLTPTWRVYSTG; HADTLTPTWRVYSTGS; HAGTDLEGNFYGPFVD; HANLDSCKRVLNVVC; HAPATVCGPKKSTNLVKNKCVN; HCGETSWQTGDFVKAT; HDFFKFRIDGDMVPHI; HDIGNPKAIKCVPQAD; HDYQIGGYTEKW; HDYQIGGYTEKWESGV; HECFVKRVDWTIEYP; HEFCSQHTMLVKQGD; HEGKTFYVLPNDDTLR; HEHEIAWYTERSEKS; HEIAWYTERSEKSY; HEIAWYTERSEKSYE; HFGAGSDKGVAPGTA; HFPREGVFVSNG; HFPREGVFVSNGTHWFVTQR; HFSMMILSDDAVVCF; HGGGVAGALNKATNNA; HGLAAVNSVPWDTIAN; HGTFTCASEYTGNYQ; HGTILTRPLLESELV; HGVVFLHVTYVP; HHLGRCDIKDLPKEI; HHLGRCDIKDLPKEIT; HHMELPTGVHAGTDL; HHMELPTGVHAGTDLE; HIDHPNPKGFCDLKGK; HIGTRNPANNAAIVL; HIGTRNPANNAAIVLQ; HIQWMVMFTPLVPFW; HKHAFLCLFLLPSLA; HKPIVWHVNNATNKAT; HLAKALNDFSNSGSDV; HLDGEVITFDNLKTLL; HLGRCDIKDLPKEIT; HLGRCDIKDLPKEITV; HLMGWDYPKCDRAMPN; HLMSFPQSAPHGVVF; HLRIAGHHLGR; HLRIAGHHLGRCDIK; HLRIAGHHLGRCDIKD; HLYFDKAGQKTYERH; HMELPTGVHAGTDLEG; HNQDVNLHSSRLSFKE; HNSEVGPEHSLAEYH; HNSEVGPEHSLAEYHN; HPDSATLVSDIDITF; HPNCVNCLDDRCILHC; HQKLLKSIAATRGATV; HRFYRLANECAQVLS; HRLYECLYRNRDVDT; HSKGLYRKCVKSREE; HSKGLYRKCVKSREET; HSLSHFVNLDNLRAN; HSMQNCVLKLKVDTAN; HSSSSDNIALLVQ; HSTQDLFLPFFS; HSTQDLFLPFFSNVT; HSWNADLYKLMGHFA; HSWNADLYKLMGHFAW; HSYGADLKSFDLGDE; HTANKWDLIISD; HTDLMAAYVDNSSLT; HTDLMAAYVDNSSLTI; HTSPDVDLGDISGIN; HTSPDVDLGDISGINA; HTTCCSLSHRFYRLA; HTTDPSFLGRYMSAL; HTVMPLSAPTLVPQE; HVDTDLTKPYIKWDL; HVETFYPKLQSSQAWQ; HVGEIPVAYRKVLLR; HVICTSEDMLNPNYE; HVNNATNKATYKPNTW; HVTYVPAQEKNF; HVTYVPAQEKNFTTAP; HVVAFNTLLFLMSFT; HWFVTQRNFYEP; HWFVTQRNFYEPQII; HWFVTQRNFYEPQIITT; HYTPSFKKGAKLLHK; IAATRGATVVIGTSKF; IADTTDAVRDPQ; IADTTDAVRDPQTLE; IADTTDAVRDPQTLEI; IAEIPKEEVKPFITES; IAGHHLGRCDIKDLP; IAGHHLGRCDIKDLPK; IAKNTVKSVGKFCLE; IALKGGKIVNNWLKQ; IAMACLVGLMWLSYF; IANQFNSAIGKIQDS; IANQFNSAIGKIQDSL; IANYAKPFLNKVVST; IANYAKPFLNKVVSTT; IAQFAPSASAFFGMS; IAQVDVVNFNLT; IAQYTSALLAGTITSG; IASEFSSLPSYAAFA; IASFRLFARTRSMWS; IAWYTERSEKSYELQ; IAYIICISTKHFYWF; ICHDGKAHFPRE; ICHDGKAHFPREGVF; ICHDGKAHFPREGVFV; ICISTKHFYWFFSNY; ICQAVTANVNALLST; ICTSEDMLNPNYEDL; ICTSEDMLNPNYEDLL; ICYTPSKLIEYTDFAT; IDGALLTKSSEYKGP; IDGDMVPHISRQRLT; IDHPNPKGFCDLKGK; IDHPNPKGFCDLKGKY; IDIGNYTVSCLPFTI; IDITFLKKDAPYIVG; IDITFLKKDAPYIVGD; IDLLLDDFVEIIKSQD; IDLVPNQPYPNASFD; IDLVPNQPYPNASFDN; IDPKLDNYYKKDNSY; IDPKLDNYYKKDNSYF; IDRLITGRLQSL; IDRLITGRLQSLQTY; IDSYFVVKRHTFSNY; IDTKRGVYCCREHEH; IDYKHYTPSFKKGAK; IEELFYSYATHSDKF; IEFLKRGDKSVYYTSN; IELCVDEAGSKSPIQY; IELKFNPPALQDAYY; IELKFNPPALQDAYYR; IERFVSLAIDAYPLT; IETISLAGSYKDWSY; IEVNSFSGYLKLTDNV; IEVTGDSCNNYMLTYN; IEYPIIGDELKINAAC; IEYTDFATSACVLAAE; IFFITGNTLQCIMLV; IFGTTLDSKTQSLLI; IFGTTLDSKTQSLLIVNNAT; IFTIGTVTLKQGEIK; IFTIGTVTLKQGEIKD; IFTQTTETAHSCNV; IFTQTTETAHSCNVN; IGAGICASYQTQ; IGDCATVHTANKWDL; IGDCATVHTANKWDLI; IGDELKINAACRKVQ; IGDELKINAACRKVQH; IGDPAQLPAPRTLLTK; IGEYTFEKGDYGDAVV; IGFDVEGCHATREAV; IGGYTEKWESGVKDCV; IGIVNNTVYDPL; IGIVNNTVYDPLQPEL; IGKIQDSLSSTASA; IGKIQDSLSSTASAL; IGMEVTPSGTWLTYT; IGNYKLNTDHSSSSD; IGNYKLNTDHSSSSDN; IGNYTVSCLPFTINC; IGPDMFLGTCRRCPA; IGPERTCCLCDRRAT; IGTRNPANNAAIVLQ; IGTRNPANNAAIVLQL; IGTVTLKQGEIK; IGTVTLKQGEIKDATP; IGVTQNVLYENQKLI; IGVTQNVLYENQKLIA; IHADQLTPTWRV; IHADQLTPTWRVYSTG; IHADQLTPTWRVYSTGSN; IHADQLTPTWRVYSTGSNVF; IHADQLTPTWRVYSTGSNVFQTR; IHFGAGSDKGVAPGT; IIAMSAFAMMFVKHK; IIGDELKINAACRKVQ; IIIGGAKLKALNLGET; IIKNLSKSLTENKYS; IIKNLSKSLTENKYSQ; IIKTIQPRVEKKKLD; IIKTIQPRVEKKKLDG; IINLIIKNLSKSLTE; IINLIIKNLSKSLTEN; IINLVQMAPISAMVR; IIPARARVECFDKFKV; IIPLTTAAKLMVVIP; IIRGWIFGTTLDSKT; IISDMYDPKTKNVTKE; IISNEKQEILGTVSWN; IITTDNTFVSGNCDV; IITTDNTFVSGNCDVV; IIWVATEGALNTPKD; IIWVATEGALNTPKDH; IKASMPTTIAKNTVK; IKCVPQADVEWKFYDA; IKDLPKEITVATSRT; IKDLPKEITVATSRTL; IKESVQTFFKLVNKF; IKFADDLNQLTGYKKP; IKGLNNLNRGMVLGSL; IKITEHSWNADLYKLM; IKLDDKDPNFKDQVI; IKLDDKDPNFKDQVIL; IKLIFLWLLWPVTLA; IKLLGVGGKPCIKVA; IKLLGVGGKPCIKVAT; IKNADIVEEAKKVKP; IKNLSKSLTENKYSQL; IKRSDARTAPHGHVMV; IKTIQPRVEKKKLDGF; IKVATVQSKMSDVKC; IKVATVQSKMSDVKCT; IKVTLVFLFVAAIFY; IKWDLLKYDFTEERL; IKWDLLKYDFTEERLK; ILASFSASTSAFVET; ILAYCNKTVGELGDV; ILAYCNKTVGELGDVR; ILCIMSDRDLYDKLQ; ILCIMSDRDLYDKLQF; ILDITPCSFGGVSVITPGTN; ILDITPCSFGGVSVITPGTNTSNQVAVLY; ILGAGCFVDDIVKTDG; ILGLPTQTVDSSQGSE; ILLAKDTTEAFEKMV; ILLLDQALVSDVGDSA; ILLNVPLHGTILTRP; ILTLTRALTAESHVD; IMASLVLARKHTTCC; IMMNVAKYTQLCQYL; IMRTFKVSIWNLDYI; IMSDRDLYDKLQFTSL; INAACRKVQHMVVKA; INANQVIVNNLDKSA; INANQVIVNNLDKSAG; INLIIKNLSKSLTENK; INRPQIGVVREFLTR; INTNSSPDDQIG; INTNSSPDDQIGYYR; INTNSSPDDQIGYYRR; INVFAFPFTIYSLLL; IPARARVECFDKFKVN; IPCTCGKQATKYLVQQ; IPDYNTYKNTCDGTTF; IPFAMQMAYRFN; IPFAMQMAYRFNGIGV; IPIGAGICASYQTQT; IPIGAGICASYQTQTN; IPKDMTYRRLISMMG; IPKEEVKPFITESKP; IPKEEVKPFITESKPS; IPLTTAAKLMVVIPD; IPLTTAAKLMVVIPDY; IPRRNVATLQAENVTG; IPTITQMNLKYAISA; IPTITQMNLKYAISAK; IPTTCANDPVGFTLKN; IPVAYRKVLLRKNGNK; IQDSLSSTASALGKL; IQFPNTYLEGSVRVV; IQLSSYSLFDMSKFPL; IQPRVEKKKLDGFMG; IQPRVEKKKLDGFMGR; IQQVVDADSKIVQLS; IQWMVMFTPLVPFWI; IQYIDIGNYTVSCLP; IRAAEIRASANLAAT; IRAAEIRASANLAATK; IRASANLAATKMSEC; IRASANLAATKMSECV; IRCLWSTKPVETSNS; IRCLWSTKPVETSNSF; IRGGDGKMKDLSPRW; IRGGDGKMKDLSPRWY; IRKLHDELTGHMLDM; IRKSNHNFLVQAGNV; IRQGTDYKHWPQIAQ; IRQLLFVVEVVDKYF; IRQLLFVVEVVDKYFD; IRSAAKKNNLPFKLT; IRSAAKKNNLPFKLTC; IRSVYPVASPNECNQ; IRSVYPVASPNECNQM; ISAARQGFVDSDVET; ISAKNRARTVAGVSI; ISDEFSSNVANYQKV; ISDEFSSNVANYQKVG; ISDMYDPKTKNVTKE; ISDMYDPKTKNVTKEN; ISDYDYYRYNLPTMC; ISDYDYYRYNLPTMCD; ISEHDYQIGGYTEKW; ISEHDYQIGGYTEKWE; ISMATNYDLSVVNAR; ISMATNYDLSVVNARL; ISMMGFKMNYQVNGY; ISMMGFKMNYQVNGYP; ISNEKQEILGTVSWN; ISPYNSQNAVASKIL; ISPYNSQNAVASKILG; ISRQRLTKYTMADLV; ISTKHFYWFFSNYLK; ITDVFYKENSYTTTIK; ITEEVGHTDLMAAYV; ITEHSWNADLYKLM; ITEHSWNADLYKLMGH; ITESKPSVEQRKQDDK; ITFGGPSDSTGSNQNG; ITFLKKDAPYIVGDV; ITGLHPTQAPTHLSV; ITGLYPTLNISDEFSS; ITGRLQSLQTYVTQQ; ITHDVSSAINRPQIG; ITHDVSSAINRPQIGV; ITIAYIICISTKHFY; ITLATCELYHYQECV; ITLKKRWQLALSKG; ITNCVKMLCTHTGTGQ; ITPGTNTSNQVAVLY; ITPGTNTSNQVAVLYQ; ITQMNLKYAISAKNRA; ITREEAIRHVRAWIG; ITREEAIRHVRAWIGF; ITSGDGTTSPISEHD; ITSGDGTTSPISEHDY; ITSGWTFGAGAALQIPFAMQMAYRFNGIGV; ITTDNTFVSGNCDVV; ITTDNTFVSGNCDVVI; ITTYPGQGLNGYTVE; ITVTPEANMDQESFG; ITVTPEANMDQESFGG; IVCRFDTRVLSNLNL; IVDSVTVKNGSIHLY; IVESCGNFKVTKGKA; IVGDVVQEGVLTAVV; IVITSGDGTTSPISE; IVITSGDGTTSPISEH; IVLQLPQGTTLPKGF; IVLQLPQGTTLPKGFY; IVNNATNVVIKVCEF; IVNNATNVVIKVCEFQ; IVNNLDKSAGFPFNK; IVNNLDKSAGFPFNKW; IVNNTVYDPLQPELD; IVQLSEISMDNSPNL; IVQLSEISMDNSPNLA; IVQMLSDTLKNLSDRV; IVTRCLNRVCTNYMP; IVVFDEISMATNYDL; IVVFDEISMATNYDLS; IVWHVNNATNKATYKP; IWNLDYIINLIIKNL; IWNVKDFMSLSEQLR; IWVATEGALNTPKDHI; IWVATEGALNTPKDHIGTRN; IYCPACHNSEVGPEH; IYLYLTFYLTNDVSF; IYNLLKDCPAVAKHD; IYTELEPPCRFVTDTP; KAALLADKFPVLHDIG; KAKKGAWNIGEQKSIL; KALNDFSNSGSDVLY; KATCEFCGTENLTKE; KATNNAMQVESDDYIA; KATYKPNTWCIRCLW; KAVFISPYNSQNAVAS; KAYKDYLASGGQPIT; KAYKDYLASGGQPITN; KAYNVTQAFGRRGPE; KAYNVTQAFGRRGPEQ; KCDHCGETSWQTGDF; KCDLQNYGDSATLPK; KCDLQNYGDSATLPKG; KCKSAFYILPSIISN; KCVPQADVEWKFYDAQ; KCWTETDLTKGPHEF; KCWTETDLTKGPHEFC; KDAPYIVGDVVQEGV; KDCSKVITGLHPTQAP; KDEDDNLIDSYFVVKR; KDFMSLSEQLRKQIR; KDFYDFAVSKGFFKE; KDFYDFAVSKGFFKEG; KDGHVETFYPKLQSS; KDGHVETFYPKLQSSQ; KDGIIWVATEGA; KDGIIWVATEGALNT; KDHIGTRNPANNAA; KDHIGTRNPANNAAIV; KDLPKEITVATSRTLS; KDLSPRWYFY; KDNSYFTEQPIDLVPN; KDPNFKDQVILLNKHI; KDQVILLNKHIDAYKT; KDWSYSGQSTQLGIE; KDWYDFVENPDILRVY; KEELDKYFKNHT; KEELDKYFKNHTSPD; KEELDKYFKNHTSPDV; KEELDKYFKNHTSPDVD; KEELDKYFKNHTSPDVDLGDIS; KEEVKPFITESKPSV; KEEVKPFITESKPSVE; KEITVATSRTLSYYK; KEKVNINIVGDFKLNE; KENSYTTTIKPVTYKL; KEPCSSGTYEGNSPF; KEPCSSGTYEGNSPFHP; KFADDLNQLTGYKKPA; KFDEDDSEPVLKGVK; KFDEDDSEPVLKGVKL; KFDTFNGECPNFVFP; KFLALCADSIIIGGA; KFLPFQQFGRDIADTT; KFPRGQGVPINTNSSP; KFVCDNIKFADDLNQL; KFYDAQPCSDKAYKIE; KGAGGHSYGADLKSFD; KGAWNIGEQKSILSPL; KGDYGDAVVYRGTTTY; KGFCDLKGKYVQIPTT; KGFCKLHNWNCVNCD; KGFCKLHNWNCVNCDT; KGFFKEGSSVELKHFF; KGGRFVLALLSDLQD; KGIMMNVAKYTQLCQY; KGKAKKGAWNIGEQKS; KGKYVQIPTTCANDPV; KGLNNLNRGMVLGSL; KGLNNLNRGMVLGSLA; KGLPWNVVRIKIVQM; KGPITDVFYKENSYT; KGPITDVFYKENSYTT; KGSFLNGSCGSVGFNI; KGTLEPEYFNSVCRLM; KGVEAVMYMGTLSYEQ; KGVITHDVSSAINRPQ; KGVLPQLEQPYVFIKR; KGVQIPCTCGKQATK; KGVQIPCTCGKQATKY; KGYHLMSFPQSA; KGYHLMSFPQSAPHG; KGYHLMSFPQSAPHGV; KHDFFKFRIDGDMVPH; KHFFFAQDGNAAISDY; KHGGGVAGALNKATN; KHGGGVAGALNKATNN; KHGTFTCASEYTGNYQ; KHWPQIAQFAPSASAFFGMS; KIADKYVRNLQHRLY; KIAEIPKEEVKPFIT; KIAEIPKEEVKPFITE; KIKPHNSHEGKTFYV; KILGLPTQTVDSSQG; KIQDSLSSTASALGK; KITEHSWNADLYKLMG; KIVQMLSDTLKNLSD; KKDNSYFTEQPIDLV; KKDNSYFTEQPIDLVP; KKDWYDFVENPDILRV; KKFDTFNGECPNFVFP; KKFLPFQQFGR; KKFLPFQQFGRD; KKFLPFQQFGRDIAD; KKFLPFQQFGRDIADT; KKGVQIPCTCGKQATK; KKKLDGFMGRIRSVY; KKLDGFMGRIRSVYPV; KKPRQKRTATKAYNVT; KKPRQKRTATKAYNVTQAFG; KKSAAEASKKPRQKR; KKSAAEASKKPRQKRT; KKSTNLVKNKCVN; KKSTNLVKNKCVNFNF; KKWKYPQVNGLTSIK; KLAKKFDTFNGECPNF; KLALGGSVAIKITEH; KLCEEMLDNRATLQAI; KLDDKDPNFKDQVILL; KLDDKDPNFKDQVILLNKH; KLDGFMGRIRSVYPV; KLDNYYKKDNSYFTEQ; KLEKMADQAMTQMYK; KLGASQRVAGDSGFA; KLGASQRVAGDSGFAA; KLGSLVVRCSFYEDF; KLHDELTGHMLDMYSV; KLIANQFNSAIGKIQ; KLIANQFNSAIGKIQD; KLLHKPIVWHVNNATN; KLLKSIAATRGATVV; KLLKSIAATRGATVVI; KLMVVIPDYNTYKNT; KLNTDHSSSSDNIAL; KLNTDHSSSSDNIALL; KLNTDHSSSSDNIALLVQ; KLQFTSLEIPRRNVAT; KLQSSQAWQPGVAMP; KLSHQSDIEVTGDSC; KLTCATTRQVVNVVTT; KLTDNVYIKNADIVEE; KLWAQCVQLHNDILL; KMKDLSPRWYFYYLG; KMLCTHTGTGQAITV; KMNYQVNGYPNMFITR; KMQRMLLEKCDLQNY; KMSECVLGQSKR; KNFTTAPAICHD; KNFTTAPAICHDGKAH; KNGNKGAGGHSYGADL; KNHTSPDVDLGDISG; KNHTSPDVDLGDISGI; KNKCVNFNFNGL; KNLSDRVVFVLWAHG; KNLSKSLTENKYSQL; KNLSKSLTENKYSQLD; KNSKVQIGEYTFEKGD; KNVSLDNVLSTFISA; KNVSLDNVLSTFISAA; KPCIKVATVQSKMSDV; KPFITESKPSVEQRK; KPFITESKPSVEQRKQ; KPFLNKVVSTTTNIVT; KPGGTSSGDATTAYAN; KPGNFNKDFYDFAVSK; KPIVWHVNNATNKAT; KPIVWHVNNATNKATY; KPREQIDGYVMHANY; KPREQIDGYVMHANYI; KPRQKRTATKAYNVTQ; KPVETSNSFDVLKSED; KQARSEDKRAKVTSA; KQATKYLVQQESPFVM; KQEILGTVSWNLREM; KQGDDYVYLPYPDPSR; KQIRSAAKKNNLPFKL; KQRRPQGLPNNTASWF; KQVEQKIAEIPKEEVK; KQYGDCLGDIAARDLI; KRFDNPVLPFNDGVY; KRFDNPVLPFNDGVYF; KRGDKSVYYTSNPTTF; KRGVYCCREHEHEIAW; KRNRATRVECTTIVN; KRNVIPTITQMNLKYA; KRPINPTDQSSYIVD; KRPINPTDQSSYIVDS; KRRVVFNGVSFSTFE; KRRVVFNGVSFSTFEE; KRSDARTAPHGHVMV; KRTATKAYNVTQAFGR; KRTATKAYNVTQAFGRRGPE; KRVDWTIEYPIIGDEL; KRVLNVVCKTCGQQQT; KSAAEASKKPRQKRT; KSAAEASKKPRQKRTA; KSAAEASKKPRQKRTATK; KSAFYILPSIISNEK; KSAFYILPSIISNEKQ; KSDGTGTIYTELEPP; KSEDAQGMDNLACED; KSEDAQGMDNLACEDL; KSEFDRDAAMQRKLE; KSEKQVEQKIAEIPK; KSEKQVEQKIAEIPKE; KSFDLGDELGTDPYEDFQENWNTKH; KSIAATRGATVVIGT; KSIAATRGATVVIGTS; KSLTENKYSQLDEEQ; KSLTENKYSQLDEEQP; KSNIIRGWIFGTTLD; KSNIIRGWIFGTTLDS; KSPIQYIDIGNYTVSC; KSSEYKGPITDVFYKE; KSTNLVKNKCVNFNF; KSTNLVKNKCVNFNFN; KSVLYYQNNVFMSEAK; KSYELQTPFEIKLAKK; KTCGQQQTTLKGVEAV; KTFYVLPNDDTLRVE; KTFYVLPNDDTLRVEA; KTIQPRVEKKKLDGFM; KTQSLLIVNNATNVV; KTQSLLIVNNATNVVI; KTQSLLIVNNATNVVIKVCEFQ; KTVGELGDVRETMSY; KTVGELGDVRETMSYL; KVATVQSKMSDVKCTS; KVCEFQFCNDPF; KVGILCIMSDRDLYDK; KVGMQKYSTLQGPPGT; KVKPTVVVNAANVY; KVKYLFYKGLNNLN; KVLLRKNGNKGAGGHS; KVNINIVGDFKLNEE; KVNINIVGDFKLNEEI; KVQIGEYTFEKGDYGD; KVTFFPDLNGDVVAI; KVTFFPDLNGDVVAID; KVVKVTIDYTEISFM; KVYYGNALDQAISMWA; KWDLIISDMYDPKTKN; KWDLLKYDFTEERLKL; KWGKARLYYDSMSYE; KYDFTEERLKLFDRY; KYFDCYDGGCINANQV; KYFKNHTSPDVD; KYFKNHTSPDVDLGDI; KYFSGAMDTTSYREAA; KYLVQQESPFVMMSAP; KYPQVNGLTSIKWAD; KYPQVNGLTSIKWADN; KYVQIPTTCANDPV; KYVQIPTTCANDPVGF; KYVRNLQHRLYECLY; KYWDQTYHPNCVNCLD; LAATVRLQAGNATEV; LAATVRLQAGNATEVP; LAAVNSVPWDTIANYA; LAAVYRINWITGGIA; LADKFPVLHDIGNPKAIKCVP; LAGSYKDWSYSGQSTQ; LAGTITSGWTFG; LAGTITSGWTFGAGAA; LAKALNDFSNSGSDV; LAKALNDFSNSGSDVL; LAKKFDTFNGECPNFV; LAKNVSLDNVLSTFI; LALCADSIIIGGAKL; LALLSDLQDLKWARF; LALLSDLQDLKWARFP; LALSKGVHFVCNLLL; LALYNKYKYFSGAMD; LARKHTTCCSLSHRF; LATCELYHYQECVRG; LATHGLAAVNSVPWD; LAWLYAAVINGDRWF;

