Review Article

Recent research progress and current understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Running title: Current understanding of SARS-CoV-2

Abstract

Coronaviruses (CoVs) are a large group of enveloped viruses with a positive-sense RNA that have characteristic spikes that project from their surface. CoVs are well known for their large RNA genome (26-32 kb). They primarily affect mammals and birds, causing infections of the respiratory and gastrointestinal tracts. The emergence of human CoVs (HCoVs) has been reported once every ten years for the last three decades. The most recent emergence occurred in December 2019, when a new strain of CoVs named SARS-CoV-2 caused the coronavirus disease 2019 (COVID-19) pandemic, leaving a devastating impact on the global healthcare. The early cases were associated with the Huanan seafood market in Wuhan, although the exact origin of the virus is still being debated. Phylogenetic analysis reveals bats to be the reservoir hosts, but the intermediate host responsible for spill-over into the human population remains debatable. Accumulating evidence cites pangolins based on the similarity of receptor binding domain in spike protein; however, the search for a conclusive intermediate host that aided in the inter-species crossover is still underway. Advances have been made in our understanding of the functions of each structural protein, but certain non-structural proteins and accessory proteins are yet to be characterised. Owing to the large genetic diversity of CoVs that arise through recombination, genetic variation, or gene gains/losses, future re-emergence of CoVs are most likely. In this review, we provide an introduction to CoVs and discuss the origin, virology, genetics, phylogeny, and pathogenesis of SARS-CoV-2 based on relevant literature.

Keywords: 2019-nCoV, coronavirus, HCoV, SARS-CoV-2

Introduction

Coronaviruses (CoVs) belong to a family that contain the largest known RNA viral genome. They are spherical and enveloped viruses, whose size ranges between 80-120 nm in diameter[1]. CoVs primarily affect mammals causing infections of the respiratory and gastrointestinal tracts. They are usually known to cause the seasonal common cold in humans and were not given much importance until the outbreaks of the severe acute respiratory syndrome (SARS) in 2003, middle eastern respiratory syndrome (MERS) in 2012, and the most recent coronavirus disease in 2019 (COVID-19)[2]. A new strain of CoVs called the '2019 novel coronaviruses' (2019-nCoV) or more formally, 'severe acute respiratory syndrome coronavirus 2' (SARS-CoV-2) was found to be causative of the COVID-19 outbreak[3].

The SARS epidemic, caused by the SARS-CoV virus, occurred in Guangdong province, China. Affecting 8096 people and causing 774 deaths in 29 countries, SARS had a mortality rate of ~9%[4]. Comparative genomic analysis indicated Chinese horseshoe bats to be the reservoir host, owing to the high similarity between bat SARS-like CoV (SL-CoV) and SARS-CoV[5]. Further analysis confirmed *Paguma larvata* civet to be the 'amplifying' or intermediate host[6], [7]. The SARS epidemic was controlled successfully in 2003 and there has been no recurrence since then. Ten years after the SARS epidemic, the highly virulent MERS-CoV, emerged in the Middle East countries in 2012, causing the MERS epidemic. The outbreak affected over 2519 individuals, causing close to 866 deaths and had a mortality rate of 34.3%[4]. Genome analysis identified bats to be the likely reservoir; however, the natural separation between bats and human population indicated the presence of an intermediate host. The close relation of MERS-CoV to CoV of dromedary camels and the replication of MERS-CoV in camel cell-lines provided evidence for the intermediate host[8], [9]. MERS-CoV did not spread quite widely and the number of cases did not, therefore, accelerate considerably.

Comment [S1]: This paragraph has been moved here to the introduction. This was originally a separate section, but I have added it here, according to the reviewer's suggestion.

Less than a decade after the MERS epidemic, the emergence of SARS-CoV-2 responsible for the COVID-19 outbreak sent shock waves across the globe. The first cases were found to be associated with a seafood market in Wuhan city, Hubei Province of China[10],[11]. What started as a cluster of patients with idiopathic pneumonia in the local hospitals in Wuhan soon resulted in a massive explosion in the number of cases globally. To date, the aetiology of the infection remains unclear, as there are controversies circling the origin of SARS-CoV-2 and the intermediate host responsible for their spill-over to the human population. Phylogenetic analysis revealed a close similarity of SARS-CoV-2 to a bat CoV (RaTG13)[12]. As of August 31, 2020, SARS-CoV-2 has infected 20.1 million people across 188 countries, causing 819,934 deaths[13]. Although the mortality rate to date is markedly lower (~3%) than SARS-CoV and MERS-CoV, SARS-CoV-2 has the capacity to spread undetected within the population[14], placing a massive burden on the global healthcare.

Comment [S2]: As suggested, I've inserted the reference supporting the sentence.

This review aims to provide an overview on our current understanding of the SARS-CoV-2 virus, by covering various aspects such as the origin, virology, phylogeny, genetics, and pathogenesis based on the existing relevant literature.

Origin

The early cases of SARS-CoV-2 outbreak were reported to have occurred in late December, but retrospective analyses have identified the first patient to have shown symptoms as early as December 1, 2019. Since then, the number of confirmed cases has been on a steady incline. However, the total number of reported cases is expected to be markedly lower than the actual number, as many pre-symptomatic and asymptomatic cases are likely to be unintentionally excluded from the final count. The unforeseen emergence and sudden worldwide spread of SARS-CoV-2 provides a clue to their origin from a family of viruses that has the propensity to cross species barrier. Human coronaviruses (HCoVs) are generally zoonotic in nature and bats are the reservoir for all known HCoVs[15]. Phylogenetic analysis validated the zoonotic origin of SARS-CoV-2 and bats were found to be the natural host, based on the 96.3% nucleotide homology of SARS-CoV-2 with RaTG13, a *Rhinolophus*

affinis bat CoV[12]. Based on such close similarity, researchers suggested bats could have been responsible for the direct transmission to humans. However, SARS-CoV-2 and RaTG13 were later reported to differ in crucial spike receptor-binding domain (RBD) that is critical for viral entry into the cell, suggesting that they may not bind efficiently to the human receptor[16]. Since then several studies have attempted to identify the intermediate host responsible for the spill-over to the human population. A study reported that the RBD of SARS-CoV-2 resembled closely to pangolin-CoVs differing in only one amino acid, when tested against *Manis javanica*[17]. Another study identified a high similarity of codon usage of SARS-CoV-2 with that of snakes[18], but the results were mostly overlooked as this method of determining the host was questionable. To check if domestic animals may have played a role in the transmission, a study analysed the replication of SARS-CoV-2 in dogs, ferrets, cats, pigs, ducks, and chickens[19]. Of these, ferrets and cats were found to effectively host viral replication compared to other tested animals. Moreover, cats were highly susceptible to the airborne transmission of SARS-CoV-2. Despite several studies, there is no conclusive identification and efforts are still underway since the early days of the outbreak to identify the intermediate host that helped in the transmission of SARS-CoV-2 to humans.

