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

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

Genomics 112 (2020) 4993–5004

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Molecular phylogeny and missense mutations at envelope proteins across coronaviruses

Sk. Sarif Hassan[a,*], Pabitra Pal Choudhury[b], Bidyut Roy[c]

[a] Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India
[b] Applied Statistics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India
[c] Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, West Bengal, India

ARTICLE INFO

Keywords:
Envelope protein
SARS-CoV-2
COVID-19
Viroporin
Amino acid conservation
Phylogeny

ABSTRACT

Envelope (E) protein is one of the structural viroporins (76–109 amino acids long) present in the coronavirus. Sixteen sequentially different *E*-proteins were observed from a total of 4917 available complete SARS-CoV-2 genomes as on 18th June 2020 in the NCBI database. The missense mutations over the envelope protein across various coronaviruses of the β-genus were analyzed to know the immediate parental origin of the envelope protein of SARS-CoV-2. The evolutionary origin is also endorsed by the phylogenetic analysis of the envelope proteins comparing sequence homology as well as amino acid conservations.

1. Introduction

A novel coronavirus has been causing the ongoing pandemic which is certainly life threatening as our world is experiencing since December 2019 [1]. Coronaviruses (CoV), containing positive-sense RNA as genetic material, cause primarily respiratory infections in humans and a broad range of animals. Recently several new human coronaviruses, including severe acute respiratory syndrome coronavirus (SARS-CoV), MERS-CoV and SARS-CoV-2, were identified, which attract scientists in comprehensive understanding of viruses and identification of antiviral targets for development of therapeutic treatments. A CoV contains several proteins (structural, non-structural, accessory, etc.) among which two major structural proteins of the coronaviruses (CoVs) are spike (S) and membrane (M) glycoproteins [2]. Every Coronavirus of the $\beta - genus$ does contain an envelope (E) protein, containing 75 to 84 amino acids, which plays essential roles in virus assembly, budding, morphogenesis, entry in the host cell and regulation of other cellular functions [3]. This E protein is an integral membrane protein mainly found in the ERGIC (Endoplasmic Reticulum-Golgi Intermediate Compartment) of cells transfected with a plasmid encoding E protein or infected with SARS-CoV [4]. Envelope protein of SARS-CoV-2 is 75 amino acids long, and it possesses three important domains viz. (N)-terminus containing 7–9 hydrophilic region, transmembrane domain (TMD) containing 29 amino acid residues with a high leucine/isoleucine/valine content (hydrophobic region) and (C)-terminus with hydrophilic region (Fig. 1) [5].

The envelope (E) protein of the coronavirus (CoV) of the β-genus famiy forms ion channels [6]. The transmembrane domain (TMD) of the E protein is responsible for the observed ion channel activity which may attenuate the infectivity. Missense mutations in the E protein which inhibited ion channel activity engendered attenuation [7,8]. It is reported that TMD forms stable pentamers and is confirmed by the molecular simulation and in vitro oligomerization [9]. It is reported that mutation of the hydrophobic amino acid residues in the TMD of the E protein with charged amino acids significantly alter the migrating properties of the E protein [3]. Analysis by Y. Liao et al. (2006) established that the TMD is essential for the membrane permeabilizing activity of the protein and also delineates that any missense mutations in the TMD of the E protein disrupt the function of the protein [3]. It is found that the envelope protein of SARS-CoV as well as SARS-CoV-2 contains three cysteine residues at positions 40, 43, and 44 respectively [10]. The first and third cysteine residues, at amino acid positions 40 and 44, respectively, were previously reported to play roles in oligomerization of the E protein [11]. Furthermore, from bio-chemical characterization it was learned that it undergoes translational modification by palmitoylation on all three cysteine residues [12]. Again, it may be noted from mutagenesis studies that the transmembrane domain is responsible for the membrane permeabilizing activity of the SARS-CoV E protein [13]. The (C)-terminal domain of envelope protein in SARS-CoV-2 binds to human PALS1, a tight junction-associated protein, which is essential for the establishment and maintenance of epithelial polarity in mammals [14].

* Corresponding author.
E-mail addresses: sarimif@gmail.com (S.S. Hassan), broy@isical.ac.in (B. Roy).

https://doi.org/10.1016/j.ygeno.2020.09.014
Received 3 July 2020; Accepted 3 September 2020
Available online 12 September 2020
0888-7543/ © 2020 Elsevier Inc. All rights reserved.

Fig. 1. Domains of the envelope protein of β-CoVs.

Table 1
Envelope protein of different host-CoVs.

Host	Total	Distinct	% of Variability of the E protein
Bat	79	25	31.646%
Camel	269	9	3.346%
Cat	42	17	40.476%
Cattle	22	2	9.090%
Pangolin	1	1	0%
Chimpanzee	1	1	0%
SARS-CoV2	4917	19	0.3864%

Almost all the proteins embedded in the SARS-CoV-2 are being mutating as evidenced over the past few months [15–17]. It is hard to infer whether the mutations in E protein infect and sicken people deferentially due to COVID-19. In order to comprehend the effect of mutation over various proteins, one needs to accumulate all the mutations over the proteins from a large number of SARS-CoV-2 genomes available worldwide. On the other hand, most unsettled, controversial issue is the source/proximal origin of the SARS-CoV-2. Pattern of the genetic differences and motifs of the proteins present in SARS-CoV-2 distinguish it from any other known coronavirus E protein [18,19]. Zhang, Wu et al. (2020) showed that the natural reservoirs of SARS-CoV-2 are Bat and Pangolin [20]. Recently, based on genomic and protein sequences from few coronoviruses of different hosts including human, it was reported that Pangolin may not be intermediate host for coronavirus transmission from bat to human [21]. Presently, we wish to transact the transmission issue by analyzing mutations in one of most conserved proteins (E protein) over the SARS-CoV-2 and other host-CoV genomes.

