EXAMINATION OF COVID-19 OUTBREAK DYNAMICS, AND IDENTIFICATION OF BETTER VACCINE AND VIRAL DRUG TARGETS THROUGH GENOMIC ANALYSES OF SARS-CoV-2

A Thesis Submitted to
the Graduate School of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Biotechnology

by
Bahar Anıl GÜRBÜZ

July 2022
İZMİR

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Assoc. Dr. Efe SEZGİN. I am grateful that he did not withhold his support throughout the process, motivated me, and responded quickly to my every question. This thesis would not have been finished without his guidance.

Privately, I would like to thank my mother, nephew, and boyfriend for encouraging me. I am grateful to them for their constant motivation and support.

ABSTRACT

EXAMINATION OF COVID-19 OUTBREAK DYNAMICS, AND IDENTIFICATION OF BETTER VACCINE AND VIRAL DRUG TARGETS THROUGH GENOMIC ANALYSES OF SARS-CoV-2

Since the emergence of the SARS-CoV-2 virus, which causes the COVID-19 disease, it continues despite the application of many vaccines and drug treatments. It has continuously mutated, therefore the effect of vaccines and drug treatments has begun to decrease and a permanent solution has not been found.

The main hypothesis of thesis is that the conserved regions in the SARS-CoV-2 genome can be potential targets for new vaccines and drugs to eradicate the Covid-19 pandemic. In this study, a total of 807 sequences of the first emerging clades (L, O, S) of the SARS-CoV-2 human virus and its variants in the category of Variants of Concern (Alpha, Beta, Gamma, and Delta) were taken from different dates, and population genetic statistical tests were conducted. Human specific SARS-CoV-2 sequence analyses showed that the evolution of all viral proteins are primarily driven by negative selection. Interspecies tests using the RaTG13 Bat coronavirus, which has the most similar genome to the SARS-CoV-2 virus genome, showed that there was no fixed amino acid change divergence between the bat and human virus sequences for Membrane, Nsp8, Nsp10, and Nsp16, indicating high conservation.

Then, a list of the amino acid changes among the SARS-CoV-2 human clades and variants was prepared for Membrane, Nsp8, Nsp10, and Nsp16. Since the regions outside of these changes are the most conserved, the functions of the Membrane, Nsp8, Nsp10, and Nsp16 and and interactions with other viral proteins should be investigated as potential targets for new vaccines and drug treatments.

ÖZET

SARS-CoV-2'NİN GENOMİK ANALİZLERİ YOLUYLA COVID-19 SALGIN DİNAMİKLERİNİN İNCELENMESİ VE DAHA İYİ AŞI VE VİRAL İLAÇ HEDEFLERİNİN BELİRLENMESİ

COVİD-19 hastalığına neden olan SARS-CoV-2 virüsü ortaya çıkmasından bu yana üzerinden birçok aşı ve ilaç tedavisi uygulanmasına rağmen devam etmektedir. Süreç içerisinde sürekli olarak mutasyona uğramış bu nedenle bulunan aşı ve ilaç tedavilerinin etkisi azalmaya başlamıştır ve hala bu virüse karşı kalıcı bir çözüm bulunamamıştır.

Ana hipotezimiz, SARS-CoV-2 genomundaki korunmuş bölgelerin COVİD-19 pandemisini ortadan kaldırmak amacıyla geliştirilen aşı ve ilaç stratejileri için potansiyel hedefler olduğudur. Bu çalışmada, SARS-CoV-2 insan virüsünün ilk ortaya çıkan kladları (L, O, S) ve endişe verici varyant kategorisinde yer alan (Alfa, Beta, Gama ve Delta) varyantlarının farklı tarihlerden toplam 807 adet sekansı alınmış ve popülasyon genetik istatistiksel testleri uygulanmıştır. Uygulanan tür içi testler sonucunda genom proteinlerinde negatif seçilimin hakim olduğu görülmüştür. SARS-CoV-2 virüs genomuna en benzer genoma sahip olan RaTG13 yarasa koronavirüsü kullanılarak yapılan türler arası istatistiksel analizler sonucunda ise; Membrane, Nsp8, Nsp10 ve Nsp16 proteinlerinde amino asit değişimine yol açan değişimler bakımından sabitlenmiş ayrışma olmadığı görülmüştür.

Ardından Membrane, Nsp8, Nsp10 ve Nsp16 proteinlerinin SARS-CoV-2 insan virüsünün kladları ve varyantları arasında amino asit değişimine yol açan değişimlerinin listesi çıkarılmış ve bu değişimlerin dışında kalan bölgelerin en korunmuş bölgeler olmasından dolayı Membrane, Nsp8, Nsp10 ve Nsp16 proteinlerin fonksiyonları ve diğer proteinlerle etkileşimleri de araştırılarak potansiyel hedefler olabileceği ve bu bilgilere göre yeni aşı ve ilaç tedavileri ortaya konulabileceği belirtilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

Table **Page**

Table **Page**

LIST OF ABBREVIATION

A	Adenine
A	Alanine
ACE2	Angiotensin-Converting Enzyme 2
C	Cysteine
C	Cytosine
D	Aspartic Acid
E	Glutamic Acid
F	Phenylalanine
G	Glycine
G	Guanine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
MERS	Middle East Respiratory Syndrome
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
R_0	The Basic Reproduction Number
RBD	Receptor Binding Domain
RNA	Ribonucleic Acid
S	Serine
SARS	Severe Acute Respiratory Syndrome Coronavirus 1
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
T	Threonine
T	Thymine
TMPRSS2	Transmembrane Serine Protease 2
V	Valine

W	Tryptophan
Y	Tyrosine

CHAPTER 1

INTRODUCTION

Coronavirus (CoV) is a family of viruses with single-stranded, positively poled, enveloped RNA ranging in size from 80 to 120 nm, and are capable of becoming human pathogens. The name coronavirus began to be pronounced with the word "corona" which means crown in Latin, by the analogy of the rod-like extensions on its surface. CoV has taken place on *Nidovirales* order, *Cornidovirineae* suborder, *Coronaviridae* family, *Orthocoronavirinae* subfamily. *Orthocoronavirinae* is divided into four genus, *Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus* [1]. HCoV-229E and HCoV-NL63 are in the *Alphacoronavirus* genus, HCoV-OC43, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2 are in the genus *Betacoronavirus*.

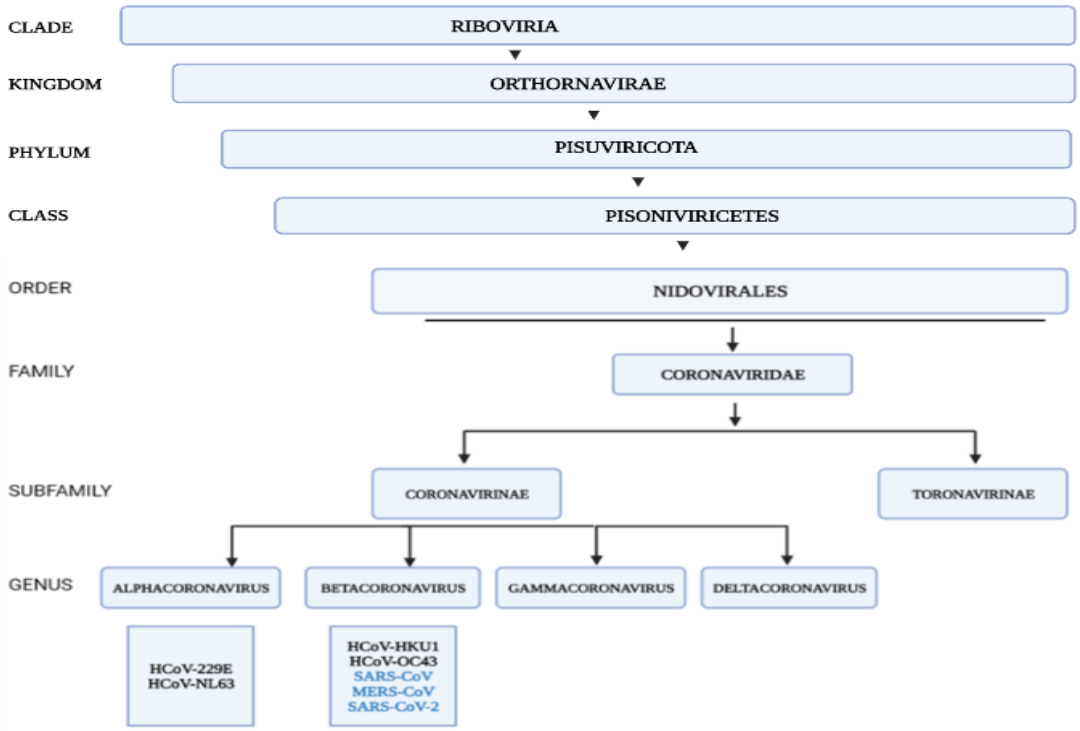

Figure 1. The Classification of SARS-CoV-2. (Created with Biorender)

Coronaviruses are found in animals and are transmitted to humans through intermediate hosts, recombination, or mutations. Human coronaviruses, which have increased public awareness in the last 20 years, were first described in the mid-1960s. Alpha and betacoronaviruses' all identified species are of animal origin, mainly from domestic animals, bats, or mice. HKU1, OC43, HCOV-229E, and HCOV-NL63, which are responsible for 35 percent of upper respiratory tract infections generally seen in winter in humans, some species cause severe acute syndromes such as SARS and MERS. SARS-CoV, which was first seen in Guandong province of China and spread to 29 countries, caused more than 8,000 people to have severe pneumonia and death [2]. SARS-CoV, which had caused widespread alarm, disappeared by the end of the year. It was determined that it was transferred from bats to the intermediate host, the civet, and subsequently from these animals to people after undergoing specific alterations. Then, In April 2012, MERS-CoV was first detected in Jordan in April 2012. In September 2012, patients were reported from Saudi Arabia. It has been determined that the intermediate host of MERS-CoV, which causes severe pneumonia cases as SARS-CoV, has spread to 27 countries and infected more than 2000 people, is dromedary camels [3].

1.1. Emergence

On December 31, 2019, cases of pneumonia occurred in Huanan Seafood Wholesale Market in Wuhan, located in the Hubei region of China, and then on February 11, a new coronavirus that caused the disease was identified and named SARS-CoV-2. When COVID-19 was declared a pandemic disease, there were 118,000 cases of COVID-19 in 114 countries as of March 11, 2020. The World Health Organization data show that the number of cases caused by Covid-19 worldwide has reached 250 million 715 thousand 502, and the number of deaths is 5 million 062 thousand 106. On March 3, 2020, the World Health Organization reported the estimated death rate due to the new Coronavirus infection as 3.4 percent [4]. With the emergence and transmission of SARS-CoV-2, various R_0 values were obtained at various periods and in different geographical locations. For pandemics, it is vital to know the reproduction number which shows the formation of the foreseen number of cases directly generated by a case in a population where all people are susceptible to infection [5]. As time goes by, the virus mutates, and different variants emerge. Along with this, the R_0 values began to differ. For the first known virus as an

ancestral strain, which originated in Wuhan, the R_0 value was revealed as 2.4–3.4 [6]. On the other hand, the Alpha variant and Delta variant were reported to have mutations that increase the rate of spread and have the R_0 values; of 4-5 and 5-8, respectively [7] [8].

Studies showed that the transmission of SARS-CoV-2 between people occurs via droplets. As a result of coughing, sneezing, and talking of infected people, it has been determined that the virus spreads to the environment, and this virus is transmitted. It is stated that particles suspended in the air can remain in the background for at least three hours. SARS-CoV-2 has also been detected in urine, feces, and blood samples. The virus takes roughly 5.2 days to incubate, and individuals begin to show symptoms around 11.5-15.5 days [9]. Apart from this, asymptomatic cases occur with no indications. Based on the most reports, fever is the most common symptom, followed by cough, dyspnea, muscular pains, weakness, expectoration, sore throat, headache, cold, lack of appetite, diarrhea, nausea-vomiting, and runny nose [10] [11].

SARS-CoV-2 intermediate host searches are still ongoing. The coronaviruses detected in pangolin studies have been confirmed to be highly similar to SARS-CoV-2. Regarding the similarity between Pangolin-CoV and SARS-CoV-2, the genetically closest species to SARS-CoV-2 is RaTG13, according to the data so far, with a sequence similarity of 96.2% [12].

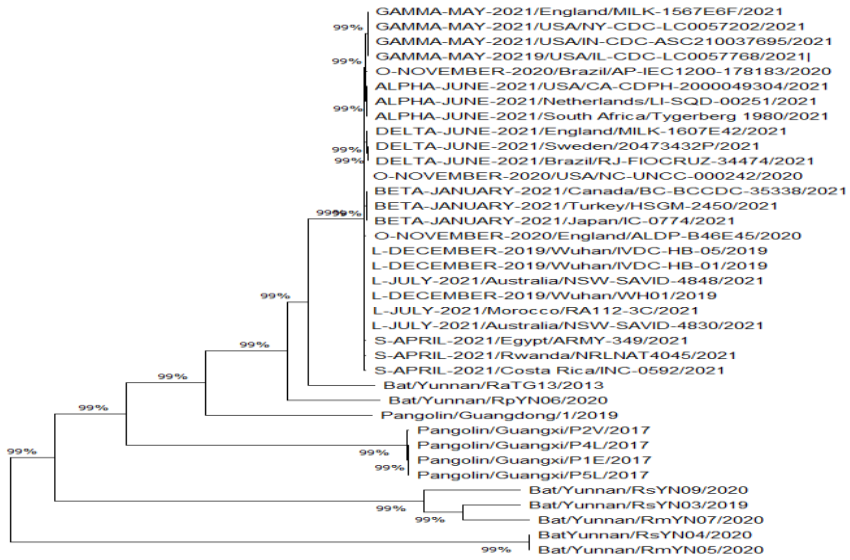

Figure 2. Phylogenetic tree based on maximum likelihood estimation with nucleotide sequences of genomes of coronaviruses using MEGAX

1.2. SARS-CoV-2 Genome

SARS-CoV-2 genome, with approximately 30k nucleotides, is the second-largest known RNA virus consisting of a single strand, and positive polarity genome. SARS-CoV-2 consists of a cap at the 5’ end and a non-coding sequence followed by a reader sequence of approximately 70 bases, 11 open reading frames encoding 27 proteins, and a containing the non-coding sequence at the 3’ end. Their genes are ORF1a, ORF1b, S, ORF3a, ORF3b, E, M, ORF6, ORF7a, ORF7b, ORF8, N, ORF10, respectively. Structural proteins are Spike, Membrane, Envelope, and Nucleocapside, and nonstructural proteins are involved in ORF1ab, and other proteins in the genome are accessory proteins.

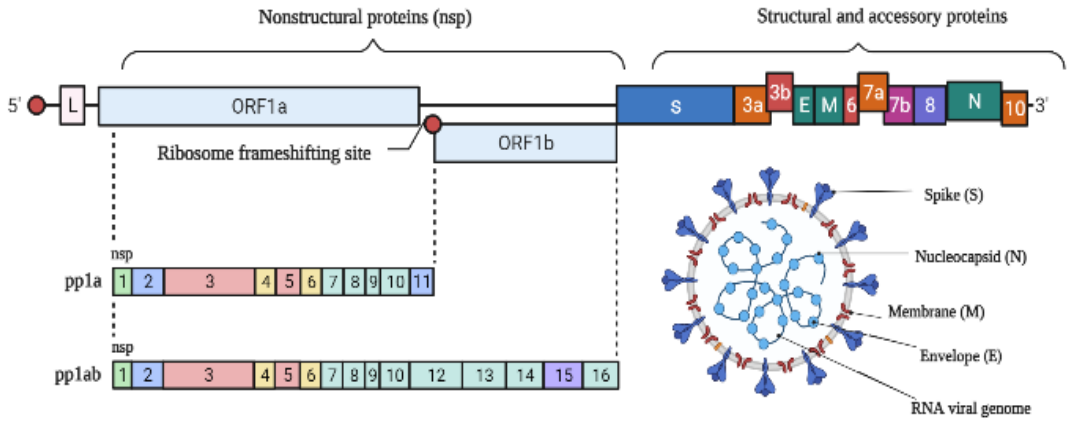

Figure 3. The genomic organization of SARS-CoV-2. (Created with BioRender)

1.2.1. Membrane

The membrane has three transmembrane segments which is critical in viral production and also known as the most abundant structural protein of coronaviruses. Membrane builds virions, enhances membrane arch, binds to nucleocapsid, and stabilizes it. As a result, it supports the formation and maintenance of the nucleocapsid-RNA complex. Membrane, an essential protein, sensitizes the host cell to viruses and activates the Toll-like receptor-dependent process and the Interferon-Beta pathway which are important for host immune regulation [13].

1.2.2. Nucleocapsid

The Nucleocapsid (N) plays a critical role in viral production and has two domains that can bind the viral genome in various ways. This protein interacts with the Nsp3 and contributes to virion production. Also, it is involved in the replication and transcription of viral RNA. The Nucleocapsid also performs as an interferon antagonist, preventing the immune system from destroying the virus [14].

1.2.3. Envelope

The Envelope (E) has a role in the viral organization, budding, formation, and pathogenicity. Membrane and Envelope interact to form virus-like particles [15]. While it is abundantly expressed within the infected cell during the replication cycle, only a small part is added to the virion envelope. [16].

1.2.4. Spike

The Spike (S) consists of two functional subunits; The S1 subunit binds to the host cell detector, while the S2 subunit is responsible for viral and cell membrane fusion. While N-Terminal Domain and Receptor Binding Domain are found in the S1 subunit, the Trans-Membrane region, Fusion Peptides, hexadecimal repeat regions HR1 and HR2 are present in the S2 subunit. The S1 part of the S protein, separated into two polyproteins by the protease enzyme, forms the receptor binding site, whereas the S2 part forms the stem portion of the S protein and is responsible for viral membrane fusion. Via the N-terminus of the S-proteins, viruses bind to specific surface receptors on the host cell plasma membrane. ACE-2 is the binding receptor for SARS-CoV-2 [17].

1.2.5. ORF1ab

ORF1ab is cleaved into 16 nonstructural proteins by viral proteases to form nonstructural proteins such as the enzymes RNA-dependent RNA polymerase (RdRp) and Helicase, which are involved in transcription and replication and form the replication transcription complex. The papain-like cysteine protease that breaks down polyproteins

is contained in Nsp3, while 3Clpro is present in nsp5 [18]. ORF1ab Nsps has essential for innate host response, vesicle membrane formation, mRNA capping, and replication transcription process, and their functions are summarized in Table 1 [19].

Table 1. List of Nsps of SARS-CoV-2 and their functions

Protein	**Function**
Nsp1	Cellular mRNA degradation and preventing host mRNA translation through ribosome binding
Nsp2	Modulation of host cell survival signaling pathway
Nsp3	As a protease to cleave the translated polyprotein into its different proteins.
Nsp4	Binds to viral replication-transcriptional complex and altered ER membranes.
Nsp5	Involves viral polyprotein processing in replication.
Nsp6	Involves in the initial induction of autophagosomes from the host endoplasmic reticulum.
Nsp7/Nsp8	It forms a hexadecameric supercomplex with each other, which adopts a hollow cylinder-like structure containing replication.
Nsp9	It participates in viral replication by acting as an ssDNA binding protein.
Nsp10	Plays a vital role in viral mRNA methylation as a cofactor of nsp14
Nsp11	Unknown
Nsp12	RNA-dependent RNA polymerase
Nsp13	It is a helical core domain that binds ATP and unwinds the double-strand RNA for nsp12 polymerase
Nsp14	Exoribonuclease activity and N7-guanine methyltransferase activity act in the 3’ to 5’ direction.
Nsp15	Mn^{+2} dependent endoribonuclease activity
Nsp16	The mRNA cap exhibits methyltransferase activity, which mediates 2’-0-ribose methylation and the 5th-cap structure of viral mRNAs.

1.2.6. ORF3a/3b

ORF3a leads the way to viral release through the use of lysosomal trafficking [20]. ORF3b modulates cytokine production and also triggers apoptosis in host cells [21].

1.2.7. ORF6

By attaching Karyopherin Alpha 2 and Karyopherin Beta 1 to the membrane, ORF6 prevents cell nuclear import complex synthesis [22]. In addition, ORF6 is essential for inhibiting the expression of interferon-stimulated genes with various antiviral movements [23].

1.2.8. ORF7a/7b

ORF7a acts as an opponent of host tetherin and takes its place, interfering with its antiviral activity. Considered taking part in bonding and regulation of leukocyte bind to Integrin Alpha L (ITGAL) of the host [24].

ORF7b is vital for Golgi complex placement, and has a transmembrane helix domain since substituting it with the transmembrane domain from the human endoprotease furin leads to abnormal positioning [25].

1.2.9. ORF8

ORF8 contributes to the modulation of the host immunological response. It binds to the Interleukin 17 receptor A for initiating the Interleukin 17 pathway and increasing pro-inflammatory factor production, and it makes an essential contribution to the cytokine storm that occurs in COVID-19 infection [26].

1.2.10. ORF10

The function of ORF10 is still unclear. However, some studies determined that it reduces the innate immune response via binding to the mitophagy receptor Nip3-like protein X that triggers mitophagy-mediated mitochondrial antiviral signaling protein breakdown. [27].

1.3. SARS-CoV-2 Life Cycle

The virus enters the host cell with the assistance of the ACE2 receptor. When the receptor-binding domain of the spike protein's S1 subunit binds to ACE2, the Spike is broken down into two subunits by protease: S1 and S2 subunits. On the cell's surface, TMPRSS2 also triggers the S2 subunit. Then, a fusion occurs between the cell membrane and the viral membrane. When a virus penetrates a cell, it completes its replicating cycle in the cytoplasm. Because it possesses positive polarity, the virus initiates direct translation after releasing its genomic RNA into the cytoplasm. Positive-stranded viral RNA is turned into polypeptide using host cell machinery. From the polypeptides, 16 Nsp

proteins are produced and then. Then, with the formation of the replication-transcription complex (RTC), genomic RNA and subgenomic RNAs are synthesized. Subgenomic mRNAs have common 5'-leader and 3'-terminal sequences. Other ORFs in one-third of the genome near the 3′-terminal encode the four major structural proteins first: Spike (S), Membrane (M), Envelope €, and Nucleocapsid (N) proteins. All structural and accessory proteins are synthesized from the subgenomic RNAs of coronaviruses.

The synthesized viral proteins and replication products are then transported to the endoplasmic reticulum and the Golgi via the secretory pathway with help of Membrane and ORF3a. Also, The Nucleocapsid links free viral genomes and proteins together to form virus particles [28].

1.4. Immunology of SARS-CoV-2 Infection

Toll-like receptors are activated and produce chemokines when the virus is picked up within the cell and are began to spread. Polymorphonuclear leukocytes, monocytes, natural killer cells, and dendritic cells are attracted to the location by these chemokines. Monokines induced by interferon-γ (MIG), interferon-γ-inducible protein-10 (IP10), and monocyte chemoattractant protein-1 (MCP-1) are chemokines generated by these cells. These chemokines cause lymphocytes to gather together to be identified by antigens addressed by dendritic cells [29 30].

The cytokine group contains interferons generated by lymphocytes and epithelial cells, and a cytokine storm can develop when their release is extreme or unregulated. There is a link between severe COVID-19 cases and proinflammatory dysregulated cytokine release, according to several types of research [31]. In severe COVID-19 occasions, a cytokine storm can occur quickly, leading to respiratory distress and, in some cases, death due to multiple organ failures [32].

1.5. Interaction of SARS-CoV-2 Proteins with Human Proteins

The process of viral entry into the cell is accompanied by the beginning of the interaction.

Although Spike is thought to be the main and important element in this process, there are significant interactions with other proteins that have been revealed in Table 2 [33].

Table 2. SARS-CoV-2 proteins and their interactions with host proteins

Host Protein	Virus Protein	Life Cycle Phase	Notes
ACE2	Spike	Virus Entry	Main entry receptor
Neuropilin	Spike	Virus Entry	Secondary entry receptor
Integrins	Spike	Virus Entry	Secondary entry receptor
TMPRSS2	Spike	Virus Entry	Cleavage of spike for fusion capacity
Furin	Spike	Virus Entry	Cleavage of spike for fusion capacity
Heparan sulfate	Spike	Virus Entry	Host cell adhesion
VPS39	ORF3a	Virus Budding	Facilitates virion egress in lysosomes
TOMM70	Nucleocapsid	Inhibition of host RLR pathway by virus	Prevents the activation of IFN upon viral infection

1.6. SARS-CoV-2 Variants

Previously, Nextstrain, GISAID, and Pango devised methods for defining and tracking SARS-CoV-2 genetic lineages. The World Health Organization created a more straightforward categorization system based on the Greek alphabet to reduce misunderstanding. The first emerged viruses were named in the S clade for Gisaid's naming scheme, followed by the L, O, V, G, GH, and GR clades [34]. On June 15, 2021, the WHO classified the SARS-CoV-2 variants into three basic types; Variant of Concern, Variant of Interest, and Variant of Monitoring [35].

1.6.1. Variants of Concern (VOCs)

Variants that cause severe disease with increased contagiousness and virulence. As a result of this, condition leads to significant adverse changes in treatment and prevention methods. At the moment, the Omicron variant is the currently circulating variant of concern according the WHO. Previously circulating variant of concerns are listed in Table 3 [35].

Table 3. List of Currently Designated Variants Of Concern (VOCs)

WHO Label	Pango Lineage	GISAID Clade	Next strain Clade	Earliest Documented Samples	Date of Designation
Alpha	B.1.1.7	GRY	0I (V1)	UK, Sep-2020	8-Dec-20
Beta	B.1.351	GH/501Y.V2	0H (V2)	South Africa, May-2020	8-Dec-20
Gamma	P.1	GR/501Y.V3	0J (V3)	Brazil, Nov-2020	11-Jan-2021
Delta	B.1.617.2	G/478K.V1	1A, 21I, 21J	India, Oct-2020	4-Apr-2021

1.6.2. Mutations of Concern

A rise in a mutation cluster was identified in the UK. B.1.1.7, named the Alpha variant by the WHO, emerged in September 2020. The effect of the D614G mutation in the G clade on the virus's course has been seen, and with this mutation, the virus's binding to the ACE2 receptor is enhanced, and therefore the viral load has risen [36]. This version features a total of 23 mutations, including the D614G mutation, and contains critical alterations. The H69/V70 deletion mutation was found as a consequence of convalescent plasma treatment in a person with COVID-19, a lymphoma sufferer, and allowed him to avoid certain antibodies and change the Spike's form. The Y453F mutation, which was reported to be transmitted from minks to humans in Denmark in November 2020, and the N501Y mutation, which was detected in laboratory tests in mice, were both shown to enhance binding to the ACE2 receptor [37] [38]. The ORF8 becomes deactivated as a result of the Q27stop alteration, demonstrating that mutations may aggregate in diverse areas [39]. The mutations N501Y, K417N, and E484K were prevalent in B.1.351, called the 'Beta' variant, which is thought to have appeared in South Africa after the UK variant. Among these, the E484K mutation was revealed to be the most frequent and the one with the highest success in evading antibodies [40]. The P1 variant, also known as the Gamma variant, has the most amount of alterations in Spike and important ORF1ab, ORF8, and Nucleocapsid mutations. These alterations lead to spreading from one person to another, catching the virus again, and escaping antibody immune response [41]. Lineage B.1.617 which emerged in India in October 2020 and is also known as the Delta variant, became a worldwide sensation. Three mutations in the RBD region of the Spike of this variation which differ from the other variants have gained attention. The mutations L452R and P681R speed up viral uptake into the host cell and boost ACE2 binding, which has a significant impact on virus transmission between people [42].

1.6.3. Variants of Interest (VOIs)

Variants detected in different countries differ from the original virus and are suspected of altering the course of the disease. Previously circulating variant of concerns are listed in Table 4 [35].

Table 4. List of Previously Designated Variants of Interest (VOIs)

WHO Label	Pango Lineage	GISAID Clade	Next strain Clade	Earliest Documented Samples	Date of Designation
Lambda	C.37	GR/452Q.V1	21G	Peru, Dec-2020	14-Jun-2021
Mu	B.1.621	GH	21H	Colombia, Jan-2021	30-Aug-2021
Epsilon	B.1.427 B.1.429	GH/452R.V1	21C	United States of America, Mar-2020	VOI: 5-Mar-2021 Previous VOI: 6-Jul-2021
Zeta	P.2	GR/484K.V2	20B/S.484K	Brazil, Apr-2020	VOI: 17-Mar-2021 Previous VOI: 6-Jul-2021
Eta	B.1.525	G/484K.V3	21D	Multiple countries, Dec-2020	VOI: 17-Mar-2021 PreviousVOI: 20-Sep-2021
Theta	P.3	GR/1092K.V1	21E	Philippines, Jan-2021	VOI: 24-Mar-2021 Previous VOI: 6-Jul-2021
Iota	B.1.526	GH/253G.V1	21F	United States of America, Nov-2020	VOI: 24-Mar-2021 Previous VOI: 20-Sep-2021
Kappa	B.1.617.1	G/452R.V3	21B	India, Oct-2020	VOI: 4-Aprl-2021 Previous VOI: 20-Sep-2021

1.6.4. Formerly monitored variants (VOMs)

Variants formerly under monitoring are thought to be dangerous in the past owing to changes in the viral genetics of SARS-CoV-2. As a result of the observations, it was indicated that there is no proof of their impacts on the overall epidemiological situation and the global public health significance.

They were not associated with any concerning properties and they lost their importance due to observations. According to current WHO data, Formerly monitored variants are listed in Table 5 [35].

Table 5. List of Formerly Designated Variants of Monitoring (VOMs)

Pango Lineage	GISAID Clade	Next strain Clade	Earliest Documented Samples	Date of Designation
B.1.1.318	GR	-	Multiple countries, Jan-2021	2-Jun-2021
C.1.2	GR	-	South Africa, May 2021	1-Sep-2021
B.1.640	GH/490R	-	Multiple countries, Sep-2021	2-Nov-2021
AV.1	GR	-	United Kingdom, Mar-2021	26-May-2021

1.7. Diagnostic Methods

Nucleic acid amplification, serological testing, computed tomography, and CRISPR-based methods are all utilized to diagnose COVID-19 [43]. The nucleic acid amplification method, RT-PCR is widely used and standardized by WHO, it is taken as a basis [44]. This procedure involves extracting virus samples from affected people's respiratory tracts using formulated dyes. Simultaneously, the presence of nucleic acids is tested [45]. The benefits of this procedure are quantitative results and high analytical sensitivity. The drawbacks are limited accuracy, the possibility of false-negative results, and the need for specialized equipment [46]. The lungs of patients are scanned using computed tomography, and a definite diagnosis is obtained as a consequence of imaging. It is considered that this method is to minimize error rates in the findings of RT-PCR and allow for early diagnosis [47]. Another important diagnostic method is serological tests. It is used to detect asymptomatic people and the immune status of individuals [48]. For SARS-CoV-2, it is utilized to find Spike and Nucleocapsid or detect the host's immune response to these proteins. Nucleocapsid plays a role in virus replication, and it can be observed at high levels in the first 14 days [49]. The Spike provides attachment to the host cell which is essential for antibody studies. Now, FDA-approved recombinant Nucleocapsid and Spike tests are also available. These tests have the advantage of detecting immunity levels and tracking patients who have antibodies. The disadvantages are that it produces inaccurate results based on the sample type and limited sensitivity [50]. Finally, the CRISPR/Cas technique, which is revolutionary and alternative, has been modified to diagnose SARS-CoV-2. Samples collected from the patient's respiratory tract with viral RNA and amplifying targeted gene, suitable effector Cas protein with complementary RNA to the target gene area are getting together and then reading test strips for results. The FDA has

certified two CRISPR-based therapies for emergency use in COVID-19 [51] [52]. Its benefits include quickness, low cost, portability, and high sensitivity [53].