Virology

SARS-CoV-2 is an enveloped virus that is sized between 80-120 nm diameter, carrying a large RNA genome. The most characteristic feature of this virus is the presence of spikes (~20 nm in length) that projects from the surface[20]. SARS-CoV-2 contains four main structural proteins - spike surface glycoprotein (S), an envelope glycoprotein (E), nucleocapsid (N), and membrane/matrix protein (M) (Fig 1)[21]. Spike glycoprotein is a transmembrane protein that forms a homotrimer, but is predominantly seen in the outer surface of the virus. They have a molecular weight of ~150 kDa. One major function of S protein of SARS-CoV-2 is to bind to the angiotensin-converting enzyme 2 (ACE2) host receptor mostly seen in respiratory and gastrointestinal tracts; the binding facilitates the entry of the virus into the host cell. The N-terminal domain (NTD) of S protein harbours a S1 sub-unit and the C-terminal domain (CTD) hosts a highly conserved S2 sub-unit. S1 has the RBD and, therefore, functions to determine host tropism while S2 has two tandem domains (Heptad repeats (HR) 1 and 2) and is responsible for the fusion of viral and host membranes[22]–[24]. The second structural protein, envelope (E), is the smallest of the four. They have a molecular weight of about ~8-10 kD consisting of 110 amino acids. They are an integral protein with a hydrophobic domain[25]. They play an established role in the assembly and maturation of SARS-CoV-2 post-replication[26]. Some in-vitro studies also indicate that E protein of CoVs can inhibit the stress response of host cell to ensure successful establishment of infection[27]. Another structural protein is the highly-conserved nucleocapsid (N) protein has a molecular weight of ~46 kDa and have over 400 amino acids[28]. The main function of N is to dynamically bind to the RNA to form a highly ordered conformation[29]. The NTD of CoV N protein has many positively charged amino acids that was found to help in RNA-binding, while the CTD is predicted to function in nuclear localisation[28]. The N protein of CoVs is crucial for processes associated with replication and transcription. Finally, the M protein is the most structured of all four proteins. Having 3 transmembrane domains, the M protein is considered to give CoVs their shape[30]. It also plays an important role in stabilising the nucleocapsid-RNA complex[31]. Fig. 2 shows the S, E, and N structural proteins. Apart from the four structural proteins, SARS-CoV-2 also has 15 non-structural proteins (NSPs 1-10 and 12-16) and 8 accessory proteins[21]. The function of the 15 NSPs of SARS-CoV-2 and their similarity to SARS-CoV are given in Table 1[32]–[34].

Genome characterisation

CoVs are a large group of viruses that are part of order *Nidovirales.* SARS-CoV-2 belongs to the *Coronaviridae* family of the *Nidovirales* order and is a member of the *Orthocoronavirinae* sub-family[11], [35]. Based on the genomic composition, the sub-family is further classified into four genera: Alphacoronavirus (αCoV), Betacoronavirus (βCoV), Gammacoronavirus (γCoV), and Deltacoronavirus (δCoV). The members of the αCoV and βCoV primarily infect mammals, whereas γCoV and δCoV are known pathogens of avian species[36]. Since the early 1960s, six types of

HCoVs have been reported. Four of these are mild pathogens (NL63, 229E, OC43, and HKU1) that belong to αCoV, while the other two (SARS-CoV and MERS-CoV) belong to βCoV and are highly pathogenic[37], [38]. SARS-CoV-2, is also a member of the βCoV genera[21]. γCoVs like infectious bronchitis coronavirus (IBV) truly affect avian species, while δCoV like the porcine delta coronavirus (PdCV) can affect both avian and mammalian species[39].

In early January, the genome of SARS-CoV-2 was sequenced with samples from nine patients in the Chinese Centre for Disease Control and Prevention (CCDC)[35]. The sequencing showed that SARS-CoV-2 is a single-stranded positive-sense RNA genome with a 5' cap and a 3' poly A tail that is approximately 29.8 kb in size. A 99.98% sequence similarity was observed between the eight complete genomes from nine patients, with differences in only 4 nucleotides (nt) of the ~30kb long sequence. Such near-identical sequence homology is suggestive of a recent species-crossover event. Apart from the 5' and 3' untranslated regions (UTRs) that constitute 265nt and 229nt respectively, the genome has 11 open reading frames (ORFs) that encode 27 proteins. The orf1a/b is a little over 20 kb and constitutes about 67% of the viral genome, encoding 15 NSPs. The rest of the genome (~10 kb) is annotated to code for 4 structural proteins and 8 accessory proteins[21], [32], [40] (Fig 3).

Each of the structural and accessory genes have a transcription regulatory sequence (TRS) at the beginning of their sequences, which helps in the expression of these genes. The structural proteins are the S, E, M, and N and the accessory proteins are the orf3a, orf3b, orf6, orf7a, orf7b, orf8b, orf9b, and orf14[21]. The ORF located near the 5' end (orf1a/b) codes for pp1ab and pp1a proteins that constitute the 15 NSPs. The expression of these NSPs happen by ribosomal frameshifting. The order of the genome is highly conserved throughout *Nidovirales* with the ORF1ab gene preceding the structural and accessory genes[41].