LCADSIIIGGAKLKA; LCEEMLDNRATLQAI; LCEEMLDNRATLQAIA; LCEKALKYLPIDKC; LCIMSDRDLYDKLQFT; LCQYLNTLTLAVPYN; LCRMNSRNYIAQVDV; LCRMNSRNYIAQVDVV; LCTFLLNKEMYLKLR; LCTHTGTGQAITVTP; LCVDEAGSKSPIQY; LCVDEAGSKSPIQYI; LCVDEAGSKSPIQYID; LCVDEAGSKSPIQYIDIGNY; LDDKDPNFKDQVILLN; LDDRCILHCANFNVL; LDGEVITFDNLKTLL; LDGEVITFDNLKTLLS; LDGVVCTEIDPKLDN; LDITPCSFGGVSVIT; LDKRTTCFSVAALTN; LDKVEAEVQIDRLITG; LDKVEAEVQIDRLITGRLQSLQTYVTQQ; LDKYFKNHTSPDVDLG; LDMYSVMLTNDNTSR; LDNLRANNTKGSLPI; LDNLRANNTKGSLPIN; LDNVLSTFISAARQG; LDNYYKKDNSYFTE; LDNYYKKDNSYFTEQP; LDQALVSDVGDSAEV; LDQALVSDVGDSAEVA; LDSFKEELDKYF; LDSFKEELDKYFKNH; LDSFKEELDKYFKNHT; LDSFKEELDKYFKNHTSPDVDLG; LDSFKEELDKYFKNHTSPDVDLGDISGI; LDYKAFKQIVESCGN; LDYKAFKQIVESCGNF; LEASFNYLKSPNFS; LECNVKTTEVVGDII; LEGIQYGRSGETLGV; LEGNFYGPFVDRQTAQ; LEGSVAYESLRPDTRY; LEILDITPCSFGGVSV; LEIPRRNVATLQAEN; LEIPRRNVATLQAENV; LEKCDLQNYGDSATL; LEKMADQAMTQMYKQ; LEMELTPVVQTIEVN; LEMKSEKQVEQKIAEI; LEQPYVFIKRSDART; LEQWNLVIGFLFLTW; LEQYVFCTVNALPET; LEQYVFCTVNALPETT; LFDMSKFPLKLRGTA; LFDRYFKYWDQTYHP; LFDRYFKYWDQTYHPN; LFVVEVVDKYFDCYD; LFYSYATHSDKFTDGV; LGAGCFVDDIVKTDG; LGASQRVAGDSGFAA; LGASQRVAGDSGFAAY; LGASQRVAGDSGFAAYSRYRIGNYKLNTDHSS; LGDELGTDPYEDFQ; LGDELGTDPYEDFQEN; LGDIAARDLICAQKF; LGIEFLKRGDKSVYYT; LGKPREQIDGYVMHA; LGKPREQIDGYVMHAN; LGLPTQTVDSSQGSE; LGLPTQTVDSSQGSEY; LGRCDIKDLPKEITVA; LGTCRRCPAEIVDTV; LGTCRRCPAEIVDTVS; LGTDPYEDFQENWNTK; LGTGPEAGLPYGANKD; LGTVSWNLREMLAHA; LGVGGKPCIKVATVQS; LGVLVPHVGEIPVAY; LHAPATVCGPKKSTNL; LHDIGNPKAIKCVPQ; LHFLPRVFSAVGNIC; LHGTILTRPLLESEL; LHKPIVWHVNNATNK; LHNDILLAKDTTEAF; LHVTYVPAQEKNFTT; LIANQFNSAIGKIQD; LIANQFNSAIGKIQDS; LIELCVDEAGSKSPIQYID; LIEYTDFATSACVLA; LIGCNYLGKPREQID; LIGCNYLGKPREQIDG; LIGDCATVHTANKWD; LIGDCATVHTANKWDL; LIIKNLSKSLTENKYS; LIIMRTFKVSIWNLD; LIISDMYDPKTKNVT; LIISDMYDPKTKNVTK; LILAYCNKTVGELGDV; LILMTARTVYDDGAR; LIRAAEIRASANLAA; LIRQGTDYKHWPQIAQFAPS; LISMMGFKMNYQVNGY; LITGRLQSLQTYVTQQLIRA; LIVNNATNVVIKVCE; LKDGTCGLVEVEKGV; LKFPRGQGVPINTNS; LKFPRGQGVPINTNSS; LKGKYVQIPTTCANDP; LKGVEAVMYMGTLSY; LKGVEAVMYMGTLSYE; LKHGTFTCASEYTGN; LKHGTFTCASEYTGNY; LKINAACRKVQHMVV; LKLFDRYFKYWDQTY; LKLRGTAVMSLKEGQ; LKLTDNVYIKNADIV; LKLTDNVYIKNADIVE; LKRGDKSVYYTSNP; LKRGDKSVYYTSNPTT; LKSEDAQGMDNLACE; LKSEDAQGMDNLACED; LKSFDLGDELGTDPY; LKSFDLGDELGTDPYE; LKSIAATRGATVVIG; LKSIAATRGATVVIGT; LKTLATHGLAAVNSVP; LKTLVATAEAELAKN; LKVDTANPKTPKYKF; LKVDTANPKTPKYKFV; LKYAISAKNRARTVA; LKYDFTEERLKLFDRY; LLADKFPVLHDIGNP; LLAGTITSGWTFGAG; LLAGTITSGWTFGAGA; LLARAGKASCTLSEQ; LLARAGKASCTLSEQL; LLCRMNSRNYIA; LLCRMNSRNYIAQVDV; LLDDFVEIIKSQDLS; LLDDFVEIIKSQDLSV; LLDKRTTCFSVAALT; LLDQALVSDVGDSAEV; LLEILQKEKVNINIV; LLEQWNLVIGFLFLT; LLESELVIGAVILRG; LLGVGGKPCIKVATVQ; LLHAPATVCGPKKST; LLHKPIVWHVNNATNK; LLKDCPAVAKHDFFKF; LLKDCPAVAKHDFFKFRIDGDMVPHISRQRLTKYTMADL; LLKEPCSSGTYEGNS; LLKEPCSSGTYEGNSP; LLKSIAATRGATVVIG; LLKYDFTEERLKLFD; LLKYDFTEERLKLFDR; LLLCRMNSRNYIAQV; LLLCRMNSRNYIAQVD; LLLDDFVEIIKSQDLS; LLLDQALVSDVGDSA; LLLDQALVSDVGDSAE; LLMPILTLTRALTAE; LLNVPLHGTILTRPL; LLNVPLHGTILTRPLLESELVIGAVI; LLQFAYANRNRFLYI; LLRKNGNKGAGGHSY; LLRKNGNKGAGGHSYG; LLSDLQDLKWARFPKS; LLSMQGAVDINKLCEE; LLSTDGNKIADKYVRN; LLTKGTLEPEYFNSVC; LLTKSSEYKGPITDV; LLTKSSEYKGPITDVF; LLVDSDLNDFVSDAD; LLVDSDLNDFVSDADS; LLYIDINGNLHPDSA; LLYIDINGNLHPDSAT; LMCQPILLLDQALVS; LMDGSIIQFPNTYLE; LMGWDYPKCDRAMPNM; LMIERFVSLAIDAYP; LMSFPQSAPHGVVFL; LMSFTVLCLTPVYSF; LMVVIPDYNTYKNTCD; LMWLIINLVQMAPIS; LNDFSNSGSDVLYQP; LNDFSNSGSDVLYQPP; LNDFVSDADSTLIGD; LNDNLLEILQKEKVN; LNDNLLEILQKEKVNI; LNGDVVAIDYKHYTPS; LNGLWLDDVVYCPRH; LNGSCGSVGFNIDYD; LNGSCGSVGFNIDYDC; LNGYTVEEAKTVLKK; LNHTKKWKYPQVNGL; LNISDEFSSNVANYQ; LNISDEFSSNVANYQK; LNKATNNAMQVESDDY; LNNLNRGMVLGSLAA; LNNLNRGMVLGSLAAT; LNPNYEDLLIRKSNHN; LNRFTTTLNDFNLVAM; LNRGMVLGSLAATVR; LNSIIKTIQPRVEKKK; LNTDHSSSSDNIALLV; LNTLTLAVPYNMRVI; LNTPKDHIGTRNPAN; LNTPKDHIGTRNPANN; LNVVCKTCGQQQTTL; LNVVCKTCGQQQTTLK; LPAPRTLLTKGTLEPE; LPFAMGIIAMSAFAM; LPFKLTCATTRQVVN; LPFKLTCATTRQVVNV; LPFQQFGRDIAD; LPFQQFGRDIADT; LPFQQFGRDIADTTDA; LPFQQFGRDIADTTDAVRDPQTLEIL; LPFTINCQEPKLGSLV; LPPAYTNSFTRGVYYPDKVF; LPPLLTDEMIAQ; LPPLLTDEMIAQYTS; LPQGTTLPKGFY; LPQGTTLPKGFYAEG; LPQGTTLPKGFYAEGS; LPQGTTLPKGFYAEGSRGGS; LPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNSTP; LPSYAAFATAQEAYEQ; LPTGTLLVDSDLNDFV; LPTGVHAGTDLEGNFY; LPTMCDIRQLLFVVE; LPTQTVDSSQGSEYD; LPTQTVDSSQGSEYDY; LPWNVVRIKIVQMLS; LPYGANKDGIIWVAT; LPYGANKDGIIWVATE; LPYGANKDGIIWVATEGALN; LQAGNATEVPANSTV; LQAGNATEVPANSTVL; LQAIASEFSSLPSYAA; LQCIMLVYCFLGYFC; LQDAYYRARAGEAANF; LQDLKWARFPKSDGTG; LQFKRPINPTDQSSYI; LQFTSLEIPRRNVAT; LQFTSLEIPRRNVATL; LQIPFAMQMAYRFNGI; LQIPFAMQMAYRFNGIGVTQ; LQKEKVNINIVGDFK; LQLPQGTTLPKGFYAE; LQLPQGTTLPKGFYAEGSRGGS; LQNYGDSATLPKGIMM; LQPELDSFKEELDKY; LQPELDSFKEELDKYF; LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNI; LQSLQTYVTQQLIRA; LQTPFEIKLAKKFDTF; LRAKHYVYIGDPAQLP; LREMLAHAEETRKLMP; LRGHLRIAGHHLGRC; LRGTAVMSLKEGQIN; LRIAGHHLGRCDIKDL; LRIMASLVLARKHTT; LRKNGNKGAGGHSYGA; LRKVPTDNYITTYP; LRKVPTDNYITTYPGQ; LRPDTRYVLMDGSII; LRPDTRYVLMDGSIIQ; LSAPTLVPQEHYVRIT; LSDDAVVCFNSTYAS; LSDLQDLKWARF; LSDLQDLKWARFPKSD; LSDRVVFVLWAHGFE; LSFCAFAVDAAKAY; LSFKELLVYAADPAM; LSFKELLVYAADPAMH; LSHFVNLDNLRANNT; LSHRFYRLANECAQV; LSKGVHFVCNLLLLF; LSKSLTENKYSQLDE; LSKSLTENKYSQLDEE; LSMQGAVDINKLCEEM; LSNLNLPGCDGGSLYV; LSPRWYFYYLGTGPE; LSSYSLFDMSKFPLK; LSSYSLFDMSKFPLKL; LSTDGNKIADKYVRN; LSTDGNKIADKYVRNL; LSTFISAARQGFVDS; LSVDTKFKTEGLCVD; LSVVNARLRAKHYVY; LSYEQFKKGVQIPCT; LSYEQFKKGVQIPCTC; LSYYKLGASQRVAGDS; LTAESHVDTDLTKPYI; LTCATTRQVVNVVTT; LTDEMIAQYTSALLA; LTDNVYIKNADIVEE; LTDNVYIKNADIVEEA; LTENKYSQLDEEQPM; LTENKYSQLDEEQPME; LTESNKKFLPFQQFG; LTESNKKFLPFQQFGR; LTFYLTNDVSFLAHI; LTKEGATTCGYLPQNA; LTKGTLEPEYFNSVC; LTKGTLEPEYFNSVCR; LTKSSEYKGPITDVFY; LTNDVSFLAHIQWMV; LTPTWRVYSTGS; LTPTWRVYSTGSNVF; LTPTWRVYSTGSNVFQ; LTPVVQTIEVNSFSGY; LTRPLLESELVIGAV; LTSHTVMPLSAPTLVP; LTSIKWADNNCYLAT; LTSIKWADNNCYLATA; LTSSSKTPEEHFIET; LTSSSKTPEEHFIETI; LTTAAKLMVVIPDYN; LTTAAKLMVVIPDYNT; LTVLPPLLTDEMIA; LTYNKVENMTPRDLGA; LTYTGAIKLDDKDPN; LTYTGAIKLDDKDPNF; LTYTGAIKLDDKDPNFKDQV; LVAELEGIQYGRSGE; LVAVPTGYVDTPNNTD; LVDFQVTIAEILLII; LVDSDLNDFVSDADST; LVEVEKGVLPQLEQPY; LVKNKCVNFNFNGLT; LVKQGDDYVYLPYPDP; LVLARKHTTCCSLSH; LVPNQPYPNASFDNF; LVPNQPYPNASFDNFK; LVPQEHYVRITGLYP; LVPQEHYVRITGLYPT; LVQQESPFVMMSAPPA; LVRGFGDSVEEVLSEA; LVSDIDITFLKKDAP; LVSDIDITFLKKDAPY; LVSDVGDSAEVAVKM; LVVRCSFYEDFLEYH; LVYAADPAMHAASGN; LVYAADPAMHAASGNL; LWAHGFELTSMKYFV; LWCKDGHVETFYPKLQ; LWEIQQVVDADSKIVQ; LWSTKPVETSNSFDVL; LYCIDGALLTKSSE; LYENQKLIANQFNSA; LYENQKLIANQFNSAI; LYENQKLIANQFNSAIGKIQ; LYIDINGNLHPDSAT; LYIDINGNLHPDSATL; LYIIKLIFLWLLWPV; LYKLMGHFAWWTAFV; LYLTFYLTNDVSFLA; LYNKYKYFSGAMDTTS; LYPTLNISDEFSSNVA; LYRNRDVDTDFVNEFY; LYVKPGGTSSGDATTA; LYYDSMSYEDQDALF; MAPISAMVRMYIFFA; MCDIRQLLFVVEVVDK; MCYKRNRATRVECTT; MDLFMRIFTIGTVTL; MDQESFGGASCCLYC; MEKLKTLVATAEAEL; MELPTGVHAGTDLEGN; MELTPVVQTIEVNSF; MELTPVVQTIEVNSFS; MEVTPSGTWLTYTGA; MFDAYVNTFSSTFNV; MFDAYVNTFSSTFNVP; MFLGTCRRCPAEIVD; MGFKMNYQVNGYPNMF; MGHFAWWTAFVTNVN; MGRIRSVYPVASPNEC; MGTLSYEQFKKGVQIP; MGTLSYEQFKKGVQIPC; MGWDYPKCDRAMPNML; MHANYIFWRNTNPIQ; MIAQYTSALLAGTIT; MIAQYTSALLAGTITS; MIAQYTSALLAGTITSGWTFGA; MKDLSPRWYFYY; MKDLSPRWYFYYLGTGPEAG; MLDNRATLQAIASEF; MLNPNYEDLLIRKSN; MLSDTLKNLSDRVVFV; MLTNDNTSRYWEPEF; MLTYNKVENMTPRDL; MLTYNKVENMTPRDLG; MLVKQGDDYVYLPYP; MLVYCFLGYFCTCYF; MMGFKMNYQVNGYPNM; MNLKYAISAKNRART; MNSRNYIAQVDV; MNSRNYIAQVDVVNFN; MPILTLTRALTAESH; MPLSAPTLVPQEHYVR; MPNLYKMQRMLLEKCD; MPNMLRIMASLVLAR; MPTTIAKNTVKSVGK; MQGAVDINKLCEEMLD; MQKYSTLQGPPGTGKS; MQMAYRFNGIGVTQNV; MRIFTIGTVTLKQGE; MRNAGIVGVLTLDNQ; MSCAAGTTQTACTDD; MSDRDLYDKLQFTSLE; MSEAKCWTETDLTKG; MSEAKCWTETDLTKGP; MSECVLGQSKRV; MSFPQSAPHGVV; MSRIGMEVTPSGTWLT; MSYEDQDALFAYTKRN; MSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLS; MTQMYKQARSEDKRAK; MTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGV; MVPHISRQRLTKYTM; MVRMYIFFASFYYVW; MVVIPDYNTYKNTCDG; MVYMPASWVMRIMTW; MWLSYFIASFRLFAR; MWSFNPETNILLNVP; MWSFNPETNILLNVPL; MYASAVVLLILMTAR; MYDPKTKNVTKENDSK; MYKGLPWNVVRIKIV; MYMGTLSYEQFKKGVQ; MYSVMLTNDNTSRYW; NAACRKVQHMVVKAA; NAAISDYDYYRYNLPT; NAAIVLQLPQGTTLP; NAAIVLQLPQGTTLPK; NADIVEEAKKVKPTV; NAGIVGVLTLDNQDL; NALLSTDGNKIADKYV; NANQVIVNNLDKSAGF; NAQALNTLVKQLSSN; NARLRAKHYVYIGDP; NARLRAKHYVYIGDPA; NATNKATYKPNTWCIR; NCCDDDYFNKKDWYD; NCCDDDYFNKKDWYDF; NCDTFCAGSTFISDE; NCDVVIGIVNNTVYD; NCLDDRCILHCANFN; NCTEVPVAIHADQLT; NCTEVPVAIHADQLTP; NCTEVPVAIHADQLTPTWRVYSTGSNVFQ; NCYLATALLTLQQIE; NDFSNSGSDVLYQPPQ; NDFVSDADSTLIGDCA; NDILLAKDTTEAFEK; NDNLLEILQKEKVNIN; NDVSFLAHIQWMVMF; NEKQEILGTVSWNLR; NFGDQELIRQGTDYKH; NFKDQVILLNKHIDAY; NFKFVCDNIKFADDLN; NFNKDFYDFAVSKGF; NFVFPLNSIIKTIQP; NFYEPQIITTDNTFV; NFYEPQIITTDNTFVS; NFYGPFVDRQTAQAAG; NGDFLHFLPRVFSAV; NGDVVAIDYKHYTPSF; NGIGVTQNVLYENQK; NGIGVTQNVLYENQKL; NGIGVTQNVLYENQKLIANQFNSAIGKIQD; NGLTSIKWADNNCYL; NGLTSIKWADNNCYLA; NGLTVLPPLLTDEMI; NGLTVLPPLLTDEMIAQYTS; NGNKGAGGHSYGADL; NGNLHPDSATLVSDI; NGNWYDFGDFIQTTPG; NGSCGSVGFNIDYDCV; NGTKRFDNPVLP; NGTKRFDNPVLPFNDG; NGTKRFDNPVLPFNDGVYFA; NGYLTSSSKTPEEHFI; NHTSPDVDLGDISGIN; NICYTPSKLIEYTDF; NICYTPSKLIEYTDFA; NIDYDCVSFCYMHHM; NIGEQKSILSPLYAF; NIIPLTTAAKLMVVI; NIIRGWIFGTTLDSKTQSLLIV; NIKFADDLNQLTGYKK; NINIVGDFKLNEEIA; NISDEFSSNVANYQKV; NIVGDFKLNEEIAII; NKATNNAMQVESDDY; NKATNNAMQVESDDYI; NKDFYDFAVSKGFFK; NKDFYDFAVSKGFFKE; NKDGIIWVATEGALN; NKDGIIWVATEGALNTP; NKGAGGHSYGADLKSF; NKKDWYDFVENPDILR; NKKFLPFQQFGRDIA; NKKFLPFQQFGRDIAD; NKKFLPFQQFGRDIADTTDAV; NKTVGELGDVRETMSY; NKWGKARLYYDSMSYE; NLDKSAGFPFNKWGK; NLDNLRANNTKGSLPI; NLGETFVTHSKGLYR; NLHPDSATLVSDIDI; NLHSSRLSFKELLVY; NLIIKNLSKSLTENKY; NLLEILQKEKVNIN; NLLEILQKEKVNINIV; NLLKDCPAVAKHDFF; NLLLYIDINGNLHPD; NLNRGMVLGSLAATV; NLPLQLGFSTGVNLVA; NLQHRLYECLYRNRD; NLSKSLTENKYSQLDE; NLTKEGATTCGYLPQ; NLVAVPTGYVDTPNNT; NLVKNKCVNFNFNGL; NLVKNKCVNFNFNGLT; NLVQMAPISAMVRMY; NMFITREEAIRHVRA; NMLRIMASLVLARKH; NMRVIHFGAGSDKGVA; NNAAIVLQLPQGTTLP; NNAAIVLQLPQGTTLPKG; NNAAIVLQLPQGTTLPKGFYAEGS; NNAAIVLQLPQGTTLPKGFYAEGSRG; NNATNKATYKPNTWCI; NNATNVVIKVCE; NNATNVVIKVCEFQF; NNLNRGMVLGSLAATV; NNSYECDIPIGA; NNSYECDIPIGAGIC; NNSYECDIPIGAGICA; NNTASWFTALTQH; NNTDFSRVSAKPPPGD; NNTKGSLPINVIVFD; NNTVYDPLQPELDSFKEELDKYF; NNTVYDPLQPELDSFKEELDKYFKNHT; NNVAFQTVKPGNFNKD; NNVFMSEAKCWTETDL; NPANNAAIVLQLPQGT; NPETNILLNVPLHGTI; NPHLMGWDYPKCDRAM; NPIQLSSYSLFDMSKF; NPKGFCDLKGKYVQI; NPNYEDLLIRKSNHNF; NPTDQSSYIVDSVTVK; NPTIQKDVLECNVKT; NPTTFHLDGEVITFD; NPVLPFNDGVYFAST; NQDVNLHSSRLSFKEL; NQDVNLHSSRLSFKEL; NQFNSAIGKIQDSLS; NQFNSAIGKIQDSLSS; NQFNSAIGKIQDSLSSTASALG; NQKLIANQFNSAIGK; NQPYPNASFDNFKFVC; NQVIVNNLDKSAGFPF; NRATRVECTTIVNGVR; NRDVDTDFVNEFYAY; NRDVDTDFVNEFYAYL; NRFLYIIKLIFLWLL; NRFTTTLNDFNLVAMK; NRNRFLYIIKLIFLW; NSAIGKIQDSLSSTA; NSAIGKIQDSLSSTAS; NSFDVLKSEDAQGMDN; NSFDVLKSEDAQGMDNL; NSFSGYLKLTDNVYI; NSFSGYLKLTDNVYIK; NSFTRGVYYPDKVFRSSVLH; NSHEGKTFYVLPNDDT; NSIIKTIQPRVEKKK; NSIIKTIQPRVEKKKL; NSKVQIGEYTFEKGD; NSKVQIGEYTFEKGDY; NSRNYIAQVDVVNFN; NSSPDDQIGYYRRAT; NSVPWDTIANYAKPFL; NSYECDIPIGAG; NSYECDIPIGAGICA; NSYECDIPIGAGICAS; NSYFTEQPIDLVPNQ; NSYFTEQPIDLVPNQP; NSYTTTIKPVTYKLDG; NTASWFTALTQHGKE; NTDFSRVSAKPPPGDQ; NTDHSSSSDNIALLV; NTDHSSSSDNIALLVQ; NTFSSTFNVPMEKLKT; NTKHSSGVTRELMRE; NTLQCIMLVYCFLGY; NTNPIQLSSYSLFDM; NTNSSPDDQIGY; NTNSSPDDQIGYYRR; NTNSSPDDQIGYYRRA; NTPKDHIGTRNPANN; NTPKDHIGTRNPANNA; NTVYDPLQPELDSFK; NTVYDPLQPELDSFKE; NTVYDPLQPELDSFKEELDKYFKNHTSPDVD; NTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDIS; NTWCIRCLWSTKPVE; NTYKNTCDGTTF; NTYLEGSVRVVTTFD; NVAFQTVKPGNFNKDF; NVAITRAKVGILCI; NVANYQKVGMQKYST; NVANYQKVGMQKYSTL; NVFMSEAKCWTETDLT; NVFQTRAGCLIG; NVLAWLYAAVINGDR; NVLEGSVAYESLRPDT; NVPLHGTILTRPLLE; NVTQAFGRRGPEQTQG; NVVCKTCGQQQTTLK; NVVCKTCGQQQTTLKG; NVVIKVCEFQFCNDP; NVVTTKIALKGGKIV; NYAKPFLNKVVSTTTN; NYDLSVVNARLRAKHY; NYEDLLIRKSNHNFL; NYGDSATLPKGIMMNV; NYKLNTDHSSSSDNIA; NYLGKPREQIDGYVMH; NYLKSPNFSKLINII; NYMLTYNKVENMTPRD; NYMLTYNKVENMTPRDV; NYQCGHYKHITSKET; NYQKVGMQKYSTLQGP; NYQVNGYPNMFITRE; NYTVSCLPFTINCQE; NYYKKDNSYFTEQPI; NYYKKDNSYFTEQPID; PAICHDGKAHFP; PAICHDGKAHFPREGV; PAICHDGKAHFPREGVFVSNGT; PALQDAYYRARAGEAA; PANNAAIVLQLPQGT; PANNAAIVLQLPQGTTLPKG; PAPRTLLTKGTLEPE; PAPRTLLTKGTLEPEY; PAQEKNFTTAPAICH; PARARVECFDKFKVNS; PASRELKVTFFPDLN; PATVCGPKKSTN; PATVCGPKKSTNLVK; PATVCGPKKSTNLVKNKC; PATVSVSSPDAVTAY; PATVSVSSPDAVTAYN; PAVAKHDFFKFRIDGD; PAWRKAVFISPYNSQN; PCIKVATVQSKMSDVK; PCRFVTDTPKGPKVK; PCSDKAYKIEELFYS; PCSFGGVSVITP; PCSFGGVSVITPGTN; PCSFGGVSVITPGTNT; PCSFGGVSVITPGTNTSNQVAV; PCSSGTYEGNSPFHPL; PCTCGKQATKYLVQQE; PDAVTAYNGYLTSSSK; PDDQIGYYRRATRRI; PDKVFRSSVLHSTQD; PDKVFRSSVLHSTQDL; PDLNGDVVAIDYKHYT;