1.8. Vaccines

Vaccine development research has been increased as a result of the pandemic's onset. With the advancement of technology, the vaccine development time which used to take several years has been significantly reduced. Previously, the vaccine with the shortest approval rate was the mumps vaccine, which took roughly 5 years. The World Health Organization has sped the licensing procedure for vaccine development to avert the devastation caused by the pandemic. Initially, vaccine development studies are divided into stages, the first of which is to conduct extensive laboratory and computer studies to identify antigens. Animal studies are performed in the second step to assess the vaccine's safety and effectiveness. Human practices begin in the third stage, which is separated into four phases. The first step is the safety phase, which is normally administered to less than 100 healthy persons and then followed upon. The number of persons in the second phase grows to hundreds, and it is applied to people with various statistical characteristics. The appropriate dosage, safety, and immunological state are all evaluated. Tens of thousands of individuals get vaccinated in the third stage. People are separated into placebo and vaccination groups, although they are not fully informed of which group they are in. The illness status of vaccinated people is compared to that of a placebo. The vaccine is presented for approval with the reduction in illness rates among those who have been vaccinated, as well as the determination of the vaccine's efficacy. Vaccines given way for approval to institutions such as the FDA and EMA are licensed as a consequence of the institutions' assessments. This procedure, which ordinarily takes more than a year owing to the COVID-19 outbreak, was completed in less than a year under the scope of emergency use approval. Licenses are secured for authorized vaccinations, and they are subsequently made accessible to communities. In the fourth and final stage, information such as the vaccine's efficiency and adverse effects are still being gathered. Vaccines licensed for emergency use have been deployed in several nations since the end of 2020 [54]. Reportedly, 31.2 percent of the global population has received at least one dose of COVID-19 vaccination. Fully vaccinated people make up 23.5 percent of the world's population. 4.7 billion doses of vaccination have been delivered globally. Currently,

36.67 million vaccination doses are provided every day. In low-income nations, just 1.2 percent of individuals have gotten at least one dose of the vaccination [55]. As of 2 December 2021, a total of 7.864.123.038 vaccine doses have been administered. 168 vaccine candidates, 536 vaccine trials, 62 countries with vaccine trials, 40 vaccines at Phase 1, 58 vaccines at Phase 2, 62 vaccines at Phase 3, 28 vaccines approved by at least one country, and 9 vaccines approved by WHO [56].

1.8.1. Complete Virion Vaccines

This category contains two types of vaccines; live attenuated vaccines and inactivated vaccines. Live attenuated vaccines are generated with viral strains that have been devitalized by mutations in animal and human cells and have lost their virulence. Live attenuated vaccines can trigger both humoral and cellular immune responses in actual viral infections and can be applied to the nasal cavity. The vaccination provides mucosal immunity at the point of viral entry into the upper respiratory tract. However, it is known that the potential of re-virulence by changing the vaccine virus in coronaviruses makes this form of vaccination challenging for COVID-19. Some live attenuated COVID-19 vaccines are now under preclinical testing. Nonetheless, none have progressed to human trials [57].

Inactivated Vaccines are created as a consequence of destroying viral strains via a physical or chemical approach. Usually, it is made with an adjuvant to boost the immune response. Inactivated vaccinations are applied intramuscularly. A biosafety level 3 facility is required for manufacture. A disadvantage is that multiple dosages are necessary, and the treatment is not remarkably effective and does not take an active role in the generation of immunological responses. Bharat Biotech Covaxin, Sinopharm (Beijing) Covilo, and Sinovac CoronaVac are inactivated vaccines that are used for COVID-19 treatment and approved by WHO.

Coronavac manufactured by Sinovac, a Chinese corporation, completed Phase 3 trials in Brazil, Chile, Indonesia, and Turkey. All countries' research showed that vaccine efficacy is above 60 percent [58] [59]. The impact of two doses of the Coronavac vaccine against variants is as follows: from 38.7 to 53.8 percent against the Beta and Gamma variants, 59 percent against the Delta variant [60] [61].

Covaxin, also known as BBV152, is manufactured by Bharat Biotech which is an Indian company. Symptomatic and non-symptomatic patients included in phase 3 data indicated effectiveness ranging from 57 to 100 percent, and it has from 33 to 83 percent efficacy against the Delta variant [62].

Covilo is produced by Sinopharm, and phase trials were made in Egypt, Bahrain, Argentina, and the United Arab Emirates. Based on the trial results of the vaccine, it is from 78 to 86 percent effective [63] [64].

1.8.2. Nucleic Acid DNA and mRNA Vaccines

The Nucleic acid-based DNA and mRNA vaccines are antigen-encoding plasmid DNA, RNA as mRNA, or viral replicons used in nucleic acid-based technologies. Because of their cellular absorption and expression, antigens encoded by nucleic acid trigger both humoral and cellular immune reactions.

DNA-based vaccines are made up of antigen-encoding DNA plasmid molecules and are preferable to mRNA vaccines due to their stable formulation and strong transfer efficiency. Although they are similar to the mRNA vaccines, they are affordable and, can be produced rapidly, can induce mutations in the host cell since they must enter the nucleus to be effective [65].

An individual is injected with RNA containing the virus's genetic code. This genetic information is used by the injected person's cells to produce the Spike and activate the immune system. Since mRNA is easily destroyed, it is enclosed in a lipid nanoparticle and delivered into the cell. After entering the cell, host lipases dissolve the lipid nanoparticle structures and let mRNA free. The targeted immunity is created by decoding the codes in the mRNA spike generated inside the cell, which activates both humoral (antibody) and cellular (T-cell) immunity. mRNA cell stays only in the cytoplasm and does not access the nucleus, and it is destroyed in the cytoplasm within 72 hours. mRNA is quickly damaged and must be preserved at low temperatures. Their main benefit is that they can be manufactured fast.

Moderna "mRNA-1273, and Biontech-Pfizer "BNT162b2" vaccines are the mRNA COVID-19 vaccines approved by WHO. Moderna was produced by a US-based company and after finished the genetic sequencing for the vaccine, which was labeled as "mRNA-1273," on January 23. This vaccine has low stability, a 10-fold lower

transfection rate than viral vectors, and a short half-life. The first phase of the trial began on March 16. Phase 2 studies began in October, and Phase 3 investigations began at the end of July. In phase 3 clinical testing, the effectiveness and safety of this vaccine, which got emergency use approval on December 8, 2020, was reported to be 94.1 percent. The impact of two doses of the Moderna vaccine against variants is as follows: 82-100 percent against the Alpha variant, from 91.9 to 98.7 against the Beta and the Gamma variants, and from 75.9 to 90.8 percent against the Delta variant, based on different researches [66] [67].

While Biontech, located in Germany, had begun vaccination tests against Covid-19, they opted to continue clinical trials by cooperating with Pfizer. It was shown that the effectiveness of the "BNT162b2" coded vaccinations chosen for phase 3 research was 95 percent. After getting emergency use authorization in the United Kingdom on December 2, 2020, the vaccine got an FDA license on December 11, 2020, making it the first COVID-19 vaccine. The impact of two doses of the Biontech vaccine against variants is as follows: from 74 to 96 percent against the Alpha variant, from 69 to 92 percent against the Beta and Gamma variants, and from 43.9 to 61.4 percent against the Delta variant, based on different researches [66] [68] [69] [70].

1.8.3. Viral Vector Vaccines

There are two kinds of viral vector vaccines; the Ebola vaccine can replicate, whereas the second cannot reproduce because vital genes, such as adenoviruses, have been deactivated. Adenoviruses are commonly utilized as vectors and have been genetically modified to be capable of producing coronavirus proteins in the body. These viruses do not cause an infection because they are subsided. It is risk-free for the individual; the genetic material carried ensures that the immune system is activated and antibodies are formed [71]. The development of immunity against the vector by the host may reduce vaccine effectiveness. Janssen (Johnson &Johnson) Ad26.COV2.S, Oxford/ Astrazeneca, and Covishield are the non-replicating viral vector vaccines that the WHO approved for COVID-19 treatment. Janssen (Johnson &Johnson) Ad26.COV2.S vaccine got conditional permission for usage in the EU In March 2021. This vaccine carries vector rAd26, and after a single dosage in phase 2 research, 90 percent of antibodies were reported in the first weeks, and in the later phase, antibodies were in close to 100 percent

of cases. Antibody levels rose 2.7-fold after the second dosage. The vaccine's efficacy was 66 percent in a phase 3 trial [72]. The impact of two doses of the Oxford/ Astrazeneca vaccine against variants is as follows: 68.4–79.4 percent against the Alpha variant, 69–92 percent against the Beta and Gamma variants, 61.3–71.8 percent against the Delta variant, based on different researches [73].

1.8.4. Recombinant Vaccines

Recombinant vaccines are divided into protein subunit and virus-like particle vaccines. Protein subunit vaccines deliver all the SARS-CoV-2 proteins to the person and boost the immune system. Subunit vaccinations prevent adverse effects including re-virulence in attenuated vaccines and antigenic peptide denaturing in inactivated vaccines but subunit vaccinations have lower immunogenicity [74]. Virus-like particle vaccine was developed based on the idea that blank viral shells lacking genetic material mimic the structure of the virus that causes the disease and induce antibody production in the live object in which it is inserted [75].

Novavax Nuvaxovid and Serum Institute of India COVOVAX are protein subunit vaccines approved by WHO and share the same formula. A study that used the Novavax vaccine found that 95 percent efficacy was reported in non-variant strains, 85.6 percent in the Alpha variant strains, 60 percent in the Beta strains, and 89.3 percent in all groups [76].

1.8.5. BCG Vaccine

Another mentioned approach for the COVID-19 disease is the BCG vaccine, often known as the tuberculosis vaccine, which was produced around the beginning of the twentieth century. It had been shown to reduce child mortality with the treatment of tuberculosis and give protection against various agents. The vaccination is known to boost cytokine levels. Although it has been claimed to boost natural immunity quickly and efficiently, significant proof has yet to be obtained [77].

1.9. Drugs and Other Treatments

It is vital to understand the virus genome, the processes involved in the virus's entry into the cell, and its life cycle to create an antiviral treatment. Having control of the first phases of the viral cycle is as important as the virus's entry into the host cell with the assistance of ACE-2, TMPRSS2, 3Clpro, the RNA-dependent RNA polymerase, fusion, and endocytosis[78]. Proteins which are vital for the viral life cycle, have similarities to some viruses such as HIV and Hepatitis C. Antivirals used for treating these diseases can be significant for COVID-19; therefore, similarities of duplication [79] [80].

The medications that will be utilized to address the patient's circumstances should also be considered. It is also possible to achieve host-based antiviral development. Essential elements of this are the interaction of ACE2 with viruses and the modulation of cytokine storms [81].

The first antiviral drug that is approved by the Food and Drug Administration for the treatment of COVID-19 is Remdesivir. It was developed for the treatment of the Ebola and Marburg viruses. It transforms to the adenosine triphosphate analog, which has the task of infecting the viral RNA polymerase. By blocking viral replication, viral load is lowered and pulmonary function is improved [82]. Paxlovid, is recently approved by Food and Drug Administration, is an oral antiviral. Results of Paxloid usage in trials indicate that it helps decrease mortality caused by COVID-19 [83].

Convalescent Plasma Therapy is another aspect of the COVID-19 treatment. This approach has been performed since the 1930s and used to treat SARS-CoV and MERS-CoV. It is based on the isolation of the serum-containing antibodies collected from the people who recovered [84].

Monoclonal Antibody is another approach for the COVID-19 preventing process. With this approach, the Spike of the virus can be neutralized by the antibodies specific for only one epitope. RBD region activity can be controlled and block the entrance of the host cell. In addition, a TMPRSS2 serine protease, PIKfyve, TPC2, and cathepsin L, which are essential for endocytosis, are also of great importance [85] [86]. It is found that the use combined with camostat mesylate and substance E-64d is highly effective for blocking the virus's entry into the host cell.

Merck and Ridgeback's oral antiviral drug Molnupiravir is a nucleoside analog of β-D-N4 hydroxycytide (NHC). The drug increased the frequency of viral RNA mutations

and the incidence of viral RNA mutations thus SARS-CoV-2 replication in a human was disrupted. In the Phase III study, Molnupiravir decreased by 50 percent the risk of hospitalization or death compared to placebo for mild or moderate COVID-19 patients [87]. Molnupiravir has been approved for emergency use by the FDA.

1.10. Aims & Hypothesis of Thesis

The main hypothesis of this thesis is that molecular evolutionary and population genetic analyses of the SARS-CoV-2 genome can identify the most conserved genome regions indicated by the strongest negative selection, and these highly conserved genome regions can be targets for small molecules (or drugs) or used for new vaccine approaches.

To test this hypothesis, firstly, comprehensive molecular population genetic analyses of virus genomes were performed with all viral genes for populations of viral variants of concern (Alpha, Beta, Gamma, Delta) associated with faster spread, and the very first clades seen at the beginning of the pandemic (L, S, O). Intraspecific comparisons focused only on human-specific SARS-CoV-2 variants, and for interspecific comparisons, bat RaTG13 is used as an outgroup.

Secondly, based on these evaluations, certain genes are chosen as the most conserved genes, and their structural information, functions, and interaction with other virus genes were used to select potential new drug and vaccine targets.

CHAPTER 2

MATERIALS AND METHODS

2.1. Preparation of Dataset

All genome sequences in this study were collected from the GISAID site [8889]. Genome sequences that are at least 29 kb or larger in sequence size, high coverage (less than 1% N content, 0.05 percent unique mutations, no unconfirmed indel mutations), human host isolates, and with recorded collection dates are selected for analyses. Only the genome sequences that belong to L, S, O clades and Alpha, Beta, Gamma, and Delta variants were analyzed. Then, the Bat Coronavirus RaTG13 sequence was added to the collected human sequences. The prepared data was aligned with the MAFFT (v7.450) alignment software with the recommended parameters [90].

Initially, over one thousand genome sequences were downloaded and aligned. However, the sequences that contained letters other than A, T, C, G, and N's were excluded from the analyses, and analyses were conducted with the remaining 808 sequences. After the alignment and selection of viral genome sequences, gene and nsp positions were found from the GISAID reference genome and were cut and trimmed [91].

Table 6. Genome positions of SARS-CoV-2 genes

Gene	Annotation (nt position)
ORF1a	1-13206
ORF1b	13206-21293
S	21301-25159
ORF3a	25168-25995
ORF3b	25540-25995
E	26020-26247
M	26299-26967
ORF6	26978-27163
ORF7a	27170-27535
ORF7b	27532- 27663
ORF8	27670-28037
N	28056-29315
ORF10	29340-29456

Table 7. Nsp positions within the ORF1ab gene

ORF1ab NSPS	Annotation (nt position)
Nsp1	1-540
Nsp2	541-2454
Nsp3	2455-8289
Nsp4	8290-9789
Nsp5	9790-10707
Nsp6	10708-11580
Nsp7	11581-11829
Nsp8	11830-12423
Nsp9	12424-12762
Nsp10	12763-13179
Nsp12	13180-15974
Nsp13	15975 – 17777
Nsp14	17778-19358
Nsp15	19359-20396
Nsp16	20397 -21290

2.2. Population Genetic Analyses

Aligned data was exported to the DnaSP, to conduct population genetic statistics and molecular evolution tests [92]. Population genetic parameters include nucleotide and haplotype diversity, Watterson theta estimator, and neutrality tests which are Tajima's D, Fu-Li's D, Fu-Li's F, Fu-Li's D*, Fu-Li's F*, and Fay and Wu's Hn and McDonald-Kreitman (MK) test were used for intraspecific and interspecific assessment and for deciding the pattern of selection or diversity in populations. Nucleotide diversity, represented by π, is used to evaluate the level of polymorphism and is the average amount of nucleotide differences given two DNA sequences [93]. The letter Hd stands for haplotype diversity, which assesses the distinctiveness of a certain haplotype in a population [94]. The Watterson estimator (θ) is a formula for calculating the nucleotide proportion of polymorphic sites [95]. In addition to these parameters, some essential tests were used for understanding the selection and demographic dynamics of populations. Tajima's D is a method for figuring out how selection works and it is estimated utilizing theta and pi. If Tajima's D value is zero, it means there's no indication for selection. A positive Tajima D value shows that heterozygosity has a selective advantage and a declining population. In contrast, a negative one shows that a specific allele has a survival benefit over the other allele and that the population is rapidly expanding [96]. Also, Fu-Li's population genetic tests were conducted. Fu-Li's test is similar to Tajima's D statistic in that a negative result shows an excess of singletons, while a positive result demonstrates a lack of singletons. The data of Fu-Li's D* and Fu-Li's F* tests are used for within species (only within

human viruses) sequence comparisons, whereas it is necessary to have intraspecific and outgroup data for D and F tests [97]. Another aspect that is commonly utilized in population genetics is Fay and Wu's Hn. This test gives information about selection occurring owing to population increase or decrease or a recent selective sweep [98]. Moreover, the McDonald-Kreitman method analyzes the levels of polymorphism and divergence within and between populations at two types of sites. For MK Test, negative selection means the neutrality index is lower than one and occurs when the rate of nonsynonymous to synonymous variation across species is less than the rate of nonsynonymous to synonymous variation within species; nevertheless, positive selection means the neutrality index is higher than one and occurs when the inverse proportion of the explained variations is observed, also determine the P-value based on the Chi-square value [99] [100] [101]. After the analysis, a Sliding Window analysis was performed using the Polymorphism and Divergence option over DnaSP to graph the nonsynonymous nucleotide diversity of some selected regions[92].

2.3. Protein Structure Prediction And Related Analyses

The 3-dimensional structure of selected regions was revealed using the I-TASSER which is a bioinformatics tool that uses amino acid sequences to predict the three-dimensional structure of protein molecules with a C-score representing the convergence parameters of the structure assembly simulations. It ranges between -5 and 2 and the best C-scored model represents the highest value of confidence. By structurally comparing the target protein's structural patterns with known proteins in protein function databases, it has been expanded for structure-based protein function predictions, offering further explanations regarding the ligand-binding site, gene ontology, and enzyme commission [102–104]. Also, UCSF Chimera is used for the demonstration of molecular structures with associated data in an interactive environment [105].

CHAPTER 3

RESULTS AND DISCUSSION

In total, 808 sequences were analyzed; 185 sequences belong to the L clade, another 185 sequences belong to the S clade, 175 sequences belong to the O clade, 82 sequences belong to the Alpha variant, 55 sequences belong to the Beta variant, 72 sequences belong to the Gamma variant, and 53 sequences belong to the Delta variant. Every calender month starting from December 2019 to August 2021 is represented by at least one sequence. For interspecific tests, RaTG13 is used as an outgroup.

Table 8. List of date SARS-CoV-2 sequences chosen from GISAD with clades/variants

Clade/Variant	**Date of Sample Sequences**
L	December 2019 – August 2021
S	December 2019 – August 2021
O	January 2020 – August 2021
Alpha	September 2020 – August 2021
Beta	October 2020 – August 2021
Gamma	October 2020 – August 2021
Delta	September 2020 – August 2021

3.1. Population Genetics and Selection Tests on SARS-CoV-2 Genome

Looking at population genetic summary statistics for the SARS-CoV-2 genome among its population of clades and variants, the highest number of segregating sites were seen on the O clade among the clades, and the highest number of segregating sites were seen on the Alpha variant among the variants. For all clades and variants, the number of singleton changes was higher than parsimony informative sites. The O clade had the highest number of singletons compared to other clades and the Alpha variant had the highest number of singletons compared to other variants. Overall, more replacement polymorphisms compared to synonymous polymorphisms were observed for the genome-wide data. Whereas the O clade had the highest number of replacement polymorphisms, the S clade and the Delta variant showed the highest number of replacement site diversity.

Also, the nucleotide diversity with Jukes-Cantor correction applied estimates showed that among all variants and clades the overall nucleotide diversity is primarily driven by synonymous sites. For nucleotide diversity on synonymous sites, the O clade had the highest value among the clades, and the Delta variant had the highest value among the variants.

Table 9. Population genetic summary statistics for nucleotide diversity of SARS-CoV-2 genome among SARS-CoV-2 population of clade/variants[1]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Syn. sites	6660.49	6642.05	6639.63	6650.57	6659.23	6663.94	6653.24
Nonsyn. sites	22697.51	22625.95	22640.37	22692.43	22704.77	22706.06	22683.76
S	367	807	1111	365	329	315	228
Eta	370	814	1126	366	329	318	231
Sing.	301	549	812	317	257	242	155
Par.	64	258	299	48	72	73	73
Syn. Pol.	125	267	377	141	107	100	81
Rep. Pol.	239	533	736	223	219	217	144
π (Pi) All Sites	1.9	10.3	10.2	4	6.1	4.3	7.8
Theta-W All Sites	21.5	47.5	10.2	25	24.5	22.1	17.3
π (JC) All Sites	1.9	10.3	10.2	4	6.1	4.3	7.8
π (JC) Syn. Sites	2.9	13.2	14	6	9.4	6	11.1
π (JC) Nonsyn. Sites	1.6	9.5	9.1	3.4	5	3.8	6.9
H	114	164	168	80	53	66	48
Hd	0.89	1.00	1.00	1.00	1.00	1.00	1.00

With population genetic summary statistics, neutrality tests also were performed for assessment of the selection on viral genomes. For Tajima's D, genome data considering all clades and variants was negative and significant. Moreover, Fu-Li's D* and Fu-Li's F* test values were also significant and negative, in agreement with the Tajima's D tests. When the clades and variants were compared, the L clade had the most negative value for Tajima's D, Fu-Li's D*, and Fu-Li's F* tests. The test values are negative when there is an excess of rare variants compared to intermediate frequency variants as a consequence of background selection or population growth.

Table 10. Neutrality tests summary statistics for the SARS-CoV-2 genome among viral clades and variants[2]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
TD	-2.93 ***	-2.54***	-2.76***	-2.90***	-2.70***	-2.81***	-2.00*
TD - Cod.	-2.93 ***	- 2.55 ***	-2.78***	-2.90***	-2.70***	-2.81***	-1.98*
TD – Syn.	-2.86 ***	-2.61***	-2.78***	-2.92***	-2.58***	-2.72***	-2.09*
TD – Nonsyn.	- 2.91***	- 2.51***	-2.75***	-2.84***	-2.72***	-2.80***	-1.88*
TD - Silent	-2.89***	-2.61***	-2.78***	-2.92***	-2.58***	-2.72***	-2.09*
Fu-Li's D*	-11.50**	-9.33**	-9.99**	-8.26**	-5.72**	-6.53**	-4.42**
Fu-Li's F*	-8.71**	-7.09**	-7.65**	-7.19**	-5.45**	-6.00**	-4.19**

When Fu-Li's F and Fu-Li's D tests were conducted using Bat coronavirus RaTG13 as an outgroup, there was still an excess of rare variants. The results of neutrality tests and Fu-Li's F and Fu-Li's D were consistent. Although Fay and Wu's Hn estimates for all clades and variants were negative, the test results were not significant.

Table 11. Fu-Li's Tests with an Outgroup (RaTG13) for human SARS-CoV-2 genomes among viral clades and variants[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Fu and Li's D	-11.84**	-9.30 **	-9.88**	-8.07**	-5.74**	-6.76**	-4.29**
Fu and Li's F	-8.66**	-6.86**	-7.36**	-6.92**	-5.42**	-6.08**	-4.09**
Fay and Wu's Hn	-30.34	-109.75	-176.74	-75.05	-62.15	-77.11	-40.33
Fay and Wu's Hn Normalized	-1.13	-1.37	-1.57	-1.81	-1.54	-2.09	-1.39

McDonald-Kreitman tests showed that the rate of nonsynonymous divergence to synonymous divergence is less than the rate of nonsynonymous polymorphism to synonymous polymorphism in all clades and variants indicating negative selection acting on all genomes. Moreover, both the neutrality index and the adaptive protein evaluation parameter alpha also showed negative selection acting on the genomes of all clades and variants.

Table 12. The McDonald-Kreitman Tests of SARS-CoV-2 genomes among human clades and variants with an outgroup (RaTG13)[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
NI	3.32	3.42	3.33	2.64	3.42	3.71	3.00
Alpha Value	-2.32	-2.46	-2.33	-1.64	-2.42	-2.71	-2.00
Fisher's exact test. P-value (two tailed)	0***	0***	0***	0***	0***	0***	0***
G test G value	93.85	165.92	183.48	62.54	90.34	99.58	54.61
G test P value	0***	0***	0***	0***	0***	0***	0***
Syn. Fixed differences between species	700	662	640	685	685	696	694
Syn. Polymorphic sites	124	265	374	141	107	100	81
NonSyn. Polymorphic sites	238	532	734	223	219	215	144
NonSyn. Fixed differences between species	405	384	377	411	410	403	411

3.2. Population Genetics And Selection Tests On Individual SARS-CoV-2 Genes

After evaluation of genome-wise data using population statistical tests, all genes on SARS-CoV-2 human sequences were also analyzed individually to understand the nature of selection acting on these genes. All population genetic tests that were conducted for genome-wise analyses for every SARS-CoV-2 gene; ORF1a, ORF1b, Spike, ORF3a, ORF3b, E, M, ORF6, ORF7a, ORF7b, ORF8, N, ORF10.

In the detail of the individual population genetic analyses of all genes, the population summary statistics, and neutrality tests results showed some differences between clades and variants. The ORF1a gene of the O clade had the highest number of segregating sites. On the other hand, the Envelope gene in the L clade and the Alpha variant, and the ORF7b and ORF10 in the Beta variant had no segregating sites. In addition, the ORF7b in the Delta variant had the highest number of nucleotide diversity compared to other genes based on clades and variants. The Envelope gene in the Alpha variant, the ORF7b in the Beta variant and the L clade, and the ORF10 in the Beta variant and the Delta variant had no nucleotide diversity. Also, the ORF1a in the O clade had the highest number of synonymous and replacement polymorphisms.

In neutrality tests, Tajima's D, Fu-Li's D*'s, and Fu-Li's F*, all clades and variants for all viral genes had a negative value. The ORF1a in the Alpha variant had the

most negative value. Moreover, the ORF1b in the L clade had the most negative Fu-Li's D* and Fu-Li's F* values.

When Fu-Li's F and Fu-Li's D tests were conducted using Bat coronavirus RaTG13 as an outgroup, there was still an excess of rare variants for all genes based on clades and variants. The values of neutrality tests, and Fu-Li's F and Fu-Li's D were consistent. The L clade on the ORF1a had the highest value for Fu-Li's F and Fu-Li's D tests. Also, it was observed that Fay and Wu's Hn values were not significant.

Population summary statistics, neutrality, Fu-Li's F and Fu-Li's D, and the Mc-Donald Kreitman test results for ORF1a, ORF1b, Spike, ORF3a, ORF3b, Envelope, ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid and ORF10 genes for all clades and variants are shown in Supplement Tables 10-18.

There are some differences among the genes based on McDonald-Kreitman tests that utilized the RaTG13 sequence as an outgroup, however, almost all genes from all clades and variants showed negative selection acting on them. But for the Membrane gene, there were no fixed nonsynonymous changes between the RaTG13 genome sequence and SARS-CoV-2 human sequences considering all clades and variants.

Spike protein is the most highlighted protein because of its function in cell entrance. As a result, various approaches have been explored using the Spike as the focus point. Although various therapies such as RBD-targeted inhibitors, S2 targeted antibodies, and recombinant RBD vaccines are used, the COVID-19 spread still could not be prevented. Therefore, instead of focusing on the highly variable spike protein focusing on the conserved viral proteins where no amino acid changes are observed can be the main point of the treatments for prevention against the virus.

3.3. Population Genetics And Selection Tests On Individual Nonstructural Proteins (Nsps)

ORF1ab, the largest gene in the virus genome and has an essential role in the viral life cycle, is composed of Nsp1-Nsp16. Population genetic tests were conducted for each Nsp.

Looking at the population genetic summary statistics, Nsp3 of the O clade had the highest number of segregating sites, synonymous and replacement polymorphisms compared to the other Nsps. On the other hand, Nsp8 of the Beta variant had no

segregating sites. As a result of that, replacement polymorphisms and nucleotide diversity values were zero for Nsp8 of the Beta variant.

Looking at the nucleotide diversity of the Nsps, the Nsp6 of the Delta variant had the highest number of nucleotide diversity. Nucleotide diversity on synonymous sites results showed that Nsp9 the S clade had the highest value.

On the other hand, the Nsp7 in the Alpha and Gamma variants had the lowest value. Nucleotide diversity on nonsynonymous sites results showed that Nsp6 in the S clade had the highest value, in contrast, the Nsp4 in the Gamma variant value was not estimated.

Neutrality test results also show differences between the Nsps but overall, there is negative selection on all Nsps for all clades and variants. Nsp3 in the Alpha variant had the most negative significant value for Tajima's D tests and Nsp3 in the O clade had the most negative significant value for Fu-Li's F* and Fu-Li's D* tests.

When Fu-Li's F and Fu-Li's D tests were conducted using Bat coronavirus RaTG13 as an outgroup, there was still an excess of rare variants for all genes from all clades and variants. The values of neutrality tests and Fu-Li's F and Fu-Li's D were consistent. The Nsp3 in the O clade and the Alpha variant had the highest value for Fu-Li's F and Fu-Li's D tests. All Fay and Wu's Hn test results were also negative, and in agreement with Tajima's D, Fu-Li's F, and Fu-Li's D. But the results were not significant.

Population genetic tests summary statistics, neutrality test and Fu-Li's F and Fu-Li's D test, and McDonald Kreitman test results for Nsp1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp9, Nsp12, Nsp13, Nsp14, Nsp15 based on clades and variants are shown in the Supplement Tables 1-9.

Using RaTG13 as an outgroup, considerable contrasts between McDonald-Kreitman test results of different Nsps were observed. Almost all nsps showed negative selection acting on all clades and variants.

However, for Nsp8, Nsp10, and Nsp16 there were no fixed nonsynonymous changes between the RaTG13 and SARS-CoV-2 human sequences considering all clades and variants.

3.4. The Most Conserved SARS-CoV-2 Genome Regions Based On Mc-Donald Kreitman Test

3.4.1. Population Genetics And Selection Tests On the Membrane

Evaluation of all data by the Mc-Donald Kreitman test indicated that the Membrane gene is highly conserved. Except for the Delta variant, there were no fixed nonsynonymous changes between the RaTG13 and human virus sequences. On the Delta variant, there was one fixed difference for the Membrane gene but the neutrality index value was not significant. However, the Mc-Donald Kreitman test results were significant for the L, S, O, Alpha and Beta. Also, the Direction of Selection test results showed that the Membrane gene was primarily driven by negative selection considering all clades and variants, confirming the high conservation on the Membrane with the Mc-Donald Kreitman tests.

Table 13. The McDonald-Kreitman Tests of SARS-CoV-2 Membrane gene among human clades and variants with an outgroup (RaTG13)[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
DoS	-0.56	-0.33	-0.43	-0.46	-0.22	-0.25	-0.29
NI	-	-	-	-	-	-	13.5
Alpha Value	-	-	-	-	-	-	-12.5
Fisher's exact test. P-value (two tailed)	0***	0**	0***	0***	0.06	0.13	0.19
Syn. Fixed differences between species	27	25	24	25	27	27	27
Syn. Polymorphic sites	4	10	13	7	7	3	2
NonSyn. Polymorphic sites	5	5	10	6	2	1	1
NonSyn. Fixed differences between species	0	0	0	0	0	0	1

Looking at the population genetic summary statistics for SARS-CoV-2 Membrane among its population of the clades and variants, the highest number of segregating sites were in the O clade, and the Alpha variant had the highest number of segregating sites among the variants. For all clades and variants, the number of singleton changes was higher than parsimony informative sites. O clade had the highest number of singletons

among the clades and the Alpha variant had the highest number of singletons among the variants. In addition, the O clade had more replacement polymorphisms compared to other clades and variants. The nucleotide diversity values for Membrane showed that the S clade had the highest value among the clades and the Beta variant had the highest value among the variants. Also, the nucleotide diversity with Jukes-Cantor correction applied estimates showed that all variants and clades were primarily driven by synonymous sites except for the Gamma and Delta variants.

