Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Genetic comparison among various coronavirus strains for the identification of potential vaccine targets of SARS-CoV2

Navpreet Kaur a, Rimaljot Singh a, Zahid Dar b, Rakesh Kumar Bijarnia c, Neelima Dhingra b, Tanzeer Kaur a,*

a Department of Biophysics, Panjab University, Chandigarh, India
b University Institute of Pharmaceutical Sciences, Panjab University, Chandigarh, India
c Medzome Life Sciences Pvt.Ltd., Baddi, Himachal Pradesh, India

ARTICLE INFO

Keywords: COVID-19 SARS-CoV2 Genomic comparison ER stress Target antigenic sites Vaccine development

ABSTRACT

On-going pandemic pneumonia outbreak COVID-19 has raised an urgent public health issue worldwide impacting millions of people with a continuous increase in both morbidity and mortality. The causative agent of this disease is identified and named as SARS-CoV2 because of its genetic relatedness to SARS-CoV species that was responsible for the 2003 coronavirus outbreak. The immense spread of the disease in a very small period demands urgent development of therapeutic and prophylactic interventions for the treatment of SARS-CoV2 infected patients. A plethora of research is being conducted globally on this novel coronavirus strain to gain knowledge about its origin, evolutionary history, and phylogeny. This review is an effort to compare genetic similarities and diversifications among coronavirus strains, which can hint towards the susceptible antigen targets of SARS-CoV2 to come up with the potential therapeutic and prophylactic interventions for the prevention of this public threat.

1. Introduction

Coronaviruses (CoVs) are single-stranded, positive-sense RNA viruses belonging to the order Nidovirales, family Coronaviridae, and subfamily Coronavirinae (Spaan et al., 2012). CoVs possess the largest genomes among all RNA viruses ranging from 26 to 32 kilobases in length, with G + C contents varying from 32% to 43% (Woo et al., 2009; Woo et al., 2007; Woo et al., 2005). CoVs are predominantly associated with enteric and respiratory diseases in animals and humans (Cheng et al., 2004; Gélinas et al., 2001).

Subfamily Coronavirinae further diverges into three major genera or groups -the alpha-CoVs (group 1), the beta-CoVs (group 2), and the gamma-CoVs (group 3) characterized by varying genetic makeup and antigenic cross-reactivity(Cleri et al., 2010; Gorbalenya et al., 2004; Khan et al., 2020; Woo et al., 2010). Delta-CoVs, representing a novel genus of Coronaviruses were also later found in birds and pigs. Out of these, only alpha-CoV strains (HCoV-229E and HCoV-NL63) along with beta-CoVs (HCoV-HKU1, HCoV-OC43, severe acute respiratory syndrome coronavirus [SARS-CoV], and Middle East respiratory syndrome coronavirus [MERS-CoV]) have been identified as human pathogenic strains (Cortellis, 2020). Swine acute diarrhea syndrome (SADS-CoV) (Zhou et al., 2018), porcine transmissible gastroenteritis virus (Brian and Baric, 2005), and porcine enteric diarrhea virus (PEDV) (Lin et al., 2016) are a few examples of alpha and beta-CoVs that hold the potential to pose heavy disease burden on livestock.

Currently, an ongoing pneumonia outbreak caused by a novel Coronavirus strain named SARS-CoV2 belonging to the beta-coronavirus genera is an emerging threat to public health all over the world. Ongoing extensive research has suggested that bats are the natural reservoir and pangolins are probably the potential intermediate host of the virus, as 99% homology was found between SARS-CoV2 and coronavirus strain isolated from pangolins(Zheng, 2020).

A plethora of research is being conducted globally on this novel coronavirus strain to gain knowledge about its origin, evolutionary history, and phylogeny. This review is an effort to compare genetic similarities and diversifications among coronavirus strains, which can...
hint towards the susceptible antigen targets of SARS-CoV2 to come up with the potential therapeutic and prophylactic interventions for the prevention of this public threat.

1.1. Genomic organization of the viral genome

Virion structure of coronavirus is made up of four major structural proteins namely, spike, envelope, membrane, nucleocapsid (Fig. 1). Distinct CoV strains share a common genetic organization for the coding region encoding for a canonical set of genes in the order 5′-end- ORF1a/b replicase, spike, envelope, membrane, nucleocapsid-3′ end (Fig. 2), although the number and location of accessory ORFs present in different CoV species vary (6–11 ORFs) (Brian and Baric, 2005). Subgenomic (sg) mRNAs that are responsible for the translation of genes form a 5′ and 3′ co-terminal nested set with viral genome. A common 5′ leader sequence and a 3′ terminal sequence is present along with subgenomic mRNAs. Both 3′ and 5′ ends of the genome carry small untranslated regions (UTRs). Besides, several nonstructural proteins (nsps) including RNA-dependent RNA polymerase (RdRp), coronavirus main protease (3CLpro), and papain-like protease (PLpro) are also encoded by the viral genome (Anand Gaurav, 2020; Baez-Santos et al., 2014; Lai, 1990).

A typical CoV genomic and subgenomic sequence consists of at least six ORFs. The first ORF representing almost 67% of the entire viral genome (ORFa/b) encodes 16 nsps (non-structural proteins) (1–16 nsp), except gamma-CoVs that lack nsp1. A papain-like protease (PL2pro) in nsp3 and a 3C-like protease (3CLpro; also known as the “main protease”) in nsp5 are the two protease domains conserved in all types of CoVs which are encoded by ORF1a sequence in the genome (Van Boheemen et al., 2012). Other ORFs encode for few accessory proteins and some major structural proteins including envelope (E), spike (S), membrane (M), and nucleocapsid (N) proteins (Van Boheemen et al., 2012; Czub et al., 2005; Woo et al., 2010).

1.2. Distinct characteristics of Novel Coronavirus SARS-CoV2

1.2.1. Receptor binding domain

The RBD sequence in the spike protein of a coronavirus is the most variable region and prone to mutations (Tao et al., 2020; Zhou et al., 2020a). A total of six amino acid residues are suggested to be essential for binding to the human ACE-2 receptor (Yushun Wan, Jian Shang 1, Rachel Graham, Ralph S. Baric, 2020). As per the amino acid sequence coordinates of SARS-CoV, these are Y442, L472, N479, D480, T487, and Y491, whereas in SARS-CoV2 these corresponds to L455, F486, Q493, S494, N501, and Y505 residues respectively (Walls et al., 2020; Yushun Wan, Jian Shang 1, Rachel Graham, Ralph S. Baric, 2020). Interestingly, studies have suggested that out of these six residues of SARS-CoV2 five does not share similarity to SARS-CoV counterparts and the change is probably due to mutations, deletions or insertions in the S1-S2 region of Coronaviruses (Walls et al., 2020; Wrapp et al., 2020; Yushun Wan, Jian Shang 1, Rachel Graham, Ralph S. Baric, 2020).

1.2.2. Polybasic furin cleavage site and O-linked glycans

Polybasic cleavage site (RRAR) is the second distinct characteristic of SARS-CoV2 located at the junction of two subunits i.e., S1 and S2 of Spike (Walls et al., 2020). Proteases like furin can effectively cleave this site. In SARS-CoV2, a leading proline is also inserted at the junction (between subunits S1 and S2 of Spike) that leads to the formation of a turn/stem-loop structure, which consequentially results in the insertion of O-linked glycans to cleavage site residues S686, S673, and T678. The presence of O-linked glycans is very unique and different from other CoV variants (Chan et al., 2008; Vankadari and Wilce, 2020).

This polybasic proteolytic cleavage of S glycoprotein is responsible for determining the viral infectivity as well as host range because whether the virus can jump across species e.g., from bats to humans, is determined by this cleavage (Nao, 2017). The S protein isolated from Ugandan bats MERS-like CoV species, for instance, was found to dully cohere to human cells, but viral entry was not mediated. However, if during coherence of virus to the human cell, protease trypsin is inserted, it could cleave the S glycoprotein and allow the virus to enter into the cells (Menachery et al., 2019). This phenomenon demonstrates the importance of cleavage of S glycoprotein for mediating viral entry into human cells and thus determining viral infectivity and that restricting its cleavage by manipulating the expression of S protein or ACE-2 receptor binding motif might be a great therapy.

Although the protein sequence analysis has suggested the presence of furin cleavage sites in SARS-CoV2 protein, however, RaTG-13 (Zhou et al., 2020b), the most closely related bat CoV strain to SARS-CoV2 isolated from a bat in Yunnan in 2013 does not constitute this furin cleavage site (Zhou et al., 2020a). However, there are abundant furin proteases present in respiratory tracts of humans suggesting that S glycoprotein cleavage takes place while exiting from RBCs and thus infect other cells (Izaguirre, 2019). This furin cleavage site might have been acquired by RNA recombination and its presence in SARS-CoV2 might be responsible for infecting human cells. Also, this cleavage site might have allowed bat CoV to jump into humans and thus initiate the ongoing pneumonia outbreak COVID-19 (Andersen et al., 2020; WO, 2020; Wrapp et al., 2020; Zhou et al., 2020a).

