Proceedings of the Indian National Science Academy (2023) 89:60–76
https://doi.org/10.1007/s43538-022-00140-y

RESEARCH PAPER

Comparative sequence analysis of SARS nCoV and SARS CoV genomes for variation in structural proteins

Jyoti Sangwan[1] · Sandhya Tripathi[2] · Nisha Yadav[1] · Yogesh Kumar[3] · Neelam Sangwan[1]

Received: 1 November 2022 / Accepted: 25 November 2022 / Published online: 20 December 2022

Abstract

SARS-nCoV was identified as corona virus had spread worldwide very quickly and affected more than million people worldwide. To halt this acceleration and for efficient control the knowledge on genomic information is of utmost importance. We attempted to determine the nature of variation i.e., insertion, deletion, substitution, among structural sequences required to code for membrane, spike, nucleocapsid, envelope protein and glycosylation variation between SARS CoV and SARS nCoV spike glycoproteins, respectively. Comparative sequence analysis was performed by using retrieved sequences from the NCBI database. The analyzed sequences revealed, that the sequences coding for envelope protein show minor substituting amino acids. SARS CoV showed 94.74 percent amino acid identities with SARS nCoV amino acid sequences coding for envelope protein. In comparison to SARS nCoV, distinct amino acid residues vary in SARS CoV sequences coding for membrane, nucleocapsid, and spike proteins, respectively. S protein coding sequences of SARS CoV exhibited one deletion, six insertion and six hundred three substitutions in SARS nCoV sequence. Insertion of valine was found in receptor binding domain of SARS nCoV at position 487, and NSPR amino acid residues at position 683–686. Deletions and substitutions were also found in nucleotide sequences of strain B.1.617.2 of SARS nCoV. Additionally, binding interaction pattern of ACE2 receptor protein with original wild-type SARS-CoV-2 strain with the recently evolved Omicron variant was also evaluated. The docking results substantiated that the specific variation in binding residues is likely to impact virulence pattern of both variants.

Keywords SARS nCoV · SARS CoV · Sequence analysis · Insertion · Deletion · Substitution

Introduction

SARS nCoV is novel coronavirus, with names as SARS-CoV2, SARS-CoV19, and COVID19 (Paraskevis et al. 2020). Coronaviruses (CoVs) are spherical to pleomorphic enveloped single stranded positive sense RNA viruses having club shaped spike glycoproteins, projected from their surfaces. Spike projections from the surfaces of CoVs appear like crown, thus given the name, coronavirus (Tyrrell and Myint 1996) (Fehr and Perlman 2015). Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus was the cause of worldwide pandemic 2019 and it imposed huge health, socio-economic burden with unparallel consequences (Gómez et al. 2021). Taken aback origin of corona virus, it was begun in early 1930 when a respiratory infection was shown in domesticated chicken by a virus, known as IBV. The history of HCoV began in the 1960s when two researchers Bynoe and Tyrrell found the virus, known as human corona virus HCoV (Hamre and Procknow 1966) (McIntosh et al. 1967). With the emergence of SARS CoV, some more HCoVs (HCoV-NL63 and HCoV-HKU1) were added to the identification list of coronavirus, which infect respiratory tract in approximately all age groups (Drosten et al. 2003). SARS-CoV had emerged and transmitted to the human from bats with the help of intermediate host (e.g. civets & bats) and then led to worldwide outbreaks of novel respiratory disease. Earliest confirmed SARS nCoV case was reported in China (Wuhan) on December 2019 (Kopecky-Bromberg et al. 2007) and caused a new infectious disease named as Corona Virus Disease 2019 (Zhu et al. 2020). As

✉ Neelam Sangwan
drneelamsangwan@gmail.com; nsangwan@cuh.ac.in

1 Department of Biochemistry, School of Interdisciplinary and Applied Sciences, Central University of Haryana, Mahendergarh, India

2 ICAR-Indian Institute of Pulse Research, Kanpur, India

3 Department of General, Visceral and Thoracic Surgery, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

per the WHO report, a new type of coronavirus, was identified early on January 2020 and its genomic sequences were shared for studies. SARS nCoV is a zoonotic infection same as MERS (Middle East Respiratory Syndrome) and SARS (severe acute respiratory syndrome) (Hui et al. 2020). Structural proteins include membrane, envelope, nucleocapsid protein and spike protein, required for virions assembly and helps CoVs to cause infection. Since, it's start, a large number of people from all over the world suffered the Covid infection (Gómez et al. 2021). COVID-19 has infected estimated 130 million persons as of April 2021, resulting in more than 2.8 million fatalities in 219 nations. Globally, almost 104 million patients have recovered (Böttcher et al. 2021). COVID-19 testing kits were being developed to rapidly and efficiently check the coronavirus infection. With the publication of the genetic sequence for COVID-19 on 11th Jan 2020, the response to prepare the vaccine for COVID-19 started globally (Li et al. 2021). Our present study is focused on analyzing viral genomic characteristics and understanding the structure and nature of sequences coding for structural proteins (Boheemen et al. 2012). In addition, the variability and identity between SARS nCoV& SARS CoV and other variants were also analyzed using bioinformatics approaches. Our goal is to investigate the nature of variations and locate the probable variable sites in the SARS nCoV genome compared to previously reported SARS CoV genomic sequences (Malik et al. 2020) and to find out the variations in the sequences of SARS nCoV variants. Our detailed investigation unfolds the viral sequences coding for structural proteins viz spike glycoprotein, membrane glycoprotein, nucleocapsid protein, small envelope proteins respectively (S, M, N, E) and to analyze the nature of variation located at specific sites, allows estimating the similarity of function of SARS nCoV when compared with previously identified sequences (Masters 2006; Tan et al. 2006). Usually, RNA virus's nucleotide substitution rates are faster than their hosts. Gene mutations such as insertions, deletion and substitutions have been computed while comparing SARS nCoV with SARS CoV, and SARS nCoV variants (Kumar et al. 2020). The analysis also investigated the differences in N-glycosylation sites of 3 coronavirus isolates as well.

Materials and methods

Dataset

To compare spike protein sequences of SARS-nCoV with other corona viruses, we have used 11 sequences retrieved from NCBI Virus Genome database (https://www.ncbi.nlm.nih.gov/genome/viruses/) with accession numbers:—YP_009194639, YP_003767, YP_006908642, YP_009824967, YP_173238, YP_007188579, YP_009724390.1, NP_828851, YP_009824990, YP_009824990 and YP_009701451.

We have done the comparative genomic and proteomic analysis of structural sequences coding for spike, membrane, envelope and nucleocapsid protein of SARS-nCoV and SARS-CoV with accession number NC_045512.2, MT499203.1 and NC_004718.3 (https://www.ncbi.nlm.nih.gov/genome/viruses/). In addition to this, comparative genomic and proteomic analysis of structural sequences coding for spike, membrane, envelope and nucleocapsid protein of SARS-nCoV variant B.1.617 (MZ157006.1), B.1.617.2 (MZ208926.1) and B.1.351 (MZ068161.1) with reference sequence (NC_045512.2) was also done. To study the variations within the B.1.617.2 strain sequences, 8 sequences of B.1.617.2 strain with accession number MZ157012.1, MZ157012.2, MZ157010.1, MZ157009.1, MZ157008.1, MZ157007.1, MZ157005.1 and MZ208926.1 (taken as reference sequence) were retrieved.

Phylogenetic tree construction

Codon based sequence alignment of ten amino acid sequences of S glycoprotein from different species of corona virus and one from Nor Virus (out-group) was performed for the conserved domain sequence using Multiple Sequence Comparison by Log-Expectation [MUSCLE] program in MEGAX (Edgar 2004). The aligned sequence file was used for phylogeny tree construction by MEGAX using Neighbor joining clustering method, and 1000 bootstrap replicates (Edgar 2004). Phylogenetic analysis with same sequences was also performed using Phylogenyfr (http://www.phylogeny.fr/simplephylogeny.cgi) (Kumar et al. 2018).

