

SARS-CoV-2: Review of Structure, Genome, Genetic Variants, and Vaccines

Amir Jalali* and Mahsa Khoramipour

Department of Biology, Faculty of Sciences, Arak University, Arak, Iran

ARTICLE INFO

Article history:

Received 17 July 2021

Accepted 22 September 2021

Available online 02 November 2021

Keywords:

Angiotensin-converting enzyme 2

COVID-19

RNA viruses

Spike protein

Virulence

Supplementary information:

Supplementary information for

this article is available at

<http://sc.journals.umz.ac.ir/>

*Corresponding authors:

✉ A. Jalali

a-jalali@araku.ac.ir

p-ISSN 2423-4257

e-ISSN 2588-2589

ABSTRACT

The emergence and outbreak of the deadly novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019, became a global health problem in the last two years. SARS-CoV-2 is an enveloped virus with a non-segmented single-strand positive-sense RNA. It has some similarities with other coronaviruses, especially the SARS-CoV. However, the specific features of this virus have changed its pathogenicity and transmissibility compared to other coronaviruses. The distinctive structural differences of the SARS-CoV-2 spike protein have a key role in the kinetics of viral load and a broad range of virus tissue tropism. Because of these differences, SARS-CoV-2 has a greater affinity for binding host cell receptor angiotensin-converting enzyme 2 than SARS-CoV. Since its emergence, the SARS-CoV-2 genome has undergone several mutations. However, a small number can alter the virus antigenicity and clinical features of the disease, leading to the formation of different SARS-CoV-2 variants. Some of these variants have higher transmissibility, severity, and impact on host immunity than the original SARS-CoV-2. Although there are currently no specific therapeutic interventions for coronavirus disease 2019 (COVID-19), manufacturers are working to make and update vaccines by continuously monitoring antigenic and genetic changes in the SARS-CoV-2 population worldwide. Some of these vaccines are very effective against different variants of the virus, therefore in some countries, with widespread vaccination and compliance with health protocols, people have largely returned to normal living conditions. To better understand SARS-CoV-2, in this article, we reviewed its classification, genome organization, structure, and life cycle. In addition, an overview of the mutations occurring in the spike protein and the characteristics of the different variants is given here. Finally, vaccines produced against this coronavirus were briefly introduced.

© 2022 UMZ. All rights reserved.

Please cite this paper as: Jalali A, Khoramipour M. 2022. SARS-CoV-2: Review of structure, genome, genetic variants, and vaccines. *J Genet Resour* 8(1): 16-34. doi: 10.22080/jgr.2021.21980.1270.

Introduction

Since December 2019, the emergence of a new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from Wuhan province in China and its rapidity worldwide pandemic has had a devastating impact on the human population deaths many million people. SARS-CoV-2 (termed 2019 novel coronavirus or 2019-nCoV) was assigned this name because its genome is 85% identical to SARS-CoV. The pandemic of coronavirus disease 2019 (COVID-19) is the third large-scale outbreak of CoV-induced

disease in the last three decades after severe acute respiratory syndrome (SARS) and Middle Eastern respiratory syndrome (MERS) (Zhu *et al.*, 2020). The severe acute respiratory syndrome is caused by the SARS-CoV-2 is milder but more contagious than SARS-CoV (2002) and MERS-CoV (2013) (Munster *et al.*, 2020). There is some evidence that these viruses are zoonotic coronaviruses and appear to originate from bat CoVs. The considerable genomic identity (up to 96%) between these human CoVs and bat CoVs supports this idea.



Furthermore, based on the high similarity (90.3%) of the complete genome sequence of pangolin CoV and SARS-CoV-2, some researchers have also proposed a pangolin origin for SARS-CoV-2 (Zhang *et al.*, 2020a).

The virus is primarily transmitted by coughing or sneezing and nasal discharge or droplets of saliva from an infected individual. It mainly affects the respiratory system and causes influenza-like symptoms such as dry cough, fever, fatigue, and feeling of not being able to breathe well in acute cases (Zou *et al.*, 2020). However, along with respiratory disease, COVID-19 can cause unusual and unpredictable disorders such as systemic inflammation, development of acute kidney injury (AKI), cardiovascular disorders, and multi-organ dysfunction in high-risk patients. Multi-organ tropism is because of SARS-CoV-2 entering the host cells through angiotensin-converting enzyme 2 (ACE2) as an endogenous viral receptor. According to available statistics, the death rate due to COVID-19 is higher in people older than 60 and those with other diseases.

Here, we aim to briefly review the classification, structural properties, genome organization, life cycle, and molecular mechanisms of SARS-CoV-2 pathogenesis. We have introduced the most common SARS-CoV-2 variants and also provided useful data about them, such as the classification, origin, and the most important functional mutations that have a significant impact on the pathogenesis, transmissibility, and clinical demonstrations of the virus. Finally, we have provided updated information on COVID-19 vaccines, which includes the type of vaccine, storage conditions, effectiveness against different variants, etc. In general, this review article can provide a proper reference for the improvement of control and medical care programs for Covid-19 infection.

Classification of coronaviruses

Coronaviruses (CoVs) are spherical, enveloped pathogenic viruses with a diameter of 80-160 nanometers. This genus falls into the Coronaviridae family of the Nidovirales order. Within the Nidovirales order, which was introduced for the first time by ICTV in 1996, eight suborders have been developed: *Nanidovirineae*, *Ronidovirineae*,

Tornidovirineae, *Arnidovirineae*, *Mesnidovirineae*, *Abnidovirineae*, *Monidovirineae*, and *Cornidovirineae* (Walker *et al.*, 2019). In between, the *Cornidovirineae* suborder consists of one family, Coronaviridae, with two subfamilies, *Letovirinae* and *Orthocoronavirinae*. The *Orthocoronavirinae* subfamily referred to as "CoVs" in this review, is the largest in the Nidovirales order, with four genera: *Alpha-*, *Beta-*, *Gamma-*, and *Delta-* coronaviruses. *Alpha-* and *Beta-* CoVs affect only mammals, *Gamma-* CoVs infect some birds species, and *Delta-* CoVs influence both mammals and birds (Woo *et al.*, 2012).

Shape and structure of coronaviruses

The presence of club-like protein spikes (S) on the surface of spherical virus particles creates a crown-like (corona in Latin) shape on electron micrographs, giving them the name coronavirus. The four major structural proteins in coronavirus particles are the envelope (E), spike (S), membrane (M), and nucleocapsid (N) proteins, which share significant sequence homology in MERS-CoV, SARS-CoV, and SARS-CoV-2. The S, E, and M proteins are on the surface, while N protein complexes with a positive single-stranded RNA genome create a helically symmetrical nucleocapsid inside the envelope. In some beta-CoVs, another structural helically symmetrical nucleocapsid protein, the hemagglutinin esterase (HE), constructs secondary spikes at the membrane. All structural proteins participate in the host's innate immune response, and their coding sequence is organized at the 3' end of the SARS-CoV-2 genome.

The N protein is needed for the synthesis and packaging of RNA, and finally, virion assembly. This protein is extremely conserved in CoV and has about 90% sequence similarity between SARS-CoV and SARS-CoV-2. The RNA binding domain of the N protein is composed of 140 amino acids and bind to the viral genome in a beaded conformation (Fehr and Perlman, 2015). The phosphorylated status of the N proteins is essential for their attachment to viral versus non-viral RNA. Due to its high immunogenicity, this protein may be an attractive candidate for a possible vaccine target. The E protein acts as an ion channel (viroporin) and is the smallest (~8-12 kDa, 75 amino acids)

structural protein. This protein is required for different steps of the viral cycle including the etiology, virus assembly, intracellular transportation, and release of the virus (Fallah *et al.*, 2021). The amino acid composition of the E protein is considerably variable, but its structure is preserved across various genera of β -coronaviruses (Schoeman and Fielding, 2019). The M protein (~25-30 kDa, 222 amino acids) is the most frequent structural protein in the virus particles, participating in virus assembly and morphogenesis (Song *et al.*, 2019). Although the amino acid sequence is different, the structure of M protein is conserved in the various genera. This protein shows O-linked glycosylation in β - and δ -CoVs and N-linked glycosylation in other coronaviruses. The M protein glycosylation is critical for interferon signaling and organ tropism (Jacobs *et al.*, 1986). For example, de Haan *et al.* (2003) reported the effect of the glycosylation situation of the M protein on the ability of the murine hepatitis coronavirus to replicate in the liver but not in brain cells.

