Microbiology Research Journal International

31(2): 64-85, 2021; Article no.MRJI.66560
ISSN: 2456-7043
(Past name: British Microbiology Research Journal, Past ISSN: 2231-0886, NLM ID: 101608140)

Comparative Analyses of ACE2 Receptor Binding Corona Viruses that Cause Mild versus Severe Acute Respiratory Syndrome in Humans

Peramachi Palanivelu[1*]

[1]Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India.

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/MRJI/2021/v31i230299
Editor(s):
(1) Prof. Essam Hussein Abdel-Shakour, Al-Azhar University, Egypt.
(2) Dr. Zahid Anwar, University of Gujrat, Pakistan.
(3) Prof. En Tao Wang, Instituto Politécnico Nacional, México and China Agricultural University, China.
Reviewers:
(1) Kleber Juvenal Silva Farias, Federal University of Campina Grande, Brazil.
(2) Phelipe Magalhães Duarte, Universidade de Cuiabá, Brazil.
(3) Basil Abdulzahra Abbas, University of Basrah, Iraq.
(4) Maria Cristina Teixeira Cangussu, Federal University of Bahia, Brazil.
(5) Alicja Forma, Medical University of Lublin, Poland.
Complete Peer review History: http://www.sdiarticle4.com/review-history/66560

Original Research Article

Received 01 March 2021
Accepted 19 April 2021
Published 23 April 2021

ABSTRACT

Aim: To analyze the spike proteins and Replication-Transcription Complexes (RTCs) of the Mild and Severe Acute Respiratory Syndrome (SARS) and SARS-related coronaviruses (CoVs) to find out the similarities and differences between them, as both of groups bind to angiotensin-converting enzyme 2 (ACE2) receptor for human cell entry.
Study Design: Bioinformatics, Biochemical, Site-Directed Mutagenesis (SDM), X-ray crystallographic, cryo-Electron microscopic (cryo-EM) and Mass Spectrometric (MS) data were analyzed.
Methodology: The protein sequence data for spike proteins and the proteins of the RTCs, viz. the RNA- dependent RNA polymerases (RdRps), primases and the nonstructural protein 7 (NSP7) were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. Along with the conserved motifs identified by the bioinformatics analysis, the data already available by biochemical and SDM experiments and X-ray

**Corresponding author: E-mail: ppmkupp@gmail.com;*

crystallographic and cryo-EM studies on these proteins were used to confirm the possible amino acids involved in ACE2 receptor binding and active sites of the RTCs. For identification of probable N-linked and O-linked glycosylation sites, NetNGlyc 1.0 and NetOGlyc 4.0 tools of Technical University of Denmark were used. ExPASy tool was used for pI analysis.

Results: The spike protein of human CoV (HCoV)-NL63 is ~90 amino acids longer than the spike proteins of SARS and SARS-related CoVs. The additions are mostly found in the N-terminal regions and few insertions are also found in the crucial receptor binding domain (RBD). The SARS and SARS-related CoVs and HCoV-NL63 showed several conserved residues, motifs and large peptide regions. The most important aspect between the recent pandemic causing SARS-CoV-2 and HCoV-NL63 is a unique but different tetrapeptide insertions very close to the S1/S2 cleavage region, i.e., -PRRA- and -IPVR-, respectively. The next cleavage point S2' and the transmembrane domains are conserved between the two groups. The RdRps are highly conserved between the two groups. The catalytic regions, catalytic amino acids and the NTP selection tripeptide regions are completely conserved between SARS-CoVs and HCoV-NL63. However, one of the metal binding sites, viz. the universal –G**D**D- reported in all RdRps is aligning with– K**D**G- in the RdRp of HCoV-NL63. The other metal binding site, viz. –SDD- is completely conserved in both the groups. The NiRAN domains of the RdRps differed from the possible catalytic amino acid and NTP selection tripeptide regions. The primases (NSP8) and the NSP7 subunits of the RTC are highly conserved in both the groups. The NSP8 and NSP7 subunits exhibit closer similarities between the MERS-CoV and HCoV-NL63. Unlike other SARS and SARS-related CoVs, the HCoV-NL63 possesses only a single accessory protein. Interestingly, a large number of amino acids are replaced with Ns in the spike proteins (which is also reflected in the number of N-linked glycosylation sites in it) as well as in the RTC.

Conclusions: Detailed analysis revealed several unique features in the HCoV-NL63 pathogen. As all the pandemic strains like SARS-CoV-1, SARS-CoV-2 and the milder HCoV-NL63 strain, use the same ACE2 receptor for entry into human cells, the frequent infection of humans by HCoV-NL63, especially in children, suggests that there is an ample opportunity for highly pathogenic variants to evolve in the future**.**

Keywords: Coronaviruses; SARS-CoVs; HCoV-NL63; angiotensin-converting enzyme 2; spike proteins; transcription-replication complex.

1. INTRODUCTION

RNA viruses employing positive, single-stranded RNA genomes form the most abundant class and its members are known to infect all types of hosts except Archaea. Large numbers of human and animal pathogens belong to this class and have been causing major global health-care crisis and unprecedented economic losses. Human coronaviruses (HCoVs) belong to positive-strand RNA virus and have attracted renewed interest, because of the emergence of SARS-CoV-2, the causative agent of the recent pandemic. In fact, the recent pandemic caused by SARS-CoV-2 has affected more than 100 million people, resulting in ~2.4 million deaths worldwide. The types of RNA viruses and their classifications have been summarized by Palanivelu [1]. SARS-CoV (SARS-CoV-1) and SARS-CoV-2 use angiotensin-converting enzyme 2 (ACE2) for binding onto human cells and entry. However, one more HCoV and HCoV-NL63 also employs the same ACE2 receptor for entry into human cells, but cause mostly mild respiratory diseases. Therefore, it will be especially interesting to find out whether the differences between the mild and severe acute versions of the HCoVs reside in their mode of ACE2 binding and their replication potentials or both. Characterization of HCoV-NL63 spike protein interactions with ACE2 receptor and its RTC might be useful not only for vaccine development but also for development of antivirals as they are the main targets for therapeutic interventions. Furthermore, the apparent similarities and differences between the SARS-CoVs and HCoV-NL63 and the frequent HCoV-NL63 infection of humans suggest that the pathogenic HCoVs could evolve into new variants. Analyses of the spike protein interactions with the ACE2 receptor and the RTC by these two different types of HCoVs might reveal important insights into their RBDs, RTCs and evolution of these viruses.

1.1 Classification of HCoVs

CoVs are zoonotic viruses and belong to the group of Nidoviruses which possess a positive, single-stranded RNA genome with a characteristic lipid bilayer envelope protecting the

genome. The CoVs are classified under the order *Nidovirales*, family *Coronaviridae* and subfamily *Coronavirinae* while the CoVs belonging to the family *Coronaviridae* are host specific and primarily cause enzootic infections in birds and mammals. However, in the last few decades, they are shown to be capable of infecting humans as well. The outbreaks of SARS-CoV-1in 2003, MERS-CoV in 2013 and currently the SARS-CoV-2 in 2020 have demonstrated not only their adaptability and lethality when they cross the species barrier and start infecting humans but also their efficient human to human transfer. Another important problem with the current SARS-CoV-2 is their ability to jump back and forth between humans and animals and the notable one is the recent outbreak of the virus in mink farms in Denmark. The infected minks started reinfecting humans. (in August and September 2020, a SARS-CoV-2 variant linked to infection among farmed minks and subsequently transmitted to humans). During such bidirectional interchange of SARS-CoV-2 between animals and humans, modifications in the virus can occur and may possibly generate new and more deadly variants.

Based on genetic and antigenic criteria, the *Coronavirinae* subfamily consists of four genera: α-CoVs, β-CoVs, γ-CoVs and δ-CoVs. Evolutionary analyses have shown that bats and rodents are the gene sources for most of the α-CoVs and β-CoVs, whereas avian and some mammalian species are the gene sources for most of the γ-CoVs and δ-CoVs [2]. It is interesting to note that only the α- and β-CoVs are known to infect humans. The HCoV-NL63 belongs to α-CoV and is not dangerous as they cause only mild, self-limiting respiratory infections in humans, which last for 3 to 4 weeks. It is interesting to note that all the known HCoVs do not use ACE2 receptor to enter into human cells but use different type of receptors for human cell entry. Though the SARS-CoV and HCoV-NL63 use the same ACE2 receptor for human cell entry, yet they differ dramatically in their ability to induce respiratory diseases [3,4]. Therefore, an in-depth analysis of HCoV-NL63 might unravel its pathogenicity factors in humans and its evolution.

1.2 Infection and Associated Diseases by HCoVs

Seven members make the HCoV family in which only 3, viz. SARS-CoV-1, MERS-CoV and SARS-CoV-2 belonging to β-CoV genus have been found to be dangerous and deadly to cause SARS in epidemic and pandemic proportions. As the three human CoVs that caused SARS have been studied extensively, only the rest of the HCoVs including HCoV-OC43 (β), HCoV-HKU1(β), HCoV-229E (α) and HCoV-NL63 (α), which are less dangerous and cause mild respiratory symptoms in humans are described here. They cause most of the colds every year, but they are not a serious threat to otherwise healthy individuals. Out of the four milder versions, only HCoV-NL63 uses the same ACE2 receptor like the other 3 SARS-CoVs, but it belongs to the genus α-CoV. HCoV-NL63 was first described in 2004, soon after the SARS-CoV-1, (as it was first described from the Netherlands, it was also named NL-63) epidemic, in a clinical sample of a child suffering from a respiratory condition that tested negative for all the known respiratory pathogens. Later, HCoV-NL63 was found to be a major cause of bronchiolitis and pneumonia in newborns and could cause severe lower respiratory tract infections among young children and immune-compromised adults [5]. Presently, HCoV-NL63 infections have spread worldwide, and are detected frequently in the winter season. Further information about these 4 HCoVs is described below and in Table 1.

HCoV-OC43 (β) infects humans and cattle. It normally causes common cold in humans. It uses *N*-Acetylneuraminic acid (NANA) as the receptor for cell entry with an additional shorter spike protein called hemagglutinin esterase (HE) [6]. NANA is the predominant sialic acid found in human and many mammalian cells. NANA also acts as a receptor for influenza viruses where the viruses get attached to mucous cells via hemagglutinin [7].

HCoV-HKU1 (β) infects humans. It usually causes an upper respiratory disease with symptoms of common cold, but can advance to pneumonia and bronchiolitis. It was first discovered in January 2004 in Hong Kong from a person suffering from upper respiratory tract infection and subsequent investigations revealed that it has global distribution. Like the HCoV-OC43, it also enters human cells via the NANA receptor and use additionally HE for receptor binding [6].

HCoV-229E (α) infects humans and bats [6]. Unlike the other two HCoVs discussed above, HCoV-229E binds to a different receptor for cell entry, i.e., it binds to the aminopeptidase N

(APN) receptor [8] HCoV-229E is associated with respiratory symptoms ranging from the common cold to high-morbidity conditions such as pneumonia and bronchiolitis. However, such high morbidity outcomes are almost always associated with co-infection with other respiratory infections.

HCoV-NL63 (α) was first discovered in a seven-month-old child suffering from bronchiolitis and conjunctivitis in the Netherlands in late 2004 [9] and is now reported worldwide [10,11]. The associated diseases include mild to moderate upper respiratory tract infections (URTI), sometimes severe lower respiratory tract infections (LRTI), croup, (laryngo-tracheobronchitis) and bronchiolitis [12]. It has a seasonal association in temperate climates and frequently detected in the winter season. Intriguingly, a novel HCoV infection in children similar to HCoV-NL63 was also reported to be associated with Kawasaki disease [13], which affects the coronary arteries and is a major cause of acquired heart disease in young children. Furthermore, the expression of ACE2 in coronary vessels further supports a possible role of HCoV-NL63 in Kawasaki disease. It is likely that the HCoV-NL63 might have evolved by recombinant events in bats infected by two very closely related viruses like HCoV-229E and other types. Since both the SARS-CoVs and HCoV-NL63 use the same receptor, it provides an opportunity for a double infection of the same cell and evolution of more pathogenic variants by recombination events between these two different viral groups [9]. Therefore, an analysis of the HCoV-NL63 genome and proteome might reveal important insights into the evolution and pathogenesis of this human pathogen.

1.3 HCoVs' Receptors in Human Cells for Entry

As discussed elsewhere, not all HCoVs use the same receptor for human cell entry, e.g., different CoV use different receptors for human cell entry. Table 1 shows the family of HCoVs, their receptors and diseases caused in humans.