1.3. Human coronavirus serotypes

In 1965, B814 was described as the first human Coronavirus (HCoV) strain. In subsequent years, about 30 more types of HCoV strains have been identified. Out of these, seven HCoV strains i.e., two alpha-CoVs (HCoV-229E; HCoV-NL63), and five beta-CoVs (HCoV-OC43; HCoV-HKU1; MERS-CoV; SARS-CoV including novel SARS-CoV-2) (Table 1) (Fig. 3) are commonly circulating in the human population and predominantly responsible for cold symptoms and other respiratory diseases in healthy individuals. The Middle East respiratory syndrome CoV and Severe acute respiratory syndrome CoV are two zoonotic viruses that pose major public health threats as they are associated with severe lower respiratory tract infections, while most other HCoVs cause comparatively mild upper respiratory tract infections (common cold) (Jonsdottir and Dijkman, 2016).

1.3.1. Alpha coronavirus

The genus of CoVs contains two human pathogenic viral strains i.e., HCoV-229E and HCoV-NL63. HCoV-229E utilizes angiotensin-converting enzyme-2 (ACE-2) (Hofmann et al., 2005) receptors to enter in the host cell, similar to SARS-CoV, which is otherwise a beta-CoV.
1.3.2. Beta coronavirus

MERS-CoV and SARS-CoV (Shults et al., 2012; Zaki et al., 2012) are two bat-viruses included in this genus. Novel COVID-2019 disease-causing virus is also classified as a beta-CoV and has been named as SARS-CoV2. The angiotensin-converting enzyme 2 (ACE2) (Gorbalenya et al., 2004; Woo et al., 2010) acts as one of the main receptors in SARS-CoV to enter the host cell, whereas, MERS-CoV, on the other hand, utilizes dipeptidyl peptidase 4 (DPP4, also known as CD26) (Song et al., 2014) as the primary receptor. Other than these, HCoV-OC43 and HCoV-HKU1, are two non-SARS CoV species included in this genus that probably utilize sialic acid residues as receptors and have hemagglutinin-esterase activity (Vlasak et al., 1988).

1.3.3. Gamma coronavirus

Mostly Avian Coronaviruses are included in this genus; the infectious bronchitis virus of chickens is the most prominent of all. This virus is responsible for causing respiratory and reproductive tract disease in chickens.

1.3.4. Delta coronavirus

This group represents a novel genus of Coronaviruses found in birds and pigs.

2. Origin and evolution of SARS-CoV and SARS-CoV2

Current sequence database survey analysis shows an animal origin of all diverse sorts of human CoVs: HCoV-NL63, HCoV-229E, SARS-CoV and MERS-CoV are mulled over to have originated from bats, whereas, HCoV-OC43 and HCoV-HKU1 are considered to have evolved from rodents (Forni et al., 2017; Su et al., 2016) (Fig. 4). The transmission of the virus from natural hosts to humans requires an intermediate host; domesticated animals might serve as essential hosts for viral transmission, additionally, domestic animals themselves can endure the disease caused by bat-borne or closely associated CoVs. Of the various coronavirus strains found so far, most of them are considered to have evolved from bats, thus, bats make the major natural reservoirs of alpha and beta-CoVs. (Woo et al., 2012)

Initially, when the SARS epidemic originated, most of the diseased patients were found with a record of animal exposure before the disease development. In 2005, two groups independently documented that horseshoe bats (genus Rhinolophus) (Lau et al., 2005; Li et al., 2005) were the main natural reservoir of novel CoVs related to human SARS-CoV. Many other CoV strains which are phylogenetically related to SARS-CoV were also discovered to have bat-origin, isolated from various locations including diverse European, African and Southeast Asian countries and also from, China (Gouilh et al., 2011; Hu et al., 2017; Lacroix et al., 2017; Lau et al., 2005; Li et al., 2005; Rihtari et al., 2010; Tong et al., 2009; Wacharapluesadee et al., 2015). As per studies, the SARS-CoV strains from Hipposideros bats (Tong et al., 2009) in Africa were found to be phylogenetically distant from those of Rhinolophus SARS-CoV strains suggesting that SARS-CoV have wide geographical spread and bats might have been a host of SARS-CoV from a very long period. A cave of Yunnan province, China, is designated as a diversity hot spot as this location contains a highly diverse SARSr-CoVs in bat populations (Ge et al., 2013; Hu et al., 2017; Lacroix et al., 2017; Lau et al., 2005; Li et al., 2005; Rihtari et al., 2010; Tong et al., 2009; Wacharapluesadee et al., 2015). The SARSr-CoV in this location consists of all the genetic diversity of SARS-CoV found in all other places of china. All these pieces of evidence point towards an immensely diverse nature of SARS-CoV, additionally, despite 15 years of extensive searching, researchers are unable to locate the ultimate progenitor of SARS-CoV in bat populations (Ge et al., 2013; Hu et al., 2017; Wang et al., 2016). The SARSr-CoV in this location consists of all the genetic diversity of SARS-CoV found in all other places of china. All these pieces of evidence point towards an immensely diverse nature of SARS-CoV, additionally, despite 15 years of extensive searching, researchers are unable to locate the ultimate progenitor of SARS-CoV in bat populations (Ge et al., 2013; Hu et al., 2017; Wang et al., 2016). The SARSr-CoV in this location consists of all the genetic diversity of SARS-CoV found in all other places of china. All these pieces of evidence point towards an immensely diverse nature of SARS-CoV, additionally, despite 15 years of extensive searching, researchers are unable to locate the ultimate progenitor of SARS-CoV in bat populations (Ge et al., 2013; Hu et al., 2017; Wang et al., 2016).
be-identified.

The S gene and upstream orf8 are the two regions of the viral genome suggested to be essential break-points for viral RNA recombination, where the former encodes the spike (S) protein containing the receptor-binding domain (RBD) whereas, later encodes an accessory protein (Hon et al., 2008; Hu et al., 2017; Wu et al., 2016). Frequent RNA recombination of CoVs and immense genetic diversity, as well as the prevalence of SARS-CoV, is likely to be one of the reasons for the emergence of novel coronavirus variant i.e., SARS-CoV2 and subsequently COVID-19 outbreak.

2.1. Natural selection

Andersen et al. (2020) have tried to explain the proximal origin of SARS-CoV2 by providing two hypotheses of its emergence. First, natural selection in an animal host prior to zoonotic transfer, and the second hypothesis is natural selection in humans after the zoonotic transfer (Andersen et al., 2020). Natural selection is a process that allows organisms to adapt to their environment by selectively reproducing suitable variations in their genotype or genetic constitution (Gregory, 2009).

2.1.1. Natural selection in an animal host prior to zoonotic transfer

Andersen et al. (2020) have tried to explain the proximal origin of SARS-CoV2 by providing two hypotheses of its emergence. First, natural selection in an animal host prior to zoonotic transfer, and the second hypothesis is natural selection in humans after the zoonotic transfer (Andersen et al., 2020). Natural selection is a process that allows organisms to adapt to their environment by selectively reproducing suitable variations in their genotype or genetic constitution (Gregory, 2009).

2.1.1. Natural selection in an animal host prior to zoonotic transfer

The RBD region of S protein in SARS-CoV2 is optimized for binding to human-like ACE2 receptor, and the most probable reason for this change is natural selection (Tao et al., 2020; Zhang et al., 2020a; Zhou et al., 2020b). As bat SARS-CoV-like Coronaviruses share the closest homology to SARS-CoV2 therefore, bats are thought to be the natural reservoir of SARS-CoV2 (Tao et al., 2020). Additionally, the genome of Rhinolophus affinis bat (Zhou et al., 2020b) was found to share 96% similarity to the novel coronavirus with non-homologous regions belonging to the RBD region of S protein suggesting incompatibility of this strain with human ACE2 receptor (Wan et al., 2020). This suggests
natural selection as the reason for the optimization of S protein in novel coronavirus.

Furthermore, no strain of bat beta-CoV is found to have a polybasic cleavage site as found in the genome of novel coronavirus SARS-CoV2 strain. Also, no direct progenitor of SARS-CoV2 is identified until now. All these shreds of evidence suggest that mutations, insertions, and deletions at the S1 and S2 junction of Spike protein of SARS-CoV2 are responsible for evolutionary changes (Andersen et al., 2020; Yamada and Liu, 2009).

A high population density of an animal host is a must for a precursor virus to achieve both mutations and polybasic cleavage site in S protein of SARS-CoV2 (Andersen et al., 2020).

2.1.2. Natural selection in humans after zoonotic transfer

Another hypothesis of adaptation of above mentioned genomic features and consequently the origin of SARS-CoV2 could be that a progenitor of SARS-CoV2 containing all the genomic characteristics (polybasic cleavage site and mutations in S protein) might have jumped into humans via yet unknown human to human transmission (Andersen et al., 2020; Tu et al., 2020; Zhou et al., 2020b).

Various strains of SARS-CoV2 sequenced so far carry similar genomic characteristics as described above, pointing towards a common progenitor from which they have adapted these features. The RBD region in SARS-CoV2 shows features very similar to the viral strains isolated from pangolins suggesting that the virus that jumped to humans probably consisted of this region and polybasic cleavage insertion might have occurred during human to human transmission (Andersen et al., 2020). This is true in the case of MERS-CoV where repeated jumps of the virus from dromedary camels were the reason for all human cases producing short transmission chains or single infections that without adapting to sustained transmission, eventually resolved (Dudas et al., 2018).

Retrospective serological studies and examination of banked human samples can be informative and assist in determining whether such cryptic spread has occurred or not (Andersen et al., 2020; Dudas et al., 2018; Wang et al., 2018).