PDLNGDVVAIDYKHYTPSFKKGAKLLHKPIVWHVNNATN; PDPSRILGAGCFVDD; PDTRYVLMDGSIIQF; PDYNTYKNTCDGTTFT; PEAGLPYGANKD; PEAGLPYGANKDGIIW; PEAGLPYGANKDGIIWVAT; PEHSLAEYHNESGLK; PEHSLAEYHNESGLKT; PELDSFKEELDKYFK; PELDSFKEELDKYFKN; PELDSFKEELDKYFKNHTSPD; PELDSFKEELDKYFKNHTSPDVD; PELDSFKEELDKYFKNHTSPDVDLG; PEQTQGNFGDQELIRQ; PETTADIVVFDEISM; PETTADIVVFDEISMA; PFAMQMAYRFNG; PFAMQMAYRFNGIGVT; PFAMQMAYRFNGIGVTQNVLYE; PFITESKPSVEQRKQD; PFLNKVVSTTTNIVTR; PFNKWGKARLYYDSMS; PFQQFGRDIADT; PFQQFGRDIADTTDA; PFQQFGRDIADTTDAV; PFQQFGRDIADTTDAVRDPQTL; PFQQFGRDIADTTDAVRDPQTLEILDIT; PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVIT; PFTINCQEPKLGSLVV; PFTIYSLLLCRMNSRNYI; PGGTSSGDATTAYANS; PGNFNKDFYDFAVSK; PGNFNKDFYDFAVSKG; PGQGLNGYTVEEAKT; PHGVVFLHVTYVPAQEKNFTTAPAICHDGK; PHLMGWDYPKCDRAMP; PIDKCSRIIPARARV; PIDLVPNQPYPNASFD; PIGAGICASYQTQTN; PIIGDELKINAACRKV; PILLLDQALVSDVGD; PILTLTRALTAESHVD; PINPTDQSSYIVDSV; PINTNSSPDDQI; PINTNSSPDDQIGYY; PINTNSSPDDQIGYYR; PIQASLPFGWLIVGVA; PIQLSSYSLFDMSKF; PIQLSSYSLFDMSKFP; PIQYIDIGNYTVSCLP; PISEHDYQIGGY; PISEHDYQIGGYTEKW; PITDVFYKENSYTTT; PIVWHVNNATNKATYK; PKCDRAMPNMLRIMA; PKDHIGTRNPANNAA; PKDHIGTRNPANNAAI; PKEEVKPFITESKPSV; PKKSTNLVKNKCVNF; PKKSTNLVKNKCVNFN; PKLDNYYKKDNSYFTE; PLECIKDLLARAGKAS; PLHGTILTRPLLESE; PLLTDEMIAQYT; PLLTDEMIAQYTSAL; PLNSIIKTIQPRVEKK; PLQPELDSFKEELDKY; PLQPELDSFKEELDKYFKNHTSPDV; PLQPELDSFKEELDKYFKNHTSPDVDLGDIS; PLSAPTLVPQEHYVR; PLSAPTLVPQEHYVRI; PLTTAAKLMVVIPDYN; PLVRKIFVDGVPFVV; PNASFDNFKFVCDNIK; PNDDTLRVEAFEYYH; PNFKDQVILLNKHID; PNFKDQVILLNKHIDA; PNNTASWFTALTQHG; PNNTDFSRVSAKPPPG; PNQPYPNASFDNFKF; PNQPYPNASFDNFKFV; PPALQDAYYRARAGE; PPALQDAYYRARAGEA; PPAQYELKHGTFTCAS; PPCRFVTDTPKGPK; PPLLTDEMIAQY; PPLLTDEMIAQYTSA; PQADVEWKFYDAQPCS; PQEHYVRITGLYPTL; PQGTTLPKGFYA; PQGTTLPKGFYAEGS; PQGTTLPKGFYAEGSR; PQIAQFAPSASAFFGM; PQIITTDNTFVSGNCD; PQLEMELTPVVQTIE; PQNAVVKIYCPACHN; PQSAPHGVVFLHVTY; PQVNGLTSIKWADNNC; PRDLGACIDCSARHIN; PREGVFVSNGTHWFV; PRGQGVPINTNSSPD; PRGQGVPINTNSSPDD; PRITFGGPSDSTGSN; PRITFGGPSDSTGSNQ; PRPPLNRNYVFTGYR; PRQKRTATKAYNVTQ; PRQKRTATKAYNVTQA; PRRNVATLQAENVTGL; PRTLLTKGTLEPEY; PRTLLTKGTLEPEYF; PRTLLTKGTLEPEYFN; PRVEKKKLDGFMGRI; PRVEKKKLDGFMGRIR; PRWYFYYLGTGPEAG; PSARIVYTACSHAAVD; PSASAFFGMSRIGMEV; PSFLGRYMSALNHTK; PSGTWLTYTGAIKLD; PSIISNEKQEILGTVS; PSRILGAGCFVDDIV; PSVEQRKQDDKKIKAC; PSYAAFATAQEAYEQA; PTGTLLVDSDLNDFV; PTGTLLVDSDLNDFVS; PTGVHAGTDLEGNFYG; PTGYVDTPNNTDFSR; PTITQMNLKYAISAKN; PTLNISDEFSSNVAN; PTLNISDEFSSNVANY; PTQTVDSSQGSEYDYV; PTTCANDPVGFTLKNT; PTWRVYSTGSNV; PTWRVYSTGSNVFQT; PTWRVYSTGSNVFQTRAGCLIG; PVAIHADQLTPT; PVAIHADQLTPTWRV; PVAIHADQLTPTWRVY; PVASPNECNQMCLSTL; PVAYRKVLLRKNGNKG; PVETSNSFDVLKSED; PVETSNSFDVLKSEDA; PVSEEVVENPTIQKD; PVSEEVVENPTIQKDV; PVVDSYYSLLMPILT; PVVQTIEVNSFSGYL; PWDTIANYAKPFLNKV; PYEDFQENWNTK; PYEDFQENWNTKHSSG; PYGANKDGIIWV; PYGANKDGIIWVATE; PYGANKDGIIWVATEG; PYIKWDLLKYDFTEE; PYIKWDLLKYDFTEER; PYNSQNAVASKILGLP; PYPDPSRILGAGCFV; PYPNASFDNFKFVCDN; PYRVVVLSFELLHAPATVCGPKKSTNLVKN; QADVEWKFYDAQPCSD; QAFGRRGPEQTQGNFG; QAIASEFSSLPSYAAF; QAITVTPEANMDQESF; QALLKTVQFCDAMRN; QALVSDVGDSAEVAV; QAMTQMYKQARSEDKR; QATKYLVQQESPFVM; QATKYLVQQESPFVMM; QAVTANVNALLSTDG; QCGHYKHITSKETLY; QCGHYKHITSKETLYC; QCVQLHNDILLAKDTT; QDALFAYTKRNVIPT; QDALFAYTKRNVIPTI; QDAYYRARAGEAANF; QDAYYRARAGEAANFC; QDGNAAISDYDYYRY; QDGNAAISDYDYYRYN; QDLFLPFFSNVTWFH; QDLNGNWYDFGDFIQT; QDVNLHSSRLSFKELL; QEHYVRITGLYPTLNI; QEILGTVSWNLREMLA; QEKNFTTAPAICHDG; QEKNFTTAPAICHDGK; QELIRQGTDYKHWPQ; QENWNTKHSSGVTRE; QENWNTKHSSGVTREL; QEPKLGSLVVRCSFY; QESFGGASCCLYCRC; QESPFVMMSAPPAQY; QFAPSASAFFGMSRI; QFCDAMRNAGIVGVLT; QFGRDIADTTDA; QFGRDIADTTDAVRDP; QFGRDIADTTDAVRDPQTLE; QFKHLIPLMYKGLPW; QFKRPINPTDQSSYI; QFNSAIGKIQDSLSST; QFNSAIGKIQDSLSSTASAL; QFTSLEIPRRNVATLQ; QGAVDINKLCEEMLD; QGAVDINKLCEEMLDN; QGDDYVYLPYPDPSR; QGLNGYTVEEAKTVLK; QGLPNNTASWFTALTQ; QGLPNNTASWFTALTQ; QGMDNLACEDLK; QGNFGDQELIRQGTD; QGNFGDQELIRQGTDY; QGNFGDQELIRQGTDYK; QGTDYKHWPQIAQFAP; QGTTLPKGFYAEGSR; QGTTLPKGFYAEGSRG; QGTTLPKGFYAEGSRGGS; QGVPINTNSSPD; QGVPINTNSSPDDQI; QGVPINTNSSPDDQIG; QHLKDGTCGLVEVEK; QHLKDGTCGLVEVEKG; QHRLYECLYRNRDVD; QHTMLVKQGDDYVY; QHTMLVKQGDDYVYLP; QIDGYVMHANYIFWR; QIDRLITGRLQSLQTYVTQQLIRAA; QIELKFNPPALQDAY; QIGEYTFEKGDYGDA; QIGEYTFEKGDYGDAV; QIGGYTEKWESGVKDC; QIITTDNTFVSGNCD; QIITTDNTFVSGNCDV; QIPCTCGKQATKYLVQ; QIPFAMQMAYRF; QIPFAMQMAYRFNGI; QIPFAMQMAYRFNGIG; QIPTTCANDPVGFTLK; QIRSAAKKNNLPFKLT; QKDVLECNVKTTEVV; QKIAEIPKEEVKPFIT; QKLIANQFNSAIGKIQ; QKLLKSIAATRGATV; QKLLKSIAATRGATVV; QKRTATKAYNVTQAFG; QKSILSPLYAFASEAA; QKVGMQKYSTLQGPPG; QLGIEFLKRGDKSVYY; QLIKVTLVFLFVAAI; QLIRAAEIRASANLA; QLIRAAEIRASANLAA; QLLFVVEVVDKYFDC; QLLFVVEVVDKYFDCYDGGC; QLPAPRTLLTKGTLE; QLPAPRTLLTKGTLEP; QLPQGTTLPKGFYAE; QLPQGTTLPKGFYAEG; QLPQGTTLPKGFYAEGSRG; QLPQGTTLPKGFYAEGSRGG; QLSEISMDNSPNLAW; QLSSYSLFDMSKFPL; QLSSYSLFDMSKFPLK; QLTPTWRVYSTGSNV; QLTPTWRVYSTGSNVF; QLTPTWRVYSTGSNVFQTRA; QMAYRFNGIGVT; QMAYRFNGIGVTQNV; QMAYRFNGIGVTQNVL; QMAYRFNGIGVTQNVLYENQKLIANQFNSA; QMCLSTLMKCDHCGE; QMLSDTLKNLSDRVVF; QMNLKYAISAKNRART; QNAVASKILGLPTQTV; QNNVFMSEAKCWTETD; QNVLYENQKLIANQF; QNVLYENQKLIANQFN; QNYGDSATLPKGIMM; QNYGDSATLPKGIMMN; QPCSDKAYKIEELFYS; QPELDSFKEELD; QPELDSFKEELDKYF; QPELDSFKEELDKYFK; QPELDSFKEELDKYFKN; QPELDSFKEELDKYFKNH; QPELDSFKEELDKYFKNHTSP; QPGQTFSVLACYNGS; QPGQTFSVLACYNGSP; QPIDLVPNQPYPNAS; QPIDLVPNQPYPNASF; QPITNCVKMLCTHTGT; QPRVEKKKLDGFMGRI; QPYPNASFDNFKFVCD; QPYVFIKRSDARTAP; QQFGRDIADTTD; QQFGRDIADTTDAVRD; QQFGRDIADTTDAVRDPQ; QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGP; QQKLALGGSVAIKIT; QQQTTLKGVEAVMYMG; QQTTLKGVEAVMYMGT; QQVVDADSKIVQLSEI; QRKLEKMADQAMTQM; QRKLEKMADQAMTQMY; QRKQDDKKIKACVEE; QRMLLEKCDLQNYG; QRMLLEKCDLQNYGD; QRNFEYPQIITT; QRNFEYPQIITTDNT; QRNFEYPQIITTDNTF; QRNFEYPQIITTDNTFVSGN; QRRPQGLPNNTASWFT; QRVAGDSGFAAYSRY; QRVAGDSGFAAYSRYR; QSAPHGVVFLHV; QSAPHGVVFLHVTYV; QSDIEVTGDSCNNYM; QSLLIVNNATNVVIK; QSLQTYVTQQLIRAAEIRAS; QSSQAWQPGVAMPNL; QSTQLGIEFLKRGDKS; QTAQAAGTDTTITVN; QTLEILDITPCSFGG; QTPFEIKLAKKFDTF; QTPFEIKLAKKFDTFN; QTQGNFGDQELIRQGT; QTQGNFGDQELIRQGTDYKH; QTQGNFGDQELIRQGTDYKHW; QTTETAHSCNVNRFNV; QTTPGSGVPVVDSYY; QTTPGSGVPVVDSYYS; QTVKPGNFNKDFYDF; QTVKPGNFNKDFYDFA; QTVTHPNCVNCLDDR; QTYHPNCVNCLDDRCI; QTYVTQQLIRAAEIR; QVAVLYQGVNCTEVP; QVEQKIAEIPKEEVKP; QVILLNKHIDAYKTF; QVIVNNLDKSAGFPFN; QVNGLTSIKWADNNC; QVNGLTSIKWADNNCY; QVNGYPNMFITREEA; QVRDVLVRGFGDSVE; QVTCGTTTLNGLWLD; QVTIAEILLIIMRTF; QVVDADSKIVQLSEI; QWNLVIGFLFLTWIC; QYGDCLGDIAARDLICAQKF; QYGRSGETLGVLVPH; QYIDIGNYTVSCLPF; QYTSALLAGTITSGWTFGAG; RAKHYVYIGDPAQLPA; RALTAESHVDTDLTKP; RAMPNMLRIMASLVL; RANNTKGSLPINVIV; RANNTKGSLPINVIVF; RARVECFDKFKVNSTL; RASANLAATKMSECV; RASANLAATKMSECVL; RASANLAATKMSECVLGQSK; RATATIPIQASLPFGW; RATRRIRGGDGKMKDL; RCDIKDLPKEITVATS; RCHIDHPNPKGFCDLK; RCLWSTKPVETSNSFD; RCSFYEDFLEYHDVR; RDIADTTDAVRDPQT; RDIADTTDAVRDPQTL; RDIADTTDAVRDPQTLEILDITPCS; RDLGACIDCSARHIN; RDLSLQFKRPINPTD; RDLSLQFKRPINPTDQ; RDLYDKLQFTSLEIPR; RDPQTLEILDIT; RDPQTLEILDITPCS; RDPQTLEILDITPCSF; RDVDTDFVNEFYAYLR; RDVLVRGFGDSVEEV; REAACCHLAKALNDFS; REGVFVSNGTHW; REHEHEIAWYTERSE; RELGVVHNQDVNLHS; RELKVTFFPDLNGDVV; REQIDGYVMHANYIF; RFNGIGVTQNVL; RFNGIGVTQNVLYENQ; RFQEKDEDDNLIDSYF; RFTTTLNDFNLVAMKY; RGFGDSVEEVLSEARQ; RGGDGKMKDLSPRWVF; RGPEQTQGNFG; RGPEQTQGNFGDQEL; RGPEQTQGNFGDQELI; RGQGVPINTNSSPDD; RGQGVPINTNSSPDDQ; RGTAVMSLKEGQINDM; RGTTTYKLNVGDYF; RGTTVLLKEPCSSGTY; RGVYCCREHEHEIAW; RGVYYPDKVFRSSVL; RGVYYPDKVFRSSVLHSTQD; RGVYYPDKVFRSSVLHSTQDLFLPFFSNVT; RHFDEGNCDTLKEIL; RHTFSNYQHEETIYN; RHVRAWIGFDVEGCH; RIAGHHLGRCDIKDLP; RIFTIGTVTLKQGEIK; RIGMEVTPSGTWLTY; RIGNYKLNTDHSSSSD; RIRGGDGKMKDL; RIRGGDGKMKDLSPRW; RIRSVYPVASPNECNQ; RITFGGPSDSTGSNQN; RITFGGPSDSTGSNQNG; RITGLYPTLNISDEF; RITGLYPTLNISDEFS; RKHTTCCSLSHRFYR; RKLEKMADQAMTQMYK; RKLHDELTGHMLDMYS; RKNGNKGAGGHSYGAD; RKQIRSAAKKNNLPF; RKQIRSAAKKNNLPFK; RKVLLRKNGNKGAGGH; RKVPTDNYITTYPGQG; RLANECAQVLSEMVM; RLDKVEAEVQIDRLI; RLFARTRSMWSFNPE; RLFARTRSMWSFNPET; RLISMMGFKMNYQVN; RLITGRLQSLQTYVTQQLIRAA; RLQAGNATEVPANSTV; RLQSLQTYVTQQLIR; RLRAKHYVYIGDPAQL; RLSFKELLVYAADPAM; RLYECLYRNRDVDTD; RMNSRNYIAQVD; RMNSRNYIAQVDVVN; RMNSRNYIAQVDVVNF; RNFYEPQIITTDNTF; RNFYEPQIITTDNTFVSGNC; RNFYEPQIITTDNTFVSGNCDV; RNLQHRLYECLYRNRD; RNPANNAAIVLQLPQG; RNPAWRKAVFISPYN; RNPAWRKAVFISPYNS; RNRDVDTDFVNEFYAY; RNTNPIQLSSYSLFD; RNYVFTGYRVTKNSK; RPDTRYVLMDGSIIQF; RPLLESELVIGAVIL; RPQGLPNNTASWFTA; RPQGLPNNTASWFTAL; RPQGLPNNTASWFTALTQHG; RPQGLPNNTASWFTALTQHGK; RQFHQKLLKSIAATR; RQGFVDSDVETKDVV; RQKRTATKAYNVTQAF; RQVVNVVTTKIALKG; RQWLPTGTLLVDSDL; RRATRRIRGGDGKMK; RRCPAEIVDTVSALVY; RRGPEQTQGNFGDQE; RRGPEQTQGNFGDQEL; RRGPEQTQGNFGDQELIRQGTDYK; RRIRGGDGKMKDLSP; RRIRGGDGKMKDLSPR; RRNVATLQAENVTGL; RRNVATLQAENVTGLF; RRPQGLPNNTASWFT; RRPQGLPNNTASWFTA; RRSFYVVYANGGKGFC; RSAAKKNNLPFKLTCA; RSEKSYELQTPFEIKL; RSGARSKQRRPQGLPN; RSKQRRPQGLPNNT; RSKQRRPQGLPNNTA; RSKQRRPQGLPNNTAS; RSMWSFNPETNILLN; RSMWSFNPETNILLNV; RSMWSFNPETNILLNVPLHG; RSSVLHSTQDLFLPFFSNVT; RSVASQSIIAYTMSLGAENSVAYSN; RSVYPVASPNECNQMC; RTATKAYNVTQAFGR; RTATKAYNVTQAFGRR; RTLLTKGTLEPEYFNS; RTRSMWSFNPETNIL; RTTCFSVAALTNNVA; RTVYDDGARRVWTLMN; RVAGDSGFAAYSRYRI; RVDWTIEYPIIGDEL; RVDWTIEYPIIGDELK; RVEAFEYYHTTDPSF; RVECFDKFKVNSTLEQ; RVECTTIVNGVRRSF; RVIHFGAGSDKGVAP; RVLGLKTLATHGLAAV; RVLNVVCKTCGQQQTT; RVLSNLNLPGCDGGSL; RVSAKPPPGDQFKHL; RVTKNSKVQIGEYTFE; RVVFNGVSFSTFEEAA; RVVRSIFSRTLETAQ; RVVTTFDSEYCRHGT; RVVTTFDSEYCRHGTC; RVYANLGERVRQALLK; RVYSTGSNVFQTRAG; RVYSTGSNVFQTRAGC; RVYSTGSNVFQTRAGCLIGA; RWFLNRFTTTLNDFN; RWFLNRFTTTLNDFNL; RYFKYWDQTYHPNCVN; RYRIGNYKLNTDHSSS; RYRIGNYKLNTDHSSSSDNIA; RYVDNNFCGPDGYPL; RYVDNNFCGPDGYPLE; SAAEASKKPRQKRTA; SAAEASKKPRQKRTAT; SAAKKNNLPFKLTCA; SAAKKNNLPFKLTCAT; SAEVAVKMFDAYVNTF; SAFFGMSRIGMEVTPS; SAFFGMSRIGMEVTPSGTWL; SAIGKIQDSLSSTAS; SAIGKIQDSLSSTASA; SAIGKIQDSLSSTASALGKLQD; SAINRPQIGVVREFL; SAKPPPGDQFKHLIP; SALWEIQQVVDADSKI; SAMVRMYIFFASFYY; SANLAATKMSECVLGQ; SAPTLVPQEHYVRIT; SAPTLVPQEHYVRITG; SARIVYTACSHAAVDA; SASAFFGMSRIGMEV; SASAFFGMSRIGMEVT; SASTSAFVETVKGLDY; SASVYYSQLMCQPIL; SATLVSDIDITFLKKD; SAVGNICYTPSKLIE; SCCKFDEDDSEPVLK; SCCKFDEDDSEPVLKG; SCCKFDEDDSEPVLKGVKL; SCCKFDEDDSEPVLKGVKLHYT; SCLPFTINCQEPKLG; SCNNYMLTYNKVENMT; SDADSTLIGDCATVHT; SDIDITFLKKDAPYI; SDIDITFLKKDAPYIV; SDIEVTGDSCNNYML; SDIEVTGDSCNNYMLT; SDKAYKIEELFYSYA; SDLNDFVSDADSTLI; SDLNDFVSDADSTLIG; SDLQDLKWARFPKSD; SDLQDLKWARFPKSDG; SDMYDPKTKNVTKEND; SDTLKNLSDRVVFVL; SDVENPHLMGWDYPKC; SDVETKDVVECLKLS; SDVGDSAEVAVKMFD; SDYDYYRYNLPTMCDI; SEAARVVRSIFSRTL; SEAKCWTETDLTKGPH; SEARQHLKDGTCGLV; SEDAQGMDNLACEDL; SEDKRAKVTSAMQTM; SEEVVENPTIQKDVL; SEEVVENPTIQKDVLE; SEEVVENPTIQKDVLECN; SEFSSLPSYAAFATAQ; SEGLNDNLLEILQKEK; SEHDYQIGGYTEKWES; SEISMDNSPNLAWPL; SEKQVEQKIAEIPKEE; SEKSYELQTPFEIKLA; SELVIGAVILRGHLR; SEPVLKGVKLHY; SEPVLKGVKLHYT; SEQLRKQIRSAAKKN; SEQLRKQIRSAAKKNN; SEVGPEHSLAEYHNE; SEYDYVIFTQTTETAH; SEYKGPITDVFYKENS; SFCAFAVDAAKAYKDY; SFCYMHHMELPTGVH; SFDLGDELGTDPYED; SFDNFKFVCDNIKFAD; SFDVLKSEDAQGMDN; SFDVLKSEDAQGMDNL; SFELLHAPATVCGPKKSTNL; SFGGASCCLYCRCHI; SFGGVSVITPGTNTSN; SFKEELDKYFKN; SFKEELDKYFKNHT; SFKEELDKYFKNHTS; SFKEELDKYFKNHTSP; SFKELLVYAADPAMHA; SFKKGAKLLHKPIVW; SFLEMKSEKQVEQKI; SFLEMKSEKQVEQKIA; SFLNGSCGSVGFNID; SFLNGSCGSVGFNIDY; SFMLWCKDGHVETFY; SFNPETNILLNVPLH; SFNPETNILLNVPLHG; SFNYLKSPNFSKLIN; SFPQSAPHGVVFLHV; SFPQSAPHGVVFLHVTYVPA; SFPQSAPHGVVFLHVTYVPAQE; SFRLFARTRSMWSFN; SFSGYLKLTDNVYIKN; SFTRGVYYPDKV; SFTRGVYYPDKVFRS; SFTRGVYYPDKVFRSSVLHS; SGARSKQRRPQGLPN; SGARSKQRRPQGLPNN; SGDGTTSPISEHDYQI; SGDGTTSPISEHDYQIG; SGFAAYSRYRIGNYKL; SGNCDVVIGIVN; SGNCDVVIGIVNNTV; SGTNGTKRFDNP; SGTNGTKRFDNPVLP; SGTNGTKRFDNPVLPFNDGVYF; SGTWLTYTGAIKLDDK; SGTYEGNSPFHPLADN; SGYLKLTDNVYIKNAD; SHEGKTFYVLPNDDTL; SHQSDIEVTGDSCNN; SHVDTDLTKPYIKWDL; SIAATRGATVVIGTS; SIAATRGATVVIGTSK; SIFSRTLETAQNSVR; SIFSRTLETAQNSVRV; SIHLYFDKAGQKTYE; SIIKTIQPRVEKKKLD; SIIQFPNTYLEGSVRV; SIISNEKQEILGTVS; SIKNFKSVLYYQNNV; SKCVCSVIDLLLDDF; SKFPLKLRGTAVMSL; SKGFFKEGSSVELKHF; SKIVQLSEISMDNSP; SKLIEYTDFATSACV; SKPSVEQRKQDDKKI; SKPSVEQRKQDDKKIK; SKQRRPQGLPNN; SKQRRPQGLPNNTAS; SKQRRPQGLPNNTASW; SKSLTENKYSQLDEEQ; SKSPIQYIDIGNYTVS; SKTPEEHFIETISLA; SKTQSLLIVNNATN; SKTQSLLIVNNATNV; SKTQSLLIVNNATNVV; SKVQIGEYTFEKGDYG; SLAGSYKDWSYSGQST; SLEIPRRNVATLQAEN; SLLLCRMNSRNY; SLLLCRMNSRNYIA; SLPSYAAFATAQEAYE; SLQTYVTQQLIRAAE; SLQTYVTQQLIRAAEI; SLRPDTRYVLMDGSII; SLSEQLRKQIRSAAK; SLTENKYSQLDEEQPM; SLYVKPGGTSSGDATT; SLYVKPGGTSSGDATTAYA; SMATNYDLSVVNARLR; SMMGFKMNYQVNGYPN; SMMGFKMNYQVNGYPNM; SMQGAVDINKLCEEML; SMSYEDQDALFAYTKR; SMWSFNPETNILLNVP; SNGTHWFVTQRN; SNGTHWFVTQRNFYE; SNKKFLPFQQFG; SNKKFLPFQQFGRDI; SNKKFLPFQQFGRDIA; SNKKFLPFQQFGRDIADTTD; SNLNLPGCDGGSLYV; SNLNLPGCDGGSLYVN; SNSFDVLKSEDAQGMD; SNVANYQKVGMQKYST; SNVFQTRAGCLIGAE; SPDDQIGYYRRA;