In this study, using protein sequences from a large number of coronaviruses from different hosts including human, we analyzed the phylogenetic relationship among them. A comparative investigation of the envelope (E) protein of CoVs of the β-genus family including SARS-CoV-2 from the perspective of missense mutations as well as molecular organization of the amino acids in the envelope proteins has been performed in order to gain an insight and discover the intermediate hosts.

2. Materials and methods

This study considered all the envelope proteins of coronaviruses from different hosts viz. Bat, Camel, Cat, Cattle, Pangolin, Chimpanzee and human SARS-CoV-2. In the Table 1, total number of available CoV genomes of respective hosts as well as distinct numbers of envelope proteins in them are presented. (See Table 2.)

From the NCBI virus database, all the protein sequences of 4917 complete SARS-CoV-2 genomes as on date 18th June 2020 as well as other host CoV genomes were fetched. Then the amino acid sequences of envelope protein of all the CoVs from different hosts viz. Bat, Cat, Cattle, Pangolin, Chimpanzee, Human, are exported in fasta format using file management operations through MATLAB ver. R2020a [22]. The following is the complete list of seventy-four distinct envelope (E) proteins from different host CoVs and their respective protein IDs (Table-2).

Amino Acid Conservation Shannon Entropy: For each E protein, Shannon entropy of amino acid conservation over the amino acid sequence of E protein is computed using the following formula [23]:

For a given amino acid sequence of E protein of length l, the conservation of amino acids is calculated as follows:

Table 2
List of distinct envelope (E) proteins from different host CoVs and their respective protein ID.

Protein ID	Host	Protein ID	Host	Protein ID	Host
AIA62357	Bat-CoV	AIA62302	Bat-CoV	ADO39821	Feline-CoV
AIA62348	Bat-CoV	AVP78044	Bat-CoV	ACT10858	Feline-CoV
AHY61342	Bat-CoV	AVI15004	Bovine-CoV	ACT10869	Feline-CoV
ASL68958	Bat-CoV	AVZ61113	Bovine-CoV	ACT10909	Feline-CoV
ASL68947	Bat-CoV	ALA50082	Camel-CoV	ACT10941	Feline-CoV
ATQ39391	Bat-CoV	QCI31474	Camel-CoV	ACT10974	Feline-CoV
AUM60029	Bat-CoV	QBM11741	Camel-CoV	ACT10920	Feline-CoV
QDF43841	Bat-CoV	ASU89926	Camel-CoV	AWW13513	Chimpanzee-CoV
YP_009072442	Bat-CoV	ASU90554	Camel-CoV	QIG55947	Pangolin CoV
YP_009273007	Bat-CoV	ANI69894	Camel-CoV	QHZ00381	Human-SARS-CoV-2
ABD75324	Bat-CoV	ALA49346	Camel-CoV	QKI36855	Human-SARS-CoV-2
AGC74167	Bat-CoV	ALA49390	Camel-CoV	QKG87268	Human-SARS-CoV-2
AKZ19089	Bat-CoV	ASU90334	Camel-CoV	QKE45838	Human-SARS-CoV-2
ADK66843	Bat-CoV	QDM36990	Feline-CoV	QJR88103	Human-SARS-CoV-2
QDF43816	Bat-CoV	AYF53097	Feline-CoV	YP_009724392	Human-SARS-CoV-2
ATO98160	Bat-CoV	AXE71624	Feline-CoV	QKI36831	Human-SARS-CoV-2
ATO98184	Bat-CoV	ASU62492	Feline-CoV	QJS53352	Human-SARS-CoV-2
QDF43821	Bat-CoV	ASU62503	Feline-CoV	QJA42107	Human-SARS-CoV-2
ATO98135	Bat-CoV	AUG98123	Feline-CoV	QJQ84210	Human-SARS-CoV-2
AHX37560	Bat-CoV	AMD11134	Feline-CoV	QJR89447	Human-SARS-CoV-2
AIA62280	Bat-CoV	AGT52084	Feline-CoV	QJI54124	Human-SARS-CoV-2
ABD75313	Bat-CoV	AEK25514	Feline-CoV	QKU31207	Human-SARS-CoV-2
AIA62312	Bat-CoV	AEK25525	Feline-CoV	QKU37035	Human-SARS-CoV-2
				QKV07065	Human-SARS-CoV-2
				QKU32371	Human-SARS-CoV-2
				QKU28584	Human-SARS-CoV-2
				QKU52835	Human-SARS-CoV-2
				QKV06741	Human-SARS-CoV-2