3. RNA virus mutations

Most of the evolutionary changes in organisms occur as a consequence of mutations. The mutation is an alteration of the nucleotide sequence in the genome of an organism, these variations bring novelty that we observe in the course of evolution; also, natural selection can act upon these alterations (Baer, 2008). Usually, small count of these variations is beneficial, some are inconsequential (neutral) and most of them are harmful or non-beneficial for the organisms. Whether an organism has a low mutation rate or high, the count of harmful mutations always have a low mutation rate or high, the count of harmful mutations always outnumbers beneficial mutation (Loewe and Hill, 2010).

The largest group of molecular parasites known to infect humans, animals, and plants are RNA viruses. In-depth study of different RNA virus genomes, as well as their replication pathways, suggests their heterogeneous nature in using a wide variety of strategies to enhance their stability and thus evolvability (Murphy et al., 2016; Steinhauer, 1987). Eradication of existing viruses and emergence or re-emergence of new viruses is likely to be balanced by this process of natural selection (Murphy et al., 2016). Environmental alterations along with genetic plasticity of RNA viruses favour the emergence of several new RNA viruses, consequently, new viral pathogens come in contact with potential hosts and facilitate host jumping (Morse and Schulerberg, 1990).

CoV mutations: RNA viruses usually have relatively high mutation rates than DNA viruses as well as a million times higher than their hosts; this is the reason for enhanced viral adaptability and evolvability (Duffy, 2018). Mutation rate or mis-insertion errors during RNA replication have been suggested to fall in the range of $10^{-6}$ to $10^{-5}$ substitutions per nucleotide (Domingo et al., 1988). Immensely high rates of mutations in RNA viruses results in a yield of offsprings that differ by 1–2 mutations from their parent, forming a diverse mutant cloud of descendants (Vignuzzi and Andino, 2012). Absence or lack of efficiency of proof-reading of RNA polymerases is one of the contributing factors of such high mutation rates in these viruses. In contrast, the RNA viruses with the largest known genome sizes i.e., coronaviruses, show relatively low rates of mutations than other RNA viruses. The slower mutation rates or preservation of such large genomes is probably associated with the exceptional characteristics of CoV RTC (replication-transcription complex) which contains 3′-5′ exoribonuclelease activity that probably provides proof-reading function which is unique to CoV genomes among all other variants of RNA viruses (Chen et al., 2020; Minskaia et al., 2006).

The rate of RNA virus mutations usually lies in between the range of $10^{-6}$ to $10^{-4}$ substitutions per nucleotide site per cell infection (Peck and Lauring, 2018) whereas, studies suggest that the rate of mutation of SARS-CoV falls in a range between 0.80 and $2.38 \times 10^{-3}$ nucleotide substitutions per site per year, much lower than other RNA virus counterparts (Zhao et al., 2004).

4. Genomic comparison of CoV strains

The whole-genome sequence alignment of CoV revealed 54% identity among varying CoV strains, whereas, the genome sequence alignment of the nsp-coding region alone of CoV shows 58% similarity and that of the structural protein-coding region shows 43% similarity, suggesting that nsp5 (non-structural proteins) form the conserved region of the genome with high percent identity, contrarily, structural proteins in need of adaptation to new hosts are more diverse (Chen et al., 2020).

The mutation rates in the RNA virus replication are relatively much higher than DNA viruses. RNA virus genomes are usually small, with less than 10 kb length, however, the genome length of CoVs are comparatively larger with approx. ~30 kb length permitting easier accommodation and modification of genes (Vega et al., 2004). The slower mutation rates or preservation of such large genomes is probably associated with the exceptional characteristics of CoV RTC (replication-transcription complex) which contains 3′-5′ exoribonuclelease activity of nsp14 as well as various other RNA processing enzymes (Lauber et al., 2013; Minskaia et al., 2006). The 3′-5′ exoribonuclelease activity probably provides a proof-reading function that is unique to CoV genomes among all other variants of RNA viruses (Chen et al., 2020; Minskaia et al., 2006).

4.1. Genomic comparison of SARS-CoV2 (COVID-19) with other CoV strains

In pursuance of recent studies, the genome of SARS-CoV2 is believed to have originated from bats showing 96.2% homology to a bat SARS-related coronavirus (SARs-CoV; RaTG13) collected in Yunnan province, China (Zhou et al., 2020b), besides, it was shown to possess 85% similarity to the genome of the SARS-like virus ZC45 (bat-SL-CoVZC45, MG772933.1), together these viruses along with SARS-like virus ZXC21 have been grouped into a unique Orthocoronavirinae subfamily (Ma

| S No | Viral strains | Genus | Percent identity |
|------|---------------|-------|-----------------|
| 1    | HCoV-229E     | α     | 65.04           |
| 2    | HCoV-NL63     | α     | 65.11           |
| 3    | HCoV-HKU1     | β     | 67.59           |
| 4    | HCoV-OC43     | β     | 68.93           |
| 5    | MERS-CoV      | β     | 69.58           |
| 6    | SARS-CoV      | β     | 82.45           |
| 7    | bat-SL-CoVZC45| β     | 88              |
| 8    | bat-SL-CoVZC21| β     | 88              |
| 9    | SARS-HCoV Tez2| β     | 82              |
| 10   | SARS-HCoV BJ01| β     | 82              |
| 11   | SARSr-CoV; RaTG13| β | 96.2        |

Table 2 Percent identity of novel coronavirus SARS-CoV2 strain with different CoV strains. (Chan et al., 2020; Chen et al., 2020; Gralinski and Menachery, 2020; Malik et al., 2020; Ren et al., 2020; Zakí et al., 2012; P. Zhou et al., 2020).
et al., 2020). (Table 2).

4.1.1. SARS-CoV2 and SARS-CoV

The results of zpicture comparative genomic analyses of SARS-CoV2 and SARS-CoV have revealed extremely high homology between the two strains at the nucleotide level. Additionally, the genomes of these two strains differ from each other in six regions. The first three regions of difference belong to the partial coding sequences of ORF1a/b (448 nt, 55 nt, and 278 nt, respectively). The next two regions belong to the partial coding sequences of the S gene (315 nt and 80 nt, respectively) and the last region of difference is a part of the coding sequence of the orf7b and orf8 genes (214 nt) (Jiaxu Bo, 2014). The spike gene of SARS-CoV2 shows more homology to bat-CoV, whereas two accessory genes 3a and 8b possess homology to SARS-CoV.

Proteomic similarity analyses of SARS-CoV and SARS-CoV2 have suggested that most of the proteins are highly homologous (95%–100%). RdRp and 3CLpro protease share over 95% of sequence similarity even though at the genome level these two strains share only 82% similarity (Chan et al., 2020; Dong et al., 2020; Lu et al., 2020; Morse et al., 2020). Additionally, both these strains share 76% of sequence similarity in their S proteins, a highly conserved receptor-binding domain (RBD), and a domain of S protein (Chan et al., 2020; Dong et al., 2020; Lu et al., 2020; Morse et al., 2020). Also, PLpro sequences of SARS-CoV and SARS-CoV2 share 83% similarity with a large number of similar active sites (Morse et al., 2020). All these pieces of evidence suggest a common evolutionary history of both these viral strains.

However, SARS-CoV2 possesses two proteins (orf8 and orf10) that share no homology to the SARS-CoV strain. SARS-CoV conserved sequences of orf8 are different from the amino acid sequence of orf8 derived from SARS-CoV2. Since these two proteins of SARS-CoV2 possess no homology to other CoV strains, it might be therapeutically beneficial to study the biological function of these two proteins i.e., orf8 and orf10 in SARS-CoV2 (Chan et al., 2020).

4.1.2. Structural comparison of SARS-CoV2 from SARS-CoV

Studies suggest a close relationship between SARS-CoV2 and SARS-CoV or SARS-like bat-CoV. Analysis of the structural and functional differences between SARS-CoV2 and SARS-CoV might be accomplished by shedding light on the amino acid substitutions in different proteins. Altogether, 380 amino acid substitutions occurred between the amino acid sequences and corresponding consensus sequences of SARS-CoV. Envelope, matrix, or accessory proteins p6 and 8b, nsps7, and nsps13 revealed no change in amino acid sequences. Non-structural proteins nsps2 and nsps3 manifested single amino-acid substitution at positions 61 and 102 respectively ( Wu et al., 2020). Additionally, spike protein was found to possess 27 amino acid substitutions with a length of 1273 amino acids, including amino acid region 357–528 in the RBD with six substitutions and another six at amino acid region 569–655 in the underpinning subdomain (SD). Furthermore, the C-terminal of the receptor-binding subunit S1 domain was found to have four substitutions (Q560L, S570A, F572T, and S575A) (Guo et al., 2004; Wu et al., 2020). The receptor-binding motifs that interact with the human ACE-2 receptor were found to be exactly similar to the SARS-CoV strain with no amino acid substitutions, whereas, six mutations occurred in the other RBD region (Ge et al., 2013). Due to limited knowledge about SARS-CoV2, the reasonable explanations for these amino acid substitutions are not known yet.