The S protein (~150 kDa, 1273 amino acids) is a type I transmembrane glycoprotein that is subject to host cell-surface receptor interaction and next entry by an endocytosis process. The spikes on the viral surface are made up of homotrimers of S proteins, which guide the connection to host receptors (Song *et al.*, 2018). Therefore, this protein can be one of the main targets for vaccines and anti-SARS-CoV drugs production. The S protein has two parts, an amino (N)-terminal subunit called S1 and S2 subunit at the carboxyl (C)-terminal. The most important part of the S1 subunit is the receptor-binding domain (RBD), containing amino acids 319–529 (211 amino acids). Angiotensin-converting enzyme 2 (ACE2) is a membrane exopeptidase that is used by SARS-CoV and SARS-CoV-2 as a receptor to infect human lungs, kidneys, heart, and intestines cells. Besides the human ACE2, SARS-CoV-2 can also recognize ACE2 receptors of rabbit, rhesus monkey, pig, civet, ferret, dog, cat, and pangolin (Shi *et al.*, 2020; Zhou *et al.*, 2020). Although a common receptor is used to access host cell cytosol, the RBD of the SARS-CoV-2 spike is significantly differing from the SARS-CoV. The first connection occurs between the ACE2 and the amino acids 437–507 of the S protein. This

part of RBD is varied from SARS-CoV in the five residues required for ACE2 attachment, including Y455L, L486F, N493Q, D494S, and T501N. In addition, the four amino acids 482–485 of SARS-CoV-2 RBM help stabilize S protein-ACE2 receptor interaction resulted in better interaction with the N-terminal helix of ACE2 compared with SARS-CoV (Shang *et al.*, 2020a). This difference can explain why previously developed antibodies and vaccines for SARS-CoV are inefficient against SARS-CoV-2. Although the cause is unknown, most other coronaviruses also use host cell surface peptidases as receptors to enter the cell. In MERS-CoV, the RBD is utilized to bind with the dipeptidyl peptidase 4 (DPP4) receptors, and many α -coronaviruses enter through aminopeptidase N (Zhang *et al.*, 2020b). The MERS-CoV only can employ the receptor of particular animals such as camels, rabbits, bats, horses, and humans for contagion. Because of variations in the DPP4 structure, the virus is not effective in mouse cells infection; as a result, vaccine production in this animal has become a significant challenge for researchers (Zhao *et al.*, 2009). Fusion of the cell-virus membrane is mediated by the S2 subunit, which consists of three distinct parts: a cytoplasmic domain, a transmembrane domain, and a fusion peptide. During the S1 subunit/receptor interaction, conformational changes in the S protein are followed by an acid-dependent proteolytic degradation leading to uncoating viral fusion peptide. The proteolytic process is done with the participation of host proteases, including a cysteine protease cathepsin L, transmembrane protease serine 2 (TMPRSS2), and furin (Hoffmann *et al.*, 2020; Shang *et al.*, 2020b). The viral and cellular membranes fusion as a crucial step in activating the endocytic route is necessary for the viral infection process. Because the S protein has an essential role in the immune system responses and is implicated in viral pathogenesis via stimulation of the endoplasmic reticulum (ER) stress response, its mutational alteration can lead to changes in the pathogenesis of SARS-CoV-2 (Aoe, 2020).

Genome of coronaviruses

The positive-sense single-stranded RNA (ssRNA+) genome of CoVs is the largest

recognized RNA virus genome with a length of 26 to 32 kb. The positive sense of the genome means it can translate by the host cell's ribosomes directly as a functional messenger RNA (mRNA) to synthesize viral proteins immediately after penetration into the cell. The genome of SARS-CoV-2 has about 50% identity in the sequence with MERS-CoV and 79% with SARS-CoV (Lu *et al.*, 2020).

As shown in Fig. 1, the arrangement of open reading frames (ORFs) in the SARS-CoV-2

genome is the same as SARS-CoV and MERS-CoV. There is a cap structure (m⁷GpppA₁) at the 5' end. This structure is synthesized by several enzymes encoded by the SARS-CoV-2 genome and is necessary for the survival and further replication of viral RNA in host cells (Krafcikova *et al.*, 2020; Fallah *et al.*, 2021). Leader sequence and untranslated region (UTR) at the 5' end contain seven stem-loop structures that are essential for RNA transcription and replication (Brown *et al.*, 2007).

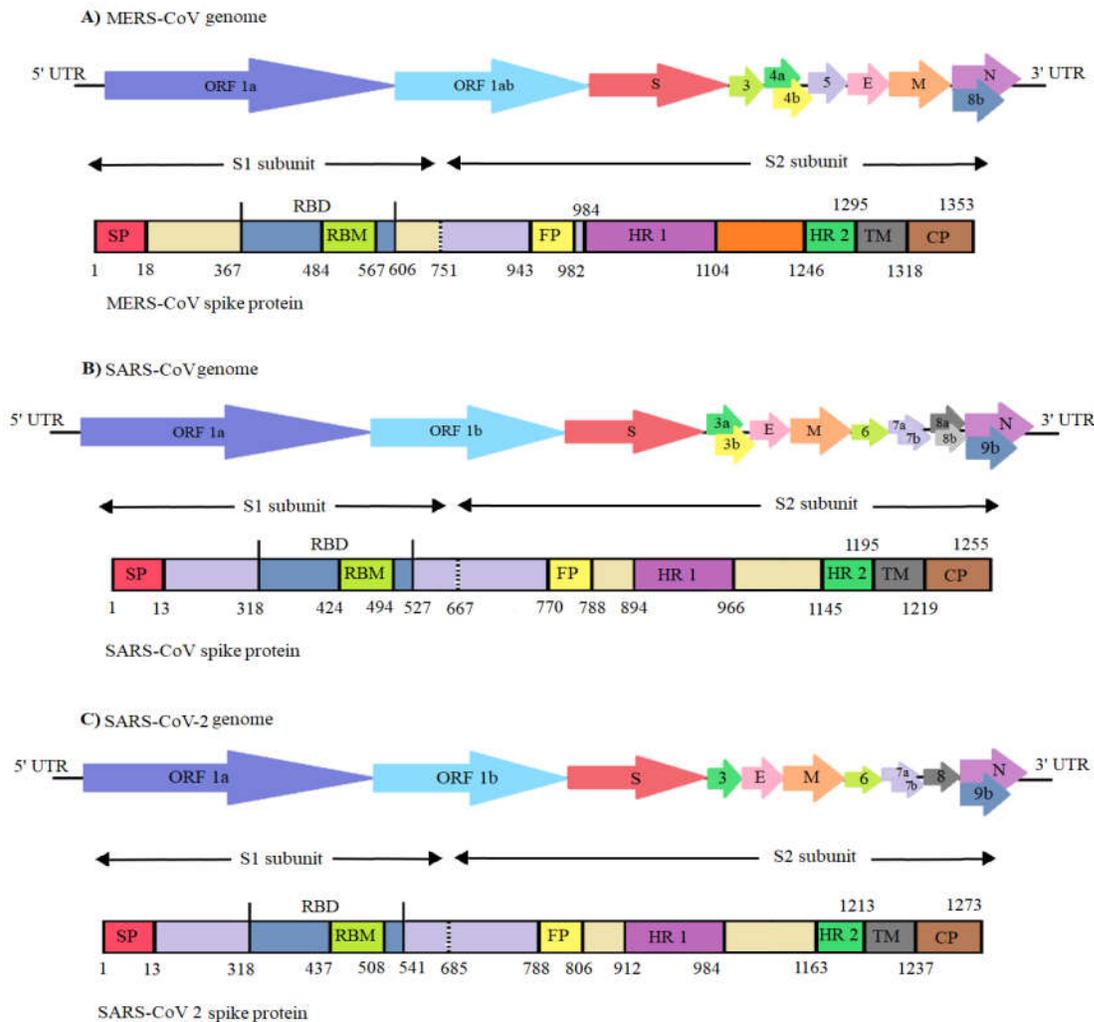


Fig 1. Schematic organization of the MERS-CoV, SARS-CoV, and SARS-CoV-2 RNA genomes: From the 5' of the genome, there are two large open reading frames (ORF1a and ORF1b) which are highly conserved between coronaviruses and encode two long polypeptide precursors, pp1a and pp1ab, respectively; The remaining 3' of the genome coding four structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), along with a set of nonstructural proteins; The structure of each CoV spike protein contains S1 and S2 subunits, indicated below each genome. CP= Cytoplasm domain, FP= Fusion peptide, HR= Heptad repeat, RBD= Receptor binding domain, RBM= Receptor binding motif, SP= Signal peptide, TM= Transmembrane domain.

