Microbiology Research Journal International

30(8): 32-50, 2020; Article no.MRJI.60411
ISSN: 2456-7043
(Past name: British Microbiology Research Journal, Past ISSN: 2231-0886, NLM ID: 101608140)

Analyses of the Spike Proteins of Severe Acute Respiratory Syndrome-Related Coronaviruses

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, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/MRJI/2020/v30i830249
Editor(s):
(1) Dr. Lachhman Das Singla, Guru Angad Dev Veterinary and Animal Sciences University, India.
(2) Dr. Mehdi Razzaghi-Abyaneh, Pasteur Institute, Iran.
(3) Dr. Grzegorz Cieslar, Medical University of Silesia in Katowice, Poland.
Reviewers:
(1) Asia Raza, Pakistan Homeopathic Medical College, Hospital and Research Center, Pakistan.
(2) Abeer Mohammed Ali Al-garawyi, Muthanna University, Iraq.
(3) Maab AL-Farwachi, University of Mosul, Iraq.
Complete Peer review History: http://www.sdiarticle4.com/review-history/60411

Original Research Article

Received 07 August 2020
Accepted 11 September 2020
Published 24 September 2020

ABSTRACT

Aim: To analyze spike proteins of Severe Acute Respiratory Syndrome (SARS)-related coronaviruses (CoVs) for their conserved motifs, Receptor-Binding Domain (RBD), Receptor Binding Motif (RBM) of SARS-CoV (CoV-1), SARS-CoV-2, Middle East Respiratory Syndrome (MERS)-CoV and their relationship to the bat, pangolin and palm civet-CoVs as possible intermediate hosts.
Study Design: Multiple sequence analysis (MSA) of spike proteins of different SARS-CoVs were studied using Clustal Omega and ExPASy tools.
Methodology: Bioinformatics, SDM and X-ray crystallographic data of the spike proteins from different CoVs including the current epidemic causing SARS-CoV-2 were analyzed. The advanced version of Clustal Omega was used for protein sequence analysis of different spike proteins from various CoVs and ExPASy tool was used for pI analysis.
Results: Spike proteins in coronaviruses play important roles in mediating receptor binding, membrane fusion, and viral entry into human cells. Furthermore, nowadays all the vaccine development programmes are mainly focused on the SARS-CoV-2 spike protein only, as it plays the crucial, first step in the infection process. Therefore, the spike proteins of the SARS-related

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

coronaviruses, the main determinant of coronavirus host specificity, are analyzed for their conserved motifs, RBD, RBM, etc. The recent epidemic causing strain, SARS-CoV-2, showed 2 dipeptide deletions and 4 peptide insertions ranging from tetra- to hepta-peptides in its spike protein as compared to its predecessor CoV-1. Most of the insertions are also found in the bat and pangolin CoVs except one unique tetrapeptide. The RBM region shows that the bats, pangolins and CoV-2 exhibit very similar to identical sequences. The overall analyses show that the latest SARS-CoV-2 is related to bats and more to pangolin-CoVs suggesting that the pangolins could be possibly the intermediate host. On the other hand, it is found that palm civet RBM sequences are highly related to CoV-1 and not CoV-2. Possibly the novel CoV-2 would have taken three insertions from bats and/or pangolins and the fourth insertion –PRRA- which is unique to SARS-CoV-2 is critically placed just in the S1/S2 cleavage region. The recently discovered G614 mutation ($D^{614}\rightarrow G$) in CoV-2, the most prevalent form in the global pandemic now, is found near the RBD towards the C-terminal. Placement of the unique tetrapeptide in the S1/S2 loop region and replacement with more positive charges on the spike protein which resulted in marked increase in the basicity of the SARS-CoV-2 spike protein may, possibly result in significant effects on the structure and function of the protein, possibly leading to rapid transmission.

Conclusions: RBD and RBM regions of the spike proteins of SARS-CoV-1 and palm civet show very close identity to each other whereas the SARS-CoV-2, pangolin- and bat-CoVs exhibit very close identities in their RBD and RBM regions. The two crucial modifications in the spike protein of SARS-CoV-2, viz. a marked increase in the basicity of the protein and the insertion of a dibasic tetrapeptide (–PRRA-) at the critical S1/S2 cleavage point possibly make it to bind to the ACE2 receptor with higher affinity and get it cleaved by the host proteases more efficiently with subsequent effective internalization of the viral genome.

Keywords: Coronaviruses; Severe Acute Respiratory Syndrome; SARS-CoV-1; MERS-CoV; SARS-CoV-2; Bat-CoV; Pangolin-CoVs; Palm civet-CoVs.

1. INTRODUCTION

Viruses are nonliving entities and hence their life depends on other life forms. They are ubiquitous in nature and infect organisms belonging to all kingdoms of life. Varieties of viruses are known to infect the same organism and therefore, they intelligently adapt their life cycle to the organism they infect. As viruses are non-life forms, after their entry into the respective host cells, they invariably make use of the host machineries to multiply in very large numbers. Based on the genomic material, viruses are broadly classified into groups, viz. RNA viruses and DNA viruses. In both the types of viruses, the genomes are organized into at least 4 major types, viz. single-stranded, double-stranded, segmented and non-segmented types. For example, the RNA viruses possess 4 major types of genome structures, e.g., i) Unimolecular, i.e., non-segmented RNA (Single-stranded RNA of either a + strand or a – strand), ii) Segmented RNA, iii) Double-stranded RNA (Reoviruses) and iv) SS Circular RNA (Hepatitis-D virus). Similarly, the DNA viruses also possess at least 4 types of genome structures, e.g., i) unimolecular, i.e., (non-segmented), ii) Segmented iii) Single-stranded linear DNA (parvoviruses) iii) Double-stranded linear DNA (adeno, poxE, herpesE viruses) and iv) DS circular DNA (papovaviruses, hepadnaE, hepatitis-B viruses). DNA viruses cause some well-known diseases like smallpox, chickenpox, herpes, hepatitis-B, etc. In viruses there are both enveloped and non-enveloped types. Interestingly, the viruses are host-specific and hence limited to a particular host range or cell type. In this communication, the positive, single-stranded RNA viruses, with the main focus on the Severe Acute Respiratory Syndrome-Related Coronaviruses (SARS-CoVs) are analyzed which come under Nidoviruses. The general classification of RNA viruses based on their genomes is shown in Fig. 1.

1.1 Nidoviruses

Nidoviruses are enveloped, positive, single-strand RNA viruses. CoVs are the largest group of viruses belonging to this order *Nidovirales*, which includes 4 families, viz. *Coronaviridae, Arteriviridae*, *Mesoniviridae*, and *Roniviridae*. The name "Coronavirus" is derived from the Latin word 'Corona', which means "Crown" or "Wreath". They derive their name because of their characteristic club-shaped spikes on their surface which look like he solar corona under an electron microscope. CoVs were first discovered by [1] in the 1930s,

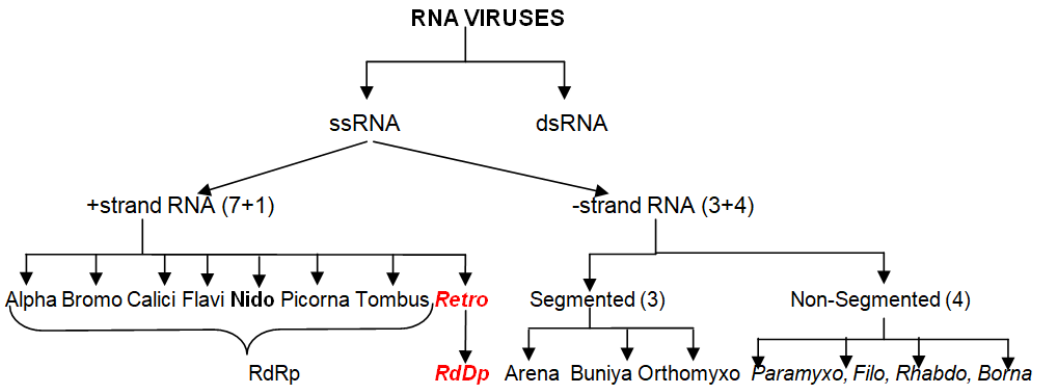


Fig. 1. Broad classification of RNA viruses

when an acute respiratory infection of domesticated chickens was shown to be caused by an infectious bronchitis virus.

Generally, large-scale infections with viruses place a major burden on the global healthcare systems and economy and the nidoviruses are no exception. The nidoviruses have caused already two global epidemics, e.g., the SARS-CoV (CoV-1) in 2002-2003, and the Middle East Respiratory Syndrome-CoV (MERS-CoV) in 2013 and currently causing the most dangerous and highly infectious SARS-CoV-2, also known as Coronavirus Disease 2019 (COVID-19). The World Health Organization has declared COVID-19 as a global epidemic.

Among the above three CoVs, the SARS-CoV-2 spreads so quickly that its transmission is not completely understood yet, e.g., just in six months, more than 20 million people are infected. Moreover, ~ 7,50,000 people have died worldwide making it one of the worst public-health crises of this century. The uncontrolled pandemic nature of this virus catalyzed a research revolution among scientists, doctors, and epidemiologists to work at breakneck speed to understand the spread of COVID-19 for its effective containment and treatment strategies. The important viral philosophy is: mutate, adapt and survive; SARS-CoV-2 is no exception and in the long battle of human history eventually, the viruses win.