Table 14. Population genetic summary statistics for nucleotide diversity of SARS-CoV-2 Membrane gene among SARS-CoV-2 clades/variants[1]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Syn. sites	163.66	163.62	163.72	163.66	163.67	163.66	163.67
Nonsyn. sites	505.34	505.38	505.28	505.34	505.33	505.34	502.33
S	9	14	21	13	9	4	3
Eta	9	15	23	13	9	4	3
Sing.	9	9	15	12	6	4	3
Par.	0	5	6	1	3	0	0
Syn. Pol.	4	10	13	7	7	3	2
Rep. Pol.	5	5	10	6	2	1	1
π (Pi) All Sites	1.4	9.9	8.8	5.1	9.6	1.7	1.8
Theta-W All Sites	23.2	36.1	54.7	39	29.4	12.3	10
π (JC) All Sites	1.4	9.9	8.8	5.1	9.6	1.7	1.8
π (JC) Syn. Sites	2.6	33.1	18	10.5	32.7	5.1	4.9
π (JC) Nonsyn. Sites	1.1	2.5	5.9	3.4	2.1	6	8
H	10	14	23	12	10	4	4
Hd	0.09	0.49	0.47	0.25	0.50	0.08	0.12

In addition to population genetic summary statistics, neutrality tests were performed for assessment of the selection on the Membrane. For Tajima's D test, Membrane data was negative and significant in all clades and variants other than the Delta variant. The Alpha variant had the most negative value for Tajima's D tests. In addition, Fu- Li's D* and Fu- Li's F* test values were also significant and negative. But for Fu- Li's D* and Fu- Li's F* tests, the L clade had the most negative value. The test values are negative when there is an excess of rare variants compared to intermediate frequency variants as a consequence of background selection or population expansion.

Table 15. Neutrality tests summary statistics for SARS-CoV-2 Membrane among viral clades/variants[2]

	Clade/Variant						
Parameters	**L (N=185**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
TD	-2.15**	-1.91*	-2.36**	-2.41**	-1.86*	-1.82*	-1.7
TD - Cod.	-2.15**	-1.91*	-2.36**	-2.41**	-1.86*	-1.82*	-1.7
TD – Syn.	-1.68	-1.62	-2.19 **	-2.13*	-1.69	-1.65	-1.46
TD – Nonsyn.	-1.82*	-1.66	-1.97*	-1.99*	-1.31	-1.06	-1.1
TD - Silent	-1.68	-1.62	-2.19**	-2.13*	-1.69	-1.65	-1.46
Fu-Li's D*	-6.10**	-4.47**	-5.90**	-5.35**	-2.69*	-3.67**	-3.02*
Fu-Li's F*	-5.59**	-4.20**	-5.40**	-5.13**	-2.85*	-3.61**	-3.05*

When Fu-Li's F and Fu-Li's D tests were conducted using Bat coronavirus RaTG13 as an outgroup, there was still an excess of rare variants. The values of neutrality tests and Fu-Li's F and Fu-Li's D tests were consistent. L clade had the most negative value for the test results. Also, Fay and Wu's Hn values of the S, O, and Alpha were negative, and the L, Beta, Gamma, and Delta values were almost zero. But, the test results were not significant.

Table 16. Fu-Li's Tests with an Outgroup (RaTG13) for human SARS-CoV-2 Membrane gene among viral clades and variants[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Fu and Li's D	-6.20**	-3.97**	-5.18**	-4.16**	-2.05 #	-3.76**	-3.11**
Fu and Li's F	-5.67**	-3.82**	-4.87**	-4.21**	-2.27 #	-3.70**	-3.14**
Fay and Wu's Hn	0.1	-3.25	-5.26	-3.64	0.5	0.11	0.12
Fay and Wu's Hn Normalized	0.07	-1.66	-1.87	-1.98	0.39	0.13	0.16

3.4.2. Sliding Window, Nature of Amino acid Changes, and Structure Analyses of Membrane

To compare nucleotide diversity of nonsynonymous changes for all clades and variants, sliding window analyses were performed. Results showed that the highest

nonsynonymous nucleotide diversity was on the 245th nucleotide which is in the codon that codes for the I82 residue.

All sequences in the L, Alpha, Beta, and Gamma had Thymine nucleotide in the 245th position, but the sequences in the Delta variant had a Cytosine nucleotide leading to a change from Isoleucine to Threonine.

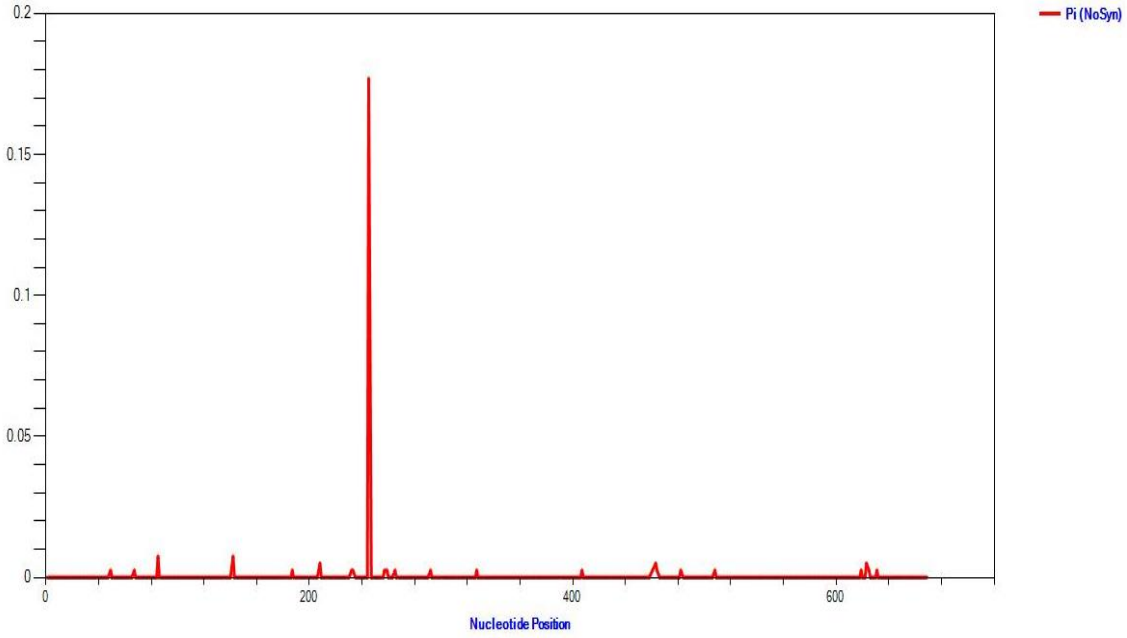

Figure 4. Sliding Window analysis of nonsynonymous nucleotide diversity for Membrane

In the O clade, some sequences had a mutation that changed the nucleotide from T to C in the 245th position.

Moreover, in the S clade, some sequences had both C and G mutation in the 245th position, and residues turned from Isoleucine to Threonine, and from Isoleucine to Serine, respectively. Also, another research showed that I82 is the most mutated residue on the Membrane [106].

Other than the I82 residue, regions that had nonsynonymous nucleotide diversity for the Membrane are listed in Table 17.

Table 17. Polymorphic Membrane residues and associated sliding window nonsynonymous nucleotide diversity estimates in the examined SARS-CoV-2 clades/variants

Window	Pi (NoSyn) (10^{-4})	Clade/Variant	Aminoacid	Change	Domain
49-49	25	Alpha	17	L-> F	Helix
67-67	25	O	23	V-> L	Helix
85-85	74	O, S	29	L->F	Helix
142-142	74	O, S	48	I-> L, I->V	Helix
187-187	25	Delta	63	A-> S	Helix
208-208	50	Alpha	70	V->L	Strand
232-232	25	L	78	G-> S	Helix
233-233	25	O	78	G-> A	Helix
245-245	1760	Delta, O, S	82	I-> T, I->S	Helix
257-257	25	O	86	C-> S	Helix
259-259	25	Alpha	87	L-> F	Helix
265-265	25	Gamma	89	G-> S	Helix
292-292	25	O	98	A-> S	Helix
327-327	25	O	109	M-> I	Coil
407-407	25	Alpha	136	S->N	Coil
463-463	50	Alpha	155	H->L	Strand
464-464	25	Beta	155	H->Y	Strand
482-482	25	L	161	I->T	Coil
508-508	25	Alpha	170	V-> L	Strand
619-619	25	L	207	N-> H	Coil
623-623	50	Beta, S	208	T-I	Coil
625-625	25	L	209	D-> H	Coil
631-631	25	L	211	S-> T	Coil

For Membrane, a 3d structure model was predicted using I-Tasser. The best C-scored model was chosen to display on UCSF Chimera and the distribution of amino acid changes was mapped on the predicted structure with Render by Attribute interface.

The percentage distribution of amino acid changes on the Membrane helix, strand, and coil regions were 53 percent, 30 percent, and 17 percent, respectively. In the figure, the blue-colored residue I82 showed the highest nonsynonymous nucleotide diversity. Other residues were not color-coded, because of the low nonsynonymous nucleotide diversity.

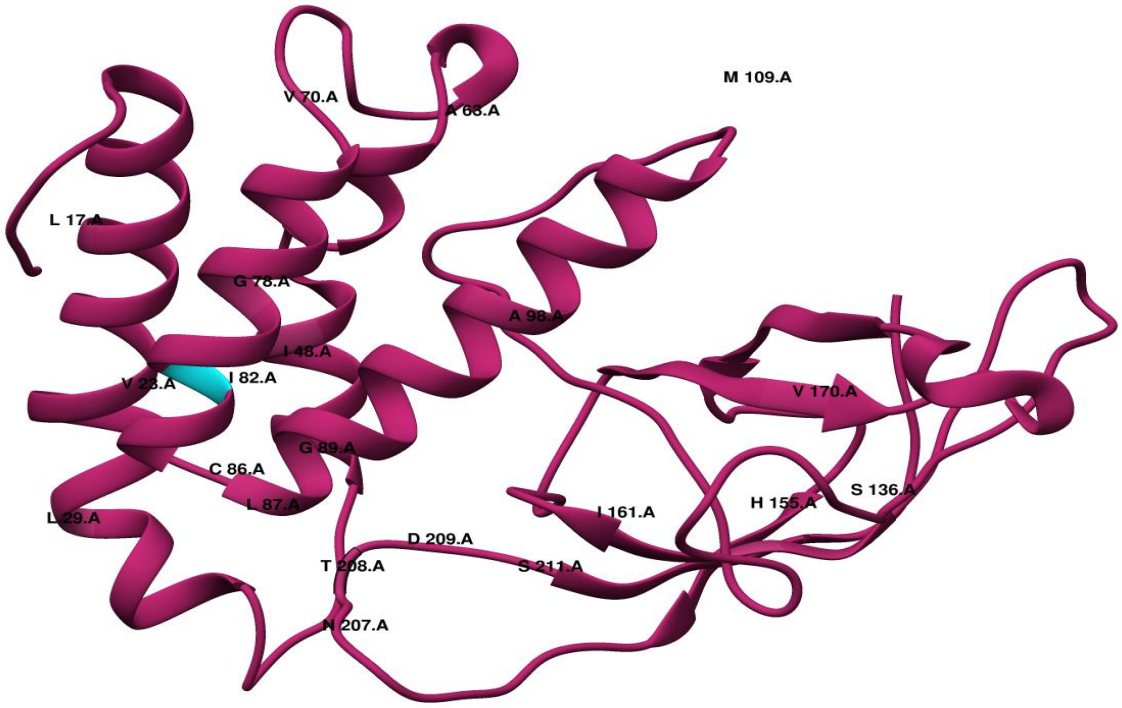

Figure 5. Display of the Membrane secondary structure using I-Tasser with UCSF Chimera. Color coding ranges from highest (blue) to lowest (maroon) nonsynonymous nucleotide diversity.

3.4.3. Interpretation Of Observed Amino Acid Changes In Terms Of Membrane Function

The Membrane is the most abundant protein in the coronaviruses, and also one of the conserved genes. The membrane has 223 amino acids and is divided into three primary domains: a short N-terminal ectodomain, three TransMembrane Helices; the first (amino acids 21–37) second (amino acids 46–68), and third (amino acids 78-100), and a long C-terminal endo-domain located on the cytoplasmic face of virions [106].

The Membrane is the core of the viral envelope and provides shape and size to the virion. Because of the envelope assembly requirement, M-M homodimer interaction is crucial. It also has a role in the processing and modification of various viral genes. In addition, the Membrane helps to form and gather virions by having a key role in interaction with structural genes Envelope, Spike, and Nucleocapsid. Membrane and Nucleocapsid linkages support the virion RNA genome. Membrane and Envelope linkages provide to form and release virus-like particles. Linkages between Membrane

and Spike continue to hold Spike in the Endoplasmic Reticulum–Golgi Intermediate Compartment. Also, the interaction adapts Spike in nascent virions [107].

An experiment on SARS-CoV Membrane (which shares 90.5% sequence identity with SARS-CoV-2 Membrane) determined that residues W19, W57, P58, W91, Y94, F95, and C158 have an essential role in homodimer interactions, proposing that homologous residues W20, W58, P59, W92, Y95, and F96 of SARS-CoV-2 may be essential for Membrane dimer interaction and stabilization [108]. In addition, another experiment indicated that key residues F96, F103, S108, S111, and F112 are preserved and able to form important interactions in the dimer [106].

Another experiment on SARS-CoV-2 Membrane, taken as a receptor, binds to all structural proteins taken as ligands showed that the L51, T55, F96, and F103 residues of Membrane have interaction with F26, R69, F26, and I33 residues of Envelope respectively. Also, M1, N5, Y71, W75, and R174 residues of Membrane have interaction with M1, F4, Q173, D198, F377, and H625 residues of Spike. Moreover, Y199, D209 and H210 residues of Membrane have interacted with G335, F314, and F286 residues of Nucleocapsid [107].

A different experiment on SARS-CoV Membrane determined that residues between 88 and 96 and between 60 and 69 play a role in the Multiple cytotoxic T-lymphocyte epitopes [109]. For SARS-CoV-2, no study confirmed that the mentioned residues are the Multiple cytotoxic T-lymphocyte epitopes but it may be essential for future information.

In this study, the D209 residue that has a role in interaction with Membrane-Nucleocapsid, and the G89 residue that has a role as an epitope of Multiple cytotoxic T-lymphocyte for SARS-CoV were among the polymorphic residues.

Other than the D209 and G89, the conserved residues that are important for homodimer interaction, interaction with other structural proteins, and immune response can be good targets for designing new vaccines or drugs.

3.4.4. Population Genetics And Selection Tests On Nsp8, Nsp10, and Nsp16

Evaluation of all data by the McDonald-Kreitman Tests indicated that the Nsp8, Nsp10, and Nsp16 are highly conserved. There were no fixed nonsynonymous changes

between the RaTG13 and human Sars-CoV-2 viruses for all clades and variants. In detail, Nsp16 of all clades and variants had significant values for the McDonald-Kreitman tests. On the other hand, MK test results were not significant for Nsp8 in the Delta variant, and the Nsp10 in the S and L clades.

Also, the Direction of Selection test results showed that the Nsp8, Nsp10, and Nsp16 were primarily driven by negative selection considering all clades and variants, confirming the high conservation on the Nsp8, Nsp10, and Nsp16 with the Mc-Donald Kreitman tests.

Table 18. The McDonald-Kreitman Tests of SARS-CoV-2 Nsp8 among human clades and variants with an outgroup (RaTG13)[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
DoS	-0.6	-0.58	-0.6	-0.63	-	-0.67	-1
NI	-	-	-	-	-	-	-
Alpha Value	-	-	-	-	-	-	-
Fisher's exact test. P-value (two tailed)	0**	0**	0.01*	0**	0**	0.03*	0.08
Syn. Fixed differences between species	10	10	10	11	11	11	11
Syn. Polymorphic sites	2	5	4	3	0	1	0
NonSyn. Polymorphic sites	3	7	6	5	0	2	1
NonSyn. Fixed differences between species	0	0	0	0	0	0	0

Table 19. The McDonald-Kreitman Tests of SARS-CoV-2 Nsp10 among human clades and variants with an outgroup (RaTG13)[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
DoS	-0.5	-0.33	-0.55	-0.75	-1	-0.6	-0.5
NI	-	-	-	-	-	-	-
Alpha Value	-	-	-	-	-	-	-
Fisher's exact test. P-value (two tailed)	0.22	0.19	0.04*	0.02*	0.03*	0.05	0.04*
Syn. Fixed differences between species	7	7	7	7	7	6	7
Syn. Polymorphic sites	1	4	5	1	0	2	1
NonSyn. Polymorphic sites	1	2	6	3	2	3	1
NonSyn. Fixed differences between species	0	0	0	0	0	0	0

Table 20. The McDonald-Kreitman Tests of SARS-CoV-2 Nsp16 among human clades and variants with an outgroup (RaTG13)[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
DoS	-0.62	-0.21	-0.41	-0.64	-0.57	-0.8	-0.5
NI	-	-	-	-	-	-	-
Alpha Value	-	-	-	-	-	-	-
Fisher's exact test. P-value (two-tailed)	0***	0.01*	0***	0***	0***	0***	0**
Syn. Fixed differences between species	38	37	36	38	39	39	39
Syn. Polymorphic sites	5	11	16	4	3	1	2
NonSyn. Polymorphic sites	8	3	11	7	4	4	2
NonSyn. Fixed differences between species	0	0	0	0	0	0	0

Looking at the population genetic summary statistics for SARS-CoV-2 Nsp8, Nsp10, and Nsp16 among their clades and variants indicated different results. Nsp8 of the S clade had the highest segregating sites number among the clades, and the Nsp8 of the Alpha variant had the highest segregating sites number among variants. Moreover, the number of singleton changes was higher than the parsimony-informative sites except for the Beta and Delta variants. The Beta variant value was zero and the Delta variant had more parsimony informative sites. In addition, all clades and variants had more replacement polymorphisms other than the Beta variant. The nucleotide diversity for the Nsp8, the L clade had the highest value among the clades and the Alpha variant had the highest value among the variants. Also, the nucleotide diversity with Jukes-Cantor correction applied estimates for Nsp8 showed that all clades and variants were primarily driven by synonymous sites except for Beta and Delta variants.

The result of the Beta variant was zero and the Delta variant was primarily driven by nonsynonymous sites. For Nsp10, the O clade had the highest segregating number among the clades, and the Gamma variant had the highest segregating number among the variants. Moreover, the number of singleton informative sites was higher than parsimony informative sites except for the Delta variant which was equal value for both sites. In addition, the S clade had more synonymous polymorphism, although the O clade, Alpha variant, Beta variant, and Gamma variant were primarily driven by replacement polymorphism. L clade and Delta variant were equal values for both polymorphisms. The nucleotide diversity of the Nsp10 showed that the L clade had the highest value among the clades and the Gamma variant had the highest value among the variants. Furthermore,

the nucleotide diversity with Jukes-Cantor correction applied estimates for Nsp10 showed that all clades and variants were primarily driven by synonymous sites except for the Alpha and Beta variants. The Alpha and Beta variants were primarily driven by synonymous sites.

Lastly, the Nsp16 data showed that the highest segregating number was in the O clade among the clades, and the Alpha variant had the highest segregating number among the variants. Moreover, the number of singleton informative sites was higher than parsimony informative sites for all clades and variants. In addition, L and S clades had more synonymous polymorphism compared to other clades and variants. The nucleotide diversity of the Nsp16 indicated that the S clade had the highest value among the clades, and the Beta variant had the highest value among the variants. Also, the nucleotide diversity with Jukes-Cantor correction applied estimates for Nsp16 showed that the L clade and Beta variant were primarily driven by synonymous sites. On the other hand, S, O, Alpha, Gamma, and Delta were primarily driven by nonsynonymous sites.

Table 21. Population genetic summary statistics for nucleotide diversity of SARS-CoV-2 Nsp8 among SARS-CoV-2 clades and variants[1]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Nonsyn. sites	462.51	462.53	462.52	462.47	462.5	462.5	462.51
S	5	12	10	8	0	3	1
Eta	5	12	10	8	0	3	1
Sing.	5	6	9	6	0	3	0
Par.	0	6	1	2	0	0	1
Syn. Pol.	2	5	4	3	0	1	0
Rep. Pol.	3	7	6	5	0	2	1
π (Pi) All Sites	9	7.9	2.1	4.9	0	1.4	1.3
Theta-W All Sites	14.5	34.8	29.3	27.1	0	10.4	3.8
π (JC) All Sites	9	7.9	2.1	4.9	0	1.4	1.3
π (JC) Syn. Sites	1.6	17.4	3.5	5.6	0	2.1	0
π (JC) Nonsyn. Sites	0.7	5.2	1.7	4.7	0	1.2	1.7
H	6	11	11	9	1	4	2
Hd	0.05	0.28	0.11	0.25	0	0.08	0.08

Table 22. Population genetic summary statistics for nucleotide diversity of SARS-CoV-2 Nsp10 among SARS-CoV-2 clades and variants[1]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Syn. sites	97.83	97.83	97.83	97.8	97.82	97.82	97.84
Nonsyn. sites	319.17	319.17	319.17	319.2	319.18	319.18	319.16
S	2	6	11	4	2	5	2
Eta	2	6	11	4	2	5	2
Sing.	2	6	10	3	2	4	1
Par.	0	0	1	1	0	1	1
Syn. Pol.	1	4	5	1	0	2	1
Rep. Pol.	1	2	6	3	2	3	1
π (Pi) All Sites	5	1.5	3.3	3.5	1.7	4.6	2.8
Theta-W All Sites	8.3	24.8	46	19.3	10.5	24.7	10.7
π (JC) All Sites	5	1.5	3.3	3.5	1.7	4.6	2.8
π (JC) Syn. Sites	1.1	4.4	5.9	2.5	0	5.7	8.1
π (JC) Nonsyn. Sites	0.3	0.7	2.5	3.8	2.3	4.3	1.3
H	3	7	11	5	3	6	3
Hd	0.02	0.06	0.12	0.14	0.07	0.19	0.12

Table 23. Population genetic summary statistics for nucleotide diversity of SARS-CoV-2 Nsp16 among SARS-CoV-2 clades and variants[1]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Syn. sites	189	189.03	189.05	189	188.94	188.97	189
Nonsyn. sites	702	701.97	701.95	702	702.06	702.03	702
S	13	17	27	11	7	5	4
Eta	13	17	27	11	7	5	4
Sing.	13	11	21	8	5	4	4
Par.	0	6	6	3	2	1	0
Syn. Pol.	8	11	11	4	2	1	0
Rep. Pol.	5	3	16	7	5	4	4
π (Pi) All Sites	1.6	5.1	5	3.8	4	1.9	1.8
Theta-W All Sites	25	32.8	52.6	24.7	17.1	11.5	10
π (JC) All Sites	1.6	5.1	5	3.8	4	1.9	1.8
π (JC) Syn. Sites	4.5	2.2	4.2	1.3	7.5	1.5	0
π (JC) Nonsyn. Sites	0.8	5.9	5.2	4.5	3.1	2	2.3
H	12	16	26	10	8	6	5
Hd	0.11	0.26	0.35	0.25	0.30	0.16	0.16

In addition to population genetic summary statistics, neutrality tests were performed for the assessment of the selection of viral genomes. Tajima's D values for Nsp8 were significant and negative other than the Beta, Gamma and Delta variants. The Gamma and Delta variant values were not significant and for the Beta variant, the result was not estimated because there was no mutation in the variant for Nsp8.

In addition, Fu-Li's D* and Fu-Li's F* test results and the Tajima's D results were consistent. The O clade had the most negative value compared to the other clades and variants. The test values are negative and significant when there is an excess of rare variants compared to intermediate frequency variants as a consequence of background selection or population expansion.

Table 24. Neutrality tests summary statistics for SARS-CoV-2 Nsp8 among viral clades and variants[2]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
TD	-1.82 *	-1.90*	-2.20**	-2.05*	NA	-1.65 #	-0.87 ##
TD - Cod.	-1.82 *	-1.90*	-2.20**	-2.05*	NA	-1.65 #	-0.87 ##
TD – Syn.	-1.29 ##	-1.43 ##	-1.69#	-1.64 #	NA	-1.06 ##	NA
TD - Nonsyn.	-1.52##	-1.72#	-1.91 *	-1.73 #	NA	-1.42 ##	-0.87 ##
TD - Silent	-1.29 ##	-1.43 ##	-1.69 #	-1.64 #	NA	-1.06 ##	NA
Fu-Li's D*	-4.71**	-2.71*	-5.54**	-3.42**	0	-3.24*	0.54 ##
Fu-Li's F*	-4.43**	-2.89*	-5.18**	-3.50**	0	-3.21**	0.16 ##

Fu-Li's F and Fu-Li's D test values using Bat coronavirus RaTG13 as an outgroup and the values of neutrality tests for the Nsp8 were consistent. The values were significant and negative other than the Beta, Gamma and Delta variants. The Gamma and Delta variant values were not significant, and the result of the Beta variant was zero because there was no mutation in the variant.

Also, Fay and Wu's Hn values were negative only for L, S, and O clades. But, the results were not significant.

Table 25. Fu-Li's Tests with an Outgroup (RaTG13) for human SARS-CoV-2 Nsp8 among viral clades and variants[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Fu and Li's D	-3.61**	-2.77 *	-4.49 **	-3.55**	0	-3.30**	-0.54 ##
Fu and Li's F	-3.57**	-2.94**	-4.36**	-3.61**	0	-3.27**	-0.16 ##
Fay and Wu's Hn	-1.94	-1.39	-1.88	0.28	0	0.08	0.08
Fay and Wu's Hn Normalized	-2.2	-0.84	-1.4	0.21	0	0.12	0.2

Secondly, Tajima's D values for Nsp10 were significant and negative for the S clade, O clade, and Gamma variant. For other clades and variants, values were not estimated since there was no mutation in the variants. Moreover, Fu-Li's D* and Fu-Li's F* test values, and the Tajima's D results were consistent. O clade had the most negative significant value compared to other clades and variants. Tajima's D, Fu-Li's D*, and Fu-Li's F* tests all reveal that the population has a high number of rare haplotypes as a consequence of background selection or population expansion if they are statistically significant and negative. Fu-Li's F and Fu-Li's D test values using Bat coronavirus RaTG13 as an outgroup and the values of neutrality tests for the Nsp10 were consistent. Again, the O clade had the most negative significant value compared to other clades and variants for the tests. The Fay and Wu's Hn values were almost zero except for the Gamma variant. But, the test results were not significant.

Table 26. Neutrality tests summary statistics for SARS-CoV-2 Nsp10 among viral clades and variants[2]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
TD	-1.29##	-1.92*	-2.25**	-1.68#	1.45 ##	-1.84*	-1.31##
TD - Cod.	-1.29##	-1.92*	-2.25**	-1.68 #	1.45 ##	-1.84*	-1.31##
TD – Syn.	-0.96##	-1.69#	-1.83 *	-1.05 #	NA	-1.42##	-0.87##
TD – Nonsyn.	-0.96##	-1.29##	-1.91*	-1.49##	1.45##	-1.49##	-1.10##
TD - Silent	-0.96##	-1.69#	-1.83*	-1.05 ##	NA	-1.42##	-0.87##
Fu-Li's D*	-3.06*	-5.10**	-5.83**	-2.58*	-2.58*	-3.01*	-0.89##
Fu-Li's F*	-2.95*	-4.77**	-5.41 **	-2.69*	-2.61*	-3.09**	-1.18##

Table 27. Fu-Li's Tests with an Outgroup (RaTG13) for human SARS-CoV-2 Nsp10 among viral clades and variants[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Fu and Li's D	-3.08**	-5.17**	-5.65**	-2.63*	-2.63*	-2.05 #	-0.92 ##
Fu and Li's F	-2.96**	-4.82**	-5.26**	-2.74*	-2.66*	-2.34*	-1.21 ##
Fay and Wu's Hn	0.02	0.06	0.12	0.14	0.07	-1.78	0.11
Fay and Wu's Hn Normalized	0.04	0.06	0.09	0.17	0.13	-1.82	0.2

Lastly, the Tajima's D values of Nsp16 were significant and negative for all clades and variants. O clade had the most negative significant value compared to other clades and variants for Tajima's D results. Furthermore, Fu-Li's D* and Fu-Li's F* test values, and the Tajima's D results were consistent. Fu-Li's D* and Fu-Li's F* test values were negative and significant. Tajima's D, Fu-Li's D*, and Fu-Li's F* tests all reveal that the population has a high number of rare haplotypes as a consequence of background selection or population expansion if they are statistically significant and negative. When Fu-Li's F and Fu-Li's D tests were conducted using Bat coronavirus RaTG13 as an outgroup for the Nsp16 gene, there was still an excess of rare variants. The values of neutrality tests and Fu-Li's F and Fu-Li's D were consistent. In addition, the L clade had the most negative significant value for Fu-Li's D*, Fu-Li F*, Fu-Li's D, and Fu-Li's F tests. The Fay and Wu's Hn values of the L, S, O, and Alpha were negative, and the Beta, Gamma, and Delta values were almost zero. But, the test results were not significant.

Table 28. Neutrality tests summary statistics for SARS-CoV-2 Nsp16 among viral clades and variants[2]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
TD	-2.34**	-2.21**	-2.56 ***	-2.27**	-1.99*	-1.90*	-1.86*
TD - Cod.	-2.34**	-2.21**	-2.56 ***	-2.27**	-1.99*	-1.90*	-1.86*
TD – Syn.	-2.09*	-0.77 ##	-1.78*	-1.05 ##	-1.18 ##	-1.06 ##	NA
TD – Nonsyn.	-1.82*	-2.20**	-2.49***	-2.22**	-1.90*	-1.75 #	-1.86*
TD - Silent	-2.09*	-0.77 ##	-1.78*	-1.05##	-1.18##	-1.06 ##	NA
Fu-Li's D*	-7.09**	-4.50**	-6.65**	-3.68**	-2.73*	-3.01*	-3.40**
Fu-Li's F*	-6.37**	-4.34**	-6.0**	-3.79**	-2.93*	-3.11**	-3.42**

Table 29. Fu-Li's Tests with an Outgroup (RaTG13) for human SARS-CoV-2 Nsp16 among viral clades and variants[3]

	Clade/Variant						
Parameters	**L (N=185)**	**S (N=185)**	**O (N=175)**	**Alpha (N=82)**	**Beta (N=55)**	**Gamma (N=72)**	**Delta (N=53)**
Fu and Li's D	-6.58**	-3.97**	-6.09**	-3.86**	-2.44 #	-3.10**	-3.11**
Fu and Li's F	-5.98**	-3.98**	-5.60**	-3.93**	-2.67*	-3.20**	-3.14**
Fay and Wu's Hn	-1.85	-3.6	-5.58	-1.59	0.31	0.16	0.12
Fay and Wu's Hn Normalized	-1.05	-1.83	-1.84	-0.93	0.27	0.17	0.16

3.4.5. Sliding Window, Nature of Amino acid Changes, and Structure Analyses of Nsp8

To compare nucleotide diversity of nonsynonymous changes for all clades and variants, sliding window analyses were performed. Results showed that the highest nonsynonymous nucleotide diversities were on the 100th and 553rd nucleotides.