After these sections, there are two large open reading frames (ORF1a and ORF1b), which make up the first two-thirds of the genome and encode nonstructural proteins (nsps) of the virus. Genes coding for the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, and nine accessory proteins that are interspersed between these structural genes, constitute the remaining third of the genome (Chan *et al.*, 2020). Each accessory and the structural gene have transcriptional regulatory sequences (TRSs) at the start that might be crucial for their expression. The 3' UTR also involves structures required for regulating alternate stages of RNA virus synthesis and replication (Liu *et al.*, 2001). Finally, there is a poly-A tail (~30-60 nucleotides long) on the 3' end that provides stability for the viral genome. The 5' and 3' terminal structures allow the CoV genome to be used as mRNA to translate the replicase polyproteins.

The giant replicase locus at the 5' end constitutes two-thirds of the genome and consists of ORF1a, and ORF1b expresses 16 nonstructural proteins. These nonstructural proteins make up two large polyproteins (pp1a and pp1ab), with various tasks contributing to the transcription and replication of the viral genome. The synthesis of pp1a with NSP1-11 and pp1ab with NSP11-16, takes place by ribosomal frameshifting from the

ORF1a into the ORF1b (Araki *et al.*, 2010). The greater than 85% amino acid sequence of the majority of the SARS-CoV-2 nonstructural proteins have an identity with SARS-CoV (Chan *et al.*, 2020). Although the exact reason for controlling protein expression using translational recoding is not well known, there are two possibilities for this: 1) to regulate the exact ratio of repla and replb proteins; 2) to stop the production of replb proteins since the repla products prepare a suitable environment for the replication of RNA (Araki *et al.*, 2010).

As shown in Fig. 1, the last one-third of the RNA preceding 3'-end coding four structural proteins: S, E, M, N, and nine accessory proteins (encoded by ORF3a, ORF3d, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF14, and ORF10 genes). The features and function of accessory proteins that are not necessary for virus replication in cell culture, but have a crucial role in virus (Perlman and Netland, 2009) are exhibited in Table 1. Different CoVs have distinct accessory and structural proteins. Except for the gene encoding the spike protein, which is highly variable and shares <75% nucleotide identity with SARS-CoV, other structural proteins of SARS-CoV-2 are more than 90% similar in amino acid sequence with SARS-CoV (Lu *et al.*, 2020; Zhou *et al.*, 2020).

Table 1. Functions of coronavirus accessory proteins.

Protein	Characteristics and Function	References
ORF3a	Its ORF located between the S and E genes, the largest accessory proteins of SARS-CoV-2 (274 aa residues), forms dimer to create ion channels in the host cell membrane, participates in virus release, apoptosis, and pathogenesis	Ren <i>et al.</i> , 2020
ORF3d	Previously recognized as 3b, comprises 154 aa residues	Nelson <i>et al.</i> , 2020
ORF6	Has a significant role in viral pathogenesis, work together nsp8	Zhao <i>et al.</i> , 2009
ORF7a	A 122 aa type I integral protein	Nelson <i>et al.</i> , 2005
ORF7b	A 44 aa transmembrane protein localized in the Golgi apparatus	Schaecher <i>et al.</i> , 2007
ORF8	Shows low homology in SARS-CoV-2 and SARS-CoV, SARS-CoV contains two ORF8 proteins (ORF8a and ORF8b) while SARS-CoV-2 has a single ORF8 protein with 121 aa lengths, interact with major histocompatibility complex I (MHC-I), thereby may help in immune evasion	Zhang <i>et al.</i> , 2020c
ORF9b	It consists of 97 aa, tends to associate with an adaptor protein, TOM70, and suppresses type I interferon signaling mediated antiviral response.	Jiang <i>et al.</i> , 2020
ORF10	Its corresponding sgRNA is rarely detected, but ORF10 protein (38 aa) with unknown function has been found in infected cells.	Bojkova <i>et al.</i> , 2020
ORF14	Consist of 73 aa residues, its function is not clearly understood	Baruah and Bose, 2020

The life cycle of coronavirus

The contact between spike protein and ACE2 receptor triggers the first binding of the virus to the host cell (Fig. 2). In this interaction which is a key determinant in host rang determination of a coronavirus, as well as the virus tissue tropism, the RBD at the C-terminal of S1 subunit (in

some coronavirus such as MHV at the S1 N-terminal), is necessary. Following the receptor binding, the spike protein undergoes acid-dependent proteolytic cleavage at the S1/S2 cut site by a cathepsin, TMPRSS2, or another protease for priming the second cut at the S'2 site within the S2 subunit (Hoffmann *et al.*, 2020).

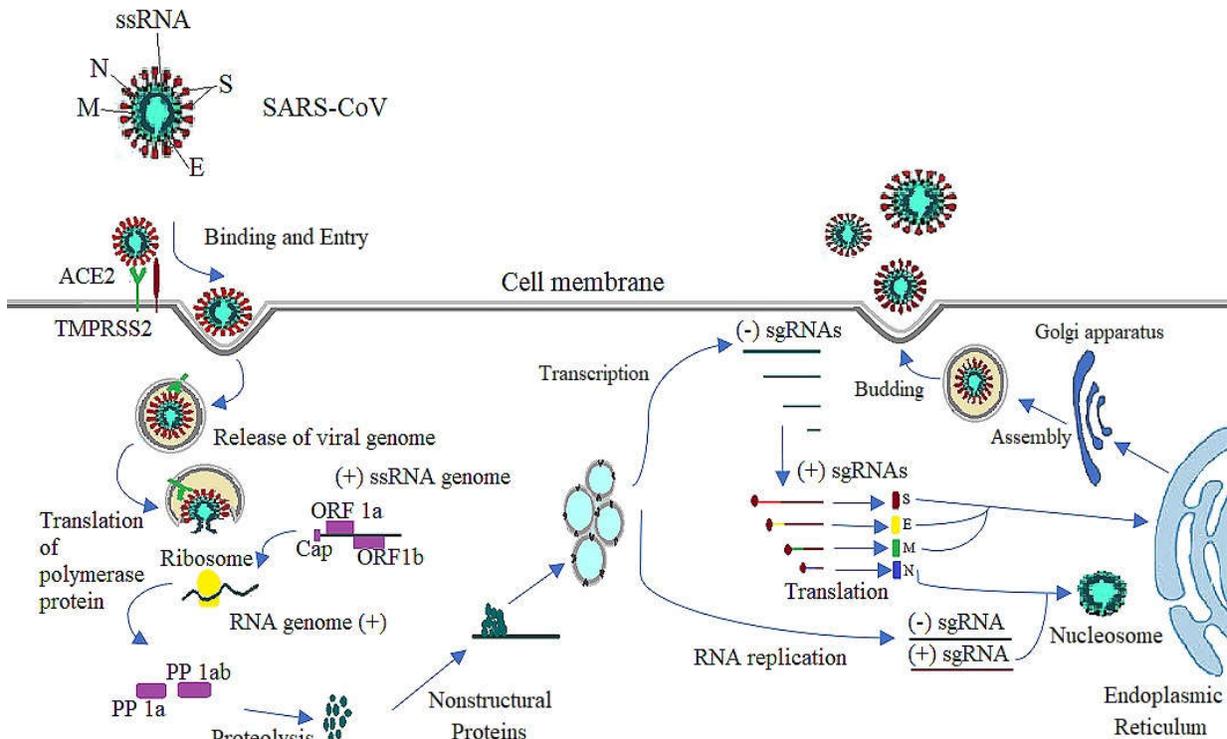


Fig 2. SARS-CoV-2 enters the host cell through the Spike protein (RBD/S1) attachment to the ACE2 receptor. In the cytosol, translation of ssRNA⁺ by host cell ribosome leads to the formation of the nonstructural polyproteins pp1a and pp1ab, which are proteolytically cleaved by a virus-encoded protease to form replicase complexes. The result of the complex activity is the formation of multiple gRNA copies and a set of sgRNAs. The sgRNA is translated into both structural and accessory proteins, which assemble into virus particles in the lumen of the ERGIC (Endoplasmic Reticulum Golgi Intermediate Compartment) and finally leave the contaminated cell by exocytosis.

The irreversible changes in the conformation of the spike protein are caused after this cleavage, allowing the virus envelope and cell membrane to fuse. Then the viral nucleocapsid is released into the cytosol. Single-cell RNA sequencing results show that TMPRSS2 is strongly expressed in various tissues and organs and is simultaneously expressed with ACE2 in bronchial branches, lung, and nasal epithelial cells, describing some of the SARS-CoV-2 tissue tropism (Lukassen *et al.*, 2020). The lysosome

immune response of the cell attempts to digest the endosome containing nucleocapsid by enzymatic activity, resulting in uncoating and releasing the viral genomic RNA (Tok and Tatar, 2017; Fallah *et al.*, 2021).