The single-stranded nature of CoVs makes them more susceptible to mutation, despite having NSP14 for their basic proofreading ability[42]. Owing to their high susceptibility to mutations, there have been concerns revolving the strains of SARS-CoV-2; whether more than a single strain has led to COVID-19. In a network-based phylogenetic analysis of 160 complete SARS-CoV-2 genomes from the GISAID NCBI database (www.gisaid.org), three main variants (A, B, and C) differing in their amino acid sequences were noted[43]. The RaTG13 bat CoV was chosen to be the outgroup. The cluster of lineages directly under this outgroup were labelled A (for ancestral) and were mostly seen in the United States, Europe, and Australia. Type B genomes were found predominantly in East Asia and were mutationally derived from Type A. Type C was found to be a derivative of Type B by nonsynonymous mutations and was majorly found across Europe and some parts of America (California and Brazil). However, apart from the three primary variants studied from the viral network in the early stages of the outbreak, SARS-CoV-2 may have evolved further and may be responsible for the difference in spread and severity of the disease between countries. In another study, 86 complete or partially complete genomes of SARS-CoV-2 were analysed[44]. The samples included viral strains from the 12 countries. Among the 86 genomes, 3 deletions (3nt, 10nt, and 24nt) and 93 substitution mutations were observed in the samples from the United States, Japan, and Australia. The deletions, notably the 3nt and 24nt, occurred in the orf1a/b polyprotein. Their effects on the functions of the proteins remain to be studied. The third deletion (10nt) occured near the 3' end. Among the 93 substitutions, 42 missense mutations were present in all major structural proteins (excluding E) and NSPs. Three mutations occurred in the RBD of S protein and these may induce conformational changes that can, in turn, lead to a changed antigenicity of the virus[45].

Evolution and phylogenetic analysis

Zoonotic HCoVs emerging from wild animals rapidly evolve through various mechanisms such as genetic variations, recombination, or even gene gains/losses altogether[39]. The large genomic size of CoVs predisposes them to frequent recombination, making it a common method of evolution. Recombination has been reported in both SARS-CoV and MERS-CoV. An in-silico analysis revealed recombination in the orf1ab region and S protein of SARS-CoV with 6 other CoVs[46]. In MERS-CoV,

recombination was seen in 28 regions with other camel CoVs[47]. However, recombination in SARS-CoV-2 still remains to be studied. Gene gains/losses is another mechanism that accounts for the divergence of CoVs[39]. As an example of gene loss, SARS-CoV-2 completely lacks the haemagglutinin-esterase (HE) and orf8a genes present in SARS-CoV, with which it shares ~80% homology. There were also minor changes seen at amino acid level. For example, orf3b has over 100 amino acids in SARS-CoV, but only 22 amino acids in SARS-CoV-2[21], although the importance of these in pathogenesis remains to be studied. To understand the origin of SARS-CoV-2 and better trace its evolution, many phylogenetic analyses were conducted by several groups of researchers. Based on one such analysis by whole genome sequence clustering, SARS-CoV-2 clustered with members of sub-genus *Sarbecovirus*, being ~80% identical with SARS-CoV, ~89% with ZXC21, a SARS-like CoV, and ~96% with bat SARS-like viruses[3], [11]. Since the S protein is primarily responsible for viral entry into the host cell, the difference in S gene was studied by aligning sequenced samples of SARS-CoV-2 and SARS-CoV against the reference S gene of SARS-CoV-2[12]. On analysis, it was found that SARS-CoV-2 aligned with a near 100% identity and a complete 100% coverage with the reference gene, and SARS-CoV aligned with 74.5% identity and 91% coverage. Furthermore, 76.5–83% identity was shared between the S protein gene of bat SARS and SARS-like coronavirus with SARS-CoV-2. Having such a high similar identity and coverage possibly suggest a recent common ancestor for SARS-CoV-2. In another study, SARS-CoV-2 samples from 18 countries were phylogenetically studied in comparison with MERS-CoV, SARS-CoV, bat SARS-CoV, bat SARS-like CoV, and other HCoVs[48]. The study showed that in all samples, SARS-CoV-2 clustered closely with bat SARS and other SARS-like CoVs, especially with Wuhan bat CoV RaTG13 and bat SARS-like CoVs (bat-SL-CoVZC45 and bat-SL-CoVZXC21), supporting the findings of a similar study conducted earlier[35]. MERS-CoV was found to be closely clustered with HCoVs - HKU4 and HKU5, but was distant from SARS-CoV-2.

Comment [S3]: A section titled 'Emergence of SARS-CoV and MERS-CoV' has been removed here. A brief version of the original content has been added to the introduction. Therefore, Table 2 has been removed - I have already summarised it in the text.

Pathogenesis

SARS-CoV-2 is thus far the most infectious HCoV known; the number of confirmed cases dwarf the infection caused by SARS-CoV and MERS-CoV. The highly infective potential of SARS-CoV-2 is determined by the establishment of successful infection in the host cell and is partly responsible for the rapid spread of the virus. The first step for a successful pathogenesis is the entry of virus into the cell, for which CoVs bind to the cell surface receptor in the host cell with their S protein. There are several studies focusing on the viral entry mechanisms of SARS-CoV and SARS-CoV-2[3], [49], [50]. Both viruses bind to the human ACE2 receptors (hACE2)[51]. However, there were subtle differences in the RBD of S protein between the two, which increased the affinity of RBD binding of SARS-CoV-2 to ACE2[52]. The RBD in both SARS-CoV and SARS-CoV-2 is quite dynamic, as they switch between two conformational states - a vertical position (the 'standing-up' phase) for binding to the host receptor and a horizontal position (the 'lying-down' phase) which is a feature of immune evasion[53], [54]. S proteins are typically class I viral fusion proteins and proteolytic cleavage of S into S1 and S2 is crucial for their activation[55]. A hallmark of the S protein of SARS-CoV-2 is the presence of an insert of four amino acids (PRRA) between the S1 and S2 sub-units[56]. While this was not present in SARS-CoV, it was found in a bat-derived CoV RmYN02[57]. This insert in SARS-CoV-2 has been shown to act as a cleavage site for furin protease. MERS-CoV has a furin activation mechanism to mediate membrane fusion. Other proteases like TMPRSS2 or cathepsins also play an important role in the proteolytic process of SARS-CoV and SARS-CoV-2[56], [58].