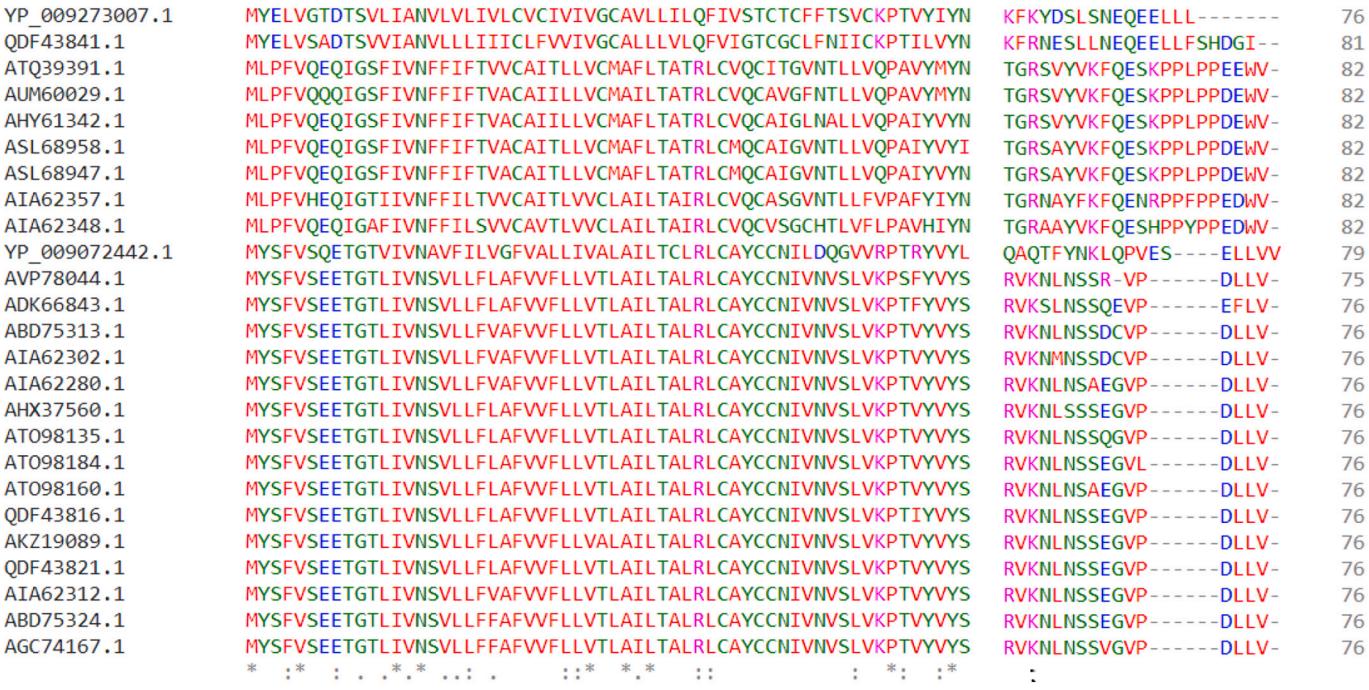

Fig. 2. Sequence alignment of the E protein of Bat CoV.

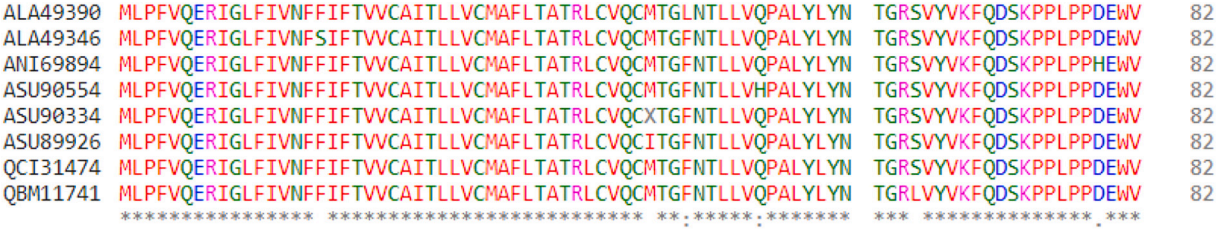

Fig. 3. Sequence alignment of the E protein of Camel CoV.

Fig. 4. Sequence alignment of the E protein of Cat CoV.

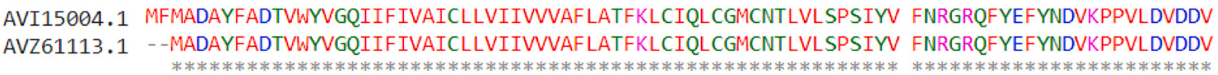

Fig. 5. Sequence alignment of the E protein of Cattle CoV.

Fig. 6. Sequence alignment of the E protein of SARS-CoV-2.

Table 3
Envelope proteins across different host CoVs.

Host-CoVs	E protein sequence (N to C terminal of protein)	Length
Human SARS-CoV2	MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV	75
Chimpanzee-CoV	MFMADAYLADTVWYVGQIIFIVAICLLVTIVVVAFLATFKLCIQLCGMCNTLVLSPSIYVFNRGRQFYEFYNDIKPPVLDVDDV	84
Pangolin-CoV	MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV	75
Feline or Cat-CoV	MMFPRAFTIIDDHGMVVSVFFWLLLIIILILFSIALLNVIKLCMVCCNLGKTIIVLPARHAYDAYKNFMHIKAYDPDEAFLV	82
Camel-CoV	MLPFVQERIGLFIVNFFIFTVVCAITLLVCMAFLTATRLCVQCITGFNTLLVQPALYLYNTGRSVYVKFQDSKPPLPPDEWV	82
Cattle or Bovine-CoV	MFMADAYFADTVWYVGQIIFIVAICLLVIIVVVAFLATFKLCIQLCGMCNTLVLSPSIYVFNRGRQFYEFYNDVKPPVLDVDDV	84
Bat-CoV	MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVYSRVKNLNSSEGVPDLLV	76

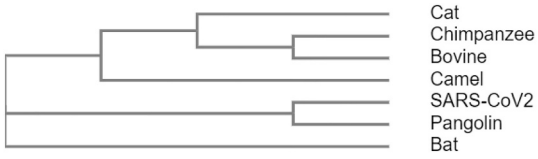

Fig. 7. Sequence homology based phylogeny of the envelope protein of different host-CoVs.

Table 4
Amino acid residues and their respective color and property used in Fig. 2.

Residue	Color	Property
A,V,F,P,M,I,L and W	RED	hydrophobic (incl.aromatic −Y)
D and E	BLUE	Acidic
Rand K	MAGENTA	Basic - H
S,T,Y,H,C,N,G and Q	GREEN	Hydroxyl + sulfhydryl + amine + G

$$SE = -\sum_{i=1}^{20} p_{s_i} log_{20}(p_{s_i})$$

where $p_{s_i} = \frac{k_i}{l}$; k_i represents the number of occurrences of an amino acid s_i in the given sequence.