HCoV, Human Coronavirus; SARS-CoV, Severe Acute Respiratory Syndrome coronavirus ; MERS-CoV, Middle East Respiratory Syndrome coronavirus; APN, Aminopeptidase N; ACE2, Angiotensin-Converting Enzyme 2,NANA, N-Acetyl-9-O-acetylneuraminic Acid; HE, Hemagglutinin Esterase DPP4, Dipeptidyl Peptidase 4.

Interestingly, all the HCoVs use an enzyme protein as receptor; HCoV-OC43 and HCoV-HKU1 use NANA, but along with an additional shorter spike protein HE, Therefore, it is likely they exploit the moonlighting function(s) of these enzymes and use them as receptor for cell entry. HCoV-NL63, like SARS-CoV-1, replicates efficiently in monkey kidney cells suggesting a use of a common receptor [9]. It should also be noted that small changes like one or two amino acids in the spike protein altogether changed the receptor binding property. For example the bat CoV-HKU4 spike cannot mediate viral entry into human cells; but two mutations enabled it to do so by allowing it to be activated by human proteases [14]. Similarly, single mutations like N439K and E484K (the RBM variants), in the spike protein of SARS-CoV-2, have markedly affected the antigenic property and efficacy of neutralizing antibodies [15].

Table 1 Family of HCoVs, their Receptors and Properties

Name	Receptor type	Disease(s) caused	Reference(s)
Alpha-Coronaviruses			
HCoV-229E	APN	Common cold	[6,8]
HCoV-NL63	ACE2	Mild to moderate URTIs, LRTIs, Bronchiolitis, Croup	[8,10-13]
Beta Coronaviruses			
HCoV-OC43	NANA/HE	Common cold	[6]
HCoV-HKU1	NANA/HE	Common cold, URTIs	[6]
SARS-CoV-1	ACE2	SARS	[8]
SARS-CoV-2	ACE2	SARS	[1]
MERS-CoV	DPP4	SARS	[8]

1.4 Properties of Human ACE2 Receptor and its Localization on Human Chromosome

As seen in Table 1, only three of the HCoVs use the ACE2 receptor for human cell entry (highlighted in yellow). The ACE2, (EC 3.4.17.23) is a homologue of angiotensin-converting enzyme ACE1 (ACE1, a peptidyl-dipeptidase that cleaves the C-terminal dipeptide from Angiotensin I (Ang I) to produce Ang II. Both ACE1 and ACE2 play an important role in the rennin-angiotensin system for blood pressure homeostasis. In other words, the ACE1 converts the Ang 1 to Ang II and the ACE2 converts Ang II into Ang 1-7 (vasodilator) and thus deactivating the Ang II. Thus, the ACE2 acts to counterbalance the activity of ACE1 to provide effective blood pressure maintenance.

ACE2 is an integral membrane protein with an ectodomain, projected on the cell surfaces of endothelial cells which are mainly present in lungs and intestine, but also in lower concentrations in other cells. ACE2 is a metallo-carboxypeptidase which requires zinc ion for maximal activity and optimally active at pH 6.5. The ACE2 enzyme cleaves a single amino acid residue at the carboxyl terminus of peptides, e.g., it cleaves the carboxyl-terminal amino acid phenylalanine (marked in red) from Ang II (Asp-Arg-Val-Tyr-Ile-His-**Pro**↓-**Phe**) into the vasodilator Ang 1-7, (Asp-Arg-Val-Tyr-Ile-His-**Pro-OH**). Thus, ACE1 and ACE2 play opposite roles in maintain blood pressure maintenance in humans. (ACE1 (EC 3.4.15.1) is also a zinc metallo-endopeptidase, but it acts as a peptidyl dipeptidase and converts Ang-I to Ang-II. ACE2 exhibits 42% sequence identity and 61% sequence similarity to its closest homolog ACE1 [16].

1.5 Genome Organization of HCoV-NL63

In general, CoVs share similar genomic, proteomic structures, replication mechanisms and overall gene sequences. Like other SARS-CoVs, HCoV-NL63 is an enveloped virus with a positive-sense, single-stranded RNA genome of ~28 kb in size and contains a cap structure at its 5' end and a poly-(A) tail at its 3' end, mimicking the human cellular mRNAs for direct translation of the viral proteins in the host cells. The complete genome of HCoV-NL63 is sequenced which is available under the accession number AY567487) [9,17,18]. The genome of HCoV-NL63 consists of 27,553 nucleotides whereas the SARS-CoV-1 and SARS-CoV-2 (Wuhan strain) consists of 29,903 and 29,751 nucleotides, respectively and a poly-(A) tail at the 3'-end [19]. Like other SARS-CoV genomes, the HCoV-NL63 genome also consists of 5'- and 3'-UTRs and the 1a and 1b genes with a pseudo-knot ribosome frame shift structure. The 1a and 1b genes encode all the nonstructural polyproteins (NSPs) NSP1-NSP16. Whereas the subgenomic mRNAs code for the four structural proteins (SPs) , viz. the spike protein (S), envelope protein (E), membrane protein (M), and nuclear capsid protein (N) as in other SARS-CoVs. Each subgenomic mRNA has a common 5′-end, derived from the 5′-portion of the genome (the 5′-leader sequence), and common 3′-coterminal parts. All subgenomic mRNAs have an identical 5'- part, i.e., an untranslated leader of 72 nt. Discontinuous transcription requires base-pairing between *cis*-acting transcription regulatory sequences (TRSs), located near the 5′-end of the viral genome (the leader TRS) and others located upstream of each of the respective ORFs (the body TRSs) [9]. The TRS motif of HCoV-NL63 is AACUAAA and is conserved in all subgenomic mRNAs except in E protein gene where it was found to be slightly different, AACUA**U**A [20]. HCoV-NL63 multiplication can be inhibited at the transcriptional level by pyrimidine nucleoside analogues: *b*-D-N4-hydroxycytidine and 6-azauridine [20]. The accessory protein gene ORF3 was found located between the S and E genes [3]. Thus, the less aggressive HCoV-NL63 genome is structurally similar to the SARS-CoVs, where ORF 1a codes for 1-11 NSPs and the ORF 1b codes for 12-16 NSPs in the same order and exhibiting the same functions. They are produced as polyproteins as in other SARS-CoVs. The polyproteins are cleaved by two different proteases, viz, papain-like (PL^{pro}, NSP3 autocatalytic) and a serine-like main protease (M^{pro}, NSP5, cleaves at LQ↓X). The PL^{pro} cuts at 3 sites between NSP1↓NSP2, NSP2↓NSP3, and NSP3↓NSP4 and the rest of the sites are cleaved by the M^{pro} [9,21]. In the same way, the subgenomic region codes for the 4 SPs, viz. S, E, M and N; - again in the same order and exhibiting the same functions as in SARS-CoVs. At the beginning of each structural or accessory gene, TRS is present for expression of each of the gene. They found the GC content of the HCoV-NL63 is only 34%, and therefore, it has the lowest GC content among the *Coronaviridae.* Short untranslated regions (UTRs) of 286 and 287 nucleotides are present at the 5'-and 3'-termini, respectively. In stark

contrast to other HCoVs, the HCoV-NL63 encodes for a single accessory protein gene.

1.6 Differences in the Accessory Proteins between Mild and SARS-CoVs

As discussed elsewhere, the striking difference between these two groups of HCoVs is in the number of accessory proteins. The recent pandemic causing SARS-CoV-2 produces as many as 27 proteins/peptides out of which 16 of them are NSPs, 4 are SPs and 7 are accessory proteins (APs) whereas the HCoVs-NL63 harbours single AP. The functions of the APs have been reported previously [21]. The number SPs and APs in HCoV-Nl63 and SARS-CoVs are shown in Fig.1. In SARS-CoV-1, SARS-CoV-2 and HCoV-229E the AP3 has additional overlapping ORF and processed as 3a and 3b. In SARS-CoV-2, it has been found that the ORF3a AP makes a hole in the membrane of infected cells to make the new viruses easier to escape. It also triggers inflammation, one of the most dangerous symptoms of COVID-19.

The ORF3b AP is a novel short protein with 4 helices. This SARS-CoV-2 AP exhibits no homology to the existing SARS-CoVs or SARS-related-CoVs. However, the ORF3b still possessed Interferon (IFN) antagonist and Interferon Regulating transcription Factor-3 modulating activities, suggesting a potential link between bats' SARS-related-CoVs ORF3b function [2].

However, in HCoV-NL63 the AP3 presented as a single gene similar to that in MERS-CoV (Fig. 1). SARS-CoV-1 3a protein, also referred to as ORF3 [22], was the first characterized SARS-CoV accessory protein. The co-localization of ORF3a with M and E proteins, which are essential for viral assembly, suggests that ORF3a also may play an important role in SARS-CoV-1 assembly or budding [23]. The tissue culture studies have shown that accessory proteins are almost exclusively nonessential for replication of the HCoVs. However, some of them have been shown to play important roles in viral pathogenesis [24]. It is therefore, conceivable that a single ORF of AP3 gene in HCoV-NL63 similar to MERS-CoV, may play similar role(s) in pathogenesis.

This communication presents the results obtained from the protein sequence analysis of the spike proteins and the proteins which form the RTC to find out the differences and similarities between the two groups of HCoVs, with special reference to the nature of interactions with the ACE2 receptor and mode of replication.

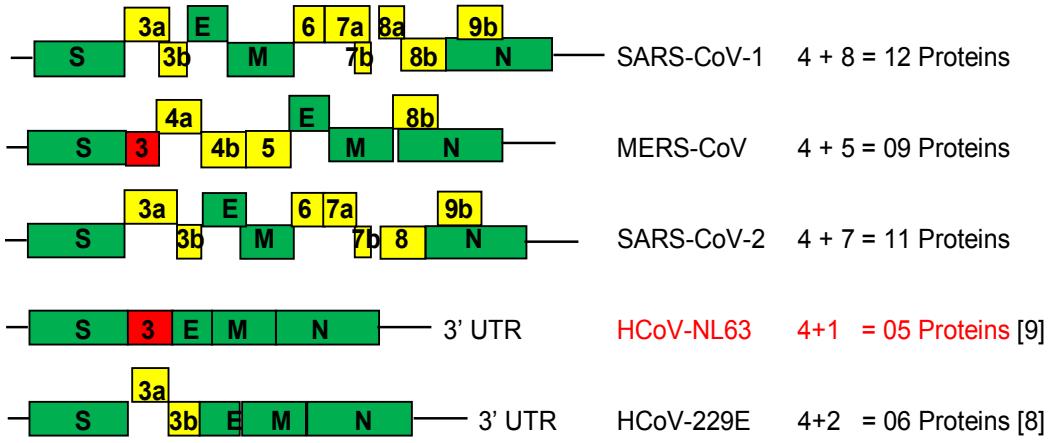


Fig. 1. Organization of APs in different Mild HCoVs and SARS-CoVs

Adapted from Palanivelu [21].

The structural proteins are shown in green boxes and the accessory proteins in yellow boxes. The red ones show the similarity. Upper ones show the overlapping/internal ORFs

A fifth structural protein, the hemagglutinin-esterase is present in a subset of β-coronaviruses. The protein acts as a hemagglutinin and binds onto the sialic acids of surface glycoproteins. These activities are thought to enhance S protein-mediated cell entry and virus spread through the mucosa

2. MATERIALS AND METHODS

A large number of spike protein and RdRp sequences of SARS, SARS--related CoVs and HCoV-NL63 are available in PUBMED and SWISS-PROT databases. For MSA analysis of spike protein and RTC protein sequences were retrieved from SWISS-PROT and PUBMED sites and analyzed using Clustal Omega tool, an accurate, fast and widely accepted algorithm, available on their website. For identification of probable N-linked and O-linked glycosylation sites, NetNGlyc 1.0 and NetOGlyc 4.0 tools of Technical University of Denmark were used. ExPASy tool was used for pI analysis.