Once the S protein is proteolytically cleaved, membrane fusion follows. Generally, fusion occurs within endosomes, and in SARS-CoV-2, it happens along with an internalisation of ACE2 receptors that causes a downregulation of available ACE2 receptors on the cell surface. Once the virus is inside the cell, the nucleocapsid is released and the viral genome is translated by ribosomal frameshifting to produce polyproteins pp1a and pp1ab[59]. Following this, the polyproteins are processed by papain-

like protease (PLpro) and a main serine-like protease (MPro) encoded by NSPs 3 and 5. These proteases cause cleavage events to produce other NSPs[60]. The 15 NSPs along with other viral proteins and possibly, host proteins, form the replication-transcription complex (RTC) which accumulate in double-membrane vesicles. This is followed by the synthesis of mRNAs that code for structural proteins (S, E, M, N). Once they are produced, the membrane-bound structural proteins (S, E, and M) are integrated into the endoplasmic reticulum (ER). The NTD of S protein is heavily glycosylated and the signal sequence is used for accessing the ER. Along with the S protein, E and M proteins also move to ER-Golgi intermediate compartment (ERGIC) which is a site of viral assembly. While M protein regulates interaction between proteins for viral assembly, they are not solely responsible for producing envelopes and forming virions. They work in conjunction with E protein to produce viral envelopes, although the exact mechanism underlying this step is unclear. Concurrently, genome replication occurs to produce a full-length positive-sense single-strand RNA. Once they are produced, the N protein gets encapsulated onto the RNA to form a nucleocapsid, which then enters the ERGIC complex. By the time this happens, the membrane formed by M and E proteins are readily available in the complex. Once the RNA-nucleocapsid is inside the ERGIC, the M protein binds to them to stabilise the genome and the S proteins attach to the virions. The ability of S proteins to interact with M proteins fuels this step. After the viral particles are assembled, they are carried in smooth-walled vesicles to the plasma membrane, where they are released by exocytosis[24] (Fig 4). However, whether the virus uses the conventional export pathway from Golgi or if they have their own mechanism is unknown.

Immune response against SARS-CoV-2

The anti-viral response in the host typically begins when the virus enters the host cell. While research on immune response against SARS-CoV-2 is still in its infancy, studies on anti- SARS-CoV and MERS-CoV response will help in understanding the plausible mechanisms. Once the virus enters the cell using the host cell receptor, their antigenic peptides are presented by major histocompatibility complex (MHC, or the human leukocyte antigen (HLA) in humans). These are then recognised by the cytotoxic T lymphocytes (CTLs). For SARS-CoV, the MHC I predominantly presents the antigen[61], and for MERS-CoV, MHC II presents the antigenic peptides[62]. Some polymorphisms in the HLA gene, such as HLA-DR0301 and HLA-A∗0201, protect against SARS infection[63], while variants like HLA-DRB1∗11:01 modulates the susceptibility to MERS infection[62]. Once the presented antigen is recognised by CTL, humoral and cell-mediated immunity are stimulated. SARS-CoV and SARS-CoV-2 elicit a characteristic IgM and IgG response pattern. In SARS infection, IgM antibodies typically decline after week 12 of first infection[64] and IgG, particularly anti-S and anti-N IgG, continue to stay elevated[65]. On the other hand, the cellular immunity is characterised by a sharp drop in CD4+ and CD8+ cells in both SARS-CoV and SARS-CoV-2 infected individuals[66], [67]. Despite the magnitude of anti-viral immune response, SARS-CoV, MERS-CoV, and most likely SARS-CoV-2, have developed strategies to circumvent them. For example, the orf4a, orf4b, and orf5 of MERS-CoV interfere with the interferon regulatory factor 3 (IRF3), a protein which usually increases the type 1 interferon (T1IFN) and mitigates viral infection[68]. Additionally, the viruses can induce the formation of double-membrane vesicles and replicate within them to avoid being recognised by pattern recognition receptors[69]. Future studies focusing on understanding the immune evasion mechanisms of SARS-CoV-2 may provide important insights for developing effective therapeutics against COVID-19.

Conclusion

Over the last century, CoVs have rapidly evolved and several different strains emerged that caused a range of diseases in mammals and avian species. CoVs have the potential to infect multiple species and cell types, which in combination with their large genome makes them more likely to re-emerge in the future. The outbreak of SARS-CoV-2 that started in December 2019 in Wuhan, China has rapidly

spread across the globe placing a massive burden on healthcare. Phylogenetic analysis confirmed the zoonotic origin of the outbreak, with bats identified to be the reservoir host. The search an intermediate host is still underway, but recently accumulating evidence points to pangolins to be the potential intermediate host responsible for the species crossover. Rapid advances have been made in our understanding of SARS-CoV-2 characteristics, but we are only beginning to understand the pathogenesis and the anti-SARS-CoV-2 immune response in the host. For example, some NSPs and accessory proteins remain uncharacterised and studying their function and the role they play in viral replication will help in better understanding of the pathogenesis. Studies on the anti-viral response against SARS-CoV-2 will also help in identifying effective potential therapeutic targets for controlling COVID-19. Furthermore, completely understanding the ability of CoVs to cross species barrier will drastically help in predicting such outbreaks in the future.

Comment [S4]: Directions for future study have been provided here.

Consent to participate

Not applicable

Ethical approval

Not applicable

References

[1] S. R. Weiss and S. Navas-Martin, “Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus,” *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 4, pp. 635–664, 2005, doi: 10.1128/MMBR.69.4.635-664.2005.

[2] Z. Song *et al.*, “From SARS to MERS, thrusting coronaviruses into the spotlight,” *Viruses*, vol. 11, no. 1, 2019, doi: 10.3390/v11010059.

[3] P. Zhou *et al.*, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *nature.com*.

[4] N. Petrosillo, G. Viceconte, O. Ergonul, … G. I.-C. M., and undefined 2020, “COVID-19, SARS and MERS: are they closely related?,” *Elsevier*.

[5] S. K. P Lau *et al.*, “Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats,” 2005.