3. Results

3.1. Mutations in the E protein of CoVs

It is noted that the envelope (E) protein of the CoVs of Pangolin and Chimpanzee are found to be 100% conserved as presented in Table 1 and consequently no mutation was found over there. In order to detect the missense mutations, we have made the multiple sequence alignment of the E protein sequences (Table-3) using the *Clustal-Omega server* [24,25]. In the following Table 4, description of the amino acid residues and their respective color and property are mentioned. These notations are also used in Fig. 2, 3, 4, 5 and 6

It may be noted that an * (asterisk) indicates positions which have a single, fully conserved residue. Colon (:) indicates conservation between groups of strong similarity. Period (.) indicates conservation between groups of weak similarity [25].

3.1.1. Missense mutations of the E protein of bat CoV

Among 79 available complete CoV genomes of Bat, twenty-five unique sequences possess various mutations in the three domains of the E protein as presented in the Fig. 2.

The missense mutations over the E proteins of Bat-CoV with the respective domains are described in the Table 5. There exists variety of mutations in the envelope proteins of Bat-CoV.

The most of the frame-shift mutations occurred in the C-terminal domain of the protein. There are also mutations in other two domains viz. TMD and N-terminal. Clearly, changes in the R-group property from Hydrophobic/Acidic to Hydrophilic/Basic of the amino acid residues of the three domains of the E protein may affect the function of the envelope protein. It is to be noted that envelope protein sequence of the protein QDF43841, YP_009273007, AIA62348 and ATQ39391 possess mutations at the cysteine residue such as C40V, C40I, C44V, C44I respectively. E protein sequence of the proteins AIA62357, ASL68958, AHY61342, AUM60029 contain the mutation C44A. These missense mutations at the cysteine residue may affect virus growth, release, entry, protein transport, and stability [26]. There is an important mutation V25C which is found in the TMD of E protein in the genome YP_009273007, which might stop the ion channel activity and led to in vivo attenuation. The TMD of the E protein for Bat CoV genomes

Table 5
Missense mutations in the envelope protein of the Bat CoV.

Protein ID	Mutation	Domain
ATQ39391, AUM60029, AHY61342, ASL68958, ASL68947, AIA62357, AIA62348	Y2L	N-terminal
YP_009072442, AUM60029	E7Q	N-terminal
QDF43841	E7A	N-terminal
YP_009273007	E7T	N-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	E8Q	N-terminal
QDF43841, YP_009273007	E8D	N-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	T9I	N-terminal
AIA62348	T11A	N-terminal
QDF43841, YP_009273007	T11V	N-terminal
AIA62348	F20S	TMD
ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	F20T	TMD
YP_009072442	A22G	TMD
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391, QDF43841, YP_009273007	F23C	TMD
YP_009273007	V25C	TMD
AIA62348, ASL68947, AIA62357, ASL68958, ATQ39391	F26T	TMD
AKZ19089, YP_009072442,	T30A	TMD
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	T30C	TMD
QDF43841, YP_009273007	T30G	TMD
QDF43841, YP_009273007	L31C	TMD
QDF43841, YP_009273007	T35L	TMD
YP_009072442	A36C	TMD
ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	L37T	C-terminal
QDF43841	C40V	C-terminal
YP_009273007	C40I	C-terminal
ASL68947, ASL68958	A41M	C-terminal
AIA62348	C44V	C-terminal
AIA62357, ASL68958, AHY61342, AUM60029	C44A	C-terminal
ATQ39391	C44I	C-terminal
AIA62357, ASL68958, AHY61342	N45I	C-terminal
AUM60029	N45V	C-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	I46G	C-terminal
YP_009072442, AUM60029	I46C	C-terminal
AIA62348	V47C	C-terminal
YP_009072442	N48D	C-terminal
YP_009072442, AUM60029	N48F	C-terminal
YP_009072442	V49Q	C-terminal
AIA62348, ASL68947, AIA62357, ASL68958	V49T	C-terminal
QDF43841	V49N	C-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	S50L	C-terminal
QDF43841	S50I	C-terminal
QDF43841, YP_009273007	V52C	C-terminal
AIA62348	K53L	C-terminal
AIA62357	K53V	C-terminal
YP_009072442	V56R	C-terminal
QDF43841	Y57L	C-terminal
YP_009072442	S60L	C-terminal
ASL68958	S60I	C-terminal
YP_009072442	R61Q	C-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	R61T	C-terminal
AIA62348, ASL68947, AIA62357, ASL68958, AHY61342, AUM60029, ATQ39391	V62G	C-terminal
YP_009072442	K63Q	C-terminal
YP_009072442	N64A	C-terminal
YP_009273007	L65D	C-terminal
QDF43841	L65E	C-terminal
AIA62348, ASL68947, ASL68958, AHY61342, AUM60029, ATQ39391	S67V	C-terminal
AIA62357	S67F	C-terminal
QDF43841, YP_009273007	S67L	C-terminal
ATO98160, AIA62280	S68A	C-terminal
YP_009072442, AIA62348, ASL68947, AIA62357, ASL68958,	S68K	C-terminal
AHY61342, AUM60029, ATQ39391		
QDF43841, YP_009273007	S68L	C-terminal
AGC74167	E69V	C-terminal
ATO98135	E69Q	C-terminal
AVP78044	E69R	C-terminal
YP_009072442	E69L	C-terminal
AIA62348, ASL68947, ASL68958, AHY61342, AUM60029, ATQ39391	E69F	C-terminal
QDF43841, YP_009273007	E69N	C-terminal
QDF43841, YP_009273007	G70E	C-terminal
AIA62348, ASL68947, ASL68958, AHY61342, AUM60029, ATQ39391	V71E	C-terminal
QDF43841, YP_009273007	V71Q	C-terminal
AIA62348, ASL68947, ASL68958, AHY61342, AUM60029, ATQ39391	P72S	C-terminal
AIA62357	P72N	C-terminal
QDF43841, YP_009273007	P72E	C-terminal
AIA62348, AIA62357	L73D	C-terminal
ASL68947, ASL68958, AHY61342, AUM60029, ATQ39391	L73E	C-terminal
QDF43841	L73G	C-terminal

Table 6
Missense mutation of the envelope protein of the Cat CoV.