3. RESULTS AND DISCUSSION

3.1 Spike Protein Analysis of SARS-CoVs and HCoV-NL63

The spike protein assumes greater importance among the structural proteins as it mediates the very first steps in the infection process, viz. receptor binding, membrane fusion, and viral entry. Furthermore, it is the primary immunogenic target for vaccine development. It is a glycoprotein and is heavily glycosylated. The glycosylation is both N-linked and O-linked and mainly achieved by the host enzymes. In the first step of glycosylation, the N-linked, high mannose oligosaccharide side chains are added co-translationally during protein synthesis [25]. In the next step, the partially glycosylated spike protein is transported from the ER to the Golgi complex where the glycan side chains are modified to more complex forms, and O-linked oligosaccharide side chains are also added [26,27]. Cryo-EM analysis found that the spike protein of HCoV-NL63 is heavily masked by N-linked glycans obstructing the protein surface [28]. By heavily glycosylating their spike proteins; the HCoV-NL63 and other HCoVs evade host immune responses and the N-linked glycans cover a substantial amount of the accessible surface of the trimer [28]. The higher glycan density per accessible surface area detected for the S2 subunits (819 $Å^2$/glycan) compared with the S1 subunits (1,393 $Å^2$/glycan) may explain why most CoV neutralizing antibodies isolated to-date target the latter region. Their cryo-EM and MS data provided evidence for glycosylation at 34 out of 39 possible -**N**XS/T- glycosylation sequons. Most of the HCoV-NL63 S2′ trigger loop, which connects the upstream helix to the fusion peptide, participates in fusion activation. The S1/S2 and S2′ processing occurs at topologically equivalent positions in HCoV-NL63 as in other SARS-CoVs. An unusually high glycan content with 102 N-linked oligosaccharides as compared to 22 N-linked glycan in SARS-CoV-2 spike protein may possibly shield and obstruct the protein surface in HCoV-NL63 [28,29]. A large number of amino acid replacements in the HCoV-NL63 spike protein with Ns support this finding. In fact, most of the Ds and S/Ts are replaced with N in HCoV-NL-63 spike protein (marked in red) (Fig. 2). Codon analysis has shown that both transition and transversion mutations have occurred during evolution for preferentially replacing the regular amino acids with Ns. However, the MSA of HCoV-NL63 spike protein has shown that the HCoV-NL-63 has placed a different set of tetrapeptide, viz. an –^{743}**IPVR**- instead of –-^{681}**PRRA**- which could be one of the reasons for its lesser infectivity as compared to the current epidemic causing strain, SARS-CoV-2.

The spike protein is a trimeric, class I fusion protein which does not require any other viral surface *proteins* for *fusion with* host cells*;* e.g., CoVs, influenza viruses, retroviruses, etc. The spike proteins are essentially composed of two functional domains;- one responsible for receptor binding (S1 domain) and the other is the membrane fusion domain (S2 domain) which facilitates membrane fusion between the viral and host cell membranes. The S1 domain contains the signal peptide followed by the N-terminal domain (NTD) and RBD (333 to 527) whereas the S2 domain consists of the membrane fusion part.

A loop region covering, residues 437 to 506, termed receptor binding motif (RBM), is the only region that is known to make the direct contact with ACE2. The distinct domains in the spike proteins of various SARS and SARS-related CoVs and HCoV-NL63 are shown (marked by arrows) in Fig. 2. In this analysis MERS-CoV spike protein is not included in the MSA, as it binds to a different receptor, viz. DPP4, for human cell entry).

The HCoV-NL63 spike protein is the longest among the SARS and SARS-related CoVs with 1356 amino acids, i.e., 83 amino acids longer than the present epidemic strain SARS-CoV-2 which consists of only 1273 amino acids. The insertions are mainly found in the N-terminal region (highlighted in light green) and in the crucial RBM of the RBD region that makes another important difference between SARS-

CoV-2 and HCoV-NL63. The RBD region contains the main three insertions; out of which two are found in the RBM region. Interestingly, about one-fourth (24 amino acids) of the total insertions are found exclusively in the RBM of the RBD. The inserts found in HCoV-NL63 spike protein appears to be unique as they were not detected in other HCoVs like, HCoV-229E, HCoV-OC43, HCoV-HKU1 spike proteins and also not detected in MERS-CoV spike protein (data not shown) even though MERS-CoV and HCoV-NL63 RTCs showed similarities. The C-terminal contains the addition of two small peptides and a deletion of a peptide (Fig. 2). The above data suggests that the HCoV-NL63 spike protein may bind the ACE2 receptor less efficiently and a different tetrapeptide just in front of the S1/S2 cleavage may also lead to less efficient cleavage for subsequent human cell entry. However, considering the frequent HCoV-NL63 infections of humans and high mutation rates of CoVs, there is an ample opportunity for the HCoV-NL63 to evolve further.

There are two deletions just before and one after the S1/S2 cleavage site. The most intriguing finding is the tetrapeptide insertion (highlighted in light blue) in HCoV-NL63 **(-IVPR**-) just before the S1/S2 cleavage site as in SARS-CoV-2(-**PRRA**-). The S1/S2 cleavage site is completely conserved in both the groups. The next cleavage site, viz. the S2' is also completely conserved though the region is not highly conserved cleavage at S2' lead to conformational rearrangements, which ultimately result in fusion of the host and viral cell membranes, and delivery of the viral genome into the infected host cell. Very close to the S2' cleavage site is the fusion peptide which is conserved. Presence of highly conserved Cs (highlighted in orange) towards the C-terminal end, suggests a Zn binding motif in both the group of HCoVs (Fig. 2). The unique tetrapeptide found in HCoV-NL63 spike protein, viz. –IVPR- was also not detected in all the above 4 spike proteins (data not shown)

Lin et al. [30] have mapped a minimal RBD region in HCoV-NL63 that is required to bind to ACE2 receptor. They found, it consisted of 141 residues (amino acids from 476 to 616) towards the C-terminus of the S1 domain. Interestingly, their data also demonstrated that the HCoV-NL63 RBD bound to hACE2 more efficiently than its full-length counterpart and had a binding efficiency comparable to the S1 or RBD of SARS-CoV-1. Crystallographic analysis [31] of the HCoV-NL63 RBD showed a novel β-sandwich core structure consisting of 2 layers of β-sheets which stack tightly by hydrophobic interactions and 3 discontinuous RBMs to bind ACE2 (the RBMs consisting of ~70 amino acids in SARS-CoVs are continuous). It is interesting to note that the HCoV-NL63 and SARS-CoVs did not show any structural homology in RBD cores or RBMs; yet the 2 viruses recognize common ACE2 regions, largely because of a ''virus-binding hotspot'' on ACE2. The lack of structural homology in the RBMs indicates two independent ways in which HCoV-NL63 and SARS-CoVs recognize their common ACE2 receptor protein [31].

3.2 N- and O-Linked Glycosylation sites in the Spike Proteins of SARS-CoVs and HCoV-NL63

Table 2 shows the probable N-and O-linked glycosylation sites in the SARS-CoVs and HCoV-NL63 spike proteins as predicted by NetNGlyc (-**N**-X-S/T- sequons are used by the prediction tool) and NetOGlyc tools. The large number of replacements by Ns in the spike protein sequence of the HCoV-NL63 are reflected in the number of glycosylation sites too, i.e., the probable N-linked glycosylation sites are almost doubled in HCoV-NL63 spike protein as compared to the current epidemic causing strain SARS-CoV-2 and 5 times more in the RBD region alone (Table 2). Except SARS-CoV-1, all the three HCoVs showed three O–linked glycosylations at different positions, whereas the SARA-CoV-1 showed only one O–linked glycosylation site (Table 2). The number of N-linked glycosylation sites in SARS-CoV-2 spike protein also agrees with MS data [27].

3.3 Basicity of HCoV Spike Proteins and Infection Rates

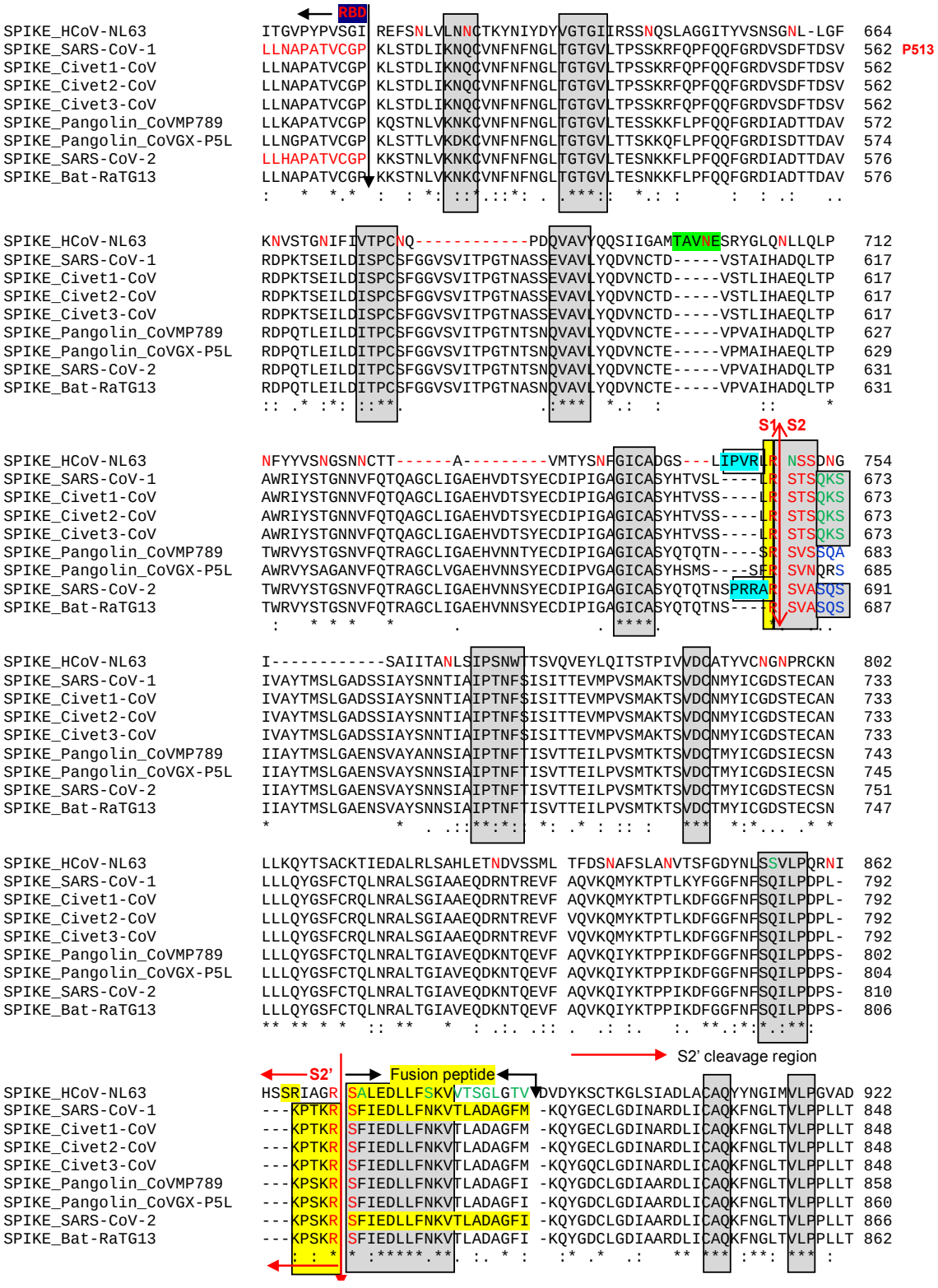
As MSA analysis revealed replacements of a number of amino acids with basic amino acids in SARS-CoV-2 spike protein, the pI values of the spike proteins were analyzed (Table 3). It seems from the Table 3, that as the pI value and basicity of the spike proteins increase, the infection and death rates also increase dramatically suggesting a possible correlation between them. It is interesting to note that the optimum pH for the ACE2 carboxypeptidase is 6.5 [32] and with increasing pI values, the substrate of the enzyme (the spike protein in this case), is nearing ACE2's optimum pH leading to higher-affinity binding to the receptor which may enable effective cell entry and hence possibly leading to a rapid

transmission of CoV-2 as compared to CoV-1. In the presently circulating variant D614G, there is an increase in pI value nearing the optimum pH of the ACE2 receptor which may alter the binding affinity favourably. The CoV-1's pI is only 5.56, and is far away from the optimal pH of the ACE2 receptor enzyme and hence could bind with lesser affinity resulting in less number of cases and deaths. From the C-terminal end sequences of the spike proteins of SARS-CoV-1 and SARS-CoV-2 (-PVLKGVKLHY**T**) and HCoV-NL63 (-PYYEFEKVHV**Q**) and from the ACE-2 enzyme's sequence specificity for activity, it seems that it is not going to make possibly any cleavage at the C-terminal end of the spike proteins.