SPDDQIGYYRRATRR; SPDDQIGYYRRATRRI; SPDVDLGDISGI; SPDVDLGDISGINAS; SPDVDLGDISGINASV; SPDVDLGDISGINASVV; SPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLID; SPFVMMSAPPAQYEL; SPIQYIDIGNYTVSCL; SPISEHDYQIGGYTE; SPISEHDYQIGGYTEK; SPYNSQNAVASKILGL; SQAWQPGVAMPNLYK; SQGLVASIKNFKSV; SQHTMLVKQGDDYVY; SQHTMLVKQGDDYVYL; SQLMCQPILLLDQAL; SQRVAGDSGFAAYSR; SQRVAGDSGFAAYSRY; SRELKVTFFPDLNGDV; SRIGMEVTPSGTWLTY; SRIKASMPTTIAKNT; SRLDKVEAEVQID; SRLDKVEAEVQIDRL; SRNYIAQVDVVN; SRVLGLKTLATHGLAA; SRYRIGNYKLNTDHSS; SSEYKGPITDVFYKE; SSEYKGPITDVFYKEN; SSGTYEGNSPFHPLA; SSGTYEGNSPFHPLAD; SSKLWAQCVQLHNDI; SSKLWAQCVQLHNDIL; SSKTPEEHFIETISLA; SSLPSYAAFATAQEAY; SSNVANYQKVGMQKYS; SSPDDQIGYYRRATRR; SSSKTPEEHFIETIS; SSSKTPEEHFIETISL; SSSSDNIALLVQ; SSTCMMCYKRNRATR; SSTFNVPMEKLKTLVA; SSVLHSTQDLFLPFF; SSYSLFDMSKFPLKLR; STDGNKIADKYVRNLQ; STFISDEVARDLSLQ; STGSNVFQTRAGCLI; STKPVETSNSFDVLKS; STLEQYVFCTVNALP; STLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYV; STLMKCDHCGETSWQ; STLQGPPGTGKSHF; STNLVKNKCVNFNFNG; STNSRIKASMPTTIAK; STQLGIEFLKRGDKSV; STSAFVETVKGLDYKA; SVAIKITEHSWNADL; SVAYESLRPDTRYVL; SVAYESLRPDTRYVLM; SVCRLMKTIGPDMFL; SVEEVLSEARQHLKDG; SVELKHFFFAQDGNA; SVELKHFFFAQDGNAA; SVEQRKQDDKKIKACV; SVFNICQAVTANVNA; SVGFNIDYDCVSFCY; SVIDLLLDDFVEIIKS; SVITPGTNTSNQVAVL; SVIYLYLTFYLTNDV; SVLACYNGSPSGVYQ; SVLHSTQDLFLPFFS; SVMLTNDNTSRYWEP; SVMLTNDNTSRYWEPE; SVPWDTIANYAKPFLN; SVRVVTTFDSEYCRH; SVRVVTTFDSEYCRHG; SVSSPDAVTAYNGYL; SVYPVASPNECNQMC; SVYPVASPNECNQMCL; SVYYSQLMCQPILLL; SVYYTSNPTTFHLDG; SWLMWLIINLVQMAP; SWNADLYKLMGHFAWW; SWNLREMLAHAEETR; SWNLREMLAHAEETRK; SWVMRIMTWLDMVDT; SYAAFATAQEAYEQAV; SYECDIPIGAGICASY; SYEDQDALFAYTKRNV; SYELQTPFEIKLAK; SYELQTPFEIKLAKKF; SYEQFKKGVQIPCTCG; SYFIASFRLFARTRS; SYFTEQPIDLVPNQPY; SYFVVKRHTFSNYQH; SYGADLKSFDLGDELG; SYIVDSVTVKNGSIH; SYKDWSYSGQSTQLG; SYREAACCHLAKALND; SYSLFDMSKFPLKLRG; SYTTTIKPVTYKLDGV; SYVHVVDGCNSSTCM; SYYKLGASQRVAGDS; SYYKLGASQRVAGDSG; TAAKLMVVIPDYNTY; TAEAELAKNVSLDNVL; TAESHVDTDLTKPYI; TAESHVDTDLTKPYIK; TAFVTNVNASSSEAF; TAHSCNVNRFNVAIT; TANKWDLIISDMYDP; TANVNALLSTDGNKIA; TAPHGHVMVELVAEL; TASWFTALTQHGKED; TATKAYNVTQAFGRRG; TAYNGYLTSSSKTP; TAYNGYLTSSSKTPE; TAYNGYLTSSSKTPEE; TCANDPVGFTLKNTVC; TCASEYTGNYQCGHY; TCATTRQVVNVVTTKI; TCCSLSHRFYRLANE; TCFSVAALTNNVAFQ; TCGKQATKYLVQQESP; TCGQQQTTLKGVEAV; TCGQQQTTLKGVEAVM; TCGTTTLNGLWLDDV; TCMMCYKRNRATRVE; TCRRCPAEIVDTVSA; TCRRCPAEIVDTVSAL; TDAVRDPQTLEI; TDAVRDPQTLEILDI; TDAVRDPQTLEILDIT; TDAVRDPQTLEILDITPCS; TDEMIAQYTSALLAGTITSGW; TDFATSACVLAAECT; TDFSRVSAKPPPGDQ; TDFSRVSAKPPPGDQF; TDFVNEFYAYLRKHFS; TDGNKIADKYVRNLQ; TDGTLMIERFVSLAI; TDHSSSSDNIALLVQ; TDLEGNFYGPFVDRQ; TDLEGNFYGPFVDRQT; TDLMAAYVDNSSLTIK; TDLTKPYIKWDLLKY; TDNTFVSGNCDVVIG; TDNTFVSGNCDVVIGI; TDNVYIKNADIVEEAK; TDNYITTYPGQGLNG; TDPSFLGRYMSALNH; TDPYEDFQENWNTKHS; TDQSSYIVDSVTVKN; TDVFYKENSYTTTIK; TDVFYKENSYTTTIKP; TDYKHWPQIAQFAPS; TEERLKLFDRYFKYW; TEGALNTPKDHIGTRN; TEHSWNADLYKLMGHF; TEIDPKLDNYYKKDN; TEIDPKLDNYYKKDNS; TEKSNIIRGWIF; TELEPPCRFVTDTPKG; TEMLAKALRKVPTDN; TENKYSQLDEEQPMEI; TENLTKEGATTCGYL; TEQPIDLVPNQPYPN; TEQPIDLVPNQPYPNA; TERSEKSYELQTPFE; TERSEKSYELQTPFEI; TESKPSVEQRKQDDKK; TESNKKFLPFQQ; TESNKKFLPFQQFGR; TESNKKFLPFQQFGRD; TESNKKFLPFQQFGRDIA; TESNKKFLPFQQFGRDIAD; TESNKKFLPFQQFGRDIADTTDAVRD; TEVPVAIHADQLTPTW; TEVPVAIHADQLTPTWRVYST; TEVVGDIILKPANNSL; TFDSEYCRHGTCER; TFDSEYCRHGTCERS; TFDSEYCRHGTCERSE; TFEEAALCTFLLNKE; TFEKGDYGDAVVYRGT; TFFPDLNGDVVAIDY; TFFPDLNGDVVAIDYK; TFGGPSDSTGSN; TFGGPSDSTGSNQNG; TFGGPSDSTGSNQNGE; TFHLDGEVITFDNLKT; TFISAARQGFVDSDV; TFLLNKEMYLKLRSDV; TFVSGNCDVVIGIVN; TFVTHSKGLYRKCVK; TFYPKLQSSQAWQPGV; TGAIKLDDKDPN; TGAIKLDDKDPNFKD; TGAIKLDDKDPNFKDQ; TGHMLDMYSVMLTND; TGLYPTLNISDEFSS; TGLYPTLNISDEFSSN; TGNTLQCIMLVYCFL; TGPEAGLPYGANKDG; TGPEAGLPYGANKDGI; TGQAITVTPEANMDQE; TGTGQAITVTPEANMD; TGTGVLTESNKKFLP; TGTGVLTESNKKFLPF; TGTGVLTESNKKFLPFQQFGRD; TGTGVLTESNKKFLPFQQFGRDIA; TGTIYTELEPPCRFV; TGTLLVDSDLNDFVSD; TGVHAGTDLEGNFYG; TGVHAGTDLEGNFYGP; TGVLTESNKKFLPFQ; TGVLTESNKKFLPFQQ; TGVLTESNKKFLPFQQFG; TGVLTESNKKFLPFQQFGRD; TGYHFRELGVVHNQD; TGYRVTKNSKVQIGEY; TGYVDTPNNTDFSRVS; THGLAAVNSVPWDTIA; THLSVDTKFKTEGLC; THWFVTQRNFYE; TIANYAKPFLNKVVST; TIANYAKPFLNKVVSTTT; TIEYPIIGDELKINAA; TIGTVTLKQGEIKDAT; TILTRPLLESELVIG; TIPIQASLPFGWLIVGVALLAVFQSASKIITL; TIQKDVLECNVKTTE; TIQPRVEKKKLDGFMG; TITQMNLKYAISAKN; TITQMNLKYAISAKNR; TITSGWTFGAGAALQ; TIVEVQPQLEMELTP; TIYNLLKDCPAVAKHD; TIYSLLLCRMNSRNY; TIYTELEPPCRFVTD; TKAYNVTQAFGRRGP; TKAYNVTQAFGRRGPE; TKDVVECLKLSHQSD; TKGSLPINVIVFDGK; TKGTLEPEYFNSVCRL; TKHFYWFFSNYLKRR; TKIALKGGKIVNNWL; TKKSAAEASKKPRQKR; TKMSECVLGQSK; TKMSECVLGQSKRVD; TKNSKVQIGEYTFEKG; TKPVETSNSFDVLKS; TKPVETSNSFDVLKSE; TKPYIKWDLLKYDFT; TKRFDNPVLPFNDGV; TKRFDNPVLPFNDGVY; TKRGVYCCREHEHEI; TKRNVIPTITQMNLKY; TKSSEYKGPITDVFY; TKSSEYKGPITDVFYK; TKYLVQQESPFVMMS; TKYLVQQESPFVMMSA; TKYTMADLVYALRHF; TLATHGLAAVNSVPW; TLDSKTQSLLIVNNA; TLDSKTQSLLIVNNAT; TLDSKTQSLLIVNNATNVVIKVCEFQFCND; TLEILDITPCSF; TLEILDITPCSFGGV; TLETAQNSVRVLQKAA; TLIGDCATVHTANKW; TLIGDCATVHTANKWD; TLKGVEAVMYMGTLSY; TLLTKGTLEPEYFNSV; TLLVDSDLNDFVSDAD; TLNISDEFSSNVANYQ; TLQQIELKFNPPALQD; TLSYEQFKKGVQIPCT; TLSYYKLGASQRVAG; TLSYYKLGASQRVAGD; TLSYYKLGASQRVAGDSGFA; TLTLAVPYNMRVIHF; TLTRALTAESHVDTD; TLVATAEAELAKNVSL; TLVSDIDITFLKKDAP; TLYCIDGALLTKSSE; TMCDIRQLLFVVEVV; TMLVKQGDDYVYLPYP; TNCVKMLCTHTGTGQA; TNDNTSRYWEPEFYE; TNGTKRFDNPVL; TNGTKRFDNPVLPFN; TNGTKRFDNPVLPFND; TNIVTRCLNRVCTNY; TNPIQLSSYSLFDMS; TNPIQLSSYSLFDMSK; TNSRIKASMPTTIAKN; TNSSPDDQIGYYRRA; TNVLEGSVAYESLRP; TNVLEGSVAYESLRPD; TNVVIKVCEFQFCND; TNVVIKVCEFQFCNDP; TPEEHFIETISLAGS; TPFEIKLAKKFDTFNG; TPGSGVPVVDSYYSL; TPGTNTSNQVAVLYQ; TPGTNTSNQVAVLYQD; TPKDHIGTRNPANNA; TPKDHIGTRNPANNAA; TPKDHIGTRNPANNAAIVLQ; TPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGG; TPNNTDFSRVSAKPP; TPNNTDFSRVSAKPPP; TPRDLGACIDCSARH; TPRDLGACIDCSARHI; TPSGTWLTYTGAIKL; TPSKLIEYTDFATSA; TPSKLIEYTDFATSAC; TPTWRVYSTGSNVFQT; TPVVQTIEVNSFSGYL; TQAFGRRGPEQTQGN; TQAFGRRGPEQTQGNF; TQAFGRRGPEQTQGNFGDQE; TQDLFLPFFSNVTWF; TQGNFGDQELIRQGTD; TQLGIEFLKRGDKSVY; TQMNLKYAISAKNRA; TQMNLKYAISAKNRAR; TQMYKQARSEDKRAKV; TQNVLYENQKLIANQF; TQQLIRAAEIRASAN; TQQLIRAAEIRASANLAATK; TQQLIRAAEIRASANLAATKMS; TQRNFYEPQIITTDN; TQRNFYEPQIITTDNT; TQSLLIVNNATNVVIK; TQTTETAHSCNVNRFN; TQTVDSSQGSEYDYV; TRAKVGILCIMSDRD; TRALTAESHVDTDLT; TRCLNRVCTNYMPYF; TREAVGTNLPLQLGFS; TREEAIRHVRAWIGF; TRNPANNAAIVLQLP; TRNPANNAAIVLQLPQ; TRNPANNAAIVLQLPQGTTLPKGFYAEGSR; TRPLLESELVIGAVI; TRRIRGGDGKMKDLS; TRRIRGGDGKMKDLSP; TRRIRGGDGKMKDLSPRWYF; TRSMWSFNPETNILL; TRSTNSRIKASMPTTI; TRVLSNLNLPGCDGGS; TSALLAGTITSGWTF; TSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNV; TSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKL; TSEDMLNPNYEDLLI; TSFGPLVRKIFVDGV; TSGDGTTSPISEHDYQ; TSGWTFGAGAALQIP; TSGWTFGAGAALQIPFAMQMAY; TSHTVMPLSAPTLVP; TSLEIPRRNVATLQAE; TSNPTTFHLDGEVIT; TSNSFDVLKSEDAQG; TSNSFDVLKSEDAQGM; TSPDVDLGDISG; TSPDVDLGDISGINAS; TSPISEHDYQIGGYTE; TSRTLSYYKLGASQR; TSRYWEPEFYEAMYT; TSSGDATTAYANSVF; TSSGDATTAYANSVFN; TSSSKTPEEHFIETIS; TTADIVVFDEISMAT; TTAPAICHDGKA; TTAPAICHDGKAHFPREGVF; TTCANDPVGFTLKNTV; TTDAVRDPQTLE; TTDAVRDPQTLEILDI; TTDNTFVSGNCDVVI; TTDNTFVSGNCDVVIG; TTEMLAKALRKVPTDN; TTETAHSCNVNRFNVA; TTEVVGDIILKPANN; TTEVVGDIILKPANNS; TTFDSEYCRHGTCERS; TTFHLDGEVITFDNL; TTFHLDGEVITFDNLK; TTKGGRFVLALLSDL; TTLKGVEAVMYMGTL; TTNGDFLHFLPRVFS; TTNIVTRCLNRVCTN; TTQTACTDDNALAYY; TTRQVVNVVTTKIAL; TTSPISEHDYQI; TTSPISEHDYQIGGY; TTSPISEHDYQIGGYT; TTTIKPVTYKLDGVV; TTTLNDFNLVAMKYNY; TVATSRTLSYYKLGA; TVAYFNMVYMPASWV; TVCGPKKSTNLV; TVCGPKKSTNLVKNKC; TVDSSQGSEYDYVIF; TVDSSQGSEYDYVIFT; TVGELGDVRETMSYL; TVKGLDYKAFKQIVE; TVKPGNFNKDFYDFAV; TVLLKEPCSSGTYEGN; TVLPPLLTDEMIAQY; TVMPLSAPTLVPQEHY; TVSVSSPDAVTAYNGY; TVSWNLREMLAHAEE; TVTKKSAAEASKKPRQKRTATKAYNVTQAF; TVTPEANMDQESFGGA; TVVVNAANVYLKHGGG; TVYDPLQPELDS; TVYDPLQPELDSFK; TVYDPLQPELDSFKEELDKY; TVYSDVENPHLMGWD; TWLTYTGAIKLDDKDP; TWRVYSTGSNVF; TYASALWEIQQVVDA; TYERHSLSHFVNLDN; TYHPNCVNCLDDRCIL; TYKLDGVVCTEIDPKL; TYNCCDDDYFNKKDW; TYNKVENMTPRDLGAC; TYPGQGLNGYTVEEA; TYPGQGLNGYTVEEAK; TYTGAIKLDDKD; TYTGAIKLDDKDPNF; TYTGAIKLDDKDPNFK; TYVPAQEKNFTTAPA; TYVPAQEKNFTTAPAI; TYVPAQEKNFTTAPAICHDG; TYVTQQLIRAAEIRA; TYVTQQLIRAAEIRASANLA; VAFNTLLFLMSFTVL; VAGALNKATNNAMQVE; VAGDSGFAAYSRYRI; VAGDSGFAAYSRYRIG; VAIDYKHYTPSFKKGA; VAIHADQLTPTW; VAIHADQLTPTWRVY; VAIHADQLTPTWRVYS; VAIKITEHSWNADLYK; VAKHDFFKFRIDGDM; VAKHDFFKFRIDGDMV; VANYQKVGMQKYSTLQ; VARDLSLQFKRPINPT; VASPNECNQMCLSTL; VASQSIIAYTMSLGAENSVAYS; VATEGALNTPKDHIG; VATEGALNTPKDHIGT; VATEGALNTPKDHIGTRNPA; VATLQAENVTGLFKDC; VATVQSKMSDVKCTSV; VAVKMFDAYVNTFSS; VAVPTGYVDTPNNTDF; VAYESLRPDTRYVLMD; VAYRKVLLRKNGNKG; VAYRKVLLRKNGNKGA; VCDNIKFADDLNQLTG; VCGPKKSTNLVKN; VCGPKKSTNLVKNKCV; VCGPKKSTNLVKNKCVN; VCGPKKSTNLVKNKCVNFNF; VCKTCGQQQTTLKGV; VCKTCGQQQTTLKGVE; VCRLMKTIGPDMFLGT; VCTEIDPKLDNYYKKD; VDAAKAYKDYLASGGQ; VDADSKIVQLSEISM; VDALCEKALKYLPID; VDALCEKALKYLPIDK; VDDIVKTDGTLMIER; VDDIVKTDGTLMIERE; VDEAGSKSPIQY; VDEAGSKSPIQYIDI; VDEAGSKSPIQYIDIG; VDINKLCEEMLDNRAT; VDKYFDCYDGGCINAN; VDNNFCGPDGYPLEC; VDNNFCGPDGYPLECI; VDSDLNDFVSDADST; VDSDLNDFVSDADSTL; VDSDVETKDVVECLK; VDSSQGSEYDYVIFTQ; VDSYYSLLMPILTLT; VDTANPKTPKYKFVR; VDTDFVNEFYAYLRKH; VDTDLTKPYIKWDLLK; VDTPNNTDFSRVSAK; VDTPNNTDFSRVSAKP; VDWTIEYPIIGDELKI; VEAEVQIDRLITGRLQSLQT; VEAVMYMGTLSYEQFK; VECFDKFKVNSTLEQY; VECLKLSHQSDIEVT; VEEAKKVKPTVVVNA; VEEAKKVKPTVVVNAA; VEEVLSEARQHLKDGT; VEKGVLPQLEQPYVF; VEKGVLPQLEQPYVFI; VEKKKLDGFMGRIRS; VELKHFFFAQDGNAAI; VELVAELEGIQYGRS; VELVAELEGIQYGRSG; VENMTPRDLGACIDC; VENPDILRVYANLGE; VENPDILRVYANLGER; VENPHLMGWDYPKCDR; VENPTIQKDVLECNVK; VEQKIAEIPKEEVKPF; VEQRKQDDKKIKACV; VEQRKQDDKKIKACVE; VETFYPKLQSSQAWQP; VETSNSFDVLKSEDAQ; VETVKGLDYKAFKQI; VEVEKGVLPQLEQPY; VEVEKGVLPQLEQPYV; VEVQPQLEMELTPVV; VEVVDKYFDCYDGGCI; VEWKFYDAQPCSDKAY; VFCTVNALPETTADI; VFDEISMATNYDLSV; VFDEISMATNYDLSVV; VFHLYLQYIRKLHDE; VFISPYNSQNAVASK; VFMSEAKCWTETDLT; VFMSEAKCWTETDLTK; VFNGVSFSTFEEAA; VFPLNSIIKTIQPRV; VFRSSVLHSTQDLFLPFFSNVTWFH; VFVSNGTHWFVT; VFYKENSYTTTIKPVT; VGDFKLNEEIAIILA; VGDSAEVAVKMFDAY; VGELGDVRETMSYLF; VGGKPCIKVATVQSKM; VGHTDLMAAYVDNSS; VGLMWLSYFIASFRL; VGNICYTPSKLIEYTD; VGPEHSLAEYHNESG; VGVLTLDNQDLNGNW; VHAGTDLEGNFYGPF; VHAGTDLEGNFYGPFV; VHNQDVNLHSSRLSF; VHTANKWDLIISDMY; VHVVDGCNSSTCMMC; VIDLLLDDFVEIIKS; VIDLLLDDFVEIIKSQ; VIFTQTTETAHSCNV; VIFTQTTETAHSCNVN; VIGAVILRGHLRIAG; VIGFLFLTWICLLQF; VIGIVNNTVYDPLQP; VIGIVNNTVYDPLQPEL; VIKVCEFQFCND; VIKVCEFQFCNDPFL; VIPDYNTYKNTCDGTT; VITFDNLKTLLSLRE; VITHDVSSAINRPQIG; VITPGTNTSNQVAVL; VITPGTNTSNQVAVLY; VITSGDGTTSPISEHD; VIVNNLDKSAGFPFNK; VKAALLADKFPVLHDI; VKIGPERTCCLCDRR; VKIGPERTCCLCDRRA; VKMFDAYVNTFSST; VKMFDAYVNTFSSTF; VKMLCTHTGTGQAITV; VKNGSIHLYFDKAGQ; VKNKCVNFNFNGLT; VKPFITESKPSVEQRK; VKPGGTSSGDATTAYA; VKPGNFNKDFYDFAVS; VKQGDDYVYLPYPDP; VKQGDDYVYLPYPDPS; VKRHTFSNYQHEETI; VKRVDWTIEYPIIGDE; VKTTEVVGDIILKPA; VKVTIDYTEISFMLW; VLAAECTIFKDASGKP; VLEGSVAYESLRPDT; VLEGSVAYESLRPDTR; VLHSTQDLFLPFFSN; VLKSEDAQGMDNLACE; VLLKEPCSSGTYEGNS; VLLRKNGNKGAGGHS; VLLRKNGNKGAGGHSY; VLNVVCKTCGQQQTTL; VLPNDDTLRVEAFEY; VLPPLLTDEMIA; VLQLPQGTTLPKGFY; VLQLPQGTTLPKGFYA; VLQVRDVLVRGFGDS; VLSNLNLPGCDGGSL; VLSNLNLPGCDGGSLY; VLSTFISAARQGFVD; VLTESNKKFLPF; VLTESNKKFLPFQQF; VLTESNKKFLPFQQFG; VLTLDNQDLNGNWYD; VLTSHTVMPLSAPTL; VLVPHVGEIPVAYRK; VLVRGFGDSVEEVL; VLVRGFGDSVEEVLS; VLVRGFGDSVEEVLSE; VLYENQKLIANQFNSA; VMPLSAPTLVPQEHY; VMPLSAPTLVPQEHYV; VMYMGTLSYEQFKKGV; VNALLSTDGNKIADK; VNALLSTDGNKIADKY; VNALPETTADIVVFD; VNARLRAKHYVYIGDP; VNCTEVPVAIHADQL; VNCTEVPVAIHADQLT; VNGLTSIKWADNNCYL; VNKFLALCADSIIIG; VNLDNLRANNTKGSLP; VNNATNKATYKPNTWC; VNNATNVVIKVCEFQF; VNNSYECDIPIGAGI; VNNSYECDIPIGAGIC; VNNSYECDIPIGAGICA; VNNTVYDPLQPELDS; VNNTVYDPLQPELDSF; VNNTVYDPLQPELDSFKEELDKYFKNHT; VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI; VNSFSGYLKLTDNVYI; VNSVPWDTIANYAKPF; VNTFSSTFNVPMEKLK; VPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWF; VPFWITIAYIICIST; VPHVGEIPVAYRKVL; VPINTNSSPDDQIGY; VPINTNSSPDDQIGYY; VPMEKLKTLVATAEA; VPNQPYPNASFDNFKF; VPQADVEWKFYDAQPC; VPQEHYVRITGLYPTL; VPTDNYITTYPGQGLN; VPTGYVDTPNNTDFSR; VPVAIHADQLTPTWRV; VPWDTIANYAKPFLN; VPWDTIANYAKPFLNK; VQIDRLITGRLQ; VQIDRLITGRLQSLQ; VQIGEYTFEKGDYGDA; VQIPCTCGKQATKYLV; VQIPTTCANDPVGFT; VQIPTTCANDPVGFTL; VQLHNDILLAKDTTE; VQMAPISAMVRMYIF; VQMLSDTLKNLSDR; VQPQLEMELTPVVQT; VQQESPFVMMSAPPA; VQQESPFVMMSAPPAQ; VQTIEVNSFSGYLKL; VRATATIPIQAS; VRATATIPIQASLPFG; VRCSFYEDFLEY; VRCSFYEDFLEYHDV; VRDPQTLEILDI; VRDPQTLEILDITPC; VRDPQTLEILDITPCS; VRETMSYLFQHANLD; VRGFGDSVEEVLSEA; VRGFGDSVEEVLSEAR; VRGTTVLLKEPCSSGT; VRITGLYPTLNISDEF; VRLQAGNATEVPANS; VRNLQHRLYECLYRN; VRVVTTFDSEYCRHGT;

VSCLPFTINCQEPKL; VSDADSTLIGDCATV; VSDADSTLIGDCATVH; VSDIDITFLKKDAPYI; VSDVGDSAEVAVKMFD; VSEEVVENPTIQKDVL; VSFLAHIQWMVMFTP; VSFSTFEEAALCTFL; VSKVVKVTIDYTEIS; VSLDNVLSTFISAAR; VSNGTHWFVTQRNFY; VSNGTHWFVTQRNFYEPQII; VSSAINRPQIGVVRE; VSTGYHFRELG; VSTGYHFRELGVVHN; VSVITPGTNTSNQVA; VSVITPGTNTSNQVAV; VSVSSPDAVTAYNGYL; VTANVNALLSTDGNKI; VTAYNGYLTSSSKTPE; VTHSKGLYRKCVKSR; VTIAEILLIIMRTFKV; VTIDYTEISFMLWCKD; VTKGKAKKGAWNIGE; VTKKSAAEASKKPRQKRTATKA; VTKNSKVQIGEYTFEK; VTNVNASSSEAFLIG; VTNVNASSSEAFLIGC; VTPEANMDQESFGGAS; VTPSGTWLTYTGAIK; VTQAFGRRGPEQTQG; VTQAFGRRGPEQTQGN; VTQNVLYENQKLIAN; VTQNVLYENQKLIANQ; VTQNVLYENQKLIANQFNSAIG; VTQRNFYEPQIITTDN; VTSNYSGVVTTVMFL; VTTFDSEYCRHGTCE; VTTFDSEYCRHGTCER; VTTKIALKGGKIVNN; VTVKNGSIHLYFDKA; VTYKLDGVVCTEIDP; VTYVPAQEKNFTTAP; VTYVPAQEKNFTTAPA; VTYVPAQEKNFTTAPAICHDGK; VVAIDYKHYTPSFKKG; VVCKTCGQQQTTLKG; VVCKTCGQQQTTLKGV; VVCTEIDPKLDNYYKK; VVENPTIQKDVLECNV; VVEVVDKYFDCYDGGC; VVFDEISMATNYDLSV; VVFLHVTYVPAQEKN; VVIKVCEFQFCNDPF; VVIPDYNTYKNTCDGT; VVKIYCPACHNS; VVNAANVYLKHGGGVA; VVNARLRAKHYVYIGD; VVNVVTTKIALKGGK; VVTTFDSEYCRHGTCE; VVYRGTTTYKLNVGD; VWHVNNATNKATYKPN; VYAADPAMHAASGNLL; VYALRHFDEGNCDTL; VYALRHFDEGNCDTLK; VYANGGKGFCKLHNW; VYCCREHEHEIAWYT; VYCPRHVICTSEDMLN; VYDDGARRVWTLMNV; VYDPLQPELDSF; VYDPLQPELDSFKEE; VYIGDPAQLPAPRTLL; VYIKNADIVEEAKKV; VYKVYYGNALDQAIS; VYKVYYGNALDQAISM; VYLKHGGGVAGALNKA; VYPVASPNECNQMCLS; VYRGTTTYKLNVGDY; VYSDVENPHLMGWDYP; VYSTGSNVFQTRAGCL; VYSTGSNVFQTRAGCLIGAE; VYYGNALDQAISMWAL; VYYHKNNKSWMESEF; VYYHKNNKSWMESEFR; VYYPDKVFRSSVLHS; WADNNCYLATALLTL; WAQCVQLHNDILLAKD; WCKDGHVETFYPKLQS; WDLLKYDFTEERLKL; WDLLKYDFTEERLKLF; WDQTYHPNCVNCLDD; WDQTYHPNCVNCLDDR; WDTIANYAKPFLNKVV; WESGVKDCVVLHSYF; WFLNRFTTTLNDFNLV; WFVTQRNFYEPQIIT; WHVNNATNKATYKPNT; WICLLQFAYANRNRF; WKYPQVNGLTSIKWA; WLIINLVQMAPISAM; WLPTGTLLVDSDLND; WLTYTGAIKLDDKDP; WLTYTGAIKLDDKDPN; WNADLYKLMGHFAWW; WNIGEQKSILSPLYA; WNLREMLAHAEETRKL; WNTKHSSGVTRELM; WNVVRIKIVQMLSDT; WPQIAQFAPSASAFF; WPQIAQFAPSASAFFGMSRI; WRKAVFISPYNSQNA; WRVYSTGSNVFQ; WRVYSTGSNVFQTRA; WRVYSTGSNVFQTRAG; WSLFFFLYENAFLPF; WSTKPVETSNSFDVLK; WSYSGQSTQLGIEFL; WTAFVTNVNASSSEA; WTAFVTNVNASSSEAF; WTETDLTKGPHEFCSQ; WTFGAGAALQIPFAM; WTIEYPIIGDELKINA; WVATEGALNTPKDHIG; WWTAFVTNVNASSSEA; WYDFVENPDILRVYA; WYTERSEKSYELQTP; WYTERSEKSYELQTPF; YAAFATAQEAYEQAVA; YAAVINGDRWFLNRF; YAKPFLNKVVSTTTNI; YALRHFDEGNCDTLKE; YCIDGALLTKSSEYK; YCRCHIDHPNPKGFCD; YCYDTNVLEGSVAYE; YDAQPCSDKAYKIEE; YDFAVSKGFFKEGSS; YDFGDFIQTTPGSGVP; YDFTEERLKLFDRYFK; YDKLQFTSLEIPRRN; YDLSVVNARLRAKHY; YDLSVVNARLRAKHYV; YDPKTKNVTKENDSK; YDPKTKNVTKENDSKE; YDPLQPELDSFKEEL; YDPLQPELDSFKEELDKYFK; YDPLQPELDSFKEELDKYFKNHTS; YDPLQPELDSFKEELDKYFKNHTSPDVDLGDI; YDSMSYEDQDALFAY; YDTNVLEGSVAYESL; YDYVIFTQTTETAHSC; YECDIPIGAGICASYQ; YECLYRNRDVDTDFV; YEDFQENWNTKHSSGV; YEDLLIRKSNHNFLV; YEDQDALFAYTKRNVI; YELKHGTFTCASEYT; YELQTPFEIKLAKKF; YELQTPFEIKLAKKFD; YENQKLIANQFN; YENQKLIANQFNSAIG; YEPQIITTDNTFVSGN; YESLRPDTRYVLMDGS; YFCTCYFGLFCLLNR; YFDCYDGGCINANQVI; YFDKAGQKTYERHSL; YFDKAGQKTYERHSLS; YFIKGLNNLNRGMVLG; YFKNHTSPDVDLGDI; YFKNHTSPDVDLGDIS; YFKNHTSPDVDLGDISGINA; YFKNHTSPDVDLGDISGINASVVNIQKEID; YFKYWDQTYHPNCVN; YFKYWDQTYHPNCVNC; YFNKKDWYDFVENPDI; YFTEQPIDLVPNQPY; YFTEQPIDLVPNQPYP; YFVLTSHTVMPLSAP; YFYYLGTGPEAGLPY; YGADLKSFDLGDELG; YGADLKSFDLGDELGT; YGDCLGDIAARDLIC; YGDSATLPKGIMMNVA; YHFRELGVVHNQDVN; YHKNNKSWMESE; YHKNNKSWMESEFRV; YHPNCVNCLDDRCILH; YIAQVDVVNFNLT; YIDINGNLHPDSATL; YIDINGNLHPDSATLV; YIFWRNTNPIQLSS; YIGDPAQLPAPRTLL; YIIKLIFLWLLWPVTL; YIINLIIKNLSKSLTE; YIKWDLLKYDFTEERL; YILPSIISNEKQEIL; YILPSIISNEKQEILG; YKAFKQIVESCGNFK; YKDWSYSGQSTQLGIE; YKDYLASGGQPITNC; YKDYLASGGQPITNCV; YKENSYTTTIKPVTYK; YKGPITDVFYKENSYT; YKHITSKETLYCIDG; YKKDNSYFTEQPIDL; YKKDNSYFTEQPIDLV; YKLGASQRVAGDSGF; YKLGASQRVAGDSGFA; YKLNTDHSSSSDNIAL; YKVYYGNALDQAISMW; YKYFSGAMDTTSYRE; YLEGSVRVVTTFDSE; YLGKPREQIDGYVMHA; YLGTGPEAGLPYGAN; YLGTGPEAGLPYGANK; YLKLTDNVYIKNADIV; YLKRRVVFNGVSFST; YLPYPDPSRILGAGC; YLTSSSKTPEEHFIET; YLVQQESPFVMMSAP; YMGTLSYEQFKKGVQI; YNCCDDDYFNKKDWYD; YNGYLTSSSKTPEEH; YNGYLTSSSKTPEEHF; YNKVENMTPRDLGACI; YNKYKYFSGAMDTTSY; YNLLKDCPAVAKHDFF; YNSQNAVASKILGLPT; YNVTQAFGRRGPEQT; YNVTQAFGRRGPEQTQ; YNVTQAFGRRGPEQTQGNFG; YPGQGLNGYTVEEAKT; YPIIGDELKINAACRK; YPKLQSSQAWQPGV; YPLECIKDLLARAGK; YPNASFDNFKFVCDNI; YPNMFITREEAIRHV; YPQVNGLTSIKWAD; YPQVNGLTSIKWADNN; YPTLNISDEFSSNVAN; YPVASPNECNQMCLST; YQCGHYKHITSKET; YQCGHYKHITSKETLY; YQIGGYTEKWESGVKD; YQKVGMQKYSTLQG; YQKVGMQKYSTLQGPP; YQNNVFMSEAKCWTE; YQNNVFMSEAKCWTET; YQVNGYPNMFITREEA; YREAACCHLAKALND; YREAACCHLAKALNDF; YRFNGIGVTQNVLYE; YRIGNYKLNTDHSSS; YRIGNYKLNTDHSSSS; YRKVLLRKNGNKGA; YRKVLLRKNGNKGAGG; YRNRDVDTDFVNEFY; YRNRDVDTDFVNEFYA; YRRATRRIRGGDGKMK; YRRATRRIRGGDGKMKDLSPRW; YRRLISMMGFKMNYQ; YRVTKNSKVQIGEYTF; YSDVENPHLMGWDYP; YSDVENPHLMGWDYPK; YSFLPGVYSVIYLYL; YSLLLCRMNSRN; YSQLDEEQPMEI; YSQLDEEQPMEID; YSRYRIGNYKLNTDH; YSRYRIGNYKLNTDHS; YSTGSNVFQTRAGCL; YSTLQGPPGTGKSHFA; YSYATHSDKFTDGVCL; YTELEPPCRFVTDTPK; YTERSEKSYELQTPFE; YTFEKGDYGDAVVYR; YTFEKGDYGDAVVYRG; YTGAIKLDDKDP; YTGAIKLDDKDPNFKD; YTKRNVIPTITQMNL; YTMADLVYALRHFDE; YTMADLVYALRHFDEG; YTNSFTRGVYYPDKV; YTPSFKKGAKLLHKPI; YTPSKLIEYTDFATSA; YTRYVDNNFCGPDGY; YTSALLAGTITSGWT; YTVEEAKTVLKKCKSA; YVDNNFCGPDGYPLEC; YVDTPNNTDFSRVSAK; YVFIKRSDARTAPHG; YVIFTQTTETAHSCNV; YVKPGGTSSGDATTA; YVKPGGTSSGDATTAY; YVNTFSSTFNVPMEK; YVNTFSSTFNVPMEKL; YVPAQEKNFTTAPAIC; YVQIPTTCANDPVGF; YVTQQLIRAAEIRASANLAATKMSECVL; YVWKSYVHVVDGCNS; YWDQTYHPNCVNCLDD; YYGNALDQAISMWALI; YYHKNNKSWMESEFRV; YYHTTDPSFLGRYMSA; YYKKDNSYFTEQPIDL; YYKLGASQRVAGDSG; YYKLGASQRVAGDSGF; YYLGTGPEAGLPYGAN; YYLGTGPEAGLPYGANKDGI; YYPDKVFRSSVL; YYPDKVFRSSVLHST; YYQNNVFMSEAKCWTE; YYSQLMCQPILLLDQ; YYTSNPTTFHLDGEV;