1.2 Classification of SARS-CoVs and their Host Ranges

CoVs are large, roughly spherical shaped particles with an average diameter of ~125 nm. They exhibit a wider spectrum of infectivity and infect birds, animals, and humans. Based on their host, genome structure, and antigenic criteria, they are classified into 4 groups as α, β, γ and δ. α and β use bats as a reservoir, whereas γ and δ use birds as a reservoir. Evolutionary analyses have shown that bats and rodents are the gene sources for most α- and β-CoVs, while avian species are the gene sources for most of the γ- and δ-CoVs. It is interesting to note that CoVs have repeatedly crossed species barriers and infected humans too and some have emerged as important human pathogens. Human CoVs were first discovered in the 1960s by Kahn and McIntosh [2]. CoVs primarily cause enzootic infections in birds and mammals but, in the last few decades, they have successfully breached the species barrier many times and became true human pathogens. However, most CoVs are not dangerous, but only a few of them are highly contagious and extremely dangerous as they cause SARSs. For example, out of the 7 family members of CoVs, only 3 (CoV-1, MERS-CoV and CoV-2) caused SARS in humans and also spread in epidemic proportions in recent decades and interestingly, the above 3 belong to β-coronaviral group. The other 4 human CoVs, viz. HCoV-OC43 (β), HCoV-HKU1 (β), HCoV-29E (α), and HCoV-NL63 (α) cause most of the colds throughout the year and are not a serious threat to otherwise healthy people. However, some CoVs are found to be dangerous and caused major global epidemics. The best-known examples include CoV-1 which emerged in the Guangdong province of China in 2002–2003 which caused a large-scale epidemic and then the second one, the MERS-CoV, which caused a persistent epidemic in the Arabian Peninsula during 2012. The current one, viz. the SARS-CoV-2, which was first reported in the Wuhan region of the

Hubei province in China in December 2019, has emerged as one of the most dangerous among them. The current epidemic causing strain is also known as COVID-19 as it was reported in 2019.

1.3 SARS-CoVs and their Mortality Rates

The SARS-CoV-1 outbreak in 2002-2003 infected 8,096 people with 774 deaths, resulting in a mortality rate of 9.6%. This rate was much higher in elderly individuals with mortality rates approaching ~50% in individuals over 60 years of age [3]. Furthermore, the outbreak resulted in the loss of nearly $40 billion dollars in economic activity, while the SARS-CoV-1 epidemic was controlled in 2003 and the virus has not since returned. However, another human CoV emerged in the Middle East in 2012. This virus, named MERS-CoV, was found to be the causative agent in a series of highly pathogenic respiratory tract infections in Saudi Arabia and other countries in the Middle East. As of August 27th, 2014 there were a total of 2494 cases with 912 deaths with a fatality rate of ~40%, according to the European Center for Disease Prevention and Control [4]. The novel SARS-CoV-2, reported on Dec. 31, 2019 in the Wuhan province of China, has become an uncontrolled epidemic and so far infected more than ~20 million people and caused ~750,000 deaths worldwide, and thus, devastating public health and global economy.

1.4 SARS-CoV-2 and Mode of Pathogenesis

Generally, the SARS–CoVs bind to their receptor protein on human cells and enter into them. The current pandemic causing SARS-CoV-2 primarily infects the nose, sinuses, or throat, windpipe and lungs. In the lungs, it primarily infects the epithelial cells and causes pneumonia and respiratory failure. It also causes kidney failure, sometimes multiple organ failure, and septic shocks. The virus is also capable of entering immune cells such as macrophages and dendritic cells but that leads to an abortive infection. However, infection of these cell-types may be important in inducing pro-inflammatory cytokines (also called 'cytokine storm', involving IL-6, IL-10, and TNF-α) that are shown to be contributing to the severity of the disease. (Cytokines are a large group of small molecular weight proteins and peptides (5 kDa to 20 kDa) that are secreted by specific cells of the immune system. They serve as signaling molecules and mediate and regulate immunity, inflammation and hematopoiesis) and it also involves the chemokines (chemokines are also signaling proteins, secreted by cells of the immune system that stimulate the movement of other cells, i.e., migration of cells through venules from blood into tissues and *vice versa* and hence also known as chemotactic cytokines). They are produced by the immune cell types and show elevated levels in the serum of SARS-CoV-2 infected patients [5]. The reproduction number, known as R_0, is the average number of people each person can infect and it differs from virus to virus. (SARS-CoV-2's normal transmission rate, R_0, is between 2.2 and 3.9).

1.5 SARS-CoVs and their Mode of Entry

Worldwide, researchers are wondering whether the widespread prevalence of COVID-19, unlike the other two predecessors, is due to the virus's spike protein, which plays the initial and decisive role in infection by binding to specific human cell receptor. In human cells, it is known to bind to the angiotensin-converting enzyme-2 (ACE2) receptor. (The renin-angiotensin system is a signaling pathway that regulates vascular functions and maintains the blood pressure). The ACE2 is nothing but a carboxypeptidase, an enzyme that involves mainly in the metabolism of angiotensin-I and II. The highest expression of ACE2 is observed in kidneys, lungs, and heart cells. The ACE2 receptors are abundant, especially in the alveolar cells of the lungs, which seem to be main target of the SARS-CoV-2. It is interesting to note that the CoV-1 and CoV-2 use ACE2 receptor similar to bat coronaviruses. However, the MERS-CoV uses a different type of human receptor, i.e., a dipeptidepeptidase-4 (DPP4). Generally, the α-CoVs use an aminopeptidase (APN) receptor, e.g., the less dangerous CoVs which belong to α-CoVs like HCoV-229E use an APN receptor but HCoV-NL63 use the ACE2 receptor for entering into the cells. Table 1 shows the different CoVs and their human receptors. Therefore, this communication mainly focuses on the analyses of the spike proteins of different SARS-related CoVs.

It is interesting to note from Table-1, that the CoVs use invariably a peptidase enzyme for initial binding and entry into the human cells. As there are hundreds of different receptors on human cells why CoVs choose invariably a peptidase is intriguing. They might be choosing their peptidase receptors after random collision with so many other types of receptors on the human cells. Therefore, obviously there should

be a specific interaction between the spike protein and peptidase receptors before enabling entry of the CoVs into the human cells.

Table 1. Human CoVs and their receptors

Human CoVs	Receptor	Reference
Human α-coronaviruses		
HCoV-229E	APN	[6]
HCoV-NL63	ACE2	[7]
Human β-coronaviruses		
SARS-CoV-1	ACE2	[8]
MERS-CoV	DPP4	[9]
SARS-CoV-2	ACE2	[10]

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; DPP4, Dipeptidyl Peptidase 4. Adapted from [11]

1.6 Coronaviral Entry, Uncoating and Translation of the Viral RNA

Different CoVs use different type of receptors for binding and entry into human cells (Table 1). After binding to their specific receptor, further access to the cytosol is generally accomplished by the proteolytic cleavages of the spike protein by a subtilisin-like serine protease, furin (EC 3.4.21.75) and a trypsin-like transmembrane serine protease 2 (TMPRSS2) (EC 3.4.21.109), followed by fusion of the viral and cellular membranes. (The furin which acts at S1/S2 site, belongs to the proprotein convertase family which cleaves precursor proteins and facilitates their conversion to a biologically active state and TMPRSS2 is a plasma membrane-anchored protease that activates several substrates in the cells, including the SARS-CoV-2 spike protein at S2' site). The proteolytic cleavages of the spike protein promote viral uptake. The uptake process follows the regular endocytic pathway of the host cells and using this uptake pathway, the virus moves into the cytosol. In the cytosol, it uncoats and releases the viral RNA which is then translated using the host machinery. The host ribosomes translate the viral RNA into a polyprotein, which is then processed into individual proteins/enzymes by autocatalytic domains in the translated polyprotein, and the individual proteins now assume their functions. Some are structural proteins (SPs) and most of them are non-structural proteins (NSPs) that possess enzymatic activities.

2. SALIENT FEATURES OF THE SPIKE PROTEIN OF SARS-COV-2

The spike protein assumes greater importance among the structural proteins because all the vaccine development programmes mainly focus on the spike protein only, as it mediates receptor binding, membrane fusion, and viral entry. The spike protein is a trimeric class I fusion protein, composed of two functional domains responsible for receptor binding (S1 domain) and membrane fusion (S2 domain).

2.1 Organization and Activities of Spike Proteins

The organization of the genome of CoVs is as follows: 5′Cap-5'-UTR→ORFs 1a and 1b→S (Spike protein)→E (Envelope protein)→M (Membrane protein)→N (Nuclear capsid protein) →3′-UTR-poly-(A) tail. The accessory protein genes are interspersed within the structural protein genes at the 3′ end of the genome. The 1a and 1b codes for polyproteins, which are cleaved into 16 non-structural proteins (NSPs) by autocatalytic proteases, embedded on the polyprotein itself. The NSPs possess various enzymatic activities to transcribe, and replicate the genome. The major structural proteins are S, E, M and N. (All CoVs exhibit a characteristic, strictly conserved genome organization with the essential genes occurring in the following order: 5′-Polyproteins-S-E-M-N-3′).

The spike protein is a heavily glycosylated protein (exhibit N-linked glycosylations) and composed of two structural domains, viz. S1 and S2. The spike protein can be dissected into 7 different domains, viz., the N-terminal domain (NTD), an RBD, a fusion peptide (FP), a heptad repeat (HR), a central helix (CH), a connector domain (CD), and a transmembrane domain (TM). Fig. 2 illustrates the various domains of the spike protein of CoV-2 [12]. The S1 domain contains the signal peptide (Sp) followed by the N-terminal domain (NTD) and RBD. A loop region covering, residues ~440 to 510, termed RBM, is the only domain that is known to make the contact with ACE2 directly. The RBM seems to be unique to SARS-CoVs. A recombination event that occurred in bats, pangolins, palm civets, or another host, perhaps with a group 1 virus similar to HCoV-NL63, may have given rise to SARS-CoV-2. Consistent with a role for palm civets in transmitting the virus, palm civet ACE2 supported SARS-CoV-1 infection as efficiently as human ACE2. These results correlated with the

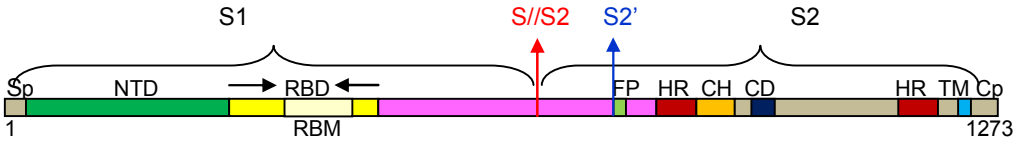


Fig. 2. A schematic diagramme showing various domains of the spike protein of SARS-CoV-2
The S1 and S2 domains are dividing the spike around half-way, i. e., residues between R685 and S686 in CoV-2. S1, Surface domain (Receptor binding); S2, Transmembrane domain (Fusion); Sp, Signal peptide; NTD, N-terminal domain; RBD, Receptor binding domain, RBM, Receptor binding motif; FP, Fusion peptide; HR, Heptad repeat; CH, Central helix; CD, Connector domain; TM, Transmembrane domain; Cp, Cytoplasmic domain

affinity of each of these receptors for the spike protein and its RBD [8]. Spike protein cleavage occurs at two sites, one at S1/S2 junction, and the second one at S2'. The first cleavage separates the RBD and fusion domains of the spike protein whereas the second one exposes the fusion peptide for membrane fusion (Fig. 2). The first proteolytic cleavage is performed by the proprotein convertase, furin, which unmasks the fusion peptide S2. In CoV-2, the S2 is further cleaved by another plasma membrane-anchored transmembrane protease, TMPRSS2, at the so-called S2′ site, located immediately upstream of the fusion peptide [13] (Fig. 2), which results in the activation of membrane fusion within endosomes (Endosomes are small, dynamic vesicles that sort and deliver signaling molecules to the correct location in cells). The S2' domain is known to mediate fusion of the virion and cellular membrane by acting as a class I viral fusion protein.