[6] N. Lee *et al.*, “A major outbreak of severe acute respiratory syndrome in Hong Kong,” *N. Engl. J. Med.*, vol. 348, no. 20, pp. 1986–1994, May 2003, doi: 10.1056/NEJMoa030685.

[7] D. Hui, A. Z.-I. D. Clinics, and undefined 2019, “Severe acute respiratory syndrome: historical, epidemiologic, and clinical features,” *id.theclinics.com*.

[8] A. Zumla, D. Hui, S. P.-T. Lancet, and undefined 2015, “Middle East respiratory syndrome,” *Elsevier*.

[9] I. Eckerle, V. Corman, M. Müller, … M. L.-E. infectious, and undefined 2014, “Replicative capacity of MERS coronavirus in livestock cell lines,” *ncbi.nlm.nih.gov*.

[10] N. Zhu *et al.*, “A novel coronavirus from patients with pneumonia in China, 2019,” *N. Engl. J. Med.*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[11] F. Wu *et al.*, “A new coronavirus associated with human respiratory disease in China,” *nature.com*.

[12] D. Paraskevis, E. G. Kostaki, G. Magiorkinis, G. Panayiotakopoulos, G. Sourvinos, and S. Tsiodras, “Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event,” *Infect. Genet. Evol.*, vol. 79, no. January, p. 104212, 2020, doi: 10.1016/j.meegid.2020.104212.

[13] “WHO Coronavirus Disease (COVID-19) Dashboard.” https://covid19.who.int/ (accessed Jun. 25, 2020).

[14] Y. Bai *et al.*, “Presumed asymptomatic carrier transmission of COVID-19,” *jamanetwork.com*.

[15] J. Cui, F. Li, Z. S.-N. R. Microbiology, and undefined 2019, “Origin and evolution of pathogenic coronaviruses,” *nature.com*.

[16] Y. Wan, R. Graham, R. Baric, F. L.-J. Virol., and undefined 2020, “An analysis based on decade-long structural studies of SARS 3, JVI Accepted Manuscript Posted Online 29 January 2020.”

[17] K. Xiao *et al.*, “Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan

Pangolins," *bioRxiv*, p. 2020.02.17.951335, 2020, doi: 10.1101/2020.02.17.951335.
[18] W. Ji, W. Wang, X. Zhao, J. Zai, and X. Li, "Cross-species transmission of the newly identified coronavirus 2019-nCoV," *J. Med. Virol.*, vol. 92, no. 4, pp. 433–440, Apr. 2020, doi: 10.1002/jmv.25682.
[19] J. Shi, Z. Wen, G. Zhong, H. Yang, … C. W.-, and undefined 2020, "Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2," *science.sciencemag.org*, doi: 10.1126/science.abb7015.
[20] S. P.-C. and experimental pediatrics and undefined 2020, "Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19)," *ncbi.nlm.nih.gov*.
[21] A. Wu *et al.*, "Commentary Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China," *Cell Host Microbe*, vol. 27, pp. 325–328, 2020, doi: 10.1016/j.chom.2020.02.001.
[22] A. Walls, Y. Park, M. Tortorici, A. Wall, A. M.- Cell, and undefined 2020, "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein," *Elsevier*.
[23] Y. R. Guo *et al.*, "The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status," *Military Medical Research*, vol. 7, no. 1. BioMed Central Ltd., Mar. 13, 2020, doi: 10.1186/s40779-020-00240-0.
[24] A. R. Fehr and S. Perlman, "Coronaviruses: An overview of their replication and pathogenesis," in *Coronaviruses: Methods and Protocols*, Springer New York, 2015, pp. 1–23.
[25] T. R. Ruch and C. E. Machamer, "The coronavirus E protein: Assembly and beyond," *Viruses*, vol. 4, no. 3. Multidisciplinary Digital Publishing Institute (MDPI), pp. 363–382, Mar. 2012, doi: 10.3390/v4030363.
[26] D. Schoeman and B. C. Fielding, "Coronavirus envelope protein: Current knowledge," *Virology Journal*, vol. 16, no. 1. BioMed Central Ltd., May 27, 2019, doi: 10.1186/s12985-019-1182-0.
[27] M. L. DeDiego *et al.*, "Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis," *PLoS Pathog.*, vol. 7, no. 10, Oct. 2011, doi: 10.1371/journal.ppat.1002315.
[28] M. Surjit, S. L.- Infection, genetics and evolution, and undefined 2008, "The SARS-CoV nucleocapsid protein: a protein with multifarious activities," *Elsevier*.
[29] M. H. Verheije *et al.*, "The Coronavirus Nucleocapsid Protein Is Dynamically Associated with the Replication-Transcription Complexes † Downloaded from," *J. Virol.*, vol. 84, no. 21, pp. 11575–11579, 2010, doi: 10.1128/JVI.00569-10.
[30] M. Ujike, F. Taguchi, M. Johnson, and S.-L. Liu, "Incorporation of Spike and Membrane Glycoproteins into Coronavirus Virions," *Viruses*, vol. 7, pp. 1700–1725, 2015, doi: 10.3390/v7041700.
[31] I. Astuti and Ysrafil, "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 4, pp. 407–412, 2020, doi: 10.1016/j.dsx.2020.04.020.
[32] J. Fuk-Woo Chan *et al.*, "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan," *Taylor Fr.*, vol. 9, no. 1, pp. 221–236, Jan. 2020, doi: 10.1080/22221751.2020.1719902.
[33] S. Angeletti, D. Benvenuto, M. Bianchi, M. Giovanetti, S. Pascarella, and M. Ciccozzi, "COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis," *J. Med. Virol.*, vol. 92, no. 6, pp. 584–588, Jun. 2020, doi: 10.1002/jmv.25719.
[34] B. Krichel, S. Falke, R. Hilgenfeld, L. Redecke, and C. Uetrecht, "Processing of the SARS-CoV pp1a/ab nsp7-10 region," *Biochem. J.*, vol. 477, no. 5, pp. 1009–1019, Mar. 2020, doi: 10.1042/BCJ20200029.
[35] R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *Elsevier*.
[36] Q. Tang *et al.*, "Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition," *nature.com*.
[37] C. Drosten *et al.*, "Identification of a novel coronavirus in patients with severe acute respiratory syndrome," *N. Engl. J. Med.*, vol. 348, no. 20, pp. 1967–1976, May 2003, doi: 10.1056/NEJMoa030747.
[38] A. M. Zaki, S. Van Boheemen, T. M. Bestebroer, A. D. M. E. Osterhaus, and R. A. M. Fouchier, "Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia," *N. Engl. J. Med.*, vol. 367, no. 19, pp. 1814–1820, Nov. 2012, doi: 10.1056/NEJMoa1211721.
[39] D. Forni, R. Cagliani, M. Clerici, M. S.-T. in microbiology, and undefined 2017, "Molecular evolution of human coronavirus genomes," *Elsevier*.