Protein ID	Missense Mutations	Domain
AXE71624	K51N, L81S	C-terminal
AEK25514	W22L	TMD
ADO39821	N48D	C-terminal
ACT10869	V19G, R59C	TMD, C-terminal
ACT10909	L81M	C-terminal
ACT10941	V19G	TMD

AIA62348, ASL68947, AIA62357, ASL68958, ATQ39391 contains a mutation F26T and it may also cause stopping the ion channel activity [27–29]. Mutations in the motif"DFLV" might also affect its binding to the PALS1 protein and accordingly may influence replication and/or infectivity of the virus [30].

3.1.2. missense mutations of the E protein from camel CoV

Among 269 available complete CoV genomes of Camel, only 9 of them possess mutations as presented in the Fig. 3.

Most of the envelope proteins of the Camel CoV do not contain any mutations, only nine E proteins among the 269 Camel-CoV genomes possess few mutations. The envelope (E) protein possesses only three missense mutations viz. F17S in TMD of the protein ALA49346, S64L and D79H in C-terminal of the proteins QBM11741 and ANI69894 respectively. It is to be noted that the motif is *"DEWV"* in the C-terminal end is absolutely conserved within the host-CoV except in ANI69894.

3.1.3. Missense mutations of the E protein of cat CoV

The highest amount (40.476%) of variability among the E proteins is found in the case of Cat-CoV although the mutations over the sequences is limited to seven different positions with 8.536% over the three domains as presented in the Fig. 4.

These missense mutations over TMD and C-terminal domains of the envelope protein of Cat CoV are shown in Table 6. It is worth noting that though the amount of variability of E proteins is too high comparatively, but the N-terminal of each E protein is absolutely conserved.

The mutations in the TMD and C-terminal in the E protein across the Cat CoV would possibly affect the functions of the protein. The mutations in the TMD of the E protein would impact on ion channel activity

Table 7
Protein ID and respective location of mutation of the E proteins over SARS-CoV-2.

Protein ID and Respective Geo-location	Mutations	Domain	R-Group
QKO24093 (USA: San Diego, California)	E8K	N-terminal	Acidic to Basic
QKU52835 (USA: WA)	E7Q	N-terminal	Acidic to Basic
QKN20885 (USA), QJQ84210 (USA: New Orleans, LA)	F26L	TMD	Hydrophobic to Hydrophobic
QKI36831 (China: Guangzhou)	D72Y	C-terminal	Hydrophilic to Hydrophobic
QKI36855 (China: Guangzhou)	S68C	C-terminal	Hydrophilic to Hydrophobic
QKG87268, QKG88576 (USA: Massachusetts)	S68F	C-terminal	Hydrophilic to Hydrophobic
QKE45838 (USA:CA), QKE45886 (USA:CA)	P71L	C-terminal	Hydrophobic to Hydrophobic
QKE45898 (USA:CA), QKE45910 (USA:CA)	P71L	C-terminal	Hydrophobic to Hydrophobic
QJE38284 (USA:CA), QIU81527 (USA:WA), QKV06741 (USA: WA)	P71L	C-terminal	Hydrophobic to Hydrophobic
QKU32371 (USA: CA)	P71L	C-terminal	Hydrophobic to Hydrophobic
QJS53352 (Greece: Athens)	L39M	TMD	Hydrophobic to Hydrophobic
QJR88103 (Australia: Victoria)	L73F	C-terminal	Hydrophobic to Hydrophobic
QJA42107 (USA: VA)	A36V	TMD	Hydrophobic to Hydrophobic
QHZ00381 (South Korea)	L37H	TMD	Hydrophobic to Hydrophilic
QKU31207 (USA: CA)	T9I	TMD	Hydrophilic to Hydrophobic
QKU37035 (Saudi Arabia: Jeddah)	L19F	TMD	Hydrophobic to Hydrophobic
QKV07065 (USA: WA)	S55F	C-terminal	Hydrophilic to Hydrophobic
QKU28584 (USA: WA)	A41S	C-terminal	Hydrophobic Hydrophilic

Table 8
Amino acid counts over the envelope proteins over the different host CoVs.

Host-CoVs	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
SARS-CoV-2	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	13
Chimpanzee-CoV	6	2	3	6	4	3	1	3	0	8	9	2	3	7	3	2	4	1	5	12
Pangolin-CoV	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	13
Cat-CoV	7	2	3	5	3	0	1	2	3	11	11	4	5	7	3	2	2	1	3	7
Camel-CoV	4	3	3	2	4	4	2	3	0	5	11	2	2	8	6	2	7	1	3	10
Bovine-CoV	6	2	3	6	4	3	1	3	0	8	8	2	3	8	3	2	3	1	5	13
Bat-CoV	4	2	5	1	3	0	3	2	0	3	14	2	1	4	2	7	5	0	4	14

of the envelope protein in the Cat CoV.

3.1.4. Missense mutations of the E protein of cattle CoV

Among 22 available complete CoV genomes of Cattle, only two of them had variations due with frame-shifts as shown in Fig. 5.

The envelope proteins of the cattle CoV are highly conserved as shown in Fig. 5. It is noted that there are two frame-shifts in the N-terminal sequence.

3.1.5. Missense mutations of the E protein of human SARS-CoV-2

The E protein is present over all the available 4917 SARS-CoV-2 genomes as on 18th June 2020 in the NCBI database. There are only sixteen distinct E proteins over the 4917 available SARS-CoV-2 genomes. The mutations of the E proteins (presented in Table 7) are determined through the multiple sequence alignment as shown in Fig. 6. It is to be noted that the mutations in the C-terminal domain of E protein from SARS-CoV to SARS-CoV-2 is already described in the unpublished article [31].