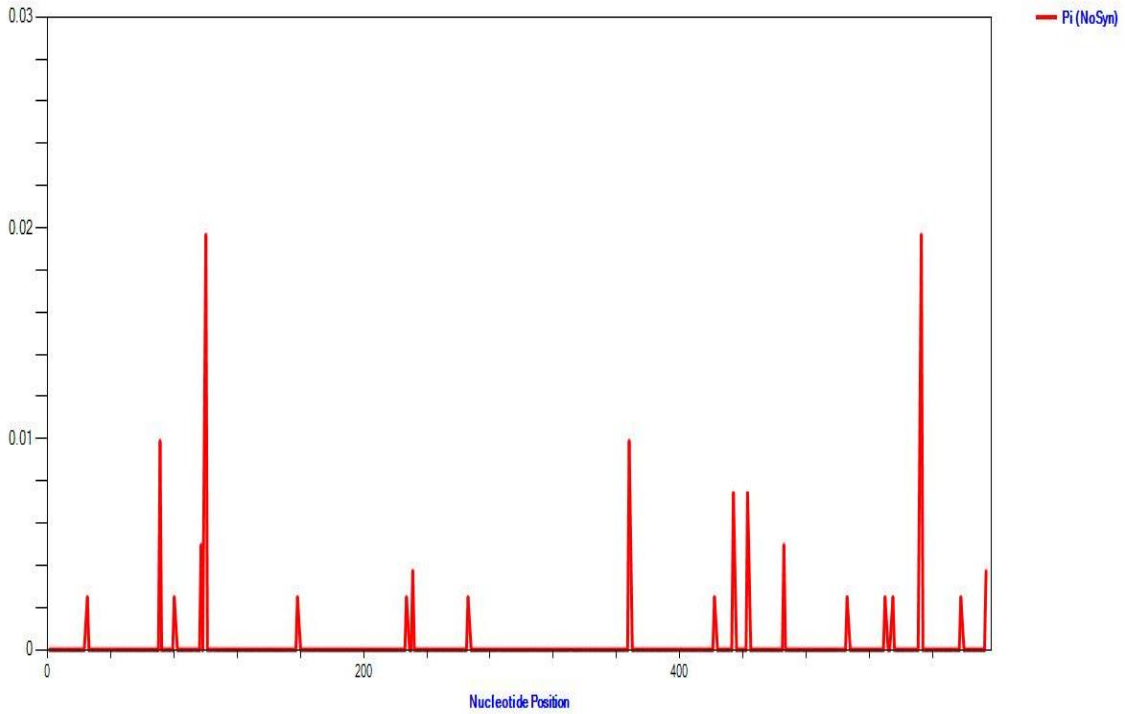

Figure 6. Sliding Window analysis of nonsynonymous nucleotide diversity for Nsp8

The 100th nucleotide was in the codon that codes for the V34 residue, and the 553rd nucleotide was in the codon that codes for the I185 residue. The 100th nucleotide and the 553rd nucleotide of some sequences mutated in the S clade. The residues changed from Valine to Phenylalanine for the 100th nucleotide and changed from Isoleucine to Valine for the 553rd nucleotide. Other than the V34 and I185 residues, regions that had nonsynonymous nucleotide diversity in Nsp8 are listed in Table 30.

Table 30. Polymorphic Nsp8 residues and associated sliding window nonsynonymous nucleotide diversity estimates in the examined SARS-CoV-2 clades/variants

Window	**Pi (NoSyn) (10^{-4})**	**Clade/Variant**	**Aminoacid**	**Change**	**Domain**
25-25	25	S	9	L-> F	Coil
71-71	99	Alpha	24	Q-> R	Helix
80-80	25	Alpha	29	A-> V	Coil
97-97	50	O, S	33	V-> F	Helix
100-100	197	S	34	V-> F	Helix
158-158	25	O	53	A-> V	Helix
227-227	25	Alpha	76	S-> F	Helix
231-231	37	O	77	E->D	Helix
266-266	25	Alpha	89	T-> I	Helix
368-368	99	S, O	123	T-> I	Helix
422-422	25	S	141	T-> M	Strand
434-434	74	O, Delta	145	T-> I	Coil
443-443	74	Alpha, L	148	T-> I	Strand
466-466	50	S	156	I-> V	Strand
506-506	25	O	169	L-> H	Strand
530-530	25	L	177	S-> L	Coil
535-535	25	Gamma	179	N-> H	Coil
553-553	197	S	185	I-> V	Strand
578-578	25	L	193	S-> F	Coil
594-594	37	Gamma	198	Q-> H	Coil

For Nsp8, a 3d structure model was predicted using I-Tasser. The best C-scored model was chosen to display on UCSF Chimera and the distribution of amino acid change was mapped on the predicted structure with Render by Attribute interface.

The percentage distribution of amino acid changes on the Nsp8 helix, strand, and coil regions were 40 percent, 25 percent, and 35 percent, respectively. In the figure, blue-colored residues V34 and I185 showed the highest nonsynonymous nucleotide diversity. Other residues with lower nonsynonymous nucleotide diversity had a color range between pink to maroon, respectively.

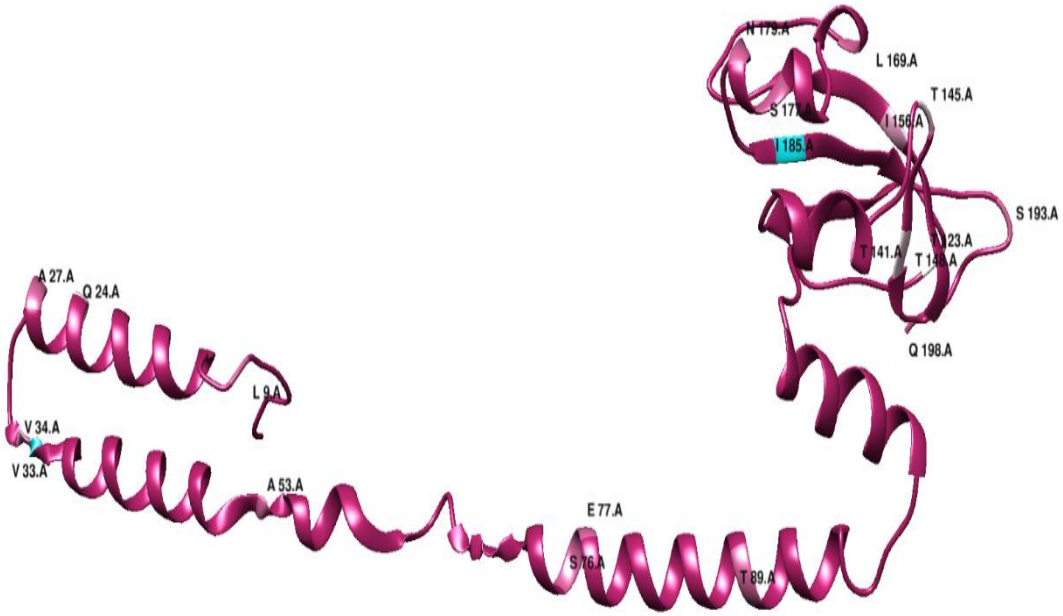

Figure 7. Display of the Nsp8 secondary structure using I-Tasser with UCSF Chimera. Color coding ranges from highest (blue) to lowest (maroon) nonsynonymous nucleotide diversity

3.4.6. Interpretation Of Observed Amino Acid Changes In Terms Of Nsp8 Function

Nsp8 consists of 198 amino acids and contains a helical N-terminal domain and a C-terminal domain. Nsp8 and Nsp7 work together to increase the efficiency of Nsp12, also known as RNA-dependent RNA polymerase (RdRp) [110]. These three components carry out the replication of the virus essential to ensure viral life sustainability. An experiment determined that R80, V83, T84, A86, M87, M90, L91, F92, M94, L95, N108, D112, G113, C114, V115, P116, L117, N118, P121, A125, K127, L128, M129, V130, V131, P133, P183, I185, R190 residues of Nsp8 have interaction with Nsp12, and M87, G88, T89, L91, F92, M94, L95, R96, N100, L103, I106, I107, P116, L117, I119, I120, L122 residues of Nsp8 have interaction with Nsp7 [111].

In this study, the I185 residue which has a role in interaction with Nsp8-Nsp12, and the T89 residue which has a role in interaction with Nsp8-Nsp7 were among the polymorphic residues.

Other than the I185 and T89, the conserved residues that are important in interacting with Nsp7 and Nsp12 can be targetted with small molecules to inhibit the activity of RdRp.

3.4.7. Sliding Window, Nature of Amino acid Changes, and Structure Analyses of Nsp10

To compare nucleotide diversity of Nsp10 nonsynonymous changes for clades and variants, sliding window analyses were performed. Results showed that the highest nonsynonymous nucleotide diversity was on the 305th nucleotide.

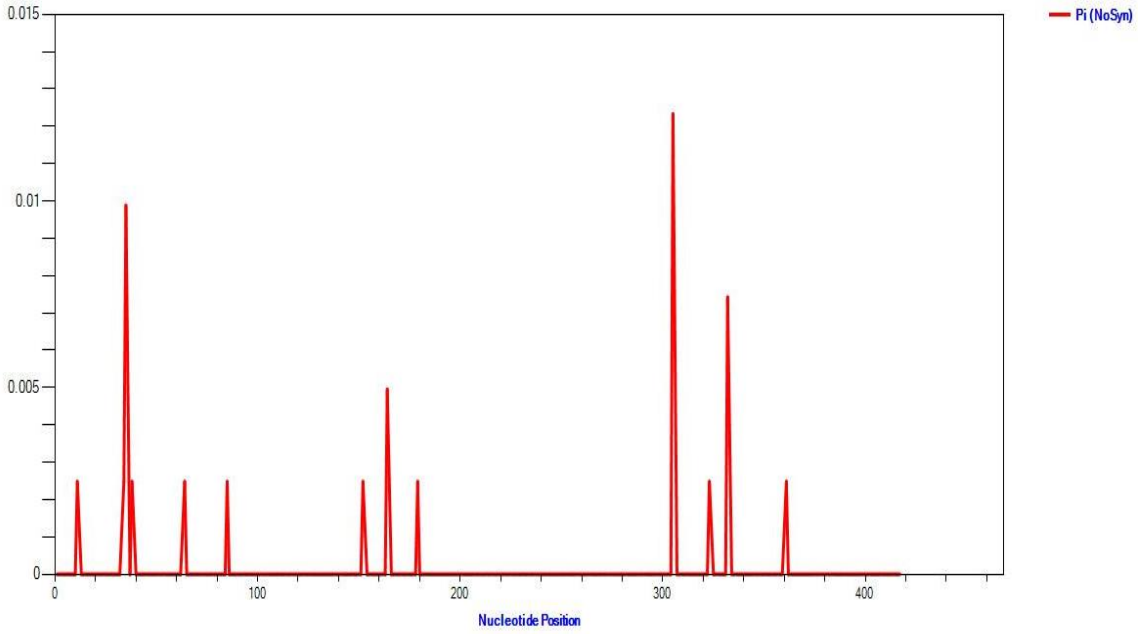

Figure 8. Sliding Window analysis of nonsynonymous nucleotide diversity for Nsp10

The 305th nucleotide was in the codon that codes for the T102 residue. In the O clade and the Gamma variant, some sequences changed the residue from Threonine to Isoleucine. Other than the T102 residue, regions that had nonsynonymous nucleotide diversity for Nsp10 are listed in Table 31.

Table 31. Polymorphic Nsp10 residues and associated sliding window nonsynonymous nucleotide diversity estimates in the examined SARS-CoV-2 clades/variants

Window	Pi (NoSyn) (10^{-4})	Clade/Variant	Aminoacid	Change	Domain
11-11	25	Gamma	4	A-> V	Coil
35-35	99	Alpha, O, S	12	T->I T->A	Coil
38-38	25	L	13	V-> E	Helix
64-64	25	Alpha	22	D-> N	Coil
85-85	25	Gamma	29	D-> N	Helix
152-152	25	Beta	51	T->I	Coil
164-164	50	Delta, O	55	I->T	Strand
179-179	25	O	60	E->G	Coil
305-305	120	Gamma, O	102	T->I	Coil
323-323	25	O	108	V->A	Coil
332-332	74	Beta, O, S	111	T->I	Strand
361-361	25	Alpha	121	G->S	Coil

For Nsp10, a 3d structure model was predicted using I-Tasser. The best C-scored model was chosen to display on UCSF Chimera and the distribution of amino acid change was mapped on the predicted structure with Render by Attribute interface.

The percentage distribution of amino acid changes on the Nsp10 helix, strand, and coil regions were 16.7 percent, 16.7 percent, and 66.6 percent, respectively. In the figure, the blue-colored residue T102 showed the highest nonsynonymous nucleotide diversity. Other residues with lower nonsynonymous nucleotide diversity had a color range between white to pink and maroon, respectively.

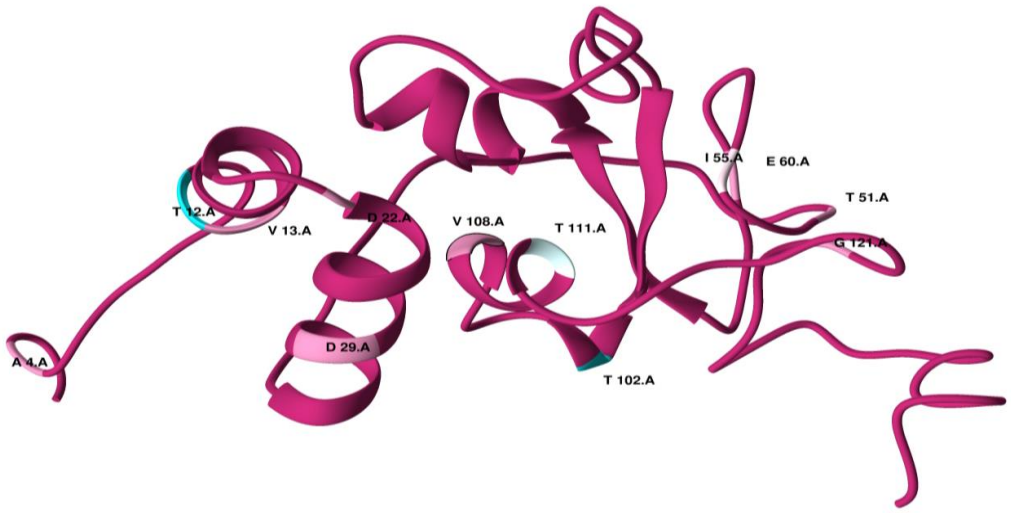

Figure 9. Display of the Nsp10 secondary structure using I-Tasser with UCSF Chimera. Color coding ranges from highest (blue) to lowest (maroon) nonsynonymous nucleotide diversity

3.4.8. Interpretation Of Observed Amino Acid Changes In Terms Of Nsp10 Function

Nsp10 consists of 140 amino acids, an N-terminal domain and a C-terminal domain includes. On the secondary structure, the protein has 5 α-helices and an antiparallel β-strands. Also, Nsp10 has two zinc fingers. Helices α3 and α4 coordinate the first zinc finger (aminoacids C74, C77, H83, and C90) and the second zinc finger (amino acids C117, C120, C128, and C130) is located at the C terminus[112].

Nsp10 interacts with both the Nsp14 and Nsp16. The linkage between the ExoN region of nsp14 and Nsp10 is essential because the lack of the interaction leads to damage to the proofreading mechanism of the Nsp14 and increases the mutation rates. Also, zinc finger structures are associated with the stability and enzymatic activity of ExoN [112].

A study on Sars-CoV showed that A1, N3, E6, F16, F19, V21, N40, K43, T58, S72, H80, C90, K93 and Y96 residues of Nsp10 have interaction with K9, D10, T5, F60, M62, Y64, T25, C39, D41, A23, D16, N19, Y51 and H19 residues of Nsp14, respectively [113]. The residues also can be essential for interaction with SARS-CoV-2 Nsp10 and Nsp14.

In this study, none of the residues had a role in interaction with Nsp10-Nsp14 were the polymorphic residues. The conserved residues that are important in interacting with Nsp10 and Nsp14 can be targetted with small molecules to inhibit the activity of ExoN.

3.4.9. Sliding Window, Nature of Amino acid Changes, and Structure Analyses of Nsp16

To compare nucleotide diversity of Nsp16 nonsynonymous changes for clades and variants, sliding window analyses were performed. Results showed that the highest nonsynonymous nucleotide diversities are on 479th and 646th nucleotides.

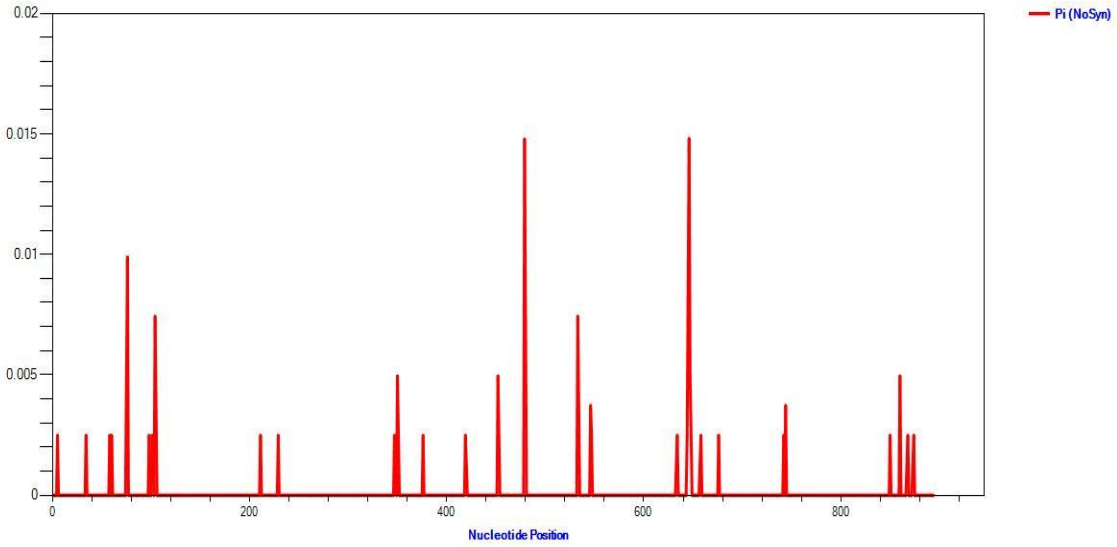

Figure 10. Sliding Window analysis of nonsynonymous nucleotide diversity for Nsp16

The 479th nucleotide was in the codon that codes for the K160 residue. and the 646th nucleotide was in the codon that codes for the R148 residue. Changes on the 479th nucleotide of some sequences turned the residue from Lysine to Arginine in the L, O, S, and Alpha, and also changes on the 646th nucleotide of some sequences on L, O, S, and Alpha turned aminoacid from Asparagine and Cysteine. Other than K160 and R148, regions of the Nsp16 that had nonsynonymous nucleotide diversity are listed in Table 32.

Table 32. Polymorphic Nsp16 residues and associated sliding window nonsynonymous nucleotide diversity estimates in the examined SARS-CoV-2 clades/variants

Window	**Pi (NoSyn) (10^{-4})**	**Clade/Variant**	**Aminoacid**	**Change**	**Domain**
5-5	25	Delta	2	S->N	Coil
34-34	25	L	12	P-> S	Helix
58-58	25	Beta	20	M-> I	Coil
60-60	25	L	20	M-> V	Coil
76-76	99	S	26	D-> Y	Coil
98-98	25	Alpha	33	S-> I	Coil
101-101	25	O	34	A-> V	Coil
104-104	74	Alpha	35	T-> I	Coil
211-211	25	O	71	G-> C	Coil
229-229	25	L	77	G->R	Coil
347-347	25	Gamma	116	A-> V	Coil
350-350	50	Gamma, L	117	T-> I	Coil
376-376	25	O	126	L-> F	Strand
419-419	25	O	140	T->I	Coil
452-452	50	Beta	151	T->I	Helix

(cont. on next page)

Table 32 (cont.)

479-479	148	Alpha, L, O, S	160	K-> R	Helix
533-533	74	Beta	178	A-> V	Coil
546-546	37	Delta	182	K-> N	Helix
547-547	25	S	183	L-> I	Helix
634-634	25	L	212	L-> I	Coil
644-644	25	Alpha	215	P-> L	Coil
646-646	148	Alpha, L, O, S	216	R-> N, R-> C	Coil
647-647	50	L	220	D-> N	Coil
658-658	25	O	226	A-> S	Strand
676-676	25	O	238	Q-> H	Coil
742-742	25	Alpha	248	S -> G	Coil
744-744	37	Gamma	248	S-> R	Coil
850-850	25	O	284	E-> K	Coil
860-860	50	Gamma	287	R-> I	Coil
868-868	25	Beta	290	I-> V	Coil
874-874	25	O	292	S-> G	Coil

For Nsp16, a 3d structure model was predicted using I-Tasser. The best C-scored model was chosen to display on UCSF Chimera and the distribution of amino acid changes was mapped on the predicted structure with Render by Attribute interface.

The percentage distribution of amino acid changes on the Nsp16 helix, strand, and coil regions were 16 percent, 6 percent, and 78 percent, respectively. In the figure, blue-colored residues K148 and R216 showed the highest nonsynonymous nucleotide diversity. Other residues with lower nonsynonymous nucleotide diversity had a color range between white to pink and maroon, respectively.

Figure 11. Display of the Nsp16 secondary structure using I-Tasser with UCSF Chimera. Color coding ranges from highest (blue) to lowest (maroon) nonsynonymous nucleotide diversity

3.4.10. Interpretation Of Observed Amino Acid Changes In Terms Of Nsp16 Function

Nsp16, also known as 2′- O-methyltransferase, consists of 298 amino acids with an N-terminal domain and C-terminal domain included, and on the secondary structure, the protein has 12 β-strands and 12 α-helices [114]. Nsp16 plays a role in mRNA capping which is essential for giving stability and reducing an innate immune response in coronaviruses [115]. The last stage of the mRNA cap synthesis process occurred in Nsp16. Nsp16 methylates the ribose2′-O of the first nucleotide in Cap-0 mRNA (m7G0pppA1-RNA) to produce Cap-1 mRNA (m7G0pppA1m-RNA) using S-adenosyl methionine (SAM) with the presence of Nsp10 [116]. Nsp10 functions as a co-factor for Nsp16, stabilizing the SAM-binding pocket and boosting the enzymatic activity considerably. N43, Y47, G71, G81, D99, N101, C115, and D130 residues of Nsp16 engage with the SAM-binding pocket and all these residues are conserved in SARS-CoV and MERS-CoV and SARS-CoV-2 [114] [117].

In this study, only the G71 residue was polymorphic. The rest of the residues were conserved. Considering the nonchanging residues on the SARS-CoV-2 clades and variants, the conserved SAM-interaction residues can be the key to developing pan-antiviral inhibitors by targeting the SAM-binding pocket.

N40, V42, K43, M44, L45, C46, T47, P59, G70, A71, C77, R78, K93, G94, and Y96 residues of Nsp10 have interaction with K38, G39, I40, M41, V44, K76, V78, P80, A83, R86, Q87, V104, S105, D106, L244, and M247 residues of Nsp16, respectively [118].

In this study, none of the residues has a role in interaction with Nsp10-Nsp16 were polymorphic. The conserved residues that are important in interacting with Nsp10 and Nsp16 can be targetted with small molecules. Also, to inhibit the activity of 2′-O-methyltransferase can be used Sinefungin which is a pan inhibitor of Mtase and interacts with SAM-engaging residues of Nsp16 [119]. Finally, the construction of Nsp16 mutant viruses using a live attenuated virus vaccine can be considered based on conserved residues [120].

CHAPTER 4

CONCLUSION

The main hypothesis of this thesis was that molecular evolutionary and population genetic analyses of the SARS-CoV-2 genome can identify the most conserved genome regions indicated by the strongest negative selection, and these highly conserved genome regions can be targets for small molecules (or drugs) or used for new vaccine approaches. Therefore, the molecular evolutionary dynamics of SARS-CoV-2 human virus clades and variants that cause COVID-19 disease were examined by using population genetic statistical tests, and the DNA sequence differences between human and bat viral sequences and within human viruses were revealed.

To test the hypothesis, firstly, comprehensive molecular population genetic analyses of virus genomes were performed with all viral genes for populations of viral variants of concern (Alpha, Beta, Gamma, Delta) associated with faster spread, and the very first clades seen at the beginning of the pandemic (L, S, O). Intraspecific comparisons focused only on human-specific SARS-CoV-2 variants. Considering the whole genome sequences, within human viral sequence (intraspecies) analyses showed that S clade had the highest overall nucleotide diversity and amino acid changing nucleotide diversity. When individual viral genes were analyzed, the ORF7b protein found in the Delta variant showed the highest nucleotide and amino acid changing nucleotide diversity. For population genetic tests between Nsps of the ORF1ab protein, the Nsp6 protein of the Delta variant had the highest nucleotide diversity, and the Nsp6 protein of the S clade had the highest amino acid changing nucleotide diversity.

In neutrality tests such as Tajima's D, Fu-Li's D*'s, and Fu-Li's F*, all clades and variants for whole-genome proteins had a negative value. Considering the whole genome sequences, the L clade had the most negative Tajima's D, Fu-Li D*'s, and Fu-Li's F* values. When individual viral genes were analyzed, the Alpha variant on the ORF1a protein had the most negative value. Moreover, the L clade on the ORF1b protein had the most negative Fu-Li D*'s and Fu-Li's F* values. For population genetic tests between Nsps of the ORF1ab protein, the Nsp3 protein of the Alpha variant had the most negative Tajima's D, and the Nsp3 protein of the O clade had the most negative Fu-Li's D*'s and

Fu-Li's F* values. The results showed that negative selection is highly effective on all genomes of viral clades and variants.

For interspecific comparisons, bat RaTG13 is used as an outgroup and several interspecific population genetic tests are conducted. Analyses showed that there was no fixed amino acid change divergence between the human SARS-CoV-2 and RaTG13 Membrane, Nsp8, Nsp10, and Nsp16, suggesting high evolutionary constraints on these proteins. Although sequence comparisons with respect to RaTG13 suggested evolutionary conservation, considering human SARS-CoV-2 Membrane, Nsp8, Nsp10, and Nsp16 sequence analyses, several amino acid changing variants were observed in all examined clades and variants.

The second part of testing the main hypothesis of the thesis involved evaluation of amino acid sequence, protein structure, function, and interaction with other virus genes of these identified most conserved genes. Nearly none of the observed amino acid changing variants in these proteins coincide with functionally important residues in human viral sequences. Regions of these four proteins, other than aminoacid-changing mutations, have been identified as the most conserved regions, suggesting that these regions could be potential targets for vaccine and drug strategies.

It will be essential to carry out this study by including other variants of concern such as the Omicron that have emerged over time. Our study contributes to the viral molecular evolutionary research in terms of population genetic statistical tests and draws attention to the need for more examination of the most conserved regions. This work could be a helpful resource for creating small molecule libraries to be used in future studies.

[1] Syn: Synonymous sites, Nonsyn: Nonsynonymous sites, S: Number of segregating sites, Eta: Number of mutations, Sing: Singleton variable sites, Par. : Parsimony informative sites, Syn. Pol. : Number of synonymous polymorphisms, Rep. Pol. : Number of replacement (nonsynonymous) polymorphisms, JC: Jukes-Cantor correction applied estimates, H: Number of haplotypes, Hd: Haplotype diversity. θ and π values represent percent sequence diversity and for exact estimates, table values should be multiplied by 10-4. N: Sample Size.

[2] TD: Tajima's D test, Cod.: Coding sites, Syn: Synonymous sites, Nonsyn: Nonsynonymous sites, Silent: Silent sites. * represent $P < 0.05$, ** represent $P < 0.01$, *** represent $P < 0.001$, # represent $0.10 > P > 0.05$.

[3] ** represent $0.001 < P < 0.01$, *** represent $P < 0.001$

REFERENCES

(1) Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. L.; Lauber, C.; Leontovich, A. M.; Neuman, B. W.; Penzar, D.; Perlman, S.; Poon, L. L. M.; Samborskiy, D. v; Sidorov, I. A.; Sola, I.; Ziebuhr, J. The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2. *Nature Microbiology* **2020**, *5* (4), 536–544.

(2) Zhong, N. S.; Zheng, B. J.; Li, Y. M.; Poon, L. L. M.; Xie, Z. H.; Chan, K. H.; Li, P. H.; Tan, S. Y.; Chang, Q.; Xie, J. P.; Liu, X. Q.; Xu, J.; Li, D. X.; Yuen, K. Y.; Peiris, J. S. M.; Guan, Y. Epidemiology and Cause of Severe Acute Respiratory Syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **2003**, *362* (9393), 1353–1358.

(3) 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. Structure of MERS-CoV Spike Receptor-Binding Domain Complexed with Human Receptor DPP4. *Cell Research* **2013**, *23* (8), 986–993.

(4) WHO. *Coronavirus Disease (COVID-19) 11 October 2020*. Weekly Situation Report. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20201012-weekly-epi-update-9.pdf.

(5) Delamater, P. L.; Street, E. J.; Leslie, T. F.; Yang, Y. T.; Jacobsen, K. H. Complexity of the Basic Reproduction Number (R0). *Emerging Infectious Diseases* **2019**, *25* (1), 1.

(6) Achaiah, N. C.; Subbarajasetty, S. B.; Shetty, R. M. R0 and Re of COVID-19: Can We Predict When the Pandemic Outbreak Will Be Contained? *Indian Journal of Critical Care Medicine : Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine* **2020**, *24* (11), 1125.

(7) Liu, Y.; Rocklöv, J. The Reproductive Number of the Delta Variant of SARS-CoV-2 Is Far Higher Compared to the Ancestral SARS-CoV-2 Virus. *J Travel Med* **2021**, *28* (7).

(8) Billah, M. A.; Miah, M. M.; Khan, M. N. Reproductive Number of Coronavirus: A Systematic Review and Meta-Analysis Based on Global Level Evidence. *PLoS ONE* **2020**, *15* (11 November).

(9) Azer, S. A. COVID-19: Pathophysiology, Diagnosis, Complications and Investigational Therapeutics. *New Microbes and New Infections*. Elsevier Ltd September 1, 2020.

(10) Rodriguez-Morales, A. J.; Cardona-Ospina, J. A.; Gutiérrez-Ocampo, E.; Villamizar-Peña, R.; Holguin-Rivera, Y.; Escalera-Antezana, J. P.; Alvarado-Arnez, L. E.; Bonilla-Aldana, D. K.; Franco-Paredes, C.; Henao-Martinez, A. F.; Paniz-Mondolfi, A.; Lagos-Grisales, G. J.; Ramírez-Vallejo, E.; Suárez, J. A.; Zambrano, L. I.; Villamil-Gómez, W. E.; Balbin-Ramon, G. J.; Rabaan, A. A.; Harapan, H.; Dhama, K.; Nishiura, H.; Kataoka, H.; Ahmad, T.; Sah, R. Clinical, Laboratory and Imaging Features of COVID-19: A Systematic Review and Meta-Analysis. *Travel Med Infect Dis* **2020**, *34*.

(11) Cevik, M.; Bamford, C. G. G.; Ho, A. COVID-19 Pandemic—a Focused Review for Clinicians. *Clinical Microbiology and Infection*. Elsevier B.V. July 1, 2020, pp 842–847.

(12) Zhou, P.; Yang, X.-L.; 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.-D.; Liu, M.-Q.; Chen, Y.; Shen, X.-R.; Wang, X.; Zheng, X.-S.; Zhao, K.; Chen, Q.-J.; Deng, F.; Liu, L.-L.; Yan, B.; Zhan, F.-X.; Wang, Y.-Y.; Xiao, G.-F.; Shi, Z.-L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2012**, *579*.

(13) Boson, B.; Legros, V.; Zhou, B.; Siret, E.; Mathieu, C.; Cosset, F. L.; Lavillette, D.; Denolly, S. The SARS-CoV-2 Envelope and Membrane Proteins Modulate Maturation and Retention of the Spike Protein, Allowing Assembly of Virus-like Particles. *Journal of Biological Chemistry* **2021**, *296*, 100111–100112.

(14) Bai, Z.; Cao, Y.; Liu, W.; Li, J.; Lundstrom, K.; Davidson, A. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. **2021**.

(15) Schoeman, D.; Fielding, B. C. Coronavirus Envelope Protein: Current Knowledge. *Virology Journal 2019 16:1* **2019**, *16* (1), 1–22.

(16) Zheng, M.; Williams, E. P.; Subbarao Malireddi, R. K.; Karki, R.; Banoth, B.; Burton, A.; Webby, R.; Channappanavar, R.; Jonsson, C. B.; Kanneganti, T. D.