CLUSTAL O (1.2.4) MSA of HCoV spike proteins which use ACE2 receptor for human cell entry

```
SPIKE_HCoV-NL63            MKLFLILLVLPLA--SCFFTCNSNANLSMLQLGVPDNSSTIVTGLLPTHWICANQSTSVY   58
SPIKE_SARS-CoV-1           --MFIFLLFLTLTSGSDLDRCTTFDDVQAPN---YTQ--------------------HTS   35
SPIKE_Civet1-CoV           --MFIFLLFLTLTSGSDLDRCTTFDDVQAPN---YTQ--------------------HTS   35
SPIKE_Civet2-CoV           --MFIFLLFLTLTSGSDLDRCTTFDDVQAPN---YTQ--------------------HTS   35
SPIKE_Civet3-CoV           --MFIFLLFLTLTSGSDLDRCTTFDDVQAPN---YTQ--------------------HTS   35
SPIKE_Pangolin_CoVMP789    -MLFFFFLHFALVNS----QCVNLTGRAAIQ---PSF--------------------TNS   32
SPIKE_Pangolin_CoVGX-P5L   --MFVFLFVLPLVSS----QCVNLTTRTGIP---PGY--------------------TNS   31
SPIKE_SARS-CoV-2           --MFVFLVLLPLVSS----QCVNLTTRTQLP---PAY--------------------TNS   31
SPIKE_Bat-RaTG13           --MFVFLVLLPLVSS----QCVNLTTRTQLP---PAY--------------------TNS   31
                             :*.::. : *.       * .

SPIKE_HCoV-NL63            SANGFFYIDVGNHRSAFALHTGY----YDVNQYYIYVTNEIGLNASVTLKICKFGINTTF   114
SPIKE_SARS-CoV-1           SMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGF-HTIN-------HTFGNPVIPFKDGI   87
SPIKE_Civet1-CoV           SMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGF-HTIN-------HTFDNPVIPFKDGI   87
SPIKE_Civet2-CoV           SMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGF-HTIN-------HTFDNPVIPFKDGI   87
SPIKE_Civet3-CoV            SMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGF-HTIN-------HTFDNPVIPFKDGI  87
SPIKE_Pangolin_CoVMP789    SQRGVYYPDTIFRSNTLVLSQGYFLPFYSNVSWY-YALTKTN-SAEKRVDNPVLDFKDGI   90
SPIKE_Pangolin_CoVGX-P5L   STRGVYYPDKVFRSSILHLTQDLFLPFFSNVTWF-NTINYQG--GFKKFDNPVLPFNDGV   88
SPIKE_SARS-CoV-2           FTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWF-HAIHVSGTNGTKRFDNPVLPFNDGV   90
SPIKE_Bat-RaTG13           STRGVYYPDKVFRSSVLHLTQDLFLPFFSNVTWF-HAIHVSGTNGIKRFDNPVLPFNDGV   90
                             .*.:* *   : . :    .     :.    :  .           .    : ::  .

SPIKE_HCoV-NL63            DFLSNSSSSFDCIVNLLFTEQLGAPLGITISGETVRLHLYNV------------------   156
SPIKE_SARS-CoV-1           YFAATEK--SNVVRGWV--------FGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPF   137
SPIKE_Civet1-CoV           YFAATEK--SNVVRGWV--------FGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPF   137
SPIKE_Civet2-CoV           YFAATEK--SNVVRGWV--------FGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPF   137
SPIKE_Civet3-CoV           YFAATEK--SNVVRGWV--------FGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPF   137
SPIKE_Pangolin_CoVMP789    YFAATEK--SNIVRGWI--------FGTTLDNTSQSLLIVNNATNVIIKVCNFQFCYDPY   140
SPIKE_Pangolin_CoVGX-P5L   YFASTEK--SNIIRGWI--------FGTTLDARTQSLLIVNNATNVVIKVCEFQFCTDPF   138
SPIKE_SARS-CoV-2           YFASTEK--SNIIRGWI--------FGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPF   140
SPIKE_Bat-RaTG13           YFASTEK--SNIIRGWI--------FGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPF   140
                            * :...   : : . :        :* *:.  :  : : *

SPIKE_HCoV-NL63            -------TRTFYVPAAYKLTKLSVKCYFNYSCVFSVVNATVTVNVTTHNGRVVNY--TVC   207
SPIKE_SARS-CoV-1           FAVSKPMGT----QTHTMIFDNAF------NCTFEYISDAFSLDVSEKSGNFKHLREFVF   187
SPIKE_Civet1-CoV           FVVSKPMGT----RTHTMIFDNAF------NCTFEYISDAFSLDVSEKSGNFKHLREFVF   187
SPIKE_Civet2-CoV           FVVSKPMGT----QTHTMIFDNAF------NCTFEYISDAFSLDVSEKSGNFKHLREFVF   187
SPIKE_Civet3-CoV           FVVSKPMGT----QTHTMIFDNAF------NCTFEYISDAFSLDVSEKSGNFKHLREFVF   187
SPIKE_Pangolin_CoVMP789    LSGYYHN-NKTWSTREFAVYSSYA------NCTFEYVSKSFMLDIAGKSGLFDTLREFVF   193
SPIKE_Pangolin_CoVGX-P5L   LGVYYHNNNKTWVENEFRVYSSAN------NCTFEYISQPFLMDLEGKQGNFKNLREFVF   192
SPIKE_SARS-CoV-2           LGVYYHKNNKSWMESEFRVYSSAN------NCTFEYVSQPFLMDLEGKQGNFKNLREFVF   194
SPIKE_Bat-RaTG13           LGVYYHKNNKSWMESEFRVYSSAN------NCTFEYVSQPFLMDLEGKQGNFKNLREFVF   194
                                           : .          .*.*. :.  . :::  :.* .      *

SPIKE_HCoV-NL63            DDCNGYTDN---IFSVQQDGRIPNGFPFNNW-FLLTNGSTLVDGVSRLYQPLRLTCLWPV   263
SPIKE_SARS-CoV-1           KNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIN----ITNFRAILTAFS----   239
SPIKE_Civet1-CoV           KNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIN----ITNFRAILTAFS----   239
SPIKE_Civet2-CoV           KNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIK----ITNFRAILTAFS----   239
SPIKE_Civet3-CoV           KNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIK----ITNFRAILTAFS----   239
SPIKE_Pangolin_CoVMP789    RNVDGYFKIYSKYTPVNVNSNLPIGFSALEPLVEIPAGIN----ITKFRTLLTIHRGDPM   249
SPIKE_Pangolin_CoVGX-P5L   KNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGIN----ITRFQTLLALHRSYLT   248
SPIKE_SARS-CoV-2           KNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGIN----ITRFQTLLALHRSYLT   250
SPIKE_Bat-RaTG13           KNIDGYFKIYSKHTPINLVRDLPPGFSALEPLVDLPIGIN----ITRFQTLLALHRSYLT   250
                            : :*:         ::    :* **   :  . :  * .    ::.:   *
```

```
SPIKE_HCoV-NL63            PGLKSSTGFVYFNATGSDVNCNGYQHNSVADVMRYNLNFSANSVDNLKSGVIVFKTLQYD   323
SPIKE_SARS-CoV-1           --PAQD-IWGTS----AAAYFVGYLKP-----TTFMLKYDEN-----------------   269
SPIKE_Civet1-CoV           --PAQD-TWGTS----AAAYFVGYLKP-----TTFMLKYDEN-----------------   269
SPIKE_Civet2-CoV           --PAQG-TWGTS----AAAYFVGYLKP-----TTFMLKYDEN-----------------   269
SPIKE_Civet3-CoV           --PAQG-TWGTS----AAAYFVGYLKP-----TTFMLKYDEN-----------------   269
SPIKE_Pangolin_CoVMP789    P---NN-GWTVF----SAAYYVGYLAP-----RTFMLNYNEN-----------------   278
SPIKE_Pangolin_CoVGX-P5L   PGKLES-GWTTG----AAAYYVGYLQQ-----RTFLLSYNQN-----------------   280
SPIKE_SARS-CoV-2           PGDSSS-GWTAG----AAAYYVGYLQP-----RTFLLKYNEN-----------------   282
SPIKE_Bat-RaTG13           PGDSSS-GWTAG----AAAYYVGYLQP-----RTFLLKYNEN-----------------   282
                               ..  :       : .   **          : *.:. *

SPIKE_HCoV-NL63            VLFYCSNSSSGVLDTTI--PFGPSSQPYYCFINSTINTTHVSTFVGVLPPTVREIVVART   381
SPIKE_SARS-CoV-1           ----------GTITDAVDCSQNPLAE-LKCSVKS--------------------F-EIDK   297
SPIKE_Civet1-CoV           ----------GTITDAVDCSQNPLAE-LKCSVKS--------------------F-EIDK   297
SPIKE_Civet2-CoV           ----------GTITDAVDCSQNPLAE-LKCSVKS--------------------F-EIDK   297
SPIKE_Civet3-CoV           ----------GTITDAVDCSQNPLAE-LKCSVKS--------------------F-EIDK   297
SPIKE_Pangolin_CoVMP789    ----------GTITDAVDCALDPLSE-AKCTLKS--------------------L-TVEK   306
SPIKE_Pangolin_CoVGX-P5L   ----------GTITDAVDCSLDPLSE-TKCTLKS--------------------L-TVEK   308
SPIKE_SARS-CoV-2           ----------GTITDAVDCALDPLSE-TKCTLKS--------------------F-TVEK   310
SPIKE_Bat-RaTG13           ----------GTITDAVDCALDPLSE-TKCTLKS--------------------F-TVEK   310
                                     *.:  ::    .* ::   * ::*                    :    .

                                                RBD →
SPIKE_HCoV-NL63            GQFYINGFKYFDLGFIE-----|------AVNFNVTTASATDFWTVAFATFVDVLVNVSAT   430
SPIKE_SARS-CoV-1           GIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWERKKI------SN--CV   349
SPIKE_Civet1-CoV           GIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWERKRI------SN--CV   349
SPIKE_Civet2-CoV           GIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWERKRI------SN--CV   349
SPIKE_Civet3-CoV           GIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWERKRI------SN--CV   349
SPIKE_Pangolin_CoVMP789    GIYQTSNFRVQPTESIVRFPNI|TNLCPFGEVFNATTFASVYAWNRKRI------SN--CV   358
SPIKE_Pangolin_CoVGX-P5L   GIYQTSNFRVQPTISIVRFPNI|TNLCPFGEVFNASKFASVYAWNRKRI------SN--CV   360
SPIKE_SARS-CoV-2           GIYQTSNFRVQPTESIVRFPNI|TNLCPFGEVFNATRFASVYAWNRKRI------SN--CV   362
SPIKE_Bat-RaTG13           GIYQTSNFRVQPTDSIVRFPNI|TNLCPFGEVFNATTFASVYAWNRKRI------SN--CV   362
                           * :  ..*:     :          .  **.:   :.  *           *  ..
                                                    |
SPIKE_HCoV-NL63            KIQNLLYCDSPFEKLQCEHLQFGLQ-DGFYSANFLDDNVLPETYV---ALPIYYQHTDIN   486
SPIKE_SARS-CoV-1           ADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYN   409
SPIKE_Civet1-CoV           ADYSVLYNSTSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYN   409
SPIKE_Civet2-CoV           ADYSVLYNSTSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYN   409
SPIKE_Civet3-CoV           ADYSVLYNSTSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYN   409
SPIKE_Pangolin_CoVMP789    ADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVRQIAPGQTGRIADYN   418
SPIKE_Pangolin_CoVGX-P5L   ADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQTGVIADYN   420
SPIKE_SARS-CoV-2           ADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYN   422
SPIKE_Bat-RaTG13           ADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVRQIAPGQTGKIADYN   422
                              .:** .: *..::*  :.     *  ::  : *. *:    *   *       :* *

                                     RBM →
                                     |
SPIKE_HCoV-NL63            FTATASFGGSCYVC|KPHQVNISLNGNTSVCVRTSHFSIRYIYNRVKSGSPGDSSWHIYLK   546
SPIKE_SARS-CoV-1           YKLPDDFMGCVLAW|NTRNIDATSTGN---------Y--NYKY-------------RYLRH   445
SPIKE_Civet1-CoV           YKLPDDFMGCVLAW|NTRNIDATSTGN---------Y--NYKY-------------RYLRH   445
SPIKE_Civet2-CoV           YKLPDDFMGCVLAW|NTRNIDATSTGN---------Y--NYKY-------------RYLRH   445
SPIKE_Civet3-CoV           YKLPDDFMGCVLAW|NTRNIDATSTGN---------Y--NYKY-------------RYLRH   445
SPIKE_Pangolin_CoVMP789    YKLPDDFTGCVIAW|NSNNLDSKVGGN---------Y--NYLY-------------RLFRK   454
SPIKE_Pangolin_CoVGX-P5L   YKLPDDFTGCVIAW|NSVKQDALTGGN---------Y--GYLY-------------RLFRK   456
SPIKE_SARS-CoV-2           YKLPDDFTGCVIAW|NSNNLDSKVGGN---------Y--NYLY-------------RLFRK   458
SPIKE_Bat-RaTG13           YKLPDDFTGCVIAW|NSKHIDAKEGGN---------F--NYLY-------------RLFRK   458
                           :.   .* *.  . :  : :    **         :   * *             :   :
                                                                  ← RBM
                                                                  |
SPIKE_HCoV-NL63            SGTCPFSFSKLN-NFQKFKTICFSTVAVPGSCNFPLEATWHYTSYTIVGA|LYVTWSEGNS   605
SPIKE_SARS-CoV-1           GKLRPFERDISNVPFSPDGKPCTP---PALNCYWPLNDYGFYTTTGIGYQ|PYRVVVLSFE   502
SPIKE_Civet1-CoV           GKLRPFERDISNVPFSPDGKPCTP---PAPNCYWPLRGYGFYTTSGIGYQ|PYRVVVLSFE   502
SPIKE_Civet2-CoV           GKLRPFERDISNVPFSSDGKPCTP---PAPNCYWPLRGYGFYTTSGIGYQ|PYRVVVLSFE   502
SPIKE_Civet3-CoV           GKLRPFERDISNVPFSSDGKPCTP---PAPNCYWPLRGYGFYTTSGIGYQ|PYRVVVLSFE   502
SPIKE_Pangolin_CoVMP789    SNLKPFERDISTEIYQAGSTPCNG--VEGFNCYFPLQSYGFHPTNGVGYQ|PYRVVVLSFE   512E484K
SPIKE_Pangolin_CoVGX-P5L   SKLKPFERDISTEIYQAGSTPCNG--QVGLNCYYPLERYGFHPTTGVNYQ|PFRVVVLSFE   514
SPIKE_SARS-CoV-2           SNLKPFERDISTEIYQAGSTPCNG--VEGFNCYFPLQSYGFQPTNGVGYQ|PYRVVVLSFE   516N501Y
SPIKE_Bat-RaTG13           ANLKPFERDISTEIYQAGSKPCNG—-QTGLNCYYPLYRYGFYPTDGVGHQ|PYRVVVLSFE   516
                           .   **. .  .  :.   . *        .* :**    .  :  :   : .   . .
```