Genomic and proteomic variations

The genomic analysis was performed to find out the percent identity and statistical variability among three isolates NC_045512.2 (SARS nCoV ref seq), MT499203.1 (SARS nCoV) and NC_004718.3 (SARS CoV). It was implied that the composition of both SARS nCoV strains must contain 6 ORFs so that they can be excluded from each of the two isolates and should have structural sequences coding for E, M, N and S proteins. The percentage identity, GC % content variation, and statistical analysis for all ORFs and structural sequences were analysed. To study deletion, insertion and substitution of nucleotide and amino acids, we specifically focus our study on structural sequences coding for E, M, N and S protein. Codon based sequence alignment of three structural sequences coding for E, M, N and S protein was performed for the CDS (Conserved domain sequence) using MUSCLE program in MEGAX (https://www.megasoftware.net/). By analyzing MSA (Multiple sequence alignment), we found the variabilities among nucleotide and amino

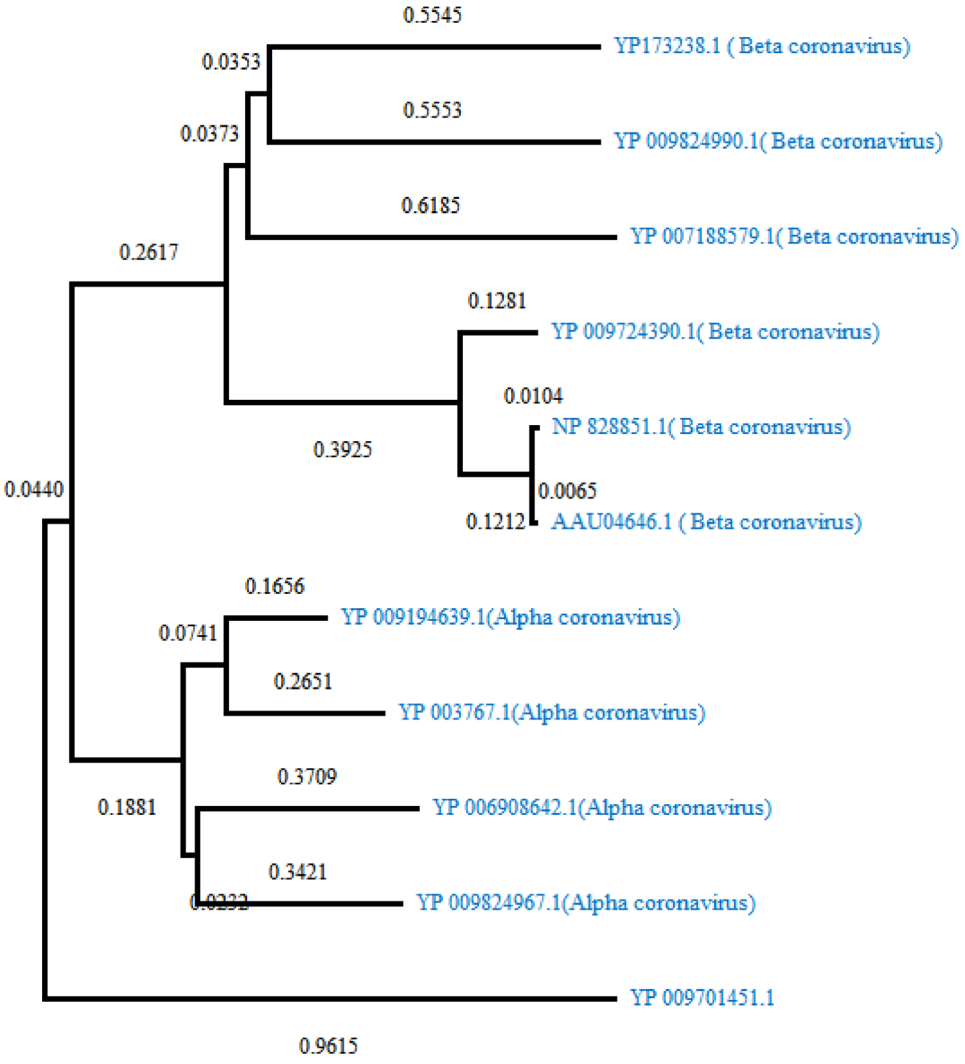

Fig. 1 Phylogenetic analysis of of SARS nCoV

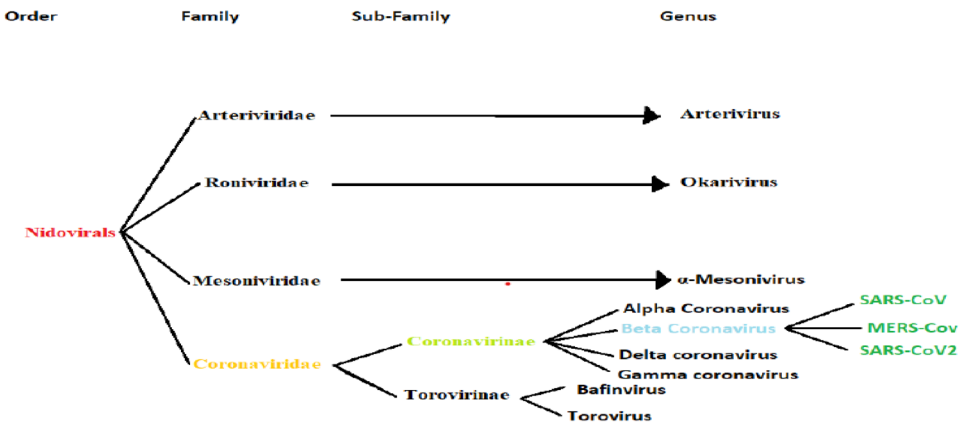

Fig. 2 Classification of Coronavirus-CoVs belongs to the order *Nidovirales* which includes 4 families: *Mesoniviridae, Arteriviridae, and Roniviridae* and *Coronaviridae* families. *Coronavirinae* family comprises in *Coronaviridae* sub-family *Torovirinae*. 4 Genra *alpha, beta, gamma and delta* CoVs are grouped under *Coronavirinae. Beta* viruses of *coronavirince* subfamily splitted into four lineages

acid sequences. We have also used the MSA to search the synonymous and non-synonymous substitutions and then analyzed, which structural sequences have more number of synonymous substitutions and conservative missense mutations ((Lokman et al. 2020; Chatterjee 2020)). With the same method as mentioned above, comparative genomic and proteomic analysis of structural sequences coding for spike, membrane, envelope and nucleocapsid protein of

Table 1 Description of retrieved sequences of spike proteins

S. no	Virus	Host	Nucleotide bases	No. of amino acids	Spike protein id	Location	Year of samples	Reference
1	Alpha CoV	Camel	27,395	1169	YP_009194639.1	S. Arabia	2018	Sabir et al. (2016)
2	Bat HKU1	*Chiropteran*	28,494	1349	YP_006908642.1	China	2018	Lau et al. (2012)
3	MERS	Human	30,111	1353	YP_007188579.1	England	2020	Groot et al. (2013)
4	SARS nCoV	Human	29,903	1273	YP_009724390.1	China	2020	Wu et al. (2020)
5	HCoVHKU1	Human	29,926	1356	YP_173238.1		2020	Woo et al. (2005)
6	HCoV NL63	Bat	28,679	1373	YP_009824967.1	Kenya	2020	Tao et al. (2017)
7	HCoV NL63	Human	27,553	1356	YP_003767.1		2018	Hoek et al. (2004)
8	SARS	Human	29,751	1259	NP_828851.1	Canada	2020	He et al. (2004)
9	SARS CoV	Civet	29,540	1255	AAU04646.1	China	2005	Wang et al. (2005)
10	Bat CoV	Bat	28,975	1269	YP_009824990.1	Cameroon	2020	Yinda et al. (2018)
11	Nor virusGII	Human	7525	128	YP_009701451.1	Peru	2019	Tohma et al. (2018)

SARS-nCoV variant B.1.617 (MZ157006.1), B.1.617.2 (MZ208926.1) and B.1.351 (MZ068161.1) concerning reference sequence (NC_045512.2) was also done. Similarly, we also analyzed the variations present in the 8 sequences of SARSnCoV variant B.1.617.2 with accession numbers MZ157012.1, MZ157011.1, MZ157010.1, MZ157009.1, MZ157008.1, MZ157007.1, MZ157005.1 and MZ208926.1 (taken as reference sequence).

Table 2 Description of retrieved sequence for comparative genomic and proteomic analysis of structural sequences coding for S, M, E and N protein

	SARS nCoV	SARS nCoV	SARS CoV
Geological location	Wuhan, China	USA	Canada
Accession no	NC_045512.2	MT499203.1	NC_004718.3
Protein ID	YP_009724390.1	QJX45223.1	NP_828851.1
5'UTR (position)	1–265	1–265	1–264
Orf1ab	266–21,555	266–21,555	265–21,485
S(spike)	21,563–25,384	21,563–25,384	21,492–25,259
Orf3a	25,393–26,220	25,393–26,220	25,268–26,092
E (envelope)	26,245–26,472	26,245–26,472	26,117–26,347
M (membrane)	26,523–27,191	26,523–27,191	26,398–27,063
Orf6	27,202–27,387	27,202–27,387	26,913–27,265
Orf7a	27,394–27,759	27,394–27,759	27,273–27,641
Orf7b	27,756–27,887	27,756–27,887	27,638–27,772
Orf8	27,894–28,259	27,894–28,259	27,779–28,118
N (nucleocapsid)	28,274–29,533	28,274–29,533	28,120–29,388
3'UTR	29,675–29,903	29,675–29,903	29,389–29,751

Glycosylation site variations on S (spike) glycoproteins

To find out the variations in the attachment sites of S glycoproteins of SARS nCoV and SARS CoV to the surface of the host cell, the glycosylation site was determined using NetNGly 1.0 software (http://www.cbs.dtu.dk/services/NetNGlyc/) and validated these glycosylation sites by N-GlyDE software (http://www.cbs.dtu.dk/services/NetNGlyc/) (Kumar et al. 2020).

Three dimensional structure and docking analysis

The 3D structures of SARS-CoV-2 strain and Omicron variants were downloaded from RCSB PDB database (https://www.rcsb.org/) and the protein–protein docking was performed using Cluspro server (https://cluspro.bu.edu/). The results were visualized using Pymol software version 2.5.1.