Impaired NLRP3 Inflammasome Activation/Pyroptosis Leads to Robust Inflammatory Cell Death via Caspase-8/RIPK3 during Coronavirus Infection. *Journal of Biological Chemistry* **2020**, *295* (41), 14040–14052.

(17) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 215–220.

(18) Fehr, A. R.; Perlman, S. *Coronaviruses: An Overview of Their Replication and Pathogenesis*; Humana Press, New York, NY, 2015; Vol. 1282.

(19) Robson, F.; Khan, K. S.; Le, T. K.; Paris, C.; Demirbag, S.; Barfuss, P.; Rocchi, P.; Ng, W. L. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol Cell* **2020**, *79* (5), 710–727.

(20) Miao, G.; Zhao, H.; Li, Y.; Ji, M.; Chen, Y.; Shi, Y.; Bi, Y.; Wang, P.; Zhang, H. ORF3a of the COVID-19 Virus SARS-CoV-2 Blocks HOPS Complex-Mediated Assembly of the SNARE Complex Required for Autolysosome Formation. *Dev Cell* **2021**, *56* (4), 427-442.e5.

(21) Konno, Y.; Kimura, I.; Uriu, K.; Fukushi, M.; Irie, T.; Koyanagi, Y.; Sauter, D.; Gifford, R. J.; Nakagawa, S.; Sato, K. SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Reports* **2020**, *32* (12).

(22) Xia, H.; Cao, Z.; Xie, X.; Zhang, X.; Chen, J. Y. C.; Wang, H.; Menachery, V. D.; Rajsbaum, R.; Shi, P. Y. Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* **2020**, *33* (1), 108234.

(23) Miorin, L.; Kehrer, T.; Sanchez-Aparicio, M. T.; Zhang, K.; Cohen, P.; Patel, R. S.; Cupic, A.; Makio, T.; Mei, M.; Moreno, E.; Danziger, O.; White, K. M.; Rathnasinghe, R.; Uccellini, M.; Gao, S.; Aydillo, T.; Mena, I.; Yin, X.; Martin-Sancho, L.; Krogan, N. J.; Chanda, S. K.; Schotsaert, M.; Wozniak, R. W.; Ren, Y.; Rosenberg, B. R.; Fontoura, B. M. A.; García-Sastre, A. SARS-CoV-2 Orf6 Hijacks Nup98 to Block STAT Nuclear Import and Antagonize Interferon Signaling. *Proc Natl Acad Sci U S A* **2020**, *117* (45), 28344–28354.

(24) Martin-Sancho, L.; Lewinski, M. K.; Pache, L.; Stoneham, C. A.; Yin, X.; Becker, M. E.; Pratt, D.; Churas, C.; Rosenthal, S. B.; Liu, S.; Weston, S.; de Jesus, P. D.; O'Neill, A. M.; Gounder, A. P.; Nguyen, C.; Pu, Y.; Curry, H. M.; Oom, A. L.; Miorin, L.; Rodriguez-Frandsen, A.; Zheng, F.; Wu, C.; Xiong, Y.; Urbanowski, M.; Shaw, M. L.; Chang, M. W.; Benner, C.; Hope, T. J.; Frieman,

M. B.; García-Sastre, A.; Ideker, T.; Hultquist, J. F.; Guatelli, J.; Chanda, S. K. Functional Landscape of SARS-CoV-2 Cellular Restriction. *Mol Cell* **2021**, *81* (12), 2656-2668.e8.

(25) Schaecher, S. R.; Mackenzie, J. M.; Pekosz, A. The ORF7b Protein of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Is Expressed in Virus-Infected Cells and Incorporated into SARS-CoV Particles. *Journal of Virology* **2007**, *81* (2), 718–731.

(26) Lin, X.; Fu, B.; Yin, S.; Li, Z.; Liu, H.; Zhang, H.; Xing, N.; Wang, Y.; Xue, W.; Xiong, Y.; Zhang, S.; Zhao, Q.; Xu, S.; Zhang, J.; Wang, P.; Nian, W.; Wang, X.; Wu, H. ORF8 Contributes to Cytokine Storm during SARS-CoV-2 Infection by Activating IL-17 Pathway. *iScience* **2021**, *24* (4).

(27) 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. SARS-CoV-2 ORF10 Suppresses the Antiviral Innate Immune Response by Degrading MAVS through Mitophagy. *Cell Mol Immunol* **2022**, *19* (1), 67–78.

(28) Shereen, M. A.; Khan, S.; Kazmi, A.; Bashir, N.; Siddique, R. COVID-19 Infection: Origin, Transmission, and Characteristics of Human Coronaviruses. *Journal of Advanced Research* **2020**, *24*, 91–98.

(29) Vabret, N.; Britton, G. J.; Gruber, C.; Hegde, S.; Kim, J.; Kuksin, M.; Levantovsky, R.; Malle, L.; Moreira, A.; Park, M. D.; Pia, L.; Risson, E.; Saffern, M.; Salomé, B.; Esai Selvan, M.; Spindler, M. P.; Tan, J.; van der Heide, V.; Gregory, J. K.; Alexandropoulos, K.; Bhardwaj, N.; Brown, B. D.; Greenbaum, B.; Gümüş, Z. H.; Homann, D.; Horowitz, A.; Kamphorst, A. O.; Curotto de Lafaille, M. A.; Mehandru, S.; Merad, M.; Samstein, R. M.; Agrawal, M.; Aleynick, M.; Belabed, M.; Brown, M.; Casanova-Acebes, M.; Catalan, J.; Centa, M.; Charap, A.; Chan, A.; Chen, S. T.; Chung, J.; Bozkus, C. C.; Cody, E.; Cossarini, F.; Dalla, E.; Fernandez, N.; Grout, J.; Ruan, D. F.; Hamon, P.; Humblin, E.; Jha, D.; Kodysh, J.; Leader, A.; Lin, M.; Lindblad, K.; Lozano-Ojalvo, D.; Lubitz, G.; Magen, A.; Mahmood, Z.; Martinez-Delgado, G.; Mateus-Tique, J.; Meritt, E.; Moon, C.; Noel, J.; O'Donnell, T.; Ota, M.; Plitt, T.; Pothula, V.; Redes, J.; Reyes Torres, I.; Roberto, M.; Sanchez-Paulete, A. R.; Shang, J.; Schanoski, A. S.; Suprun, M.; Tran, M.; Vaninov, N.; Wilk, C. M.; Aguirre-Ghiso, J.; Bogunovic, D.; Cho, J.; Faith, J.; Grasset, E.; Heeger, P.; Kenigsberg, E.; Krammer, F.; Laserson, U. Immunology of COVID-19: Current State of the Science. *Immunity* **2020**, *52* (6), 910–941.

(30) Tay, M. Z.; Poh, C. M.; Rénia, L.; MacAry, P. A.; Ng, L. F. P. The Trinity of COVID-19: Immunity, Inflammation and Intervention. *Nature Reviews Immunology* **2020**, *20* (6), 363–374.

(31) Coperchini, F.; Chiovato, L.; Croce, L.; Magri, F.; Rotondi, M. The Cytokine Storm in COVID-19: An Overview of the Involvement of the Chemokine/Chemokine-Receptor System. *Cytokine and Growth Factor Reviews* **2020**, *53*, 25–32.

(32) Jose, R. J.; Manuel, A. COVID-19 Cytokine Storm: The Interplay between Inflammation and Coagulation. *Lancet Respir Med* **2020**, *8* (6), e46–e47.

(33) *SARS-CoV-2 interactome ~ ViralZone*. https://viralzone.expasy.org/9077 (accessed 2022-03-20).

(34) Alm, E.; Broberg, E. K.; Connor, T.; Hodcroft, E. B.; Komissarov, A. B.; Maurer-Stroh, S.; Melidou, A.; Neher, R. A.; O'Toole, Á.; Pereyaslov, D.; Beerenwinkel, N.; Posada-Céspedes, S.; Jablonski, K. P.; Ferreira, P. F.; Topolsky, I.; Avšič-Županc, T.; Korva, M.; Poljak, M.; Zakotnik, S.; Zorec, T. M.; Bragstad, K.; Hungnes, O.; Stene-Johansen, K.; Reusken, C.; Meijer, A.; Vennema, H.; Ruiz-Roldán, L.; Bracho, M. A.; García-González, N.; Chiner-Oms, Á.; Cancino-Muñoz, I.; Comas, I.; Goig, G. A.; Torres-Puente, M.; López, M. G.; Martínez-Priego, L.; D'Auria, G.; Ruíz-Hueso, P.; Ferrús-Abad, L.; de Marco, G.; Galan-Vendrell, I.; Carbó-Ramirez, S.; Ruiz-Rodriguez, P.; Coscollá, M.; Polackova, K.; Kramna, L.; Cinek, O.; Richter, J.; Krashias, G.; Tryfonos, C.; Bashiardes, S.; Koptides, D.; Christodoulou, C.; Bartolini, B.; Gruber, C. E.; di Caro, A.; Castilletti, C.; Stefani, F.; Rimoldi, S. G.; Romeri, F.; Salerno, F.; Polesello, S.; Nagy, A.; Jirincova, H.; Vecerova, J.; Novakova, L.; Cordey, S.; Murtskhvaladze, M.; Kotaria, N.; Schär, T.; Beisel, C.; Vugrek, O.; Rokić, F.; Trgovec-Greif, L.; Jurak, I.; Rukavina, T.; Sučić, N.; Schønning, K.; Karst, S. M.; Kirkegaard, R. H.; Michaelsen, T. Y.; Sørensen, E. A.; Knutson, S.; Brandt, J.; Le-Quy, V.; Sørensen, T.; Petersen, C.; Pedersen, M. S.; Larsen, S. L.; Skov, M. N.; Rasmussen, M.; Fonager, J.; Fomsgaard, A.; Maksyutov, R. A.; Gavrilova, E. V. E.; Pyankov, O. V.; Bodnev, S. A.; Tregubchak, T. V.; Shvalov, A. N.; Antonets, D. V.; Resende, P. C.; Goya, S.; Perrin, A.; Lee, R. T.; Yadahalli, S.; Han, A. X.; Russell, C. A.; Schmutz, S.; Zaheri, M.; Kufner, V.; Huber, M.; Trkola, A.; Antwerpen, M.; Walter, M. C.; van der Werf, S.; Gambaro, F.; Behillil, S.; Enouf, V.; Donati, F.; Ustinova, M.; Rovite, V.; Klovins, J.; Savicka, O.; Wienecke-Baldacchino, A. K.; Ragimbeau, C.; Fournier, G.; Mossong, J.; Aberle, S. W.; Haukland, M.; Enkirch, T.; Advani, A.; Karlberg, M. L.; Lindsjö, O. K.; Broddesson, S.; Sláviková, M.; Ličková, M.; Klempa, B.; Staroňová, E.; Tichá, E.; Szemes, T.; Rusňáková, D.; Stadler, T.; Quer, J.; Anton, A.; Andres, C.; Piñana, M.; Garcia-Cehic, D.; Pumarola, T.; Izopet, J.; Gioula, G.; Exindari, M.; Papa, A.; Chatzidimitriou, D.; Metallidis, S.; Pappa, S.; Macek, M.; Geryk, J.; Brož, P.; Briksí, A.; Hubáček, P.; Dřevínek, P.; Zajac, M.; Kvapil, P.; Holub, M.; Kvapilová, K.; Novotný, A.; Kašný, M.; Klempt, P.; Vapalahti, O.; Smura, T.; Sironen, T.; Selhorst, P.; Anthony, C.; Ariën, K.; Simon-Loriere, E.; Rabalski, L.; Bienkowska-Szewczyk, K.; Borges, V.; Isidro, J.; Gomes, J. P.; Guiomar, R.; Pechirra, P.; Costa, I.; Duarte, S.; Vieira, L.; Pyrc, K.; Zuckerman, N. S.; Turdikulova, S.; Abdullaev, A.; Dalimova, D.; Abdurakhimov, A.;

Tagliabracci, A.; Alessandrini, F.; Melchionda, F.; Onofri, V.; Turchi, C.; Bagnarelli, P.; Menzo, S.; Caucci, S.; di Sante, L.; Popa, A.; Genger, J. W.; Agerer, B.; Lercher, A.; Endler, L.; Smyth, M.; Penz, T.; Schuster, M.; Senekowitsch, M.; Laine, J.; Bock, C.; Bergthaler, A.; Shevtsov, A.; Kalendar, R.; Ramanculov, Y.; Graf, A.; Muenchhoff, M.; Keppler, O. T.; Krebs, S.; Blum, H.; Marcello, A.; Licastro, D.; D'Agaro, P.; Laubscher, F.; Vidanovic, D.; Tesovic, B.; Volkening, J.; Clementi, N.; Mancini, N.; Rupnik, M.; Mahnic, A.; Walker, A.; Houwaart, T.; Wienemann, T.; Vasconcelos, M. K.; Strelow, D.; Jensen, B. E. O.; Senff, T.; Hülse, L.; Adams, O.; Andree, M.; Hauka, S.; Feldt, T.; Keitel, V.; Kindgen-Milles, D.; Timm, J.; Pfeffer, K.; Dilthey, A. T.; Moore, C.; Ozdarendeli, A.; Pavel, S. T. I.; Yetiskin, H.; Aydin, G.; Holyavkin, C.; Uygut, M. A.; Cevik, C.; Shchetinin, A.; Gushchin, V.; Dinler-Doganay, G.; Doganay, L.; Kizilboga-Akgun, T.; Karacan, I.; Pancer, K.; Maes, P.; Martí-Carreras, J.; Wawina-Bokalanga, T.; Vanmechelen, B.; Thürmer, A.; Wedde, M.; Dürrwald, R.; von Kleist, M.; Drechsel, O.; Wolff, T.; Fuchs, S.; Kmiecinski, R.; Michel, J.; Nitsche, A.; Casas, I.; Caballero, M. I.; Zaballos, Á.; Jiménez, P.; Jiménez, M.; Fernández, S. M.; Fernández, S. V.; de la Plaza, I. C.; Fadeev, A.; Ivanova, A.; Sergeeva, M.; Stefanelli, P.; Estee Torok, M.; Hall, G.; da Silva Filipe, A.; Turtle, L.; Afifi, S.; McCluggage, K.; Beer, R.; Ledesma, J.; Maksimovic, J.; Spellman, K.; Hamilton, W. L.; Marchbank, A.; Southgate, J. A.; Underwood, A.; Taylor, B.; Yeats, C.; Abudahab, K.; Gemmell, M. R.; Eccles, R.; Lucaci, A.; Nelson, C. A.; Rainbow, L.; Whitehead, M.; Gregory, R.; Haldenby, S.; Paterson, S.; Hughes, M. A.; Curran, M. D.; Baker, D.; Tucker, R.; Green, L. R.; Feltwell, T.; Halstead, F. D.; Wyles, M.; Jahun, A. S.; Ahmad, S. S. Y.; Georgana, I.; Goodfellow, I.; Yakovleva, A.; Meredith, L. W.; Gavriil, A.; Awan, A. R.; Fisher, C.; Edgeworth, J.; Lynch, J.; Moore, N.; Williams, R.; Kidd, S. P.; Cortes, N.; Brunker, K.; McCrone, J. T.; Quick, J.; Duckworth, N.; Walsh, S.; Sloan, T.; Ludden, C.; George, R. P.; Eltringham, G.; Brown, J. R.; Aranday-Cortes, E.; Shepherd, J. G.; Hughes, J.; Li, K. K.; Williams, T. C.; Johnson, N.; Jesudason, N.; Mair, D.; Thomson, E.; Shah, R.; Parr, Y. A.; Carmichael, S.; Robertson, D. L.; Nomikou, K.; Broos, A.; Niebel, M.; Smollett, K.; Tong, L.; Miah, S.; Wittner, A.; Phillips, N.; Payne, B.; Dewar, R.; Holmes, A.; Bolt, F.; Price, J. R.; Mookerjee, S.; Sethi, D. K.; Potter, W.; Stanley, R.; Prakash, R.; Dervisevic, S.; Graham, J. C.; Nelson, A.; Smith, D.; Young, G. R.; Yew, W. C.; Todd, J. A.; Trebes, A.; Andersson, M.; Bull, M.; Watkins, J.; Birchley, A.; Gatica-Wilcox, B.; Gilbert, L.; Kumžiene-Summerhayes, S.; Rey, S.; Chauhan, A.; Butcher, E.; Bicknell, K.; Elliott, S.; Glaysher, S.; Lackenby, A.; Bibby, D.; Platt, S.; Mohamed, H.; Machin, N. W.; Mbisa, J. L.; Evans, J.; Perry, M.; Pacchiarini, N.; Corden, S.; Adams, A. G.; Gaskin, A.; Coombs, J.; Graham, L. J.; Cottrell, S.; Morgan, M.; Gifford, L.; Kolyva, A.; Rudder, S. J.; Trotter, A. J.; Mather, A. E.; Aydin, A.; Page, A. J.; Kay, G. L.; de Oliveira Martins, L.; Yasir, M.; Alikhan, N. F.; Thomson, N. M.; Gilroy, R.; Kingsley, R. A.; O'Grady, J.; Gutierrez, A. V.; Diaz, M.; Viet, T. le; Tedim, A. P.; Adriaenssens, E. M.; Patrick Mcclure, C.; Moore, C.; Sang, F.; Clark, G.; Howson-Wells, H. C.; Debebe, J.; Ball, J.; Chappell, J.; Khakh, M.; Carlile, M.; Loose, M.; Lister, M. M.; Holmes, N.; Tsoleridis, T.; Fleming, V. M.; Wright, V.; Smith, W.; Gallagher, M. D.; Parker, M.; Partridge, D. G.; Evans, C.; Baker, P.; Essex, S.; Liggett, S.; Keeley, A. J.; Bashton, M.; Rooke, S.; Dervisavic, S.; Meader, E. J.; Lopez, C. E. B.; Angyal, A.; Kristiansen, M.; Tutill, H. J.; Findlay, J.; Mestek-Boukhibar, L.; Forrest, L.; Dyal, P.; Williams, R. J.; Panchbhaya, Y.;

Williams, C. A.; Roy, S.; Pandey, S.; Stockton, J.; Loman, N. J.; Poplawski, R.; Nicholls, S.; Rowe, W. P. M.; Khokhar, F.; Pinckert, M. L.; Hosmillo, M.; Chaudhry, Y.; Caller, L. G.; Davidson, R. K.; Griffith, L.; Rambaut, A.; Jackson, B.; Colquhoun, R.; Hill, V.; Nichols, J.; Asamaphan, P.; Darby, A.; Jackson, K. A.; Iturriza-Gomara, M.; Vamos, E. E.; Green, A.; Aanensen, D.; Bonsall, D.; Buck, D.; Macintyre-Cockett, G.; de Cesare, M.; Pybus, O.; Golubchik, T.; Scarlett, G.; Loveson, K. F.; Robson, S. C.; Beckett, A.; Lindsey, B.; Groves, D. C.; Parsons, P. J.; McHugh, M. P.; Barnes, J. D.; Manso, C. F.; Grammatopoulos, D.; Menger, K. E.; Harrison, E.; Gunson, R.; Peacock, S. J.; Gonzalez, G.; Carr, M.; Mihaela, L.; Popovici, O.; Brytting, M.; Bresner, C.; Fuller, W.; Workman, T.; Mentis, A. F.; Kossyvakis, A.; Karamitros, T.; Pogka, V.; Kalliaropoulos, A.; Horefti, E.; Kontou, A.; Martinez-Gonzalez, B.; Labropoulou, V.; Voulgari-Kokota, A.; Evangelidou, M.; Bizta, P.; Belimezi, M.; Lambrechts, L.; Doymaz, M. Z.; Yazici, M. K.; Cetin, N. S.; Karaaslan, E.; Kallio-Kokko, H.; Virtanen, J.; Suvanto, M.; Nguyen, P. T.; Ellonen, P.; Hannula, S.; Kangas, H.; Sreenu, V. B.; Burián, K.; Terhes, G.; Gombos, K.; Gyenesei, A.; Urbán, P.; Herczeg, R.; Jakab, F.; Kemenesi, G.; Tóth, G. E.; Somogyi, B.; Zana, B.; Zeghbib, S.; Kuczmog, A.; Földes, F.; Lanszki, Z.; Madai, M.; Papp, H.; Nagy, Á.; Pereszlényi, C. I.; Babinszky, G. C.; Dudás, G.; Csoma, E.; Abou Tayoun, A. N.; Alsheikh-Ali, A. A.; Loney, T.; Nowotny, N.; Abdul-Wahab, O.; Gonzalez-Candelas, F.; Andersen, M. H.; Taylor, S. Geographical and Temporal Distribution of SARS-CoV-2 Clades in the WHO European Region, January to June 2020. *Eurosurveillance* **2020**, *25* (32), 2001410.

(35) *Tracking SARS-CoV-2 variants*. https://www.who.int/activities/tracking-SARS-CoV-2-variants/tracking-SARS-CoV-2-variants (accessed 2022-03-20).

(36) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; Angyal, A.; Brown, R. L.; Carrilero, L.; Green, L. R.; Groves, D. C.; Johnson, K. J.; Keeley, A. J.; Lindsey, B. B.; Parsons, P. J.; Raza, M.; Rowland-Jones, S.; Smith, N.; Tucker, R. M.; Wang, D.; Wyles, M. D.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182* (4), 812-827.e19.

(37) Gu, H.; Chen, Q.; Yang, G.; 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.; Luo, D.-Y.; Wang, H.; Yang, X.; Li, Y.; Han, G.; He, Y.; Zhou, X.; Geng, S.; Sheng, X.; Jiang, S.; Sun, S.; Qin, C.-F.; Zhou, Y. *Adaptation of SARS-CoV-2 in BALB/c Mice for Testing Vaccine Efficacy*.

(38) Tada, T.; Dcosta, B. M.; Zhou, H.; Vaill, A.; Kazmierski, W.; Landau, N. R. Decreased Neutralization of SARS-CoV-2 Global Variants by Therapeutic Anti-Spike Protein Monoclonal Antibodies. *bioRxiv* **2021**.

(39) Pereira, F. Evolutionary Dynamics of the SARS-CoV-2 ORF8 Accessory Gene. *Infect Genet Evol* **2020**, *85*.

(40) Jangra, S.; Ye, C.; Rathnasinghe, R.; Stadlbauer, D.; Krammer, F.; Simon, V.; Martinez-Sobrido, L.; Garcia-Sastre, A.; Schotsaert, M. The E484K Mutation in the SARS-CoV-2 Spike Protein Reduces but Does Not Abolish Neutralizing Activity of Human Convalescent and Post-Vaccination Sera. *medRxiv* **2021**.

(41) Maggi, F.; Novazzi, F.; Genoni, A.; Baj, A.; Spezia, P. G.; Focosi, D.; Zago, C.; Colombo, A.; Cassani, G.; Pasciuta, R.; Tamborini, A.; Rossi, A.; Prestia, M.; Capuano, R.; Azzi, L.; Donadini, A.; Catanoso, G.; Grossi, P. A.; Maffioli, L.; Bonelli, G. Imported SARS-CoV-2 Variant P.1 in Traveler Returning from Brazil to Italy. *Emerg Infect Dis* **2021**, *27* (4), 1249–1251.

(42) Cherian, S.; Potdar, V.; Jadhav, S.; Yadav, P.; Gupta, N.; Das, M.; Rakshit, P.; Singh, S.; Abraham, P.; Panda, S.; Patil, S.; Jagtap, P.; Kasabe, B.; Shah, U.; Sanjeev, T.; Divekar, G.; Korabu, K.; Shelkande, S.; Shinde, P.; Zakiuddin, S.; Vipat, V.; Jadhav, S.; Iyengar, K.; Malik, V.; Bhorekar, S.; Kumar, A.; Sahay, R. Convergent Evolution of SARS-CoV-2 Spike Mutations, L452R, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *bioRxiv* **2021**, 2021.04.22.440932.

(43) Li, D.; Wang, D.; Dong, J.; Wang, N.; Huang, H.; Xu, H.; Xia, C. False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases. *Korean J Radiol* **2020**, *21* (4), 505–508.

(44) *Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: interim guidance, 19 March 2020*. https://apps.who.int/iris/handle/10665/331501 (accessed 2022-03-20).

(45) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K. W.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M. L.; Mulders, D. G. J. C.; Haagmans, B. L.; van der Veer, B.; van den Brink, S.; Wijsman, L.; Goderski, G.; Romette, J. L.; Ellis, J.; Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M. P. G.; Drosten, C. Detection of 2019 Novel Coronavirus (2019-NCoV) by Real-Time RT-PCR. *Eurosurveillance* **2020**, *25* (3), 2000045.

(46) Sethuraman, N.; Jeremiah, S. S.; Ryo, A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA* **2020**, *323* (22), 2249–2251.

(47) He, J. L.; Luo, L.; Luo, Z. D.; Lyu, J. X.; Ng, M. Y.; Shen, X. P.; Wen, Z. Diagnostic Performance between CT and Initial Real-Time RT-PCR for Clinically Suspected 2019 Coronavirus Disease (COVID-19) Patients Outside Wuhan, China. *Respiratory Medicine* **2020**, *168*, 105980.

(48) Patel, R.; Babady, E.; Theel, E. S.; Storch, G. A.; Pinsky, B. A.; George, K. S.; Smith, T. C.; Bertuzzi, S. Report from the American Society for Microbiology Covid-19 International Summit, 23 March 2020: Value of Diagnostic Testing for Sars–Cov-2/Covid-19. *mBio* **2020**, *11* (2).

(49) Vashist, S. K. In Vitro Diagnostic Assays for COVID-19: Recent Advances and Emerging Trends. *Diagnostics (Basel)* **2020**, *10* (4).

(50) Tang, Y. W.; Schmitz, J. E.; Persing, D. H.; Stratton, C. W. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *Journal of Clinical Microbiology* **2020**, *58* (6).

(51) Berber, B.; Aydin, C.; Kocabas, F.; Guney-Esken, G.; Yilancioglu, K.; Karadag-Alpaslan, M.; Caliseki, M.; Yuce, M.; Demir, S.; Tastan, C. Gene Editing and RNAi Approaches for COVID-19 Diagnostics and Therapeutics. *Gene Therapy 2020 28:6* **2020**, *28* (6), 290–305.

(52) Broughton, J. P.; Deng, X.; Yu, G.; Fasching, C. L.; Servellita, V.; Singh, J.; Miao, X.; Streithorst, J. A.; Granados, A.; Sotomayor-Gonzalez, A.; Zorn, K.; Gopez, A.; Hsu, E.; Gu, W.; Miller, S.; Pan, C. Y.; Guevara, H.; Wadford, D. A.; Chen, J. S.; Chiu, C. Y. CRISPR–Cas12-Based Detection of SARS-CoV-2. *Nature Biotechnology 2020 38:7* **2020**, *38* (7), 870–874.

(53) Bhattacharyya, R. P.; Thakku, S. G.; Hung, D. T. Harnessing CRISPR Effectors for Infectious Disease Diagnostics. *ACS Infectious Diseases* **2018**, *4* (9), 1278–1282.

(54) *COVID-19 vaccines: development, evaluation, approval and monitoring | European Medicines Agency*. https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-development-evaluation-approval-monitoring (accessed 2022-03-20).

(55) WHO. *Coronavirus Disease (COVID-19) 11 October 2020.* Strategy to Achieve Global Covid-19 Vaccination by Mid-2022.

(56) WHO. *Coronavirus Disease (COVID-19). Vaccines – COVID19 Vaccine Tracker*. https://covid19.trackvaccines.org/vaccines/approved/ (accessed 2022-03-20).

(57) WHO. *Coronavirus Disease (COVID-19). COVID-19 vaccine tracker and landscape*. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines (accessed 2022-03-20).

(58) Tanriover, M. D.; Doğanay, H. L.; Akova, M.; Güner, H. R.; Azap, A.; Akhan, S.; Köse, Ş.; Erdinç, F. Ş.; Akalın, E. H.; Tabak, Ö. F.; Pullukçu, H.; Batum, Ö.; Şimşek Yavuz, S.; Turhan, Ö.; Yıldırmak, M. T.; Köksal, İ.; Taşova, Y.; Korten, V.; Yılmaz, G.; Çelen, M. K.; Altın, S.; Çelik, İ.; Bayındır, Y.; Karaoğlan, İ.; Yılmaz, A.; Özkul, A.; Gür, H.; Unal, S.; Kayaaslan, B.; Hasanoğlu, İ.; Dalkıran, A.; Aydos, Ö.; Çınar, G.; Akdemir-Kalkan, İ.; İnkaya, A. Ç.; Aydin, M.; Çakir, H.; Yıldız, J.; Kocabıyık, Ö.; Arslan, S.; Nallı, B.; Demir, Ö.; Singil, S.; Ataman-Hatipoğlu, Ç.; Tuncer-Ertem, G.; Kınıklı, S.; Önal, U.; Mete, B.; Dalgan, G.; Taşbakan, M.; Yamazhan, T.; Kömürcüoğlu, B.; Yalnız, E.; Benli, A.; Keskin-Sarıtaş, Ç.; Ertosun, M. G.; Özkan, Ö.; Emre, S.; Arıca, S.; Kuşçu, F.; Candevir, A.; Ertürk-Şengel, B.; Ayvaz, F.; Aksoy, F.; Mermutluoğlu, Ç.; Demir, Y.; Günlüoğlu, G.; Tural-Önür, S.; Kılıç-Toker, A.; Eren, E.; Otlu, B.; Mete, A. Ö.; Koçak, K.; Ateş, H.; Koca-Kalkan, İ.; Aksu, K. Efficacy and Safety of an Inactivated Whole-Virion SARS-CoV-2 Vaccine (CoronaVac): Interim Results of a Double-Blind, Randomised, Placebo-Controlled, Phase 3 Trial in Turkey. *Lancet* **2021**, *398* (10296), 213–222.

(59) Jara, A.; Undurraga, E. A.; González, C.; Paredes, F.; Fontecilla, T.; Jara, G.; Pizarro, A.; Acevedo, J.; Leo, K.; Leon, F.; Sans, C.; Leighton, P.; Suárez, P.; García-Escorza, H.; Araos, R. Effectiveness of an Inactivated SARS-CoV-2 Vaccine in Chile. *N Engl J Med* **2021**, *385* (10), 875–884.

(60) Ranzani, O. T.; Hitchings, M. D. T.; Dorion, M.; D'Agostini, T. L.; de Paula, R. C.; de Paula, O. F. P.; Villela, E. F. D. M.; Torres, M. S. S.; de Oliveira, S. B.; Schulz, W.; Almiron, M.; Said, R.; de Oliveira, R. Di.; Vieira Da Silva, P.; de Araújo, W. N.; Gorinchteyn, J. C.; Andrews, J. R.; Cummings, D. A. T.; Ko, A. I.; Croda, J. Effectiveness of the CoronaVac Vaccine in Older Adults during a Gamma Variant Associated Epidemic of Covid-19 in Brazil: Test Negative Case-Control Study. *The BMJ* **2021**, *374*, 2015.

(61) Li, X. N.; Huang, Y.; Wang, W.; Jing, Q. L.; Zhang, C. H.; Qin, P. Z.; Guan, W. J.; Gan, L.; Li, Y. L.; Liu, W. H.; Dong, H.; Miao, Y. T.; Fan, S. J.; Zhang, Z. bin; Zhang, D. M.; Zhong, N. S. Effectiveness of Inactivated SARS-CoV-2

Vaccines against the Delta Variant Infection in Guangzhou: A Test-Negative Case–Control Real-World Study. *Emerging Microbes and Infections* **2021**, *10* (1), 1751–1759.