RBD
S1
S2
S2' cleavage region
S2'
Fusion peptide

```
SPIKE_HCoV-NL63            AERMAMYTGSLIGGMVLGGLT----SAAAIPFSLALQARLNYVALQTDVLQENQKILAAS   978
SPIKE_SARS-CoV-1           DDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQ   908
SPIKE_Civet1-CoV           DDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQ   908
SPIKE_Civet2-CoV           DDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQ   908
SPIKE_Civet3-CoV           DDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQ   908
SPIKE_Pangolin_CoVMP789    DEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ   918
SPIKE_Pangolin_CoVGX-P5L   DEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ   920
SPIKE_SARS-CoV-2           DEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ   926
SPIKE_Bat-RaTG13           DEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ   922
                            : :* **.:*:.*   .* *    :*  ***:: :  *:* :.:  :** **** :* .

SPIKE_HCoV-NL63            FNKAINNIVASFSSVNDAITQTAEAIHTVTIALNKIQDVVNQQGSALNHLTSQLRHNFQA   1038
SPIKE_SARS-CoV-1           FNKAISQIQESLT-------------TTSTALGKLQDVVNQNAQALNTLVKQLSSNFGA   954
SPIKE_Civet1-CoV           FNKAISQIQESLT-------------TTSTALGKLQDVVNQNAQALNTLVKQLSSNFGA   954
SPIKE_Civet2-CoV           FNKAISQIQESLT-------------TTSTALGKLQDVVNQNAQALNTLVKQLSSNFGA   954
SPIKE_Civet3-CoV           FNKAISQIQESLT-------------TTSTALGKLQDVVNQNAQALNTLVKQLSSNFGA   954
SPIKE_Pangolin_CoVMP789    FNSAIGKIQDSLS-------------STASALGKLQDVVNQNAQALNTLVKQLSSNFGA   964
SPIKE_Pangolin_CoVGX-P5L   FNSAIGKIQDSLS-------------STASALGKLQDVVNQNAQALNTLVKQLSSNFGA   966
SPIKE_SARS-CoV-2           FNSAIGKIQDSLS-------------STASALGKLQDVVNQNAQALNTLVKQLSSNFGA   972
SPIKE_Bat-RaTG13           FNSAIGKIQDSLS-------------STASALGKLQDVVNQNAQALNTLVKQLSSNFGA   968
                           **.**.:*  *::             :.: **.*:******:..*** *..**  ** *

SPIKE_HCoV-NL63            ISNSIQAIYDRLDSIQADQQVDRLITGRLAALNAFVSQVLNKYTEVRSSRRLAQQKVNEC   1098
SPIKE_SARS-CoV-1           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1014
SPIKE_Civet1-CoV           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1014
SPIKE_Civet2-CoV           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1014
SPIKE_Civet3-CoV           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1014
SPIKE_Pangolin_CoVMP789    ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1024
SPIKE_Pangolin_CoVGX-P5L   ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1026
SPIKE_SARS-CoV-2           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1032
SPIKE_Bat-RaTG13           ISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSEC   1028
                           **. :: * .***.::*: *:******** :*:::*:* * : :*:*:* .**  *:.**

SPIKE_HCoV-NL63            VKSQSNRYGFCGNGTHIFSIVNSAPDGLLFLHTVLLPTDYKNVKAWSGICVDGIYGYVLR   1158
SPIKE_SARS-CoV-1           VLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPRE   1074
SPIKE_Civet1-CoV           VLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPRE   1074
SPIKE_Civet2-CoV           VLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPRE   1074
SPIKE_Civet3-CoV           VLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPRE   1074
SPIKE_Pangolin_CoVMP789    VLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPSQEKNFTTTPAICHEGKAHFPRE   1084
SPIKE_Pangolin_CoVGX-P5L   VLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHEGKAHFPRE   1086
SPIKE_SARS-CoV-2           VLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPRE   1092
SPIKE_Bat-RaTG13           VLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPRE   1088
                           * .**:* .***:* *::*: ::**.*::***.. :*:: :*..:  .** :*   :  .

SPIKE_HCoV-NL63            QPNLVLYSDNGVFRVTSRVMFQPRLPVLSDFVQIYNCNVTFVNISRVELHTVIPDYVDVN   1218
SPIKE_SARS-CoV-1           ---GVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYD-------PLQ   1124
SPIKE_Civet1-CoV           ---GVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYD-------PLQ   1124
SPIKE_Civet2-CoV           ---GVFVFSGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYD-------PLQ   1124
SPIKE_Civet3-CoV           ---GVFVFSGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYD-------PLQ   1124
SPIKE_Pangolin_CoVMP789    ---GVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGSCDVVIGIVNNTVYD-------PLQ   1134
SPIKE_Pangolin_CoVGX-P5L   ---GVFVSNGTHWFITQRNFYEPQIITTDNTFVSGSCDVVIGIVNNTVYD-------PLQ   1136
SPIKE_SARS-CoV-2           ---GVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYD-------PLQ   1142
SPIKE_Bat-RaTG13           ---GVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGSCDVVIGIVNNTVYD-------PLQ   1138
                              *:  ..  : :*.* ::.*:: . .: .   .*:*.:  :...  .        ::

SPIKE_HCoV-NL63            KTLQEFAQNLPKYVKP----NFDLT---PFNLTYLNLSSELKQLEAKTASLFQTTVELQG   1271
SPIKE_SARS-CoV-1           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQK--------------EIDRLNE   1170
SPIKE_Civet1-CoV           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQE--------------EIDRLNE   1170
SPIKE_Civet2-CoV           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQE--------------EIDRLNE   1170
SPIKE_Civet3-CoV           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQE--------------EIDRLNE   1170
SPIKE_Pangolin_CoVMP789    PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQK--------------EIDRLNE   1180
SPIKE_Pangolin_CoVGX-P5L   PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQK--------------EIDRLNE   1182
SPIKE_SARS-CoV-2           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQK--------------EIDRLNE   1188
SPIKE_Bat-RaTG13           PELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQK--------------EIDRLNE   1184
                             *:.* ::* **.*     :.**     :* : :*:..                .*:
```

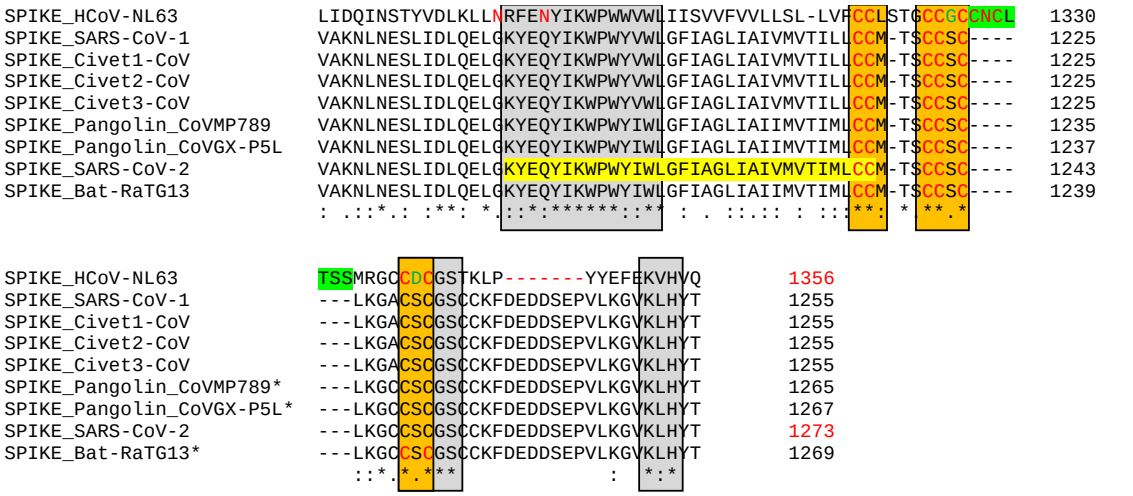


Fig. 2. MSA analysis of the spike proteins of SARS-related CoVs and HCoV-NL63

AFO70497.1, HCoV-NL63; P59594, SARS-CoV-1; AAU04646.1 Civet1 007; AAV49722.1, Civet2 PC4199; AAV91631.1, Civet3 A022; P0DTC2, SARS-CoV-2.

**Pangolin and Bat sequences are from the corresponding genomic sequences.*

Insertions in HCoV-NL63 spike protein are highlighted in green.

SARS-CoV-2 RBD covers amino acids from 333 to 527 and RBM covers the region from 437 to 506.

The spike protein domains S1 and S2 divide at 685/686, marked by a red arrow.

In SARS-CoV-2 Spike protein amino acids 1-685 and 686-1273 form the S1 and S2 domains, respectively.

The 685 and 814 are the S1 and S2' cleavage sites, respectively.

Recent immune evading variants are found in the crucial RBM is shown in the margin (marked in magenta).

Table 2. N-linked glycosylation sites in the SARS-CoVs and HCoV-NL63

Spike protein Receptor	N-Glycosyln sites	O-Glycosyln site(s)/Position
SARS-CoV-1 (ACE2)	22 (2 in RBD)	1/336 (in RBD)
SARS-CoV-2 (ACE2)	22 (1 in RBD*)	3/673,678,686
HCoV-NL63	40 (5 in RBD)	3/864,865,1008
MERS-CoV (DPP4)	25	3/135,206,1243

**In SARS-CoV-2, RBD is mapped from 333 to 527 and RBM covers the region from 437 to 506*

Table 3. pI and basicity values of the spike proteins of HCoVs and death rates

Virus	pI value	Basicity	Year of epidemic	Cases/Deaths
SARS-CoV-1	5.56 (Pc, pI 5.57)		2003	8098/774
*MERS-CoV**	5.73		2013	855/333
SARS-CoV-2 (Wuhan)	6.24 (Pango, pI 6.21)		Dec 2019	>100x10^6/~2.2x10^6
SARS-CoV-2 (D614G)	6.32		July 2020	-----
HCoV-NL63	**6.85**		2004	-----

**MERS-CoV does not use ACE2 receptor but a dipeptidase receptor; included to show only the frequency of the epidemics. Pc, Palm civet (199)-CoV; Pango, Pangolin (GX-P1)-CoV.*

The loop region covering, amino acid residues 437 to 506, termed the receptor-binding motif (RBM) in SARS-CoV-2, is the only region that is known to make the contact with ACE2 directly

From the early February 2020, the original Wuhan strain was replaced by D614G strain and became the dominant form of the virus circulating globally.

Currently, thousands of variants of SARS-CoV-2 are documented. Among them, three variants of concern are the UK B.1.1.7 (N501Y, Δ69/70, P681H) South Africa B.1.351 (N501Y, E484K, K417N) and Brazil P.1 (N501Y, E484K, K417T) where a change of two to three amino acids in the spike protein markedly reduced the efficacy of current vaccines. (The RBM variants are shown in red).