[40] C. Ceraolo and F. M. Giorgi, “Genomic variance of the 2019-nCoV coronavirus,” *J. Med. Virol.*, vol. 92, no. 5, pp. 522–528, May 2020, doi: 10.1002/jmv.25700.
[41] D. A. Brian and R. S. Baric, “Coronavirus genome structure and replication,” *Current Topics in Microbiology and Immunology*, vol. 287. pp. 1–30, 2005, doi: 10.1007/3-540-26765-4_1.
[42] F. Ferron *et al.*, “Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA,” *Natl. Acad Sci.*, p. 2017, doi: 10.1073/pnas.1718806115.
[43] P. Forster, L. Forster, C. Renfrew, and M. Forster, “Phylogenetic network analysis of SARS-CoV-2 genomes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 17, pp. 9241–9243, Apr. 2020, doi: 10.1073/pnas.2004999117.
[44] T. P.- Infection, genetics and evolution, and undefined 2020, “Genetic diversity and evolution of SARS-CoV-2,” *Elsevier*.
[45] T. S. Fung and D. X. Liu, “Human Coronavirus: Host-Pathogen Interaction,” *Annu. Rev. Microbiol.*, vol. 73, no. 1, pp. 529–557, Sep. 2019, doi: 10.1146/annurev-micro-020518-115759.
[46] X. W. Zhang, Y. L. Yap, and A. Danchin, “Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus,” *Arch Virol*, vol. 150, no. 1, pp. 1–20, Jan. 2005, doi: 10.1007/s00705-004-0413-9.
[47] Z. Zhang, L. Shen, X. G.-S. reports, and undefined 2016, “Evolutionary dynamics of MERS-CoV: potential recombination, positive selection and transmission,” *nature.com*.
[48] Y. A. Helmy, M. Fawzy, A. Elaswad, A. Sobieh, S. P. Kenney, and A. A. Shehata, “Clinical Medicine The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control,” *mdpi.com*, doi: 10.3390/jcm9041225.
[49] J. Shang *et al.*, “Cell entry mechanisms of SARS-CoV-2,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 21, pp. 11727–11734, May 2020, doi: 10.1073/pnas.2003138117.
[50] M. Letko, A. Marzi, V. M.-N. microbiology, and undefined 2020, “Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses,” *nature.com*.
[51] Y. Wan, J. Shang, R. Graham, R. Baric, F. L.-J. of virology, and undefined 2020, “Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus,” *Am Soc Microbiol*.
[52] J. Shang *et al.*, “Structural basis of receptor recognition by SARS-CoV-2,” *Nature*, vol. 581, no. 7807, pp. 221–224, May 2020, doi: 10.1038/s41586-020-2179-y.
[53] M. Gui *et al.*, “Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding,” *nature.com*.
[54] Y. Yuan *et al.*, “Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains,” *nature.com*.
[55] J. Bosch, R. Van Der Zee, C. A. M. De Haan, and P. J. M. Rottier, “The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex Downloaded from,” *J. Virol.*, vol. 77, no. 16, pp. 8801–8811, 2003, doi: 10.1128/JVI.77.16.8801-8811.2003.
[56] J. A. Jaimes, J. K. Millet, and G. R. Whittaker, “Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site,” *iScience*, vol. 23, no. 6, p. 101212, 2020, doi: 10.1016/j.isci.2020.101212.
[57] H. Zhou, X. Chen, A. C. Hughes, Y. Bi, and W. Shi, “A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein,” *Curr. Biol.*, vol. 30, pp. 2196-2203.e3, 2020, doi: 10.1016/j.cub.2020.05.023.
[58] M. Hoffmann *et al.*, “SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor,” *Cell*, vol. 181, no. 2, pp. 271-280.e8, Apr. 2020, doi: 10.1016/j.cell.2020.02.052.
[59] P. Baranov, C. Henderson, C. A.- Virology, and undefined 2005, “Programmed ribosomal frameshifting in decoding the SARS-CoV genome,” *Elsevier*.
[60] J. Ziebuhr, E. J. Snijder, and A. E. Gorbalenya, “Virus-encoded proteinases and proteolytic processing in the Nidovirales,” *Journal of General Virology*, vol. 81, no. 4. Society for General Microbiology, pp. 853–879, 2000, doi: 10.1099/0022-1317-81-4-853.
[61] J. Liu *et al.*, “Novel Immunodominant Peptide Presentation Strategy: a Featured HLA-A*2402-Restricted Cytotoxic T-Lymphocyte Epitope Stabilized by Intrachain Hydrogen Bonds from Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein,” *J. Virol.*, vol. 84, no. 22, pp. 11849–11857, 2010, doi: 10.1128/jvi.01464-10.
[62] A. Hajeer, H. Balkhy, … S. J.-A. of thoracic, and undefined 2016, “Association of human leukocyte antigen class II alleles with severe Middle East respiratory syndrome-coronavirus