4.2. Genomic Divergence from other beta-CoV strains

Novel SARS-CoV2 is currently being researched the most around the globe to cope up with the current pandemic scenario affecting a great number of people worldwide for coming up with efficient vaccines and other therapeutic interventions. Globally a bunch of research groups are working to deduce an ample amount of information from the viral genome and to find out the most potential target for the development of the viral vaccine. The search for a potential vaccine has led the researchers to conduct SARS-CoV2 genomic comparison with the previous CoVs, to know if the vaccine targets for previous CoVs can be useful for developing a potential vaccine for the novel virus.

Ceraolo and Giorgi (2020) performed an interspecies genomic divergence investigation for all proteins encoded by SARS-CoV2 and its relatives by aligning all the sequences using MUSCLE (Ceraolo and Giorgi, 2020). They stated a close homology (>80%) of SARS-CoV2 with bat BCoV bat-SL-CoVZXC21 and the high conservation of E, M, and A structural proteins across all beta-CoV strains. As specified by the results of the above-stated research, we performed a multiple sequence alignment along with the assessment of the percent identity matrix using CLUSTAL OMEGA 1.2.4 and constructed phylogenetic trees using JALVIEW 2.10.5.

In compliance with the results stated by Ceraolo and Giorgi, 2020, we observed a very high genomic conservation between the Wuhan isolated SARS-CoV2 strain and SARS Bat coronavirus RaTG13 strain, with a percent identity of 97.71% and 85.41% in S protein and N protein respectively (Table 4). The ACE-2 RBD subdomain of S2 subunit of SARS-CoV2 spike protein displayed 74.41% of genomic similarity to bat SARS RaTG13 strain showing high conservation level in the RBD region.

Furthermore, Robson, 2020 stated that subsequence KRSFELDLNFNV of the S2’ spike glycoprotein proteolytic cleavage site, is an important motif and is found to be responsible for initial binding of previous SARS-CoV to lung cells and their activation of the spike protein by proteolytic cleavage (Robson, 2020). So, we also carried sequence alignment for this particular region and found it to be completely conserved as we observed a 100% match between the SARS-CoV2 strain and SARS bat RaTG13 strain, whereas the motif differs at 3 amino acid positions in MERS strain. For further visual confirmation, of the level of genomic divergence and similarity between SARS-CoV2, SARS Bat coronavirus RaTG13, and other common beta CoV strains, via MSA, in the S and N structural proteins, please refer to Fig. 7.

5. CoV introduction into cell and its replication

The adherence of the host receptor to the virus via S protein marks the initiation of the viral infection. As soon as the virus adheres to the host receptor, the process of cleavage of viral S protein into two subunits i.e., S1 (N-terminal receptor binding domain) and S2 (C-terminal domain) are triggered by host proteases (Huang et al., 2006; Qiu et al., 2006; Yamada et al., 2009). The S1 subunit interaction with the host receptor plays a major role in determining the host range of CoVs (Kuo et al., 2000). Additionally, this interaction brings about a conformational change in the S2 subunit of S protein thus exposing the hidden fusion peptide which aids in viral entry into the host cellular membrane. The conformational change leads to the formation of a six-helix bundle fusion core that brings the virus and hosts cellular membrane nearby hence, fusing the lipid bilayers. This fusion ensures the entry of viral nucleocapsid into the host cellular cytoplasm (Fung and Liu, 2014; Masters, 2006).

Once the viral nucleocapsid is inside the host cell, it is uncoated thus revealing the viral genomic RNA. This genomic RNA now acts as mRNA for initiating translation of replicase polyprotein. Two open reading frames i.e., ORF1a and ORF1b are present in the replicase gene. Translation of these open reading frames i.e., ORF1a and ORF1b results respectively in the development of polyprotein 1a (pp1a) and a larger polyprotein 1ab (pp1ab) (Brierley et al., 1987). Newly synthesized pp1a and pp1ab undergo autoproteolytic cleavage. Cleavage of pp1a produces 11 non-structural proteins (nsp5) i.e., nsp1–nsp11, and that of pp1ab results in the formation of 15 nsp5 (nsp1-nsp10 and nsp12–nsp16). The functions of many nsp5 are not completely understood, however, the papain-like protease in nsp3 and main protease in nsp5 are responsible for this autoproteolytic cleavage of pp1a and pp1ab, whereas the RNA-dependent RNA polymerase (RdRp) is contained within nsp12 (Baker et al., 1993).
After translation, the replicase now utilizes the viral genomic RNA as a template to synthesize negative sense genomic RNAs which will later serve as a template for synthesis of progeny positive-sense genomic RNAs. Additionally, replicase synthesizes a nested set of subgenomic RNAs (sgRNAs) via discontinuous transcription process (Sawicki et al., 2007). The processes of replication and transcription of CoVs results in the formation of replication-transcription complexes (RTC), membrane proteins nsp3, nsp4, and nsp6 aids in anchoring these complexes to the intracellular membrane (Oostra et al., 2007).

The translation of sgRNAs leads to the formation of transmembrane proteins (S, M, and E) as well as accessory proteins. After synthesis, the structural proteins are inserted into the endoplasmic reticulum, folded, and then conveyed to ER-Golgi intermediate compartment (ERGIC). Further, the formation of nucleocapsid takes place by encapsulation of progeny genomic RNA which is synthesized by translation of N proteins. With the aid of M proteins, protein-protein interactions occur inside ERGIC which consequently results in assembly of the complete virion (Masters, 2006). Secretory pathways present inside the smooth wall vesicles then help in transporting the newly formed virions and thus fusing them with the plasma membrane to release the mature virus particles. The S proteins of these viruses further infect the other uninfected host cells resulting in an escalation of viral growth inside the host body (Krijnse-Locker et al., 1994; Masters, 2006).

6. CoV infection and er stress

The endoplasmic reticulum is one of the main organelle involved in the synthesis and proper folding of newly synthesized proteins and thus the growth of viral infection (GM, 2000). However, under certain circumstances, the protein folding capacity of ER might not be adequate to carry out proper folding of all the newly synthesized proteins which can result in the accumulation of unfolded proteins in ER leading to ER stress. Unfolded protein response (UPR) is the signaling pathway evolved by the cells to combat ER stress and maintain cellular homeostasis (Shapiro et al., 2017; Ron and Walter, 2007). The accumulation of unfolded proteins in ER triggers the activation of three ER stress transducers, namely: PKR-like ER protein kinase (PERK), activating transcriptional factor-6 (ATF6), and inositol-requiring protein-1 (IRE1) (Tabas and Ron, 2011). The activation of these transducers further initiates UPR signaling. Once activated, this signaling pathway tries to regain cellular homeostasis by increasing the pace of proper protein folding and decreasing the protein synthesis inside ER and if the stage of homeostasis is impossible to achieve, UPR activates apoptotic pathways for the betterment of the organism (Fung and Liu, 2014; Tabas and Ron, 2011).

Research studies around the globe have proposed that SARS-CoV infected cells exhibited an up-regulated expression of genes associated with ER stress i.e., glucose-regulated protein 94 (GRP94) and glucose-regulated protein 78 (GRP78) (Jiang et al., 2003; Yeung et al., 2008). Moreover, current pieces of evidence suggest three chief phenomenons involved in the induction of ER stress by CoV namely: Formation of double-membrane vesicles, glycosylation of CoV structural proteins, and depletion of ER lipid (Fung and Liu, 2014).

1. Formation of double-membrane vesicles (DMVs): CoV infected cells have been shown to induce the formation of DMVs. Cellular membrane modifications are suggested to occur during the replication of various plus-strand RNA viruses and CoV is one among them (David-Ferreira and Manakar, 1965).

As specified by the data gained using electron microscopy, DMVs are found to be located probably in the sites near RTCs in the vicinity of major CoV replicase proteins (Gosert et al., 2002; Snijder et al., 2006). The source of DMV formation is not yet clear but the late endosomes, autophagosomes, and the early secretory pathway are thought to be the originators of DMVs (Prentice et al., 2004; van der Meer et al., 1999; Verheije et al., 2008). A recent study by Reggiori et al., 2010, has suggested that to form DMVs, CoVs seize the EDEMosomes to obtain the ER membrane. The COPII-independent vesicles, EDEMosomes, are usually found in the ER and are responsible for maintaining the level of mannosiadase alpha-like1 (EDEM1), a regulator of ER-linked degradation (Cali et al., 2008). So, the above pieces of evidence point towards ER-origin of CoV induced DMVs (Reggiori et al., 2010).

2. Glycosylation of CoV structural proteins: During the process of CoV replication, transmembrane structural proteins (S, E, and M) are synthesized by ER in a tremendous amount to assemble progeny virions. The production, folding, and modifications of such a massive number of proteins put the immense workload on ER, thus enhancing the probability of generating ER stress (Shapiro et al., 2017; Fung and Liu, 2014).

Protein glycosylation is an integral part of protein folding in the ER. The reaction involves the addition of a carbohydrate moiety to the protein molecule (Roth et al., 2012). Excluding N proteins, all other structural proteins are synthesized in a massive amount by ER. In CoVs, based on amino-acid side-chain atoms to which glycans are attached, two types of protein glycosylation are found to take place in ER i.e., O-linked (in beta-CoVs) and N-linked (in alpha and gamma-CoVs) (Cavanagh, 2007; Jacobs et al., 1986; Nal et al., 2005). M protein is one of the most abundant proteins present in CoV and glycosylation of this protein is associated with induction of alpha interferon (IFN) function as well as in vivo tissue tropism (Charley and Laude, 1988; De Haan et al., 2003; Laude et al., 1992). S protein is also highly glycosylated (Masters, 2006), moreover, the glycans on S protein in SARS-CoV have been shown to interact with two alternative receptors of SARS-CoV independent of major ACE-2 receptor namely: L-SIGN (liver lymph node-specific intercellular adhesion molecule-3-grabbing non-integrin), and DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin) (Han et al., 2007).