Immediately after the viral genome access to the host cytosol, the replicase genes are translated and two polyproteins named pp1a (440-500 kDa) and pp1ab (740-810 kDa) synthesized. Virus and host proteases cleave these polyproteins into individual nsps to produce the replicase-

transcriptase complex (RTC). ORF1a encodes two proteases: 1) the papain-like protease (PLpro), encoded within nsp3 and is responsible for the three proteolytic cleavages at nsp1-nsp4 boundaries, 2) the 3chymotrypsin-like protease (3CLpro), encoded by nsp5 and cleaved the polyprotein at 11 other sites, which results in nsp5-nsp16 (Qiu and Xu, 2020; Fallah *et al.*, 2021). Except for MERS-CoV, SARS-CoV, and

γ -coronaviruses, which express only one PLpro, most coronaviruses encode two PLpros at nsp3. Because these enzymes are essential for the reproduction, release, and transmission of new virus particles, they can be studied as possible targets for designing medical strategies. As shown in Table 2, the nsps of RTC have several catalytic activities that prepare an appropriate environment for RNA synthesis.

Table 2. Features and functions of nonstructural proteins of coronavirus.

Protein	Characteristics and Function	References
nsp1	By blocking protein synthesis through attachment to 40S ribosomal subunit and promoting cellular mRNA degradation, suppress the host's innate immune response to virus infection.	Huang <i>et al.</i> , 2011
nsp2	The exact function is unknown; it is binding to two host proteins, prohibitin1 and 2, which may disrupt the host cell signaling for virus infection.	Graham <i>et al.</i> , 2005
nsp3	The biggest protein encoded by the SAR-CoV genome contains TM1 and TM2 transmembrane regions and eight domains involving: ubiquitin-like (Ubl1) and Ubl2, Glu-rich, PL2pro, nsp3 ectodomain (3Ecto), macrodomain (X), CoV-Y, and Y1 domains. Its activities include: - Participation in the formation of the RTC complex - Interaction with N protein through Ubl1 and Ac domains - Promoting cytokine expression by ADP-ribose-1'-phosphatase (ADRP) activity - Cleaving viral polyprotein along with papain-like protease (PLpro) - Modifying host proteins, to aid viral survival and block the host's innate immune response by de-ubiquitination	Lei <i>et al.</i> , 2018
nsp4	A potential membrane-spanning protein of the scaffold, essential for the correct structure of the vesicles with a double membrane (along with NSP3 and NSP6); Moreover, its interaction with nsp3 is necessary for the replication of the virus.	Sakai <i>et al.</i> , 2017
nsp5	The main viral protease, named 3CLpro or main protease (Mpro), is located between two hydrophobic membranes and split viral polyprotein	Anand <i>et al.</i> , 2003
nsp6	A possible transmembrane protein of the scaffold, induce autophagosome formation	Oostra <i>et al.</i> , 2008
nsp7 and nsp8	Forms hexadecameric complex with nsp8, can operate as processivity clamp for RNA polymerase. Furthermore, in SARS-CoV, nsp8 can have primase activity.	Zhai <i>et al.</i> , 2005
nsp9	ssDNA/RNA-binding protein serves as a crucial component that interacts with other proteins in the RTC complex	Miknis <i>et al.</i> , 2009
nsp10	Participation in the synthesis of viral RNA and polyprotein processing; Interaction with NSP5 protease; Forming heterodimer with nsp14 and nsp16 to stimulate and enhance their 3'→5' exoribonuclease (ExoN) and RNA Cap 2'-O-methyltransferase operations, respectively.	Bouvet <i>et al.</i> , 2014
nsp11	It consists of 13 aa, its function and structural dynamics are unknown.	Yoshimoto, 2020
nsp12	RNA-dependent RNA polymerase that builds a complex with an nsp7-nsp8 heterodimer and an nsp8 monomer to enhance the processivity	Subissi <i>et al.</i> , 2014
nsp13	Possess RNA helicase (stimulated in the presence of nsp12) and 5' triphosphatase activities	Ivanov <i>et al.</i> , 2004a; Ivanov and Ziebuhr, 2004b
nsp14	A bifunctional protein with ExoN activity in the N terminal domain and guanine-N7-methyltransferase activity in the C-terminal domain. ExoN action is necessary for proofreading, repair, and recombination of the coronavirus genome.	Chen <i>et al.</i> , 2009
nsp15	A viral Mn ²⁺ -dependent endoribonuclease also named U viral endoribonuclease (NendoU)	Deng <i>et al.</i> , 2017
nsp16	Or 2'-O-Ribose-methyltransferase protects RNA virus from identification by the melanoma differentiation-associated protein 5 (MDA5), active only in the presence of nsp10	Zust <i>et al.</i> , 2011

Intriguingly, the activities of the nsp15-NendoU and nsp14-ExoN ribonucleases are exclusive to the Nidovirales order and are considered as the genetic markers of these viruses (Snijder *et al.*, 2003).

The formation and function of RTC lead to the synthesis of multiple copies of both sub-genomic and genomic RNAs via discontinuous transcription. Sub-genomic RNAs (sgRNAs) act as mRNA and attach to the host cell ribosomes to translate various accessory and structural proteins. The production of both sub-genomic and genomic RNAs is mediated by negative-strand RNAs that are only approximately 1% as much as their positive-sense equivalents (Brown *et al.*, 2007).

Between structural proteins, only N protein is translated by free cytosolic ribosomes of host cells. Accessory and membrane-associated structural (S, M, and E) proteins synthesized by endoplasmic reticulum-bound ribosomes. These proteins converge in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC), where M protein acts as a scaffold and conducts most of the protein-protein interactions essential for virion formation. For this purpose, protein M requires the accompaniment and participation of protein E, as well as interaction with N protein (V’Kovski *et al.*, 2021). Finally, the encapsidated viral genomes bud on the ERGIC membrane to form mature virions. Then, the progeny viral particles are transferred to the cell surface via secretory vesicles and fuse with the plasma membrane, where they are exocytosed, secreted, and spread to other parts of the body.

Pathogenesis

The function of the nsps and structural proteins is linked to CoV pathogenesis. During the SARS-CoV-2 infection, S and N proteins are the two most widely expressed immunogenic proteins. In addition, scientists have described the nsps function in suppressing the responses of the host’s innate immune system. The increased numbers of leukocyte, respiratory abnormalities, and higher plasma pro-inflammatory cytokines levels (including IL2, IL7, IL10, IP10, TNF α , GCSF, MCP1, and MIP1 α) are some pathological symptoms of COVID-19 infected patients (Huang *et al.*, 2020).

Besides a broad host range, SARS-CoVs exhibit a diverse tissue tropism. In addition to the lung epithelial cell infection, other predominant pathogenesis of SARS-CoV-2 is severe pneumonia, acute cardiac injury, the formation of ground-glass opacities, and detectable serum SARS-CoV-2 RNA (RNAemia) (Huang *et al.*, 2020). However, the precise mechanism of SARS-CoV-2 lung damage and the etiology of significant illness in people has yet to be discovered. Although the invasion to dendritic cells and macrophages only results in an unsuccessful infection, these cell infections could be essential to activate the pro-inflammatory cytokines, which may cause disease (Peiris *et al.*, 2003).

SARS-CoV-2 variants

Such as other ssRNA viruses, coronaviruses exhibit a high mutation rate of up to 10^{-4} substitutions per nucleotide per cell replication (Su *et al.*, 2016). Although most changes in the RNA genome of SARS-CoV-2 are probably to be neutral or lead to the extinction of the virus, a small fraction can significantly affect functional characteristics of the virus and changes the tissue tropism, interactions with host immunity, disease severity, transmissibility, or vaccines performance (Harvey *et al.*, 2021). The most frequent mutations found in SARS-CoV-2 variants compared to the original form observed in Wuhan were shown in Fig. 3.

The first whole-genome sequences of SARS-CoV-2 were provided in the Global Initiative on Sharing Avian Influenza Data (GISAID) in early January 2020. As of September 2021, more than 3.3 million SARS-CoV-2 genome sequences from 172 different countries have now been shared on the GISAID database. In between, there are several variants different from the Wuhan MN908947 original isolate. The study of SARS-CoV-2 genetic diversity is important for understanding the pathophysiology, illness course, as well as choosing appropriate prevention and treatment strategies for COVID-19. Reduction of virus transmissibility using confirmed methods for disease control and preventing entries into animal populations are key features of the worldwide strategy to decrease the incidence of mutations that hurt public health. Furthermore, knowing the

properties of antigenic proteins and predicting their possible mutations, can play a major role in the design of effective vaccines against the SARS-CoV-2 variants.