The N-terminal domain (NTD) of the spike protein is known to bind to the sialic acid residues in the RBD of ACE2 receptor [1]. Therefore, the marked increase in the basicity would favour enhanced binding of the SARS-CoV-2. Only the SA and BR variants have all the modifications in the RBM region where it is known to make the contact with ACE2 directly. The UK variant has only one modification in the RBM region and did not also affect the pI value of the spike protein (Table 3). The recent neutralizing antibody variants from SA and BR along with the D614G variant are very close to the optimal pH (6.5) of the ACE2 activity. Though the HCoV-NL63 spike protein pI value is also slightly above the optimum pH of ACE2, a different tetrapeptide at S1/S2 cleavage point and large insertions in the crucial RBM possibly lead to less efficient binding and cleavage, leading to milder disease conditions.

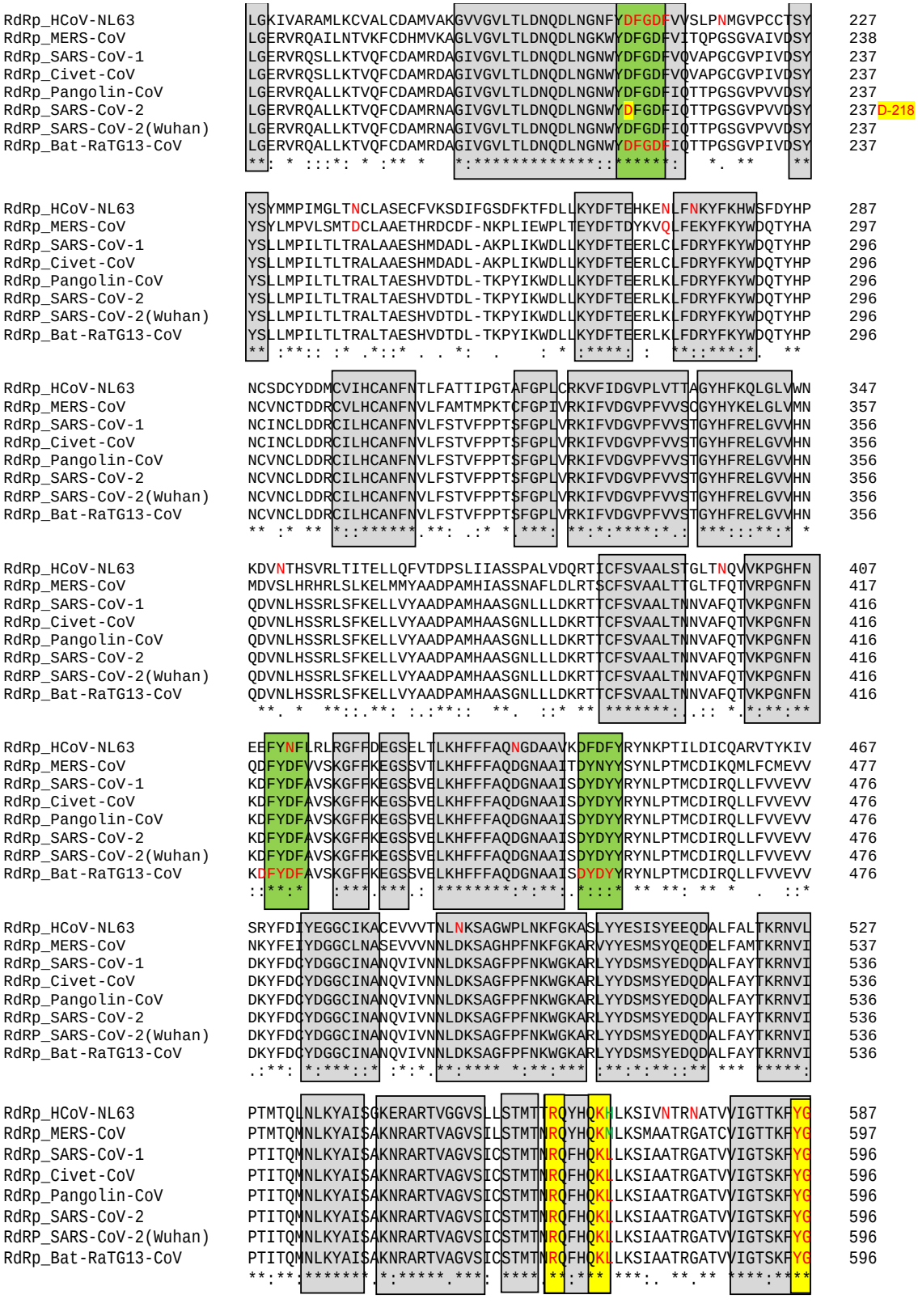
3.4 MSA Analysis of RdRps SARS and SARS-related CoVs and HCoV-NL63

Both (+) and (–) strand RNA viruses employ the enzyme RdRp for multiplication in animal and human cells. RdRp is a crucial enzyme as it performs both transcription and replication processes in all of these RNA viruses (except retro viruses). Therefore, RdRps have been the main target for antiviral drug development to control the RNA based viral pathogens. An in-depth analyses and characterization of the HCoVs will pave way for the development of effective antivirals to contain the present and future pandemics.

Fig. 3 shows the MSA analysis of the RdRps (NSP12) of the SARS, SARS-Related CoVs and HCoV-NL63. The RdRp sequence analysis from both groups of HCoVs reveals conservation of large numbers of amino acids, motifs and peptide regions among them (MERS-CoV sequence is also included in the analysis). In fact, the RdRp protein consists of two nucleotidyltransferase activities, i.e., one is identified in the NiRAN domain (amino acids 1-250, numberings are from the CoV-2 RdRp sequence) and second activity is in the regular RdRp domain (251-932). Some important differences are observed in the NiRAN domains of these two groups, viz. SARS, SARS-related CoVs and HCoV-NL63.

CLUSTAL O (1.2.4) MSA of RdRps of SARS and SARS-related CoVs and HCoV-NL63

```
RdRp_HCoV-NL63            -----SYLNRARG-SSAARLEPCN-GTDIDKCVRAFDIYN--KNVSFLGKCLKMNCVRFK    51
RdRp_MERS-CoV             -SKDSNFLNRVRGSIVNARIEPCSSGLSTDVVFRAFDICNYKAKVAGIGKYYKTNTCRFV    59
RdRp_SARS-CoV-1           SADASTFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NEKVAGFAKFLKTNCCRFQ    57
RdRp_Civet-CoV            SADASTFFKRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NEKVAGFAKFLKTNCCRFQ    57
RdRp_Pangolin-CoV         SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ    57
RdRp_SARS-CoV-2(Std)      SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ    57
RdRP_SARS-CoV-2(Wuhan)    SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ    57
RdRp_Bat-RaTG13-CoV       SADAQSFLNRVCG-VSAARLTPCGTGTSTDVVYRAFDIY--NDKVAGFAKFLKTNCCRFQ    57
                                .:::*. *    **: **. * . *   *****     :*: :.*  * *  **

RdRp_HCoV-NL63            NA----DLKDGYFVIKRCTKSVMEHEQSMYNLLNFSGALAEHDFFTWKDGRVIYGNVSRH    107
RdRp_MERS-CoV             ELDDQGHHLDSYFVVKRHTMENYELEKHCYDLLRDCDAVAPHDFFIFDVDKVKTPHIVRQ    119
RdRp_SARS-CoV-1           EKDEEGNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQ    117
RdRp_Civet-CoV            EKDEEGNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQ    117
RdRp_Pangolin-CoV         EKDEDGNLIDSYFIVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ    117
RdRp_SARS-CoV-2           EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ    117
RdRP_SARS-CoV-2(Wuhan)    EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ    117
RdRp_Bat-RaTG13-CoV       EKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQ    117
                          :     .  *.**::** * .  : *:  *:*:. . *:* **** :  .   :: *:

RdRp_HCoV-NL63            NLTKYTMMDLVYAMRNFDEQNCDVLKEVLVLTGCCDNSYFDSKGWYDPVENEDIHRVYAS    167
RdRp_MERS-CoV             RLTEYTMMDLVYALRHFDQN-SEVLKAILVKYGCCDVTYFENKLWFDFVENPSVIGVYHK    178
RdRp_SARS-CoV-1           RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN    177
RdRp_Civet-CoV            RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN    177
RdRp_Pangolin-CoV         RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDEYFNKKDWYDFVENPDILRVYAN    177
RdRp_SARS-CoV-2           RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN    177
RdRP_SARS-CoV-2(Wuhan)    RLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN    177
RdRp_Bat-RaTG13-CoV       RLTKYTMADLVYALRHFDEGNCDTLREILVTYNCCDDDYFNKKDWYDFVENPDILRVYAN    177
                          .**:*** *****:*:**:  .:.*: :**  .***  **:.* *:* *** .:  ** .
```


RdRp_HCoV-NL63 LGKIVARAMLKCVALCDAMVAKGVVGVLTLDNQDLNGNFYDFGDFVVSLPNMGVPCCTSY 227
RdRp_MERS-CoV LGERVRQAILNTVKFCDHMVKAGLVGVLTLDNQDLNGKWYDFGDFVITQPGSGVAIVDSY 238
RdRp_SARS-CoV-1 LGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSY 237
RdRp_Civet-CoV LGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSY 237
RdRp_Pangolin-CoV LGERVRQALLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY 237
RdRp_SARS-CoV-2 LGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY 237D-218
RdRP_SARS-CoV-2(Wuhan) LGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSY 237
RdRp_Bat-RaTG13-CoV LGERVRQALLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPIVDSY 237
RdRp_HCoV-NL63 YSYMMPIMGLTNCLASECFVKSDIFGSDFKTFDLLKYDFTEHKENLFNKYFKHWSFDYHP 287
RdRp_MERS-CoV YSYLMPVLSMTDCLAAETHRDCDF-NKPLIEWPLTEYDFTDYKVQLFEKYFKYWDQTYHA 297
RdRp_SARS-CoV-1 YSLLMPILTLTRALAAESHMDADL-AKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHP 296
RdRp_Civet-CoV YSLLMPILTLTRALAAESHMDADL-AKPLIKWDLLKYDFTEERLCLFDRYFKYWDQTYHP 296
RdRp_Pangolin-CoV YSLLMPILTLTRALTAESHVDTDL-TKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHP 296
RdRp_SARS-CoV-2 YSLLMPILTLTRALTAESHVDTDL-TKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHP 296
RdRP_SARS-CoV-2(Wuhan) YSLLMPILTLTRALTAESHVDTDL-TKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHP 296
RdRp_Bat-RaTG13-CoV YSLLMPILTLTRALTAESHVDTDL-TKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHP 296
RdRp_HCoV-NL63 NCSDCYDDMCVIHCANFNTLFATTIPGTAFGPLCRKVFIDGVPLVTTAGYHFKQLGLVWN 347
RdRp_MERS-CoV NCVNCTDDRCVLHCANFNVLFAMTMPKTCFGPIVRKIFVDGVPFVVSCGYHYKELGLVMN 357
RdRp_SARS-CoV-1 NCINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRp_Civet-CoV NCINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRp_Pangolin-CoV NCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRp_SARS-CoV-2 NCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRP_SARS-CoV-2(Wuhan) NCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRp_Bat-RaTG13-CoV NCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHN 356
RdRp_HCoV-NL63 KDVNTHSVRLTITELLQFVTDPSLIIASSPALVDQRTICFSVAALSTGLTNQVVKPGHFN 407
RdRp_MERS-CoV MDVSLHRHRLSLKELMMYAADPAMHIASSNAFLDLRTSCFSVAALTTGLTFQTVRPGNFN 417
RdRp_SARS-CoV-1 QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRp_Civet-CoV QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRp_Pangolin-CoV QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRp_SARS-CoV-2 QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRP_SARS-CoV-2(Wuhan) QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRp_Bat-RaTG13-CoV QDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFN 416
RdRp_HCoV-NL63 EEFYNFLRLRGFFDEGSELTLKHFFFAQNGDAAVKDFDFYRYNKPTILDICQARVTYKIV 467
RdRp_MERS-CoV QDFYDFVVSKGFFKEGSSVTLKHFFFAQDGNAAITDYNYYSYNLPTMCDIKQMLFCMEVV 477
RdRp_SARS-CoV-1 KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRp_Civet-CoV KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRp_Pangolin-CoV KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRp_SARS-CoV-2 KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRP_SARS-CoV-2(Wuhan) KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRp_Bat-RaTG13-CoV KDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVV 476
RdRp_HCoV-NL63 SRYFDIYEGGCIKACEVVVTNLNKSAGWPLNKFGKASLYYESISYEEQDALFALTKRNVL 527
RdRp_MERS-CoV NKYFEIYDGGCLNASEVVVNNLDKSAGHPFNKFGKARVYYESMSYQEQDELFAMTKRNVI 537
RdRp_SARS-CoV-1 DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRp_Civet-CoV DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRp_Pangolin-CoV DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRp_SARS-CoV-2 DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRP_SARS-CoV-2(Wuhan) DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRp_Bat-RaTG13-CoV DKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVI 536
RdRp_HCoV-NL63 PTMTQLNLKYAISGKERARTVGGVSLLSTMTTRQYHQKMLKSIVNTRNATVVIGTTKFYG 587
RdRp_MERS-CoV PTMTQMNLKYAISAKNRARTVAGVSILSTMTNRQYHQKNLKSMAATRGATCVIGTTKFYG 597
RdRp_SARS-CoV-1 PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596
RdRp_Civet-CoV PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596
RdRp_Pangolin-CoV PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596
RdRp_SARS-CoV-2 PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596
RdRP_SARS-CoV-2(Wuhan) PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596
RdRp_Bat-RaTG13-CoV PTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYG 596