Chaperones are the proteins present inside ER that assist in proper protein folding, maturation as well as assembly. S proteins are highly dependent on calnexin, an ER protein chaperone; moreover, studies have suggested that S2 subunit in S glycoproteins of SARS-CoV interact with calnexin and a decrease in the infectivity of pseudo-type lentivirus carrying SARS-CoV S protein was observed after knocking out calnexin (Fukushi et al., 2012). So, striking out chaperones will hamper proper protein folding and consequently assembly of virions, thus decreasing viral spread and infectivity (Fung and Liu, 2014).

3. Depletion of ER lipid: After synthesis and folding of proteins in the ER, they are transported to ERGIC for assembly of the virion. Mature virions are released by the process of exocytosis leaving behind an ER with depleted lipid levels (Fung and Liu, 2014).

The above-mentioned factors play a vital role in the generation of ER stress, which in turn activates the UPR pathway that brings stress response factors like PERK, IRE1, and ATF6 into action (Fig. 5). Although the detailed mechanism of molecular interactions between CoV S proteins and stress factors (PERK/IRE1/ATF6) has not been determined, but the layout of CoV induced UPR activation pathways are demonstrated below:

CoV induced activation of the PERK pathway: PERK is the first branch to get activated in the UPR pathway after ER stress (Szegedi et al., 2012). The translation of PERK protein is entirely collapsed as the activation of PERK (protein kinase RNA-like endoplasmic reticulum kinase) pathway results in phosphorylation of α-subunit of eukaryotic initiation factor2 (eIF2α), that regulates the mRNA translation machinery, bringing a halt to protein translation (Ron and Walter, 2007). Additionally, the transcription factor GADD153 is responsible for the initiation of CoV induced apoptosis and it was found that the expression of GADD153 is up-regulated during late CoV infection stages (Fung and Liu, 2014; Marciniak et al., 2004; Puthalakath et al., 2007).
CoV induced activation of the IRE1 pathway: Inositol-requiring kinase 1 (IRE1), self-activates its free luminal domain by homodimerization and transautophosphorylation. The activated domain, in turn, activates the transcription factor XBP1 (Xbox binding protein) mRNA by splicing and removing a 26 bp intron (Calfon et al., 2002; Yoshida et al., 2001). Activated XBP1 then induces upstream UPR gene expression (Bechill et al., 2008; Versteeg et al., 2007). Besides, IRE1 also modulates JNK and Akt kinase activities that help infected cells escape virus-induced apoptosis (Fung and Liu, 2014; Urano et al., 2000).

CoV induced activation of the ATF6 pathway: Activating transcription factor (ATF6) is basic in nature consisting of leucine residues. As per studies, CoV infected cells lead to activation of ATF6 pathways which in turn results in the up-regulation of ER chaperone proteins to combat ER stress (Fung and Liu, 2014; Sung et al., 2009).

7. Key SARS-CoV2 targets for vaccine development

Recent pandemic pneumonia outbreak COVID-19, in Wuhan, China is spreading globally and has raised an urgent public health issue worldwide impacting millions of people with a continuous increase in both morbidity and mortality. The causative agent of this disease is identified and named as SARS-CoV2 because of its genetic relatedness to SARS-CoV species that was responsible for the 2003 coronavirus outbreak. The immense spread of the disease in a very small period demands urgent development of therapeutic and prophylactic interventions for the treatment of SARS-CoV2 infected patients.

A vaccine is a biological preparation designed to protect humans from viral and bacterial infections. Vaccines stimulate the production of antibodies inside the human body before disease generation, in the same manner as antibodies are produced after the individuals are exposed to

---

**Table 3**

| Target protein | Accession | FASTA Sequence |
|----------------|-----------|----------------|
| Spike protein S2 subunit | 6LXT_F | GVTQNVLYENQKIAQNSAIGKIQDSLSTASALGKLDQIVNVQNAQALNTLYQKLSSN |
| Receptor Binding domain (RBD) | 6VW1_F | BVPSGIDYRVPRNPITCLCPIGEVENALKQPSVAYWERNKISNPCVDNVYNSTTSTFK |
| N-Terminal Domain (NTD) | 6YI3_A | CYGVSTKNLDSFNSVTAIDPHYVQDAAPQQTGVIAADYNLYLPLDFAWCGVLAWN |
| M protein | Q1Z97065 | MADSSNGTTVEEILKELQWNLVGGFLVTGWLCQLQAYANRNYFLVILYLFLWLLWPWVL |
| E protein | QHU36866 | MYSFVSEETGTLVILFSLAPVVLVFLVTLAIALRLACYCNVNVSLVKSFFYVYSRVKNLNSRVPDLLV |
| N protein | QIQ22768 | MSDINGQIQNQRNAPRITPGPSDDTSNQNGERSGARSQKRRQGQLPNATSWATLQ |
| PKR | | |
| PERK | | |
| ER stress | | |
| ATP6 | | |
| ER Folding capacity | | |
| dsRNA | | |
| PKR | | |
| PERK | | |
| ER stress | | |
| ATF6 | | |
| CoV infection | | |
| Translation attenuation | | |
| Apoptosis | | |

---

Fig. 5. Flowchart displaying CoV induced generation of ER stress and activation of the UPR signaling cascade.
the disease pathogen (Siegrist, 2013). In the last decade, vaccine development technology has evolved significantly, involving the formulation of cell-culture based vaccines (e.g., Flucelvax Tetra cell-based vaccine (Bühler and Ramharter, 2019)), licensed vectored vaccines ((e.g., Ervebo, a vesicular stomatitis virus [VSV]-vectorized ebola virus vaccine (Dumiak, 2019)), several DNA and RNA vaccine candidates and recombinant protein vaccines (e.g., a vaccine against hepatitis B virus (Huzair and Sturdy, 2017; Shang et al., 2020). There is limited knowledge of how the immune system of humans reacts naturally against SARS-CoV2. The fact that SARS-CoV2 displays a high level of homology to SARS-CoV, the target epitopes for vaccine development of SARS-CoV might prove to be effective for SARS-CoV2 as well (Hoffmann et al., 2020; Letko et al., 2020; Lu et al., 2020). Despite the high level of similarity between both, there are certain genetic variations as well. Studies reveal that only 16% of B cell epitopes and 23% of T cell epitopes of the SARS-CoV map identically to SARS-CoV2 (Ahmed et al., 2020). Key SARS-CoV2 target antigenic sites for vaccine development (Table 3) are enlisted below:

### Table 4
Percent identity matrix of major proteins and domains of novel coronavirus SARS-CoV2 strain with other beta-CoVs obtained using CLUSTAL O (1.2.4).

| PROTEIN                  | SARS-CoV | MERS-CoV | HCoV-OC43 | HCoV-HKU1 |
|-------------------------|----------|----------|-----------|-----------|
| S (spike)               | 97.71%   | 32.79%   | 30.50%    | 31.26%    |
| E (Envelope)            | 96.00%   | 36.00%   | 28.00%    | 20.00%    |
| M (Membrane)            | 89.59%   | 39.27%   | 35.29%    | 38.74%    |
| N (Nucleocapsid)        | 85.41%   | 46.47%   | 34.28%    | 35.20%    |
| Receptor (ACE-2) binding domain | 74.41%   | 18.75%   | 24.44%    | 22.83%    |
| N-terminal domain       | 52.55%   | 21.67%   | 21.49%    | 20.26%    |

7.1. Spike protein

Earlier vaccine development studies of SARS-CoV and MERS-CoV identified surface S glycoprotein of CoVs to be the most ideal target for vaccine development (Du et al., 2009; Schindewolf and Menachery, 2019). During host cell receptor interaction with the S glycoprotein of the CoV, S protein undergoes cleavage directed by host cell proteases and divides into two subunits: S1 and S2 (Huang et al., 2006; Kuo et al., 2000; Yamada et al., 2009). The S2 subunit then undergoes some conformational change to reveal the hidden fusion peptide that aids in fusing S protein to host cell membrane and thus making a path for viral RNA to enter the host cell cytoplasm (Masters, 2006). Since this surface glycoprotein (S protein) plays the most significant role in the initiation and spread of CoV infection inside an organism, therefore, S protein serves as the most vulnerable target for vaccine development of CoVs (Wrapp et al., 2020).

Developing vaccines against full length S protein is beneficial as full-length proteins can provide more target epitopes along with maintaining proper protein conformation and consequently higher immunogenicity (Pallesen et al., 2017; Zhang et al., 2020b). To date, a SARS-CoV-2S protein trimmer (S-Trimer) vaccine prepared using its patented Trimer-Tag® technology is reported to have been developed by Clover Biopharmaceuticals, which will be launched within next 4-5 weeks (Biopharmaceuticals, 2020). Other than full-length protein-based vaccines, specific regions of S protein such as S2 subunit, RBD, NTD, and FP also serve as potential targets for vaccine development (Fig. 6).