# T Cell - MHC Class 1 epitopes

ADKYVRNLQHRLYEC; AEIRASANLAATKMSECV; AESHVDTDLTKPY; AGAALQIPFAMQMAYRFN; ALFAYTKRNVIPTIT; ALLAGTITSGWTFGAGAA; ALLSTDGNKIADKYV; ALRKVPTDNYITTY; ANKDGIIWVATEGAL; ANLGERVRQALLKTV; ANVYLKHGGGVAGAL; APHGVVFLHVTYV; APHGVVFLHVTYVPA; AQFAPSASAFFGMSR; ARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEI; ASFNYLKSPNFSK; ASQGLVASIKNFKSV; ASQRVAGDSGFAAY; ATEGALNTPKDHI; ATIPIQASLPFGWLIVGVALLAVFQSASKIITLKKRWQ; ATKMSECVLGQSKRV; ATSRTLSYYKLGASQ; AVRDPQTLEILDITPCSF; CILHCANFNVLFSTV; CLDDRCILHCANFNV; CLVGLMWLSYFIASF; CSLSHRFYRLANECA; DAVVCFNSTYASQGL; DDQIGYYRRATRRIRGGDG; DFAVSKGFFKEGSSV; DFIQTTPGSGVPVVD; DFYDFAVSKGFFK; DGNKIADKYVRNLQH; DGTLMIERFVSLAID; DGVPFVVSTGYHFRE; DIVKTDGTLMIERFV; DLNGNWYDFGDFIQT; DLSPRWYFYY; DLSPRWYFYYL; DLSPRWYFYYLGTGP; DLSPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALN; DLVYALRHFDEGNCD; DMYSVMLTNDNTSRY; DQTYHPNCVNCLDDR; DSLSSTASALGKLQD; DVDTDFVNEFYAYLR; DWYDFVENPDILRVY; DYKHWPQIAQF; DYKHWPQIAQFAPSASAFF; DYVYLPYPDPSRI; EAEVQIDRLITGRLQSLQ; ECDIPIGAGICASYQ; EEVGHTDLMAAY; EGSSVELKHFFFAQD; EKSNIIRGWIFGTTL; ELKHFFFAQDGNAAI; EPQIITTDNTFVSGN; EVTGDSCNNYMLTY; EVTPSGTWLTY; EYADVFHLYLQYIRK; FAYANRNRFLYIIKL; FDNPVLPFNDGVYF; FFAQDGNAAISDYDY; FGPLVRKIFVDGVPF; FHLYLQYIRKLHDEL; FHQKLLKSIAATRGA; FIASFRLFARTRSMW; FNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKL; FNKWGKARLYYDSMS; FNSTYASQGLVASIK; FPLTSFGPLVRKIFV; FPQSAPHGVVF; FRSSVLHSTQDLFLPFFS; FSVAALTNNVAFQTV; FTEQPIDLVPNQPY; FTSLEIPRRNVATLQ; FVCNLLLLFVTVYSHLLLV; FVNEFYAYLRKHFSM; FVSNGTHWFV; GAALQIPFAMQMAYR; GAALQIPFAMQMAYRF; GAGAALQIPFAMQMAYRFN; GAVILRGHLRIAGHH; GCFVDDIVKTDGTLM; GCINANQVIVNNLDK; GDGKMKDLSPRWYFYYLGT; GELGDVRETMSYLF; GMEVTPSGTWLTY; GMEVTPSGTWLTYTG; GMEVTPSGTWLTYTGAIKL; GMEVTPSGTWLTYTGAIKLD; GMSRIGMEVTPSGTW; GNAAISDYDYYRYNL; GNYTVSCLPFTI; GPKKSTNLVKNKCVN; GPPGTGKSHFAIGLA; GSLYVKPGGTSSGDA; GVAPGTAVLRQW; GVFVSNGTHW; GVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRW; GWDYPKCDRAMPNML; HAASGNLLLDKRTTC; HLRIAGHHLGR; HNDILLAKDTTEAFE; HTMLVKQGDDYVYLP; HVDTDLTKPYIKWDL; HVTYVPAQEKNF; HWFVTQRNFYEPQII; ICQAVTANVNALLST; IDGDMVPHISRQRLT; IERFVSLAIDAYPLT; IHADQLTPTW; IIKNLSKSLTENKY; IKDLPKEITVATSRT; IKLDDKDPNFKDQVI; IKWDLLKYDFTEERL; ILLIIMRTFKVSIWNLDYII; ILRVYANLGERVRQA; IPFAMQMAYRFNGI; IPFAMQMAYRFNGIG; IRGWIFGTTLDSKTQ; ITEEVGHTDLMAAY; ITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV; ITLKKRWQLAL; IVGVLTLDNQDLNGN; IYNLLKDCPAVAKHD; KARLYYDSMSYEDQD; KCDRAMPNMLRIMAS; KEITVATSRTLSYYK; KENSYTTTIKPVTY; KFLPFQQFGR; KGFFKEGSSVELKHF; KHFSMMILSDDAVVC; KHTTCCSLSHRFYRL; KLFDRYFKYWDQTYH; KLIANQFNSAIGKIQ; KLLHKPIVWHV; KMKDLSPRWYFYYLGTGPEAGLPYGANK; KNRARTVAGVSICST; KPGGTSSGDATTAYA; KPGNFNKDFYDFAVS; KPRQKRTATKAYNVTQAFG; KQGDDYVYLPYPDPS; KRTTCFSVAALTNNV; KVFRSSVLHSTQDLF; KYTMADLVYALRHFD; KYTQLCQYLNTLTLAVPYNMRVIHF; LFSTVFPLTSFGPLV; LITGRLQSLQTYV; LITLATCELYHYQECV; LKHGGGVAGALNKAT; LKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLE; LKSIAATRGATVVIG; LLKEPCSSGTY; LLVYAADPAMHAASG; LMIERFVSLAIDAYPLTK; LMPILTLTRALTAES; LMSFPQSAPHGVVFLHVTY; LPPLLTDEMIAQYTS; LPYGANKDGIIWVATEGAL; LQFAYANRNRFLY; LQGPPGTGKSHFAIG; LQPELDSFKEELDKY; LQSLQTYVTQQLIRA; LQTYVTQQLIRAAEI; LRAKHYVYIGDPAQL; LRGHLRIAGHHLGRC; LRIAGHHLGRCDIKD; LSDDAVVCFNSTY; LSFKELLVYAADPAM; LSPRWYFYY; LSPRWYFYYL; LSPRWYFYYLGTGPEAGL; LSYYKLGASQRVAGD; LSYYKLGASQRVAGDSGFAAYSRYRIGNYKLNTDHSS; LTAESHVDTDLTKPY; LTKPYIKWDLLKYDF; LTNNVAFQTVKPGNF; LYRNRDVDTDFVNEF; LYYQNNVFMSEAKCW; MFHLVDFQVTIAEILLIIMRTFKVSIWNLDY; MGYINVFAFPFTIYSL; MGYINVFAFPFTIYSLLLCRMN; MIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGV; MILSDDAVVCFNSTY; MKDLSPRWYFYYLGT; MLKTVYSDVENPHLM; MLTNDNTSRYWEPEF; MRNAGIVGVLTLDNQ; MSRIGMEVTPSGTWL; NCTEVPVAIHADQLT; NFKDQVILLNKHIDAYKTF; NFKSVLYYQNNVFMS; NIIRGWIFGTTLDSKTQS; NKDFYDFAVSKGFFK; NKKFLPFQQFGRDIA; NKLCEEMLDNRATLQ; NLLLDKRTTCFSVAA; NNSYECDIPIGAGICASYQ; NPHLMGWDYPKCDRA; NQDVNLHSSRLSFKE; NQVIVNNLDKSAGFP; NRDVDTDFVNEFYAY; NRNRFLYIIKLIFLW; NSSPDDQIGYY; NSVFNICQAVTANVN; NSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVK; NTCDGTTFTYASALWEIQQVV; NTNSSPDDQIGYY; NTSRYWEPEFY; NTSRYWEPEFYEAMY; NVFMSEAKCWTETDL; NYIAQVDVVNFNL; PFAMQMAYRFNGIGVTQN; PFAMQMAYRFNGIGVTQNV; PLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQE; PNCVNCLDDRCILHC; PPAYTNSFTRGVYYPDKVF; PPLLTDEMIAQYTSALLA; PTMCDIRQLLFVVEV; PTWRVYSTGSNVFQTRAG; QELIRQGTDYKHW; QFAYANRNRFLYIIK; QFCDAMRNAGIVGVL; QHEETIYNLLKDCPA; QIDGYVMHANYIFWR; QKYSTLQGPPGTGKS; QLIRAAEIRASANLA; QMNLKYAISAKNRAR; QPYVFIKRSDARTAP; QQLIRAAEIRASANL; QRNFYEPQIITTDNT; QSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFV; QSDIEVTGDSCNNY; QSLQTYVTQQLIRAAEIR; QVILLNKHIDAYKTF; REGVFVSNGTHW; REQIDGYVMHANY; RFDNPVLPFNDGVYFAST; RFYRLANECAQVLSE; RGHLRIAGHHLGRCD; RILGAGCFVDDIVKT; RIMASLVLARKHTTC; RLFARTRSMWSFNPE; RLITGRLQSLQTYVTQQL; RLYECLYRNRDVDTD; RNLQHRLYECLYRNR; RPQGLPNNTASWFTALTQH; RQRLTKYTMADLVYA; RRCPAEIVDTVSALVY; RRGPEQTQGNF; RTFKVSIWNLDY; RVAGDSGFAAY; RVAGDSGFAAYSRY; RVAGDSGFAAYSRYR; RVRQALLKTVQFCDA; SASAFFGMSRIGMEVTPSGTWLTYT; SASKIITLKKRW; SDYDYYRYNLPTMCD; SEHDYQIGGYTEKW; SELVIGAVILRGHLR; SELVIGAVILRGHLRIAGHHLGR; SEYKGPITDVFYKENSY; SEYTGNYQCGHY; SFNPETNILLNVPLH; SFPQSAPHGVVF; SFPQSAPHGVVFLHV; SFTRGVYYPDKVFRS; SGFAAYSRYRIGNYK; SGTNGTKRFDNPVLP; SIWNLDYIINL; SLLLCRMNSRNYIA; SMWSFNPETNIL; SNEKQEILGTVSWNL; SNGTHWFVTQRNFYEPQII; SPRWYFYYL; SPRWYFYYLGTGPEA; SPRWYFYYLGTGPEAGLPY; SQSIIAYTMSLGAENSVAY; SSGDATTAYANSVFN; SSPDDQIGYYR; SVLHSTQDLFLPFFS; SYECDIPIGAGICASYQT; SYYKLGASQRVAGDS; SYYSLLMPILTLTRA; TANVNALLSTDGNKI; TEERLKLFDRYFKYW; TESNKKFLPFQQF; TESNKKFLPFQQFGRDIAD; TETDLTKGPHEF; TETDLTKGPHEFCSQ; TGHMLDMYSVMLTND; TINCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDFI; TKAYNVTQAFGRRGPEQTQ; TKGPHEFCSQHTMLV; TKRFDNPVLPFNDGV; TKRNVIPTITQMNLK; TLDNQDLNGNWYDFG; TLTRALTAESHVDTD; TNGTKRFDNPVLPFNDGVY; TNSFTRGVYYPDKVFRSS; TPGSGVPVVDSYYSL; TPSGTWLTYTGAIKL; TQDLFLPFFSNVTWF; TQQLIRAAEIRASANLAA; TRSMWSFNPETNILL; TRSMWSFNPETNILLNVPLHGTILTRPLLESELVIGAVILR; TTAYANSVFNICQAV; TTYPGQGLNGY; TVAGVSICSTMTNRQ; TVATSRTLSYYKL; TVVIGTSKFYGGWHN; TVYDPLQPELDSFK; TWLTYTGAIKLDDKDPNFK; TYVPAQEKNFTTAPA; TYVTQQLIRAAEIRA; VAIHADQLTPTW; VAKHDFFKFRIDGDM; VASIKNFKSVLYYQN; VEAEVQIDRLITGR; VENPDILRVYANLGE; VENPHLMGWD; VFLHVTYVPAQEKNF; VFVSNGTHWF; VLPPLLTDEMIAQYT; VLQLPQGTTLPKGFYA; VNNSYECDIPIGAGI; VNRFNVAITRAKVGI; VPHVGEIPVAYRKVLL; VPVVDSYYSLLMPIL; VQIDRLITGRLQSLQ; VTQNVLYENQKLIAN; VVIKVCEFQFCNDPF; VYDPLQPELDSF; VYLPYPDPSRI; VYYPDKVFRSSVLH; VYYPDKVFRSSVLHS; VYYPDKVFRSSVLHSTQD; WPQIAQFAPSASAFFGMSRIGMEVT; WYDFGDFIQTTPGSG; YAISAKNRARTVAGV; YAYLRKHFSMMILSD; YECDIPIGAGI; YEDFLEYHDVRVVL; YEDQDALFAYTKRNV; YFKYWDQTYHPNCVN; YFNKKDWYDFVENPD;

YLKHGGGVAGALNKA; YLPYPDPSRIL; YMHHMELPTGV; YPDPSRILGAGCFVD; YRYNLPTMCDIRQLL; YSDVENPHLMGW; YSRYRIGNYK; YSRYRIGNYKLNTDH; YSTGSNVFQTRAGCL; YVYLPYPDPSRILGA; YYRARAGEAANF

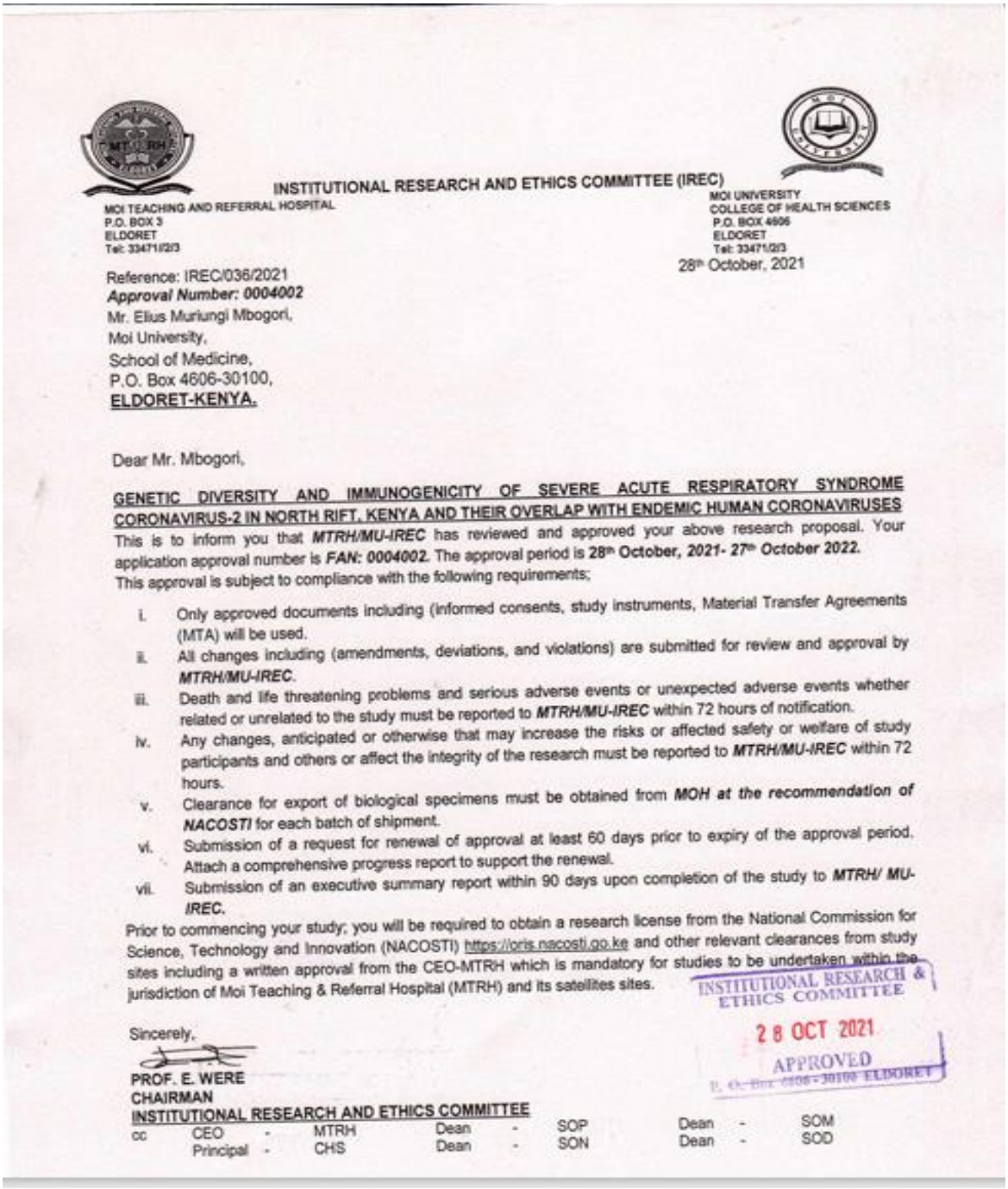
# T Cell - MHC Class 2 epitopes

AAEASKKPRQKRTAT; AAEIRASANLAATKM; ACLVGLMWLSYFI; ADKYVRNLQHRLYEC; AEILLIIMRTFKVSI; AEIRASANLAATKMS; AEIRASANLAATKMSECV; AFFGMSRIGMEVTPS; AFFGMSRIGMEVTPSGTWLT; AIKLDDKDPNFKDQV; AIVLQLPQGTTLPKG; ALFAYTKRNVIPTIT; ALLAGTITSGWTFGA; ALLAVFQSASKIITL; ALLSTDGNKIADKYV; ALQIPFAMQMAYRFNGIGV; ANECAQVLSEMVMCG; ANKDGIIWVATEG; ANKDGIIWVATEGAL; ANKDGIIWVATEGALNTPK; ANLGERVRQALLKTV; ANVYLKHGGGVAGAL; APHGVVFLHVTYVPA; AQFAPSASAFFGM; AQFAPSASAFFGMSRIGMEV; AQYTSALLAGTITSG; ASAFFGMSRIGMEVT; ASFRLFARTRSMWSFN; ASQGLVASIKNFKSV; ASRELKVTFFPDLNG; ASWFTALTQHGK; ATKMSECVLGQSKRV; ATSRTLSYYKLGASQ; AVTANVNALLSTDGN; AVYRINWITGGIA; CCHLAKALNDFSNSG; CILHCANFNVLFSTV; CLDDRCILHCANFNV; CLVGLMWLSYFIASF; CNLLLLFVTVYSHLL; CSLSHRFYRLANECA; CTFLLNKEMYLKLRS; CVDEAGSKSPIQYID; DALFAYTKRNVIPTITQMNLKY; DAVVCFNSTYASQGL; DDQIGYYRRATRRIR; DDQIGYYRRATRRIRGG; DDQIGYYRRATRRIRGGDG; DDQIGYYRRATRRIRGGDGK; DDSEPVLKGVKLHYT; DFAVSKGFFKEGSSV; DFCGKGYHLMSFPQS; DFIQTTPGSGVPVVD; DFLEYHDVRVVLDFI; DGNKIADKYVRNLQH; DGTLMIERFVSLAID; DGVPFVVSTGYHFRE; DITPCSFGGVSVITP; DIVKTDGTLMIERFV; DKVEAEVQIDRLITGRLQSL; DLNGNWYDFGDFIQT; DLSPRWYFYYLGTGP; DLVYALRHFDEGNCD; DMYSVMLTNDNTSRY; DNFKFVCDNIKFADD; DQIGYYRRATRRIRGG; DQTYHPNCVNCLDDR; DQVILLNKHIDAY; DSFKEELDKYFKNHT; DSKTQSLLIVNNATN; DSLSSTASALGKLQD; DTLRVEAFEYYHT; DTRYVLMDGSIIQFP; DTTDAVRDPQTLEIL; DVDTDFVNEFYAYLR; DVFHLYLQYIRKLHD; DVFYKENSYTTTI; DWYDFVENPDILRVY; DYKHWPQIAQFAPSA; DYKHWPQIAQFAPSASAFF; DYKHWPQIAQFAPSASAFFG; EAEVQIDRLITGRLQSLQ; ECDIPIGAGICASYQ; EDFLEYHDVRVVLDF; EFCSQHTMLVKQGDD; EFYAYLRKHFSMMIL; EGALNTPKDHIGTRN; EGSSVELKHFFFAQD; EIIKSQDLSVVSKVV; EILLIIMRTFKVSIW; EIRASANLAATKM; EIRASANLAATKMSECVLGQ; EKNFTTAPAICHDGKAHF; EKSNIIRGWIFGTTL; ELDKYFKNHTSPDVD; ELKHFFFAQDGNAAI; ELLHAPATVCGPKKSTNLVK; ENQKLIANQFNSAIGKI; EPQIITTDNTFVSGN; ESNKKFLPFQQFGRDIAD; ESPFVMMSAPPAQYE; EVQIDRLITGRLQSL; EYADVFHLYLQYIRK; FAFPFTIYSL; FDAYVNTFSSTFNVP; FFAQDGNAAISDYDY; FGDQELIRQGTDYKH; FGEYSHVVAFNTLLF; FGPLVRKIFVDGVPF; FGTTLDSKTQSLLIV; FHLYLQYIRKLHDEL; FHQKLLKSIAATRGA; FIASFRLFARTRSMW; FKDQVILLNKHIDAY; FLFLTWICLLQFAYA; FLHFLPRVFSAVGNI; FMRIFTIGTVTLKQG; FNICQAVTANVVNALL; FNSTYASQGLVASIK; FPLTSFGPLVRKIFV; FPRGQGVPINTNSSP; FRSSVLHSTQDLFLPFFS; FSNSGSDVLYQPPQT; FSVAALTNNVAFQTV; FTGYRVTKNSKVQIG; FTINCQEPKLGSLVV; FTSLEIPRRNVATLQ; FVNEFYAYLRKHFSM; FVSGNCDVVIGIV; FVSGNCDVVIGIVNN; FVSLAIDAYPLTKHP; FVVEVVDKYFDCYDG; FYEPQIITTDNTFVSGNCD; GAALQIPFAMQMAYR; GAALQIPFAMQMAYRF; GAGAALQIPFAMQ; GAGAALQIPFAMQMAYRFN; GAGAALQIPFAMQMAYRFNG; GAVILRGHLRIAG; GAVILRGHLRIAGHH; GAVILRGHLRIAGHHLG; GAVILRGHLRIAGHHLGR; GCFVDDIVKTDGTLM; GCINANQVIVNNLDK; GDFKLNEEIAIIL; GDGKMKDLSPRWYFYYLGT; GDSGFAAYSRYRI; GDVVAIDYKHYTP; GETFVTHSKGLYRKC; GGDGKMKDLSPRWYF; GHHLGRCDIKDLPKEITVAT; GIIWVATEGALN; GIIWVATEGALNT; GIIWVATEGALNTPK; GIIWVATEGALNTPKDHIGT; GIVNNTVYDPLQPEL; GKIQDSLSSTASALGKLQ; GLMWLSYFIASFRLFARTRS; GLTVLPPLLTDEMIAQYTSA; GLVASIKNFKSVLYY; GMEVTPSGTWLTYTG; GMEVTPSGTWLTYTGAIKL; GMEVTPSGTWLTYTGAIKLD; GMSRIGMEVTPSGTW; GPEAGLPYGANKDGI; GPKKSTNLVKNKCVN; GPKKSTNLVKNKCVNFNFNG; GPPGTGKSHFAIGLA; GRCDIKDLPKEIT; GRDIADTTDAVRDPQ; GSLVVRCSFYEDFLE; GSLYVKPGGTSSGDA; GTILTRPLLESELVI; GTWLTYTGAIKLDDK; GVPINTNSSPDDQIG; GVYYPDKVFRSS; GWDYPKCDRAMPNML; GWTFGAGAALQIP; GWTFGAGAALQIPFA; GYKKPASRELKVTFF; GYYRRATRRIRG; GYYRRATRRIRGG; GYYRRATRRIRGGDG; HAASGNLLLDKRTTC; HADQLTPTWRVYSTGSNVFQ; HCANFNVLFSTVFPP; HGTILTRPLLESE; HLRIAGHHLGR; HSTQDLFLPFFSN; HTMLVKQGDDYVYLP; HTTDPSFLGRYMSAL; HVDTDLTKPVIKWDL; HWFVTQRNFYEPQ; HWFVTQRNFYEPQII; HWFVTQRNFYEPQIITTDN; IAATRGATVVIGTSK; IAEILLIIMRTFKVS; IAGHHLGRCDIKDLP; IASFRLFARTRSMWSFNPET; ICQAVTANVNALLST; IDGDMVPHISRQRLT; IDRLITGRLQSLQTYVTQQL; IERFVSLAIDAYPLT; IGKIQDSLSSTASAL; IGNYKLNTDHSSSSD; IGNYKLNTDHSSSSDNIALL; IGNYTVSCLPFTINC; IGYYRRATRRIRGGD; IKDLPKEITVATSRT; IKDLPKEITVATSRTLSYYK; IKLDDKDPNFKDQVI; IKWDLLKYDFTEERL; ILRVYANLGERVRQA; INVFAPFTIYSLLL; IPFAMQMAYRF; IPFAMQMAYRFNG; IPFAMQMAYRFNGI; IPFAMQMAYRFNGIG; IPFAMQMAYRFNGIGVTQNV; IQWMVMFTPLVPFWI; IQYIDIGNYTVSCLP; IRGWIFGTTLDSKTQ; IRQLLFVVEVVDKYF; ISPYNSQNAVASKIL; IVGVLTLDNQDLNGN; IWNLDYIINLIIKNL; IWVATEGALNTPKDH; IYNLLKDCPAVAKHD; KAHFPREGVFVSNGT; KAHFPREGVFVSNGTHWFVT; KARLYYDSMSYEDQD; KAYNVTQAFGRRGPE; KCDRAMPNMLRIMAS; KDCPAVAKHDFFKFR; KDGIIWVATEGALNT; KEELDKYFKNHTSPD; KEITVATSRTLSYYK; KGFFKEGSSVELKHF; KGGKIVNNWLKQLIK; KHFSMMILSDDAVVC; KHFYWFFSNYLKRRV; KHTTCCSLSHRFYRL; KIITLKKRWQLALSK; KKFLPFQQFGR; KKFLPFQQFGRDI; KKPRQKRTATKAYNV; KKRWQLALSKGVHFV; KLFDRYFKYWDQTYH; KLIANQFNSAIGK; KLIANQFNSAIGKIQ; KLIANQFNSAIGKIQDSLS; KLLKSIAATRGATVV; KMKDLSPRWYFYYLGTGPEAGLPYGANK; KNFTTAPAICHDGKAHFPRE; KNRARTVAGVSICST; KPGGTSSGDATTAYA; KPGNFNKDFYDFAVS; KPRQKRTATKAYNVT; KPRQKRTATKAYNVTQAFG; KPRQKRTATKAYNVTQAFGR; KPRQKRTATKAYNVTQAFGRRGPEQTQ; KQGDDYVYLPYPDPS; KRTATKAYNVTQAFG; KRTTCFSVAALTNNV; KSAFYILPSIISNEK; KSNIIRGWIFGTTLDSKTQS; KSTNLVKNKCVNF; KTDGTLMIERFVSLA; KVCEFQFCNDPFL; KVEAEVQIDRLITGR; KVEAEVQIDRLITGRL; KVEAEVQIDRLITGRLQSLQT; KVFRSSVLHSTQDLF; KVFRSSVLHSTQDLFLPFFS; KVTFFPDLNGDVVAI; KYTMADLVYALRHFD; KYTQLCQYLNTLTLAVPYNMRVIHF; LDDFVEIIKSQDLSV; LEASFNYLKSPNFSK; LFSTVFPLTSFGPLV; LGASQRVAGDSGFAA; LGDIAARDLICAQKF; LGRCDIKDLPKEITV; LGTGPEAGLPYGA; LGVVHNQDVNLHSSR; LIIMRTFKVSIWNLD; LIRQGTDYKHWPQIA; LITGRLQSLQTYVTQ; LKHGGGVAGALNKAT; LKKRWQLALSKGV; LKSIAATRGATVVIG; LLEQWNLVIGFLFLT; LLTDEMIAQYTSALLAGTI; LLVYAADPAMHAASG; LMIERFVSLAIDAYP; LMIERFVSLAIDAYPLTK; LMPILTLTRALTAES; LNQLTGYKKPASREL; LNTDHSSSSDNIALL; LPNDDTLRVEAFEYY; LPNNTASWFTALTQHGKED; LPPLLTDEMIAQYTS; LPQGTTLPKGFYAEG; LPQGTTLPKGFYAEGSRGGS; LPYGANKDGIIWVAT; LPYGANKDGIIWVATEGAL; LPYGANKDGIIWVATEGALN; LQGPPGTGKSHFAIG; LQPELDSFKEELDKY; LQSLQTYVTQQLIRA; LQSLQTYVTQQLIRAAEIRA; LQTYVTQQLIRAAEI; LRAKHYVYIGDPAQL; LRGHLRIAGHHLG; LRGHLRIAGHHLGRC; LRGHLRIAGHHLGRCDIKDL; LRHFDEGNCDTLKEI; LRIAGHHLGRCDIKD; LRKHFSMMILSDDAV; LSFELLHAPATVCGPKKSTN; LSFKELLVYAADPAM; LSYFIASFRLFAR; LSYYKLGASQRVAGD; LTAESHVDTDLTKPY; LTDEMIAQYTSALLA; LTDEMIAQYTSALLAGTIT; LTDEMIAQYTSALLAGTITS; LTKPYIKWDLLKYDF; LTNNVAFQTVKPGNF; LTQYNRYLALYNKYK; LTYTGAIKLDDKDPN; LYENQKLIANQFNSA; LYIIKLIFLWLLWPV; LYRNRDVDTDFVNEF; LYYQNNVFMSEAKCW; MAYRFNGIGVTQNVLY; MCVEYCPIFFITG; MFHLVDFQVTIAEIL; MFTPLVPFWITIAYI; MIAQYTSALLAGTIT; MILSDDAVVCFNSTY; MKDLSPRWYFYYLGT; MKDLSPRWYFYYLGTGPEAG; MLKTVYSDVENPHLM; MLTNDNTSRYWEPEF; MPNMLRIMASLVLAR; MRNAGIVGVLTLDNQ; MWLSYFIASFRLFAR; NELSRVLGLKTLATH; NFGDQELIRQGTDYKHWPQ; NFGDQELIRQGTDYKHWPQIAQF; NFKDQVILLNKHIDAYKTF; NFKDQVILLNKHIDAYKTFP; NFKSVLYYQNNVFMS; NHNFLVQAGNVQLRV; NHTKKWKYPQVNGLT; NIIRGWIFGTTLDSKTQ; NIIRGWIFGTTLDSKTQS; NILLNVPLHGTIL; NKDFYDFAVSKGFFK; NKDGIIWVATEGALN; NKKFLPFQQFGRDIA; NLLLDKRTTCFSVAA; NLVIGFLFLTWICLL; NMFITREEAIRHVRA; NNATNVVIKVCEFQF; NNLDKSAGFPFNKWG; NNSYECDIPIGAG; NNSYECDIPIGAGIC; NNSYECDIPIGAGICASYQ; NPHLMGWDYPKCDRA; NPVLPFNDGVYFAST; NQDVNLHSSRLSFKE; NQFNSAIGKIQDS; NQVIVNNLDKSAGFP; NRFLYIIKLIFLWLLWPVTL; NRNRFLYIIKLIF; NRNRFLYIIKLIFLW; NSVFNICQAVTANVN; NTCDGTTFTYASALWEIQQVV; NTLLFLMSFTVLCLT; NTSRYWEPEFYEAMY; NTVKSVGKFCLEASF; NVFMSEAKCWTETDL; NVNRFNVAITRAKVG; NVPLHGTILTRPLLE; NVVIKVCEFQFCNDP; PAICHDGKAHFPREGVFV; PANNAAIVLQLPQGT; PDILRVYANLGERVR; PDLNGDVVAIDYKHY; PEAGLPYGANKDG; PEEHFIETISLAG; PFAMQMAYRFNGIGV; PFAMQMAYRFNGIGVTQNV; PFAMQMAYRFNGIGVTQNVL; PFQQFGRDIADTTDA; PHGVVFLHVTYVP; PHGVVFLHVTYVPAQEKNFT; PIGAGICASYQTQ; PLMYKGLPWNVVRIK; PLNSIIKTIQPRVEK; PNCVNCLDDRCILHC; PNFKDQVILLNKHIDAYK; PNMLRIMASLVLARK; PNNTASWFTALTQHG; PPAYTNSFTRGVYYPDKVF; PPAYTNSFTRGVYYPDKVFR; PRITFGGPSDSTG; PRPPLNRNYVFTGYR; PRWYFYYLGTGPEAG; PSASAFFGMSRIGMEVTPS; PSGTWLTYTGAIKLD; PTMCDIRQLLFVVEV; PTWRVYSTGSNVFQT; PVYSFLPGVYSVIYL; QDALFAYTKRNVIPT; QEKNFTTAPAICHDG; QELIRQGTDYKHW; QELIRQGTDYKHWPQ; QELIRQGTDYKHWPQIAQF; QELIRQGTDYKHWPQIAQFA; QEPKLGSLVVRCSFY; QFAPSASAFFGMSRI; QFAYANRNRFLYIIK; QFAYANRNRFLYIIKLI; QFCDAMRNAGIVGVL; QHEETIYNLLKDCPA; QIGYYRRATRRIR; QIITTDNTFVSGN; QIITTDNTFVSGNCDVVIGI; QIPFAMQMAYRFNGI; QKYSTLQGPPGTGKS; QLIRAAEIRASANLA; QLIRAAEIRASANLAATK; QLIRAAEIRASANLAATKM; QMAYRFNGIGVTQNV; QMNLKYAISAKNRAR; QNCVLKLKVDTANPK; QNVLYENQKLIAN; QNVLYENQKLIANQF; QNVLYENQKLIANQFNSAI; QNVLYENQKLIANQFNSAIG; QQLIRAAEIRASANL; QRNFYEPQIITTDNT; QSLQTYVTQQLIRAAEIR; QTFFKLVNKFLALCA; QTQGNFGDQELIRQG; QTQGNFGDQELIRQGTDYKH; QTYVTQQLIRAAEIR; QVILLNKHIDAYK; QVILLNKHIDAYKTF; QYGDCLGDIAARDLI; QYIRKLHDELTGHML; RAMPNMLRIMASLVL; RATRRIRGGDGKMKD; RATRRIRGGDGKMKDLSPRW; RCSFYEDFLEYHDVR; REEAIRHVRAWIGFD; REGVFVSNGTHWFVT; RELKVTFFPDLNG; RFDNPVLPFNDGVYFAST; RFNGIGVTQNVLYENQKLIA; RFYRLANECAQVLSE; RGHLRIAGHHLGRCD; RGHLRIAGHHLGRCDIK; RGHLRIAGHHLGRCDIKDLP; RIKIVQMLSDTLKNL; RILGAGCFVDDIVKT; RIMASLVLARKHTTC; RKIFVDGVPFVVSTG; RLFARTRSMWSFNPE; RLITGRLQSLQTYVTQQL; RLYECLYRNRDVDTD; RNFYEPQIITTDNTFVSG; RNLQHRLYECLYRNR; RPLLESELVIGAVIL; RPQGLPNNTASWFTA; RPQGLPNNTASWFTALTQH; RPQGLPNNTASWFTALTQHG; RQGTDYKHWPQIAQFAPSA; RQRLTKYTMADLVYA; RRCPAEIVDTVSALVY; RRGPEQTQGNFGDQE; RSKQRRPQGLPNNTA; RTRSMWSFNPETNILLNVPL; RVAGDSGFAAYSRYR; RVRQALLKTVQFCDA; RWYFYYLGTGPEAGL; RYLALYNKYKYFSGA; SAFFGMSRIGME; SAFFGMSRIGMEVTPSGTW; SAGFPFNKWGKARLY; SAIGKIQDSLSSTASALGKL; SANLAATKMSECVLG; SASAFFGMSRIGMEVTPSGTWLTYT; SCGSCCKFDEDDSEPVLKGV; SDFVRATATIPIQAS; SDYDYYRYNLPTMCD; SEFSSLPSYAAFATA; SELVIGAVILRGHLR; SELVIGAVILRGHLRIAGHHLGR; SFLAHIQWMVMFTPL; SFLGRYMSALNHTKK; SFNPETNILLNVPLH; SFPQSAPHGVVFLHV; SFTRGVYYPDKVFRS; SGFAAYSRYRIGNYK; SGTNGTKRFDNPVLP; SGTWLTYTGAIKLDDKDPNFK; SHRFYRLANECAQVL; SICSTMTNRQFHQKL; SKLWAQCVQLHNDIL; SLLIVNNATNVVIKV; SLLMPILTLTRALTA; SNGTHWFVTQRNFYEPQII; SNGTHWFVTQRNFYEPQIIT; SPDVDLGDISGINAS; SPLYAFASEAARVVR; SPRWYFYYL; SPRWYFYYLGTGP; SPRWYFYYLGTGPEA; SPRWYFYYLGTGPEAGLPY; SQSIIAYTMSLGAENSVAY; SQSIIAYTMSLGAENSVAYS; SRTLSYYKLGASQRV; SSGDATTAYANSVFN; SSVLHSTQDLFLP; STQDLFLPFFSNVTWFH; SVLHSTQDLFLPFFS; SYYKLGASQRVAG; SYYKLGASQRVAGDS; SYYKLGASQRVAGDSGFAAY; SYYSLLMPILTLTRA; TANVNALLSTDGNKI; TAPAICHDGKAHFPR; TASWFTALTQHGK; TASWFTALTQHGKED; TDEMIAQYTSALLAG; TDNTFVSGNCDVVIG; TDYKHWPQIAQFAPS; TEERLKLFDRYFKYW; TESNKKFLPFQQFGRDIAD; TESNKKFLPFQQFGRDIADT; TETDLTKGPHEFCSQ; TFGGPSDSTGSNQNG; TFKVSIWNLDYIINL; TFSNYQHEETIYNLL; TGHMLDMYSVMLTND; THWFVTQRNFYEPQIITT; TIAYIICISTKHFYW; TITSGWTFGAGAALQ; TKAYNVTQAFGRR; TKAYNVTQAFGRRGPEQTQ; TKAYNVTQAFGRRGPEQTQG; TKGPHEFCSQHTMLV; TKRFDNPVLPFNDGV; TKRNVIPTITQMNLK; TLDNQDLNGNWYDFG; TLEILDITPCSFGGV; TLRVEAFEYYHTTDP; TLTRALTAESHVDTD; TNGTKRFDNPVLPFNDGVY; TNILLNVPLHGTILT; TNILLNVPLHGTILTRP; TNLVKNKCVNFNFNG; TNSFTRGVYYPDKVFRSS; TNSRIKASMPTTIAK; TNSSPDDQIGYYRRA; TPGSGVPVVDSYYSL; TPKDHIGTRNPANNA; TPSFKKGAKLLHKPI; TQAFGRRGPEQTQGN; TQDLFLPFFSNVTWF; TQLCQYLNTLTLAVP; TQQLIRAAEIRASANLAA; TQQLIRAAEIRASANLAATK; TQRNFYEPQIITTDNTFVSG; TRGVYYPDKVFRS; TRGVYYPDKVFRSSV; TRGVYYPDKVFRSSVLHSTQ; TRRIRGGDGKMKDLS; TRSMWSFNPETNILL; TSRTLSYYKLGASQRVA; TTAPAICHDGKAHFP; TTAYANSVFNICQAV; TTDNTFVSGNCDVVI; TVAGVSICSTMTNRQ; TVATSRTLSYYKLGASQRVA; TVVIGTSKFYGGWHN; TVYDPLQPELDSFKE; TWLTYTGAIKLD; TWLTYTGAIKLDDKD; TWLTYTGAIKLDDKDPNF; TWLTYTGAIKLDDKDPNFK; TWLTYTGAIKLDDKDPNFKD; TYVPAQEKNFTTAPA; TYVTQQLIRAAEIRA; VAKHDFFKFRIDGDM; VASIKNFKSVLYYQN; VDKYFDCYDGGCINA; VENPDILRVYANLGE; VFLHVTYVPAQEKNF; VLAAVYRINWITGGI; VLSFCAFAVDAAKAY; VLTESNKKFLPFQQF; VNARLRAKHYVYIGD; VNNATNVVIKVCEFQFC; VNNSYECDIPIGAGI; VNNTVYDPLQPELDS; VNRFNVAITRAKVGI; VPHISRQRLTKYTMA; VPVVDSVYSLLMPIL; VQIDRLITGRLQSLQ; VQIDRLITGRLQSLQT; VRDPQTLEILDITPC; VSNGTHWFVTQRNFY; VTKNSKVQIGEYTFE; VTQNVLYENQKLIAN; VVIKVCEFQFCNDPF; VVLSFELLHAPATVCGPKKS; VWHVNNATNKATYKP; VYYPDKVFRSSVLHS; VYYPDKVFRSSVLHSTQ; VYYPDKVFRSSVLHSTQD; WICLLQFAYANRNRF; WPQIAQFAPSASAFF; WQLALSKGVHFVC; WRVYSTGSNVFQTRAGCLIG; WTFGAGAALQIPFAM; WYDFGDFIQTTPGSG; WYFYYLGTGPEA; WYFYYLGTGPEAGLP; YAISAKNRARTVAGV; YAYLRKHFSMMILSD; YCIDGALLTKSSEYK; YEDQDALFAYTKRNV; YENQKLIANQFNSAI; YENQKLIANQFNSAIGKIQ; YFKYWDQTYHPNCVN; YFNKKDWYDFVENPD; YIINLIIKNLSKSLT; YKHWPQIAQFAPSAS; YLKHGGGVAGALNKA; YLTFYLTNDVSFLAH; YMSALNHTKKWKYPQ; YPDPSRILGAGCFVD; YPKCDRAMPNMLRIM;