(62) Ella, R.; Reddy, S.; Blackwelder, W.; Potdar, V.; Yadav, P.; Sarangi, V.; Aileni, V. K.; Kanungo, S.; Rai, S.; Reddy, P.; Verma, S.; Singh, C.; Redkar, S.; Mohapatra, S.; Pandey, A.; Ranganadin, P.; Gumashta, R.; Multani, M.; Mohammad, S.; Bhatt, P.; Kumari, L.; Sapkal, G.; Gupta, N.; Abraham, P.; Panda, S.; Prasad, S.; Bhargava, B.; Ella, K.; Vadrevu, K. M.; Aggarwal, P.; Aglawe, V.; Ali, A.; Anand, N.; Awad, N.; Bafna, V.; Balasubramaniyam, G.; Bandkar, A.; Basha, P.; Bharge, V.; Bhate, A.; Bhate, S.; Bhavani, V.; Bhosale, R.; Chalapathy, D. v.; Chaubal, C.; Chaudhary, D.; Chavan, A.; Desai, P.; Dhodi, D.; Dutta, S.; Garg, R.; Garg, K.; George, M.; Goyal, P.; Guleria, R.; Gupta, S.; Jain, M.; Jain, M. K.; Jindal, S.; Kalra, M.; Kant, S.; Khosla, P.; Kulkarni, P.; Kumar, P.; Kumar, Y.; Majumdar, A.; Meshram, P.; Mishra, V.; Mohanty, S.; Nair, J.; Pandey, S.; Panigrahi, S. K.; Patil, B.; Patil, V.; Rahate, P.; Raj, V.; Ramanand, S.; Rami, K.; Ramraj, B.; Rane, S.; Rao, E. v.; Rao, N.; Raphael, R.; Reddy, G.; Redkar, V.; Redkar, S.; Sachdeva, A.; Saha, J.; Sahoo, J.; Sampath, P.; Savith, A.; Shah, M.; Shanmugam, L.; Sharma, R.; Sharma, P.; Sharma, D.; Singh, A.; Singh, J.; Singh, P.; Sivaprakasam, S.; Subramaniam, S.; Sudheer, D.; Tandon, S.; Tariq, M.; Tripathi, V.; Vable, M.; Verma, R.; Waghmare, S. Efficacy, Safety, and Lot-to-Lot Immunogenicity of an Inactivated SARS-CoV-2 Vaccine (BBV152): Interim Results of a Randomised, Double-Blind, Controlled, Phase 3 Trial. *Lancet* **2021**, *398* (10317), 2173–2184.

(63) al Kaabi, N.; Zhang, Y.; Xia, S.; Yang, Y.; al Qahtani, M. M.; Abdulrazzaq, N.; al Nusair, M.; Hassany, M.; Jawad, J. S.; Abdalla, J.; Hussein, S. E.; al Mazrouei, S. K.; al Karam, M.; Li, X.; Yang, X.; Wang, W.; Lai, B.; Chen, W.; Huang, S.; Wang, Q.; Yang, T.; Liu, Y.; Ma, R.; Hussain, Z. M.; Khan, T.; Saifuddin Fasihuddin, M.; You, W.; Xie, Z.; Zhao, Y.; Jiang, Z.; Zhao, G.; Zhang, Y.; Mahmoud, S.; Eltantawy, I.; Xiao, P.; Koshy, A.; Zaher, W. A.; Wang, H.; Duan, K.; Pan, A.; Yang, X. Effect of 2 Inactivated SARS-CoV-2 Vaccines on Symptomatic COVID-19 Infection in Adults: A Randomized Clinical Trial. *JAMA* **2021**, *326* (1), 35–45.

(64) Farooqi, T.; Malik, J. A.; Mulla, A. H.; al Hagbani, T.; Almansour, K.; Ubaid, M. A.; Alghamdi, S.; Anwar, S. An Overview of SARS-COV-2 Epidemiology, Mutant Variants, Vaccines, and Management Strategies. *J Infect Public Health* **2021**, *14* (10), 1299–1312.

(65) Zhang, J.; Zeng, H.; Gu, J.; Li, H.; Zheng, L.; Zou, Q. Progress and Prospects on Vaccine Development against SARS-CoV-2. *Vaccines 2020, Vol. 8, Page 153* **2020**, *8* (2), 153.

(66) Chemaitelly, H.; Yassine, H. M.; Benslimane, F. M.; al Khatib, H. A.; Tang, P.; Hasan, M. R.; Malek, J. A.; Coyle, P.; Ayoub, H. H.; al Kanaani, Z.; al Kuwari, E.; Jeremijenko, A.; Kaleeckal, A. H.; Latif, A. N.; Shaik, R. M.; Abdul Rahim, H. F.; Nasrallah, G. K.; al Kuwari, M. G.; al Romaihi, H. E.; Al-Thani, M. H.; al Khal, A.; Butt, A. A.; Bertollini, R.; Abu-Raddad, L. J. MRNA-1273 COVID-19 Vaccine Effectiveness against the B.1.1.7 and B.1.351 Variants and Severe COVID-19 Disease in Qatar. *Nat Med* **2021**, *27* (9), 1614–1621.

(67) Tang, P.; Hasan, M. R.; Chemaitelly, H.; Yassine, H. M.; Benslimane, F. M.; al Khatib, H. A.; AlMukdad, S.; Coyle, P.; Ayoub, H. H.; al Kanaani, Z.; al Kuwari, E.; Jeremijenko, A.; Kaleeckal, A. H.; Latif, A. N.; Shaik, R. M.; Abdul Rahim, H. F.; Nasrallah, G. K.; al Kuwari, M. G.; al Romaihi, H. E.; Butt, A. A.; Al-Thani, M. H.; al Khal, A.; Bertollini, R.; Abu-Raddad, L. J. BNT162b2 and MRNA-1273 COVID-19 Vaccine Effectiveness against the SARS-CoV-2 Delta Variant in Qatar. *Nature Medicine 2021 27:12* **2021**, *27* (12), 2136–2143.

(68) Tenforde, M. W.; Patel, M. M.; Ginde, A. A.; Douin, D. J.; Talbot, H. K.; Casey, J. D.; Mohr, N. M.; Zepeski, A.; Gaglani, M.; McNeal, T.; Ghamande, S.; Shapiro, N. I.; Gibbs, K. W.; Files, D. C.; Hager, D. N.; Shehu, A.; Prekker, M. E.; Erickson, H. L.; Exline, M. C.; Gong, M. N.; Mohamed, A.; Henning, D. J.; Steingrub, J. S.; Peltan, I. D.; Brown, S. M.; Martin, E. T.; Monto, A. S.; Khan, A.; Hough, C. T.; Busse, L.; Lohuis, C. C. ten; Duggal, A.; Wilson, J. G.; Gordon, A. J.; Qadir, N.; Chang, S. Y.; Mallow, C.; Gershengorn, H. B.; Babcock, H. M.; Kwon, J. H.; Halasa, N.; Chappell, J. D.; Lauring, A. S.; Grijalva, C. G.; Rice, T. W.; Jones, I. D.; Stubblefield, W. B.; Baughman, A.; Womack, K. N.; Lindsell, C. J.; Hart, K. W.; Zhu, Y.; Olson, S. M.; Stephenson, M.; Schrag, S. J.; Kobayashi, M.; Verani, J. R.; Self, W. H.; Network, F. the I. and O. V. in the A. I. (IVY). Effectiveness of SARS-CoV-2 MRNA Vaccines for Preventing Covid-19 Hospitalizations in the United States. *medRxiv* **2021**, 2021.07.08.21259776.

(69) Hall, V. J.; Foulkes, S.; Saei, A.; Andrews, N.; Oguti, B.; Charlett, A.; Wellington, E.; Stowe, J.; Gillson, N.; Atti, A.; Islam, J.; Karagiannis, I.; Munro, K.; Khawam, J.; Chand, M. A.; Brown, C. S.; Ramsay, M.; Lopez-Bernal, J.; Hopkins, S.; Andrews, N.; Atti, A.; Aziz, H.; Brooks, T.; Brown, C. S.; Camero, D.; Carr, C.; Chand, M. A.; Charlett, A.; Crawford, H.; Cole, M.; Conneely, J.; D'Arcangelo, S.; Ellis, J.; Evans, S.; Foulkes, S.; Gillson, N.; Gopal, R.; Hall, L.; Hall, V. J.; Harrington, P.; Hopkins, S.; Hewson, J.; Hoschler, K.; Ironmonger, D.; Islam, J.; Kall, M.; Karagiannis, I.; Kay, O.; Khawam, J.; King, E.; Kirwan, P.; Kyffin, R.; Lackenby, A.; Lattimore, M.; Linley, E.; Lopez-Bernal, J.; Mabey, L.; McGregor, R.; Miah, S.; Monk, E. J. M.; Munro, K.; Naheed, Z.; Nissr, A.; O'Connell, A. M.; Oguti, B.; Okafor, H.; Organ, S.; Osbourne, J.; Otter, A.; Patel, M.; Platt, S.; Pople, D.; Potts, K.; Ramsay, M.; Robotham, J.; Rokadiya, S.; Rowe, C.; Saei, A.; Sebbage, G.; Semper, A.; Shrotri, M.; Simmons, R.; Soriano, A.; Staves, P.; Taylor, S.; Taylor, A.; Tengbe, A.; Tonge, S.; Vusirikala, A.; Wallace, S.; Wellington, E.; Zambon, M.; Corrigan, D.; Sartaj, M.; Cromey, L.; Campbell, S.; Braithwaite, K.; Price, L.; Haahr, L.; Stewart, S.; Lacey, E. D.;

Partridge, L.; Stevens, G.; Ellis, Y.; Hodgson, H.; Norman, C.; Larru, B.; Mcwilliam, S.; Roynon, A.; Northfield, J.; Winchester, S.; Cieciwa, P.; Pai, A.; Bakker, P.; Loughrey, C.; Watt, A.; Adair, F.; Hawkins, A.; Grant, A.; Temple-Purcell, R.; Howard, J.; Slawson, N.; Subudhi, C.; Davies, S.; Bexley, A.; Penn, R.; Wong, N.; Boyd, G.; Rajgopal, A.; Arenas-Pinto, A.; Matthews, R.; Whileman, A.; Laugharne, R.; Ledger, J.; Barnes, T.; Jones, C.; Osuji, N.; Chitalia, N.; Bailey, T.; Akhtar, S.; Harrison, G.; Horne, S.; Walker, N.; Agwuh, K.; Maxwell, V.; Graves, J.; Williams, S.; O'Kelly, A.; Ridley, P.; Cowley, A.; Johnstone, H.; Swift, P.; Democratis, J.; Meda, M.; Brake, S.; Gunn, J.; Selassi, A.; Hams, S.; Irvine, V.; Chandrasekaran, B.; Forsyth, C.; Radmore, J.; Thomas, C.; Brown, K.; Roberts, S.; Burns, P.; Gajee, K.; Lewis, T.; Byrne, T. M.; Sanderson, F.; Knight, S.; Macnaughton, E.; Burton, B. J. L.; Smith, H.; Chaudhuri, R.; Aeron-Thomas, J.; Hollinshead, K.; Shorten, R. J.; Swan, A.; Favager, C.; Murira, J.; Baillon, S.; Hamer, S.; Shah, A.; Russell, J.; Brennan, D.; Dave, A.; Chawla, A.; Westwell, F.; Adeboyeku, D.; Papineni, P.; Pegg, C.; Williams, M.; Ahmad, S.; Horsley, A.; Gabriel, C.; Pagget, K.; Maloney, G.; Ashcroft, J.; del Rosario, I.; Crosby-Nwaobi, R.; Flanagan, D.; Dhasmana, D.; Fowler, S.; Cameron, E.; Prentice, L.; Sinclair, C.; Bateman, V.; McLelland-Brooks, K.; Ho, A.; Murphy, M.; Cochrane, A.; Gibson, A.; Black, K.; Tempeton, K.; Donaldson, S.; Coke, L.; Elumogo, N.; Elliott, J.; Padgett, D.; Cross, A.; Mirfenderesky, M.; Joyce, S.; Sinanovic, I.; Howard, M.; Cowling, P.; Brazil, M.; Hanna, E.; Abdelrazik, A.; Brand, S.; Sheridan, E. A.; Wadams, B.; Lloyd, A.; Mouland, J.; Giles, J.; Pottinger, G.; Coles, H.; Joseph, M.; Lee, M.; Orr, S.; Chenoweth, H.; Browne, D.; Auckland, C.; Lear, R.; Mahungu, T.; Rodger, A.; Warren, S.; Brooking, D.; Pai, S.; Druyeh, R.; Smith, E.; Stone, S.; Meisner, S.; Delgado, D.; Underhill, E.; Keen, L.; Aga, M.; Domingos, P.; Gormley, S.; Kerrison, C.; Birch, S.; DeSilva, T.; Allsop, L.; Ambalkar, S.; Beekes, M.; Jose, S.; Tomlinson, J.; Painter, S.; Price, C.; Pepperell, J.; James, K.; Trinick, T.; Moore, L.; Day, J.; Boulos, A.; Knox, I.; Defever, E.; McCracken, D.; Gray, K.; Houston, A.; Planche, T.; Pritchard Jones, R.; Wycherley, D.; Bennett, S.; Marrs, J.; Nimako, K.; Stewart, B.; Bain, S. C.; Kalakonda, N.; Khanduri, S.; Ashby, A.; Holden, M.; Mahabir, N.; Harwood, J.; Payne, B.; Court, K.; White, N.; Longfellow, R.; Hughes, L. E.; Green, M. E.; Halkes, M.; Mercer, P.; Roebuck, A.; Wilson-Davies, E.; Gallego, L.; Lazarus, R.; Aldridge, N.; Berry, L.; Game, F.; Reynolds, T.; Holmes, C.; Wiselka, M.; Higham, A.; Booth, M.; Duff, C.; Alderton, J.; Hilton, D.; Powell, J.; Jackson, A.; Plant, A. J.; Ahmed, N.; Chin, T.; Qazzafi, M. Z.; Moody, A. M.; Tilley, R. E.; Donaghy, T.; O'Kane, M.; Shipman, K.; Sierra, R.; Parmar, C.; Mills, G.; Harvey, D.; Huang, Y. W. J.; Birch, J.; Robinson, L.; Board, S.; Broadley, A.; Laven, C.; Todd, N.; Eyre, D. W.; Jeffery, K.; Dunachie, S.; Duncan, C.; Klenerman, P.; Turtle, L.; Baxendale, H.; Heeney, J. L. COVID-19 Vaccine Coverage in Health-Care Workers in England and Effectiveness of BNT162b2 MRNA Vaccine against Infection (SIREN): A Prospective, Multicentre, Cohort Study. *Lancet* **2021**, *397* (10286), 1725–1735.

(70) Nasreen, S.; Chung, H.; He, S.; Brown, K. A.; Gubbay, J. B.; Buchan, S. A.; Fell, D. B.; Austin, P. C.; Schwartz, K. L.; Sundaram, M. E.; Calzavara, A.; Chen, B.; Tadrous, M.; Wilson, K.; Wilson, S. E.; Kwong, J. C.; Investigators, on behalf of the C. I. R. N. (CIRN) P. C. N. (PCN). Effectiveness of MRNA and ChAdOx1

COVID-19 Vaccines against Symptomatic SARS-CoV-2 Infection and Severe Outcomes with Variants of Concern in Ontario. *medRxiv* **2021**, 2021.06.28.21259420.

(71) Amanat, F.; Krammer, F. SARS-CoV-2 Vaccines: Status Report. *Immunity* **2020**, *52* (4), 583–589.

(72) Yoo, J. H. What We Do Know and Do Not Yet Know about COVID-19 Vaccines as of the Beginning of the Year 2021. *J Korean Med Sci* **2021**, *36* (6).

(73) Lopez Bernal, J.; Andrews, N.; Gower, C.; Gallagher, E.; Simmons, R.; Thelwall, S.; Stowe, J.; Tessier, E.; Groves, N.; Dabrera, G.; Myers, R.; Campbell, C. N. J.; Amirthalingam, G.; Edmunds, M.; Zambon, M.; Brown, K. E.; Hopkins, S.; Chand, M.; Ramsay, M. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med* **2021**, *385* (7), 585–594.

(74) Karch, C. P.; Burkhard, P. Vaccine Technologies: From Whole Organisms to Rationally Designed Protein Assemblies. *Biochem Pharmacol* **2016**, *120*, 1–14.

(75) Mohsen, M. O.; Zha, L.; Cabral-Miranda, G.; Bachmann, M. F. Major Findings and Recent Advances in Virus-like Particle (VLP)-Based Vaccines. *Semin Immunol* **2017**, *34*, 123–132.

(76) Mahase, E. Covid-19: Novavax Vaccine Efficacy Is 86% against UK Variant and 60% against South African Variant. *BMJ* **2021**, *372*.

(77) Escobar, L. E.; Molina-Cruz, A.; Barillas-Mury, C. BCG Vaccine Protection from Severe Coronavirus Disease 2019 (COVID-19). *Proc Natl Acad Sci U S A* **2020**, *117* (30), 17720–17726.

(78) Siddiqi, H. K.; Mehra, M. R. COVID-19 Illness in Native and Immunosuppressed States: A Clinical-Therapeutic Staging Proposal. *J Heart Lung Transplant* **2020**, *39* (5), 405–407.

(79) Barlow, A.; Landolf, K. M.; Barlow, B.; Yeung, S. Y. A.; Heavner, J. J.; Claassen, C. W.; Heavner, M. S. Review of Emerging Pharmacotherapy for the Treatment of Coronavirus Disease 2019. *Pharmacotherapy* **2020**, *40* (5), 416–437.

(80) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **2020**, *117* (21).

(81) Liao, J.; Way, G.; Madahar, V. Target Virus or Target Ourselves for COVID-19 Drugs Discovery?—Lessons Learned from Anti-Influenza Virus Therapies. *Medicine in Drug Discovery* **2020**, *5*, 100037.

(82) Brown, A. J.; Won, J. J.; Graham, R. L.; Dinnon, K. H.; Sims, A. C.; Feng, J. Y.; Cihlar, T.; Denison, M. R.; Baric, R. S.; Sheahan, T. P. Broad Spectrum Antiviral Remdesivir Inhibits Human Endemic and Zoonotic Deltacoronaviruses with a Highly Divergent RNA Dependent RNA Polymerase. *Antiviral Res* **2019**, *169*.

(83) Supinsky, S. COVID Antiviral Pills: What Scientists Still Want to Know. **2021**.

(84) Roback, J. D.; Guarner, J. Convalescent Plasma to Treat COVID-19: Possibilities and Challenges. *JAMA* **2020**, *323* (16), 1561–1562.

(85) Yin, Y.; Wunderink, R. G. MERS, SARS and Other Coronaviruses as Causes of Pneumonia. *Respirology* **2018**, *23* (2), 130–137.

(86) Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; Xiang, Z.; Mu, Z.; Chen, X.; Chen, J.; Hu, K.; Jin, Q.; Wang, J.; Qian, Z. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nature Communications 2020 11:1* **2020**, *11* (1), 1–12.

(87) *Merck and Ridgeback's Investigational Oral Antiviral Molnupiravir Reduced the Risk of Hospitalization or Death by Approximately 50 Percent Compared to Placebo for Patients with Mild or Moderate COVID-19 in Positive Interim Analysis of Phase 3 Study - Merck.com*. https://www.merck.com/news/merck-and-ridgebacks-investigational-oral-antiviral-molnupiravir-reduced-the-risk-of-hospitalization-or-death-by-approximately-50-percent-compared-to-placebo-for-patients-with-mild-or-moderat/ (accessed 2022-05-19).

(88) Khare, S.; Gurry, C.; Freitas, L.; B Schultz, M.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; TC Lee, R.; Yeo, W.; Core Curation Team, G.; Maurer-Stroh, S. GISAID's Role in Pandemic Response. *China CDC Weekly* **2021**, *3* (49), 1049–1051.

(89) Elbe, S.; Buckland-Merrett, G. Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health. *Global Challenges* **2017**, *1* (1), 33–46.

(90) Katoh, K.; Rozewicki, J.; Yamada, K. D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief Bioinform* **2019**, *20* (4), 1160–1166.

(91) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **2022**, *50* (D1), D20–D26.

(92) Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J. C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S. E.; Sanchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* **2017**, *34* (12), 3299–3302.

(93) Nei, M.; Li, W. H. Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. *Proc Natl Acad Sci U S A* **1979**, *76* (10), 5269.

(94) Nei, M.; Tajima, F. DNA Polymorphism Detectable by Restriction Endonucleases. *Genetics* **1981**, *97* (1), 145–163.

(95) Watterson, G. A. On the Number of Segregating Sites in Genetical Models without Recombination. *Theoretical Population Biology* **1975**, *7* (2), 256–276.

(96) Tajima, F. *Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism*; 1989.

(97) Fu, Y. X.; Li, W. H. Statistical Tests of Neutrality of Mutations. *Genetics* **1993**, *133* (3), 693–709.

(98) Fay, J. C.; Wu, C. I. Hitchhiking under Positive Darwinian Selection. *Genetics* **2000**, *155* (3), 1405–1413.

(99) McDonald, J. H.; Kreitman, M. Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature 1991 351:6328* **1991**, *351* (6328), 652–654.

(100) Rand, D. A.; Kann, L. M. Excess Amino Acid Polymorphism in Mitochondrial Among Genes from Drosophila, Mice, and Humans.

(101) Fay, J. C.; Wyckoff, G. J.; Wu, C. I. Testing the Neutral Theory of Molecular Evolution with Genomic Data from Drosophila. *Nature* **2002**, *415* (6875), 1024–1026.

(102) Zheng, W.; Zhang, C.; Li, Y.; Pearce, R.; Bell, E. W.; Zhang, Y. Folding Non-Homologous Proteins by Coupling Deep-Learning Contact Maps with I-TASSER Assembly Simulations. *Cell Reports Methods* **2021**, *1* (3), 100014.

(103) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, *12* (1), 7–8.

(104) Yang, J.; Zhang, Y. I-TASSER Server: New Development for Protein Structure and Function Predictions Supplementary Material.

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

(106) Uversky, V. N.; Marques-Pereira, C.; Pires, M. N.; Gouveia, R. P.; Pereira, N. N.; Caniceiro, A. B.; Rosário-Ferreira, N.; Moreira, I. S. SARS-CoV-2 Membrane Protein: From Genomic Data to Structural New Insights. **2022**.

(107) Kumar, P.; Kumar, A.; Garg, N.; Giri, R. An Insight into SARS-CoV-2 Membrane Protein Interaction with Spike, Envelope, and Nucleocapsid Proteins. *J Biomol Struct Dyn* **2021**.

(108) Tseng, Y. T.; Chang, C. H.; Wang, S. M.; Huang, K. J.; Wang, C. T. Identifying SARS-CoV Membrane Protein Amino Acid Residues Linked to Virus-like Particle Assembly. *PLoS One* **2013**, *8* (5).

(109) Liu, J.; Sun, Y.; Qi, J.; Chu, F.; Wu, H.; Gao, F.; Li, T.; Yan, J.; Gao, G. F. The Membrane Protein of Severe Acute Respiratory Syndrome Coronavirus Acts as a Dominant Immunogen Revealed by a Clustering Region of Novel Functionally and Structurally Defined Cytotoxic T-Lymphocyte Epitopes. *J Infect Dis* **2010**, *202* (8), 1171–1180.

(110) Zhang, C.; Li, L.; He, J.; Chen, C.; Su, D. Nonstructural Protein 7 and 8 Complexes of SARS-CoV-2. *Protein Science : A Publication of the Protein Society* **2021**, *30* (4), 873.

(111) Sarma, H.; Jamir, E.; Sastry, G. N. Protein-Protein Interaction of RdRp with Its Co-Factor NSP8 and NSP7 to Decipher the Interface Hotspot Residues for Drug Targeting: A Comparison between SARS-CoV-2 and SARS-CoV. *Journal of Molecular Structure* **2022**, *1257*, 132602.

(112) Rogstam, A.; Nyblom, M.; Christensen, S.; Sele, C.; Talibov, V. O.; Lindvall, T.; Rasmussen, A. A.; André, I.; Fisher, Z.; Knecht, W.; Kozielski, F. Crystal Structure of Non-Structural Protein 10 from Severe Acute Respiratory Syndrome Coronavirus-2. *International Journal of Molecular Sciences* **2020**, *21* (19), 1–15.

(113) Ma, Y.; Wu, L.; Shaw, N.; Gao, Y.; Wang, J.; Sun, Y.; Lou, Z.; Yan, L.; Zhang, R.; Rao, Z. Structural Basis and Functional Analysis of the SARS Coronavirus Nsp14-Nsp10 Complex. *Proc Natl Acad Sci U S A* **2015**, *112* (30), 9436–9441.

(114) Rosas-Lemus, M.; Minasov, G.; Shuvalova, L.; Inniss, N. L.; Kiryukhina, O.; Wiersum, G.; Kim, Y.; Jedrzejczak, R.; Maltseva, N. I.; Endres, M.; Jaroszewski, L.; Godzik, A.; Joachimiak, A.; Satchell, K. J. F. The Crystal Structure of Nsp10-Nsp16 Heterodimer from SARS-CoV-2 in Complex with S-Adenosylmethionine. *bioRxiv* **2020**, 2020.04.17.047498.

(115) Chang, L. J.; Chen, T. H. NSP16 2'-O-MTase in Coronavirus Pathogenesis: Possible Prevention and Treatments Strategies. *Viruses* **2021**, *13* (4).

(116) Bouvet, M.; Debarnot, C.; Imbert, I.; Selisko, B.; Snijder, E. J.; Canard, B.; Decroly, E. In Vitro Reconstitution of SARS-Coronavirus MRNA Cap Methylation. *PLoS Pathogens* **2010**, *6* (4), 1–13.

(117) Chen, Y.; Su, C.; Ke, M.; Jin, X.; Xu, L.; Zhang, Z.; Wu, A.; Sun, Y.; Yang, Z.; Tien, P.; Ahola, T.; Liang, Y.; Liu, X.; Guo, D. Biochemical and Structural Insights into the Mechanisms of SARS Coronavirus RNA Ribose 2'-O-Methylation by Nsp16/Nsp10 Protein Complex. *PLoS Pathog* **2011**, *7* (10).

(118) Lin, S.; Chen, H.; Ye, F.; Chen, Z.; Yang, F.; Zheng, Y.; Cao, Y.; Qiao, J.; Yang, S.; Lu, G. Crystal Structure of SARS-CoV-2 Nsp10/Nsp16 2'-O-Methylase and Its Implication on Antiviral Drug Design. *Signal Transduct Target Ther* **2020**, *5* (1).

(119) Rosas-Lemus, M.; Minasov, G.; Shuvalova, L.; Inniss, N. L.; Kiryukhina, O.; Brunzelle, J.; Satchell, K. J. F. High-Resolution Structures of the SARS-CoV-2 2'- O-Methyltransferase Reveal Strategies for Structure-Based Inhibitor Design. *Sci Signal* **2020**, *13* (651).

(120) Menachery, V. D.; Yount, B. L.; Josset, L.; Gralinski, L. E.; Scobey, T.; Agnihothram, S.; Katze, M. G.; Baric, R. S. Attenuation and Restoration of Severe Acute Respiratory Syndrome Coronavirus Mutant Lacking 2'-o-Methyltransferase Activity. *J Virol* **2014**, *88* (8), 4251–4264.

Supplement Table 1. Fu and Li's Tests and The McDonald-Kreitman Tests of Sars-CoV-2 Nsp12, Nsp13, Nsp14, Nsp15 of clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	Nsp12	-6.68**	-6.01**	-3.63	-1.3	0.47	0.534	0.23	1.74	0.19	10	7	16	49
S	Nsp12	-6.45**	-5.70**	-1.04	-0.22	0.43	0.57	0.09	3.09	0.08	10	13	28	50
O	Nsp12	-8.23**	-6.72**	-7.98	-1.1	0.71	0.29	0.5	0.52	0.47	9	14	52	47
Alpha	Nsp12	-3.23 **	-3.33**	-5.26	-2.69	0.67	0.33	0.69	0.28	0.59	10	3	10	50
Beta	Nsp12	-4.13**	-4.16 **	0.84	0.34	1.01	-0.01	1	0	0.99	11	3	14	51
Gamma	Nsp12	-3. 69 **	-3.90**	-7.35	-2.96	0.99	0.01	1	0	0.99	10	3	14	47
Delta	Nsp12	-3.35 **	-3.41**	-0.29	-0.14	0.79	0.21	1	0.1	0.75	11	3	11	51
L	Nsp13	-7.08**	-6.31**	-3.66	-1.22	56.89	-55.89	0***	28.54	0***	32	9	16	1
S	Nsp13	-5.64**	-4.90**	-4.43	-0.73	48.57	-47.57	0***	34.25	0***	30	21	34	1
O	Nsp13	-8.05 **	-6.76**	-2.8	-0.4	38.4	-37.4	0***	30.95	0***	32	30	36	1
Alpha	Nsp13	-5.04**	-4.82**	-5.2	-2.04	38.75	-37.75	0***	19.06	0***	31	8	10	1
Beta	Nsp13	-4.43**	-4.43**	0.88	0.38	29.75	-28.75	0***	14.16	0***	34	8	7	1
Gamma	Nsp13	-2.44 *	-2.77*	-5.13	-2.94	18.27	-17.29	0.01*	7.77	0**	32	7	4	1
Delta	Nsp13	-2.09 #	-2.27 #	-2.5	-1.59	115.5	-114.5	0***	22.76	0***	33	2	7	1
L	Nsp14	-5.50**	-5.18**	-3.79	-1.83	27.13	-26.13	0***	21.79	0***	37	5	11	3
S	Nsp14	-5.58 **	-5.01**	-10.06	-1.84	9.24	-8.24	0***	15.07	0***	33	25	21	3
O	Nsp14	-6.99 **	-6.09**	-4.87	-0.84	11.11	-10.11	0***	19.18	0***	36	27	25	3
Alpha	Nsp14	-5.32 **	-5.13**	-5.19	-1.72	17.33	-16.33	0***	19.28	0***	36	9	13	3
Beta	Nsp14	-2.40 #	-2.89 **	-7.21	-3.11	4.24	-3.24	0.17	2.98	0.08	35	11	4	3
Gamma	Nsp14	-4.67**	-4.61**	0.62	0.25	18.57	-17.57	0***	17.58	0***	39	7	10	3
Delta	Nsp14	-3.26**	-3.43**	-0.8	-0.26	19.73	-18.73	0***	15.5	0***	37	5	8	3
L	Nsp15	-4.13**	-4.00**	0.07	0.07	19.38	-18.38	0**	9.12	0**	31	1	5	8
S	Nsp15	-6.28**	-5.73**	-5.6	-2.35	2.21	-1.21	0.23	1.71	0.19	29	15	8	7
O	Nsp15	-6.85**	-6.16**	0.26	0.12	5.33	-4.33	0.01**	7.92	0**	31	8	11	8
Alpha	Nsp15	-3.55**	-3.65**	-1.76	-1.31	3.75	-2.75	0.18	2.6	0.11	30	4	4	8
Beta	Nsp15	-3.86**	-3.97**	-1.57	-0.94	4.43	-3.43	0.09	3.82	0.05	31	5	5	7
Gamma	Nsp15	-3.10**	-3.20**	0.16	0.17	15.5	-14.5	0.01*	6.98	0.01**	31	1	4	8
Delta	Nsp15	-1.49 ##	-1.61##	0.66	0.57	7.75	-6.75	0.04*	4.98	0.03*	31	2	4	8