7.1.1. Spike S2 subunit

Studies reveal that B cell epitopes in the S2 subunit derived from S protein of SARS-CoV2 map identically to SARS-CoV and thus epitopes of this region might prove to be promising candidates for induction of protective antibody response and thus vaccine development. Preliminary studies for confirming the compatibility of the S2 subunit for vaccine development have already been carried out and the results depict promising nature of these epitopes in generating cross-reactivity and neutralizing antibodies (Ahmed et al., 2020).

7.1.2. Spike Receptor Binding Domain (RBD)

RBD is present in the S1 subunit of SARS-CoV2 and aids in binding the host cell receptor and then fusing the membrane of the host cell to the virus with the help of S2 subunit. SARS-CoV2 binds to the same host cell receptor i.e., ACE-2 (angiotensin-converting enzyme-2) as SARS-CoV (Brielle et al., 2020; Lan et al., 2020). Therefore, like the S2 subunit of the spike, the RBD region also makes a promising candidate as a target region for the development of vaccines against SARS-CoV2. As per the findings from a recent study, the SARS-CoV RBD specific antibodies were observed to cross-react with SARS-CoV2 protein, additionally, SARS-CoV2 was found to be cross-neutralized by induction of RBD SARS-CoV antisera, suggesting that RBD-SARS CoV based vaccines have potential to prevent SARS-CoV2 infection (Tai et al., 2020). In addition, the manipulation of ACE-2 receptor binding domain will prove to be a great therapy as it will ensure preventing viral entry into the host cell.
Fig. 7. (1). (A) MSA of KRSFIEDLFLNKV motif, (B) MSA of Receptor (ACE-2) binding domain of S1 subunit of spike protein, (C) MSA of Nucleocapsid (N) protein of various beta coronavirus strains representing sequences for SARS-CoV2, SARS, MERS, HCoV HKU1 and HCoV OC43. (2). Phylogenetic trees of (A) S protein and, (B) N protein of different beta CoVs. Both the trees demonstrate close homology of novel SARS-CoV-2 with SARS-CoV.
7.1.3. Spike N-Terminal Domain (NTD)
In several CoV species, NTD like RBD is reported to constitute carbohydrate receptor binding activity. Thus, NTD also serves as a candidate target region for vaccine development. A study demonstrated that an antibody that binds to the MERS-CoV S1 subunit NTD region was found to cross-neutralize the wild-type strain EMC of MERS-CoV (2012) (Chen et al., 2017), thus, depicting the potential of spike NTD region as potential antigen target. However, so far there is not much knowledge available about the function of S1-NTD of SARS-CoV2, a gain of higher understanding might demonstrate its interaction with certain receptors and make it a potential vaccine development target (Zhang et al., 2020b).

7.2. M protein
Membrane protein is the most abundant protein present on the surface of SARS-CoV2 and is involved in virion assembly in the cells. As specified by immunogenic and structural analysis, M protein is found to harbour a T cell epitope cluster that holds the potential to generate a strong cellular immune response (van der Meer et al., 1999), thus depicting the potential of M protein to serve as a target for vaccine development (Jin et al., 2005; Li et al., 2020a; Oh et al., 2012; Zhang et al., 2020b).

7.3. E protein
Envelope glycoproteins are small proteins that are composed of 75 amino acids in SARS-CoV2. The E proteins of CoVs along with M proteins play a crucial role in virion morphogenesis and their assembly within the cell (Siu et al., 2008). As per studies, the potential of recombiant SARS and MERS-CoV with mutated E protein live attenuated vaccines have been explored earlier (Graham et al., 2013; Schoeman and Fielding, 2019; Shang et al., 2020). Moreover, E protein is believed to be a chief virulence factor as knock-out of this protein results in reduced secretion of major inflammatory factors i.e., IL-1, TNF, and IL-6 (Nieto-Torres et al., 2014; Zhang et al., 2020b).

7.4. N protein
Nucleocapsid protein is an integral protein of SARS-CoV2 that aids in the formation of ribonucleoprotein complex also known as capsid by packaging viral genome inside the virion envelope. The formation of the capsid is important for viral self-assembly and replication (Jin et al., 2005; Li et al., 2020a). As specified by the data obtained from a study, 89% of SARS infected patients were able to produce antibodies to this antigen, depicting high antigenicity of N protein (Leung et al., 2004). However, previous studies demonstrate very fluctuating observations regarding the potential of N protein as a target antigen for vaccine development. As per the results obtained from a previous study, the SARS-CoV N protein DNA vaccine was able to induce the production of antibodies in vaccinated C57BL/6 mice and was thus used for the treatment of vaccinia virus (Kim et al., 2004). In contrast, a recent study on SARS-CoV2 claims N protein to be unsuitable for vaccine development because the antibodies induced by the N protein of SARS-CoV2 were found to be incapable of providing immunity to SARS-CoV infection (Gralinski and Menachery, 2020; Shang et al., 2020).

8. Conclusion
Recent pandemic novel coronavirus emerged from Wuhan, China has caused widespread fear and concern and has turned on a global public health security alarm. The direct zoonotic origin of SARS-CoV2 is not confirmed yet, however, genetic sequence studies have revealed bat-origin of this virus as it shares 96.2% genetic homology to bat SARS-related coronavirus strain (SARSr-CoV; RaTG13) procured from Yunnan province, China. Despite the high level of genetic similarity of novel SARS-CoV2 virus to other CoV strains, it is revealed that the SARS-CoV2 genome contains a few distinct characteristics like mutations in the S1-S2 region of the receptor-binding domain, polybasic furin cleavage site, and the presence of o-linked glycans. Although, targeting the regions of difference for the development of drugs and vaccines could be of great help as these genomic variations might be the reason for increased infectivity and severity of the disease, but since the pandemic is causing immense public health damage, it is the need of the hour to come up with a prophylactic or therapeutic intervention at the earliest. So, drug/vaccine repurposing is the current hot research area and researchers are aiming for the use of potential target antigen sequences of previously known coronavirus strains to come up with a suitable vaccine for novel SARS-CoV2 strain.

Additionally, S glycoprotein is responsible for mediating viral entry into human cells and thus aids in determining host range as well as viral infectivity. Restricting its cleavage by manipulating the expression of S protein or ACE-2 receptor binding motif might be a great therapy. Moreover, S protein is the most potential target antigenic site suggesting that RBD-SARS CoV based vaccines have potential to prevent SARS-CoV2 infection. Therapeutics, Moderna Inovio Pharmaceuticals, Clover Biopharmaceuticals, Novavax, Johnson & Johnson, Codagenix, are a few among the various companies working for the development of SARS-CoV2 vaccines.

Although immense global research is being conducted on this novel coronavirus strain, based on the genetic comparison of various coronavirus strains we have enlisted a few potential antigen target sites of SARS-CoV2 for the development of vaccines. More research and in-depth understanding of pathogenic mechanisms and genomic variability of SARS-CoV2 will further help in coming up with more targets for better therapy of COVID-19.

Declaration of Competing Interest
Authors declare no conflicts of interest.

Acknowledgement
The authors duly acknowledge the support provided by the Indian Council of Medical Research (ICMR), New Delhi, India, for providing fellowship to Ms.Navpreet kaur (Grant number- 5/4-5/188/Neuro/2019-NCD-1) and Panjab University, Chandigarh, for providing essential research facilities.

References
Ahmed, S.F., Quadeer, A.A., McKay, M.R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 12. https://doi.org/10.3390/v12030254, 2020.03.10.986398.
Andam Gaurav, M.A.-N., 2020. Polymerases of Coronavirus: Structure, Function, and inhibitors.
Anderson, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 2-4 https://doi.org/10.1038/s41591-020-0820-9, Baer, C.F., 2008. Does mutation rate depend on itself. PLoS Biol. 6, 0233–0235. https://doi.org/10.1371/journal.pbio.0060052.
Barz-Santos, Y., St. John, S., Mesecar, A., 2014. The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. Antiviral Res. 115 https://doi.org/10.1016/j.antiviral.2014.12.015.
Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V., Lai, M. M., 1993. Identification of the catalytic sites of a papain-like cysteine protease of murine coronavirus. J. Virol. 67, 6056-6063. https://doi.org/10.1128/jvi.67.6056-6063.1993.
Bechill, J., Chen, Z., Brewer, J.W., Baker, S.C., 2008. Coronavirus infection modulates the unfolded protein response and mediates sustained translational repression. J. Virol. 82, 4492–4501. https://doi.org/10.1128/jvi.00161-08, Biopharmaceuticals, C., 2020. Clover Initiates Development of Recombinant Subunit-Primere Vaccine for Wuhan Coronavirus (2019-nCoV).
Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication. Curr. Top. Microbiol. Immunol. 287, 1–30. https://doi.org/10.1007/3-540-26765-4_1, Brielle, E.S., Schneiderman, D., Linial, M., 2020. The SARS-CoV-2 everts a distinctive strategy for interacting with the ACE2 human receptor. bioRxiv. https://doi.org/10.1101/2020.03.10.986398, 2020.03.10.986398.

N. Kaur et al.
Infection, Genetics and Evolution 89 (2021) 104490
Infection, Genetics and Evolution 89 (2021) 104490

Buerly, I., Boursnell, M.E., Birnies, M.M., Billmoria, B., Blik, V.C., Brown, T.D., Inglis, S.C., 1987. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the Rous sarcoma virus. EMBO J. 6, 3779-3785. https://doi.org/10.1002/j.1460-2075.1987.tb02713.x.