Most of the missense mutation occurred in the C-terminal. The E protein of QKN20885 (USA) and QJQ84210 (USA: New Orleans, LA) have a mutation at F26L in the TMD of the E protein. This particular mutation in the TMD terminate the ion channel activity and may led to in vivo attenuation. The E protein of QJS53352 (Greece: Athens), QJA42107 (USA: VA) and QHZ00381 (South Korea) contain mutations L39M, A36V and L37H respectively in the TMD of the E protein. These mutations in the TMD terminates the ion channel activity and led to in vivo attenuation. Several mutations have been found in the C-terminal of E proteins of SARS-CoV-2 and some of these mutations lead to non-synonymous R-group properties of amino acids, which might affect interaction of E protein with host proteins.

From the mutation data of different host-CoVs, it is concluded that the mutations over the E proteins of the SARS-CoV-2, Pangolin CoVs and Bat CoVs are almost similar in nature. It is to be mentioned that the SARS-CoV-2 E protein is much closer to that of the Pangolin-CoV, from the variability perspective. This closeness is also supported by sequence based homology. Here we illustrate the phylogenetic relationship among the E proteins (Table 3) across different CoVs based on sequence homology, as shown in Fig. 7.

From the phylogeny Fig. 7, it is derived that among all E proteins of all the host CoVs, the E proteins of Pangolin-CoV and SARS-CoV-2 are very much close to each other. In order to get a more intensive phylogenetic relationship among the E proteins of the host CoVs, we further did amino acid frequency based phylogeny. We determined the amino acid frequencies for each of the common E proteins from each of the host CoV as tabulated in Table 8. Based on the frequency vector for each E protein, pairwise euclidean distance has been calculated and consequently the phylogeny is derived (Fig. 8).

From the amino acid frequency based phylogeny, it is reconfirmed that the E protein of BatCoV and SARS-CoV-2 are co-evolved from the same origin. Further it is also confirmed that the E protein of Pangolin-CoV and SARS-CoV-2 are very much conserved from the point of amino acid conservation in the protein. It is worth mentioning that the

Table 9
Frequency of amino acids over the envelope proteins across the seven different host-CoVs.