The United States Department of Health and Human Services (HHS) determines three different classes of SARS-CoV-2 variants: 1) Variant of Interest (VOI) that have particular genetic markers linked to alterations in viral binding to the host cell receptor, diminished neutralization by antibodies produced in response to past infection or vaccination, lower therapeutic effectiveness, or enhanced disease transmissibility and severity. 2) Variant of Concern (VOC) is a variant with a significant reduction in the neutralizing effect of antibodies produced during previous infections or

vaccination, more transmissibility, and disease severity decreased efficacy of treatments or vaccines. 3) Variant of High Consequence (VOHC) is associated with diagnostic failure, causes greater numbers of hospitalizations, more severe disease, significantly reduced vaccine efficacy compared to two other classes of variants. At present, all identified variants are classified in the first two classes, and no SARS-CoV-2 variants have reached the level of very hazardous. Of course, this classification is not permanent. Compared with other circulating SARS-CoV-2 variants, an earlier assigned VOI or VOC, which has decisively shown to no longer pose a significantly increased risk to worldwide public health can be reclassified.

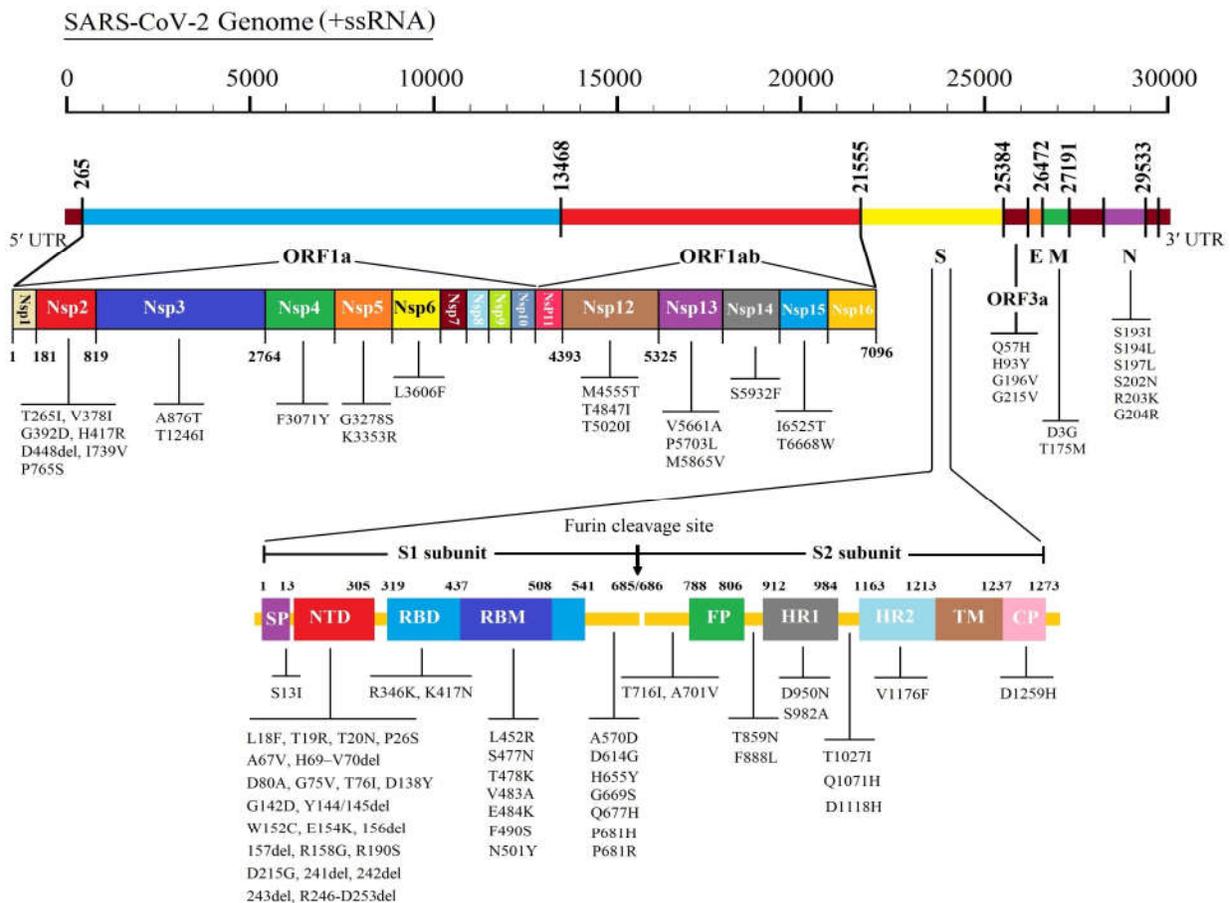


Fig 3. The most frequent mutations found in SARS-CoV-2 variants compared to Wuhan MN908947 original isolate. Most deletion mutations were observed in regions encoding proteins that interact with host responses and in spike proteins.

Along with scientific terminology, and as of May 31, 2021, World Health Organization (WHO) suggested names for worldwide SARS-CoV-2 variants of interest and concern using the Greek alphabet. The variant name is determined by at least one Pango lineage, as well as any further distinctive S protein alterations. If the variation is difficult to define using this terminology, an alternate description may be employed. SARS-CoV-2 variants and mutations associated with them are summarized in Table 3. The First Variant of Concern for SARS-CoV-2 (VOC-202012/01) was found in December 2020 in the southeast of the United Kingdom (UK) and designated Alpha lineage by WHO. This variant, also called B.1.1.7 is 43-82 percent more transmissible than earlier identified variants (Davies *et al.*, 2021a). The B.1.1.7 lineage is associated with 17 mutations throughout the genome, including three deletions and 14 nonsynonymous substitutions, eight of which are found in the spike protein (two deletions and six substitutions). The N501Y mutation is an important contact residue in the RBD that affects its conformation and increases viral binding affinity to human ACE2 (Starr *et al.*, 2020). According to some studies, this mutation may reduce neutralization by a small fraction of RBD antibodies (Collier *et al.*, 2021). P681H another important mutation is located precisely near the furin cut site in spike, which has a central role in transmission and infection (Hoffmann *et al.*, 2020). The deletion of H69/V70 in the SARS-CoV-2 spike protein is associated with immunological escape and increases viral infectivity *in vitro* (Kemp *et al.*, 2021). Generally, these mutations cause a 1.5-fold decrease in neutralization activity of antibodies toward the B.1.1.7 variant in comparison with previous lineages. Beta (B.1.351) and Gamma (P.1) lineages are the two other most notable SARS-CoV-2 VOC, with several genetic mutations same to B.1.1.7 variant that has arisen independently in these lineages. The B.1.351 or 501Y.V2 lineage was first discovered in South Africa in December 2020. The P.1 lineage (also known as 501Y.V3) was found in Brazil in December 2020. In addition to N501Y mutation, E484K is a usual substitution within the Alpha, Beta, and Gamma lineages. Furthermore, E484K is likely to have originated many times in

various SARS-CoV-2 populations across the world, such as P.2 and P.3 variants (Wise, 2021), which are spread in the Brazilian state of Rio de Janeiro and the Philippines, respectively. Some studies have shown that Glutamic acid (E) 484 plays as an immunodominant residue in spike protein, and its substitution facilitates escape from various monoclonal antibodies, including antibodies in convalescent plasma (Greaney *et al.*, 2021). The beta variant exhibited more resistance to convalescent plasma and monoclonal antibodies than the Alpha variant (Wang *et al.*, 2021; Wibmer *et al.*, 2021) found that 93% of the 44 plasma samples have a decrease in the neutralizing antibody titers, and 48% (21 of 44) showed no detectable neutralization activity. Furthermore, Gamma neutralization of convalescent plasma from individuals infected with non-Gamma variants had a 6-fold reduction compared to non-Gamma isolates. In both P.1 and B.1.351 lineages, the lysine 417 (K417) spike receptor-binding domain (RBD) has changed. It seems that substitution of this amino acid in the B.1.351 variant with Asparagine (N) and the P.1 variant with Threonine (T) also helps the virus bind more tightly to human ACE2 (Funk *et al.*, 2021). The B.1.617 lineage is a new one with rapid expansion that was discovered in India in late 2020. The variant has evolved into two distinct lineages, B.1.617.2 (Delta) and B.1.617.1 (Kappa), because of differences in spike protein (Hoffmann *et al.*, 2021).

The B.1.617 lineage was first classified as a variant of interest, when additional information about the transmissibility and possibility for antibody neutralization became available, this classification was changed and B.1.617.2 was defined as a variant of concern (Hoffmann *et al.*, 2021). It is known that this lineage has eight substitution mutations in the spike protein, among them T19R, L452R, T478K, P681R, and D950N are the genetic markers for the Delta variant.

Studies have shown that L452R substitution in the RBD and P681R that impact antibody binding and antibody-mediated neutralization (Hoffmann *et al.*, 2021). Furthermore, many Delta sequences have two deletions in ORF8 at positions D119 and F120. At present, the Delta variant that causes more infections and

contagious than the previous form of the SARS-CoV-2 is the dominant variant in most parts of the world. In people who have not been vaccinated, it can cause more serious diseases than previous strains.