YREAACCHLAKALND; YRYNLPTMCDIRQLL; YSDVENPHLMGWDYP; YSRYRIGNYK; YSRYRIGNYKLNTDH; YSTGSNVFQTRAGCL; YTGAIKLDDKDPNFK; YVYLPYPDPSRILGA; YYKLGASQRVA; YYLGTGPEAGLPYGA; YYRRATRRIRGGDGK; YYRRATRRIRGGDGKMK.

# App: 3 APPENDIX 2: ETHICAL REQUIREMENTS

## App: 3.1 IREC Approval



INSTITUTIONAL RESEARCH AND ETHICS COMMITTEE (IREC)

MOI TEACHING AND REFERRAL HOSPITAL
P.O. BOX 3
ELDORET
Tel: 33471/2/3

MOI UNIVERSITY
COLLEGE OF HEALTH SCIENCES
P.O. BOX 4606
ELDORET
Tel: 33471/2/3

28th October, 2021

Reference: IREC/038/2021
***Approval Number: 0004002***
Mr. Elius Muriungi Mbogori,
Moi University,
School of Medicine,
P.O. Box 4606-30100,
**ELDORET-KENYA.**

Dear Mr. Mbogori,

**GENETIC DIVERSITY AND IMMUNOGENICITY OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS-2 IN NORTH RIFT, KENYA AND THEIR OVERLAP WITH ENDEMIC HUMAN CORONAVIRUSES**

This is to inform you that ***MTRH/MU-IREC*** has reviewed and approved your above research proposal. Your application approval number is ***FAN: 0004002***. The approval period is **28th October, 2021- 27th October 2022.**
This approval is subject to compliance with the following requirements;

i. Only approved documents including (informed consents, study instruments, Material Transfer Agreements (MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by ***MTRH/MU-IREC***.
iii. Death and life threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to ***MTRH/MU-IREC*** within 72 hours of notification.
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to ***MTRH/MU-IREC*** within 72 hours.
v. Clearance for export of biological specimens must be obtained from ***MOH at the recommendation of NACOSTI*** for each batch of shipment.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to ***MTRH/ MU-IREC.***

Prior to commencing your study; you will be required to obtain a research license from the National Commission for Science, Technology and Innovation (NACOSTI) https://oris.nacosti.go.ke and other relevant clearances from study sites including a written approval from the CEO-MTRH which is mandatory for studies to be undertaken within the jurisdiction of Moi Teaching & Referral Hospital (MTRH) and its satellites sites.

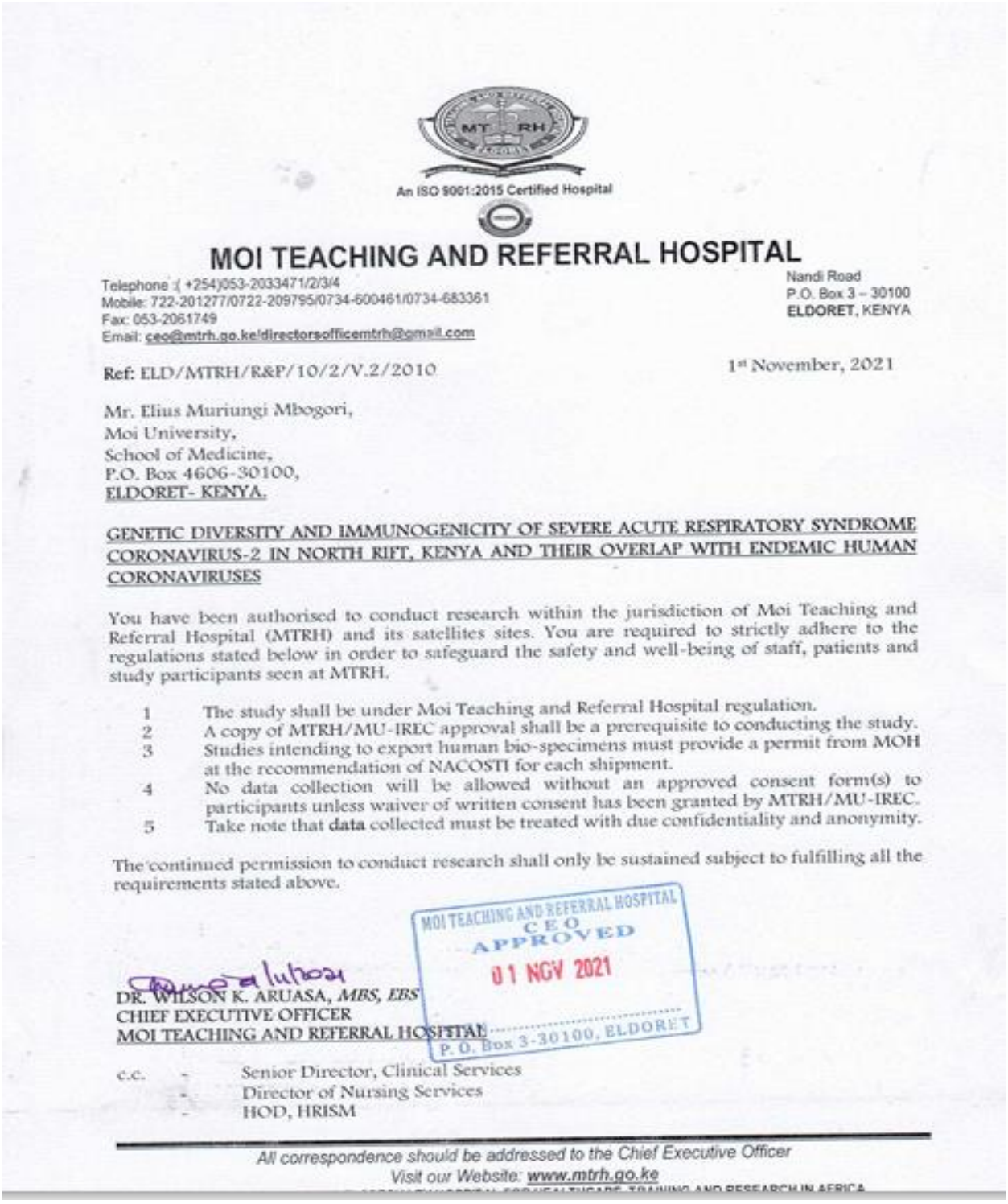
INSTITUTIONAL RESEARCH & ETHICS COMMITTEE
28 OCT 2021
APPROVED
P. O. Box 4606 - 30100 ELDORET

Sincerely,

**PROF. E. WERE**
**CHAIRMAN**
**INSTITUTIONAL RESEARCH AND ETHICS COMMITTEE**

cc CEO - MTRH; Principal - CHS; Dean - SOP; Dean - SON; Dean - SOM; Dean - SOD

**App: 3.2 MTRH Approval**



An ISO 9001:2015 Certified Hospital

MOI TEACHING AND REFERRAL HOSPITAL

Telephone : ( +254)053-2033471/2/3/4
Mobile: 722-201277/0722-209795/0734-600461/0734-683361
Fax: 053-2061749
Email: ceo@mtrh.go.ke/directorsofficemtrh@gmail.com

Nandi Road
P.O. Box 3 – 30100
ELDORET, KENYA

Ref: ELD/MTRH/R&P/10/2/V.2/2010

1st November, 2021

Mr. Elius Muriungi Mbogori,
Moi University,
School of Medicine,
P.O. Box 4606-30100,
ELDORET- KENYA.

GENETIC DIVERSITY AND IMMUNOGENICITY OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS-2 IN NORTH RIFT, KENYA AND THEIR OVERLAP WITH ENDEMIC HUMAN CORONAVIRUSES

You have been authorised to conduct research within the jurisdiction of Moi Teaching and Referral Hospital (MTRH) and its satellites sites. You are required to strictly adhere to the regulations stated below in order to safeguard the safety and well-being of staff, patients and study participants seen at MTRH.

1. The study shall be under Moi Teaching and Referral Hospital regulation.
2. A copy of MTRH/MU-IREC approval shall be a prerequisite to conducting the study.
3. Studies intending to export human bio-specimens must provide a permit from MOH at the recommendation of NACOSTI for each shipment.
4. No data collection will be allowed without an approved consent form(s) to participants unless waiver of written consent has been granted by MTRH/MU-IREC.
5. Take note that **data** collected must be treated with due confidentiality and anonymity.

The continued permission to conduct research shall only be sustained subject to fulfilling all the requirements stated above.

MOI TEACHING AND REFERRAL HOSPITAL
CEO
APPROVED
01 NOV 2021
P. O. Box 3-30100, ELDORET

DR. WILSON K. ARUASA, *MBS, EBS*
CHIEF EXECUTIVE OFFICER
MOI TEACHING AND REFERRAL HOSPITAL

c.c. - Senior Director, Clinical Services
- Director of Nursing Services
- HOD, HRISM

*All correspondence should be addressed to the Chief Executive Officer*
*Visit our Website: www.mtrh.go.ke*

# App: 4 APPENDIX 3: PUBLICATIONS

## App: 4.1 Bioinformatic Identification of conserved epitopes from SARS-CoV-2 genome isolated in Kenya

Next Research xxx (xxxx) 100215

Contents lists available at ScienceDirect

Next Research

journal homepage: www.elsevier.com

Journal logo

### Bioinformatic identification of conserved epitopes from SARS-COV-2 genome isolated in Kenya

Elius Mbogori [a, b, *], Stanslaus Musyoki [c], Richard Biegon [b], Kirtika Patel [b]

[a] Moi Teaching and Referral Hospital, Kenya
[b] Moi University, Kenya
[c] South Eastern Kenya University, Kenya

ARTICLE INFO

*Keywords:*
SARS-CoV-2
ORFfinder
BlastP
Epitopes
Conservation

ABSTRACT

The development of vaccines and diagnostic tools for SARS-CoV-2 heavily relies on identifying conserved epitopes across various virus strains. BLASTp is a pivotal bioinformatics tool for comparing protein sequences to unveil regions of similarity, aiding in understanding evolutionary relationships and functional conservation. The current study used bioinformatics methods to highlight the conserved epitopes on SARS-CoV-2 genomes isolated in Moi Teaching and Referral Hospital. To achieve this objective, the genomes were divided into their constituent genes using NCBI ORFfinder and translated to proteins using EXPASY. BlastP was then used to identify the proteins. Meanwhile, epitopes from the Wuhan genome were downloaded from IEDB and a BlastP analysis was done to identify matching epitopes. From the IEDB databank, 12,285 Wuhan genome epitopes were found and on conducting BlastP analysis, 5154 epitopes were isolated. These epitopes were deemed conserved as they had not changed despite numerous mutations. The identification and analysis of the conserved epitopes in the SARS-CoV-2 genome are crucial for the development of effective vaccines and diagnostic tools. Further laboratory experiments are however recommended to ascertain them to be conserved epitopes.

### 1. Introduction

The unprecedented health crisis that is caused by the advent of the novel coronavirus, Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2), has challenged healthcare systems, economies and societal norms throughout the world. The development of vaccines and diagnostic tools for SARS-CoV-2 heavily relies on identifying conserved epitopes across various virus strains. These epitopes are crucial because they remain unchanged despite the virus mutating, making them ideal targets for broad-spectrum vaccines and reliable diagnostic tests. Epitopes, also known as antigenic determinants, play a crucial role in the immune system's ability to recognize and respond to pathogens. The specific regions on antigens interact with immune system components, such as antibodies or T cells, to initiate the body's defense mechanisms. Binding of antibodies to epitopes is not random but is influenced by the structural flexibility of the target protein surface, with a preference for less flexible regions for more effective binding [1]. These interactions are fundamental in various applications, including the development of therapeutic antibodies, vaccines, and diagnostic assays [2,3].

Numerous studies have identified epitopes within the genome of the Wuhan strain of SARS-CoV-2. These epitopes are located in various structural proteins of the virus, including the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins [4]. The evolution of these epitopes has since been demonstrated in subsequent variants with distinct mutations [5,6]. This has been linked to decreased vaccine efficacy and the reduced effectiveness of neutralizing antibody drugs emphasizing the need for continuous epitope monitoring to enhance immunization strategies [7]. Recent studies have also highlighted the importance of B-cell epitopes in SARS-CoV-2, which are critical for eliciting neutralizing antibodies and have been explored as potential candidates for COVID-19 vaccines [8].

Investigations into lipopeptides derived from conserved epitopes across the S, N, and M proteins have shown promising results in inducing cross-protective immunity, supporting their use in vaccine development [9]. Similarly, conserved epitopes from all four structural proteins of SARS-CoV-2 have been utilized to design a vaccine construct, demonstrating the potential of these epitopes in eliciting an immune response against various SARS-CoV-2 variants [10]. Vaccines targeting conserved epitopes on non-spike structural proteins have shown the potential in providing broad and long-lasting immunity, emphasizing the importance of these regions in universal vaccine development [11]. The conserved regions in the S1 and S2 subunits of the spike protein are also

* Corresponding author at: Moi Teaching and Referral Hospital, Kenya.
*E-mail address:* eliusmbogori@mtrh.or.ke (E. Mbogori).

https://doi.org/10.1016/j.nexres.2025.100215
3050-4759/

considered immunogenic, further highlighting the role of conserved epitopes in vaccine development [12]. Additionally, computational approaches have been employed to predict B-cell epitopes, which are essential for designing effective vaccines and understanding immune responses [13,14].

The nucleocapsid (N) protein, abundant early during infection, has been identified as a key target for COVID-19 diagnosis, with conserved epitopes playing a significant role in the efficacy of antigen testing across SARS-CoV-2 variants [15]. Moreover, diagnostic tools leveraging conserved spike protein epitopes have demonstrated their utility in identifying SARS-CoV-2 infections, reinforcing the value of these stable regions [16]. Studies have also explored the role of epitopes in the development of rapid diagnostic tests, emphasizing their stability and specificity across different viral strains [17]. Collectively, these studies affirm the critical role of conserved SARS-CoV-2 epitopes in the development of effective vaccines and diagnostic tools.

Bioinformatic tools have been used to predict potential epitopes for SARS-CoV-2 by compiling known epitope sites from other coronaviruses, mapping these regions in SARS-CoV-2 sequences, and predicting likely epitopes [18]. This aligns with the work of Agarwal et al. (2022), which leveraged known epitope sequences from various coronaviruses to identify additional antigen regions in SARS-CoV-2 genomes [31] ]. BLASTp is a pivotal bioinformatics tool for comparing protein sequences to unveil regions of similarity, aiding in understanding evolutionary relationships and functional conservation [19,20]. In epitope determination, BLASTp plays a crucial role in searching for homologous proteins [21] that harbor conserved epitopes shared between the Wuhan SARS-CoV-2 genome and recent clades. Recent advancements in epitope prediction have also utilized machine learning and structural bioinformatics to enhance the accuracy of epitope identification, further supporting vaccine and diagnostic development [22].

## 2. Methods

The SARS-CoV-2 genomes were isolated from 44 patients attending Moi Teaching and Referral Hospital (MTRH), Kenya, in December 2021. Written consent was obtained from all participants, and ethical approval was granted by the institutional research and ethics committee (Ref. IREC/036/2021 and ELD/MTRH/R&P/10/2/V.2/2010). MTRH is a national referral hospital authorized by the Kenyan Ministry of Health to test and treat COVID-19 patients from the North Rift and Western regions of Kenya.

Viral RNA was extracted using the DaAn Gene RNA extraction Kit (DaAnGene, China) and sequenced using the Illumina MiSeq sequencing platform (Illumina, California USA). The raw sequencing data were processed and assembled using the Wuhan-Hu-1 strain (NCBI accession: NC_045512.2) as a reference. The assembled genomes were deposited in the Global Initiative on Sharing All Influenza Data (GISAID) under accession numbers EPI_ISL_19,004,000 to EPI_ISL_19,004,016 and EPI_ISL_19,004,981 to EPI_ISL_19,005,007.

These genomes were then subjected to protein translation and analysis to identify conserved epitopes from the Wuhan SARS-CoV-2 genome. Initially, the genomes underwent Open Reading frame (ORF) identification using ORF Finder, followed by careful annotation to determine their potential functions. Subsequently, protein-coding sequences were translated into all six frames using EXPASY to ensure comprehensive coverage. The six-frame translation approach was employed to account for potential sequencing errors or frameshift mutations that could affect ORF prediction. A BlastP analysis was then conducted, comparing these translated SARS-CoV-2 proteins with a database of known proteins from National Center for Biotechnology Information (NCBI) to identify the best protein matches across the six frames.

Epitopes from the Wuhan SARS-CoV-2 genome (NCBI Taxon ID: 2,697,049) were obtained from the Immune Epitope Database (IEDB) v2.28. The search criteria included linear epitopes with positive B-cell and T-cell assay results, focusing on human hosts. Major Histocompatibility Complex (MHC) class I and II epitopes were downloaded separately and organized into three FASTA files based on immune cell type. Each epitope FASTA file was subjected to BLASTp analysis against the translated SARS-CoV-2 proteins. Epitopes showing 100 % homology were recorded as conserved.

## 3. Results

A total of 12,285 epitopes derived from the Wuhan SARS-CoV-2 genome were acquired from the IEDB database. Of these, 5154 epitopes were conserved in the genomes isolated from Moi Teaching and Referral Hospital, Kenya in December 2021. The conserved epitopes were: B-cell linear epitopes: 4087 out of 8597 (47.5 %); T-cell MHC class I epitopes: 374 out of 2326 (16.1 %); T-cell MHC class II epitopes: 693 out of 1362 (50.9 %). The conserved epitopes were primarily located in the membrane (M), nucleocapsid (N), and spike (S) proteins, as well as various ORFs (Table 1). Notably, no conserved epitopes were identified in the envelope (E), ORF7b, or ORF9b genes. Each epitope sequence is shown in the appendix

Fig. 1 below graphically compares the total number of epitopes found in the Wuhan genome and those seen in the later MTRH genome. These conserved epitopes are attached in the appendix.