Supplement Table 2. Fu and Li's Tests and The McDonald-Kreitman Tests of Sars-CoV-2 Nsp5, Nsp6, Nsp7, Nsp9 of clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	Nsp5	-6.76**	-6.11**	0.12	0.08	48	-47	0***	21.03	0***	36	3	8	2
S	Nsp5	-5.22*	-4.83**	-7.02	-2.71	9.85	-8.85	0**	9.03	0**	32	13	8	2
O	Nsp5	-4.66**	-4.57**	-5.63	-2.15	15.13	-14.13	0***	14.55	0***	33	12	11	2
Alpha	Nsp5	-3.62**	-3.65**	0.17	0.15	18	-17	0.01*	7.17	0.01**	36	3	3	2
Beta	Nsp5	-2.87*	-3.09**	0.64	0.35	21	-20	0***	14.47	0***	36	4	7	3
Gamma	Nsp5	-4.14**	-4.05**	0.14	0.14	27	-26	0.01**	8.51	0**	36	2	3	2
Delta	Nsp5	-2.39*	-2.58*	0.19	0.22	6	-5	0.27	1.45	0.23	36	3	1	2
L	Nsp6	-4.11**	-4.03**	-1.87	-1.54	33	-32	0**	10.55	0**	33	4	4	1
S	Nsp6	-6.14**	-5.25**	-0.5	-0.15	64	-63	0***	33.05	0***	32	10	20	1
O	Nsp6	-6.53**	-5.80**	-7.1	-1.62	-	-	0***	-	-	29	16	22	0
Alpha	Nsp6	-3.83**	-3.94**	-1.57	-0.8	27.43	-26.43	0**	12.33	0***	32	7	6	1
Beta	Nsp6	-2.94**	-3.08**	0.21	0.21	165	-164	0***	19.39	0***	33	1	5	1
Gamma	Nsp6	-2.66*	-2.93**	-1.68	-1.22	41.25	-40.25	0***	13.37	0***	33	4	5	1
Delta	Nsp6	-0.36 ##	0.09 ##	0.93	0.71	85	-84	0***	16.99	0***	34	2	5	1
L	Nsp7	-3.08 **	-2.96**	0.02	0.04	-	-	-	-	-	6	1	0	1
S	Nsp7	-4.13**	-4.08**	0.16	0.11	-	-	-	-	-	6	6	4	0
O	Nsp7	-2.64 *	-2.82**	-1.81	-1.62	-	-	-	-	-	5	3	4	0
Alpha	Nsp7	-1.99 #	-1.99 #	0.02	0.07	-	-	-	-	-	6	0	1	0
Beta	Nsp7	-3.16 **	-3.18 **	0.11	0.15	-	-	-	-	-	6	2	1	0
Gamma	Nsp7	-2.73 *	-2.73*	0.05	0.1	-	-	-	-	-	6	0	2	0
Delta	Nsp7	-2.59*	-2.63 *	0.08	0.14	-	-	-	-	-	6	1	1	0
L	Nsp9	-1.31 ##	-1.67 ##	0.14	0.16	-	-	0.44	-	-	4	3	2	0
S	Nsp9	-2.57 *	-2.64*	-2.72	-1.89	-	-	0.52	-	-	2	6	4	0
O	Nsp9	-2.74*	-3.03 **	-3.66	-2.2	-	-	1	-	-	2	9	3	0
Alpha	Nsp9	-1.96 #	-2.05 #	0.17	0.25	-	-	0.14	-	-	4	1	2	0
Beta	Nsp9	-3.16**	-3.18 **	0.11	0.15	-	-	0.43	-	-	4	2	1	0
Gamma	Nsp9	-1.01 ##	-1.48##	-3.62	-3.69	-	-	0.46	-	-	3	3	2	0
Delta	Nsp9	-0.12 ##	-0.34##	-0.77	-0.87	-	-	1	-	-	3	3	1	0

Supplement Table 3. Fu and Li's Tests and The McDonald-Kreitman Tests of Sars-CoV-2 Nsp1, Nsp2, Nsp3, Nsp4 of clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	Nsp1	-4.53**	-4.36**	0.23	0.15	1.43	-0.43	0.71	0.21	0.64	12	6	5	7
S	Nsp1	-2.33 #	-2.55 *	-3.15	-1.44	3.06	-2.06	0.17	2.47	0.12	10	6	11	6
O	Nsp1	-5.92**	-5.47**	-1.64	-0.66	3.71	-2.71	0.1	3.87	0.05*	12	6	13	7
Alpha	Nsp1	-3.83**	-3.75**	0.1	0.12	0.57	0.43	1	0.2	0.64	12	3	1	7
Beta	Nsp1	-4.52**	-4.49**	0.39	0.23	1.86	-0.86	0.46	0.62	0.43	13	5	5	7
Gamma	Nsp1	-2.66	-2.88	-1.49	-1.09	1.86	-0.86	0.67	0.53	0.47	13	4	4	7
Delta	Nsp1	-3.11**	-3.14**	0.12	0.16	3.43	-2.43	0.54	0.94	0.33	12	1	2	7
L	Nsp2	-7.02**	-6.20**	-1.45	-0.45	13.32	-12.32	0***	28.14	0***	74	10	18	10
S	Nsp2	-7.4**	-6.24**	-9.07	-1.36	9.07	-8.07	0***	32.16	0***	70	27	35	10
O	Nsp2	-7.46**	-6.30**	20.12	-2.23	6.82	-5.82	0***	27.11	0***	65	41	43	10
Alpha	Nsp2	-5.81**	-5.45**	-10.26	-2.38	6.57	-5.57	0***	15.12	0***	71	18	15	9
Beta	Nsp2	-4.27**	-4.30**	-8.81	-2.03	7.68	-6.68	0***	18.15	0***	72	15	16	10
Gamma	Nsp2	-3.42**	-3.59**	-1.37	-0.65	5	-4	0.01*	6.17	0.01*	75	9	6	10
Delta	Nsp2	-2.11 #	-2.37 #	-2.7	-1.08	10.71	-9.71	0***	15.15	0***	75	7	9	9
L	Nsp3	-8.64**	-7.19**	-7.11	-1.15	7.27	-6.27	0***	40.86	0***	187	17	39	59
S	Nsp3	-8.18**	-6.57	-33.38	-2.41	4.6	-3.6	0***	45.53	0***	175	56	81	55
O	Nsp3	-9.46**	-7.33**	-45.01	-2.13	4.19	-3.19	0***	49.59	0***	168	86	118	55
Alpha	Nsp3	-7.17**	-6.45**	-15.73	-1.9	4.21	-3.21	0***	26.54	0***	180	29	42	62
Beta	Nsp3	-4.72**	-4.75**	-18.49	-2.52	4.01	-3.01	0***	20.13	0***	180	24	31	58
Gamma	Nsp3	-5.71**	-5.41**	-11.19	-1.52	5.3	-4.3	0***	30.76	0***	186	22	37	59
Delta	Nsp3	-1.81 ##	-2.16 #	-19.78	-3.92	2.72	-1.72	0.01**	7.21	0.01**	179	19	17	59
L	Nsp4	-6.10**	-5.60**	0.24	0.12	66	-65	0***	32.74	0***	60	5	11	2
S	Nsp4	-6.31**	-5.52**	-2.87	-0.73	40.71	-39.71	0***	38.22	0***	57	14	20	2
O	Nsp4	-4.17**	-4.13**	-14.51	-3.32	61.83	-60.83	0***	37.96	0***	53	18	21	1
Alpha	Nsp4	-3.04**	-3.37**	-7.41	-3.2	12.73	-11.73	0***	9.06	0***	56	11	5	2
Beta	Nsp4	-2.46*	-2.91**	-8.91	-3.46	24.44	-23.44	0***	17.78	0***	55	9	8	2
Gamma	Nsp4	-5.65 **	-5.35**	0.61	0.25	42.87	-41.86	0***	26.6	0***	60	7	10	2
Delta	Nsp4	-1.01 #	-1.01 #	-2.68	-1.02	53.17	-52.17	0***	30.31	0***	58	6	11	2

Supplement Table 4. Neutrality tests summary statistics for Sars-CoV Nsp12, Nsp13, Nsp14, Nsp15 among viral clades/variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	Nsp12	-2.51***	-2.51***	-1.83*	-2.43 **	-1.83 *	-7.38 **	-6.53**
S	Nsp12	-2.52***	-2.52***	-2.30**	-2.34**	-2.30**	-6.29**	-5.62**
O	Nsp12	-2.46**	-2.46**	-1.55 #	-2.41 **	-1.55 #	-8.14 **	-6.75 **
Alpha	Nsp12	-1.99 *	-1.99 *	-1.64 #	-1.76 #	-1.64 #	-4.78**	-4.51**
Beta	Nsp12	-2.216**	-2.22**	-1.08 ##	-2.26 **	-1.08 ##	-3.54 **	-3.65 **
Gamma	Nsp12	-2.49 **	-2.49 **	-1.65 #	-2.40 **	-1.65 #	-5.28 **	-5.10 **
Delta	Nsp12	-1.89 *	-1.89 *	-0.58 ##	-2.09 *	-0.58 ##	-2.81 *	-2.96 *
L	Nsp13	-2.58 ***	-2.53 ***	-1.95 *	-2.40 **	-1.95 *	-7.71 **	-6.77 **
S	Nsp13	-2.17 **	-2.17 **	-1.97 *	-2.06 *	-1.97 *	-5.58 **	-4.90 **
O	Nsp13	-2.73 ***	-2.73 ***	-2.34 **	-2.67 ***	-2.34 **	-7.93 **	-6.75 **
Alpha	Nsp13	-2.30 **	-2.30 **	-2.0402 *	-2.02 *	-2.04 *	-5.65 **	-5.28 **
Beta	Nsp13	-2.14 *	-2.14 *	-1.31 ##	-2.07 *	-1.31 ##	-4.09 **	-4.05 **
Gamma	Nsp13	-2.12 *	-2.12 *	-1.06 ##	-2.05 *	-1.06 ##	-2.93 *	-3.14 **
Delta	Nsp13	-1.55 ##	-1.55 ##	-1.10 ##	-1.42 ##	-1.10 ##	-2.60 *	-2.66 *
L	Nsp14	-2.41 **	-2.41 **	-1.52 ##	-2.32 **	-1.52 ##	-6.52 **	-5.94 **
S	Nsp14	-2.41 **	-2.41 **	-2.12 *	-2.30 **	-2.12 *	-6.64 **	-5.77 **
O	Nsp14	-2.67 ***	-2.67 ***	-1.95*	-2.68 ***	-1.95 *	-7.11 **	-6.22 **
Alpha	Nsp14	-2.56***	-2.56***	-2.13 *	-2.40 **	-2.13 *	-5.76 **	-5.46 **
Beta	Nsp14	-2.42 **	-2.42 **	-0.88	-2.42 **	-0.88 ##	-4.09 **	-4.16 **
Gamma	Nsp14	-2.42 **	-2.42 **	-1.66 #	-2.32 **	-1.66 #	-4.36 **	-4.36 **
Delta	Nsp14	-2.13*	-2.00 *	-1.41 ##	-2.12 *	-1.41 ##	-2.93 *	-3.15 *
L	Nsp15	-1.90*	-1.90*	-1.25 ##	-1.68 #	-1.25 ##	-4.08 **	-3.96 **
S	Nsp15	-2.51***	-2.51***	-1.52 ##	-2.45 **	-1.52 ##	-6.91 **	-6.19 **
O	Nsp15	-2.45 **	-2.45 **	-1.69 #	-2.34 **	-1.69 #	-6.53 **	-5.94 **
Alpha	Nsp15	-2.16 *	-2.16 *	-1.40 ##	-1.99 *	-1.40 ##	-4.20**	-4.15 **
Beta	Nsp15	-2.30 **	-2.30 **	-1.45 ##	-2.18 **	-1.45 ##	-4.25 **	-4.25 **
Gamma	Nsp15	-1.89*	-1.89*	-1.42 ##	-1.57 #	-1.42 ##	-3.01 *	-3.11 **
Delta	Nsp15	-1.10 ##	-1.10 ##	-1.02 ##	-0.82 ##	-1.02 ##	-1.40 ##	-1.53 ##

Supplement Table 5. Neutrality tests summary statistics for Sars-CoV-2 Nsp5, Nsp6, Nsp7, Nsp9 among viral clades/variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	Nsp5	-2.26 **	-2.26 **	-1.52 ##	-2.09 *	-1.52 ##	-6.63**	-6.02**
S	Nsp5	-2.18**	-2.18**	-1.93 *	-1.89 *	-1.93 *	-5.55**	-5.07**
O	Nsp5	-2.53***	-2.53***	-2.28**	-2.19**	-2.28**	-5.90**	-5.48**
Alpha	Nsp5	-1.99*	-1.99*	-1.56 #	-1.64 #	-1.56 #	-3.52**	-3.56**
Beta	Nsp5	-2.06*	-2.06*	-1.34 ##	-2.05*	-1.34 ##	-2.67*	-2.92*
Gamma	Nsp5	-1.95 *	-1.95 *	-1.42##	-1.65 #	-1.42 ##	-4.02 **	-3.95**
Delta	Nsp5	-1.76 #	-1.76 #	-1.58#	-1.10 ##	-1.58 #	-2.30 #	-2.50*
L	Nsp6	-2.03 *	-2.03 *	-1.56 #	-1.66 #	-1.58 #	-4.92**	-4.65**
S	Nsp6	-1.92 *	-1.92 *	-2.08 *	-1.53 ##	-2.08*	-6.08**	-5.25**
O	Nsp6	-2.54 ***	-2.54 ***	-2.26**	-2.39**	-2.26**	-7.51**	-6.51**
Alpha	Nsp6	-2.34 **	-2.34 **	-2.13*	-1.86 *	-2.13*	-4.21**	-4.22**
Beta	Nsp6	-1.71 #	-1.71 #	-0.12 ##	-1.90*	-1.90##	-2.33 #	-2.50*
Gamma	Nsp6	-2.08*	-2.08*	-1.82*	-1.66 #	-1.82*	-2.92 *	-3.12**
Delta	Nsp6	0.98 ##	0.98 ##	0.40##	1.03 ##	0.40 ##	-0.32 ##	0.11##
L	Nsp7	-1.29 ##	-1.29 ##	-0.96 ##	-0.96 ##	-0.96 ##	-3.06 *	-2.95*
S	Nsp7	-2.14 *	-2.14 *	-1.80*	-1.63 #	-1.80*	-4.06**	-4.02**
O	Nsp7	-1.84*	-1.84*	-1.42 ##	-1.52 ##	-1.42 ##	-3.54 **	-3.51**
Alpha	Nsp7	-1.05 ##	-1.05 ##	NA	-1.05 ##	NA	-1.98#	-1.98#
Beta	Nsp7	-1.69#	-1.69#	-1.45 ##	-1.09 ##	-1.45 ##	-3.08*	-3.10*
Gamma	Nsp7	-1.42##	-1.42##	NA	-1.42 ##	NA	-2.69*	-2.69 *
Delta	Nsp7	-1.46##	-1.46##	-1.10 ##	-1.10 ##	-1.10 ##	-2.53*	-2.57*
L	Nsp9	-1.60#	-1.60#	-1.22 ##	-1.29 ##	-1.22 ##	-1.29 ##	-1.66 ##
S	Nsp9	-1.57 #	-1.57 #	-1.00 ##	-1.69 #	-1.00 ##	-3.29 **	-3.19**
O	Nsp9	-2.18 **	-2.18 **	-2.01*	-1.49 ##	-2.01*	-2.67*	-2.98**
Alpha	Nsp9	-1.29 ##	-1.29 ##	-0.38 ##	-1.40 ##	-0.38 ##	-1.92 #	-2.02 #
Beta	Nsp9	-1.69 #	-1.69 #	-1.45 ##	-1.09 ##	1.45 ##	-3.08*	-3.10*
Gamma	Nsp9	-1.78 #	-1.78 #	-1.41 ##	-1.42 ##	-1.41 ##	-1.99 #	-2.26 #
Delta	Nsp9	-0.66 ##	-0.66 ##	-0.35 ##	-0.87 ##	-0.35 ##	-0.1 ##	-0.31 ##

Supplement Table 6. Neutrality tests summary statistics for Sars-CoV-2 Nsp1, Nsp2, Nsp3, Nsp4 among viral clades and variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	Nsp1	-2.11*	-2.11 *	-1.78*	-1.72#	-1.78 *	-4.44 **	-4.29**
S	Nsp1	-1.77*	-1.77*	-1.48##	-1.57#	-1.48	-3.38 **	-3.32**
O	Nsp1	-2.44 **	-2.44 **	-1.84*	-2.32**	-1.84*	-6.23 **	-5.70**
Alpha	Nsp1	-1.81*	-1.81*	-1.64#	-1.05##	-1.64#	-3.75**	-3.68**
Beta	Nsp1	-2.30**	-2.30**	-1.90 *	-1.98*	-1.90*	-4.25 **	-4.25**
Gamma	Nsp1	-1.97*	-1.97*	-1.42##	-1.82*	-1.42##	-2.54 *	-2.78*
Delta	Nsp1	-1.69#	-1.70#	-1.10##	-1.46##	-1.10##	-3.02 *	-3.05*
L	Nsp2	-2.51***	-2.51***	-2.18**	-2.30 **	-2.18**	-6.93 **	-6.16**
S	Nsp2	-2.56 ***	-2.56 ***	-2.19**	-2.59***	-2.19**	-7.3 **	-6.24**
O	Nsp2	-2.72 ***	-2.72***	-2.65 ***	-2.55 ***	-2.65***	-8.6 **	-7.13**
Alpha	Nsp2	-2.57 ***	-2.57***	-2.40 **	-2.32 **	-2.40 **	-6.79 **	-6.19**
Beta	Nsp2	-2.41 **	-2.41 **	-2.07*	-2.36**	-2.07*	-4.61**	-4.54**
Gamma	Nsp2	-2.29 **	-2.29 **	2.16*	-1.76 #	-2.16*	-3.97 **	-4.01**
Delta	Nsp2	-1.72#	-1.72#	-1.38 ##	-1.62 #	-1.38 #	-2.79*	-2.87*
L	Nsp3	-2.75***	-2.75***	-2.43**	-2.64 ***	-2.43**	-9.15**	-7.60**
S	Nsp3	-2.67 ***	-2.67 ***	- 2.47***	-2.67 ***	- 2.47***	-9.46**	-7.49**
O	Nsp3	-2.77***	-2.77***	-2.69 ***	-2.73***	-2.69 ***	-10.05**	-7.87**
Alpha	Nsp3	-2.83***	-2.83***	-2.68***	-2.73***	-2.68***	-7.52**	-6.76**
Beta	Nsp3	-2.68 ***	-2.68 ***	-2.47**	-2.61***	-2.47**	-5.43**	-5.27**
Gamma	Nsp3	-2.72***	-2.72***	-2.40 **	-2.70***	-2.40 **	-5.88**	-5.56**
Delta	Nsp3	-1.75 #	-1.75 #	-2.33**	-0.87 ##	-2.33**	-4.10**	-3.88**
L	Nsp4	-2.37**	-2.37**	-1.74 #	-2.20**	-1.74 ##	-5.94**	-5.49**
S	Nsp4	-2.25 **	-2.25 **	-2.23**	-1.91*	-2.23**	-6.74**	-5.85**
O	Nsp4	-2.44**	-2.44**	-2.21**	-2.27**	-2.21**	-5.42**	-5.00**
Alpha	Nsp4	-2.42**	-2.42**	-2.33**	-1.73 #	-2.33**	-4.84**	-4.73**
Beta	Nsp4	-2.37 **	-2.37 **	-2.15 *	-2.03 *	-2.15 *	-3.97**	-4.05**
Gamma	Nsp4	-2.43**	-2.43**	-2.00 *	-2.16 *	-2.00 *	-5.28**	-5.07**
Delta	Nsp4	-0.53 ##	-0.53 ##	-0.99 ##	-0.15 ##	-0.99 ##	-1.73 ##	-1.56 ##

Supplement Table 7 . Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 Nsp12, Nsp13, Nsp14 and Nsp15 among Sars-Cov-2 population of clade/variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	Nsp12	613.78	2179.22	23	23	20	3	7	16	1.2	14.2	1.3	2.9	0.8	23	0.27
S	Nsp12	613.76	2179.24	42	42	28	14	13	29	3.7	25.9	3.7	2.8	3.9	44	0.71
O	Nsp12	613.35	2179.65	64	66	48	16	14	52	8.2	39.9	8.2	15.5	6.1	60	0.88
Alpha	Nsp12	613.17	2179.83	13	13	11	2	3	10	2.6	9.3	2.6	1.2	3.1	14	0.54
Beta	Nsp12	613.09	2179.91	17	17	12	5	3	14	3.8	13.3	3.8	5	3.4	14	0.63
Gamma	Nsp12	612.82	2180.18	17	17	15	2	3	14	2	12.5	2	1.4	2.2	14	0.40
Delta	Nsp12	613.17	2179.83	14	14	9	5	3	11	4.2	11.2	4.2	7.9	3.2	14	0.80
L	Nsp13	382.84	1417.16	25	25	22	3	7	15	1.8	23.9	1.8	3.1	1.4	22	0.26
S	Nsp13	383.04	1416.97	55	55	32	23	10	45	14.8	52.6	14.8	7.3	16.8	50	0.87
O	Nsp13	382.89	1417.11	66	66	49	17	13	53	6.9	63.8	6.9	4.2	7.6	51	0.66
Alpha	Nsp13	382.95	1417.05	18	18	16	2	6	12	4.3	20.1	4.3	3.8	4.4	16	0.53
Beta	Nsp13	382.86	1417.14	15	15	12	3	2	13	5.3	18.2	5.3	2.8	6	16	0.63
Gamma	Nsp13	383.1	1416.9	11	11	7	4	1	10	2.9	12.6	2.9	0.7	3.5	12	0.39
Delta	Nsp13	383.41	1416.59	9	9	6	3	1	8	5	11.1	5	1	6	9	0.52
L	Nsp14	358.85	1219.15	16	16	14	2	3	13	1.2	17.4	1.2	0.9	1.3	14	0.15
S	Nsp14	358.74	1219.26	49	49	33	16	15	34	10.4	53.4	10.4	12.5	9.9	43	0.79
O	Nsp14	358.77	1219.23	52	52	37	15	11	41	6.7	57.3	6.7	10.4	5.6	48	0.69
Alpha	Nsp14	358.81	1219.19	22	22	19	3	8	14	4.1	28	4.1	6.8	3.4	21	0.47
Beta	Nsp14	358.79	1219.21	15	15	12	3	1	14	4.1	20.7	4.1	2	4.8	13	0.45
Gamma	Nsp14	358.72	1219.28	16	17	12	4	5	12	4	20.9	4	6.8	3.2	16	0.50
Delta	Nsp14	357.44	1217.56	18	21	10	8	8	8	10	25.4	10	25	5.6	13	0.73
L	Nsp15	209.82	825.18	6	6	5	1	2	4	0.7	10	0.7	1.5	0.5	7	0.07
S	Nsp15	209.84	825.16	23	23	19	4	3	20	3.4	38.2	3.4	1.5	3.9	23	0.31
O	Nsp15	209.84	825.16	19	19	16	3	4	15	2.8	31.9	2.8	2.2	3	19	0.27
Alpha	Nsp15	209.8	825.2	8	8	7	1	2	6	2.1	15.5	2.1	2.3	2.1	8	0.16
Beta	Nsp15	209.85	825.15	10	10	9	1	2	8	3.8	21.1	3.8	3.5	4	11	0.36
Gamma	Nsp15	209.83	825.17	5	5	4	1	2	3	1.6	9.9	1.6	2.7	1.3	6	0.16
Delta	Nsp15	209.89	825.11	6	6	3	3	2	4	7.3	12.9	7.3	9.1	6.9	7	0.52

Supplement Table 8. Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 Nsp5, Nsp6, Nsp7 and Nsp9 among Sars-Cov-2 population of clade/variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	Nsp5	212.65	705.35	11	11	11	0	3	8	1.3	20.6	1.3	1.5	1.2	11	0.10
S	Nsp5	212.66	705.34	21	21	15	6	13	8	7.7	39.4	7.8	23.8	2.9	21	0.56
O	Nsp5	212.64	705.36	11	11	17	6	12	11	1.3	20.6	1.3	7.5	2.6	11	0.10
Alpha	Nsp5	212.66	705.34	6	6	5	1	3	3	1.9	13.1	1.9	4.6	1	7	0.16
Beta	Nsp5	212.98	705.02	11	11	7	4	4	7	7.3	26.2	7.3	16.4	4.6	11	0.42
Gamma	Nsp5	212.66	705.34	5	5	5	0	2	3	1.5	11.2	1.5	2.6	1.2	6	0.14
Delta	Nsp5	212.67	705.33	4	4	3	1	3	1	2.2	9.7	2.2	7.5	0.6	5	0.19
L	Nsp6	202.34	667.66	7	8	6	1	4	4	1.3	13.9	1.3	3.7	0.6	9	0.11
S	Nsp6	202.48	667.52	30	30	21	9	10	20	19.5	59.5	19.5	10	22.4	25	0.80
O	Nsp6	198.02	650.98	38	38	30	8	38	22	10.4	77.7	10.4	19.1	7.9	34	0.50
Alpha	Nsp6	199.97	661.03	13	13	10	3	7	6	4.8	30.3	4.8	8.6	3.6	14	0.33
Beta	Nsp6	200.03	660.97	6	6	4	2	1	5	4.8	15.2	4.8	9.9	3.3	6	0.35
Gamma	Nsp6	199.66	658.34	9	9	6	3	4	5	4.5	21.6	4.5	5.6	4.1	10	0.31
Delta	Nsp6	202.29	667.71	7	7	2	5	2	5	24.6	18	24.6	27.1	3.9	7	0.75
L	Nsp7	55.67	193.33	2	2	2	0	1	1	9	13.8	0.9	1.9	0.6	3	0.02
S	Nsp7	55.68	193.32	9	10	6	3	6	4	6.4	62.3	6.5	17.5	3.3	10	0.15
O	Nsp7	55.64	193.36	7	7	5	2	3	4	7.2	49	7.2	12.3	5.8	8	0.17
Alpha	Nsp7	55.67	193.33	1	1	1	0	0	1	1	8.1	1	0	1.3	2	0.02
Beta	Nsp7	55.65	193.35	3	3	3	0	2	1	4.4	26.3	4.4	13.2	1.9	4	0.11
Gamma	Nsp7	55.67	193.33	2	2	2	0	0	2	2.2	16.6	2.2	0	2.9	3	0.06
Delta	Nsp7	55.68	193.32	2	2	2	0	1	1	3.2	17.9	3.2	7.3	2.1	3	0.08
L	Nsp9	83.99	255.01	5	5	2	3	3	2	4.4	25.4	4.4	15.2	0.8	6	0.14
S	Nsp9	83.86	255.14	10	10	6	4	6	4	16.9	50.9	16.9	63.9	1.7	10	0.44
O	Nsp9	83.97	255.03	12	12	6	6	9	3	7.4	61.7	7.4	24.5	1.8	15	0.22
Alpha	Nsp9	84	255	3	3	2	1	1	2	5.5	17.8	5.5	16.5	1.9	4	0.18
Beta	Nsp9	84	255	3	3	3	0	2	1	3.2	19.3	3.2	8.7	1.4	4	0.11
Gamma	Nsp9	83.97	255.03	5	5	3	2	3	2	6.5	30.4	6.5	19.7	2.2	6	0.21
Delta	Nsp9	83.87	255.13	4	4	1	3	3	1	18.7	26.3	18.8	67.1	3.1	6	0.53

Supplement Table 9. Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 Nsp1, Nsp2, Nsp3 and Nsp4 among Sars-Cov-2 population of clade/variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	Nsp1	127.31	403.69	11	11	8	3	6	5	4.4	35.7	4.4	10.8	2.4	13	0.21
S	Nsp1	124.78	394.22	17	17	9	8	6	11	18.2	56.1	18.3	23.3	16.9	17	0.62
O	Nsp1	128.99	411.01	20	20	16	4	7	13	6.5	64.5	6.5	14	4.2	21	0.28
Alpha	Nsp1	127.33	403.67	4	4	4	0	3	1	1.8	15.1	1.8	5.8	0.6	5	0.10
Beta	Nsp1	128.98	411.02	10	10	9	1	5	5	7.4	40.5	7.4	16.9	4.4	9	0.30
Gamma	Nsp1	128.98	411.02	8	8	5	3	4	4	7.1	30.6	7.1	21.1	2.7	9	0.31
Delta	Nsp1	127.32	403.68	3	3	3	0	1	2	2.3	12.6	2.3	3.2	2	4	0.12
L	Nsp2	441.24	1466.76	28	28	22	6	10	18	2.8	25.2	2.8	2.9	2.8	28	0.36
S	Nsp2	441.96	1472.04	62	62	43	19	27	35	8.8	55.8	8.8	23.5	4.4	55	0.79
O	Nsp2	441.89	1472.11	84	84	64	20	41	43	9.4	76.5	9.4	16.9	7.2	69	0.82
Alpha	Nsp2	441.88	1472.12	33	33	30	3	18	15	6.3	34.6	6.3	14.9	3.8	28	0.69
Beta	Nsp2	441.56	1472.44	31	31	24	7	15	16	9.5	35.4	9.5	23.4	5.4	20	0.80
Gamma	Nsp2	442.04	1471.96	15	15	11	4	9	6	3.4	16.2	3.4	7.5	2.2	15	0.53
Delta	Nsp2	442.09	1471.91	16	16	10	6	7	9	8.3	18.7	8.3	16.9	5.7	16	0.77
L	Nsp3	1340.46	4494.54	56	56	47	9	17	39	1.5	16.5	1.5	1.6	1.4	42	0.52
S	Nsp3	1337.78	4485.23	137	137	104	33	56	81	6.2	40.5	6.2	13.1	4.2	92	0.92
O	Nsp3	1340.33	4494.67	206	207	161	45	86	118	8.2	61.5	8.2	15.2	6.1	112	0.95
Alpha	Nsp3	1339.8	4495.2	71	71	63	8	29	42	3.5	24.4	3.5	6	2.8	46	0.83
Beta	Nsp3	1340.38	4494.62	55	55	45	10	24	31	4.5	20.6	4.5	9	3.1	35	0.94
Gamma	Nsp3	1340.12	4494.88	60	60	47	13	22	38	3.9	21.2	3.9	7.2	2.9	38	0.82
Delta	Nsp3	1340.25	4491.75	36	36	26	10	19	17	6.6	13.8	6.6	8.4	6.1	25	0.92
L	Nsp4	362.83	1137.17	16	16	13	3	5	11	1.6	18.4	1.6	2.3	1.4	15	0.18
S	Nsp4	362.9	1137.1	34	34	25	9	14	20	8.6	39.1	8.6	7.7	8.9	32	0.62
O	Nsp4	360.17	1127.83	39	39	24	15	18	21	7.8	45.6	7.8	15.2	5.5	38	0.59
Alpha	Nsp4	362.85	1137.15	16	16	13	3	11	5	3.4	21.4	3.4	8.1	1.9	15	0.41
Beta	Nsp4	362.74	1137.26	17	17	13	4	9	8	5.7	24.8	5.7	11.9	3.8	16	0.53
Gamma	Nsp4	362.84	1137.16	16	17	14	2	7	10	4.2	22	4.2	6.8	NA	15	0.44
Delta	Nsp4	362.71	1137.29	17	17	8	9	6	11	21	25.3	21	22.7	20.5	15	0.83