infection," *ncbi.nlm.nih.gov*.
[63] S.-F. Wang *et al.*, "Human-Leukocyte Antigen Class I Cw 1502 and Class II DR 0301 Genotypes Are Associated with Resistance to Severe Acute Respiratory Syndrome (SARS) Infection," *liebertpub.com*, vol. 24, no. 5, pp. 421–426, Oct. 2011, doi: 10.1089/vim.2011.0024.
[64] E. De Wit, N. Van Doremalen, … D. F.-N. R., and undefined 2016, "SARS and MERS: recent insights into emerging coronaviruses," *nature.com*.
[65] G. Li, X. Chen, and A. Xu, "Profile of specific antibodies to the SARS-associated coronavirus [6]," *New England Journal of Medicine*, vol. 349, no. 5. pp. 508–509, Jul. 31, 2003, doi: 10.1056/NEJM200307313490520.
[66] Y. Fan *et al.*, "Characterization of SARS-CoV-specific memory T cells from recovered individuals 4 years after infection," *Springer*.
[67] Z. Xu, L. Shi, Y. Wang, J. Zhang, … L. H.-T. L. respiratory, and undefined 2020, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *thelancet.com*.
[68] Y. Yang *et al.*, "Protein Cell & Protein Cell The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists," *Protein Cell*, vol. 4, no. 12, pp. 951–961, 2013, doi: 10.1007/s13238-013-3096-8.
[69] E. J. Snijder *et al.*, "Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex," *J. Virol.*, vol. 80, no. 12, pp. 5927–5940, 2006, doi: 10.1128/JVI.02501-05.
[70] A. R. Fehr and S. Perlman, "Coronaviruses: An overview of their replication and pathogenesis," in *Coronaviruses: Methods and Protocols*, vol. 1282, Springer New York, 2015, pp. 1–23.
[71] L. Ren, Y. Wang, Z. Wu, Z. Xiang, … L. G.-C. medical, and undefined 2020, "Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study," *ncbi.nlm.nih.gov*.
[72] E. Snijder, P. Bredenbeek, J. Dobbe, … V. T.-J. of molecular, and undefined 2003, "Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage," *Elsevier*.

Table 1: A comprehensive table enlisting the function, mechanism of action of non-structural proteins (NSPs) of SARS-CoV-2, and their nucleotide homology to SARS-CoV[32]–[34], [70].

NSPs	**Function**	**Mechanism**	**Nucleotide similarity to SARS-CoV**[71]
NSP 1	Suppresses immune response	Degrades mRNA and inhibits translation, blocking innate immune response	82.2%
NSP 2	Unknown	Unknown	68.3%
NSP 3	Cleaves viral polyprotein	Papain-like protease (PLpro)	72.7%
NSP 4	Suspected to act as scaffolding for transmembrane proteins	Unknown	74.8%
NSP 5	Cleaves viral polyprotein	Main protease domain (MPro)	83.6%
NSP 6	Suspected to act as scaffolding for transmembrane proteins	Unknown	79%
NSP 7	Acts as primase with NSP 8. Also processes RNA polymerase.	Forms hexadecameric complex with NSP 8	79.9%
NSP 8	Acts as primase with NSP 7. Also processes RNA polymerase.	Forms hexadecameric complex with NSP 7	86.1%
NSP 9	Acts as RNA binding protein	-	84.9%
NSP 10	Supports NSPs 14 and 16	Forms heterodimer with NSPs 14 and 16. Stimulates formation of ExoN	88.2%
NSP 12	RNA-dependent RNA polymerase (RdRP)	-	88.5%
NSP 13	RNA helicase	-	88.4%
NSP 14[a]	3'-5' exoribonuclease (ExoN)	Adds 5' cap to RNA. ExoN helps in proofreading viral genome after replication.	83.3%

NSP 15[a]	PolyU-specific endoribonuclease	-	82%
NSP 16	Immune evasion	Protects RNA from being detected by MDA5	85.1%

[a]NSPs 14 and 15 are highly conserved and considered genetic markers of viruses under Order *Nidovirales*[72].

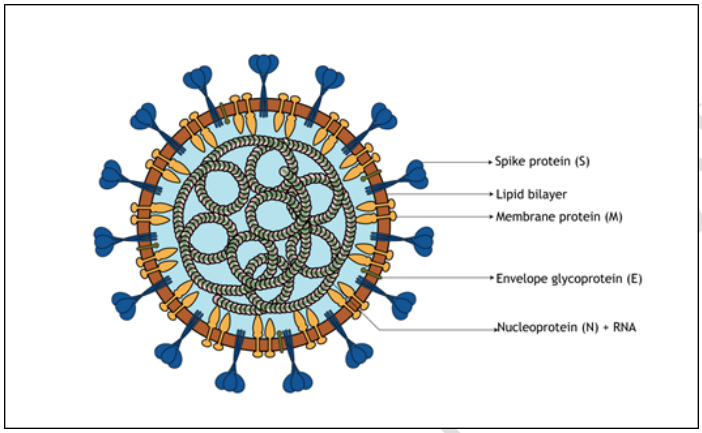

Fig 1 Schematic representation of the structure of SARS-CoV-2

This illustration depicts the location of the structural proteins, where each protein plays a critical role at various stages of events. The Spike protein (S) located on the outer surface of the virus facilitates the binding to the transmembrane ACE2 host receptor; the envelope protein (E) aids in the assembly and maturation of SARS-CoV-2 post-replication events; the nucleocapsid (N) protein plays a vital role in replication and transcription processes; the M protein promotes the stabilisation process of the nucleocapsid-RNA complex

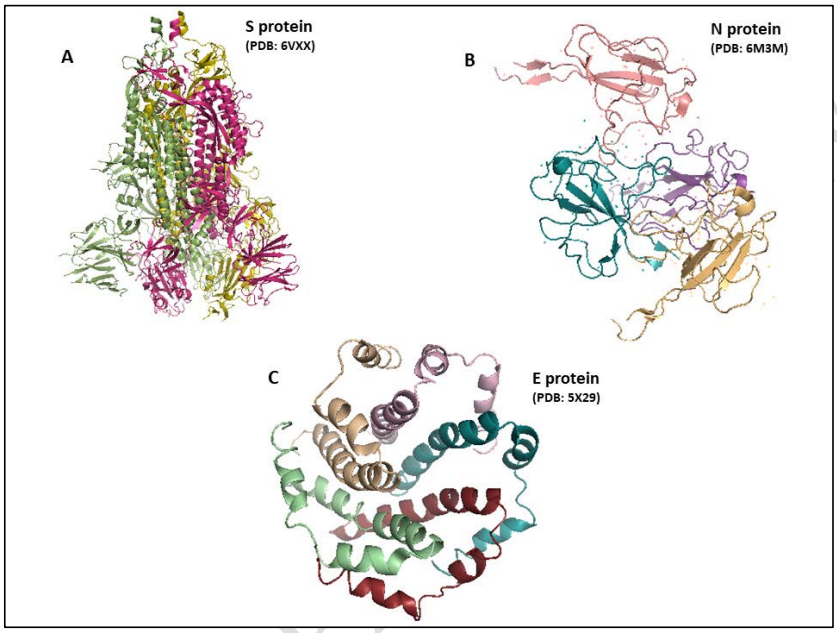

Fig 2 Cartoon representation of SARS-CoV-2 structural proteins (S,N) and SARS-CoV (E) protein.