Result and discussion

Phylogenetic tree analysis

All the 11 sequences coding for spike glycoprotein were aligned using MUSCLE program in MEGAX. Evolutionary tree was generated using NJ method as shown in Fig. 1 with the sum of branche length of tree = 5.08122990. The evolutionary distance in tree was computed in MEGAX using Poisson correction method. Phylogenyfr was used for evolutionary tree construction with the same amino acid sequences. Phylogenetic tree analysis shows that SARS

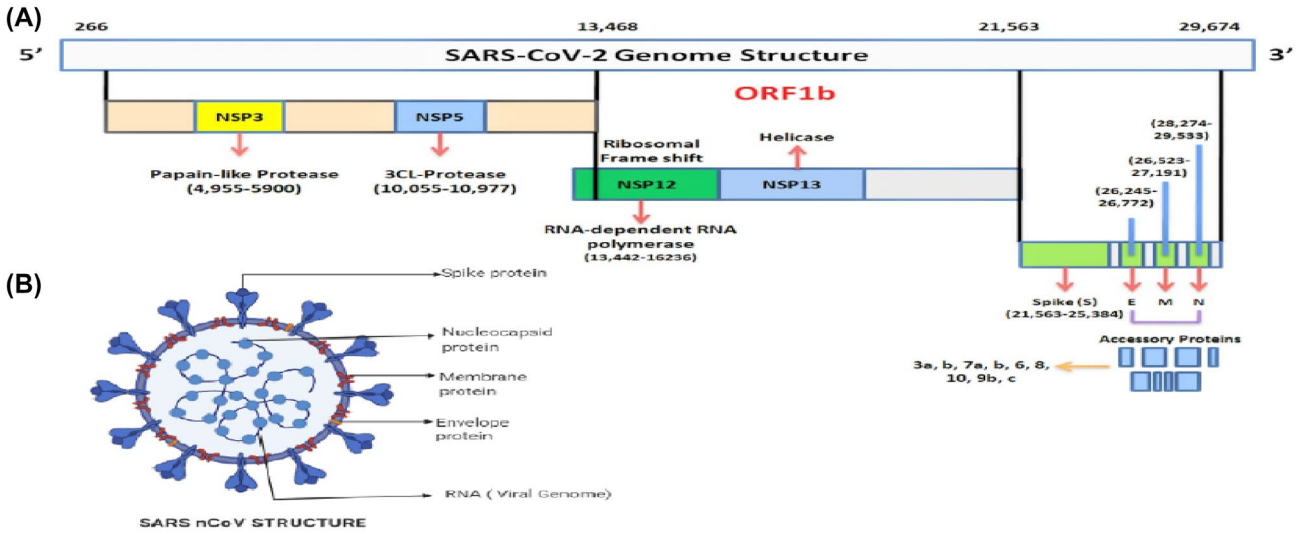

Fig. 3 Gene structure and morphology-Different genes code for different protein as shown in figure. Genes nsp3 and nsp2 code for papain like protease and 3CL-Protease respectively whereas nsp12 and nsp13 code for RNA dependent RNA polymerase and helices enzymes respectively. S, M, E and N genes code for spike, membrane, envelope and nucleocapsid

nCoV is evolutionary closely related to SARS CoV and civet SARS CoV.

SARS CoV and newly reported SARS nCoV are grouped under B lineage while MERS-CoV belongs to the C lineage of *Beta* viruses of *Coronavirinae* subfamily belonging to *Coronaviridae* families (Cui et al. 2019) (Shafique et al. 2020). Detailed classification of coronavirus is shown in Fig. 2 (See Table 1).

Genomic and proteomic variations analysis

As we have taken the three isolates of coronavirus for the variation analysis, their detailed description is given in the Table 2. We have mainly focused on the analysis of variations in structural sequences of different isolates of coronavirus. For better analysis of sequences, first we have to clearly understand the structure. Different genes code for different proteins and show some specific functions as shown in Figs. 3 and 4. Gene's nsp3 and nsp2 code for papain like protease and 3CL-Protease respectively. Gene's nsp12 and nsp13 code for RNA dependent RNA polymerase and helices enzymes respectively. S, M, E and N genes code for spike, membrane, envelope and nucleocapsid.

Homology analysis

Percent identity among structural sequences for S, E, N and M proteins of these three isolates nCoVNC_045512 (ref seq), nCoVMT499203 and CoVNC_004718 was found out through BLAST pair wise alignment.

Number of nucleotide and %I (% identity) was seen with respect to reference sequence nCoVNC_045512. Nucleotide sequence of S shows highest variability among all in comparison with reference sequence while Orf6 CoVNC_004718 has lowest % AI as compared to all amino acid sequences analyzed in comparison with reference sequence (ord Table 3). We found that by comparing gene sequences of CoVNC_004718 and nCoVNC_045512 (ref seq.), nCoVMT499203.1 CoVNC_004718 gene sequences have higher %GC content.

Variation profiling of SARS CoV with two SARS nCoV sequences

Variation profiling of sequences coding for E protein Codon based multiple sequence alignment of structural sequences coding for E protein was performed for the CDS. By analyzing MSA we found that ECoVNC_004718 is 3 bp (additional nucleotides GAA) larger as compared to EnCoVNC_045512 and EnCoVMT499203 sequences as shown in Fig. 5.

Position of nucleotide is not according to whole genome sequence, rather position No 1 means first nucleotide of gene coding for E protein which are aligned on codon based method by MUSCLE using MEGAX (same for amino acid position).

Total 13 substitutions are analyzed in the ECoVNC_004718 sequence with respect to reference sequence. Out of the 13 substitutions, only 5 are non-synonymous substitutions (results in three amino acid change) and eight are synonymous substitutions. Here, synonymous mutations are less which means that change in amino acid constitution is less. All the details about the substitution sites are mentioned in Table 4.

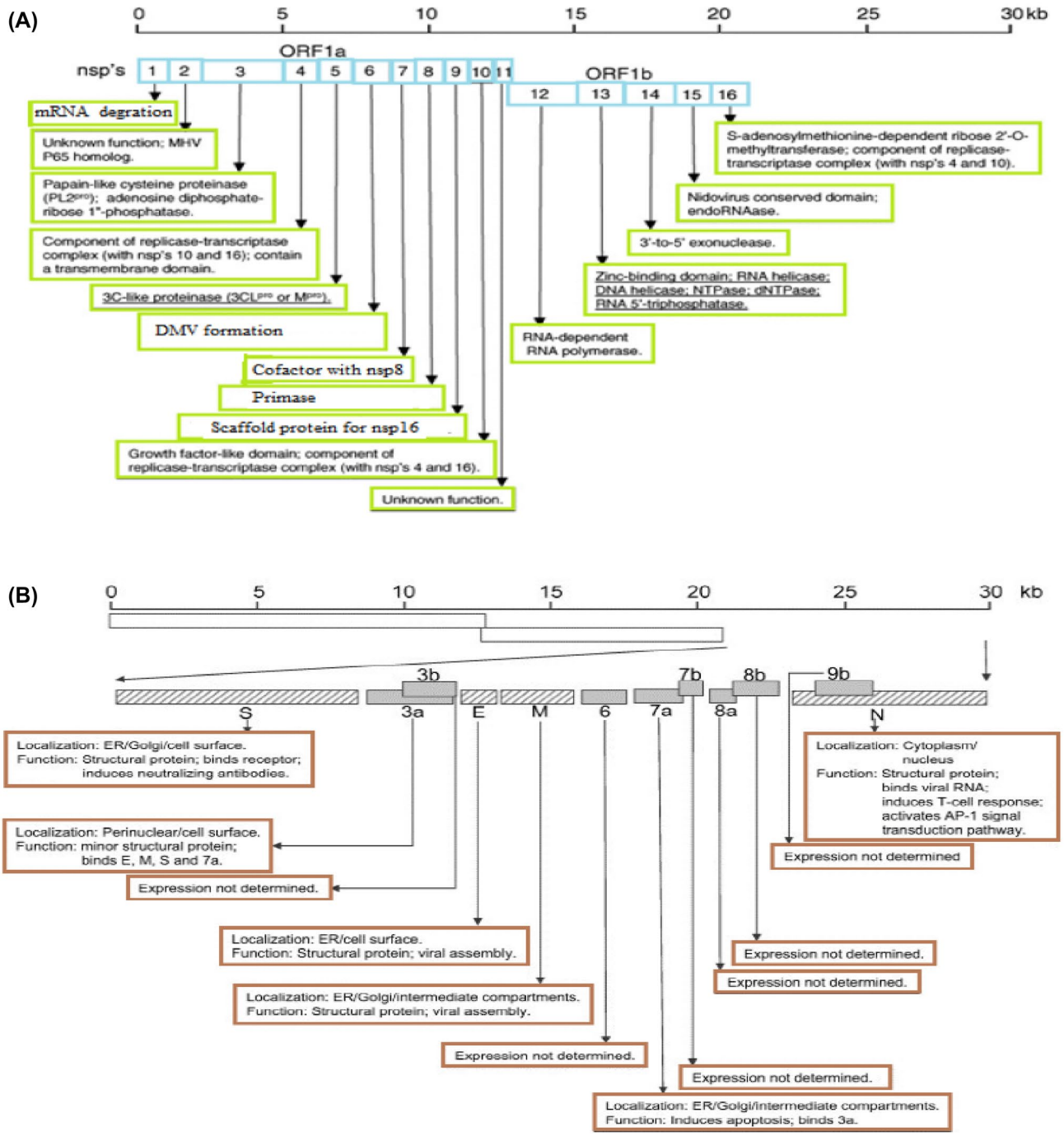

Fig. 4 Gene function- **A** ORF1a and ORF1b **B** Structural proteins and ORFs of remaining one third parts of genome-all genes and their functions are shown in boxes located below them