Supplement Table 10. Fu and Li's Tests and The McDonald-Kreitman Tests for Sars-CoV-2 ORF7b, ORF8, Nucleocapsid and ORF10 genes of clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	ORF7b	0	0	0.00	0.00	-	-	-	-	-	0	0	0	1
S	ORF7b	-4.28**	-4.06**	0.04	0.06	1	0	1	-	-	0	2	2	1
O	ORF7b	-5.31**	-4.91**	0.23	0.17	0	1	1	-	-	0	3	6	1
Alpha	ORF7b	-1.96#	-2.15#	0.10	0.14	0	1	1	-	-	0	2	1	1
Beta	ORF7b	0	0	0.00	0.00	-	-	-	-	-	0	0	0	1
Gamma	ORF7b	0.51##	0.11##	0.05	0.14	0	1	1	-	-	0	1	0	1
Delta	ORF7b	0.54##	1.01##	-0.04	-0.11	-	-	-	-	-	0	0	1	1
L	ORF8	-5.83**	-5.41**	0.20	0.10	5.42	-4.42	0.09	3.351	0.07	5	2	13	6
S	ORF8	-4.96**	-4.60**	0.59	0.24	5.33	-4.33	0.08	3.725	0.05	5	3	16	5
O	ORF8	-4.61**	-4.36**	-2.91	-0.87	10.42	-9.42	0.01*	7.2	0.01**	5	3	25	4
Alpha	ORF8	-5.16**	-4.93**	0.19	0.14	0.74	0.26	1	0.1	0.75	5	3	4	9
Beta	ORF8	-3.17**	-3.31**	-3.13	-1.61	10.00	-9.00	0.06	4.562	0.03*	5	1	10	5
Gamma	ORF8	-3.06**	-3.22**	-3.59	-2.89	5.00	-4.00	0.31	1.9	0.17	5	1	5	5
Delta	ORF8	-2.88**	-2.98**	0.29	0.29	0.83	0.17	1	0.024	0.88	5	2	2	6
L	N	-5.46**	-5.08**	0.76	0.17	12.55	-11.55	0**	23.21	0***	34	13	24	5
S	N	-7.00**	-5.72**	-2.02	-0.36	5.95	-4.95	0**	12.19	0***	31	25	24	5
O	N	-6.28**	-5.29**	-7.69	-0.68	12.22	-11.22	0***	30.66	0***	28	33	72	5
Alpha	N	-4.33**	-4.35**	-5.26	-2.16	1.87	-0.86	0.36	1.09	0.3	32	12	7	10
Beta	N	-4.42**	-4.42**	1.02	0.35	7.56	-6.56	0**	11.51	0***	34	9	12	6
Gamma	N	-4.55**	-4.54**	-13.86	-3.43	10.67	-9.67	0***	18.7	0***	32	12	20	5
Delta	N	-3.75**	-3.89**	-2.82	-1.08	45.33	-44.33	0***	30.62	0***	34	2	16	6
L	ORF10	0	0	0.00	0.00	-	-	-	-	-	0	0	0	1
S	ORF10	-3.09**	-3.17**	0.10	0.10	0	1	1	-	-	0	2	4	1
O	ORF10	-2.43*	-2.58*	0.13	0.15	0	1	1	-	-	0	1	4	1
Alpha	ORF10	-2.78*	-2.76**	0.05	0.09	-	-	-	-	-	0	0	2	1
Beta	ORF10	0	0	0.00	0.00	-	-	-	-	-	0	0	0	1
Gamma	ORF10	-1.96#	-1.97#	0.03	0.07	0	1	1	-	-	0	1	0	1
Delta	ORF10	0	0	0.00	0.00	-	-	-	-	-	0	0	0	1

Supplement Table 11. Fu and Li's Tests and The McDonald-Kreitman Tests of Sars-CoV-2 ORF3b, Envelope, ORF6 and ORF7a genes for clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	ORF3b	-3.61**	-3.55**	0.07	0.08	6	-5	0.24	1.78	0.18	9	3	2	1
S	ORF3b	-4.68**	-4.37**	0.69	0.29	-	-	0***	-	-	9	5	14	0
O	ORF3b	-7.67**	-6.65**	0.59	0.15	52.2	-51.2	0***	20.15	0***	9	5	29	1
Alpha	ORF3b	-3.53**	-3.65**	0.44	0.24	99	-98	0***	16.93	0***	9	1	11	1
Beta	ORF3b	-2.24#	-2.51*	-2.90	-1.60	80	-79	0***	14.55	0***	9	1	10	1
Gamma	ORF3b	-2.05#	-2.32#	-1.65	-1.68	-	-	0***	-	-	9	0	6	1
Delta	ORF3b	-3.53**	-3.54**	0.16	0.18	9	-8	0.18	2.50	0.11	9	2	2	1
L	E	-4.59**	-4.39**	0.08	0.08	-	-	0.4	-	-	1	1	0	3
S	E	-4.75**	-4.47**	0.05	0.06	-	-	1	-	-	1	4	1	0
O	E	-3.57**	-3.53**	0.07	0.08	-	-	0.33	-	-	1	1	4	0
Alpha	E	0	0	0.00	0.00	-	-	-	-	-	1	0	0	0
Beta	E	-1.89##	-1.93##	0.04	0.09	-	-	1	-	-	1	0	1	1
Gamma	E	-3.76**	-3.70**	0.11	0.13	-	-	0.2	-	-	1	0	4	0
Delta	E	-1.87##	-1.91##	0.04	0.10	-	-	1	-	-	1	0	1	0
L	ORF6	-4.59**	-4.39**	0.08	0.08	0	1	0.46	-	-	0	3	3	3
S	ORF6	-2.45*	-2.54*	0.18	0.20	-	-	-	-	-	0	0	5	3
O	ORF6	-2.43*	-2.60*	0.11	0.13	0	1	0.43	-	-	0	2	2	3
Alpha	ORF6	-2.78*	-2.76**	0.05	0.09	-	-	-	-	-	0	0	2	3
Beta	ORF6	-1.89##	-1.93##	0.04	0.09	0	1	0.25	-	-	0	1	0	3
Gamma	ORF6	-1.96#	-1.96#	0.03	0.07	-	-	-	-	-	0	0	1	3
Delta	ORF6	-3.53**	-3.53**	0.16	0.18	-	-	-	-	-	0	0	4	3
L	ORF7a	-2.99**	-3.01**	0.06	0.08	1.33	-0.33	1	0.46	0.83	12	3	3	1
S	ORF7a	-2.30#	-2.63*	-3.67	-2.38	6.00	-5.00	0.1	4.33	0.04*	9	4	8	3
O	ORF7a	-4.50**	-4.28**	0.42	0.24	13.00	-12.00	0.01**	9.30	0**	13	3	9	3
Alpha	ORF7a	-4.66**	-4.55**	0.24	0.16	34.67	-33.67	0**	12.58	0***	13	1	8	3
Beta	ORF7a	-4.77**	-4.67**	0.29	0.20	30.33	-29.33	0**	11.13	0***	13	1	7	3
Gamma	ORF7a	-3.76**	-3.70**	0.11	0.13	17.33	-16.33	0..02*	6.29	0.01*	13	1	4	3
Delta	ORF7a	-2.81*	-2.94**	0.46	0.35	6.5	-5.5	0.08	4.00	0.05*	13	2	5	5

Supplement Table 12. Fu and Li's Tests and The McDonald-Kreitman Tests for Sars-CoV-2 ORF1a, ORF1b, Spike and ORF3a genes of clades/variants with an outgroup (RaTG13)

Clade/Variant	Gene	Fu and Li's D	Fu and Li's F	Fay and Wu's Hn	Fay and Wu's Hn Normalized	NI	Alpha value	Fisher's exact test. P-value (two tailed)	G test G value	G test P value	Synonymous fixed differences between species	Synonymous Polymorphic sites	Nonsynonymous Polymorphic sites	Nonsynonymous Fixed differences between species
L	ORF1a	-10.76**	-8.23**	-11.60	-0.77	9.37	-8.37	0***	120.85	0***	429	52	92	81
S	ORF1a	-9.15**	-6.90**	-59.87	-1.75	6.85	-5.85	0***	149.94	0***	401	148	192	76
O	ORF1a	-9.52**	-7.26**	-101.23	-2.22	6.32	-5.32	0***	155.65	0***	384	200	247	75
Alpha	ORF1a	-7.51**	-6.59**	-34.09	-1.84	5.46	-4.46	0***	73.86	0***	415	76	83	83
Beta	ORF1a	-5.50**	-5.30**	-34.69	-1.97	6.31	-5.31	0***	78.71	0***	417	62	81	76
Gamma	ORF1a	-6.518**	-5.94**	-20.25	-1.32	7.26	-6.26	0***	86.71	0***	429	54	74	81
Delta	ORF1a	-2.40*	-2.47*	-24.42	-2.01	6.00	-5.00	0***	54.03	0***	421	43	49	80
L	ORF1b	-10.12**	-8.05**	-12.86	-1.44	22.46	-21.46	0***	101.3	0***	193	29	54	16
S	ORF1b	-7.97**	-6.31**	-24.74	-1.33	11.66	-10.66	0***	86.89	0***	185	92	87	15
O	ORF1b	-10.02**	-7.65**	-20.98	-0.90	12.1	-11.1	0***	103.4	0***	187	113	117	16
Alpha	ORF1b	-6.59**	-5.96**	-19.00	-2.15	13.97	-12.97	0***	66.05	0***	190	32	40	17
Beta	ORF1b	-5.20**	-5.07**	-6.76	-0.82	8.94	-7.94	0***	38.21	0***	196	37	27	16
Gamma	ORF1b	-5.47**	-5.25**	-11.54	-1.64	12.57	-11.57	0***	48.9	0***	194	27	28	16
Delta	ORF1b	-4.11**	-4.06**	-2.81	-0.40	15.51	-14.07	0***	51.5	0***	197	20	26	17
L	S	-7.95**	-6.83**	-7.38	-1.63	3.90	-2.899	0***	16.66	0***	176	16	28	79
S	S	-7.03**	-5.73**	-14.85	-1.23	4.75	-3.75	0***	44.25	0***	171	36	78	78
O	S	-7.89**	-6.36**	-39.02	-2.28	4.15	-3.145	0***	46.23	0***	164	58	107	73
Alpha	S	-7.28**	-6.56**	-10.56	-1.71	4.22	-3.217	0***	21.94	0***	175	18	36	83
Beta	S	-4.64**	-4.61**	-15.39	-2.33	4.43	-3.431	0***	21.53	0***	174	16	33	81
Gamma	S	-4.76**	-4.66**	-23.40	-3.39	5.05	-4.053	0***	28.03	0***	174	17	39	79
Delta	S	-3.53**	-3.60**	-11.78	-2.67	4.06	-3.061	0***	14.11	0***	175	12	22	79
L	ORF3a	-6.02**	-5.54**	0.14	0.09	3.47	-2.472	0.12	2.68	0.1	25	6	5	6
S	ORF3a	-5.95**	-5.25**	-2.66	-0.61	13.80	-12.8	0***	22.95	0***	23	10	30	5
O	ORF3a	-8.99**	-7.36**	-0.75	-0.10	23.33	-22.333	0***	39.51	0***	25	12	56	5
Alpha	ORF3a	-4.26**	-4.31**	-3.16	-1.01	22.80	-21.8	0***	23.86	0***	24	4	19	5
Beta	ORF3a	-2.97**	-3.25**	-4.30	-1.39	30.40	-29.4	0***	26.34	0***	24	3	19	5
Gamma	ORF3a	-4.40**	-4.39**	-5.00	-1.92	14.38	-13.375	0***	16.66	0***	23	4	15	6
Delta	ORF3a	-5.24**	-5.10**	0.43	0.23	9.52	-8.524	0**	9.11	0**	25	3	8	7

Supplement Table 13. Neutrality tests summary statistics for Sars-CoV-2 ORF7b, ORF8, Nucleocapsid, and ORF10 genes among viral clades variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	ORF7b	NA	NA	NA	NA	NA	0	0
S	ORF7b	-1.69	-1.69	-1.29	-1.29	-1.29	-4.24**	-4.02**
O	ORF7b	-1.96*	-1.96*	-1.53	-1.66	1.53	-5.22**	-4.84**
Alpha	ORF7b	-1.56	-1.56	-1.31	-1.05	-1.31	-1.92	-2.12
Beta	ORF7b	NA	NA	NA	NA	NA	0	0
Gamma	ORF7b	-0.9	-0.9	-0.90	NA	-0.90	0.51	0.11
Delta	ORF7b	1.72	1.72	NA	1.72	NA	0.54	1.02
L	ORF8	-2.36**	-2.36**	-1.29	-2.29 **	-1.29	-5.69**	-5.31**
S	ORF8	-2.09*	-2.09*	-1.52	-1.94*	-1.52	-4.81**	-4.49**
O	ORF8	-2.22**	-2.22**	-1.49	-2.14**	-1.49	-4.82**	-4.52**
Alpha	ORF8	-2.20**	-2.13*	-1.64	-1.81*	-1.64	-4.98**	-4.78**
Beta	ORF8	-2.01*	-2.03*	-1.09	-1.96*	-1.09	-3.50**	-3.55**
Gamma	ORF8	-2.05*	-1.95*	-1.06	-1.84*	-1.06	-3.77**	-3.77**
Delta	ORF8	-1.75	-1.58	-1.02	-1.46	-1.02	-2.75*	-2.86*
L	N	-2.60***	-2.60***	-2.05*	-2.54***	-2.05*	-5.37**	-5.04**
S	N	-1.92*	-1.92*	-1.34	-1.91*	-1.34	-6.99**	-5.74**
O	N	-2.35**	-2.54***	-2.04*	-2.59***	-2.04*	-6.10**	-5.24**
Alpha	N	-2.5**	-2.5**	-1.40	-2.45**	-1.40	-4.91**	-4.80**
Beta	N	-2.40**	-2.40**	-1.47	-2.37**	-1.47	-4.16**	-4.20**
Gamma	N	-2.57***	-2.57***	-1.86*	-2.53***	-1.86*	-5.54**	-5.28**
Delta	N	-1.98*	-1.98*	-0.670	-2.20**	-0.67	-3.57**	-3.59**
L	ORF10	NA	NA	NA	NA	NA	0	0
S	ORF10	-1.86*	-1.86*	-0.96	-1.74	-0.96	-3.05*	-3.14**
O	ORF10	-1.65	-1.65	-1.3	-1.29	-1.3	-2.40*	-2.55*
Alpha	ORF10	-1.4	-1.4	-1.4	NA	-1.4	-2.75*	-2.73*
Beta	ORF10	NA	NA	NA	NA	NA	0	0
Gamma	ORF10	-1.06	-1.06	NA	-1.06	NA	-1.94	-1.96
Delta	ORF10	NA	NA	NA	NA	NA	0	0

Supplement Table 14. Neutrality tests summary statistics for Sars-CoV-2 ORF3b,Envelope, ORF6 and ORF7a genes among viral clades and variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	ORF3b	-1.77*	-1.77*	-1.45	-1.29	-1.45	-3.57**	-3.51**
S	ORF3b	-2.02*	-2.02*	-1.75	-1.79*	-1.75	-4.54**	-4.27**
O	ORF3b	-2.61***	-2.61***	-1.83*	-2.55***	-1.83*	-7.33**	-6.45**
Alpha	ORF3b	-2.22**	-2.22**	-1.05	-2.16*	-1.05	-3.36**	-3.51**
Beta	ORF3b	-1.84*	-1.84*	-1.09	-1.75	-1.09	-2.67*	-2.83*
Gamma	ORF3b	-1.90*	-1.90*	NA	-1.90*	NA	-2.51*	-2.72*
Delta	ORF3b	-1.86*	-1.86*	-1.46	-1.46	-1.46	-3.40**	-3.42**
L	E	-1.99*	-1.66	-0.96	-1.49	-0.96	-4.52**	-4.34**
S	E	-1.82*	-1.82*	-1.69	-0.96	-1.69	-4.70**	-4.43**
O	E	-1.80*	-1.80*	-0.90	-1.69	-0.90	-3.53**	-3.50**
Alpha	E	NA	NA	NA	NA	NA	0	0
Beta	E	-1.09	-1.09	NA	-1.09	NA	-1.87	-1.90
Gamma	E	-1.82*	-1.82*	NA	-1.82*	NA	-3.67**	-3.61**
Delta	E	-1.10	-1.10	NA	-1.10	NA	-1.84	-1.88
L	ORF6	-1.99*	-1.99*	-1.68	-1.48	-1.68	-4.52**	-4.34**
S	ORF6	-1.52	-1.52	NA	-1.52	NA	-2.43*	-2.52*
O	ORF6	-1.7	-1.7	-1.35	-1.30	-1.35	-2.40*	-2.57*
Alpha	ORF6	-1.4	-1.4	NA	-1.40	NA	-2.75*	-2.73*
Beta	ORF6	-1.09	-1.09	-1.09	NA	-1.09	-1.87	-1.90
Gamma	ORF6	-1.06	-1.06	NA	-1.06	NA	-1.94	-1.96
Delta	ORF6	-1.86*	-1.86*	NA	-1.86*	NA	-3.40**	-3.42**
L	ORF7a	-1.63	-1.63	-1.45	-1.63	-1.45	-2.96*	-2.99**
S	ORF7a	-2.07*	-2.07*	-1.52	-1.86*	-1.52	-4.09**	-4.00**
O	ORF7a	-2.02*	-2.02*	-1.21	-2.06*	-1.21	-4.39	-4.20
Alpha	ORF7a	-2.23**	-2.23**	-1.05	-2.16*	-2.04*	-4.48**	-4.40**
Beta	ORF7a	-2.23 **	-2.23 **	-1.69	-1.98*	-1.69	-4.52**	-4.45**
Gamma	ORF7a	-1.89*	-1.89*	-1.42	-1.57	-1.42	-3.01*	-3.11**
Delta	ORF7a	-1.81*	-1.81*	-0.87	-1.76	-0.87	-2.65*	-2.80*

Supplement Table 15. Neutrality tests summary statistics for Sars-CoV-2 ORF1a, ORF1b, Spike and ORF3a genes among viral clades and variants

Clade/Variant	Gene	Tajima's D	Tajima's D - Coding	Tajima's D - Synonymous	Tajima's D - Nonsynonymous	Tajima's D - Silent	Fu-Li's D*	Fu-Li's F*
L	ORF1a	-2.85***	-2.85***	-2.72***	-2.80***	-2.72***	-10.63**	-8.33**
S	ORF1a	-2.6***	-2.6***	-2.53***	-2.58***	-2.53***	-9.53**	-7.33**
O	ORF1a	-2.82***	-2.82***	-2.79***	-2.79***	-2.79***	-10.21**	-7.89**
Alpha	ORF1a	-2.86***	-2.86***	-2.82***	-2.78***	-2.82***	-7.84**	-6.92**
Beta	ORF1a	-2.71***	-2.71***	-2.54***	-2.76***	-2.54***	-5.68**	-5.44**
Gamma	ORF1a	-2.79***	-2.79***	-2.63***	-2.76***	-2.63***	-6.36**	-5.90**
Delta	ORF1a	-1.48	-1.48	-1.96*	-0.97	-1.96*	-3.51**	-3.29**
L	ORF1b	-2.85***	-2.85***	-2.55***	-2.81***	-2.55***	-10.72**	-8.56**
S	ORF1b	-2.56***	-2.56***	-2.54***	-2.50***	-2.54***	-8.14**	-6.54**
O	ORF1b	-2.78***	-2.78***	-2.40**	-2.77***	-2.40**	-9.63**	-7.60**
Alpha	ORF1b	-2.69***	-2.69***	-2.57***	-2.57***	-2.57***	-7.10**	-6.38**
Beta	ORF1b	-2.61***	-2.61***	-1.91*	-2.63***	-1.91*	-5.19**	-5.05**
Gamma	ORF1b	-2.72***	-2.72***	-2.26**	-2.67***	-2.26**	-5.79**	-5.50**
Delta	ORF1b	-2.15*	-2.11*	-1.5	-2.22**	-1.5	-3.68**	-3.73**
L	S	-2.72***	-2.72***	-2.40**	-2.583***	-2.40**	-8.44**	-7.18**
S	S	-2.33**	-2.40**	-2.32**	-2.31**	-2.32**	-7.13**	-5.85**
O	S	2.62***	-2.60***	-2.76***	-2.41**	-2.76***	-8.92**	-7.11**
Alpha	S	-2.84***	-2.84***	-2.57***	-2.73***	-2.57***	-8.06**	-7.17**
Beta	S	-2.53***	-2.53***	-2.31**	-2.44**	-2.31**	-5.43**	-5.20**
Gamma	S	-2.53***	-2.53***	-2.31**	-2.42**	-2.31**	-4.47**	-4.45**
Delta	S	-2.08*	-2.08*	-2.31**	-1.71	-2.31**	-4.17**	-4.07**
L	ORF3a	-2.23**	-2.23**	-1.88*	-1.82*	-1.88*	-5.90**	-5.45**
S	ORF3a	-2.29**	-2.29**	-2.13*	-2.10*	-2.13*	-6.31**	-5.53**
O	ORF3a	-2.67***	-2.67***	-2.28**	-2.60***	-2.28**	-8.50**	-7.09**
Alpha	ORF3a	-2.49**	-2.49**	-1.81*	-2.39**	-1.81*	-4.73**	-4.65**
Beta	ORF3a	-2.25**	-2.25**	-1.47	-2.20**	-1.47	-3.22*	-3.42**
Gamma	ORF3a	-2.44**	-2.44**	-1.69	-2.36**	-1.69	-4.70**	-4.62**
Delta	ORF3a	-2.39**	-2.39**	-1.70	-2.24**	-1.70	-4.88**	-4.79**

Supplement Table 16. Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 ORF7b, ORF8, Nucleocapsid and ORF10 genes among Sars-Cov-2 population of clades and variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	ORF7b	27	105	0	0	0	0	0	0	0	0	0	0	0	1	0.00
S	ORF7b	26.66	102.34	4	4	4	0	2	2	3.3	53.5	3.4	8.6	2.1	3	0.02
O	ORF7b	26.98	105.02	9	9	8	1	3	6	18.1	118.8	18.3	13.2	19.7	7	0.19
Alpha	ORF7b	26.97	105.03	3	3	2	1	2	1	7.3	45.7	7.4	28.1	2.3	4	0.07
Beta	ORF7b	27	105	0	0	0	0	0	0	0	0	0	0	0	1	0.00
Gamma	ORF7b	26.98	105.02	1	1	0	1	1	0	4.1	15.6	4.2	21.1	0	2	0.06
Delta	ORF7b	26.83	105.17	1	1	0	1	0	1	38.6	16.9	38.8	0	48.7	2	0.51
L	ORF8	83.46	279.54	15	15	12	3	2	13	5.5	70.6	5.6	2.6	6.5	13	0.13
S	ORF8	81.14	269.86	19	20	12	7	3	16	21.7	92.3	21.8	4	25.9	20	0.58
O	ORF8	82	275	28	28	17	11	3	25	29.9	135.5	30.1	5.6	37.7	28	0.48
Alpha	ORF8	80.17	273.83	8	8	8	0	3	4	5.4	44.6	5.4	9.2	3.6	8	0.16
Beta	ORF8	83.1	279.9	12	12	9	3	1	10	22	71.7	22.1	5.2	23.6	14	0.54
Gamma	ORF8	83.48	279.52	7	7	6	1	1	5	6.8	39.5	6.8	3.4	6.9	7	0.21
Delta	ORF8	82.16	274.84	5	5	4	1	2	2	8.6	31	8.6	23.3	2.9	6	0.29
L	N	292.75	964.25	39	40	23	16	10	30	6.2	53.3	6.2	7.6	5.8	33	0.44
S	N	292.17	964.83	51	52	35	16	6	40	25.6	69.8	25.7	12.8	29.7	40	0.88
O	N	292.22	964.78	104	108	61	43	25	79	37.6	144	37.7	44.8	35.7	87	0.96
Alpha	N	288.13	962.87	19	19	15	4	2	17	4.6	30.4	4.6	1.7	5.5	18	0.41
Beta	N	292.23	964.77	21	21	16	5	3	18	8.8	36.4	8.8	6.1	9.6	21	0.77
Gamma	N	292.41	964.59	32	32	26	6	6	26	10.2	52.4	10.2	8.1	10.8	21	0.57
Delta	N	291.55	965.45	18	18	13	5	4	14	11.8	31.9	11.8	21.6	8.9	17	0.80
L	ORF10	24.67	89.33	0	0	0	0	0	0	0	0	0	0	0	1	0.00
S	ORF10	24.66	89.34	6	6	4	2	1	5	8.2	88.4	8.3	4.5	9.6	7	0.09
O	ORF10	24.67	89.33	5	5	3	2	2	3	11.5	74.5	11.5	9.5	12.6	6	0.13
Alpha	ORF10	24.67	89.33	2	2	2	0	2	0	4.2	34.3	4.2	20.3	0	3	0.05
Beta	ORF10	24.67	89.33	0	0	0	0	0	0	0	0	0	0	0	1	0.00
Gamma	ORF10	24.67	89.33	1	1	1	0	0	1	2.4	17.6	2.4	0	3.1	2	0.03
Delta	ORF10	24.67	89.33	0	0	0	0	0	0	0	0	0	0	0	1	0.00

Supplement Table 17. Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 ORF3b, Envelope, ORF6 and ORF7a genes among Sars-Cov-2 population of cclades and variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	ORF3b	104.67	351.33	5	5	4	1	3	2	1.6	18.9	1.6	5.1	0.6	5	0.06
S	ORF3b	102.29	341.71	19	19	12	7	5	14	17.8	73.1	17.8	8.4	20.9	21	0.61
O	ORF3b	103.32	346.68	33	34	26	7	5	29	13.2	127.2	13.3	5.6	15.6	31	0.45
Alpha	ORF3b	104.67	351.33	12	12	8	4	1	11	10	52.9	10	2.3	12.3	13	0.37
Beta	ORF3b	104.3	351.7	11	11	7	4	1	10	18.9	52.7	19	3.5	23.6	13	0.54
Gamma	ORF3b	104.64	351.36	6	6	4	6	0	6	5.4	27.1	5.4	0	7	7	0.21
Delta	ORF3b	104.66	351.34	4	4	4	0	2	2	3.5	19.6	3.5	7.7	2.3	5	0.16
L	E	60.82	167.18	6	7	5	1	1	3	3.7	45.3	3.8	3.6	3.8	7	0.07
S	E	60.83	167.17	5	5	4	1	5	0	2.4	37.8	2.4	7.2	0.6	6	0.05
O	E	60.83	167.17	5	5	4	1	1	4	3	38.2	3	3.8	2.7	6	0.07
Alpha	E	60.83	167.17	0	0	0	0	0	0	0	0	0	0	0	1	0.00
Beta	E	60.83	167.17	1	1	1	0	0	1	1.6	9.6	1.6	0	2.2	2	0.04
Gamma	E	60.81	167.19	4	4	4	0	0	4	4.9	36.2	4.9	0	6.7	5	0.11
Delta	E	60.83	167.17	1	1	1	0	0	1	1.8	9.8	1.8	0	2.4	2	0.04
L	ORF6	34.52	148.48	7	7	6	1	4	3	4.6	64.8	4.6	11	3.3	5	0.05
S	ORF6	33.86	146.14	5	5	3	2	0	5	10.1	46.6	10.2	0	12.9	6	0.17
O	ORF6	34.52	148.48	5	5	3	2	3	2	6.1	46.8	6.1	24.9	1.9	5	0.09
Alpha	ORF6	34.49	148.51	2	2	2	0	0	2	2.6	21.6	2.6	0	3.3	3	0.05
Beta	ORF6	34.5	148.5	1	1	1	0	1	0	2	11.8	2	10.7	0	2	0.04
Gamma	ORF6	34.5	148.5	1	1	1	0	0	1	1.5	11.1	1.5	0	1.9	2	0.03
Delta	ORF6	34.5	148 .5	4	4	4	0	0	4	8.6	48	8.6	0	10.8	5	0.16
L	ORF7a	79.51	277.49	4	4	3	1	0	4	1.8	19	1.8	0	2.3	5	0.06
S	ORF7a	75.33	257.67	12	12	8	4	3	9	9.6	60.8	9.6	4.3	11.4	12	0.28
O	ORF7a	75.15	266.85	13	13	9	4	5	8	12.8	65.5	12.9	44.7	4.3	14	0.26
Alpha	ORF7a	81.5	281.5	9	9	8	1	6	3	6.6	49.4	6.7	18.1	3.5	10	0.23
Beta	ORF7a	81.48	281.52	8	8	8	0	3	5	7.9	47.8	8	13.5	6.5	7	0.21
Gamma	ORF7a	81.5	281.5	5	5	4	1	2	3	4.5	28.2	4.5	6.9	3.9	6	0.16
Delta	ORF7a	81.43	281.57	7	7	5	2	1	6	13.5	42.7	13.5	9.7	14.8	8	0.41

Supplement Table 18. Population genetic summary statistics for nucleotide diversity of Sars-CoV-2 ORF1a, ORF1b, Spike and ORF3a genes among Sars-Cov-2 population of clades and variants

Clade/Variant	Gene	Synonymous sites	Nonsynonymous sites	Segregating sites	Eta	Singleton informative sites	Parsimony informative sites	Synonymous Polymorphism	Replacement Polymorphism	π (Pi) All Sites	Theta-W All Sites	π (JC) All Sites	π (JC) Synonymous Sites	π (JC) Nonsynonymous Sites	Haplotype	Haplotype Diversity
L	ORF1a	3061.81	10126.19	143	144	118	25	52	92	1.8	18.7	1.8	2.7	1.5	73	0.76
S	ORF1a	3057.4	10112.6	339	340	242	97	148	192	8.5	44.4	8.5	16.3	6.2	139	0.98
O	ORF1a	3056.97	10113.04	450	451	342	108	201	247	7.6	59.5	7.6	14.5	5.5	149	0.99
Alpha	ORF1a	3059.48	10125.53	159	159	137	22	76	83	3.9	24.2	3.9	7.6	2.7	71	0.99
Beta	ORF1a	3061.64	10132.36	138	138	110	28	62	76	5.3	22.9	5.4	11.7	3.4	49	0.99
Gamma	ORF1a	3061.26	10129.74	128	129	100	28	54	75	3.8	20	3.8	7.3	2.7	54	0.95
Delta	ORF1a	3062.2	10128.8	92	92	56	36	92	49	9.1	15.6	9.1	13.4	7.8	40	0.99
L	ORF1b	1751.45	6333.55	83	83	74	9	27	53	1.3	17.7	1.3	2.5	1	59	0.55
S	ORF1b	1751.56	6333.44	185	185	122	63	41	144	7.6	39.4	7.6	5.5	8.2	112	0.97
O	ORF1b	1751.07	6333.93	228	230	170	56	47	183	6.6	49.1	6.6	9.2	5.9	132	0.98
Alpha	ORF1b	1750.9	6334.1	72	72	61	11	20	52	3.4	17.9	3.4	3.1	3.5	47	0.89
Beta	ORF1b	1750.7	6334.3	64	64	50	14	10	44	4.2	17.3	4.2	4	4.3	35	0.93
Gamma	ORF1b	1750.62	6334.38	54	55	42	12	12	43	2.5	13.8	2.5	2.5	2.5	38	0.84
Delta	ORF1b	1750.08	6331.92	51	54	32	19	14	35	5.6	14.1	5.6	9.2	4.7	34	0.97
L	S	850.85	2950.16	45	45	37	8	16	29	1.7	20.3	1.7	2.4	1.5	34	0.40
S	S	847.41	2935.59	122	124	74	46	38	79	14.5	55.3	14.5	18.9	13.3	94	0.96
O	S	842.9	2925.1	169	174	122	47	59	109	14.2	77.7	14.2	11.1	15.1	107	0.96
Alpha	S	849.8	2948.2	54	55	52	2	18	36	3.8	28.5	3.8	5.2	3.4	37	0.70
Beta	S	849.8	2945.2	51	51	42	9	16	33	7.6	29.3	7.6	10	6.5	32	0.92
Gamma	S	852.81	2948.19	56	57	36	20	17	40	7.4	30.3	7.4	8.3	7.2	36	0.82
Delta	S	845.64	2937.36	34	34	25	9	12	22	7.6	20	7.6	7	7.8	31	0.95
L	ORF3a	196.67	631.33	11	11	10	1	6	5	1.7	22.9	1.7	4.3	8	11	0.12
S	ORF3a	194.15	621.85	39	40	26	13	10	30	18.5	82	18.6	8.3	21.9	44	0.73
O	ORF3a	195.24	626.76	67	68	52	15	12	56	18.9	141.7	19	8.2	22.4	57	0.81
Alpha	ORF3a	196.64	631.36	23	23	17	6	4	19	9.9	55.8	9.9	5	11.4	23	0.57
Beta	ORF3a	196.28	631.72	22	22	14	8	3	19	16.9	58.1	17	9.2	19.4	21	0.75
Gamma	ORF3a	196.58	631.42	19	19	15	4	4	15	8.9	47.3	9	8.4	9.1	16	0.48
Delta	ORF3a	196.35	631.65	11	11	11	0	3	8	5.3	29.7	5.3	6.1	5.1	10	0.33