The spike protein's S1 domain binds to human angiotensin-converting enzyme 2 (ACE2), initiating the infection. Li et al. [8] has successfully isolated ACE2 receptor from (SARS-CoV)-permissive Vero E6 cells that efficiently bind the S1 domain of the SARS-CoV-1 spike protein and identified ACE2 as a metallopeptidase. This finding was further confirmed when the anti-ACE2 but not anti-ACE1 antibody blocked viral replication on Vero E6 cells, suggesting that ACE2 is the functional receptor for SARS-CoVs. Although many cell lines do not express the ACE2 receptor, it is expressed in the lung and in the gastrointestinal tract, the major sites of replication of the virus.

As the latest epidemic, caused by the SARS-CoV-2 is found to be much more serious global problem, affecting public health and devastating the economy in an unprecedented level, this virus is analyzed especially to understand for its rapid transmission and multiplication. As the spike protein of the virus is the first one involved in recognizing and binding to human cells, the objective of the study is, therefore, to analyze for its conserved motifs, RBD, RBM and compare it with its predecessors, viz. SARS-CoV-1, MERS-CoV and also for its relationship to the bat, pangolin and palm civet CoVs, as possible intermediate hosts.

3. MATERIALS AND METHODS

For multiple sequence alignment (MSA) analysis of various spike proteins of CoVs, the protein sequences were retrieved from SWISS-PROT and PUBMED sites. The protein sequences were analyzed using Clustal Omega, an accurate, fast and widely accepted algorithm, available on the website. For pI calculations of the spike proteins of SARS-CoVs which caused the pandemics, and other SARS-Related CoVs, ExPASy tools were used.

4. RESULTS AND DISCUSSION

Unlike the SARS-CoV-1, the SARS-CoV-2 is found to be much more lethal and highly transmissible in humans, even though both were claimed to be originated from bats. This raises the question as to what makes the CoV-2 more dangerous and enables it for rapid transmissions in humans as compared to CoV-1 or MERS-CoV. In the viral infection, the first structure that encounters with the human cell receptor is the spike protein, which is found outermost on the viral envelope. Therefore, in order to find out whether there is any difference(s) exists between the CoV-1 and CoV-2 spike proteins, the spike protein sequences were analyzed by the MSA tool, Clustal Omega.

Fig. 3 shows the MSA results of the two spike proteins of the epidemic strains CoV-1 and CoV-2. At least 6 differences between the two spike proteins were observed, i.e., two dipeptide deletions (highlighted in light red) in the N-terminal region (both starting with a Thr, as TF

and TQ) and four insertions ranging from tetra- to hexa-peptides (highlighted in yellow) which were found only in SARS-CoV-2 spike protein. Interestingly, all the insertions were in-frame (Fig. 3). The N-terminal region which is known to bind to the ACE2 receptor is less conserved (but with a conspicuous insertion of a tetrapeptide) than the C-terminal region, i.e., the tetrapeptide insertion is found in the RBD region of S1 in CoV-2. The large scale insertions would not have happened due to RNA dependent RNA polymerase (RdRp) as the replicase enzymes are very faithful and rarely makes any mistakes (makes ~1 in 10000 nucleotide additions). Furthermore, unlike other viruses, the CoVs have a proofreading mechanism too. So it suggests a possibility of an RNA recombination and editing mechanisms that may exist in CoVs. Therefore, in order to find out whether such deletions and insertions are found in the MERS-CoV too or originated from it, the spike protein sequences from MERS-CoV with the other two CoVs were subjected to MSA analysis.

CLUSTAL O (1.2.4) MSA of spike proteins of the SARS-CoV-1 and SARS-CoV-2

```
P0DTC2|SPIKE_SARS-2        MFVFLVLLPLVSSQCVNLTT--RTQLPPAY--TNSFTRGVYYPDKVFRSSVLHSTQDLFL   56
P59594|SPIKE_SARS-1        MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFL   60
                           **:**::* *.*.. ::  *       * * .*  *******::***..*: ******

P0DTC2|SPIKE_SARS-2        PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQS   116
P59594|SPIKE_SARS-1        PFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQS   113
                           **:**** **:*:       : *.***:**:**:***:*****::***:**:*::.*:**

P0DTC2|SPIKE_SARS-2        LLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFL   176
P59594|SPIKE_SARS-1        VIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAFS   169
                           ::*:**:*****:.*:*::*::**:.*    ..    ::.  ::..* ******:*: *

P0DTC2|SPIKE_SARS2         MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINIT   236
P59594|SPIKE_SARS1         LDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINIT   229
                           :*:. *.****:******** **:: :*. : **::*****.**.:*:*:..**:*****

P0DTC2|SPIKE_SARS-2        RFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPL   296
P59594|SPIKE_SARS-1        NFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPL   283
                           .*:::*: . :     :.. * :.****:****:* **:***:***********: :**

P0DTC2|SPIKE_SARS-2        SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRK   356
P59594|SPIKE_SARS-1        AELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERK   343
                           :* **::*** ::********** *: .:******************:* *****:**

P0DTC2|SPIKE_SARS2         RISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTG   416
P59594|SPIKE_SARS1         KISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTG   403
                           :**************: ********** ********:********::**:**********

P0DTC2|SPIKE_SARS-2        KIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG   476
P59594|SPIKE_SARS-1        VIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPD   463
                           ************ ***:***:.*:*:.  ***** ** :*:.:*:*******.  :. .

P0DTC2|SPIKE_SARS-2        STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKN   536
P59594|SPIKE_SARS-1        GKPCTP-PALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKN   522
                           ..**.   .:***:**:.***  *.*:***************:********* **:*:**

P0DTC2|SPIKE_SARS-2        KCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS   596
P59594|SPIKE_SARS-1        QCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVS   582
                           :**************** *.*:* ********::* **:****:* *****:********

P0DTC2|SPIKE_SARS-2        VITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHV   656
P59594|SPIKE_SARS-1        VITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHV   642
                           *******:*.:***********:* .*********:**:****.*****:**********

P0DTC2|SPIKE_SARS-2        NNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPT   716
P59594|SPIKE_SARS-1        DTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPT   698
                           :.***************:* :     **.:.:**:********:.*:*****:*****

P0DTC2|SPIKE_SARS-2        NFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK   776
P59594|SPIKE_SARS-1        NFSISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDR   758
                           **:**:***::****:******.**********:*****************:***.***:
```

```
P0DTC2|SPIKE_SARS-2        NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQ   836
P59594|SPIKE_SARS-1        NTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQ   818
                           **:********:**** :* ************* **:*********************:**

P0DTC2|SPIKE_SARS-2        YGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQI   896
P59594|SPIKE_SARS-1        YGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQI   878
                           **:***** ********************:*** **:**::** *:************

P0DTC2|SPIKE_SARS-2        PFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNA   956
P59594|SPIKE_SARS-1        PFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNA   938
                           ************************ ******.**.:**:**::*::************

P0DTC2|SPIKE_SARS-2        QALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA  1016
P59594|SPIKE_SARS-1        QALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA   998
                           ************************************************************

P0DTC2|SPIKE_SARS-2        EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFT  1076
P59594|SPIKE_SARS-1        EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFT  1058
                           ****************************************:**************:**:***

P0DTC2|SPIKE_SARS-2        TAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNT  1136
P59594|SPIKE_SARS-1        TAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNT  1118
                           *******:***:******** *** **:*****:.*********************:***

P0DTC2|SPIKE_SARS-2        VYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNES  1196
P59594|SPIKE_SARS-1        VYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNES  1178
                           ************************************************************

P0DTC2|SPIKE_SARS-2        LIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKF  1256
P59594|SPIKE_SARS-1        LIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKF  1238
                           ******************:****************:*************.*********

P0DTC2|SPIKE_SARS-2        DEDDSEPVLKGVKLHYT 1273
P59594|SPIKE_SARS-1        DEDDSEPVLKGVKLHYT 1255
                           *****************
```

Fig. 3. MSA of SARS-CoV-1 and SARS-CoV-2

(The deletions and insertions are highlighted in red and yellow, respectively. P0DTC2, CoV-2; P59594, CoV-1)

As found in Fig. 3, the spike proteins of the two SARS-CoVs differ in a few aspects. Therefore, in order to find out whether such deletions and insertions are also found or originated from MERS-CoV, all three spike protein sequences were subjected to MSA analysis. Fig. 4 shows the MSA analysis of all three spike proteins from the pandemic strains. The two dipeptide deletions and the four insertions (highlighted in yellow) observed in CoV-2 appears to be unique and it is not found in the other two strains (Fig. 4). Obviously, both the N- and C-terminal regions are not conserved in all three (but for a few small stretches of amino acids) because MERS-CoV and SARS-CoVs recognize two different types of receptors (Table 1).