Variation profiling of sequences coding for M protein Pairwise sequence alignment analysis of M sequences shows that MCoVNC_004718 has 85.52% nucleotide identity with 573 identical nt. sites when compared with MnCoVNC_045512. AI% is 90.5% which shows the preserved sequence length of MPnCoVNC_045512 with MPCoVNC_004718) as shown in Fig. 6. On analysis of M coding sequences, it was found that there is total 95 nucleotodes substitution, out of which 27 nucleotides substitutions were non-synonymous are present in MCoVNC_004718 in comparison to other 2 isolates. These 27 non-synonymous substitutions lead to the 16 amino acid change in MCoVNC_004718 seq. A complete substitution of codon TGT > ATG (start position of codon

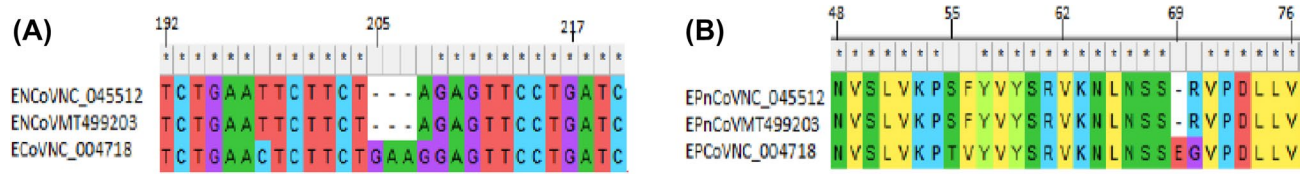

Fig. 5 Variations at specific regions of **A** nucleotide sequences and, **B** amino acid sequences of nCoVNC_045512, nCoVMT499203 in comparison to CoVNC_004718- By analyzing MSA we found that ECoVNC_004718 is 3 bp (additional nucleotides GAA) larger as compared to EnCoVNC_045512 and EnCoVMT499203 sequences. Additional nucleotides GAA code for glutamic acid (E). Total 13 substitutions are analyzed in the ECoVNC_004718 sequence with respect to reference sequence. Out of the 13 substitutions, only 5 substitutions are non-synonymous substitutions (results in 3 amino acid change) and 8 substitutions are synonymous substitution. Here, synonymous mutations are less which means that change in amino acid constitution is less none. All details about substitution sites are mentioned in Table 4

Table 3 Homology study of nCoVNC_045512 vs nCoVMT499203 and nCoVNC_045512 vs CoVNC_004718

	% AI (Amino acid identity)		% NI (Nucleotide identity)		Nt. match		% GC	
Genes	nCoVMT499203	CoVNC_004718.3	nCoVMT499203	CoVNC_004718.3	nCoVMT499203	CoVNC_004718.3	nCoVMT499203	CoVNC_004718.3
Orf1ab	99	86.12	99.99	81.47	21288/21290	14867/18249	37.5	40.8
S	100	75.46	99.97	74.48	3821/3822	2782/3735	37.3	38.8
Orf3a	100	72.36	100	75.63	828/828	627/829	39.5	40.0
E	100	94.74	100	93.50	228/228	216/231	38.1	40.2
M	100	90.54	100	85.52	669/669	573/670	42.6	45.2
Orf6	100	68.85	100	76.76	186/186	185/346	28.0	37.0
Orf7a	100	82.25	100	82.11	366/366	303/369	38.3	40.1
Orf7b	100	81.40	100	86.18	132/132	106/123	31.1	31.9
N	99.17	90.54	99.92	88.17	1259/1260		47.1	

Table 4 Position of substitution of nucleotide and respective amino acid change in 2 isolates sequences for Envelope (E) protein as compared to reference sequence

Nucleotide Substitution			Amino acid Substitution		
Position	nCoVMT499203	CoVNC_004718	Position	nCoVMT499203	CoVNC_004718
24	No change	G→A	8	No change	Silent
87	No change	T→C	29	No change	Silent
151	No change	C→T	51	No change	Silent
153	No change	T→A	51	No change	Silent
162	No change	T→A	54	No change	Silent
163	No change	T→A	55	No change	S→T
165	No change	T→G	55	No change	S→T
166	No change	T→G	56	No change	F→V
174	No change	T→C	58	No change	Silent
180	No change	T→G	60	No change	Silent
198	No change	T→C	66	No change	Silent
205	No change	A→G	69	No change	R→G
206	No change	G→A	69	No change	R→G

is 97) leads to the change in C > M in MCoVNC_004718 when compared to other 2 isolates. All details about substitution sites are mentioned in Table 5.

Variation profiling of sequences coding for N protein On a detailed analysis of coding sequences of N protein unfolded that length of N sequence of SARS nCoV is 1260 bp, while the sequence length of SARS COV were 1269 bp i.e. NCoVNC_004718 sequence, is nine bps

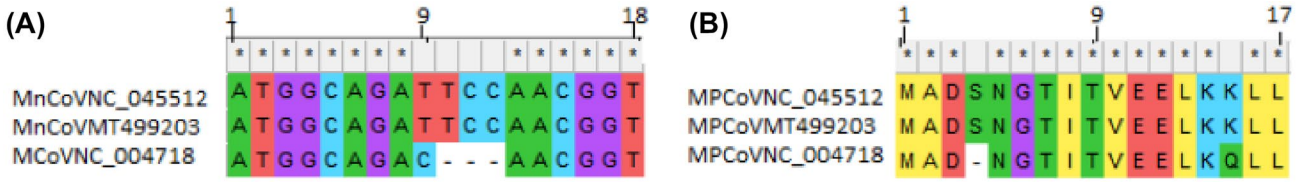

Fig. 6 Absence of 3 bps and one amino acid residue from the M sequence of MCoVNC_004718-TCC is absent in below sequence coding for Serine amino acid

Table 5 Position of non-synonymous substitution of nucleotide and respective amino acid change in 2 isolates sequence for M protein as compared to reference sequence

Nucleotide substitution			Amino acid substitution		
Position	CoVMT499203	CoVNC_004718	Position	CoVMT499203	CoVNC_004718
43	No change	A→C	15	No change	K→Q
45	No change	G→A	15	No change	K→Q
88	No change	A→G	30	No change	T→A
90	No change	A→C	30	No change	T→A
97	No change	T→A	33	No change	C→M
98	No change	G→T	33	No change	C→M
99	No change	T→G	33	No change	C→M
118	No change	G→T	40	No change	A→S
120	No change	C→T	40	No change	A→S
154	No change	A→G	52	No change	I→V
226	No change	A→G	76	No change	I→V
289	No change	A→G	97	No change	I→V
374	No change	A→G	125	No change	H→R
375	No change	T→G	125	No change	H→R
400	No change	C→A	134	No change	L→M
402	No change	A→G	134	No change	L→M
453	No change	T→G	151	No change	I→M
464	No change	C→T	155	No change	H→S
465	No change	T→C	155	No change	H→S
563	No change	C→G	188	No change	A→G
565	No change	G→A	189	No change	G→T
566	No change	G→C	189	No change	G→T
590	No change	G→A	197	No change	S→A
591	No change	T→C	197	No change	S→A
631	No change	T→G	211	No change	S→A
641	No change	G→A	214	No change	S→N
642	No change	T→C	214	No change	S→N

larger than other two isolates as shown in Fig. 7. Our pair wise sequence alignment analysis of N sequences showed that NPCoVNC_004718 has 90.52% AI when aligned with NPnCoVNC_045512. On analysis of N coding sequences, it was found that there are total 141 nucleotide. substitution, out of which 52 nt. substitutions are non-synonymous present in NCoVNC_004718. These 52 non-synonymous substitutions lead to the 37 amino acid changes in NCoVNC_004718 seq. While in case of NCoVMT499203 only one non-synonymous substitution occurs. A complete substitution of codon 577AAC > ATG Position of codon is 577, 802GCA > CAG and 1003ACA > CAT leading to the change in amino acid residues are N > G, A > Q and T > H residues in NCoVNC_004718 when compared to other two isolates. Substitution of serine to aspargine amino acid occurs in NPCoVMT499203 in comparison to other two isolates. We know that serine and aspargine are similar type of amino acid i.e. uncharged polar amino acids which may not affect much change in that protein function (Table 6).