The information about SARS-CoV-2 variants of interest is summarized in Table 3. The WHO is constantly reviewing the global epidemiology of VOIs and monitoring their global spread. Therefore, this information is continuously being updated and some of these variants may be classified as variants of concern in the future. Among VOIs, the Lambda variant (C.37) has the greatest potential for being included in the VOCs list. This variant appeared first in Peru and was classified as a variant of interest on 15 June 2021 at the world level by the WHO. The S protein of the Lambda carries a 7-amino-acid deletion in the N-terminal domain (RSYLTPGD246-253N) and six substitution mutations (G75V, T76I, L452Q, F490S, D614G, and T859N). The T76I and L452Q mutations appear to make the Lambda variant spike protein more infectious. The RSYLTPGD246253N mutation enhances the ability to evade the host's immune response and is responsible for reducing antibody neutralization (a remarkable phenotype of most VOCs) (Kimura *et al.*, 2021). The B.1.621 lineage, also known as the Mu variant, is another VOI first detected in January 2021 in Colombia. Based on the WHO's weekly bulletin on the COVID-19 pandemic (WHO, 2021) the Mu variant has a series of mutations that allow it to evade the host's immune response. Some of these mutations, such as T95I, E484K, N501Y, and D950N that are found in S protein are also seen in the other VOIs and VOCs. Public Health England (PHE) reports that this variant is at least as resistant to vaccination as the Beta variant. However, more research and real global cases are needed to confirm the characteristics of the Mu variants (Bernal *et al.*, 2021). Together with the variants detailed in Table 3, several other variants are under monitoring. There is evidence that they may have similar characteristics to VOCs, but the evidence is weak or has not been evaluated by the researchers. In general, understanding the virulence, infectivity, antigenicity, and transmissibility of different

variants depends on the study of their particular mutations, either alone or together with other mutations.

Vaccines for COVID-19

Although some treatments are effective for some patients, effective and proven universal treatments for COVID-19 or antiviral drugs for SARS-CoV-2 have not yet been developed. However, researchers and pharmaceutical manufacturers continue to conduct large-scale clinical trials to evaluate different treatments and vaccines against COVID19. Unlike bacterial infections, vaccines are more effective than drugs for viral infections. Therefore, access to efficient and safe vaccines is essential to ending the COVID-19 pandemic. Generally, vaccines come in a variety of forms, killed or live-attenuated viruses, DNA or RNA, toxoid, soluble proteins (S glycoproteins), and recombinant viral vectors. The various aspects of worldwide approved vaccines are listed in Supplement 1. Inactivated or killed vaccine contains virus particles that have been cultured in monkey kidney epithelial cells and grow in bioreactor tanks and subsequently destroyed by soaking with beta-propiolactone to disable their replication properties and thus, ability to cause disease. These particles that have intact S proteins are then blended with an aluminum-based adjuvant to enhance the immune response against the native viral antigens expressed by inactivated vaccine.

An attenuated vaccine consists of a live virus with reduced virulence in the host but keeps its ability to stimulate an effective immune response. These vaccines differ from those that are made by "killing" the virus (Minor, 2015). Viral vector vaccines are used to deliver genetic material encoding the target antigen to the host cell of the recipient. Due to the high transduction efficiency, transgene expression, broad tissue tropism, and ability to infiltrate proliferating and non-proliferating cells, adenoviral vectors are good candidates for preparing vaccines against Covid19 (Vrba *et al.*, 2020; Fallah *et al.*, 2021). Currently, several approved vaccines for Covid-19 are based on replicating and nonreplicating human adenovirus vectors (Fallah *et al.*, 2021).

Table 3. SARS-CoV-2 variants of interest and concern.

PANGO ¹ Lineage	WHO label	Spike protein mutations	Country and date first detected	Pathogenicity	References
B.1.1.7	Alpha	H69-V70del, Y144/145del, E484K, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	United Kingdom, September 2020	Impact on transmissibility (About 50% increase) and severity, No impact on immunity and susceptibility to EUA ² monoclonal antibody treatments	Davies <i>et al.</i> , 2021a; Davies <i>et al.</i> , 2021b; Collier <i>et al.</i> , 2021; Funk <i>et al.</i> , 2021
B.1.351	Beta	D80A, D215G, 241del, 242del, 243del, K417N, E484K, N501Y, D614G, A701V	South Africa, September 2020	Impact on transmissibility (About 50% increase) and severity, Impact on immunity (reduced serum neutralization during recovery and after vaccination)	Funk <i>et al.</i> , 2021
P.1	Gamma	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I	Brazil, December 2020	Impact on transmissibility and severity, Impact on immunity (reduced serum neutralization during recovery and after vaccination)	Funk <i>et al.</i> , 2021
B.1.617.2	Delta	T19R, (G142D), 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N	India, December 2020	Impact on transmissibility, impact on immunity (escape)	Public Health England, 2021
B.1.525	Eta	A67V, 69del, 70del, 144del, E484K, D614G, Q677H, F888L	Nigeria, December 2020	Impact on immunity (reduced neutralization)	Jangra <i>et al.</i> , 2021
B.1.427/ B.1.429	Epsilon	S13I, W152C, L452R, D614G	USA, September 2020	Impact on transmissibility (18.6-24% increase), Impact on immunity (reduced neutralization)	McCallum <i>et al.</i> , 2021
P.2	Zeta	E484K, (F565L ³), D614G, V1176F	Brazil, April 2020	Impact on immunity (reduced neutralization)	Jangra <i>et al.</i> , 2021
P.3	Theta	E484K, N501Y, D614G, P681H	Philippines, January 2021	Impact on transmissibility and immunity (reduced neutralization)	Davies <i>et al.</i> , 2021b; Jangra <i>et al.</i> , 2021
B.1.616		V483A, D614G, H655Y, G669S	France, February 2021		Santé Publique France, 2021
B.1.617.1	Kappa	G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H	India, December 2020	Impact on transmissibility and immunity (reduced neutralization)	Bernal <i>et al.</i> , 2021; Public Health England, 2021
C.37	Lambda	R246-D253del, G75V, T76I, L452Q, F490S, D614G, T859N	Peru, December 2020	Impact on transmissibility and severity, Impact on immunity (reduced neutralization by sera after vaccination)	Kimura <i>et al.</i> , 2021
B.1.620		S477N, E484K, D614G, P681H	Unclear, February 2021	Impact on immunity (reduced neutralization)	Public Health England, 2021; Bernal <i>et al.</i> , 2021
B.1.621	Mu	R346K, E484K, N501Y, D614G, P681H	Colombia, January 2021	Impact on transmissibility and immunity (reduced neutralization)	Bernal <i>et al.</i> , 2021

¹ Phylogenetic Assignment of Named Global Outbreak; ² Emergency Use Authorization; ³ Detected in some sequences but not all; * Alpha, Beta, Gamma, and Delta are variants of concern, other are variants of interest; ** Adapted from European Centre for Disease Prevention and Control (2021) (Available from: <https://www.ecdc.europa.eu/en>).

The human Ad5 and Ad26 adenoviruses as well as a changed form of the chimpanzee adenovirus ChAdOx1 are currently the most widely used attenuated or replication-deficient viral vectors used to encode the S-protein antigen of SARS-CoV-2 in host cells. In the host cells, non-replicating vectors release the genetic material that enter the nucleus, where the DNA is expressed and not integrated into the genome. The transcribed mRNAs exit the nucleus and are translated into spike proteins. The S proteins assembled on the surface of the infected cells are recognized by components of the immune system, which trigger the creation of specific neutralizing antibodies as well as activation of T cells and memory cells types.

The messenger ribonucleic acid (mRNA) vaccine that used for the first time in humans, is a new technology that contains non-replicating or self-replicating mRNA to trigger an immune response. mRNA molecules are not stable and to prevent degradation inside the cells, the mRNA vaccines are encapsulated by lipid nanoparticles (LNP). In the host cells, the manufactured nucleoside-modified mRNA encodes a SARS-CoV-2 specific antigen, such as S protein, that can activate the immune system without risking the serious consequences of getting sick with COVID-19 (Anand and Stahel, 2021). This technology allows the human cells to generate antigenic proteins instead of receiving them in vaccines.

As a result, the required time for vaccine production and activation of immune responses is reduced compared to that required time for classical vaccines. Given that mRNA molecules are usually degraded by cellular ribonucleases after 48 hours, it is almost impossible for the synthetic mRNA to enter the cell nucleus and integrate with the genome of the vaccinated cells. For this reason, so far there has been no concern about the mutagenic effects of these vaccines. Recombinant protein subunit vaccines are the fourth type of vaccines against SARS-CoV-2 containing whole or fragments of viral protein, and genetic materials are not used in their production.