CLUSTAL O (1.2.4) MSA of SARS-CoV-1, SARS-CoV-2, and MERS-CoV

```
K9N5Q8|SPIKE_MERS-CoV      MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRP-IDVSKADGIIYP    59
P0DTC2|SPIKE_SARS-2        ----MFVFLVLLPLV-S----------SQCVN--LTT--RTQLPPAY--TNSFTRGVYYP    39
P59594|SPIKE_SARS-1        ----MFIFLLFLTLT-S----------GSDLD--RCTTFDDVQAPNYTQHTSSMRGVYYP    43
                               :*::::.:*  . *          .  ::        *    *   *: **

K9N5Q8|SPIKE_MERS-CoV      QGRTYSNITITYQGLF-PYQGDHGDMYVYSAGHATGTTPQKLFVANYSQDVKQFANGFVV    118
P0DTC2|SPIKE_SARS-2        DKVFRSSVLHSTQDLFLPFFSNVT---WFHAIHVSGTNGTKR----FDNPVLPFNDGVYF    92
P59594|SPIKE_SARS-1        DEIFRSDTLYLTQDLFLPFYSNVT---GFHTIN-------HT----FGNPVIPFKDGIYF    89
                           :    *.     *.** *: .:      : : :       :     :.: *  * :*. .

K9N5Q8|SPIKE_MERS-CoV      RIGAAANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGT    178
P0DTC2|SPIKE_SARS-2        ASTE-----------------KSNIIRGWIFGTTLDSKTQ---------SLLIVNNATNV    126
P59594|SPIKE_SARS-1        AATE-----------------KSNVVRGWVFGSTMNNKSQ---------SVIIINNSTNV    123
                                               ::  .:::*:::.. ::         ::::: :. ..

K9N5Q8|SPIKE_MERS-CoV      LLRAF--YCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRN    236
P0DTC2|SPIKE_SARS-2        VIKVCEFQFCNDPFLGVYYHKNN--------------------KSWMESEFRVYSSANN    165
P59594|SPIKE_SARS-1        VIRACNFELCDNPFFAVSKPMGT----------------------QTHTMIFDNAFN    158
                           :::.       :*  .     ..                        .    : .  *
```

```
K9N5Q8|SPIKE_MERS-CoV      CTFMYTYNITEDEILEWFGITQTAQG-VHLFSSRYVDLYGGN---------------MFQ   280
P0DTC2|SPIKE_SARS-2        CTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEP   225
P59594|SPIKE_SARS-1        CTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKP   218
                           *** *  :    ::    *  :  :  *.   . :. :*                 :

K9N5Q8|SPIKE_MERS-CoV      FATLPVYDTIKYYSIIPHSIR---SIQSDRKAW----AAFYVYKLQPLTFLLDFSVDGYI   333
P0DTC2|SPIKE_SARS-2        LVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTI   285
P59594|SPIKE_SARS-1        IFKLPLGINITNFRAILTAFSPAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTI         272
                           :  **:  .*. :  :             . *    **::*  *:* **:*.:. :* *

K9N5Q8|SPIKE_MERS-CoV      RRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEG-VECDFSPLLSG-   391
P0DTC2|SPIKE_SARS-2        TDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNAT   345
P59594|SPIKE_SARS-1        TDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNAT   332
                             *:**. : *:: :*: :** ::.*:*..*.*.. *: .:*.  :    * *. ::..

K9N5Q8|SPIKE_MERS-CoV      TPPQVYNFKRLVFTNCNYNLTKLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSYPLS   451
P0DTC2|SPIKE_SARS-2        RFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD   405
P59594|SPIKE_SARS-1        KFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGD   392
                              .** ::*  ::**  : : * .   .. *.*  :* : : . *::.:  * *    .

K9N5Q8|SPIKE_MERS-CoV      MKSDLSVSSAGPISQFNYKQSFSNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDR   511
P0DTC2|SPIKE_SARS-2        EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFE   465
P59594|SPIKE_SARS-1        DVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFE   452
                              ::: ..:* *:::***   .   *::  .  :  :.     :* *       *   .

K9N5Q8|SPIKE_MERS-CoV      TEVPQLVNANQYSPCVSIVPST-VWEDGDYYRKQLSPLEGGGWLVASGSTVAMTEQLQMG   570
P0DTC2|SPIKE_SARS-2        RDISTEIYQAGSTPCNGVEGFNCYF-----------PLQSYGFQPTNGVGYQPYRVVVLS   514
P59594|SPIKE_SARS          RDISNVPFSPDGKPCTP-PALNCYW-----------PLNDYGFYTTTGIGYQPYRVVVLS   500
                            ::         .**      .  :           **:. *:  :.*      . : :.

K9N5Q8|SPIKE_MERS-CoV      FGITVQYGTDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCTAVGVRQQRF   630
P0DTC2|SPIKE_SARS-2        FELL----HAPATVCGP-----KKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQF   565
P59594|SPIKE_SARS-1        FELL----NAPATVCGP-----KLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQF   551
                           * :         :**      . .     .:**::.: *::* **:  ..      *:*

K9N5Q8|SPIKE_MERS-CoV      VYDAYQNLVGYYSD--DGNYYCLRACVSVPVSVIYD--KETKTHATLFGSVACEHISSTM   686
P0DTC2|SPIKE_SARS-2        GRDIADTT-DAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAI   624
P59594|SPIKE_SARS-1        GRDVSDFT-DSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAI   610
                             *  :   .   *    :   :  *    ****    : :.  *.*: .* * .:  ::

K9N5Q8|SPIKE_MERS-CoV      SQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPSTLT   746
P0DTC2|SPIKE_SARS-2        HADQLT--PTWRVYSTGSNVFQTRAGCLIGAEHVN-NSYECDIPIGAGICASYQTQT-NS   680
P59594|SPIKE_SARS-1        HADQLT--PAWRIYSTGNNVFQTQAGCLIGAEHVD-TSYECDIPIGAGICASYHTVS-L-   665
                              . :     :  .:  . :** .**::*  : .    :*.:*:* .:**  .* :

K9N5Q8|SPIKE_MERS-CoV      PRSVRSVPGEMRLASIAFNHPIQV-DQLNSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVD   805
P0DTC2|SPIKE_SARS-2        PRRARSVASQSI---IAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVD   737
P59594|SPIKE_SARS-1        ---LRSTSQKSI---VAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVD   719
                              **.  :     :*:.  : . ..:  *  .::*****:.:* * : .:: *.:**

K9N5Q8|SPIKE_MERS-CoV      CKQYVCNGFQKCEQLLREYGQFCSKINQALHGANLRQDDSVRNLFASVKSSQSSPIIPGF   865
P0DTC2|SPIKE_SARS-2        CTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDF   797
P59594|SPIKE_SARS-1        CNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYF   779
                           *. *:*..  :* :** :**.**::*:** *   .** ..:::**.**.  .:* :  *

K9N5Q8|SPIKE_MERS-CoV      GGDFNLTLLEPVSISTGSRSARSAIEDLLFDKVTIADPGYMQGYDDCMQQGPASARDLIC   925
P0DTC2|SPIKE_SARS-2        GGF-NFSQILPDP---SKPSKRSFIEDLLFNKVTLADAGFIKQYGDCL--GDIAARDLIC   851
P59594|SPIKE_SARS-1        GGF-NFSQILPDP---LKPTKRSFIEDLLFNKVTLADAGFMKQYGECL--GDINARDLIC   833
                           **  *:: : *      . : ** ******:***:** *::: *.:*:  *   ******

K9N5Q8|SPIKE_MERS-CoV      AQYVAGYKVLPPLMDVNMEAAYTSSLLGSIAGVGWTAGLSSFAAIPFAQSIFYRLNGVGI   985
P0DTC2|SPIKE_SARS-2        AQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGV   911
P59594|SPIKE_SARS-1        AQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGV   893
                           ** . * .*****:  :* * **::*:..    *** * .:   **** .: **:**:*:

K9N5Q8|SPIKE_MERS-CoV      TQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNAQALSKLASELSNTFG   1045
P0DTC2|SPIKE_SARS-2        TQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFG   971
P59594|SPIKE_SARS-1        TQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFG   953
                           **:** **** ***:**.*:. :* .:::*  *: *:**.**:*****.*..:**..**
```

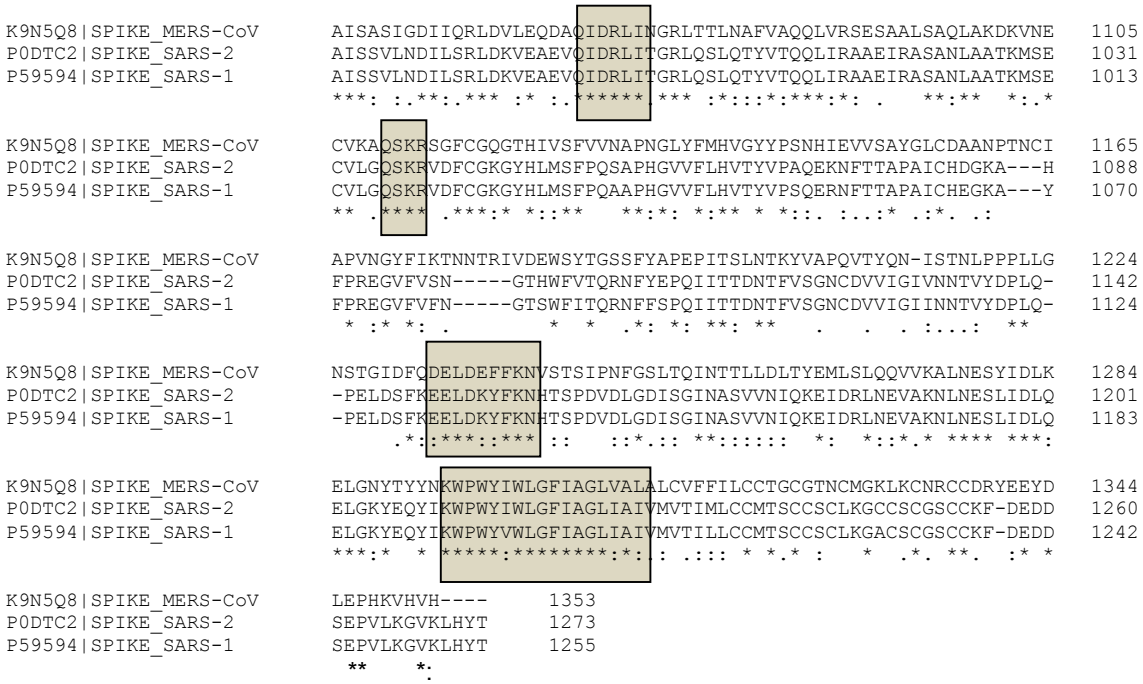


Fig. 4. MSA of SARS-CoV-1, SARS-CoV-2, and MERS-CoV

(The deletions and insertions are highlighted in red and yellow, respectively. P59594, SARS-CoV-1; P0DTC2, SARS-COV-2)

Researchers believe that pangolins when possibly got infected from bats, recombination events would have occurred *in vivo* that possibly lead to the emergence of SARS-CoV-2. In fact, pangolins are the only mammals, in addition to bats, that have been documented to be infected by a SARS-CoV-2-related coronavirus. Molecular and evolutionary analyses have shown that the pangolin-CoV is genetically associated with the SARS-CoV-2 but may not be its likely precursor [14]. However, Lam et al. [15] discovered that multiple lineages of pangolin-CoVs and their similarity to SARS-CoV-2 suggesting that pangolins should be considered as possible hosts in the emergence of new coronaviruses. Furthermore, the Malayan pangolin-CoV genomes showed 85.5% to 92.4% sequence similarity to SARS-CoV-2, and represent two sub-lineages of SARS-CoV-2-related viruses in phylogenetic tree analysis [15].