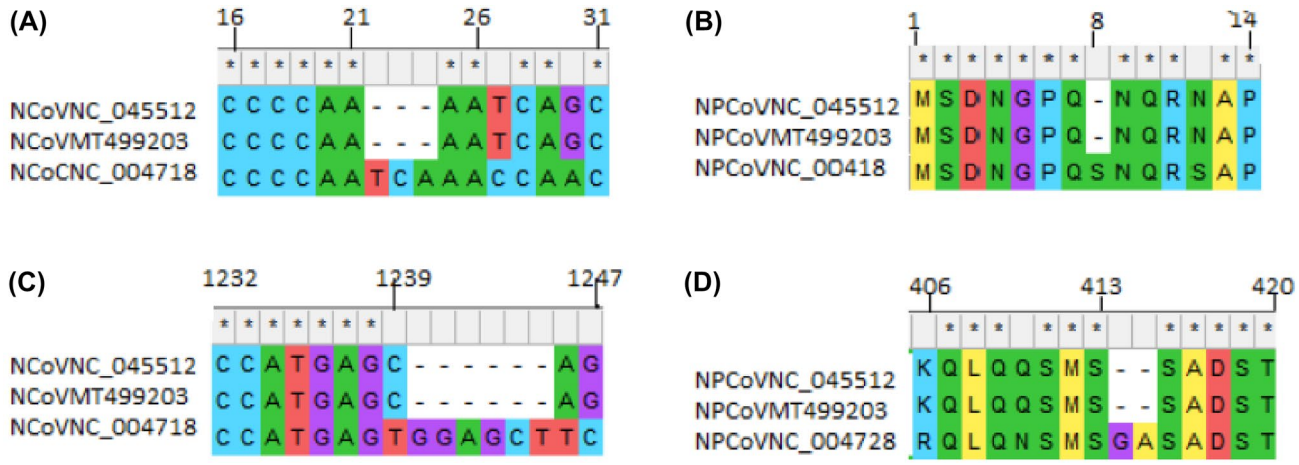

Fig. 7 Presence of extra nucleotide and amino acid residue in sequence of NCoVNC_004718- shows deletion of 8 nucleotides

Variation profiling of sequences coding for S protein The S protein of virus is homotrimer which makes the spikes on the CoV surface which is important for attachment to cellular receptor. It consists of three domains (shown in Fig. 8); large ectodomain, single pass membrane anchor and a small sized intracellular tail. The ectodomain contains 2 subunits; S1 receptor binding and S2 membrane fusion subunit. It is a homotrimer, having 3 S1 heads and S2 trimeric stalk, first S1 binds to host cell receptor for attachment to virus, second S2 mediates the fusion of virus and host membrane, initiates the infection cycle (Li 2016).

On analysis of codon-based MSA shows that SCoVNC_004718 has a total of 994 substitutions out of non-synonymous substitutions of 603, leading to 176 amino acid change. SnCoVMT499203 have only one synonymous substitution which is A > T at 1056 nucleotide position as shown in Fig. 9.

Pair wise sequence analysis of S protein sequences showed that SARS nCoV isolates and SARS CoV have nearly 75.46% similarity. It was found that 22 amino acids are present in SARS nCoV 2 isolates which results from 6 insertions shown in Fig. 10. (C) (D) (E) (F) (G) (H). Out of these insertions, 76GTNGTKR82 is major insertion shown in figure (D). N-terminal domain of spike protein contains two insertions which are 148YYHK151 and 247ALHR250 as shown in Figure (F) (H). Insertion of valine is present at 487 positions in RBD domain figure (J). Another insertion 683NSPR686 is present upstream to cleavage site of S1/S2 that leading the formation of PRRARS (furin like cleavage site) in SARS-nCoV isolates.

Glycosylation site variations on S glycoproteins The N-glycosylation sites of both SARSnCoV and SARSCoV were presented in Table 7, Fig. 11, where 0.5 N-glycosylation potential was set as cutoff. On comparison of SARS nCoV with SARS CoV we have observed that S protein have different glycosylation sites such as NLTT, NVTW, NGTK, NATN, NKSW, and NATR that can results in sequence variation. We also found some common glycosylation site in all 3 isolates such as NCTF, NITN, NASV is basis of similarity and differences in glycosylation site, it may be suggested that SARS nCoV interacts to ACE2 host receptor using these different glycosylation sites due to that internalization process may be affected [Fig. 12].

Variation profiling of SARS nCoV with its variants

Variation profiling of sequences coding for S protein- On analysis of codon based MSA shows that there is deletion of starting 1259 nucleotides from the B1 variant sequence compared to the reference sequence. Deletion in the B2 variant sequence occurs at 467AGTTCA472 and in the B3 variant sequence at 722TACTTGCTT730 compared to the reference sequence. All synonymous and non-synonymous substitutions are shown in Table 8.

Variation profiling of sequences coding for E protein- In the sequence coding for E protein, only one non synonymous substitution occurs: C > T at 212 nucleotide position, which leads to change in proline to leucine at 71 amino acid position in B3 sequence.

Variation profiling of sequences coding for M protein- There is one synonymous substitution C > T occurs in B1 variant sequence at 159 nucleotide position while one non synonymous substitution T > G and T > C at 245 nucleotide position in B1 and B2 variant sequences as compare to reference sequence respectively. This non synonymous substitution results in different amino acids

Table 6 Position of non-synonymous substitution of nucleotide and respective amino acid change in 2 isolates sequence for N protein as compared to reference sequence

Nucleotide substitution			Amino acid substitution		
Position	CoVMT499203	CoVNC_004718	Position	CoVMT499203	CoVNC_004718
35	No change	A→G	12	No change	N→S
77	No change	G→A	26	No change	G→D
80	No change	G→A	27	No change	S→N
95	No change	A→G	32	No change	E→G
101	No change	G→A	34	No change	S→N
112	No change	S→P	38	No change	S→P
192	No change	C→A	64	No change	D→E
197	No change	A→G	66	No change	K→R
238	No change	A→G	80	No change	S→G
283	No change	A→G	95	No change	I→V
312	No change	T→G	104	No change	D→E
361	No change	G→T	121	No change	G→S
362	No change	G→C	121	No change	G→S
387	No change	D→E	129	No change	D→E
394	No change	A→G	132	No change	I→V
457	No change	G→A	153	No change	A→N
458	No change	C→A	153	No change	A→N
472	No change	T→C	158	No change	I→T
577	No change	A→G	193	No change	N→G
578	No change	A→G	193	No change	N→G
579	No change	C→T	193	No change	N→G
581	No change	G→A	194	No change	S→N
608	G→A	No change	203	S→N	No change
617	No change	C→A	206	No change	T→N
637	No change	G→A	213	No change	G→S
641	No change	A→G	214	No change	N→G
642	No change	T→A	214	No change	N→G
651	No change	T→A	217	No change	D→E
652	No change	G→A	218	No change	A→T
703	No change	A→G	235	No change	M→V
705	No change	G→T	235	No change	M→V
802	No change	G→C	268	No change	A→Q
803	No change	C→A	268	No change	A→Q
804	No change	A→G	268	No change	A→Q
873	No change	A→C	291	No change	E→D
1003	No change	A→C	335	No change	T→H
1004	No change	C→A	335	No change	T→H
1005	No change	A→T	335	No change	T→H
1036	No change	C→A	346	No change	N→Q
1038	No change	T→A	346	No change	N→Q
1048	No change	C→A	350	No change	Q→N
1050	No change	A→C	350	No change	Q→N
1129	No change	G→A	377	No change	A→T
1138	No change	A→G	380	No change	T→A
1144	No change	G→C	382	No change	A→P
1146	No change	C→T	382	No change	A→P
1172	No change	A→C	391	No change	Q→P
1173	No change	A→C	391	No change	Q→P
1201	No change	T→A	401	No change	L→M

Table 6 (continued)

Nucleotide substitution			Amino acid substitution		
Position	CoVMT499203	CoVNC_004718	Position	CoVMT499203	CoVNC_004718
1217	No change	A→G	406	No change	K→R
1228	No change	C→A	410	No change	Q→N
1230	No change	A→T	410	No change	Q→N

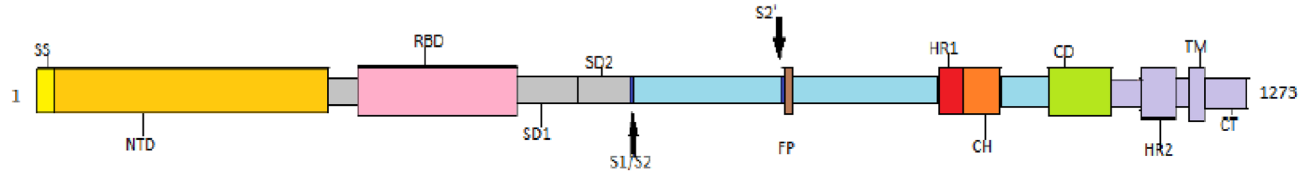

Fig. 8 Structure of spike protein- Spike consists of 3 domains; large ectodomain, single pass membrane anchor and a small sized intracellular tail. The ectodomain contains 2 subunits; S1 receptor binding and S2 membrane fusion subunit