## 4. Discussion

The identification and characterization of epitopes derived from the SARS-CoV-2 genome are critical for understanding the immune response to COVID-19 and for developing effective vaccines and diagnostic tools. In this study, we analyzed 12,285 epitopes from the Wuhan SARS-CoV-2 reference genome and identified 5154 (42 %) conserved

**Table 1**
Distribution of conserved epitopes across SARS-CoV-2 proteins.

| | Bcell | | Tcell-MHC1 | | Tcell-MHC2 | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Wuhan | MTRH | Wuhan | MTRH | Wuhan | MTRH | **Wuhan** | **MTRH** |
| Envelope | 41 | 0 | 33 | 0 | 20 | 0 | **94** | **0** |
| Membrane | 198 | 165 | 134 | 33 | 101 | 77 | **433** | **275** |
| Nucleoprotein | 784 | 405 | 189 | 45 | 198 | 134 | **1171** | **584** |
| ORF1ab | 4598 | 2487 | 1122 | 177 | 327 | 249 | **6047** | **2913** |
| ORF3a | 181 | 81 | 85 | 8 | 26 | 8 | **292** | **97** |
| ORF6 | 33 | 33 | 15 | 5 | 10 | 8 | **58** | **46** |
| ORF7a | 59 | 20 | 30 | 2 | 12 | 0 | **101** | **22** |
| ORF7b | 2 | 0 | 8 | 0 | 1 | 0 | **11** | **0** |
| ORF8 | 97 | 52 | 28 | 6 | 26 | 9 | **151** | **67** |
| ORF9b | 1 | 0 | 10 | 0 | 1 | 0 | **12** | **0** |
| Spike | 2603 | 844 | 672 | 98 | 640 | 208 | **3915** | **1150** |
| Total | **8597** | **4087** | **2326** | **374** | **1362** | **693** | **12,285** | **5154** |

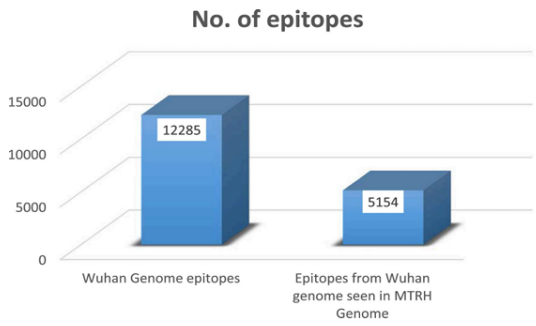



**Fig. 1.** Comparison of epitope conservation between the Wuhan and MTRH genomes.

epitopes in genomes isolated from Moi Teaching and Referral Hospital (MTRH) in Kenya. These conserved epitopes were primarily located in the spike (S), nucleocapsid (N), and membrane (M) proteins, which are known to play key roles in viral pathogenesis and immune recognition.

Our findings are consistent with previous studies that have highlighted the importance of conserved epitopes in these structural proteins. For instance, Patel et al. (2023) demonstrated that lipopeptides derived from conserved epitopes in the S and N proteins induce cross-protective immunity in mice, underscoring their potential for vaccine development [9]. Similarly, Wang et al. (2023) identified conserved epitopes in the S1 and S2 subunits of the spike protein, further supporting their utility in designing universal vaccines [11].

In addition to these findings, recent studies have emphasized the conservation of epitopes across major SARS-CoV-2 variants, including Delta and Omicron. Rodrigues-da-Silva et al. (2023) identified conserved B-cell epitopes in the nucleocapsid protein, which remain stable despite the genetic variability observed in these variants [15]. Likewise, Tarigan et al. (2023) characterized linear epitopes in the spike protein that are conserved across multiple variants, highlighting their potential for use in diagnostic assays [16].

The conservation of epitopes, particularly in structural proteins like the spike (S) and nucleocapsid (N) proteins, is a key finding of this study. The spike protein, which is the primary target for neutralizing antibodies, accounted for the largest proportion of conserved B-cell epitopes (844 out of 2603). This aligns with previous studies that have highlighted the spike protein as a critical target for vaccine development due to its role in viral entry and its immunodominant epitopes [18,23]. Similarly, the nucleocapsid protein, which is involved in viral replication and assembly, also showed significant epitope conservation (405 out of 784 B-cell epitopes). These findings are consistent with research demonstrating that conserved epitopes in these proteins can elicit robust immune responses, even against emerging variants [24]. Moreover, recent advances in phage-based vaccines, such as those targeting the spike protein, offer promising avenues for leveraging conserved epitopes. Studies by Markosian et al. (2022) and Staquicini et al. (2021) have demonstrated the efficacy of phage-displayed epitopes in eliciting robust immune responses against SARS-CoV-2 variants [25,26]. Our findings complement these efforts by identifying conserved regions that could be incorporated into phage-based vaccine platforms.

The conservation of T-cell epitopes, particularly MHC class II epitopes (693 out of 1362), further underscores the potential for cross-reactive immunity. T-cell responses are crucial for viral clearance and long-term immunity, and the identification of conserved T-cell epitopes supports the development of vaccines that can induce broad and durable immune responses [27]. In contrast, MHC class I epitopes T-cell epitopes were less conserved, showing a significant reduction in the MTRH genomes (374 out of 2326). This could reflect selective pressures from host immune responses or genetic drift, as MHC class I epitopes are critical for cytotoxic T-cell responses that eliminate infected cells [27]. The lower conservation of T-cell epitopes may have implications for cross-protective immunity, as mutations in these regions could enable immune evasion.

While this study provides valuable insights into the conservation of SARS-CoV-2 epitopes, it has several limitations. First, the analysis is based on in silico predictions and database-derived epitopes, which may not fully reflect the immunogenicity of these epitopes in vivo. Laboratory validation, including peptide binding assays and functional studies, is necessary to confirm the immunogenicity and protective potential of the identified epitopes. Second, the study focuses on a limited dataset of 44 genomes from a single geographic region, Kenya and a specific time point, December 2021, which may limit the generalizability of the findings. SARS-CoV-2 is known to exhibit regional variability due to differences in transmission dynamics, host immunity, and viral evolution [28]. Future studies with a broader geographic and temporal scope to capture the full spectrum of viral diversity.

Despite these limitations, the findings of this study have important implications for vaccine development and immune surveillance. The identification of conserved epitopes in the spike and nucleocapsid proteins supports the development of multiepitope vaccines that can provide broad protection against multiple SARS-CoV-2 variants [29]. The use of bioinformatics tools, such as machine learning, to predict epitope reactivity across variants further enhances the potential for tailored vaccine strategies [30]. Additionally, the conservation of key epitopes in the MTRH genomes suggests that vaccines targeting these regions may be effective in the Kenyan population and other regions with similar viral lineages.

Future research should focus on validating the immunogenicity of the identified epitopes through experimental studies, including peptide binding assays, T-cell activation assays, and animal models. Additionally, longitudinal studies tracking epitope conservation and immune responses over time can provide insights into the durability of vaccine-induced immunity and the potential for immune escape. Finally, the integration of genomic and immunological data from diverse populations will be critical for developing globally effective vaccines and diagnostic tools.

## 5. Conclusion

In this study, we identified 5154 conserved epitopes from the Wuhan SARS-CoV-2 genome in genomes isolated from Kenya. These epitopes, located primarily in the M, N, and S proteins, offer potential targets for vaccine and diagnostic tool development. However, the limitations of our study, including the lack of experimental validation and the narrow geographic scope of the dataset, highlight the need for further research. Future studies should focus on validating the immunogenicity of these epitopes and expanding the analysis to include a broader range of SARS-CoV-2 variants.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The SARS-CoV-2 genome sequences analysed during the current study are available in the Global Initiative on Sharing All Influenza Data (GISAID) repository database with accession numbers: EPI_IS-

L_19,004,000 to EPI_ISL_19,004,016; and EPI_ISL_19,004,981 to EPI_ISL_19,005,007 (https://gisaid.org/).

The Wuhan SARS-CoV-2 genome was gotten from National Center for Biotechnology Information (NCBI) virus repository with accession number MN908947.2 (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202,%20-taxid:2,697,049&ids=MN908947)

The epitopes Wuhan Coronavirus are available in the Immune Epitope Database (IEDB) repository with NCBI Taxon no 2,697,049 (https://www.iedb.org/)


## Funding

The research was funded by Moi Teaching and Referral Hospital Intra-mural Research Fund


## Authors' information

Elius Mbogori is a Researcher and Principal Medical Laboratory Officer at Moi Teaching and Referral Hospital

Stanslaus Musyoki is a Senior lecturer in South Eastern Kenya University in the Department of Medical Laboratory Science

Richard Biegon is a Senior Lecturer of Immunology in Moi University

Kirtika Patel is a Professor of Immunology in Moi University

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used SciSpace for grammar and references. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## CRediT authorship contribution statement

**Elius Mbogori:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Stanslaus Musyoki:** Writing – review & editing, Methodology, Conceptualization. **Richard Biegon:** Writing – review & editing, Validation, Supervision. **Kirtika Patel:** Writing – review & editing, Validation, Supervision, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


## Acknowledgments

Binhua Liang, University of Manitoba; Elijah Songok, Kenya Medical Research Institute; Caroline Wangui, Moi Teaching and Referral Hospital; Management of Moi Teaching and Referral Hospital


## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.nexres.2025.100215.

## References


[1] D.-G. Kim, Y. Choi, H.S. Kim, Epitopes of protein binders are related to the structural flexibility of a target protein surface, J. Chem. Inf. Model. 61 (2021) 2099–2107, https://doi.org/10.1021/ACS.JCIM.0C01397.

[2] Y. Kaiguang, S. Li, L. Liu, C. Yuwan, W. Zhou, P. Jiaqi, et al., Epitope imprinting technology: progress, applications, and perspectives toward artificial antibodies, Adv. Mater. 31 (2019) 1902048, https://doi.org/10.1002/ADMA.201902048.

[3] Q. Ul-Ain, R. Kandler, D. Gillespie, A. Nag, Chemical epitope targeting: review of a novel screening technology, J. Med. Chem. Drug Des. 1 (2018) 4, https://doi.org/10.16966/2578-9589.110.

[4] Y.D. Devi, H.B. Goswami, S. Konwar, C. Doley, A. Dolley, A. Devi, et al., Immunoinformatics mapping of potential epitopes in SARS-CoV-2 structural proteins, PLoS One (2021), https://doi.org/10.1371/journal.pone.0258645.

[5] Y. Chen, S. Li, W. Wu, S. Geng, M. Mao, Distinct mutations and lineages of SARS-CoV-2 virus in the early phase of COVID-19 pandemic and subsequent 1-year global expansion, J. Med. Virol. 94 (2022) 2035–2049, https://doi.org/10.1002/jmv.27580.

[6] P. Guérin, N. Yahi, H. Chahinian, J.-M. Sabatier, J. Fantini, F. Azzaz, Distinct evolution of infection-enhancing and neutralizing epitopes in the spike protein of SARS-CoV-2 variants (from alpha to omicron): a structural and molecular epidemiology study, Res. Sq. (2021), https://doi.org/10.21203/RS.3.RS-1054360/V1.

[7] X. Zhang, J.C. Wu, Y. Luo, Y. Wang, Y. Wu, X. Xu, et al., CovEpiAb: a comprehensive database and analysis resource for immune epitopes and antibodies of human coronaviruses, Brief Bioinform. 25 (2024), https://doi.org/10.1093/bib/bbae183.

[8] A.N.M. Ansori, R.V. Nidom, M.K.J. Kusala, S. Indrasari, I. Normalina, A.N. Nidom, et al., Viroinformatics investigation of B-cell epitope conserved region in SARSCoV-2 lineage B.1.1.7 isolates originated from Indonesia to develop vaccine candidate against COVID-19, J. Pharm. Pharmacogn. Res. 9 (2021) 766–779, https://doi.org/10.56499/jppres21.1101_9.6.766.

[9] R.K. Patel, B. Agrawal, Mucosal immunization with lipopeptides derived from conserved regions of SARS-CoV-2 antigens induce robust cellular and cross-variant humoral immune responses in mice, Front. Immunol. 14 (2023), https://doi.org/10.3389/fimmu.2023.1178523.

[10] A.R. Yaseen, S. Muhammad, A.S. Qadri, A. Asghar, I. Arshad, D.M. Khan, Development of conserved multi-epitopes based hybrid vaccine against SARS-CoV-2 variants, Immunoinform. Approach (2023), https://doi.org/10.21203/rs.3.rs-2919803/v1.

[11] C.Y. Wang, W.-J. Peng, B. Kuo, Y. Ho, M.-S. Wang, Y.-T. Yang, et al., Toward a pan-SARS-CoV-2 vaccine targeting conserved epitopes on spike and non-spike proteins for potent, broad and durable immune responses, PLoS Pathog. 19 (2023) https://doi.org/10.1371/journal.ppat.1010870, e1010870–e1010870.

[12] T.A. Olukitibi, Z. Ao, B.M. Warner, R. Unat, D. Kobasa, X. Yao, Significance of conserved regions in coronavirus spike protein for developing a novel vaccine against SARS-CoV-2 infection, Vaccines. (Basel) 11 (2023) 545, https://doi.org/10.3390/vaccines11030545.

[13] M. Fahmi, V.D. Kharisma, A.N.M. Ansori, M. Ito, Retrieval and investigation of data on SARS-CoV-2 and COVID-19 using bioinformatics approach, Adv. Exp. Med. Biol. 1318 (2021) 839–857 https://doi.org/10.1007/978-3-030-63761-3_47, Springer, Cham.

[14] Z.M. Aqsha, M.A. Dharmawan, V.D. Kharisma, A.N.M. Ansori, N.I. Sumantri, Reverse vaccinology analysis of B-cell epitope against Nipah virus using fusion protein, Jordan J. Pharm. Sci. 16 (2023) 499–507, https://doi.org/10.35516/jjps.v16i3.1602.

[15] R.N. Rodrigues-da-Silva, F.P. Conte, G.D.P. Silva, S.N. Kuriyama, A.A. Neto, J.C. Lima-Junior, Identification of B-cell linear epitopes in the nucleocapsid (N) protein B-cell linear epitopes conserved among the main SARS-CoV-2 variants, Viruses 15 (2023) 923, https://doi.org/10.3390/v15040923.

[16] S. Tarigan, N.I. Dharmayanti, D. Sugiartanti, R. Putri, N.H. Andriani, et al., Characterization of two linear epitopes SARS CoV-2 spike protein formulated in tandem repeat, PLoS One 18 (2023) https://doi.org/10.1371/journal.pone.0280627, e0280627–e0280627.

[17] T. Widjaja, A.N.M. Ansori, V.D. Kharisma, I. Faizal, Y. Antonius, J.P. Trinugroho, et al., B-cell conserved epitope screening and in silico cloning of envelope glycoprotein from Ebola virus (EBOV) for vaccine candidate construction, Indones. J. Pharm. 34 (2023) 193–204, https://doi.org/10.22146/ijp.4197.

[18] A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2, Cell Host. Microbe 27 (2020) 671–680 https://doi.org/10.1016/j.chom.2020.03.002, e2.

[19] Y. Mahlich, M. Steinegger, B. Rost, Y. Bromberg, HFSP: high speed homology-driven function annotation of proteins, Bioinformatics 34 (2018) i304–i312 https://doi.org/10.1093/bioinformatics/bty262, Bioinformatics.

[20] Y.-C. Li, Y-C. Lu, BLASTP-ACC: parallel architecture and hardware accelerator design for BLAST-based protein sequence alignment, IEEe Trans. Biomed. Circ. Syst. 13 (2019) 1771–1782, https://doi.org/10.1109/TBCAS.2019.2943539.

[21] N. Bhat, E.B. Wijaya, A.A. Parikesit, Use of the "DNAChecker" algorithm for improving bioinformatics research, Theory Comput. Syst. √Math. Syst. Theory 23 (2019) 72–77, https://doi.org/10.7454/MST.V23I2.3488.

[22] A.N.M. Ansori, V.D. Kharisma, A. Fadholly, M.R. Tacharina, Y. Antonius, AA. Parikesit, Severe acute respiratory syndrome coronavirus-2 emergence and its treatment with alternative medicines: a review, Res. J. Pharm. Technol. 14 (2021) 5551–5557, https://doi.org/10.52711/0974-360X.2021.00967.

[23] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell (2020), https://doi.org/10.1016/j.cell.2020.02.058.

[24] A. Tarke, J. Sidney, N. Methot, Y. Zhang, J.M. Dan, B. Goodwin, et al., Negligible impact of SARS-CoV-2 variants on CD4 + and CD8 + T cell reactivity in COVID-19 exposed donors and vaccinees, BioRxiv Prepr. Serv. Biol. (2021) https://doi.org/10.1101/2021.02.27.433180, 2021.02.27.433180.

[25] D.I. Staquicini, F.H.F. Tang, C. Markosian, V.J. Yao, F.I. Staquicini, E. Dodero-

Rojas, et al., Design and proof of concept for targeted phage-based COVID-19 vaccination strategies with a streamlined cold-free supply chain, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) e2105739118, https://doi.org/10.1073/pnas.2105739118.

[26] C. Markosian, D.I. Staquicini, P. Dogra, E. Dodero-Rojas, J.H. Lubin, F.H.F. Tang, et al., Genetic and structural analysis of SARS-CoV-2 spike protein for universal epitope selection, Mol. Biol. Evol. 39 (2022), https://doi.org/10.1093/molbev/msac091.

[27] A. Sette, S. Crotty, Adaptive immunity to SARS-CoV-2 and COVID-19, Cell 184 (2021) 861–880, https://doi.org/10.1016/j.cell.2021.01.007.

[28] Rambaut A., Loman N., Pybus O., Barclay W., Barrett J., Carabelli A., et al. on behalf of COVID-19 Genomics Consortium UK (CoG-UK). 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Rep Virol Org Website Https//Virological Org/t/Preliminary-Genomic-Characterisation-of-an-Emergent-Sars-Cov-2-Lineage-in-the-Uk-Defined-by-a-Novel-Set-of-Spike-Mutations/563 2021:2019.

[29] J.S. Tregoning, K.E. Flight, S.L. Higham, Z. Wang, B.F. Pierce, Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape, Nat. Rev. Immunol. 21 (2021) 626–636, https://doi.org/10.1038/s41577-021-00592-1.

[30] P. Wang, M.S. Nair, L. Liu, S. Iketani, Y. Luo, Y. Guo, et al., Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7, Nature 593 (2021) 130–135, https://doi.org/10.1038/s41586-021-03398-2.

[31] A Agarwal, KL Beck, S Capponi, M Kunitomi, G Nayar, E Seabolt, et al., Predicting Epitope Candidates for SARS-CoV-2, Viruses 14 (2022) 1837. https://doi.org/10.3390/V14081837/S1.

## App: 4.2 Manuscript awaiting publication

### App: 4.2.1 SARS-COV-2 mutations in North Rift, Kenya

**Authors**
Elius Mbogori1,2,3, Kelvin Thiongo3, Harrison Yunying Deng6, Winfrida Cheriro2, Stanslaus Musyoki7, Richard Biegon2, Damaris Matoke-Muhia3, Kirtika Patel2, Binhua Liang 4,5 and Elijah Songok3,
Affiliation:
1 Laboratory Department, Moi Teaching and Referral Hospital
2 School of Medicine, Moi University
3 Kenya Medical Research Institute
4 Vaccines & Therapeutics, Medical and Scientific Affairs, National Microbiology Laboratory Branch, Public Health Agency of Canada, Winnipeg, MB, CA;
5 Department of Biochemistry & Medical Genetics, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB, CA
6Faculty of Arts & Science, University of Toronto, Toronto, Ontario, Canada
7South Eastern Kenya University

**Abstract**
The rise of new SARS-CoV-2 mutations brought challenges and progress in the global fight against COVID-19. Mutations in spike and accessory genes affect transmission, vaccine efficacy, treatments, testing, and public health strategies. Monitoring emerging variants is crucial to halt re-emergency of the virus and spread. 44 nasopharyngeal/oropharyngeal swabs from Kenyan patients were sequenced with the Illumina platform. Galaxy's bioinformatic tools were used for genomic analysis. SARS-CoV-2 genome classification was done using PANGOLIN and mutation annotation with the COVID-19 Annotator tool. From this study, 5 clades of SARS-CoV-2 were shown to be circulating in the region. 38(86%) were BA.1.1; 2(5%) were BA.1.1.1; 1(2%) was BA.1; 1(2%) was BA.1.14 and 2(5%) were AY.46. In these clades we found 53 significant mutations of which 42 were non-synonymous, 10 synonymous, 7 deletions, 4 insertions and 2 extragenic. Out of the 42 non-synonymous mutations, 7 were exclusively found in symptomatic patients. Two new mutations, S:R214R, and NSP2:A555A, were also found. These findings add to the understanding of the SARS-CoV-2 virus future evolutionary in the region

Keywords:SARS-CoV-2; COVID-19; Spike protein; Reverse transcription; Pathogenesis; Vaccine efficacy; Complementary DNA; Reverse transcription-polymerase chain reaction.

1.0 Introduction

The emergence of novel mutations of SARS-CoV-2 introduced new challenges and breakthroughs in the global fight against the COVID-19 pandemic. These mutations in the viral genome, have impacted various aspects of the pandemic response [1,2]. Notably, mutations in the spike gene and other accessory genes have been identified, affecting transmission dynamics, vaccine efficacy, disease severity, therapeutic interventions, diagnostic testing, and public health measures [3,4]. Within the spike gene, several mutations such as R203K, G204R, N501Y, E484K, L452R, and E484Q have been

identified, which have the potential to increase viral replication, fitness, and immune evasion [3,4].

Additionally, the D614G mutation in the spike protein has been found to influence the pathogenesis of SARS-CoV-2 by enhancing spike protein trafficking to lysosomes, resulting in decreased spike expression on the cell surface [5]. Epistasis, the interaction between mutations, has also been studied extensively in the context of SARS-CoV-2. Research has shown that mutations in the receptor binding domain (RBD) of the spike protein can confer resistance to neutralizing antibodies, indicating complex interactions between mutations and antibody recognition [6]. Studies have identified sites under strong and weak epistatic constraints, highlighting their roles in viral replication and pathogenicity [7]. Furthermore, deep mutational scans have revealed epistatic shifts in the effects of mutations, suggesting that certain substitutions shape subsequent evolutionary changes [8]. Bernando et al. also found that substitutions in the RBD can cause epistatic shifts, influencing the virus's affinity for the ACE2 receptor [9].

In Kenya, the pandemic has been characterized by multiple waves, each with unique features. Initial waves occurred before the emergence of variants of concern (VoCs) and had lower attack rates [10]. Subsequent waves, driven by Alpha, Delta, and Omicron VoCs, exhibited higher attack rates [10]. The Omicron variant, notably first identified in South Africa, gained worldwide recognition due to its numerous mutations, particularly in the spike protein's receptor-binding domain [11]. The occurrence and rise of new strains of the SARS-CoV-2 virus have drawn global attention and caused concern. These changes can affect how easily the pathogen spreads, how severe the disease is, or its ability to escape immune system responses.

Every new mutation brings more uncertainty about what it could mean for the pandemic. Worries include the possibility of faster community spread and potential reduced efficacy of existing vaccines. The emergence of new mutations among general populations can be a cause of confusion and anxiety. This further fuels fear and panic, especially in places already severely hit by the virus. This emphasis on novel strains mirrors international anxiety to understand and tackle COVID-19 urgently. To keep the virus from spreading, there has to be constant monitoring alongside research as well as cooperation between countries

## 2.0 Methods

Swab samples were obtained from the oropharynx/nasopharynx of patients attending Moi Teaching and Referral Hospital in Eldoret, North Rift Kenya subsequent to obtaining ethics approval from Moi University/Moi Teaching and Referral Hospital Institutional Research Ethics committee (Ref. IREC/036/2021 and ELD/MTRH/R&P/10/2/V.2/2010). All study participants provided informed consent. The samples were then preserved in viral transport medium (VTM) and transported to the laboratory within a period of 6 hours, adhering strictly to the cold chain, and subsequently stored at -20°C. Using the DaAn Gene Detection kit for 2019-nCoV (DaAnGene, China) on the Rotor Gene Q real-time PCR machine, reverse transcription-polymerase chain reaction (RT-PCR) analysis was performed. Samples yielding positive results for SARS-CoV-2 with a cycle threshold (CT) value less than 30 were earmarked for sequencing.

Prior to sequencing, samples underwent preparation procedures. Initially, reverse transcription (RT) was carried out using the SuperScript™ IV Reverse Transcriptase kit

from Illumina, enabling the conversion of RNA into complementary DNA (cDNA). Subsequent to this, cDNA underwent amplification via polymerase chain reaction (PCR) targeting specific SARS-CoV-2 genes, facilitated by the NEBNext® Ultra™ II DNA Library Prep Kit. During library preparation, adapters tailored for Illumina sequencing were affixed to the amplified fragments. Cleanup procedures were then conducted utilizing the Agencourt AMPure XP system, followed by size selection to eliminate contaminants. Quantification of library concentration was performed using qubit to ensure an optimal loading concentration. Finally, sequencing was executed using the MiSeq sequencing platform (Illumina, California, USA), adhering strictly to the manufacturer's stipulated protocols.

After to sequencing, bioinformatics analyses were performed using the Galaxy platform. Initial quality assessments of sequencing data were conducted using FastQC. This was followed by sequence trimming using the trim sequence tool and alignment to the reference SARS-CoV-2 genome (NCBI accession number MN908947.2) using bowtie2 [12]. Duplicate sequences were removed via the Markduplicates tool. Variant calling was done using FreeBayes [13], with subsequent variant filtering utilizing VCFfilter to ensure criteria met for read depth (DP >30), allele frequency (AF >0.2), and variant quality (QUAL >30). The consensus tool was then employed to generate a fasta file combining variant information.

Further categorization of SARS-CoV-2 genome sequences into genetic lineages was achieved utilizing the Phylogenetic Assignment of Named Global Outbreak LiNeages (PANGOLIN). Annotation of sequences was carried out using the COVID-19 Annotation tool (http://giorgilab. unibo.it/coronannotator/) [14].

To elucidate novel mutations, sequences from previous mutations were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID) database. These mutations were annotated utilizing the COVID-19 Annotation tool (http://giorgilab. unibo.it/coronannotator/) [14] for comparison with mutations identified in the study.

## 3.0 Results

### 3.1 Sequencing data

The sequencing information from the set of 44 SARS-CoV-2 sequenced samples had an average of base coverage of 95%. Each base on average had a coverage depth of 282.73, with a maximum coverage depth reaching up to 1031.16. The sequences were deposited in GISAID and published with the accession numbers EPI_ISL_19004000 to EPI_ISL_19004016 and EPI_ISL_19004981 to EPI_ISL_19005007 .

### 3.2 SARS-CoV-2 Variants of concern

The sequencing outcomes of SARS-CoV-2 encompassed a cohort of 44 samples. Notably, the genomic analysis revealed the presence of two variants of concern: Delta (2 samples) and Omicron (42 samples). Within the Omicron variant, a nuanced genetic diversity was observed, with four discernible clades denoted as BA.1, BA.1.1, BA.1.1.1, and BA.1.14. Noteworthy is the dominance of the BA.1.1 clade, constituting a substantial majority at 86.4% of the total sequenced samples. These detailed genetic insights are graphically depicted in Figure 1, providing a comprehensive visual representation of the intricate distribution and prevalence of identified SARS-CoV-2 variants and their respective clades within the examined dataset.

Figure 1: Clade frequency: The clade naming is based on PANGO system. BA.1 represents the original Omicron variant. Each subsequent number denotes a subvariant of its preceding number. AY.46 is from the Delta variant.

Among the symptomatic patients, there were 16 males and 22 females. For symptomatic males, 14 cases had BA.1.1 variant. There is one case each for the BA.1 and AY.46 variants. No symptomatic males are reported for the BA.1.1.1 and BA.1.14 variants. For symptomatic females, 18 cases were due to the BA.1.1 variant. There were 2 cases of the BA.1.1.1 variant and one case each for the BA.1.14 and AY.46 variants. There are no symptomatic female cases for the BA.1 variant. Among the asymptomatic patients, there are 2 males and 4 females. All asymptomatic males and females were infected with the BA.1.1 variant. There were no asymptomatic cases for the BA.1.1.1, BA.1, BA.1.14, and AY.46 variants in either gender as figure 2 illustrates.

Figure 2: Distribution of the clades in relation to gender and symptoms

3.3 Frequency of mutations in the population

A total of 65 mutations were seen in various samples at varying frequencies. The "5'UTR" region showed a variant at position 241 with a frequency of 31.8%. In the NSP2 protein, a mutation A555A at position 2470 is present in 18.2% of subjects. Multiple mutations were observed in the NSP3 protein, including F106F, I388L, A889A, S1265, and A1892T with frequencies ranging from 11.4% to 97.7% as shown in the supplementary Table S2.

For the NSP4 protein, the T492I mutation at position 10029 was found in 97.7% of subjects. NSP6 showed variants like L105 and I189V, each with a frequency of 95.5%. NSP10 and NSP12b also displayed high-frequency mutations such as V57V and P314L, with the latter being present in all subjects (100%). NSP12b also includes N591N at a lower frequency of 11.4%. At NSP13, G170S mutation was observed in 63.6% of subjects. The NSP14 protein included the I42V mutation, present in 95.5% of subjects. For the Spike (S) protein, multiple mutations such as A67, T95I, I210, N211K, and others were observed, with frequencies ranging from 6.8% to 100%. Notably, the D614G variant in the Spike protein was found in all subjects (100%).

Other significant mutations included G339D, R346K, and S371L, with frequencies of 84.1%, 81.8%, and 22.7%, respectively. The ORF3a protein had a T64T mutation in 95.5% of subjects. Mutations in the Envelope (E) and Membrane (M) proteins, like T9I and Q19E, showed frequencies of 88.6% and 95.5%, respectively. The ORF6 and ORF7b regions included variants like M19M and L17L with frequencies of 95.5% and 61.4%, respectively. Out of these 65 mutations that were seen in the study, 42 were non-synonymous, 10 were synonymous, 7 were deletions, 4 were insertions, and 2 were extragenic mutations in the 3'UTR and 5'UTR regions as shown in figure 3.

Figure 3: Distribution of various mutations in respect to their type

Out of the 42 non- synonymous mutations, 7 mutations: NSP3:I388L; S:E484A; S:G446S; S:G496S; S:N211K; S:N440K; and S:N501Y were exclusively found in symptomatic patients. All the four insertions: S:I210; S:R214; S:S214; and S:V213, were found in symptomatic patients. Similarly, one synonymous mutation; S:R214R out of the 10 and one extragenic mutation: 5’UTR:241 out of the two were exclusively found in symptomatic patients. All the other mutations were seen in both symptomatic and asymptomatic patients.