Name	Host	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
ASL68947	*Bat CoV*	7	2	3	1	4	5	3	3	0	6	8	2	3	7	6	3	6	1	3	9
ASU90554	*Camel CoV*	4	3	3	2	4	3	2	3	1	4	11	2	3	8	6	2	7	1	3	10
ASL68958	*Bat CoV*	7	2	2	1	4	5	3	3	0	7	8	2	3	7	6	3	6	1	3	9
ANI69894	*Camel CoV*	4	3	3	1	4	4	2	3	1	4	11	2	3	8	6	2	7	1	3	10
AXE71624	*Feline CoV*	7	2	4	4	3	1	1	2	2	11	10	3	5	7	3	3	3	1	3	7
ALA49346	*Camel CoV*	4	3	3	2	4	4	2	3	0	4	11	2	3	7	6	3	7	1	3	10
AGT52084	*Feline CoV*	7	2	3	4	3	1	1	2	2	11	11	4	5	6	3	3	3	1	3	7
AUM60029	*Bat CoV*	6	2	3	1	4	6	2	3	0	6	8	2	3	7	6	3	5	1	3	11
ACT10909	*Feline CoV*	7	3	3	4	3	1	1	2	2	10	11	3	6	6	3	2	3	1	3	8
AHY61342	*Bat CoV*	7	2	3	1	4	5	3	3	0	7	9	2	2	7	6	3	4	1	3	10
AIA62348	*Bat CoV*	7	2	2	1	5	4	3	3	3	6	8	1	1	6	6	3	4	1	3	13
ACT10941	*Feline CoV*	7	2	3	4	3	1	1	3	2	11	12	4	5	6	3	2	3	1	3	6
AYF53097	*Feline CoV*	7	2	3	4	3	1	1	2	2	10	11	4	5	7	3	2	3	1	3	8
QCI31474	*Camel CoV*	4	3	3	2	4	4	2	3	0	4	11	2	3	8	6	2	7	1	3	10
ASU62492	*Feline CoV*	7	2	3	4	3	1	1	2	2	11	11	4	5	7	3	2	3	1	3	7
ACT10974	*Feline CoV*	6	2	3	4	3	1	1	2	2	11	11	4	5	7	3	2	3	1	3	8
ALA49390	*Camel CoV*	4	3	3	2	4	4	2	3	0	4	12	2	3	7	6	2	7	1	3	10
ASU89926	*Camel CoV*	4	3	3	2	4	4	2	3	0	5	11	2	2	8	6	2	7	1	3	10
ACT10869	*Feline CoV*	7	1	3	4	4	1	1	3	2	11	12	4	5	6	3	2	3	1	3	6
AIA62357	*Bat CoV*	6	3	5	1	4	3	3	3	1	8	8	1	1	8	6	1	6	1	3	10
ASU90334	*Camel CoV*	4	3	3	2	4	4	2	3	0	4	11	2	2	8	6	2	7	1	3	10
ASU62503	*Feline CoV*	7	2	3	4	3	1	1	2	2	11	12	4	5	6	3	2	3	1	3	7
ADO39821	*Feline CoV*	7	2	2	5	3	1	1	2	2	11	11	4	5	7	3	2	3	1	3	7
QDM36990	*Feline CoV*	7	2	4	4	3	1	1	2	2	11	11	4	4	7	3	2	2	1	3	8
ACT10858	*Feline CoV*	7	2	4	4	3	1	1	2	2	11	12	4	5	6	3	2	2	1	3	7
AWW13513	*Chimpanzee CoV*	6	2	3	6	4	3	1	3	0	8	9	2	3	7	3	2	4	1	5	12
ACT10920	*Feline CoV*	7	2	4	4	3	1	1	2	2	11	11	3	5	7	2	2	2	1	3	7
ATQ39391	*Bat CoV*	4	2	3	0	4	5	4	3	0	5	8	2	3	7	6	3	7	1	3	12
QBM11741	*Camel CoV*	4	3	3	2	4	4	2	3	0	4	12	2	3	8	6	1	7	1	3	10
AVI15004	*Bovine CoV*	6	2	3	6	4	3	1	3	0	8	8	2	3	8	3	2	3	1	5	13
AUG98123	*Feline CoV*	7	3	3	4	3	0	1	2	2	12	11	4	5	7	3	2	3	1	3	6
AMD11134	*Feline CoV*	7	2	3	5	3	0	1	2	3	11	11	4	5	7	3	2	2	1	3	7
AEK25525	*Feline CoV*	7	2	3	4	3	0	1	2	2	11	11	5	5	7	3	2	3	1	3	7
AVZ61113	*Bovine CoV*	6	2	3	6	4	3	1	3	0	8	8	2	2	7	3	2	3	1	5	13
ALA50082	*Camel CoV*	6	2	3	6	4	3	1	3	0	9	8	2	2	8	3	2	3	1	5	13
AEK25514	*Feline CoV*	7	2	3	4	3	1	1	2	2	11	12	4	5	7	3	2	3	0	3	7
YP_009072442	*Bat CoV*	6	3	3	1	4	5	3	3	0	5	12	1	1	4	2	3	5	0	5	13
QDF43841	*Bat CoV*	3	1	5	2	5	2	5	4	1	10	15	2	1	5	1	4	3	0	2	10
YP_009273007	*Bat CoV*	2	0	3	2	6	2	4	2	0	7	12	3	1	4	1	5	6	0	4	12
QHZ00381	*SARS-CoV-2*	4	3	5	1	3	0	2	1	1	3	13	2	1	5	2	8	4	0	4	13
ATO98135	*Bat CoV*	4	2	5	1	3	1	2	2	0	3	14	2	1	4	2	7	5	0	4	14
AIA62302	*Bat CoV*	4	2	5	2	4	0	2	1	0	3	12	2	2	4	2	7	5	0	4	15
QJS53352	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	13	2	2	5	2	8	4	0	4	13
QDF43816	*Bat CoV*	4	2	5	1	3	0	3	2	0	4	14	2	1	4	2	7	5	0	4	13
ABD75324	*Bat CoV*	4	2	5	1	3	0	3	2	0	3	13	2	1	5	2	7	5	0	4	14
ADK66843	*Bat CoV*	4	2	4	0	3	1	4	1	0	3	13	2	1	6	2	8	5	0	4	13
ATO98160	*Bat CoV*	5	2	5	1	3	0	3	2	0	3	14	2	1	4	2	6	5	0	4	14
AIA62280	*Bat CoV*	5	2	5	1	3	0	3	2	0	3	13	2	1	4	2	6	5	0	4	15
QJR88103	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	13	2	1	6	2	8	4	0	4	13
AKZ19089	*Bat CoV*	5	2	5	1	3	0	3	2	0	3	14	2	1	4	2	7	4	0	4	14
QDF43821	*Bat CoV*	4	2	5	1	3	0	3	2	0	3	14	2	1	4	2	7	5	0	4	14
QKI36855	*SARS-CoV-2*	4	3	5	1	4	0	2	1	0	3	14	2	1	5	2	7	4	0	4	13
AIA62312	*Bat CoV*	4	2	5	1	3	0	3	2	0	3	13	2	1	4	2	7	5	0	4	15
ABD75313	*Bat CoV*	4	2	5	2	4	0	2	1	0	3	13	2	1	4	2	7	5	0	4	15
QKG87268	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	14	2	1	6	2	7	4	0	4	13
AHX37560	*Bat CoV*	4	2	4	1	3	0	3	2	0	3	14	2	1	4	2	8	5	0	4	14
AGC74167	*Bat CoV*	4	2	5	1	3	0	2	2	0	3	13	2	1	5	2	7	5	0	4	15
AVP78044	*Bat CoV*	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	13
QIG55947	*Pangolin CoV*	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	13
YP_009724392	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	13
QJR89447	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	12
QJQ84210	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	15	2	1	4	2	8	4	0	4	13
QJA42107	*SARS-CoV-2*	3	3	5	1	3	0	2	1	0	3	14	2	1	5	2	8	4	0	4	14
ATO98184	*Bat CoV*	4	2	5	1	3	0	3	2	0	3	15	2	1	4	1	7	5	0	4	14
QKE45838	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	15	2	1	5	1	8	4	0	4	13
QKI36831	*SARS-CoV-2*	4	3	5	0	3	0	2	1	0	3	14	2	1	5	2	8	4	0	5	13
QJI54124	*SARS-CoV-2*	4	3	5	0	3	0	2	1	0	3	13	2	1	5	2	8	4	0	4	13
QKU31207	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	4	14	2	1	5	2	8	3	0	4	13
QKU37035	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	13	2	1	6	2	8	4	0	4	13
QKV07065	*SARS-CoV-2*	4	3	5	1	3	0	2	1	0	3	14	2	1	6	2	7	4	0	4	13
QKU28584	*SARS-CoV-2*	3	3	5	1	3	0	2	1	0	3	14	2	1	5	2	9	4	0	4	13
QKU52835	*SARS-CoV-2*	4	3	5	1	3	1	1	1	0	3	14	2	1	5	2	8	4	0	4	13

Table 10
Shannon entropy of the amino acid conservation of the E protein of the host CoVs.