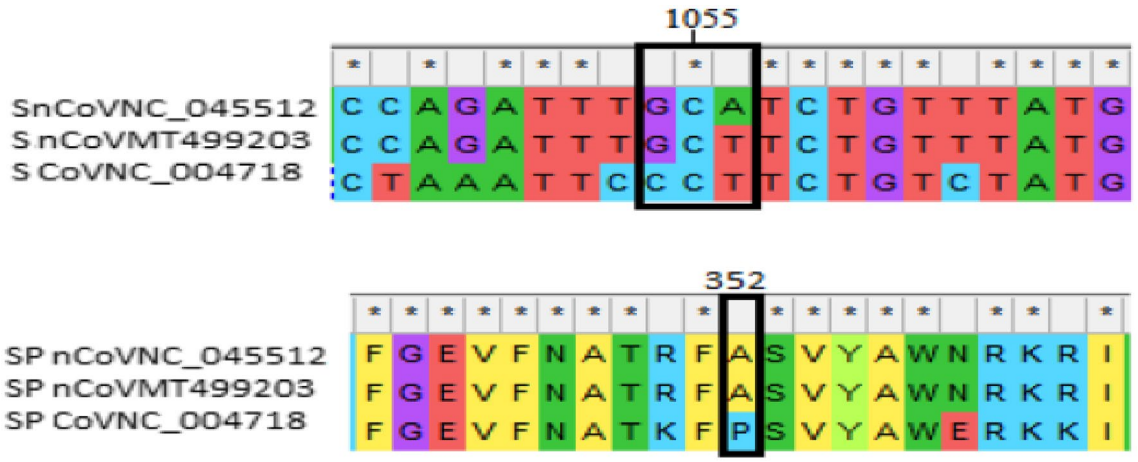

Fig. 9 Showing synonymous substitution in SnCoVMT499203- SnCoVMT499203 has only one synonymous substitution which is A > T at 1056 nucleotide position

which is Isoleucine > Serine and Isoleucine > Threonine at 82 amino acid position for B1 and B2 variant sequences as compare to reference sequence respectively.

Variation profiling of sequences coding for M protein In case of B2 variant sequence, there is deletion of last 176 nucleotides code for M protein with respect to reference sequence. All the non synonymous substitutions for different variants are mentioned in Table 9.

Profiling of variations in SARS nCoV variant B.1.617.2 sequences

Comparative analysis of nucleotide sequences of B.1.617.2 strain found out some variations in the sequences with respect to reference sequence (MZ208926.1) as mentioned in Table 10. There is only one non- synonymous substitution at 241 positions of MZ157012.1 sequence codes for membrane protein which results in substitution of Alanine to Serine at 81 positions.

As we found variations such as deletions and substitutions in the nucleotide sequences of strain B.1.617.2 of SARS nCoV shows that mutation occurs at fast rate, and this may

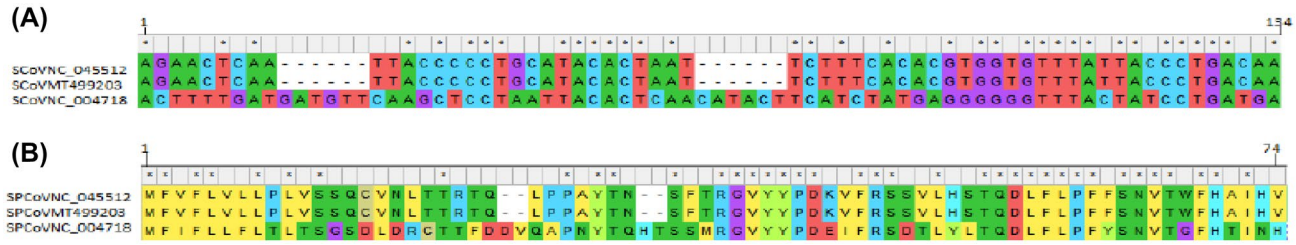

Fig. 10 Showing deletion in CoVNC_045512 and CoVMT499203-**A** 12 nucleotides deletion in SCoVNC_045512 and SCoVMT499203 **B** 4 amino acid deletion in SPCoVNC_045512 and SPCoVMT499203 as compared to SARS CoV

leads to difference in virulence, infectivity and transmissibility at different regions in the world. We believed that such type of genomic and proteomic analysis need to be done at earliest stages to make easy understanding of disease diagnosis, viral adaptability and transmission dynamics and then correlate with different clinical characteristics. If we monitor the emerging mutations, it will help develop better formulation of vaccine and design of antiviral drugs.

Three D Structure and protein–protein docking

We have here compared the binding interaction pattern of ACE2 receptor protein with original wild-type SARS-CoV-2 strain with the recently evolved variant Omicron. It was reported that Omicron comprise of 60 mutations with 37 of them present in the spike protein which present the target site for vaccine and antibodies (Yin et al.,

Table 7 Glycosylation site variations on S glycoproteins

SeqName (SARS nCoV)	Position	Potential	Jury agreement	N-Glyc result	SeqName (SARS CoV)	Position	Potential	Jury agreement	N-Glyc result
QJX45223.1	17 NLTT	0.6606	(8/9)	+	NP_828851.1	29 NYTQ	0.7751	(9/9)	+ + +
QJX45223.1	61 NVTW	0.7820	(9/9)	+ + +	NP_828851.1	65 NVTG	0.8090	(9/9)	+ + +
QJX45223.1	74 NGTK	0.7192	(9/9)	+ +	NP_828851.1	73 NHTF	0.4327	(6/9)	–
QJX45223.1	122 NATN	0.6781	(8/9)	+	NP_828851.1	109 NKSQ	0.6081	(7/9)	+
QJX45223.1	149 NKSW	0.6318	(7/9)	+	NP_828851.1	118 NNST	0.4711	(4/9)	–
QJX45223.1	165 NCTF	0.6220	(8/9)	+	NP_828851.1	119 NSTN	0.7039	(9/9)	+ +
QJX45223.1	234 NITR	0.7613	(9/9)	+ + +	NP_828851.1	158 NCTF	0.5808	(7/9)	+
QJX45223.1	282 NGTI	0.7378	(9/9)	+ +	NP_828851.1	227 NITN	0.7518	(9/9)	+ + +
QJX45223.1	331 NTIN	0.5970	(7/9)	+	NP_828851.1	269 NGTI	0.6910	(9/9)	+ +
QJX45223.1	343 NATR	0.5671	(8/9)	+	NP_828851.1	318 NITN	0.6414	(9/9)	+ +
QJX45223.1	603 NTSN	0.5783	(6/9)	+	NP_828851.1	330 NATK	0.6063	(8/9)	+
QJX45223.1	616 NCTE	0.7163	(9/9)	+ +	NP_828851.1	357 NSTF	0.5746	(8/9)	+
QJX45223.1	657 NNSY	0.4724	(6/9)	–	NP_828851.1	589 NASS	0.5778	(6/9)	+
QJX45223.1	709 NNSI	0.3528	(9/9)	–	NP_828851.1	602 NCTD	0.6882	(9/9)	+ +
QJX45223.1	717 NFTI	0.6426	(9/9)	+ +	NP_828851.1	691 NNTI	0.4604	(5/9)	–
QJX45223.1	801 NFSQ	0.6146	(8/9)	+	NP_828851.1	699 NFSI	0.5357	(7/9)	+
QJX45223.1	1074 NFTI	0.4084	(7/9)	-	NP_828851.1	783 NFSQ	0.6348	(9/9)	+ +
QJX45223.1	1098 NGTH	0.5496	(5/9)	+	NP_828851.1	1056 NFTT	0.4342	(5/9)	–
QJX45223.1	1134 NNTV	0.5800	(6/9)	+	NP_828851.1	1080 NGTS	0.5806	(7/9)	+
QJX45223.1	1158 NHTS	0.3730	(9/9)	–	NP_828851.1	1116 NNTV	0.5106	(5/9)	+
QJX45223.1	1173 NASV	0.3998	(8/9)	–	NP_828851.1	1140 NHTS	0.3739	(9/9)	–
QJX45223.1	1194 NESL	0.6791	(9/9)	+ +	NP_828851.1	1155 NASV	0.4000	(8/9)	–