In order to find out whether the insertions and deletions in SARS-CoV-2 are from the well-known reservoir of the β-CoVs, viz., the horseshoe bats or pangolins, all the 4 spike protein sequences were analyzed by MSA tool. Fig. 5 shows the MSA analysis of bats and pangolins along with the SARS-CoVs. In fact, the pangolin and bat-CoVs showed many similarities with the CoV-2 rather than with CoV-1. For example, a typical 4-amino acid deletion at the N-terminal region, the second insertion sequence are very similar in bat, pangolin-CoVs and CoV-2 (highlighted in orange) The bat-CoV and SARS-CoV-2 sequences are very similar in first and third insertions (highlighted in yellow) raising the possibility for an *in vivo* insertion and deletion events could have occurred in pangolins. However, the last block was unique and was present only in SARS-CoV-2 and not in the other three CoVs (highlighted in light blue).

Interestingly, SARS-CoV-2 has replaced four amino acids to a basic one V→K, T→K, N→H and L→K (marked in red in light green boxes) as compared to the SARS-CoV-1. This makes the SARS-CoV-2 spike protein more basic. Recently, Korber et al. [16] have reported a mutation at position 23,403 which has drawn the most attention—in part because it changed the virus's spike protein. This mutation changed the amino acid at position 614 of the spike protein from an aspartic acid (D) to a glycine (G); i.e., D^{614}→G, which is now known as G614 (marked red in the orange block), i.e., an acidic amino acid (D) is replaced by a neutral one (G), from a highly conserved motif (Fig. 5). Furthermore, they have shown that the G614 mutation is common in

almost every nation and region they looked at, whereas D614 is virtually gone. They suggested that the G614 caused a slight change in the shape of the spike protein, apparently making it easier for the protein to undergo the structural changes that could cause the membranes of the virus and the cell to fuse and making it more infectious somewhere between 3 and 10 times. In other words, if add up, the five replacements, has increased markedly the basicity of the CoV-2's spike protein (Table 2) which has possibly increased the high transmission rate.

4.1 Basicity and Infectivity of SARS-CoVs

The pIs of the all 4 SARS-CoVs were analyzed and presented in Table 2. From Table-2, it is clear as the basicity of the spike protein increases, the infection and death rate has increased dramatically. (*MERS-CoV does not use ACE2 receptor but included to show only the frequency of the epidemics) The spike proteins of SARS-CoVs first encounter with the cell is the ACE2 receptor. ACE2 is an integral membrane protein which contains a catalytically active ectodomain that is projected on the cell surfaces. ACE2 is nothing but an enzyme known as carboxypeptidase (EC 3.4.15.1), a Zn metalloprotease. (Carboxypepditases degrade proteins from the C-terminal end, ACE2 functions predominantly as a carboxypeptidase with a substrate preference for hydrolysis between proline and a hydrophobic or a basic C-terminal residue).

The ACE2 enzyme converts Angiotensin I and II to Angiotensins 1-9 and 1-7, respectively. It is interesting to note that the optimum pH for the ACE2 carboxypeptidase is 6.5 [17]. With increasing pI values, the substrate of the enzyme (the CoV-2 spike protein in this case), is nearing not only ACE2's optimum pH but also the other two crucial enzymes for host entry, viz. the furin and TMPRSS2 optimum pHs. This might enable not only tighter binding with higher affinity to the ACE2 receptor but also efficient cleavages at S1/S2 and S2' sites enabling positive cell entry which might possibly lead to a rapid transmission of CoV-2 as compared to CoV-1. The optimum pH values of the furin, which cleaves at S1/S2 cleavage site and TMPRSS2, which cleaves at S2' site are 6.5 and 7.5, respectively [18].

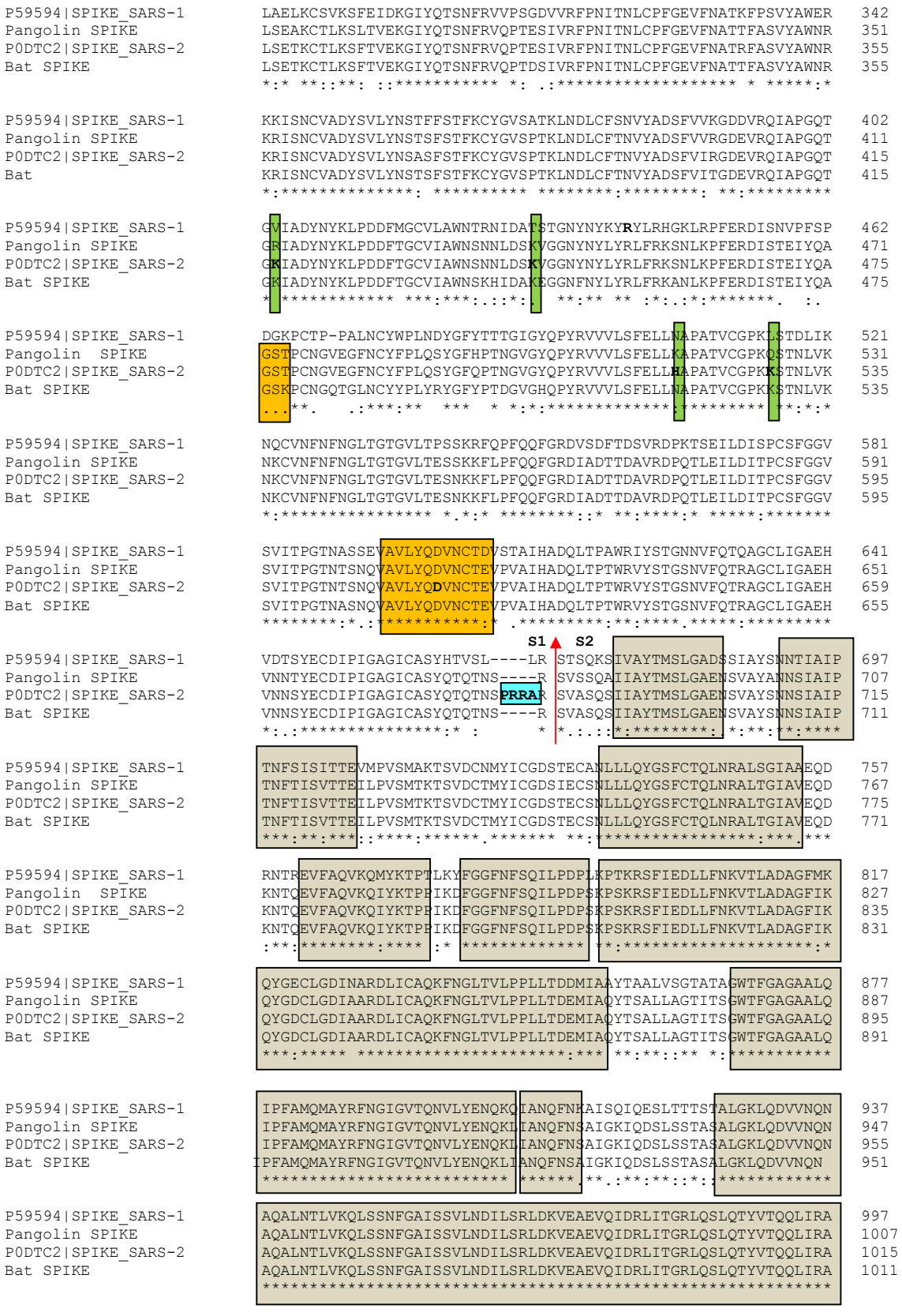
Table 2. pI values and basicty of the spike proteins of CoVs and death rates

Virus	pI Value	Basicity	Year of Epidemic	Cases/Deaths
SARS-CoV-1	5.56		2003	8096/774
MERS-CoV*	5.73		2013	2494/912
SARS-CoV-2	6.24		Dec 2019	~$21x10^6$/~$0.8x10^6$
SARS-CoV-2-G614	6.32		July 2020$	[16]

$ *It is reported that by mid-March, the D614G variant dominated around the world [16]*