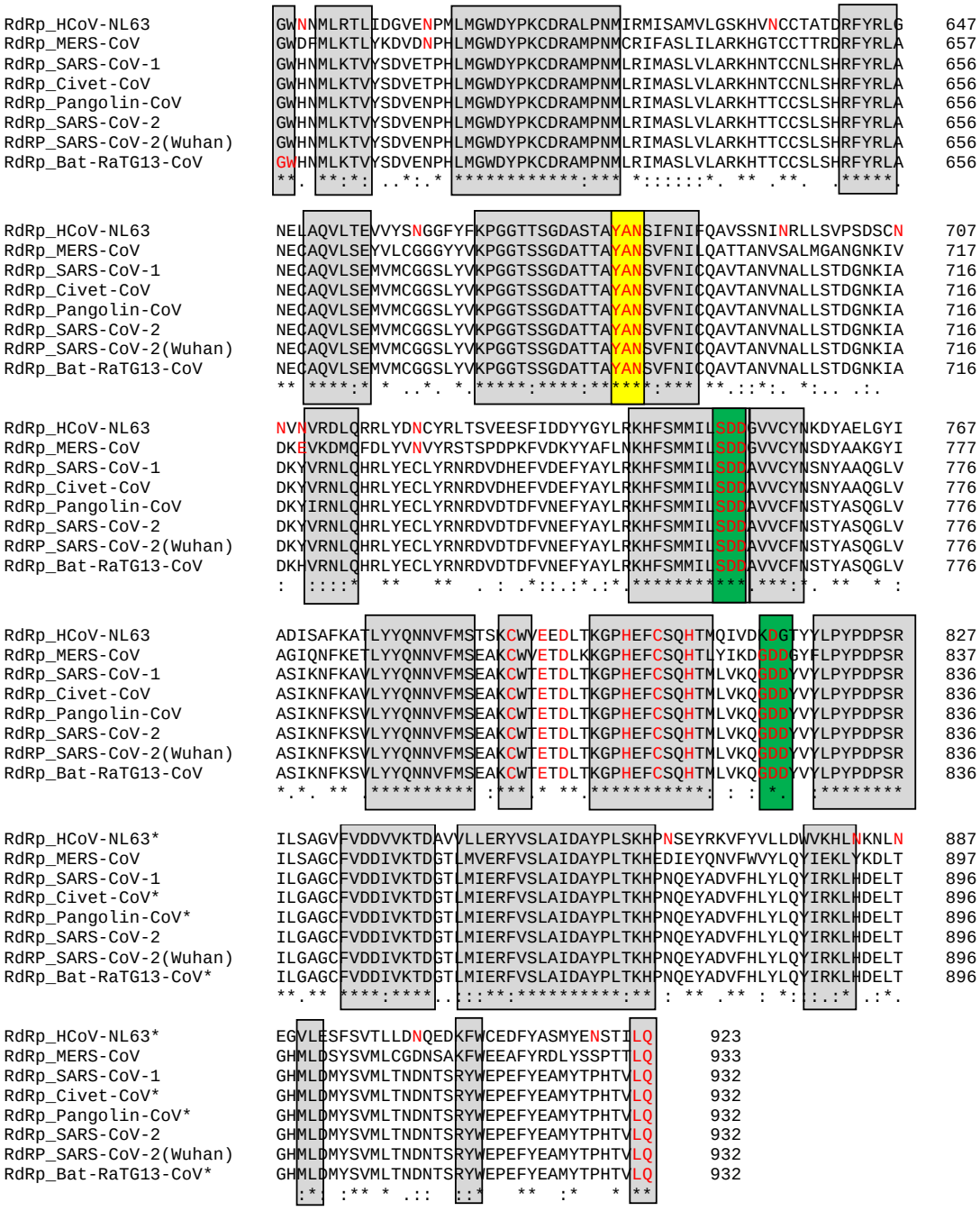


Fig. 3. MSA analysis of the RdRps of SARS and SARS-related CoVs and HCoV-NL63

YP_009047223.1 MERS; NP_828869.1 CoV1; YP_009725307.1 CoV2;; NC_045512 Wuhan Ref
**HCoV-Nl63; *Pangolin-CoV, *Civet-CoV and *Bat CoV are obtained from the corresponding genome sequences*

Firstly, the proposed catalytic **R**, found in the NiRAN domain of other SARS, SARS-related CoVs, is replaced by an **N** in HCoV-NL63. Secondly, in the second NTP selection site –YA**N**-, the N is replaced by an **S** in HCoV-NL63. Thirdly, a -KDG- is found only in the HCoV-NL63 NiRAN domain (but slightly out of phase) very close to one of the proposed metal binding sites, viz. the –DxD- that is also found in the RdRp domain of all the other RdRps. However, the second –DxxD- (as -DFGDF-) is found in the completely conserved block of the NiRAN

domains of both the types (Fig. 3), in which (the possible metal binding sites are highlighted in light green in the NiRAN domains). SDM analysis of D^{218}→A (numbering from SARS-CoV-2) in the –D^{218}xxD- box was found to be essential for NiRAN activity [33].

The active sites in the second domain of the RdRp are completely conserved in both types of the CoVs;- For example, the catalytic region amino acids are completely conserved in all the RdRps and follow the same pattern as (**R^{-5}------K----------YG^{22}**); one of the metal binding sites, –SDD-, is conserved in all whereas the second universal metal binding site, –G**D**D-, is slightly modified as –K**D**G- in HCoV-NL63; and only the second **D** is conserved in both the types (highlighted in dark green); the NTP selection site –YAN- is also completely conserved in all (highlighted in yellow). A possible Zn^{2+} binding motif is found between the two metal binding sites (marked in red) in both the groups. A few highly conserved peptide regions are found in the C-terminal region too. A detailed analysis of the NiRAN and RdRp domains has been specified previously [21]. It is interesting to note that a large number of amino acids in HCoV-NL-63 are replaced by N (marked in red) as found in the spike protein (Fig. 2).

The HCoV-NL63 and MERS-CoV show similarities in many places. The –LQ diad appearing at the very end of the protein is the main protease cleavage site during the polyprotein processing and is completely conserved in all suggesting that the M^{pro} specificity is not altered in both the groups.
The possible template-binding pair, the catalytic amino acid and the invariant basic amino acid at -5 from the catalytic amino acid for NTP selection are completely conserved in the RdRp domains of both the groups (Table 4). However, the NiRAN domains exhibit complete conservation only among the SARS-CoVs. In the HCoV-NL63, the catalytic and NTP selection amino acids R and H are replaced by Ns (marked in green and highlighted) whereas the template binding -YA- pair is conserved in all (Table 4).

3.5 MSA of the primases (NSP8) of SARS and SARS-related CoVs and HCoV-NL63

It has been shown by several investigators that the RdRp of SARS and SARS-related CoVs are a primer-dependent enzyme. It forms the active RTC holoenzyme by complexing with the primase (NSP8) and an associated small molecular weight protein NSP7 [34,2,35-37].

Fig. 4 shows the MSA analysis of the primases in SARS and SARS-related CoVs and HCoV-NL63 (MERS-CoV sequence is also included in the analysis). Primases from both the groups follow the same -**KL**----------**YA**- pattern suggesting the template-binding and catalytic pairs with the conspicuous absence of a basic NTP selection amino acid at around -5, as found in all RdRps (Figs. 3 and 4). This agrees with the finding by Imbert et al [38] that the primases are not strict in their NTP selection. The universal –GDD- motif and the –SDD- metal binding motif found in all positive strand RdRps, are not found in the primases. However, the primases possess two DxD type metal binding motifs which could possibly bind the Mg^{2+} at the catalytic site (highlighted in light green). Instead of a common YG template-binding pair, a YA pair is found in primases (highlighted in yellow). In MERS-CoV and HCoV-NL63 primases, a G follows the YA pair whereas in other SARS and SARS-related CoVs an S follows as –YAS- pair, similar to-YGS-, as found in SSU DdRps [39]. In fact, the S in the –YGS-, as found in SSU RNA polymerases, is shown to be essential for nucleotide discrimination [21]. It is interesting to note, that a single mutation modifying the **S** in YG**S** to **A** by an SDM experiment, the RNA polymerase behaved like a DNA polymerase accepting dNTPs, suggesting that the Ser–OH might involve in nucleotide discrimination by binding to the 2'-OH of the incoming NTPs [39,40].

The amino acid K in the invariant KL (marked in red) is predicted to play the catalytic role in the polymerization process. It is interesting to note that the invariant KL and YG pairs are common catalytic diads in eukaryotic DNA polymerases as well [39].

It is interesting to note that Imbert et al. [38] have shown that the primase uniquely encodes a Mn^{2+} dependent, rifampicin insensitive, second RdRp in SARS-CoVs but with relatively low fidelity that closely agrees with the MSA findings (in fact, it lacks an NTP binding R at -5 from the catalytic pair KL, which is invariably found in the high fidelity RdRps) (Figs. 4 and 5). Both HCoV-NL63 and MERS-CoV primases exhibit the same template-binding YA pair. In primases also many amino acids are replaced by N in HCoV-NL63 (marked in red). The –LQ diad appearing at the

very end is the main protease cleavage site during the polyprotein processing and is completely conserved in all suggesting that the M^{pro} specificity is not altered in both the groups.

3.6 MSA of the NSP7 of SARS and SARS-related CoVs and HCoV-NL63

As discussed elsewhere the NSP7 and the primase (NSP8) make the final primase complex and make the primer for the RdRp. Fig. 5 shows the MSA analysis of the NSP7 subunit of RTC (MERS-CoV sequence is also included in the analysis). The N-terminal regions of HCoV-NL63 and MERS-CoV are more conserved (highlighted in yellow) among the conserved blocks but the C-terminal region though conserved in all including the HCoV-NL63, but MERS-CoV deviated drastically from others. As there are no template-binding YG or YA pairs as seen in the NiRAN domains, this subunit may play only a supportive role in the replication process by the RdRps and Primases. Its purely supportive role has been further confirmed by mutational analysis by Subissi et al. [36]. For example, they have shown that several mutations in the NSP7 subunit did not affect the viral replication to any significant level. The close identity of the NSP7 in MERS-CoV and HCoV-NL63 and the template-binding and catalytic pairs in NSP8 (Fig. 4) suggest a possible common lineage. In NSP7 also many amino acids are replaced by N in HCoV-NL63 (marked in red). The –LQ diad appearing at the very end is the main protease cleavage site during the polyprotein processing and is completely conserved in all suggesting that the M^{pro} specificity is not altered in both the groups.

The MSA analysis suggests a close relationship between the MERS-CoV and HCoV-NL63 of the RTCs. The HCoV-NL63 infects mainly young children whereas the SARS-CoV-2 infects mainly older people and thus, complementing each other. From the above analyses it is clear that a large number of amino acids in both the spike protein and RTCs are replaced by Ns in HCoV-NL6. It is not clear now whether there is any relationship between the N requiring leukemia and large number of N incorporating HCoV-NL63, because both of them affecting mainly children.