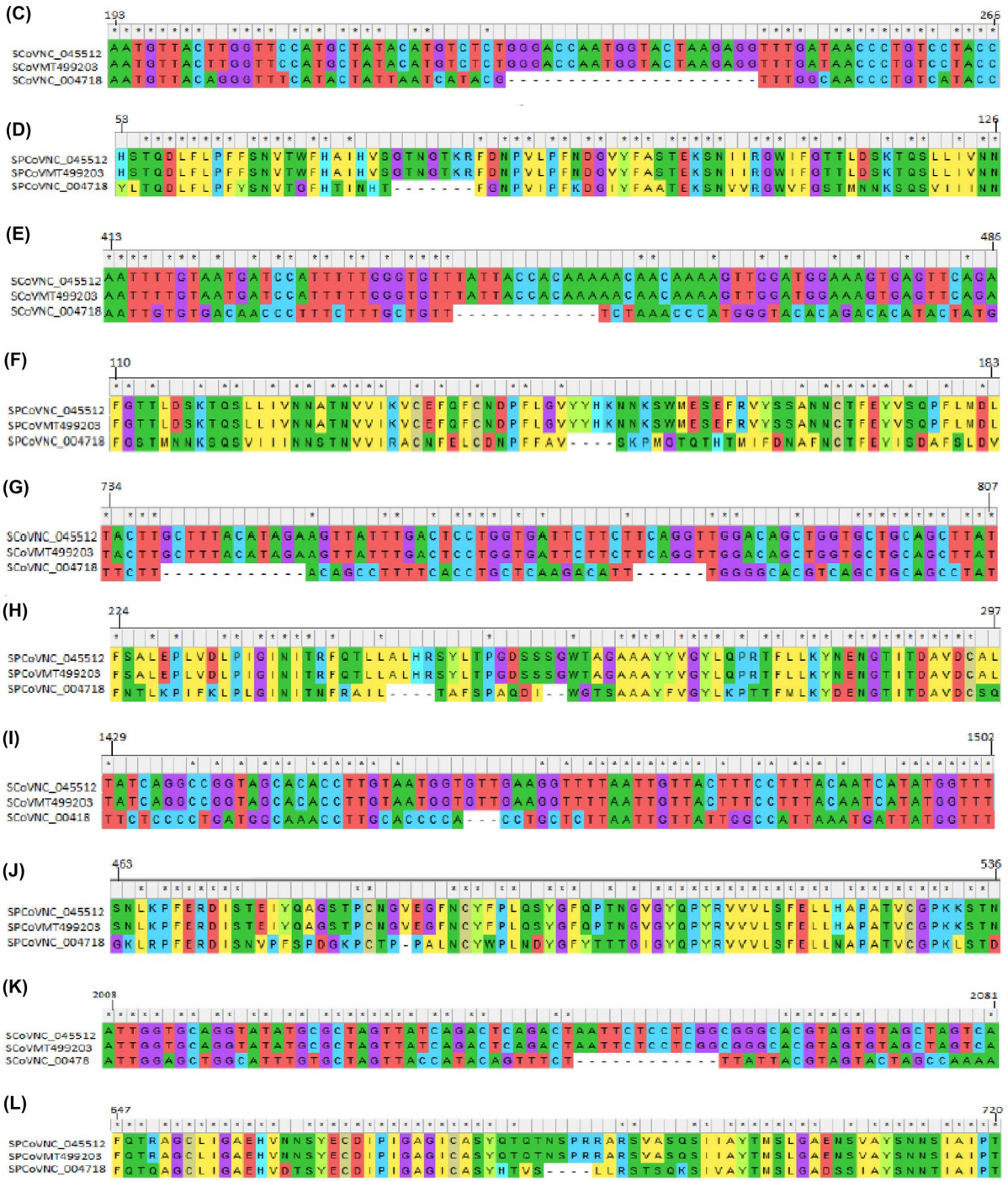

Fig. 11 Showing comparative variations in sequences of three isolates **A** 12 nucleotides deletion in SCoVNC_045512 and SCoVMT499203, **B** 4 amino acids deletion in SPCoVNC_045512 and SPCoVMT499203 as compare to SARS CoV, **C** and **D** major insertion, **E F G H** insertion at N-terminal domain, **I J** insertion at receptor binding domain, **K L** insertion at upstream of S1 S2 cleavage site

2022). Both SARS-CoV-2 (7wp9.pdb) and SARS-CoV-2 Omicron (7a4n.pdb) variant were reported to be present in homo trimer structure form (Yin et al., 2022; Juraszek et al., 2021). We docked A chain of native human angiotensin converting enzyme monomer (ACE) (1o8a.pdb) with the SARS-CoV-2 and SARS-CoV-2 Omicron variant

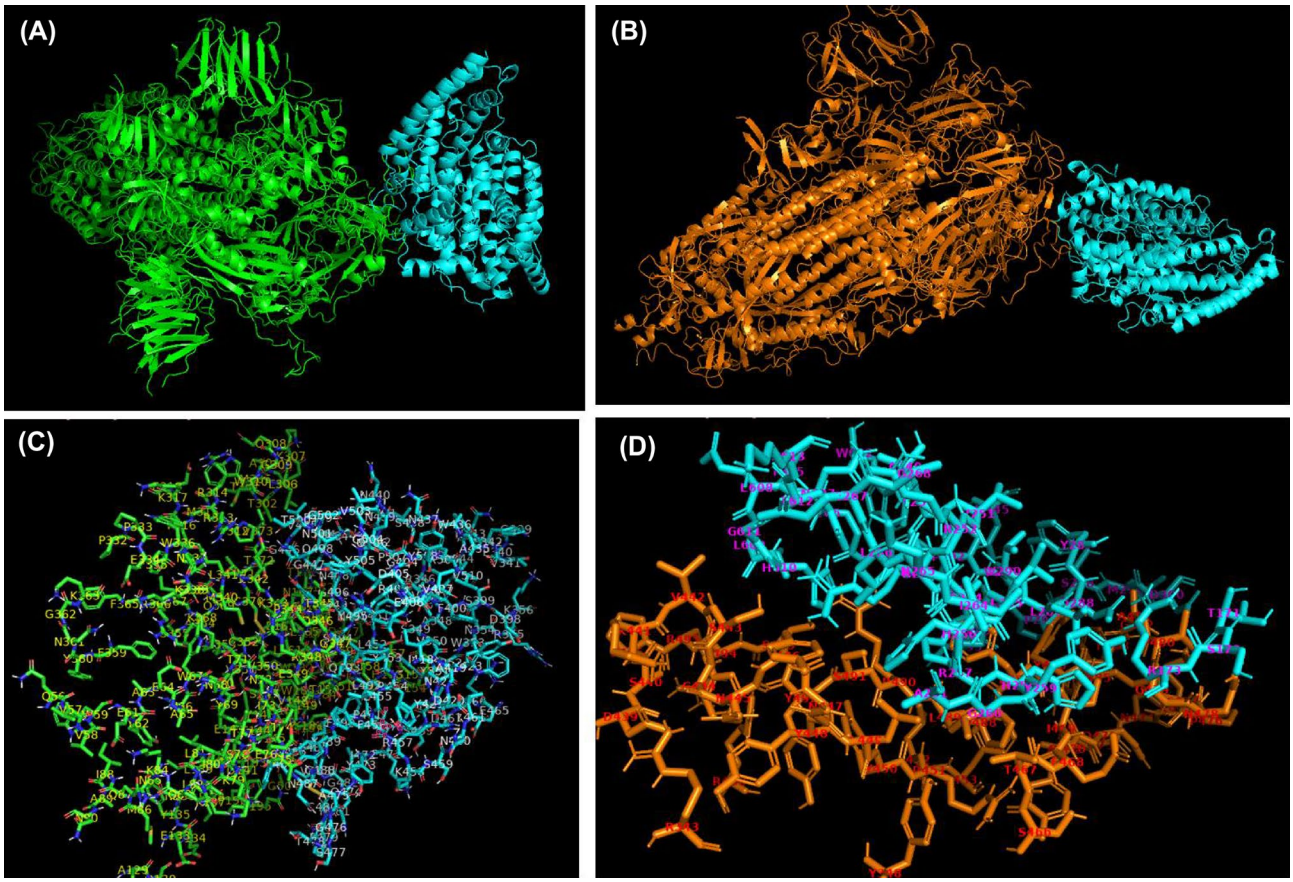

Fig. 12 Three-dimensional structure and protein–protein docking view. Docked 3-D structure view of ACE receptor protein shown in light blue ribbon structure (1o8a.pdb) against corona virus variants **a** SARS-CoV-2 (7wp9.pdb) shown in olive green ribbon structure and **b** SARS-CoV-2 Omicron (7a4n.pdb) shown in orange ribbon structure. The interacting residues for both the corona variants docked with ACE receptor protein is shown in **c** SARS-CoV-2 and ACE protein interacting residues **d** SARS-CoV-2 Omicron and ACE protein interacting residues

to compare the interactions in both the cases. Docking was performed successfully and both the variants were observed to interact in close vicinity with the ACE receptor protein. The major interacting residues between SARS-CoV-2 and ACE receptor docked molecule were V503, G502, Y505, N501, Y449, T500, R403, G496, Q498, F497, G446, V445, G447, S443, K444, N448, Y449, Y505, R403, Y495,S494, Q493, N450, L452, L492, G485, Y489, F486, F456 etc. for ACE protein and I193, A192, A189, Y146, M142, T145, E143, T144, L140, S78, Y69, I73, N70, T71, E349, V350, V351, S147, C370, K343, E342, L341, R313, P312, T371 and so on for SARS-CoV-2 protein. Similarly, in SARS-CoV-2 Omicron variant *versus* ACE docked molecule, the major participating residues in ACE protein were Y259, H258, L255, H256, R257, R2522, A254, R253, A249, Y250, V290, L289, D288, Y287, V291, P292, P294, F293, N445, D300, S298, P297 and A296. The interacting residues for the SARS-CoV-2 Omicron protein partner were T467, S466, E468, I469, F487, Y470, S491, L489, R490, Y470, P488, Y486, C485, Q471, N484, A472, N474, G473, F483, G482, A481 and so on. The best clustered pose with 31 members for SARS-CoV-2 docked protein has lowest energy score of -896.7. Further, the best clustered pose with 116 members for SARS-CoV-2 Omicron variant docked protein has lowest energy score of -1152. It is observed from the docked structures that there is variation in binding residues of both the variants to the human ACE receptor protein which therefore results in the virulence pattern of the variants.