3.4 Emerging mutations

Mutation data from previous waves in Kenya was analyzed alongside the mutations from this study. Two new mutations were noted: S:R214R accounting for 43.2% of cases within the study and NSP2:A555A, is identified in 18.2% of cases in the study group as shown in Table 2

Table 1: Newly mutations

| Name | %Cases in study |
|---|---|
| S:R214R | 43.2 |
| NSP2:A555A | 18.2 |

## 4.0 Discussion

The SARS-CoV-2 lineage has been changing over time around the world. In the beginning of the pandemic, the B.1 global parental lineage, present since the outset of the outbreak in the country, held sway, constituting 94% of all genomes in the first wave and 71% in the second wave. As the second wave progressed into the third wave, a variety of virus strains emerged, encompassing both Variants of Concern (VoCs) like B.1.351 (Beta) and B.1.1.7 (Alpha), as well as non-VoCs including B.1.3x, B.1.5x, B.1.525 (Eta), A, and A.23x. However, by the midpoint of the third wave, the Alpha variant took center stage, representing 74.9% of all sequenced genomes. The B.1.617.2 (Delta) variant and its AY.x sub-variant were identified in March 2021, rapidly surpassing other variants to become the predominant strain (99.3% of sequenced genomes) during forth wave [10]. This is study that was conducted during the fifth wave was dominated with clade BA.1.1 which is a subvariant BA.1, the clade that was dominant in the country during this wave.

The mutations S:G446S; S:E484A; S:G496S; S:N440K, and S:N501Y have been identified in symptomatic patients have been characterized in previous studies [15]. Studies have shown that such mutations play a significant role in the infectivity and immune evasion mechanisms of SARS-CoV-2 variants [16].

The presence of the S:G446S mutation in the spike protein of SARS-CoV-2 has been linked to enhanced infectivity and evasion of neutralizing antibody responses, particularly affecting the binding of class III neutralizing antibodies [17]. Additionally, the S:G446S mutation has been highlighted in the context of the Omicron variant, contributing to antibody evasion and potentially impacting the efficacy of monoclonal antibody treatments [18].

The S:E484A mutation has been identified as a significant variant, particularly in the context of the Omicron variant of concern. In combination with other mutations, S:E484A mutation, has been predicted to have profound effect to the virus; along with other mutations S:L432R, it has been shown to induce drastic changes in the folding pattern of the spike protein, affecting its structure and potentially its function [19]; alongside S371L, S373P, and S375F to impact antibody neutralization through conformational changes in the spike protein and alterations in surface charge distribution [20]; and with S:Y505H, it can destabilize the ACE2-RBD complex, impacting viral-host interactions [21]. It affects the recognition pathway of the spike protein to ACE2, influencing binding free energy and potentially modulating the infectivity of the virus [22].

The S:N440K mutation has been associated with immune evasion and increased infectivity. This mutation, has been linked to the ability of the virus to escape neutralizing antibodies and potentially impact the efficacy of antibody-based treatments [20,23]. Studies have shown that the S:N440K mutation, along with other key mutations like S:G446S, S:S477N,

and S:Q498R, can modulate the binding affinity of the spike protein to the ACE2 receptor, affecting viral entry and infectivity [24,25].

The S:N501Y mutation in the spike protein of SARS-CoV-2 has been extensively studied due to its implications for viral infectivity and immune evasion. This mutation, has been associated with increased binding affinity to the ACE2 receptor, potentially enhancing viral entry and infectivity [26,27]. Along with other key mutations it has been linked to higher infectivity of SARS-CoV-2 variants, altering the binding affinity and creating new inter-protein contacts that increase infectivity [28].

Studies have shown that the S:N501Y mutation plays a critical role in expanding the host range of SARS-CoV-2 variants, as it enhances the virus's ability to infect cells expressing mouse ACE2 receptors [29,30]. The presence of the S:N501Y has been associated with increased interaction forces between the spike protein and ACE2, contributing to the higher infectivity of certain variants [31].

The presence of the S:G496S mutation has been linked to enhanced infectivity and evasion of neutralizing antibody responses, particularly affecting the binding of class III neutralizing antibodies [32]. Additionally, the S:G496S mutation has been highlighted in the context of the Omicron variant, contributing to antibody evasion and potentially impacting the efficacy of monoclonal antibody treatments [33].

In summary, the mutations S:G446S, S:E484A, S:G496S, S:N440K, and S:N501Y in the spike protein of SARS-CoV-2 significantly contribute to the virus's enhanced infectivity and immune evasion. Each mutation plays a crucial role: S:G446S and S:E484A are particularly impactful in the context of the Omicron variant, altering antibody responses and the spike protein's structure. S:G496S and S:N440K facilitate immune escape and modulate the spike protein's functionality, while S:N501Y increases binding affinity to the ACE2 receptor and expands the virus's host range. Together, these mutations underline the importance of monitoring SARS-CoV-2's evolution to inform treatment and vaccination strategies. More research is however required to understand the importance of S:N211K and NSP3:I388L in the infectivity and pathogenicity of SARS-CoV-2

The most abundant new mutation in the region is S: R214R, a synonymous mutation, which account for 43.2% and of cases within the study. Previously, within the S1 subunit of the spike protein where this mutation lies was an insertion S:R214 [32] but progressively the mutation is changing to S:R214R. There is little information available for this change to a synonymous mutation. Synonymous mutations in SARS-CoV-2 have been found to be functionally and evolutionarily significant. In the past, synonymous mutations were viewed as phenotypically silent but recent studies indicate they can affect viral fitness with potential functional associations [34,35]. These types of mutations may alter RNA secondary structure leading to a higher probability of base pairing [36]. Moreover, there is a different frequency and regularity of mutations in various SARS-CoV-2 genes, where some genes have more non-synonymous mutations than others [37]. Better controlling pandemic requires an understanding of the impact of synonymous mutation on treatment and vaccine development vaccines [38]. Using millions of publicly accessible SARS-CoV-2 sequences could allow estimation of mutation effects which can provide crucial details for the assessment therapeutics that would be impervious to escape from newly derived variants.

## 5.0 Conclusion and recommendations

The emergence of novel mutations and variants of SARS-CoV-2 has posed significant challenges in the global battle against the COVID-19 pandemic. These mutations, particularly those affecting the spike gene, have far-reaching implications for various

aspects of the pandemic response, including transmission dynamics, vaccine efficacy, and disease severity. Variants such as Alpha, Delta, and Omicron underscore the need for continuous monitoring and international collaboration to effectively combat the spread of the virus.

To understand the genetic landscape of SARS-CoV-2, comprehensive genomic analyses are imperative. Patient swab samples are collected, preserved, and analyzed using advanced molecular techniques like RT-PCR and sequencing. Bioinformatic tools aid in the interpretation of sequencing data, allowing for the identification of novel mutations and variants. This holistic approach provides valuable insights into the virus's evolution and informs public health strategies to mitigate its impact.

The non-synonymous mutations G446S, E484A, G496S, N440K, and N501Y in the spike protein of SARS-CoV-2 seen exclusively in symptomatic patients are significant determinants of the virus's infectivity, immune evasion, and resistance to neutralizing antibodies. Each mutation contributes uniquely to the virus's ability to spread and evade immune responses. Understanding the implications of these mutations is essential for monitoring the virus's evolution and developing effective strategies to combat COVID-19. This knowledge is critical for adapting vaccination and treatment protocols to address the ongoing and dynamic challenges presented by SARS-CoV-2 variants. More studies are however required to understand the dynamics of S:N211K and NSP3:I388L in infectivity.

In conclusion, the ongoing evolution of SARS-CoV-2 underscores the need for vigilant surveillance, research, and international cooperation. By continuously monitoring the genetic landscape of the virus and understanding the implications of novel mutations, we can better adapt our public health strategies to control the spread of COVID-19 and mitigate its impact on global health and society.

Reference

1. Johnson BA, Zhou Y, Lokugamage KG, Vu MN, Bopp NE, Crocquet-Valdes PA, et al. Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. PLOS Pathog. 2022;18: e1010627–e1010627. doi:10.1371/journal.ppat.1010627

2. McGrath M, Xue Y, Dillen CA, Oldfield LM, Assad-Garcia N, Zaveri J, et al. SARS-CoV-2 variant spike and accessory gene mutations alter pathogenesis. Proc Natl Acad Sci U S A. 2022;119. doi:10.1073/pnas.2204717119

3. Hoter A, Naim HY. Biochemical Characterization of SARS-CoV-2 Spike RBD Mutations and Their Impact on ACE2 Receptor Binding. Front Mol Biosci. 2022;9. doi:10.3389/fmolb.2022.893843

4. Schürzinger H. Functional mutations of SARS-CoV-2: implications to viral transmission, pathogenicity and immune escape. Chin Med J (Engl). 2022;135: 1213–1222. doi:10.1097/cm9.0000000000002158

5. Khatri R, Siddqui G, Sadhu S, Maithil V, Vishwakarma P, Lohiya B, et al. Intrinsic D614G and P681R/H mutations in SARS-CoV-2 VoCs Alpha, Delta, Omicron and viruses with D614G plus key signature mutations in spike protein alters fusogenicity and infectivity. Med Microbiol Immunol. 2022;212: 103–122. doi:10.1007/s00430-022-00760-7

6. Witte L, Baharani VA, Schmidt F, Wang Z, Cho A, Raspe R, et al. Epistasis lowers the genetic barrier to SARS-CoV-2 neutralizing antibody escape. Nat Commun. 2023;14. doi:10.1038/s41467-023-35927-0

7. Alemrajabi M, Calix KM, Assis RC de. Epistasis-Driven Evolution of the SARS-CoV-2 Secondary Structure. J Mol Evol. 2022;90: 429–437. doi:10.1007/s00239-022-10073-1
8. Starr TN, Greaney AJ, Hannon WW, Loes AN, Hauser K, Dillen J, et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. Science (80- ). 2022;377: 420–424. doi:10.1126/science.abo7896
9. Bernando franky okto, Simpson A. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. 2022. doi:10.1101/2022.02.24.481899
10. Nasimiyu C, Matoke-Muhia D, Rono GK, Osoro E, Ouso DO, Mwangi JM, et al. Imported SARS-CoV-2 Variants of Concern Drove Spread of Infections across Kenya during the Second Year of the Pandemic. COVID. 2022;2: 586–598. doi:10.3390/covid2050044
11. Hoffmann M, Wong L, Arora P, Zhang L, Rocha C, Odle AE, et al. Omicron subvariant BA.5 efficiently infects lung cells. Nat Commun. 2023;14: null. doi:10.1038/s41467-023-39147-4
12. Kamohara M, Yagi S, Ishii Y, Nara H. Tie2 NOVEL ANTI-HUMAN TIE-2 ANTIBODY. 2017.
13. Richter F, Morton SU, Qi H, Kitaygorodsky A, Wang J, Homsy JG, et al. Whole Genome De Novo Variant Identification with FreeBayes and Neural Network Approaches. bioRxiv Genomics. 2020; 2020.03.24.994160. doi:10.1101/2020.03.24.994160
14. Mercatelli D, Holding AN, Giorgi FM. Web tools to fight pandemics: The COVID-19 experience. Briefings in Bioinformatics. 2021. pp. 690–700. doi:10.1093/bib/bbaa261
15. Dejnirattisai W, Huo J, Zhou D, Zahradník J, Supasa P, Liu C, et al. SARS-CoV-2 Omicron-B.1.1.529 Leads to Widespread Escape From Neutralizing Antibody Responses. Cell. 2022;185: 467-484.e15. doi:10.1016/j.cell.2021.12.046
16. Hunter Z, Liu X, Yang G, Zhou Y, Liu X, Cao Y, et al. The Genomic Landscape of Waldenström Macroglobulinemia Is Characterized by Highly Recurring MYD88 and WHIM-like CXCR4 Mutations, and Small Somatic Deletions Associated With B-Cell Lymphomagenesis. Blood. 2014;123: 1637–1646. doi:10.1182/blood-2013-09-525808
17. Qu P, Evans JP, Zheng Y-M, Carlin C, Saif LJ, Oltz EM, et al. Evasion of Neutralizing Antibody Response by the SARS-CoV-2 BA.2.75 Variant. 2022. doi:10.1101/2022.08.14.503921
18. Alcantara MC, Higuchi Y, Kirita Y, Matoba S, Hoshino A. Deep Mutational Scanning to Predict Escape From Bebtelovimab in SARS-CoV-2 Omicron Subvariants. Vaccines. 2023;11: 711. doi:10.3390/vaccines11030711
19. Saranya S, Sangavai C, B R, Chellapandi P. Studying Folding Kinetics of Omicron to Understand Its Hijack Mechanism in Human Host Cells. 2022. doi:10.21203/rs.3.rs-2392479/v1
20. Jawad B, Adhikari P, Podgornik R, Ching WY. Binding Interactions Between Receptor-Binding Domain of Spike Protein and Human Angiotensin Converting Enzyme-2 in Omicron Variant. J Phys Chem Lett. 2022;13: 3915–3921. doi:10.1021/acs.jpclett.2c00423
21. Zhang X, Chen J, Hong B, Xu H, Pei P, Chen L, et al. Pathogen-Host Adhesion Between SARS-CoV-2 Spike Proteins From Different Variants and Human ACE2 Probed at Single-Molecule and Single-Cell Levels. 2022. doi:10.1101/2022.06.05.493249

22. Peng C. The Recognition Pathway of the SARS-CoV-2 Spike Receptor-Binding Domain to Human Angiotensin-Converting Enzyme 2. Molecules. 2024;29: 1875. doi:10.3390/molecules29081875
23. Chen J, Wang R, Gilby NB, Wei G-W. Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance. J Chem Inf Model. 2022;62: 412–422. doi:10.1021/acs.jcim.1c01451
24. Stampfer SD, Goldwater M-S, Bujarski S, Regidor B, Zhang W, Feinstein AJ, et al. Severe Breakthrough COVID-19 With a Heavily Mutated Variant in a Multiple Myeloma Patient 10 Weeks After Vaccination. Clin Infect Pract. 2022;13: 100130. doi:10.1016/j.clinpr.2021.100130
25. Flifl AA, Singh RM, Sitotaw Y, Tarekegn TA. SARS-CoV-2 Genome Variants Epidemiology Surveillance in Ethiopia and Dynamic Mutational Changes of Structural S Spike Protein Through Computational Analysis. 2021. doi:10.21203/rs.3.rs-700282/v1
26. Gu H, Chen Q, Guan Y, He L, Fan H, Deng Y-Q, et al. Adaptation of SARS-CoV-2 in BALB/c Mice for Testing Vaccine Efficacy. Science (80- ). 2020;369: 1603–1607. doi:10.1126/science.abc4730
27. Khan A, Zia T, Suleman M, Khan T, Ali SS, Abbasi AA, et al. Higher Infectivity of the SARS-CoV-2 New Variants Is Associated With K417n/T, E484K, and N501Y Mutants: An Insight From Structural Data. J Cell Physiol. 2021;236: 7045–7057. doi:10.1002/jcp.30367
28. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, et al. Neutralization of SARS-CoV-2 Spike 69/70 Deletion, E484K and N501Y Variants by BNT162b2 Vaccine-Elicited Sera. Nat Med. 2021;27: 620–621. doi:10.1038/s41591-021-01270-4
29. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. 2020. doi:10.1101/2020.12.21.20248640
30. Tuncel G, Ergoren MC, Baddal B, Tulay P, Ozverel CS, Güler E, et al. Detection of SARS-CoV-2 N501Y Mutation Among SARS-CoV-2 Variants of Concern Circulating in Northern Cyprus. Future Virol. 2022;17: 523–528. doi:10.2217/fvl-2021-0273
31. Wang Q, Ye S, Zhou Z, Li J, Lv J, Hu B, et al. Key Mutations on Spike Protein Altering ACE2 Receptor Utilization and Potentially Expanding Host Range of Emerging SARS-CoV-2 Variants. J Med Virol. 2022;95. doi:10.1002/jmv.28116
32. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa. Nature. 2022;603: 679–686. doi:10.1038/s41586-022-04411-y
33. Wang L, Cheng G. Sequence Analysis of the Emerging SARS-CoV-2 Variant Omicron in South Africa. J Med Virol. 2021;94: 1728–1733. doi:10.1002/jmv.27516
34. Saldivar-Espinoza B, Garcia-Segura P, Macip G, Puigbò P, Cereto-Massagué A, Pujadas G, et al. The Mutational Landscape of SARS-CoV-2. Int J Mol Sci. 2023;24: 9072. doi:10.3390/ijms24109072
35. Renn O. Synonymous base substitution in SARS-CoV-2 EE.2 lineage furin arginine CGG–CGG codons. 2023. doi:10.21203/rs.3.rs-3139964/v1
36. Sun Q, Zeng J, Tang K, Long H, Zhang C, Zhang J, et al. Variation in synonymous evolutionary rates in the SARS-CoV-2 genome. Front Microbiol. 2023;14. doi:10.3389/fmicb.2023.1136386

37. Boon WX, Sia BZ, Ng CH. Prediction of the effects of the top 10 synonymous mutations from 26645 SARS-CoV-2 genomes. F1000Research. 2022;10: 1053. doi:10.12688/f1000research.72896.2
38. Bloom JD, Neher RA. Fitness effects of mutations to SARS-CoV-2 proteins. bioRxiv. 2023. doi:10.1101/2023.01.30.526314

6.0 Table S1:

Table S1: Sequencing data and Clade for each genome

| Sequence Name | Lineage | Bases with coverage | Average coverage depth | Maximum coverage depth |
|---|---|---|---|---|
| hCoV-19/Kenya/SME008_MTRH_S13/2021 | AY.46 | 97.8 | 265.6 | 1000 |
| hCoV-19/Kenya/SME026_MTRH_S37/2021 | AY.46 | 93.6 | 36.5 | 142 |
| hCoV-19/Kenya/SME028a_MTRH_S7/2021 | BA.1.1 | 82.6 | 28.7 | 247 |
| hCoV-19/Kenya/SME029_MTRH_S49/2021 | BA.1.1 | 99.1 | 386.4 | 1222 |
| hCoV-19/Kenya/SME030_MTRH_S61/2021 | BA.1.1 | 91.4 | 74.5 | 414 |
| hCoV-19/Kenya/SME036_MTRH_S85/2021 | BA.1.1 | 98.4 | 514.8 | 1510 |
| hCoV-19/Kenya/SME037_MTRH_S2/2021 | BA.1.1 | 99.78 | 427.7 | 1.175 |
| hCoV-19/Kenya/SME038_MTRH_S14/2021 | BA.1.1 | 97.9 | 167.4 | 609 |
| hCoV-19/Kenya/SME039_MTRH_S26/2021 | BA.1.1 | 86.4 | 32.7 | 177 |
| hCoV-19/Kenya/SME041_MTRH_S38/2021 | BA.1.1.1 | 99.8 | 524.3 | 1572 |
| hCoV-19/Kenya/SME042_MTRH_S50/2021 | BA.1.1 | 87.9 | 41.9 | 239 |
| hCoV-19/Kenya/SME043_MTRH_S62/2021 | BA.1.1 | 97.8 | 196.2 | 684 |
| hCoV-19/Kenya/SME045_MTRH_S74/2021 | BA.1.1 | 99.2 | 489.8 | 1519 |
| hCoV-19/Kenya/SME047_MTRH_S86/2021 | BA.1.1 | 97.4 | 431.7 | 1325 |
| hCoV-19/Kenya/SME048_MTRH_S3/2021 | BA.1.1 | 96.25 | 71.23 | 290 |
| hCoV-19/Kenya/SME066_MTRH_S15/2021 | BA.1.1 | 95.6 | 120.8 | 461 |
| hCoV-19/Kenya/SME071_MTRH_S27/2021 | BA.1.1 | 95 | 113 | 502 |
| hCoV-19/Kenya/SME072_MTRH_S39/2021 | BA.1.1.1 | 90.6 | 40.7 | 220 |
| hCoV-19/Kenya/SME077_MTRH_S51/2021 | BA.1.1 | 95.7 | 200.9 | 1041 |
| hCoV-19/Kenya/SME078_MTRH_S63/2021 | BA.1.1 | 96.3 | 382.7 | 2277 |
| hCoV-19/Kenya/SME091_MTRH_S75/2021 | BA.1.1 | 98.6 | 404.9 | 1438 |
| hCoV-19/Kenya/SME092_MTRH_S87/2021 | BA.1.1 | 99.8 | 2283.2 | 6175 |
| hCoV-19/Kenya/SME093_MTRH_S4/2021 | BA.1.1 | 98.9 | 613.3 | 2358 |
| hCoV-19/Kenya/SME095_MTRH_S16/2021 | BA.1.1 | 99.8 | 342.3 | 1298 |
| hCoV-19/Kenya/SME099_MTRH_S28/2021 | BA.1.1 | 93.3 | 99.4 | 609 |
| hCoV-19/Kenya/SME103_MTRH_S52/2021 | BA.1.1 | 90.6 | 98.9 | 610 |
| hCoV-19/Kenya/SME106_MTRH_S64/2021 | BA.1.14 | 83.5 | 77.3 | 898 |
| hCoV-19/Kenya/SME108_MTRH_S76/2021 | BA.1.1 | 95.6 | 378.8 | 2055 |
| hCoV-19/Kenya/SME109_MTRH_S88/2021 | BA.1.1 | 99.4 | 854.9 | 2741 |
| hCoV-19/Kenya/SME110_MTRH_S5/2021 | BA.1.1 | 87.8 | 36.6 | 200 |
| hCoV-19/Kenya/SME112_MTRH_S17/2021 | BA.1.1 | 99.7 | 606.6 | 1817 |
| hCoV-19/Kenya/SME116_MTRH_S41/2021 | BA.1.1 | 95.9 | 437.4 | 2126 |
| hCoV-19/Kenya/SME117_MTRH_S53/2021 | BA.1.1 | 97.4 | 122.9 | 395 |

| | | | | |
|---|---|---|---|---|
| hCoV-19/Kenya/SME118_MTRH_S65/2021 | BA.1.1 | 91 | 43.8 | 192 |
| hCoV-19/Kenya/SME119_MTRH_S77/2021 | BA.1.1 | 90.7 | 124.2 | 887 |
| hCoV-19/Kenya/SME123_MTRH_S6/2021 | BA.1.1 | 95.8 | 110.2 | 612 |
| hCoV-19/Kenya/SME124_MTRH_S18/2021 | BA.1.1 | 99.2 | 228.4 | 623 |
| hCoV-19/Kenya/SME125_MTRH_S29/2021 | BA.1.1 | 98.2 | 106.6 | 365 |
| hCoV-19/Kenya/SME126_MTRH_S30/2021 | BA.1.1 | 90.1 | 171.6 | 1453 |
| hCoV-19/Kenya/SME127_MTRH_S42/2021 | BA.1.1 | 99.4 | 340 | 1031 |
| hCoV-19/Kenya/SME128_MTRH_S54/2021 | BA.1.1 | 99.2 | 197.4 | 657 |
| hCoV-19/Kenya/SME133_MTRH_S66/2021 | BA.1 | 75.5 | 14 | 90 |
| hCoV-19/Kenya/SME135_MTRH_S78/2021 | BA.1.1 | 92.7 | 110.8 | 807 |
| hCoV-19/Kenya/SME137_MTRH_S90/2021 | BA.1.1 | 93.4 | 89.3 | 482 |

7.0 Table S2

Table S2: Mutation frequency for each mutated position in Percentage

| Ref pos | Protein | Ref var | Q var | Variant | Var name | No. of Subjects | Mutation frequency in Percentage |
|---|---|---|---|---|---|---|---|
| 241 | 5'UTR | C | T | 241 | 5'UTR:241 | 14 | 31.8 |
| 2470 | NSP2 | C | T | A555A | NSP2:A555A | 8 | 18.2 |
| 3037 | NSP3 | C | T | F106F | NSP3:F106F | 43 | 97.7 |
| 3881 | NSP3 | A | C | I388L | NSP3:I388L | 5 | 11.4 |
| 5386 | NSP3 | T | G | A889A | NSP3:A889A | 42 | 95.5 |
| 6513 | NSP3 | GTT | . | S1265 | NSP3:S1265 | 42 | 95.5 |
| 8393 | NSP3 | G | A | A1892T | NSP3:A1892T | 42 | 95.5 |
| 10029 | NSP4 | C | T | T492I | NSP4:T492I | 43 | 97.7 |
| 11286 | NSP6 | TGTCTGGTT | . | L105 | NSP6:L105 | 42 | 95.5 |
| 11537 | NSP6 | A | G | I189V | NSP6:I189V | 42 | 95.5 |
| 13195 | NSP10 | T | C | V57V | NSP10:V57V | 42 | 95.5 |
| 14408 | NSP12b | C | T | P314L | NSP12b:P314L | 44 | 100.0 |
| 15240 | NSP12b | C | T | N591N | NSP12b:N591N | 5 | 11.4 |
| 16744 | NSP13 | G | A | G170S | NSP13:G170S | 28 | 63.6 |
| 18163 | NSP14 | A | G | I42V | NSP14:I42V | 42 | 95.5 |
| 21762 | S | C | . | A67 | S:A67 | 42 | 95.5 |
| 21764 | S | A | . | A67 | S:A67 | 42 | 95.5 |
| 21767 | S | CATG | . | I68 | S:I68 | 42 | 95.5 |
| 21846 | S | C | T | T95I | S:T95I | 42 | 95.5 |
| 21987 | S | GTGTTTATT | . | G142 | S:G142 | 42 | 95.5 |
| 22193 | S | . | T | I210 | S:I210 | 19 | 43.2 |
| 22195 | S | T | G | N211K | S:N211K | 19 | 43.2 |
| 22197 | S | TA | GC | L212C | S:L212C | 19 | 43.2 |
| 22201 | S | . | AGC | S214 | S:S214 | 19 | 43.2 |
| 22202 | S | . | A | V213 | S:V213 | 19 | 43.2 |
| 22203 | S | . | A | R214 | S:R214 | 19 | 43.2 |
| 22204 | S | T | A | R214R | S:R214R | 19 | 43.2 |
| 22578 | S | G | A | G339D | S:G339D | 37 | 84.1 |
| 22599 | S | G | A | R346K | S:R346K | 36 | 81.8 |
| 22673 | S | TC | CT | S371L | S:S371L | 10 | 22.7 |
| 22679 | S | T | C | S373P | S:S373P | 10 | 22.7 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 22686 | S | C | T | S375F | S:S375F | 10 | 22.7 | |
| 22813 | S | G | T | K417N | S:K417N | 3 | 6.8 | |
| 22882 | S | T | G | N440K | S:N440K | 18 | 40.9 | |
| 22898 | S | G | A | G446S | S:G446S | 18 | 40.9 | |
| 22992 | S | G | A | S477N | S:S477N | 16 | 36.4 | |
| 22995 | S | C | A | T478K | S:T478K | 18 | 40.9 | |
| 23013 | S | A | C | E484A | S:E484A | 18 | 40.9 | |
| 23040 | S | A | G | Q493R | S:Q493R | 17 | 38.6 | |
| 23048 | S | G | A | G496S | S:G496S | 17 | 38.6 | |
| 23055 | S | A | G | Q498R | S:Q498R | 17 | 38.6 | |
| 23063 | S | A | T | N501Y | S:N501Y | 16 | 36.4 | |
| 23075 | S | T | C | Y505H | S:Y505H | 16 | 36.4 | |
| 23202 | S | C | A | T547K | S:T547K | 42 | 95.5 | |
| 23403 | S | A | G | D614G | S:D614G | 44 | 100.0 | |
| 23525 | S | C | T | H655Y | S:H655Y | 42 | 95.5 | |
| 23599 | S | T | G | N679K | S:N679K | 42 | 95.5 | |
| 23604 | S | C | A | P681H | S:P681H | 42 | 95.5 | |
| 23854 | S | C | A | N764K | S:N764K | 38 | 86.4 | |
| 23948 | S | G | T | D796Y | S:D796Y | 40 | 90.9 | |
| 24130 | S | C | A | N856K | S:N856K | 42 | 95.5 | |
| 24424 | S | A | T | Q954H | S:Q954H | 42 | 95.5 | |
| 24469 | S | T | A | N969K | S:N969K | 42 | 95.5 | |
| 24503 | S | C | T | L981F | S:L981F | 42 | 95.5 | |
| 25000 | S | C | T | D1146D | S:D1146D | 42 | 95.5 | |
| 25584 | ORF3a | C | T | T64T | ORF3a:T64T | 42 | 95.5 | |
| 26270 | E | C | T | T9I | E:T9I | 39 | 88.6 | |
| 26577 | M | C | G | Q19E | M:Q19E | 42 | 95.5 | |
| 26709 | M | G | A | A63T | M:A63T | 42 | 95.5 | |
| 27259 | ORF6 | A | C | M19M | ORF6:M19M | 42 | 95.5 | |
| 27807 | ORF7b | C | T | L17L | ORF7b:L17L | 27 | 61.4 | |
| 28271 | 3'UTR | A | T | 28271 | 3'UTR:28271 | 42 | 95.5 | |
| 28311 | N | C | T | P13L | N:P13L | 42 | 95.5 | |
| 28362 | N | GAGAACGCA | | . | E31 | N:E31 | 42 | 95.5 |
| 28881 | N | GGG | AAC | RG203KR | N:RG203KR | | 42 | 95.5 |