Table 4. Possible Catalytic Regions in NiRAN and RdRp Domains of HCoVs

HCoV type	NiRAN Domain	RdRp Domain
SARS-CoV-1	-^{113}H^{-5}ISRQRLTKYTMADLVYALR-	-R^{-5}QFHQKLLKSIAATRGATVVIGTSKFYGGW^{598}-
MERS-CoV	-^{115}H^{-5}IVRQRLTKYTMADLVYALR-	-R^{-5}QYHQKMLKSMAATRGATCVIGTTKFYGGW^{599}-
SARS-CoV-2	-^{113}H^{-5}ISRQRLTKYTMADLVYALR-	-R^{-5}QFHQKLLKSIAATRGATVVIGTSKFYGGW^{598}-
HCoV-NL63	-^{103}N^{-5}VSRHNLTKYTMMDLVYAMR-	-R^{-5}QYHQKHLKSIVNTRNATVVIGTTKFYGGW^{589}-

CLUSTAL O (1.2.4) MSA of primases (NSP8) of SARS and SARS-related CoVs and HCoVs-NL63

```
HCoV-NL63_Primase           SVASSFVSMPSYIAYENARQAYEDAIANGSS-SQLIKQLKRAMNIAKSEFDHEISVQKKI  59
MERS-CoV_Primase            ATLSEFSHLATFAELEAAQKAYQEAMDSGDTSPQVLKALQKAVNIAKNAYEKDKAVARKL  60
SARS-CoV-2_Primase          AIASEFSSLPSYAAFATAQEAYEQAVANGDS-EVVLKKLKKSLNVAKSEFDRDAAMQRKL  59
SARS-CoV-2(Wuhan_Primase    AIASEFSSLPSYAAFATAQEAYEQAVANGDS-EVVLKKLKKSLNVAKSEFDRDAAMQRKL  59
Bat-RaTG13-CoV_Primase      AIASEFSSLPSYAAFATAQEAYEQAVANGDS-EVVLKKLKKSLNVAKSEFDRDAAMQRKL  59
Pangolin-CoV_Primase        AIASEFSSLPSYAAFATAQEAYEQAVANGDS-EVVLKKLKKSLNVAKSEFDRDAAMQRKL  59
SAR-CoV-1_Primase           AIASEFSSLPSYAAYATAQEAYEQAVANGDS-EVVLKKLKKSLNVAKSEFDRDAAMQRKL  59
Civet-CoV_Primase           AIASEFSSLPSYAAYATAQEAYEQAVANGDS-EAVLKKLKKSLNVAKSEFDRDAAMQRKL  59
                            :  *.*  : ::     *::**::*: .*.:   ::* *::::*:**. :::: :: :*:

HCoV-NL63_Primase           NRMAEQAATQMYKEARSVNRKSKVISAMHSLLFGMLRRLDMSSVETVLNLARDGVVPLSV  119
MERS-CoV_Primase            ERMADQAMTSMYKQARAEDKKAKIVSAMQTMLFGMIKKLDNDVLNGIISNARNGCIPLSV  120
CoV-2_Primase               EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
SARS-CoV-2(Wuhan)_Primase   EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
Bat-RaTG13-CoV_Primase      EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
Pangolin-CoV_Primase        EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
SARS-CoV-1_Primase          EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
Civet-CoV_Primase           EKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNI  119
                            ::**:** *.***:**: ::::*: ***:::** *:::** . :: ::. **:* :**.:
```

Fig. 4. MSA analysis of the primases (NSP8) of SARS-related CoVs and HCoV-NL63

CLUSTAL O (1.2.4) MSA of the NSP7 of SARS and SARS-related CoVs and HCoV-NL63

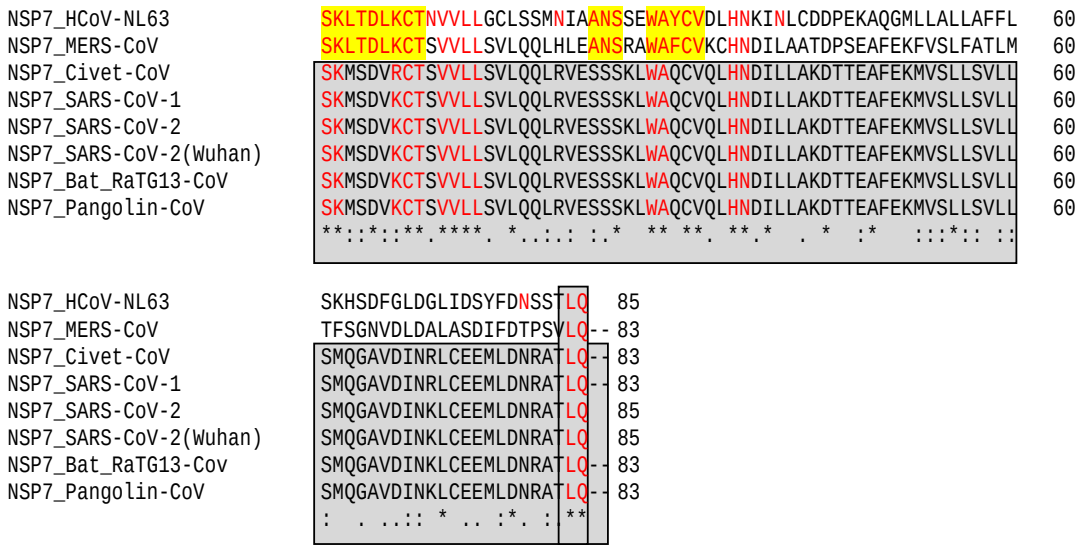


Fig. 5. MSA analysis of the NSP7 of SARS-related CoVs and HCoV-NL63

4. CONCLUSIONS

The MSA analysis of the spike proteins and RTCs of HCoVs, viz. SARS-CoVs and HCoV-NL63 has revealed both similarities and differences between these two groups. It is interesting to note that, like the spike protein of SARS-CoV-2, the HCoV-NL63 has also placed a unique tetrapeptide in its spike protein just close to the S1/S2 cleavage site. The HCoV-NL63 spike protein is abundantly N-glycosylated as compared to other HCoVs. The universal metal binding motif in RdRp domains, viz. –G**D**D- is modified to –K**D**G- in HCoV-NL63 and the second metal binding –SDD- motif and the catalytic and template-binding pairs are completely conserved. The differences in the spike protein, catalytic amino acid of the NiRAN domain and presence of a single accessory protein could possibly have led to mildness of the HCoV-NL63. Recombination events could arise when two different HCoVs infect the same cell. Thus, different regions of the two groups could possibly evolve into better variants by

recombination/mutational events in the future. These findings are important for evaluating emerging disease potentials of HCoV-NL63 and also for preventing and controlling their spread.

DISCLAIMER

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the author and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the author.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Palanivelu P. Analyses of the Spike Proteins of Severe Acute Respiratory Syndrome-Related Coronaviruses. Microbiol Res J Int. 2020;30:32-50.
2. Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9:221-236.
3. Hofmann H, Pyrc K, Hoek LVD, Geier M, Berkhout B, PÖhlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci. (USA). 2005;102:7988–7993.
4. Abdul-Rasool S, Fielding BC. Understanding human coronavirus HCoV-NL63. Open Virol J. 2010;4:76–84.
5. Chiu SS, Chan KH, Chu KW, Kwan SW, Guan Y, Poon LLM, Peiris JSM. Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. Clin Infect Dis. 2005; 40:1721–1729.
6. Lim, Yvonne Xinyi Ng, Yan Ling Tam, James P Liu, Ding Xiang. Human Coronaviruses: A Review of Virus–Host Interactions".Diseases. 2016;4(3):26-36. DOI:10.3390/diseases4030026
7. Cornelissen LA, Wierda CM, Meer FJVD et al Hemagglutinin-esterase, a novel structural protein of torovirus. J Virol. 1997; 71:5277–5286.
8. Fehr AR, Perlman S, Maier HJ, Bickerton E, Britton P (eds.). Coronaviruses: an overview of their replication and pathogenesis. Methods in Molecular Biology. Springer. 2015;1282:1–23. DOI:10.1007/978-1-4939-2438-7.
9. Hoek LVD, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, Wertheim-van Dillen PME, et al. Identification of a new human coronavirus. Nat Med. 2004;10:367-373.
10. Geng H, Cui L, Xie Z, Lu R, Zhao L, Tan W. Characterization and complete genome sequence of human coronavirus NL63 isolated in China. J Virol. 2012;86:9546–9547. Available:https://doi.org/10.1128/JVI.01457-12.
11. Han TH, Chung JY, Kim SW, Hwang ES. Human coronavirus-NL63 infections in Korean children, 2004-2006. J Clin Virol. 2007;38:27–31. Available:https://doi.org/10.1016/j.jcv.2006.10.009.
12. Hoek LVD, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Uberla K.. Croup is associated with the novel coronavirus NL63. PLoS Med. 2005;2:e240. Available:https://doi.org/10.1371/journal.pmed.0020240.
13. Esper F, Shapiro ED, Weibel C, Ferguson D, Landry ML, Kahn JS. Association between a novel human coronavirus and Kawasaki disease. J Infect Dis. 2005; 191:499–502.
14. Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, Li F. Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. J Virol. 2015;89: 9119-9123.
15. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, Filipe ADS, Wojcechowskyj JA, Davis C, et al. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. Bio Rxiv; 2020.
16. Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, Ryan D, et al. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. J Biol Chem. 2004; 279:7996–18007.

17. Pyrc K, Jebbink MF, Berkhout B, Hoek LVD. Genome structure and transcriptional regulation of human coronavirus NL63. Virol J. 2004;1:7.
Available:https://doi.org/10.1186/1743-422X-1-7.
18. Kim JM, Kim ST, Yang JS, Kim SS, Cheong HM. Complete Genome Sequence of Human Coronavirus NL63 CN0601/14, First Isolated in South Korea. genomea.asm.org. 2018;6: e00152-18.
19. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579:7798,265e269.
20. Pyrc K, Bosch BJ, Berkhout B, Jebbink MF, Dijkman R, Rottier P, Hoek LVD. Inhibition of HCoV-NL63 infection at early stages of the replication cycle. Antimicrob Agents.Chemother. 2006;50:2000–2008.
21. Palanivelu P. RNA dependent RNA polymerases of severe acute respiratory syndrome-related coronaviruses- An insight into their active sites and mechanism of action. Int J Biocem Res Rev. 2020;29:29-52.
22. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J. Asano JK, Barber SA, Chan SY, et al. The genome sequence of the SARS-associated coronavirus. Science. 2003; 300:1399–1404.
23. Yuan X, Li J, Shan Y, Yang Z, Zhao Z, Chen B, Yao Z, Dong, B, Wang S, Chen J, *et al*. Subcellular localization and membrane association of SARS-CoV 3a protein. Virus Res. 2005; 109:191–202.
24. Zhao L, Jha BK, Wu A, et al Antagonism of the interferon-induced OASRNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. Cell Host Microbe. 2012;11: 607–616.
25. Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. Curr Opin Struct Biol. 2011; 21:576–82.
DOI: 10.1016/ j.sbi.2011.08.00521.
26. Zhang X, Wang Y. Glycosylation Quality Control by the Golgi Structure. J Mol Biol. 2016;428:3183–93.
DOI: 10.1016/j.jmb.2016.02.030.
27. Watanabe Y, Allen JD, Wrapp D, McLellan JS,' Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020;369:330–333.
28. Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, DiMaio F, Bosch BJ, Veesler D**.** Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Str Mol Biol. 2016;23:899-907.
29. Duan L, Zheng Q, Zhang H, Niu Y, Lou Y, Wang H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. Frontiers in Imuunol. 2020; 11:1-12,
DOI: 10.3389/fimmu.2020.576622.
30. Lin HX, Feng Y, Wong G, Wang L, Li B, Zhao X, Li Y, Smaill F, Zhang C. Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction. Gen Virol. 2008;89:1015-1024.
31. Wu K, Li W, Peng G, Li F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc Natl Acad Sci (USA). 2009; 06:(47):19970e19974.
32. Vickers C, Hales P, Kaushik V, Dick L, Gavin J, Tang J, Godbout K, et al. Hydrolysis of Biological Peptides by Human Angiotensin-converting Enzyme-related Carboxypeptidase. J Biol Chem. 2002; 277:14838–14843.
33. Lehmann KC., Gulyaeva A., Zevenhoven-Dobbe JC, Janssen GMC, Ruben M, Overkleeft HS, van Veelen PA, Samborskiy DV, Kravchenko AA, Leontovich AM. et al. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. Nucleic Acids Res. 2015; 43:8416–8434.
34. te Velthuis AJW, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res. 2012;40:1737-17.
35. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. Structure of replicating SARS-CoV-2 polymerase. bioRxiv preprint; 2020.
doi.org/10.1101/2020.04.27.063180.
36. Subissia L, Posthumab CC, Colleta A, Dobbeb JCZ, Gorbalenyab AE, Decrolya E, et al. One severe acute respiratory

syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc Natl Acad Sci (USA). 2014; E3900–E3909.

37. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020;368:779-782.
38. Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, Ferron F, et al. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J. 2006;25:4933-4942.
39. Palanivelu P. Single Subunit RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism, In: Advances and Trends in Biotechnology and Genetics. Sciencedomain International Book Publishers, UK. 2019;1:1-38.
ISBN: 978-93-89246-59-9.
DOI: 10.9734/bpi/atbg/v1; 2019.
40. Kostyuk SM, Dragan DL, Lyakhov VO, Rechinsky VL, Tunitskaya BK, Chernov SN, Kochetkov E. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. FEBS Letters. 1995; 369:165-168.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/66560