Table 8 Positions of substitution of nucleotide and respective amino acid change in isolates sequence coding for spike protein as compared to reference sequence. R: reference sequence (NC_045512), B1: B.1.617.1 (MZ157006.1), B2: B.1.617.2 (MZ208926.1) and B3: B.1.351 (MZ068161.1)

Nucleotide substitution					Amino acid substitution				
Position	R	B1	B2	B3	Position	R	B1	B2	B3
56	C	–	G	C	19	T	–	R	T
79	G	–	G	T	27	A	–	A	S
239	A	–	A	C	80	D	–	D	A
425	G	–	A	G	142	G	–	D	G
644	A	–	A	G	215	D	–	D	G
1251	G	–	G	T	417	K	–	K	N
1355	T	G	G	T	452	L	R	R	N
1433	C	C	A	C	478	T	T	K	L
1450	G	C	G	A	484	E	Q	E	T
1501	A	A	A	T	501	N	N	N	K
1841	A	G	G	G	614	D	G	G	Y
2042	C	?	G	C	681	R	?	R	G
2102	C	?	C	T	701	A	?	A	P
2202	A	?	A	T	734	T	?	T	V
2848	G	G	A	G	950	D	D	N	D
3183	C	C	T	C	1061	V	V	V	V
3213	A	T	A	A	1071	Q	H	Q	Q

Table 9 Positions of substitution of nucleotide and respective amino acid change in isolates sequence coding for nucleocapsid protein as compared to reference sequence

Nucleotide substitution					Amino acid substitution				
Position	R	B1	B2	B3	Position	R	B1	B2	B3
53	G	T	G	G	18	G	G	V	G
188	A	G	A	A	63	D	G	D	D
608	G	T	?	G	203	R	M	?	R
614	C	C	?	T	205	T	T	?	I
1129	G	T	-	G	377	D	Y	-	D

Table 10 Position of variation in nucleotide sequences coding for spike (S), envelope (E), membrane (M) and nucleocapsid (N) protein

Position →	3177	22–228/47–228	1–22	241	759	1084–1260
Sequence →	S	E	M	M	N	N
MZ157012.1	C	No deletion	No Deletion	T	G	Deletion
MZ157011.1	C	Deletion(22)	Deletion	G	G	Deletion
MZ157010.1	C	No deletion	No Deletion	G	A	Deletion
MZ157009.1	C	Deletion(47)	Deletion	G	G	Deletion
MZ157008.1	C	No deletion	No Deletion	G	G	Deletion
MZ157007.1	C	Deletion(47)	Deletion	G	G	Deletion
MZ157005.1	C	Deletion(47)	Deletion	G	G	Deletion
MZ208926.1	T	No deletion	No Deletion	G	G	No deletion

Acknowledgements The authors thank University Grants Commission (UGC), Government of India.

Author contributions JS: conceptualization, experimental analysis, writing original draft. ST: experimental analysis, writing original draft. NY: writing-review and editing. YK: review and editing. NS: conceptualization (supporting), writing-original draft (supporting), writing-review an editing.

Data availability All the data used in the study are taken from publically available database of NCBI. All the details of genomic sequences and their accession numbers are provided in material and methods as well as wherever cited.

Declarations

Conflict of interest The authors state that there are no conflicts of interest to disclose.

References

Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T.M., Raj, V.S., Zaki, A.M., Fouchier, R.A.: Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. Mbio (2012). https://doi.org/10.1128/mBio.00473-12

Böttcher, L., D'Orsogna, M.R., Chou, T.: Using excess deaths and testing statistics to determine COVID-19 mortalities. Eur. J. Epidemiol. **36**(5), 545–558 (2021)

Chatterjee, S.: Understanding the nature of variations in structural sequences coding for coronavirus spike, envelope, membrane and nucleocapsid proteins of SARS-CoV-2. SSRN J. (2020). https://doi.org/10.2139/ssrn.3562504

Cui, J., Li, F., Shi, Z.L.: Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. **17**(3), 181–192 (2019)

De Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Ziebuhr, J.: Commentary: Middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. J. Virol. **87**(14), 7790–7792 (2013)

Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.R., Becker, S., Berger, A.: Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. **348**(20), 1967–1976 (2003)

Edgar, R.C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5), 1792–1797 (2004)

Fehr, A.R., Perlman, S.: Coronaviruses: an overview of their replication and pathogenesis. Coronaviruses **1282**, 1–23 (2015)

Gómez, C.E., Perdiguero, B., Esteban, M.: Emerging sars-cov-2 variants and impact in global vaccination programs against sars-cov-2/covid-19. Vaccines **9**(3), 243 (2021)

Hamre, D., Procknow, J.J.: A new virus isolated from the human respiratory tract. Proc. Soc. Exp. Biol. Med. **121**(1), 190–193 (1966)

He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., Li, X.: Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochem. Biophys. Res. Commun. **316**(2), 476–483 (2004)

Hui, D.S., Azhar, E.I., Madani, T.A., Ntoumi, F., Kock, R., Dar, O., Petersen, E.: The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. Int. J. Infect. Dis. **91**, 264–266 (2020)

Juraszek, J., Rutten, L., Blokland, S., Bouchier, P., Voorzaat, R., Ritschel, T., Bakkers, M.J.G., Renault, L.L.R., Langedijk, J.P.M.: Stabilizing the closed SARS-CoV-2 spike trimer. Nat. Commun. **12**(1), 1–8 (2021)

Kopecky-Bromberg, S.A., Martínez-Sobrido, L., Frieman, M., Baric, R.A., Palese, P.: Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J. Virol. **81**(2), 548–557 (2007)

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K.: MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6), 1547–1549 (2018)

Kumar, S., Maurya, V.K., Prasad, A.K., Bhatt, M.L., Saxena, S.K.: Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). Virusdisease **31**(1), 13–21 (2020)

Lau, S.K., Li, K.S., Tsang, A.K., Shek, C.T., Wang, M., Choi, G.K., Yuen, K.Y.: Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. J. Virol. **86**(21), 11906–11918 (2012)

Li, F.: Structure, function, and evolution of coronavirus spike proteins. Ann. Rev. Virol. **3**, 237–261 (2016)

Li, Y., Tenchov, R., Smoot, J., Liu, C., Watkins, S., Zhou, Q.: A comprehensive review of the global efforts on COVID-19 vaccine development. ACS Cent. Sci. **7**(4), 512–533 (2021)

Lokman, S.M., Rasheduzzaman, M., Salauddin, A., Barua, R., Tanzina, A.Y., Rumi, M.H., Hasan, M.M.: Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. Infect. Genet. Evol. **84**, 104389 (2020)

Malik, Y.S., Sircar, S., Bhat, S., Sharun, K., Dhama, K., Dadar, M., Chaicumpa, W.: Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments. Veterinary Quarterly **40**(1), 68–76 (2020)

Masters, P.S.: The molecular biology of coronaviruses. Adv. Virus Res. **66**, 193–292 (2006)

McIntosh, K., Dees, J.H., Becker, W.B., Kapikian, A.Z., Chanock, R.M.: Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. Proc. Natl. Acad. Sci. **57**(4), 933 (1967)

Paraskevis, D., Kostaki, E.G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., Tsiodras, S.: Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect. Genet. Evol. **79**, 104212 (2020)

Sabir, J.S., Lam, T.T.Y., Ahmed, M.M., Li, L., Shen, Y., Abo-Aba, S.E., Guan, Y.: Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science **351**(6268), 81–84 (2016)

Shafique, L., Ihsan, A., Liu, Q.: Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. Pathogens **9**(3), 240 (2020)

Tan, Y.J., Lim, S.G., Hong, W.: Understanding the accessory viral proteins unique to the severe acute respiratory syndrome (SARS) coronavirus. Antiviral Res. **72**(2), 78–88 (2006)

Tao, Y., Shi, M., Chommanard, C., Queen, K., Zhang, J., Markotter, W., Tong, S.: Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. J. Virol. **91**(5), e01953-e2016 (2017)

Tohma, K., Saito, M., Mayta, H., Zimic, M., Lepore, C.J., Ford-Siltz, L.A., Parra, G.I.: Complete genome sequence of a nontypeable GII norovirus detected in Peru. Genome Announc. **6**(10), e00095-e118 (2018)

Tyrrell, D. A., & Myint, S. H. Coronaviruses. In *Medical Microbiology. 4th edition*. University of Texas Medical Branch at Galveston (1996)

Van Der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Berkhout, B.: Identification of a new human coronavirus. Nat. Med. **10**(4), 368–373 (2004)

Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Xu, J.: SARS-CoV infection in a restaurant from palm civet. Emerg. Infect. Dis. **11**(12), 1860 (2005)

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Yuen, K.Y.: Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79**(2), 884–895 (2005)

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Zhang, Y.Z.: A new coronavirus associated with human respiratory disease in China. Nature **579**(7798), 265–269 (2020)

Yinda, C.K., Ghogomu, S.M., Conceição-Neto, N., Beller, L., Deboutte, W., Vanhulle, E., Matthijnssens, J.: Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. Virus Evol. (2018). https://doi.org/10.1093/ve/vey008

Yin, W., Xu, Y., Xu, P., Cao, X., Wu, C., Gu, C., He, X., Wang, X., Huang, S., Yuan, Q., Wu, K.: Structures of the OmicronSpike trimer with ACE2 and an anti-Omicron antibody. Science **375**(6584):1048–1053 (2022)

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Niu, P.: A novel coronavirus from patients with pneumonia in China, 2019. New Engl. J. Med. **382**(8), 727–733 (2020)