CLUSTAL O (1.2.4) MSA of SARS-CoV-1, SARS-CoV-2, bat, and pangolin-CoVs

```
P59594|SPIKE_SARS-1        -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF   59
Pangolin SPIKE             MLFFFFLHFALVNS----QCVNLTGRAAIQPSFTNSSQRGVYYPDTIFRSNTLVLSQGYF   56
P0DTC2|SPIKE_SARS-2        -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF   55
Bat SPIKE                  -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSSTRGVYYPDKVFRSSVLHLTQDLF   55
                            :*.*:: : *...    *..:        :  .*  ******* :***..*  :*. *

P59594|SPIKE_SARS-1        LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ   112
Pangolin  SPIKE            LPFYSNVSWYYALTKTN-SAEKRVDNPVLDFKDGIYFAATEKSNIVRGWIFGTTLDNTSQ   115
P0DTC2|SPIKE_SARS-2        LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ   115
Bat SPIKE                  LPFFSNVTWFHAIHVSGTNGIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ   115
                           ***:***: :::: :    : ..***: *:**:***:*****::***:**:*::..:*

P59594|SPIKE_SARS-1        SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF   168
Pangolin SPIKE             SLLIVNNATNVIIKVCNFQFCYDPYLSGYYHN-NKTWSTREFAVYSSYANCTFEYVSKSF   174
P0DTC2|SPIKE_SARS-2        SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF   175
Bat SPIKE                  SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF   175
                           *::*:**:***:*:.*:*::* :*::.      .    .  ::..  ******:*. *

P59594|SPIKE_SARS-1        SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI   228
Pangolin SPIKE             MLDIAGKSGLFDTLREFVFRNVDGYFKIYSKYTPVNVNSNLPIGFSALEPLVEIPAGINI   234
P0DTC2|SPIKE_SARS-2        LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI   235
Bat SPIKE                  LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPGFSALEPLVDLPIGINI   235
                            :*:  *.* *  ******:* **:: :*. : *:::  :** **.:*:*:..:* ****
P59594|SPIKE_SARS-1        TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP   282
Pangolin SPIKE             TKFRTLLTIHRGDPMP---NNGWTVFSAAYYVGYLAPRTFMLNYNENGTITDAVDCALDP   291
P0DTC2|SPIKE_SARS-2        TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP   295
Bat SPIKE                  TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP   295
                           *.*:::*: .    .. * . :***:**** * **:*:*:**********: :*
```


P59594|SPIKE_SARS-1
Pangolin SPIKE
P0DTC2|SPIKE_SARS-2
Bat SPIKE
S1
S2

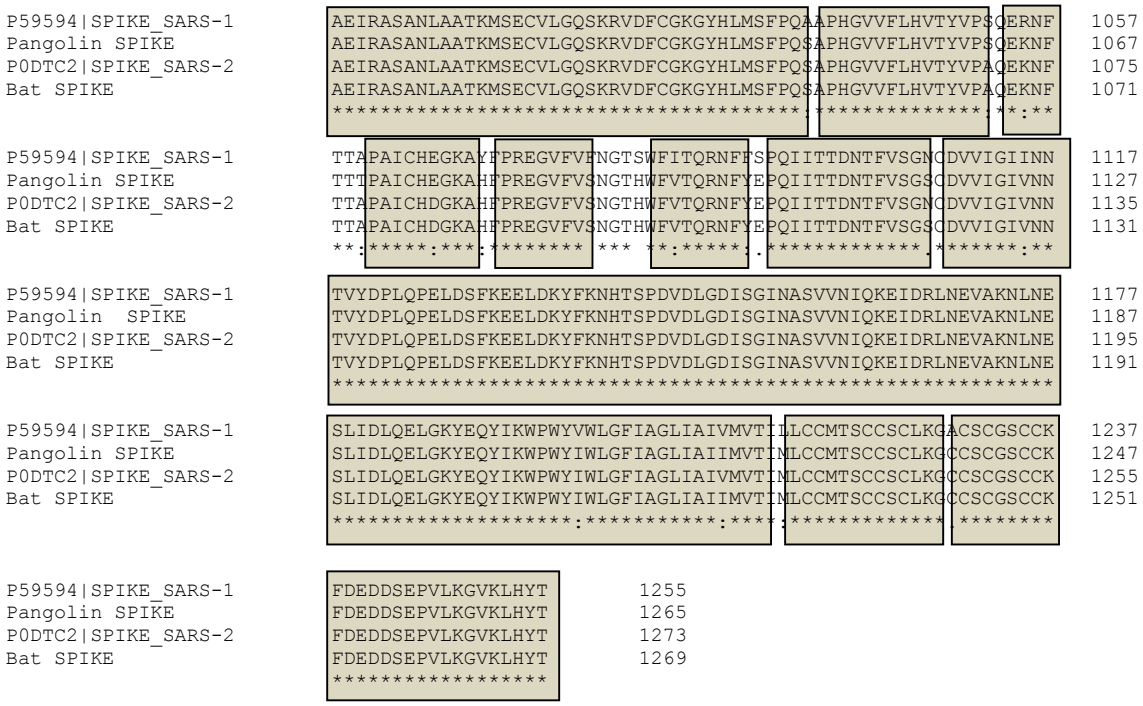


Fig. 5. MSA of SARS-CoV-1, SARS-CoV-2, bat and pangolin CoVs (SARS-CoV-1 RBD, 318-510; RBM, 424-494; S1 and S2 divide the spike protein at 679, marked in red)

(The deletions and insertions are highlighted in red and yellow, respectively. The unique insertion only in SARS-CoV-2 is highlighted in light blue and similarities in all the three or four sequences are shown in orange. Basic amino acid changes are marked in red within light green blocks. P59594, SARS-CoV-1; P0DTC2, SARS-COV-2)

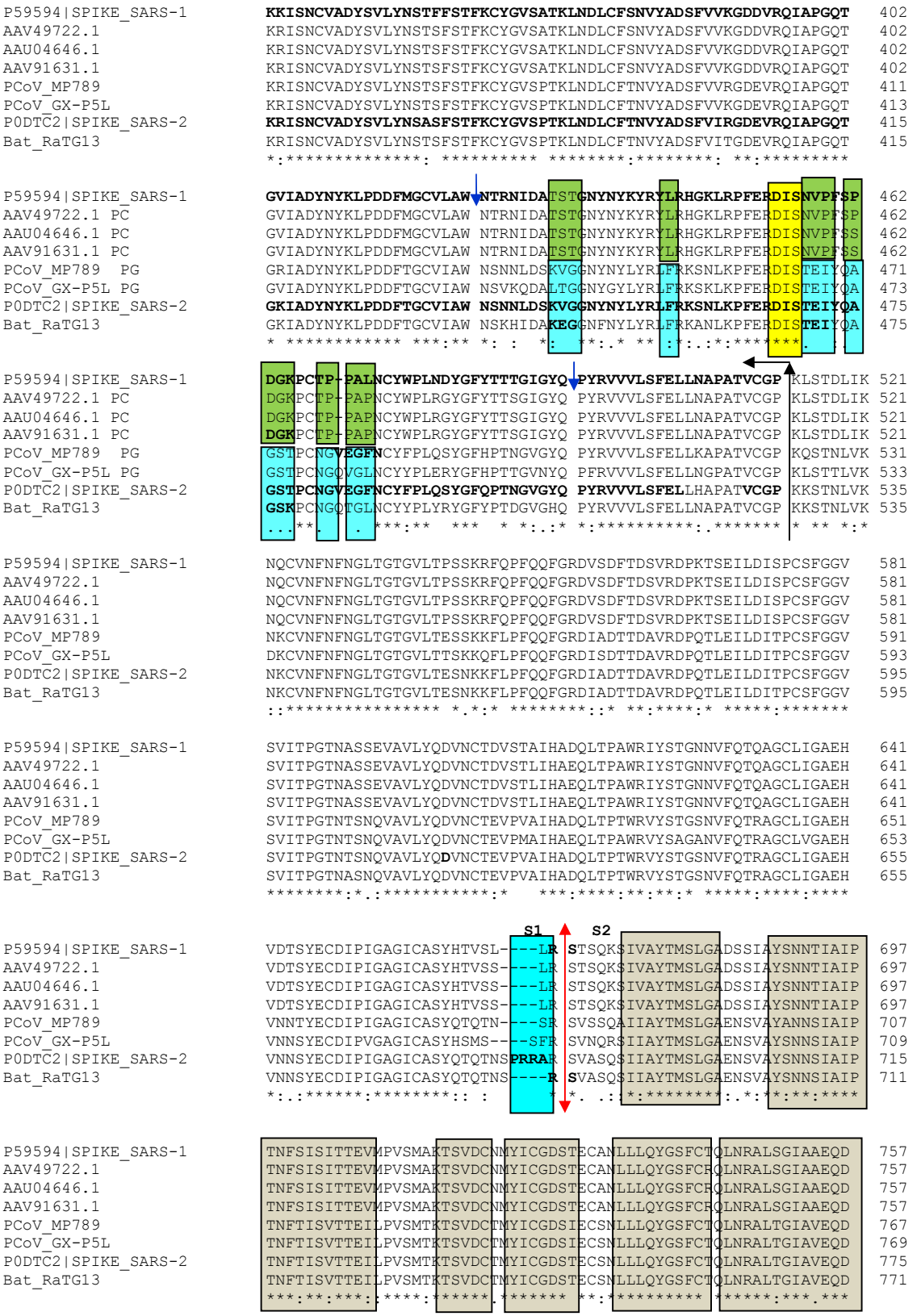
The CoV-1's pI value is only 5.56, which is far away from the optimal pH of ACE2 and other processing proteases and hence could bind with lesser affinity with ACE2, resulting in possibly less number of cases and deaths. From the C-terminal end sequences (-PVLKGVKLHY**T**) of the spike proteins of CoV-1 and CoV-2 and from the substrate specificity of the ACE2, it looks like it is not going to make possibly any cleavage on the C-terminal end of the spike protein. The NTD of the spike proteins is known to bind to the sialic acid residues in the RBD [10]. Therefore, the marked increase in the basicity would enhance favourable binding of the CoV-2 spike protein to the ACE2 receptor leading to efficient cleavages at the S1/S2 and S2' sites- a crucial step in CoV-2 evolution from a virus that infected bats and other species. Interestingly, the pIs of palm civet_PC-199-CoV with a pI, 5.57 and CoV-1 are very close to each other and similarly the pIs of panglonin_GX-5Pl-CoV with a pI, 6.21 and CoV-2 are very close to each other (Table-2).