Since the peptide subunits used for the preparation of these vaccines are inadequately immunogenic, they need adjuvants and a booster dose (Tan *et al.*, 2021). In addition, researchers

are manufacturing the S, M, and E with or without N proteins of SARS-CoV-2 to create virus-like particle (VLP) vaccines that are comparable with viruses but missed the ssRNA+ genome. The S protein of VLPs can bind and enter the human cells through the ACE2 receptors.

Conclusions

The SARS-CoV-2, outbreak in 2019, is a new strain of coronavirus that has not been recognized among humans previously and its rapid worldwide spread has become one of the greatest global health challenges.

As of September 2021, nearly 222.4 million confirmed cases of COVID-19 and more than 4.59 million deaths have been reported worldwide (worldometers.info/coronavirus; accessed on 5th July 2021). However, this virus not only caused many deaths around the world but also has a significant effect on individuals, societies, and the entire country's economic system. The origin of the SARS-CoV-2 is still debatable and there are some theories in this regard. Based on the most likely theory, SARS-CoV-2 originated from coronaviruses that infect bats. The main evidence for this theory is the overall similarity (96%) between the two coronaviruses' genomes. However, the significant differences between the type and arrangement of RBD mutations of SARS-CoV-2 and bat CoVs, which stop the effective binding of the bat CoVs spike protein to human ACE2 (Wan *et al.*, 2020), has caused some challenges for the defenders of this theory. Another theory is that SARS-CoV-2 originated from pangolin coronaviruses. There is a less overall similarity among the genomes of pangolin coronaviruses and SARS-CoV-2 than bat CoVs. However, the significant similarity between the RBD mutations of the two viruses, including the six key residues, which optimized them for binding to human ACE receptors, supports the origin of SARS-CoV-2 from Pangolin coronaviruses (Zhang *et al.*, 2020a). Although, both bat and pangolin CoVs do not have polybasic cleavage sites, however, the study of mutations in this region, as well as spike proteins between these animal viruses and SARS-CoV-2, suggests that this region could evolve through natural selection. Another theory that is being

considered these days is the emergence of SARS-CoV-2 using laboratory manipulation of coronaviruses similar to SARS-CoV. Furthermore, it is possible that SARS-CoV-2 mutated and evolved from bat or other coronaviruses through passaging in animal models or cell cultures in laboratories around the world (Ge *et al.*, 2013). Despite the pros and cons for each theory, now it is impossible to prove or disprove them. SARS-CoV-2 has several differences and similarities with other human coronaviruses that are known to cause previous outbreaks. Being aware of them can help the researcher better understand the causes of the current pandemic and design treatment strategies. Although both SARS-CoV and SARS-CoV-2 interact with the same host cell receptors (ACE2), stronger binding of the SARS-CoV-2 to the receptors due to structural differences in its surface proteins enable this virus more effective in invading the human cells. The distribution of ACE2 receptors in various tissues can explain the site of the infection and the patient's symptoms. In addition, the higher affinity of SARS-CoV-2 for the upper respiratory tract and conjunctiva facilitates infection of the upper respiratory tract and airways conduction. Extensive studies on the structure and genomic organization as well as the basic mechanisms of host cell entry, life cycle, and pathophysiological properties of SARS-CoV-2, have helped the researchers to develop and approve several vaccines for emergency use against the virus. However, the emergence of novel variants of SARS-CoV-2 as a result of higher rates of mutation and recombination has made complete control of the disease difficult and slow. Based on the WHO classification there are currently four SARS-CoV-2 variants of concern that have considerable characteristics such as increased transmissibility and severity of the disease, as well as the ability to diminish the efficacy of the vaccines. Alpha lineage is the first SARS-CoV-2 variant of concern identified in the UK in December 2020. Multiple mutations in the S protein of this variant make it more contagious compared to previously circulating SARS-CoVs. BNT162b2 (Pfizer-BioNTech) and AZD1222 (AstraZeneca/Oxford) are the most effective vaccines against the Alpha variant, 93.4% and 66.1%, respectively (Sharma *et al.*,

2021; Fallah *et al.*, 2021). The Beta variant was first detected in South Africa and contains eight characteristic mutations in the S protein. These mutations increase transmissibility (About 50%) and severity of the virus and can impact host immunity by reduction of serum neutralization during recovery and after vaccination. This variant can reduce neutralizing antibodies after vaccination with AZD1222, mRNA-1273 (Moderna), Sputnik V, Ad26.COV2. S (Johnson & Johnson), and Pfizer-BioNTech vaccines 9-, 6.4-, 6-, 5-, and 4.9-fold compared to wild-type virus, respectively (Sharma *et al.*, 2021). The Gamma variant originated from Brazil and has 20 unique mutations (seven of them are in the spike protein), some of which could be responsible for an escape of the antibodies. This variant reduced neutralizing antibody by 3.8-, 4.8-, 2.9-, 3.3-, and 3.9-fold compared to Wuhan early isolate after vaccination with BNT162b2, mRNA-1273, AZD1222, Ad26.COV2. S, and CoronaVac (Sinovac), respectively (Sharma *et al.*, 2021). The Delta variant, discovered in India in December 2020, is currently the most common lineage of SARS-CoV-2 in the world. This variant has multiple S protein mutations allowing it to increase transmissibility and decrease in neutralizing antibody and post-vaccination sera. So far, only the effectiveness of BNT162b2 and AZD1222 vaccines against the Delta variant has been reported (Sharma *et al.*, 2021). Generally, it seems that more clinical studies are required to specify the efficacy of different vaccines against different variants of the virus.

In conclusion, a better knowledge of the SARS-CoV-2 characteristics is crucial for the development of therapies, vaccines, and supportive care with COVID-19 treatment. More data is required to know the determinants of health and dysfunction responses, disease severity, and protective immune markers.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

- Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. 2003. Coronavirus main proteinase (3CLpro) Structure: basis for

- design of anti-SARS drugs. *Sci* 300(5626): 1763-1767.
- Anand P, Stahel VP. 2021. Review the safety of Covid-19 mRNA vaccines: a review. *Patient Saf* 15(1): 20. doi: 10.1186/s13037-021-00291-9.
- Aoe T. 2020. Pathological aspects of COVID-19 as a conformational disease and the use of pharmacological chaperones as a potential therapeutic strategy. *Front Pharmacol* 11: 1095. doi: 10.3389/fphar.2020.01095.
- Araki K, Gangappa S, Dillehay DL, Rouse BT, Larsen CP, Ahmed R. 2010. Pathogenic virus-specific T cells cause disease during treatment with the calcineurin inhibitor FK506: implications for transplantation. *Exp Med* 207(11): 2355-2367.
- Bernal JL, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, Ramsay M. 2021. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *N Engl J Med* <https://doi.org/10.1101/2021.05.22.21257658>.
- Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, Münch C. 2020. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583(7816): 469-472.
- Brown CG, Nixon KS, Senanayake SD, Nixon KS, Brian DA. 2007. An RNA stem-loop within the bovine coronavirus nsp1 coding region is a cis-acting element in defective interfering RNA replication. *J Virol* 81(14): 7716-7724.
- Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, Yuen KY. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* 9(1): 221-236.
- Chen Y, Cai H, Pan J, Xiang N, Tien P, Ahola T, Guo D. 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci* 106(9): 3484-3489.
- Collier DA, De Marco A, Ferreira IATM, Meng B, Datir RP, Walls AC, Gupta RK. 2021. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* 593(7857): 136-141.
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Edmunds WJ. 2021a. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372(6538): eabg3055. doi: 10.1126/science.abg3055
- Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH, Pearson CAB. 2021b. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 593(7858): 270-274.
- de Haan CA, de Wit M, Kuo L, Montalto-Morrison C, Haagmans BL, Weiss SR, Rottier PJ. 2003. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus *in vitro* and its ability to replicate in the liver but not the brain. *Virology* 312(2): 395-406.
- Deng X, Hackbart M, Mettelman RC, O'Brien A, Mielech AM, Yi G, Baker SC. 2017. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci* 114(21): 4251-4260.
- Fallah A, Razavi Nikoo H, Abbasi H, Mohammad-Hasani A, Hosseinzadeh Colagar A, Khosravi A. 2021. The features of pathobiology and clinical translation of approved treatments for coronavirus disease 2019 (COVID-19). *Intervirology* (In press): <https://doi.org/10.1159/000520234>.
- Fehr AR, Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* 1282: 1-23.
- Funk T, Pharris A, Spiteri G, Bundle N, Melidou A, Carr M, Adlhoch C. 2021. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1. data from seven EU/EEA countries, weeks 38/2020 to 10/2021 *Eurosurveillance* 26(16): 2100348.
- Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Shi ZL. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503(7477): 535-538.
- Graham RL, Sims AC, Brockway SM, Baric RS, Denison MR. 2005. The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J Virol* 79(21): 13399-13411.

- Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29(3): 463-476.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Robertson DL. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19(7): 409-424.
- Hoffmann M, Hofmann-Winkler H, Krüger N, Kempf A, Nehlmeier I, Graichen L, Pöhlmann S. 2021. SARS-CoV-2 variant B.1.617 is resistant to Bamlanivimab and evades antibodies induced by infection and vaccination. *Cell Rep* 109415. <https://doi.org/10.1016/j.celrep.2021.109415>.
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Pöhlmann S. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181(2): 271-280.
- Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. 2011. Alphacoronavirus transmissible gastroenteritis virus nsp1 protein suppresses protein translation in mammalian cells and in cell-free HeLa cell extracts but not in rabbit reticulocyte lysate. *J Virol* 85(1): 638-643.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cao B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223): 497-506.
- Ivanov KA, Thiel V, Dobbe JC, Van Der Meer Y, Snijder EJ, Ziebuhr J. 2004a. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* 78(11): 5619-5632.
- Ivanov KA, Ziebuhr J. 2004b. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J Virol* 78(14): 7833-7838.
- Jacobs L, van der Zeijst BA, Horzinek MC. 1986. Characterization and translation of transmissible gastroenteritis virus mRNAs. *J Virol* 57(3):1010-1015.
- Jangra S, Ye C, Rathnasinghe R, Stadlbauer D; Personalized Virology Initiative study group, Krammer F, Simon V. 2021. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe* 2(7): e283-e284.
- Jiang HW, Zhang HN, Meng QF, Xie J, Li Y, Chen H, Tao SC. 2020. SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol* 17(9): 998-1000.
- Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Gupta RK. 2021. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592(7853): 277-282.
- Kimura I, Kosugi Y, Wu J, Yamasoba D, Butlertanaka PE, Tanaka LY, Sato K. 2021. SARS-CoV-2 Lambda variant exhibits higher infectivity and immune resistance. *BioRxiv* [Preprint]. 07.28.454085. <https://doi.org/10.1101/2021.07.28.454085>.
- Krafcikova P, Silhan J, Nencka R, Boura E. 2020. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun* 11(1): 3717. doi:10.1038/s41467-020-17495-9.
- Lei J, Kusov Y, Hilgenfeld R. 2018. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir Res* 149: 58-74.
- Liu Q, Johnson RF, Leibowitz JL. 2001. Secondary structural elements within the 3' untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J Virol* 75(24): 12105-12113.
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Tan W. 2020. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395(10224): 565-574.
- Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Eils R. 2020. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* 39(10): e105114. doi: 10.15252/embj.20105114.
- McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, Veessler D. 2021. SARS-CoV-2 immune evasion by variant

- B.1.427/B.1.429. *BioRxiv* [Preprint]. <https://doi.org/10.1101/2021.03.31.437925>.
- Miknis ZJ, Donaldson EF, Umland TC, Rimmer RA, Baric RS, Schultz LW. 2009. Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol* 83(7): 3007-3018.
- Minor PD. 2015. Live attenuated vaccines: Historical successes and current challenges. *Virology* 479: 379-392.
- Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. 2020. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. *N Engl J Med* 382(8): 692-694.
- Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH. 2005. Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* 13(1): 75-85.
- Nelson CW, Arden Z, Goldberg TL, Meng C, Kuo CH, Ludwig C, Wei X. 2020. Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* 9: e59633. doi: 10.7554/eLife.59633.
- Oostra M, Hagemeyer MC, van Gent M, Bekker CP, Te Lintelo EG, Rottier PJ, de Haan CA. 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol* 82(24): 12392-12405.
- Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IFN, Poon LL, HKU/UCH SARS study group. 2003. clinical progression and viral load in a community outbreak of coronavirus-associated sars pneumonia: a prospective study. *Lancet* 361: 1767-1772.
- Perlman S, Netland J. 2009. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7(6): 439-450.
- Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England Technical briefing 17 2021 [Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/997418/Variants_of_Concern_VOC_Technical_Briefing_17.pdf]
- Qiu Y, Xu K. 2020. Functional studies of the coronavirus nonstructural proteins. *STEMed* 1(2): e39. doi: 10.37175/stemedicine.v1i2.39.
- Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Zhou X. 2020. The orf3a protein of sars-cov-2 induces apoptosis in cells. *Cell Mol Immunol* 17(8): 881-883.
- Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology* 510: 165-174.
- Schaefer SR, Mackenzie JM, Pekosz A. 2007. The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. *J Virol* 81(2): 718-731.
- Schoeman D, Fielding BC. 2019. Coronavirus envelope protein: Current knowledge. *Viol J* 16(1): 1-22.
- Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020b. Cell entry mechanisms of sars-cov-2. *Proc Natl Acad Sci* 117(21): 11727-11734.
- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Li F. 2020a. Structural basis of receptor recognition by sars-cov-2. *Nature* 581(7807): 221-224.
- Sharma K, Koirala A, Nicolopoulos K, Chiu C, Wood N, Britton PN. 2021, Vaccines for covid-19: where do we stand in 2021? *Paediatr Respir Rev* Mini-symposium: COVID 19: The second year. <https://doi.org/10.1016/j.prrv.2021.07.00>.
- Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, Bu Z. 2020. Susceptibility of ferrets, cats, dogs, and other domesticated animals to sars-coronavirus 2. *Science* 368(6494): 1016-1020.
- Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Gorbalenya AE. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 331(5): 991-1004.
- Song W, Gui M, Wang X, Xiang Y. 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* 14(8):

- e1007236. doi: 10.1371/journal.ppat.1007236.
- Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, Qin C. 2019. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11(1): 59. doi: 10.3390/v11010059.
- Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Bloom JD. 2020. Deep Mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182(5): 1295-1310.
- Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Gao GF. 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 24(6): 490-502.
- Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Imbert I. 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci* 111(37): 3900-3909.
- Tan HX, Juno JA, Lee WS, Barber-Axthelm I, Kelly HG, Wragg KM, Wheatley AK. 2021. Immunogenicity of prime-boost protein subunit vaccine strategies against SARS-CoV-2 in mice and macaques. *Nat Commun* 12: 1-10.
- Tok TT, Tatar G. 2017. Structures and functions of coronavirus proteins: molecular modeling of viral nucleoprotein. *Int. J Virol Infect Dis* 2(1): 001-007.
- V'Kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. 2021. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol* 19(3): 155-170.
- Vrba SM, Kirk NM, Brisse ME, Liang Y, Ly H. 2020. Development and applications of viral vectored vaccines to combat zoonotic and emerging public health threats. *Vaccines* 8(4): 680. doi: 10.3390/vaccines8040680.
- Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Dempsey DM, Dutilh BE, Davison AJ. 2019. Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the group of viruses (2019). *Arch Virol* 164(9): 2417-2429.
- Wan Y, Shang J, Graham R, Baric RS, Li F. 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol* 94(7): e00127-20.
- Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Ho DD. 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593(7857):130-135.
- Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Moore PL. 2021. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* 27(4):622-625.
- Wise J. 2021. Covid-19: the E484K mutation and the risks it poses. *BMJ* 372: n359.
- Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Yuen KY. 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus Deltacoronavirus supports bat coronaviruses as the gene source of Alphacoronavirus and Betacoronavirus and avian coronaviruses as the gene source of Gammacoronavirus and Deltacoronavirus. *J Virol* 86(7): 3995- 4008.
- World Health Organization. 2021. COVID-19 weekly epidemiological update edition 55, published 31 August 2021 [Available from: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---31-august-2021>].
- Yoshimoto FK. 2020. The Proteins of Severe Acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J* 39(3): 198-216.
- Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, Rao Z. 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* 12(11): 980- 986.
- Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Hilgenfeld R. 2020b. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* 368(6489): 409- 412.
- Zhang T, Wu Q, Zhang Z. 2020a. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol* 30(7): 1346-1351.

- Zhang, Y, Zhang J, Chen Y, Luo B, Yuan Y, Huang F, Zhang H. 2020c. The ORF8 protein of SARS-CoV-2 mediates immune evasion through potently downregulating MHC-I. *Proc Natl Acad Sci* 118(23): e2024202118.
- Zhao J, Falcon A, Zhou H, Netlan J, Enjuanes L, Brena PP, Perlman S. 2009. Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. *J Virol* 83(5): 2368-2373.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798): 270-273.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Tan W. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *NEJM* 382(8): 727- 733.
- Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Wu J. 2020. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 382(12): 1177-1179.
- Zust R, Cervantes-Barragan L, Habjan M, Maier R, Neuman BW, Ziebuhr J, Thiel V. 2011. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol* 12(2): 137-143.