Bühler, S., Ramharter, M., 2019. Fucovexa tetra: a surface antigen, inactivated influenza vaccine prepared in cell cultures. ESMO Open 4, 2018-2019. https://doi.org/10.1016/j.esmoop.2018.04.001.

Calfon, M., Zeng, H., Urano, F., Till, J.H., Hubbard, S.R., Harding, H.P., Clark, S.G., Ron, D., 2002. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. Nat. 415, 92-97. https://doi.org/10.1038/41092a.

Carl, T., Galli, C., Olivari, S., Molinari, M., 2008. Segregation and rapid turnover of EDEM1 by an autophagy-like mechanism modulates standard ERAD and folding activities. Biochem. Biophys. Res. Comm. 371, 405-410. https://doi.org/10.1016/j.bbrc.2008.04.096.

Cavanagh, D., 2007. Coronavirus avian infectious bronchitis virus. Vet. Res. 38, 281-297. https://doi.org/10.1051/vetres:2006055.

Ceraolo, C., Giorgi, F.M., 2020. Genomic variance of the 2019-nCoV coronavirus. J. Med. Virol. 92, 522-528. https://doi.org/10.1002/jmv.25790.

Chan, C.-M., Woo, P.C.Y., Lau, S.K.P., Tse, H., Chen, H.-L., Li, F., 2008. Spike Protein, S, of Human Coronavirus HKU1 : Role in Viral Life Cycle and Application in Antibody Detection, pp. 1527-1536. https://doi.org/10.1186/1866-0247-5-197.

Chan, J.F.W., Kok, K.H., Zhu, Z., Chu, H., To, K.K.W., Yuan, S., Yuen, K.Y., 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg. Microbes Infect. 9, 221-236. https://doi.org/10.1002/emmicro.202175.2020.1719902.

Charley, B., Laude, H., 1988. Induction of alpha interferon by transmissible gastroenteritis coronavirus: role of transmembrane glycoprotein E1. J. Virol. 62, 8-11. https://doi.org/10.1128/jvi.62.1.8-11.1988.

Chen, Y., Lu, S., Su, H., Deng, Y., Zhou, J., Huang, B., Yu, Y., Lan, J., Wang, W., Lou, Q., Tan, W., 2021: A novel neutrophilic monocytic antibody targeting the N-terminal domain of the MERS-CoV spike protein (emerging microbes & infections (2017) 6 5 (e37)). Emerg. Microbes Infect. 6, e60. https://doi.org/10.1002/omim.20175.00.

Chen, Y., Liu, Q., Dong, N., Yang, X.,Ye, L., Chen, K., Chan, E.W.-C., Chen, S., 2020. Genomic and protein characterization of the novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg. Microbes Infect. 9, 221-236. https://doi.org/10.1002/emmicro.202175.2020.1719902.

Cheng, V.C.C., Hung, I.F.N., Tang, B.S.F., Chu, C.M., Wong, M.M.L., Chan, K.H., Wu, A.K., Tse, D.M.W., Chan, K.S., Zheng, B.J., Peiris, J.S.M., Sung, J.J.Y., Yuen, K.Y., 2004. Virus replication in cells expressing the M protein of Human Coronavirus HKU1 : Role in Viral Life Cycle and Application in Antibody Detection, pp. 1527-1536. https://doi.org/10.1186/1866-0247-5-197.

Cheng, V.C.C., Hung, I.F.N., Tang, B.S.F., Chu, C.M., Wong, M.M.L., Chan, K.H., Wu, A.K.L., Tse, D.M.W., Chan, K.S., Zheng, B.J., Peiris, J.S.M., Song, J.Y.Y., Yuan, K.Y., 2004. Virus replication in cells expressing the M protein of Human Coronavirus HKU1 : Role in Viral Life Cycle and Application in Antibody Detection, pp. 1527-1536. https://doi.org/10.1186/1866-0247-5-197.

Cheng, V.C.C., Hung, I.F.N., Tang, B.S.F., Chu, C.M., Wong, M.M.L., Chan, K.H., Wu, A.K.L., Tse, D.M.W., Chan, K.S., Zheng, B.J., Peiris, J.S.M., Song, J.Y.Y., Yuan, K.Y., 2004. Virus replication in cells expressing the M protein of Human Coronavirus HKU1 : Role in Viral Life Cycle and Application in Antibody Detection, pp. 1527-1536. https://doi.org/10.1186/1866-0247-5-197.

Chesnutt, G., Gao, L., Besser, R.E., 1997. Identification of the viral protein recognized by antibodies in the sera of convalescent cases. Virology 324, 251-256. https://doi.org/10.1006/viro.2001.1099.

Chen, H., Yang, X.,Long, Y., Li, W., Liu, G., 2019. Alpha anti-SARS-CoV-2 IgM antibodies from COVID-19 convalescent patients. J. Intern. Med. 286, 350-358. https://doi.org/10.1111/joim.13220.

Cheng, V.C.C., Hung, I.F.N., Tang, B.S.F., Chu, C.M., Wong, M.M.L., Chan, K.H., Wu, A.K.L., Tse, D.M.W., Chan, K.S., Zheng, B.J., Peiris, J.S.M., Song, J.Y.Y., Yuan, K.Y., 2004. Virus replication in cells expressing the M protein of Human Coronavirus HKU1 : Role in Viral Life Cycle and Application in Antibody Detection, pp. 1527-1536. https://doi.org/10.1186/1866-0247-5-197.

Chen, H., Yang, X.,Long, Y., Li, W., Liu, G., 2019. Alpha anti-SARS-CoV-2 IgM antibodies from COVID-19 convalescent patients. J. Intern. Med. 286, 350-358. https://doi.org/10.1111/joim.13220.

Chen, H., Yang, X.,Long, Y., Li, W., Liu, G., 2019. Alpha anti-SARS-CoV-2 IgM antibodies from COVID-19 convalescent patients. J. Intern. Med. 286, 350-358. https://doi.org/10.1111/joim.13220.
Kim, T.W., Lee, J.H., Hung, C.-F., Peng, S., Roden, R., Wang, M.-C., Vicidivi, R., Tsai, Y.-C., Hu, L., Chen, P.-J., Boyer, D.A.K., Wu, T.-C., 2004. Generation and characterization of RNA viruses by utilizing the Nocodazol-sensitive cell line of severe acute respiratory syndrome coronavirus. J. Virol. 78, 4638-4645. https://doi.org/10.1128/jvi.78.12.4638-4645.2004.

Kiyono, J., Li, Y., Yumoto, H., Taoka, T., Takahashi, T., Yoshida, Y., Kato, K., 2004. Human poliovirus type 2 spike-mediated entry. J. Virol. 78, 6723-6731. https://doi.org/10.1128/JVI.78.12.6723-6731.2004.

Li, G., Yang, Y., Wang, X., Zhao, W., Zhai, Y., Wu, J., 2019. Identification of SARS-like Coronavirus from Patients with Pneumonia in China, 2019, pp. 727-6736(20)30251-8.

Luo, J., Guo, H., Shi, L., Chen, S., Dong, W., Lu, F., 2007. Discovery of a novel coronavirus 3′–5′ exoribonuclease that is critically involved in coronavirus RNA replication. Proc. Natl. Acad. Sci. U. S. A. 104, 5108-5113. https://doi.org/10.1073/pnas.0508200103.

Morse, S.S., Schlueterberg, A., 1990. From the National Institute of Allergy and Infectious Diseases, the Fogarty International Center of the National Institutes of Health: Emerging viruses: the evolution of viruses and viral diseases. J. Infect. Dis. 162, 1-7. https://doi.org/10.1093/infdis/162.1.1.

Morse, J.S., Lalonde, T., Xu, S., Liu, W.R., 2020. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. ChemioNanBioMed 21, 730-738. https://doi.org/10.21208/cnbmed.202009047.

Nal, B., Chan, C., Kien, F., Shi, L., Sun, J., Chu, K., Kam, J., Staopeni, sabelle, Cesnzeno, Chaigne, B., Escriou, N., van der Wef, S., Yuen, K.Y., Altmeyer, R., 2005. Differential recognition and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S and E. J. Exp. Med. 196, 1129-1139. https://doi.org/10.1084/jem.20050581.

Nao, N., 2017. Genetic Predisposition To Acquire A Polyclonal B-Cell repertoire For Highly Pathogenic Avian Influenza Virus Hemagglutinin Nagnag. 8, pp. 1-15.

Nieto-Torres, J.L., DeDiego, M.L., Verdia-Bajusca, C., Jimenez-Guardino, J.M., Regla-Nava, J.A., Fernandez-Delgado, R., Castrano-Rodriguez, C., Alcaraz, A., Torres, J., Aguillema, V.M., Enjuanes, L., 2014. Severe acute respiratory syndrome coronavirus envelope protein Ion Channel activity correlates virus Fitness and pathogenesis. PloS Path. 10 https://doi.org/10.1371/journal.ppat.1004477.

Oh, H.L.J., Ge, S.K., Bertoletti, A., Tan, Y.J., 2012. Understanding the T cell immune response in SARS coronavirus infection. Emerg. Microbes Infect. 1, 0. https://doi.org/10.1038/emi.2012.1.