The search for the intermediate host(s) of CoV-2 is essential to prevent further dissemination of the virus and more species jumps. Pangolins and palm civets are believed to be possible intermediate hosts and therefore, to find out whether the insertions are from pangolins and palm civets, both pangolin and palm civet CoV sequences were included in the analysis along with the other 4. Fig. 6 shows the results of the MSA analysis. The analysis clearly shows that SARS-CoV-1 and palm civets are very close to each other (the similarity regions in the RBM is shown in green blocks) whereas SARS-CoV-2, bats and pangolins are very close to each other (highlighted in yellow blocks) both in deletions and insertions except one unique insertion (the tetrapeptide –^{681}PRRA-), which is found only in SARS-CoV-2, not in the spike proteins of other organisms. This insertion is located at the very end of S1 and at the S1/S2 cleavage point (Fig. 6). It is interesting to note that the unique dibasic tetrapetide insertion starts with a proline. Proline residues (a cyclic, nonpolar amino acid) are known to play a prominent role in protein folding. Proline is often found at the end of α-helix or in turns or loops. It is known that when a proline is found in an α-helix, the helix will have a slight bend due to the lack of the hydrogen bond. This extension in CoV-2 may possibly lead to efficient cleavage by the furin protease at S1/S2 region with subsequent membrane fusion and

internalization of the virus. Hoffmann et al. [19] (2020) reported such S1/S2 cleavage site containing multiple arginine residues (multibasic) is not found in closely related animal coronaviruses and also suggested that such viral variants might exhibit increased cell-cell spread, potentially altering the virulence.

CLUSTAL O (1.2.4) MSA of SARS-CoV-1, SARS-CoV-2, bat, pangolin and civet-CoVs

```
P59594|SPIKE_SARS-1        -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF   59
AAV49722.1                 -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF   59
AAU04646.1                 -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF   59
AAV91631.1                 -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF   59
PCoV_MP789                  MLFFFFLHFALVNS----QCVNLTGRAAIQPSFTNSSQRGVYYPDTIFRSNTLVLSQGYF  56
PCoV_GX-P5L                -MFVFLFVLPLVSS----QCVNLTTRTGIPPGYTNSSTRGVYYPDKVFRSSILHLTQDLF   55
P0DTC2|SPIKE_SARS-2        -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF   55
Bat_RaTG13                 -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSSTRGVYYPDKVFRSSVLHLTQDLF   55
                            :*.*:. : *..:    :*..:           .*  ******* :***. *  :*. *

P59594|SPIKE_SARS-1        LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ   112
AAV49722.1                 LPFYSNVTGFHTIN-------HTFDNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ   112
AAU04646.1                 LPFYSNVTGFHTIN-------HTFDNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ   112
AAV91631.1                 LPFYSNVTGFHTIN-------HTFDNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ   112
PCoV_MP789                 LPFYSNVSWYYALTKTN-SAEKRVDNPVLDFKDGIYFAATEKSNIVRGWIFGTTLDNTSQ   115
PCoV_GX-P5L                LPFFSNVTWFNTINYQG--GFKKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDARTQ   113
P0DTC2|SPIKE_SARS-2        LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ   115
Bat_RaTG13                 LPFFSNVTWFHAIHVSGTNGIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ   115
                           ***:***: : ::       : ..***: *:**:***:*****::***:**:*::  :*

P59594|SPIKE_SARS-1        SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF   168
AAV49722.1                 SVIIINNSTNVVIRACNFELCDNPFFVVSKPMGT----RTHTMIFDNAFNCTFEYISDAF   168
AAU04646.1                 SVIIINNSTNVVIRACNFELCDNPFFVVSKPMGT----QTHTMIFDNAFNCTFEYISDAF   168
AAV91631.1                 SVIIINNSTNVVIRACNFELCDNPFFVVSKPMGT----QTHTMIFDNAFNCTFEYISDAF   168
PCoV_MP789                 SLLIVNNATNVIIKVCNFQFCYDPYLSGYYHN-NKTWSTREFAVYSSYANCTFEYVSKSF   174
PCoV_GX-P5L                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF   173
P0DTC2|SPIKE_SARS-2        SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF   175
Bat_RaTG13                 SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF   175
                           *::*:**:***:*:.*:*::* :*::       .    .  ::..  ******:*. *

P59594|SPIKE_SARS-1        SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI   228
AAV49722.1                 SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI   228
AAU04646.1                 SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIKI   228
AAV91631.1                 SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGIKI   228
PCoV_MP789                 MLDIAGKSGLFDTLREFVFRNVDGYFKIYSKYTPVNVNSNLPIGFSALEPLVEIPAGINI   234
PCoV_GX-P5L                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI   233
P0DTC2|SPIKE_SARS-2        LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI   235
Bat_RaTG13                 LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPGFSALEPLVDLPIGINI   235
                            :*:  *.* *. ******:* **:: :*. : *:::  :** ** :*:*:..:* **:*

P59594|SPIKE_SARS-1        TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP   282
AAV49722.1                 TNFRAILTAFS------PAQDTWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP   282
AAU04646.1                 TNFRAILTAFS------PAQGTWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP   282
AAV91631.1                 TNFRAILTAFS------PAQGTWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP   282
PCoV_MP789                 TKFRTLLTIHRGDPMP---NNGWTVFSAAYYVGYLAPRTFMLNYNENGTITDAVDCALDP   291
PCoV_GX-P5L                TRFQTLLALHRSYLTPGKLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP   293
P0DTC2|SPIKE_SARS-2        TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP   295
Bat_RaTG13                 TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP   295
                           *.*:::*: .        .. * . :***:****   **:*.*::**********: :*

P59594|SPIKE_SARS-1        LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWER  342
AAV49722.1                 LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWER  342
AAU04646.1                 LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWER  342
AAV91631.1                 LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNI|TNLCPFGEVFNATKFPSVYAWER  342
PCoV_MP789                 LSEAKCTLKSLTVEKGIYQTSNFRVQPTESIVRFPNI|TNLCPFGEVFNATTFASVYAWNR  351
PCoV_GX-P5L                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNI|TNLCPFGEVFNASKFASVYAWNR  353
P0DTC2|SPIKE_SARS-2        LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNI|TNLCPFGEVFNATRFASVYAWNR  355
Bat_RaTG13                 LSETKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNI|TNLCPFGEVFNATTFASVYAWNR  355
                           *:* **::**: ::*********** *: .:****** |************: * *****:*
```


S1
S2

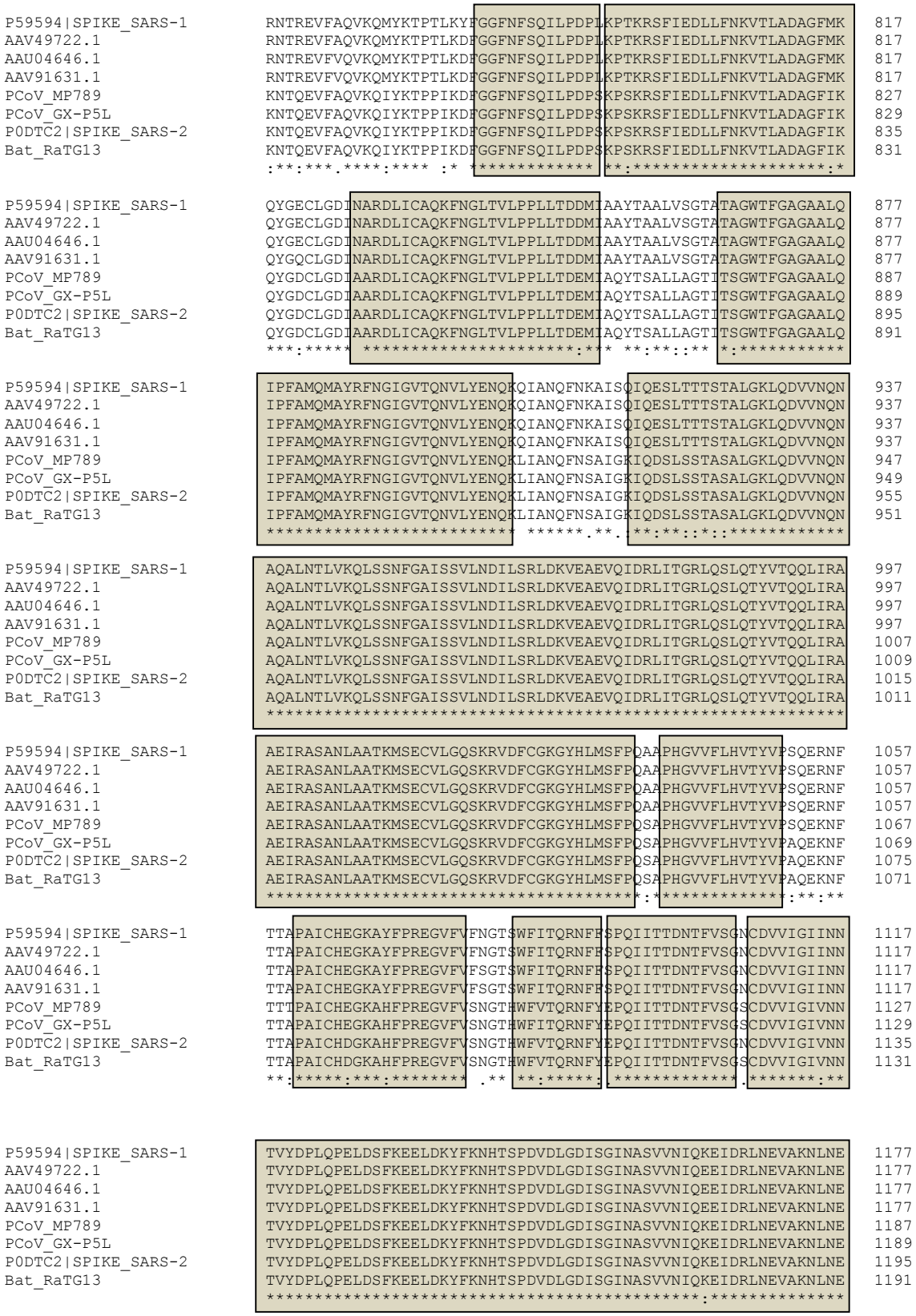

P59594|SPIKE_SARS-1 RNTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMK 817
AAV49722.1 RNTREVFAQVKQMYKTPTLKDFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMK 817
AAU04646.1 RNTREVFVQVKQMYKTPTLKDFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMK 817
AAV91631.1 RNTREVFVQVKQMYKTPTLKDFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMK 817
PCoV_MP789 KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIK 827
PCoV_GX-P5L KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIK 829
P0DTC2|SPIKE_SARS-2 KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIK 835
Bat_RaTG13 KNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIK 831
QYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQ 877
QYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQ 877
QYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQ 877
QYGQCLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQ 877
QYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ 887
QYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ 889
QYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ 895
QYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ 891
IPFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQN 937
IPFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQN 937
IPFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQN 937
IPFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQN 937
IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQN 947
IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQN 949
IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQN 955
IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQN 951
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 997
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 997
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 997
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 997
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 1007
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 1009
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 1015
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRA 1011
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNF 1057
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNF 1057
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNF 1057
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNF 1057
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPSQEKNF 1067
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF 1069
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF 1075
AEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNF 1071
TTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINN 1117
TTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINN 1117
TTAPAICHEGKAYFPREGVFVFSGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINN 1117
TTAPAICHEGKAYFPREGVFVFSGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINN 1117
TTTPAICHEGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGSCDVVIGIVNN 1127
TTAPAICHEGKAHFPREGVFVSNGTHWFITQRNFYEPQIITTDNTFVSGSCDVVIGIVNN 1129
TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGSCDVVIGIVNN 1135
TTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGSCDVVIGIVNN 1131
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE 1177
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQEEIDRLNEVAKNLNE 1177
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQEEIDRLNEVAKNLNE 1177
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQEEIDRLNEVAKNLNE 1177
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE 1187
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE 1189
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE 1195
TVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNE 1191