(A) Spike glycoprotein (PDB: 6VXX) of SARS-Cov-2 is a transmembrane protein that forms a homotrimer (depicted by three different colors), but is predominantly seen in the outer surface of the virus. (B) Nuclecapsid protein (PDB: 5X29) of SARS-CoV-2 is highly conserved and the amino acids residues present in the N-terminal domain of the protein plays a crucial role in RNA binding. (C) Envelope protein (PDB: 5X29) of the SARS-CoV is a small membrane protein that aids in ion channel formation. The above figures were illustrated using Pymol

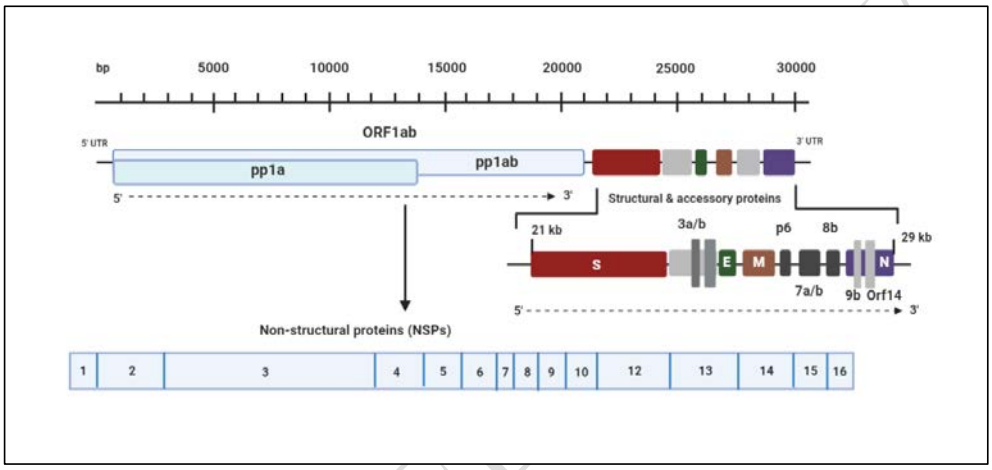

Fig 3 Genome organisation of SARS-CoV-2

The genome of SARS-CoV-2 is 29.8 kb in size. The schematic representation shows regions encoding the different structural, non-structural, and accessory proteins. Towards the 5' end are the pp1a/ab proteins that produce NSPs and towards the 3' end are the structural proteins and accessory proteins

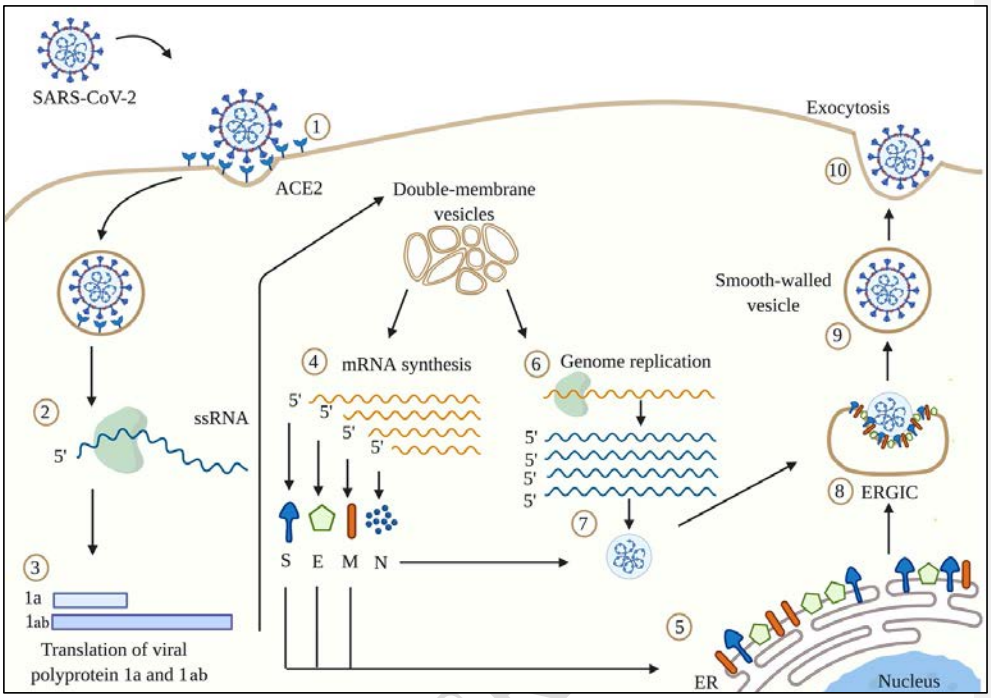

Fig 4 Potential pathogenesis of SARS-CoV-2

1. Viral entry into the host using ACE2 receptor. 2. Release of the viral genome. 3. Translation of the genome to produce polyproteins 1a and 1ab. 4. Synthesis of mRNA to produce structural proteins. 5. Integration of S, E, and M proteins into ER and entry into the site of viral assembly. 6. Replication of genome to produce full-length ssRNA. 7. Encapsulation of N protein on the newly produced RNA. 8. Movement of nucleocapsid to ERGIC to gain membrane proteins. 9 and 10. Release of newly formed viral particles by exocytosis

Comment [S5]: The image has been replaced to match the figure legend.