Oon, M., Ge, S.K., Bertoletti, A., Tan, Y.J., 2012. Understanding the T cell immune response in SARS coronavirus infection. Emerg. Microbes Infect. 1, 0. https://doi.org/10.1038/emi.2012.1.

Ostros, M., de Linto, E.G., Dejik, J., Verheije, M.H., Rottier, P.J.M., de Haan, C.A.M., 2007. Localization and membrane topology of coronavirus nonstructural protein 4: identification of the early secretory pathway in replication. J. Virol. 81, 12325-12336. https://doi.org/10.1128/JVI.01506-06.

Palleonen, J., Wang, N., Corbett, K.S., Wraap, D., Kirchdoerfer, R.N., Turner, H.L., Cottrell, A.A., Becker, M.M., Wang, L., Shi, W., Kong, W.P., Andres, E.L., Ketenbach, A.N., Denison, M.R., Chappell, J.D., Graham, B.S., Ward, A.B., McLellan, J.S., 2017. Identification and structure of a genetically engineered prefusion MERS-CoV spike antigen. Proc. Natl. Acad. Sci. U. S. A. 114, E7348-E7357. https://doi.org/10.1073/pnas.1707304114.

Peck, K.M., Lauring, A.S., 2018. Complexities of viral mutation rates. J. Virol. 92, 1–8. https://doi.org/10.1128/jvi.00476-17.

Prentice, E., Jerome, W.G., Yoshimori, M., Denison, N.M., 2004. Coronavirus replication complex formation utilizes components of cellular autophagy. J. Biol. Chem. 279, 10316-1041. https://doi.org/10.1074/jbc.M405462200.

Puthalakath, H., Orr, D., McElmam, E., McKim-McKinn, J., Motisa, N., Yotomi, V., Goore, T., Aiura, S., Pseudor, A., Wsler, A., 2007. ER stress triggers apoptosis by activating BH3-only protein Bim. Cell 129, 1337-1349. https://doi.org/10.1016/j.cell.2007.04.027.

Qiu, Z., Hingle, S.T., Simmons, G., Yu, C., Das Sarma, J., Bates, P., Weiss, S.R., 2006. Endosomal proteolysis by cathepsins is necessary for murine coronavirus murine Hepatitis virus type 2 spike-mediated entry. J. Virol. 80, 5768-5776. https://doi.org/10.1128/jvi.00442-06.

Reigoni, F., Monastyrskaya, I., Verheije, M.H., Cali, T., Ulassi, M., Bianchi, S., Bernesoni, R., De Haan, C.A.M., 2010. LC3-I-to-LC3-II conversion is required for LC3-I-positive MERS-CoV spike antigens. Proc. Natl. Acad. Sci. U. S. A. 114, 5051-5056. https://doi.org/10.1073/pnas.0710501103.

Roth, Z., Yehezkel, G., Khalaila, I., 2012. Review article identification and quantification of viral proteins of the severe acute respiratory syndrome coronavirus are required for viral replication. J. Cell Biol. 124, 574-578. https://doi.org/10.1016/S0021-9247(74)90170-0.

Schoeman, D., Fielding, B.C., 2019. Coronavirus envelope protein: current knowledge. Virol. 92, 424-432. https://doi.org/10.1016/j.jsvi.2019.04.007.

Siegrist, C.-A., 2013. Vaccines - CH 2 how do vaccines mediate protection? Vaccines 1, 1-8. https://doi.org/10.3390/vaccines1010001.

Simoensh, C., Menachery, V.D., Dinnon, K.H., Yount, B.L., McAnarney, E.T., Gralinski, L.E., Hale, A., Gralinski, L.E., Schuknecht, H.F., Hooton, T.Y., Wang, S., Tsai, Y.-N., Kaur et al. 7641 – RNA synthesis indicates a role for late endosomes in viral replication. J. Virol. 73, 193-202. https://doi.org/10.21208/cnbmed.202009047.

Siegrist, C.-A., 2013. Vaccines - CH 2 how do vaccines mediate protection? Vaccines 1, 1-8. https://doi.org/10.3390/vaccines1010001.
efficiency, targeting and trafficking, and release of virus-like particles. J. Virol. 82, 11318–11330. https://doi.org/10.1128/JVI.01532-17.

Song, W., Wang, Y., Wang, N., Wang, D., Guo, J., Fu, L., Shi, X., 2014. Identification of residues on human receptor DPP4 critical for MERS-CoV binding and entry. Virology 471, 49–53. https://doi.org/10.1016/j.virol.2014.04.031.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y., Gao, G.F., 2016. Wang, N., Yang, S.L.X., Zhang, H.H.Y., Guo, H., Luo, C., Miller, M., Zhu, G., Chmura, A. Tao, Z., Tian, J., Pei, Y., Yuan, M., Zhang, Y., Dai, F., 2020. A new coronavirus associated with human respiratory disease in China. https://doi.org/10.1038/s41586-020-0289-1.

Tao, Z., Tian, J., Pei, Y., Yuan, M., Zhang, Y., Dai, F., 2020. A new coronavirus associated with human respiratory disease in China. https://doi.org/10.1186/s12985-015-0289-1.

Wang, M.N., Zhang, W., Gao, Y.T., Hu, B., Ge, X.Y., Yang, X., Lou, Z., Zhang, Y.Z., Shi, Z.L., 2007. Comparative analysis of twelve genomes of three novel group 2c and 2d coronaviruses reveals unique group and subgroup features. J. Virol. 81, 1574–1585. https://doi.org/10.1128/JVI.01282-06. Wo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lai, K.K.Y., Huang, Y., Lee, P., Luk, G.S.M., Dyrting, K.C., Chan, K.-H., Yuan, K.-Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83, 908–917. https://doi.org/10.1128/JVI.01977-08.

Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L., Zou, Q., 2020b. Progress and prospects on vaccine development against SARS-CoV-2. Vaccines 8, 153. https://doi.org/10.3390/vaccines8030112.

Zhao, Z., Li, H., Wu, X., Zhong, Y., Zhang, K., Zhang, Y.P., Boerwinkle, E., Fu, Y.X., 2004. Moderate mutation rate in the SARS coronavirus genome and its implications. BMC Evol. Biol. 4, 1–9. https://doi.org/10.1186/1471-2148-4-1.

Zheng, X.S., Liang, X., Wang, Z., 2008. Transcriptional profiling of Vero E6 cells over-expressing SARS-CoV S2 subunit. Insights on virus infection, transcription and humoral response. Prog. Biol. 15, 518–534. https://doi.org/10.1016/j.probio.2008.08.008.

Yeager, Curtis L., Ashmun, Richard A., Williams, Richard K., Cardellichio, Christine B., Piotrowski, Paul J., Yeager, Debra A., 2008. The SARS coronavirus nucleocapsid protein. Virology 372, 59–71. https://doi.org/10.1016/j.virol.2007.11.036.

Verheije, M.H., Raaben, M., Mari, M., Te Lintelo, E.G., Reggiori, F., Van Kuppeveld, F.J.M., 2012. Closing the gap: the challenges in converging theoretical, computational, and experimental research. Viruses 4, 1–19. https://doi.org/10.3390/v4041429.

Xie, Q.M., Mani, S., Zheng, X.S., Li, B., Li, J.M., Guo, H., Pei, G.Q., An, X.P., Chen, J. 2009. SARS-CoV is a luminal ER membrane-associated protein and induces the activation of endoplasmic reticulum stress. Nat. Cell Biol. 13, 184–190. https://doi.org/10.1038/ncb1925.

Yamada, Y., Liu, D.X., 2007. Comparative analysis of twelve genomes of three novel group 2c and 2d coronaviruses reveals unique group and subgroup features. J. Virol. 81, 1574–1585. https://doi.org/10.1128/JVI.01282-06.

Woo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lai, K.K.Y., Huang, Y., Lee, P., Luk, G.S.M., Dyrting, K.C., Chan, K.-H., Yuan, K.-Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83, 908–917. https://doi.org/10.1128/JVI.01977-08.

Zhou, L., Mai, K.J., Wu, Z.X., Li, D., Anderson, D.E., Zhang, L.B., Li, S.Y., Mi, Z., He, T.T., Cong, F., Guo, P.J., Huang, R., Luo, Y., Liu, X.L., Chen, J., Huang, Q., Sun, Z., Xing, S.Z., Chen, Y.S., Sun, Y., Li, J., Daszak, P., Wang, L.F., Shi, Z.L., Tong, Y.G., Ma, J.Y., 2018. Fatal swine acute diarrhoea virus infection in humans, China. J. Virol. 93, 104–107. https://doi.org/10.1128/JVI.02501-05.
syndrome caused by an HKU2-related coronavirus of bat origin. Nature 556, 255–259. https://doi.org/10.1038/s41586-018-0010-9.
Zhou, P., Yang, X., Wang, Xian-Guang, Hu, B., Zhang, L., Zhang, W., Guo, H., Jiang, R., Liu, M., Chen, Y., Shen, X., Wang, Xi, Zhan, F., Wang, Y., Xiao, G., Shi, Z., 2020a. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2019. https://doi.org/10.1038/s41586-020-2012-7.
Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E. C., Hughes, A.C., Bi, Y., Shi, W., 2020b. A novel bat coronavirus reveals natural insertions at the S1/S2 cleavage site of the Spike protein and a possible recombinant origin of HCoV-19. bioRxiv. https://doi.org/10.1101/2020.03.02.974139, 2020.03.02.974139.