Fig. 6 MSA of SARS-CoV-1, SARS-CoV-2, bat, pangolin and palm civet CoVs

P59594 SPIKE_SARS-CoV-1; AAV49722.1 Palm civet spike protein; AAU04646.1 Palm civet spike protein AAV91631.1 Palm civet spike protein; PCoV_MP789, Pangolin CoV; PCoV_GX-P5L, Pangolin CoV; P0DTC2 SPIKE_SARS-CoV-2; Bat_RaTG13, Bat Spike protein, RBD regions are marked by black arrows; RBM regions are marked by inverted arrows in blue and the S1/S2 cleavage point marked by a red arrow. (The similarity regions between SARS-CoV-1 and palm civet are highlighted in light green and the similarity regions between pangolins, bat and SARS-CoV-2 are highlighted in yellow. The Zn^{2+} and Mg^{2+} binding motifs are highlighted in orange and dark green, respectively. The unique –PRRA- sequence is highlighted in light blue.

4.2 RBD and RBM Analyses

It is interesting to note that the RBD regions (~333-~527 in CoV-1) are remarkably conserved in all the four spike proteins. A loop region covering, residues from ~437-~506, termed the receptor-binding motif (RBM), is the only domain that is known to make the direct contact with ACE2 (marked by blue arrows) [20]. The RBM seems to be unique to SARS-CoVs. The RBM region is mapped from 424-494 in CoV-1. The two tripeptides 444-446/470—472 and a stretch from 474 to 486 in the RBM region of the RBD is very similar/identical in CoV-2, pangolin and bat, suggesting possible recombination events between bats and pangolins. Interestingly, they were almost identical between CoV-1 and palm civets. Furthermore, Bianchi et al. [21] analyzed the other two surface proteins M and E of CoV-2 with bat and pangolin sequences and found that the CoV-2 M-protein showed remarkable similarity (98% identity) and the CoV-2 and the CoV-2 E-protein sequence was almost identical to the sequences from bat- and pangolin-CoVs. These results narrow down on the possible intermediate hosts of CoV-1 and CoV-2. The C-terminal ends of 26 amino acids are completely conserved in all spike proteins suggesting a possible role in membrane fusion and internalization (Fig. 6). Zhu et al. [22] on analysis of CoV-1 spike protein found that a single amino acid substitution in the RBM (424-494) of the RBD region, viz., R441→A, (marked in red) was able to abolish the immunogenicity of RBD to induce neutralizing antibodies in immunized mice and rabbits. The RBD bearing R441→A mutation could not bind to the soluble and cell-associated ACE2, suggesting that some critical residues in the RBM region of the RBD are important in the critical neutralizing domain. The tripeptides involving this critical residue (-R^{441}YL-) are identical in CoV-1 and palm civet and also identical but as (–R^{433}LF-) in CoV-2, bat, and pangolins RBMs. However, Yi et al. [23] have demonstrated that the mutation R453→A in RBM (marked in red) abolished only viral entry, but retained the capacity for inducing neutralizing antibodies. Lan et al. [20] by X-ray crystallographic analysis found similarity in the structural features between the SARS-CoV-1 and SARS-CoV-2 RBDs, strongly indicating convergent evolution. CoVs are well-known for their ability to recombine both by homologous and nonhomologous recombination mechanisms. The ability of these viruses to recombine is mainly attributed to the strand switching ability of the key enzyme, RdRp [24,25].

A large number of highly conserved Cs (9Cs) is remarkably conserved in the last 26 amino acids stretch in all the 8 SARS-related CoV sequences. A typical, invariant –CxxC- motif, (highlighted in orange) also found in HNH endonucleases suggests a Zn binding motif [26] which may

mainly play a structural role in the spike proteins. Furthermore, the –CC- diad conservations show that the bat, pangolin and CoV-2 are very close to each other. In addition to the Zn binding motif, another metal-binding motif -DxDDxE- (likely a Mg^{2+} binding motif), mostly found HNH endonucleases, and in nucleic acid polymerases is also seen at the C-terminal end (highlighted in dark green) [26,27].

5. CONCLUSIONS

Recombination events play an important role in the evolution of viruses. CoVs are well-known for their ability to recombine both by homologous and nonhomologous recombination mechanisms. The ability of these viruses to recombine is mainly attributed to the strand switching ability of the key enzyme, viz. the viral RdRp [24,25]. The insertions and deletions found in the CoV-2 could possibly be the result of recombination events and RNA editing mechanisms.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. H. Shakila, Professor and Head, Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai for useful suggestions on the manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Estola T. Coronaviruses, a new group of animal RNA viruses. Avian Dis. 1970;14: 330–336.
2. Kahn JS, McIntosh K (November). History and recent advances in coronavirus discovery. Pediatr. Infect Dis J. 2005; 24(11 Suppl):S223–7.
3. WHO; 2004). Summary of probable SARS cases with onset of illness. Available:https://www.who.int/csr/sars/country/table2004_04_21/en/ 1 November 2002 to 31 July 2003.
4. MERS-CoV worldwide overview. European Centre for Disease Prevention and Control. Available:www.ecdc.europa.eu
5. Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, et al. Clinical and immunologic features in severe and moderate Coronavirus Disease 2019. J Clin Invest. 2020;130: 2620–2629. DOI: 10.1172/JCI137244
6. Yeager CL, Ashmun RA, Williams RK et al. Human aminopeptidase N is a receptor for human coronavirus 229E. Nature.1992; 357:420–422.
7. Hofmann H, Pyrc K, van der Hoek L et al. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proc Natl Acad Sci. USA. 2005;102:7988–7993.
8. Li W, MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003;426: 450-454.
9. Raj VS, Mou H, Smits SL, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature. 2013;495: 251–254.
10. Wu F, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579:265–269.
11. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. In Maier HJ, Bickerton E, Britton P, editors. Coronaviruses. Met Mol Biol. 1282. Springer. 2015;1–23. DOI: 10.1007/978-1-4939-2438-7_1
12. Watanabe Y, Joel D, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020; 369, 330–333.
13. Bestle D, Heindl MR, Limburg H, Lam van TV, Pilgram O, Moulton H, Stein DA, et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci Alliance. 2020;3(9);e202000786(1-14). Available:https://doi.org/10.26508/lsa.202000786.
14. Liu P , Jiang JZ , Wan XF, Hua Y, Li L, Zhou J, Wang X, Hou F, Chen J, Zou J, Chen J. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?. PLoS Pathog. 2019;16(5): e1008421.https://doi.org/10.1371/journal.ppat.1008421.
15. Lam TT, Jia N, Zhang YW, Shum MH, Jiang JF, Zhu HC, Tong YG, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020;583: 282-285.

16. Korber B, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer et al.Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. Available:https://doi.org/10.1016/j.cell.2020.06.043 26-6-2020.
17. Vickers C, Paul Hales, Virendar Kaushik, Larry Dick, James Gavin, Jin Tang, Kevin Godbout et al. Hydrolysis of Biological Peptides by Human Angiotensin-converting Enzyme-related Carboxypeptidase. J Biol Chem. 2002;277: 14838–14843.
18. Grober AS, Vey M, Angliker H, Shawl E, Thomas G, Roberts C, Kienk HD, et al. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. EMBO J. 1992;11:2407-2414.
19. Hoffmann M, Kleine-Weber H, PÖhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Molecular Cell. 2020;78: 779–784.
20. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020; 581:215-220.
21. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 Envelope and Membrane Proteins: Structural differences Linked to Virus Characteristics? BioMed Res. Int. 2020;2020: 1-6.
22. Zhu X, Qi Liu, Lanying Du, Lu Lu, Shibo Jiang. Receptor-binding domain as a target for developing SARS vaccines. J Thorac Dis. (Emerging infectious respiratory disease & the 10th Anniversary of SARS Epidemics) . 2013; 5(Supplement 2). DOI:10.3978/j.ISSN.2072-1439.2013.06.06
23. Yi CE, Ba L, Zhang L, et al. Single amino acid substitutions in the severe acute respiratory syndrome coronavirus spike glycoprotein determines viral entry and immunogenicity of a major neutralizing domain. J Virol. 2005;79: 11638-46.
24. Keck JG, Makino S, Soe LH et al. RNA recombination of coronavirus. Adv Exp Med Biol. 1987; 218:99–107.
25. Liao C, Lai MMC. RNA Recombination in a Coronavirus: Recombination between Viral Genomic RNA and Transfected RNA Fragments. J Virol. 1992;66: 6117-6124.
26. Palanivelu P. Analyses of Homing Endonucleases and Mechanism of Action of CRISPR-Cas9 HNH Endonucleases. Int J Biochem Res Rev. 2020;29:1-25.
27. Palanivelu P. Eukaryotic multi-subunit DNA dependent RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism. In: Advances and Trends in Biotechnology and Genetics. 2019;1:1-38. SCIENCEDOMAIN International Book Publishers, UK, Print ISBN: 978-93-89246-59-9. DOI: 10.9734/bpi/atbg/v1;2019.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/60411