Name	Host	SE	Name	Host	SE	Name	Host	SE
ASL68947	Bat CoV	0.933	ACT10858	Feline CoV	0.919	AIA62280	***Bat CoV***	0.851
ASU90554	Camel CoV	0.932	AWW13513	Chimpanzee CoV	0.918	QJR88103	**SARS-CoV2**	0.850
ASL68958	Bat CoV	0.930	ACT10920	Feline CoV	0.916	QKU37035	**SARS-CoV2**	0.850
ANI69894	Camel CoV	0.929	ATQ39391	Bat CoV	0.916	AKZ19089	***Bat CoV***	0.850
AXE71624	Feline CoV	0.928	QBM11741	Camel CoV	0.915	QDF43821	***Bat CoV***	0.850
ALA49346	Camel CoV	0.927	AVI15004	Bovine CoV	0.914	QKI36855	**SARS-CoV2**	0.850
AGT52084	Feline CoV	0.926	AUG98123	Feline CoV	0.912	AIA62312	***Bat CoV***	0.849
AUM60029	Bat CoV	0.926	AMD11134	Feline CoV	0.912	ABD75313	***Bat CoV***	0.848
ACT10909	Feline CoV	0.926	AEK25525	Feline CoV	0.912	QKG87268	**SARS-CoV-2**	0.848
AHY61342	Bat CoV	0.925	AVZ61113	Bovine CoV	0.912	QKV07065	**SARS-CoV-2**	0.848
AIA62348	Bat CoV	0.925	ALA50082	Camel CoV	0.909	AHX37560	***Bat CoV***	0.847
ACT10941	Feline CoV	0.924	AEK25514	Feline CoV	0.908	AGC74167	Bat CoV	0.847
AYF53097	Feline CoV	0.924	YP_009072442	Bat CoV	0.888	AVP78044	Bat CoV	0.846
QCI31474	Camel CoV	0.923	QDF43841	Bat CoV	0.881	QIG55947	***Pangolin CoV***	0.846
ASU62492	Feline CoV	0.922	YP_009273007	***Bat CoV***	0.868	YP_009724392	**SARS-CoV-2**	0.846
ACT10974	Feline CoV	0.922	QHZ00381	**SARS-CoV-2**	0.862	QKU31207	**SARS-CoV-2**	0.846
ALA49390	Camel CoV	0.921	ATO98135	***Bat CoV***	0.858	QJR89447	**SARS-CoV-2**	0.843
ASU89926	Camel CoV	0.921	AIA62302	***Bat CoV***	0.857	QKU28584	**SARS-CoV-2**	0.842
ACT10869	Feline CoV	0.920	QJS53352	**SARS-CoV-2**	0.856	QJQ84210	**SARS-CoV-2**	0.841
AIA62357	Bat CoV	0.920	QDF43816	***Bat CoV***	0.856	QJA42107	**SARS-CoV-2**	0.840
ASU90334	Camel CoV	0.920	ABD75324	Bat CoV	0.855	ATO98184	***Bat CoV***	0.840
ASU62503	Feline CoV	0.920	ADK66843	***Bat CoV***	0.852	QKE45838	**SARS-CoV2**	0.836
ADO39821	Feline CoV	0.920	QKU52835	**SARS-CoV-2**	0.852	QKI36831	**SARS-CoV-2**	0.835
QDM36990	Feline CoV	0.919	ATO98160	***Bat CoV***	0.851	QJI54124	**SARS-CoV-2**	0.824

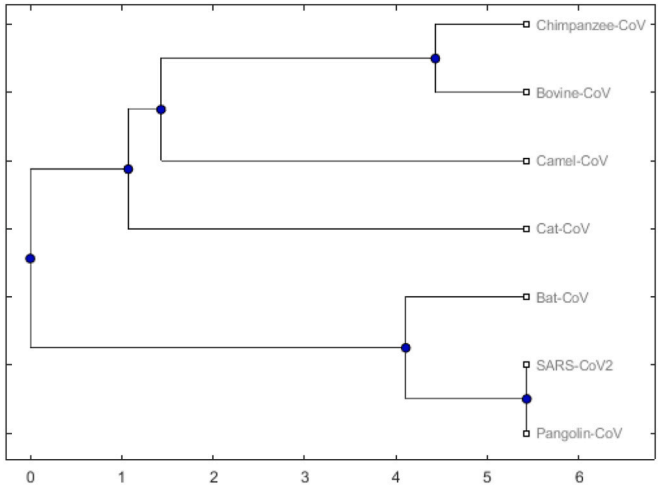

Fig. 8. Phylogenetic relationship among the different host CoVs with respect to the amino acids conservation the envelope protein.

Chimpanzee-CoV and Bovine-CoV contain the most closest E proteins as confirmed from the sequence based homology as well as amino acid conservation.

3.2. Phylogeny of the envelope proteins of host-CoVs

The sequence based homology of 74 distinct E proteins across the different host CoVs are presented in Fig. 9.

The E proteins of the Bat-CoV, Pangolin-CoV and SARS-CoV-2 belong to the left hand side of the cladogram (from root) exclusively as shown in Fig. 9. The other side contains the E proteins of the other host CoVs. It is also observed that all the sixteen different E proteins of SARS-CoV-2 and that of Pangolin belong to a nearby neighbourhood.

In Table 9, for each of the E proteins of the CoVs, frequency of each amino acids is computed, which yields the amino acids conservation based phylogeny (Fig. 10).

From the sequence homology (Fig. 9) it is observed that the E proteins of AIA62312 and ABD75324 of Bat-CoV are very much close. Based on the amino acid conservation over the E protein, the phylogeny (Fig. 10) further showed that the E proteins of QDF43821, ATO98184, AHX37560, AGC74167, AKZ19089, AIA62280, ATO98160, ATO98135 of Bat-CoV are in the same branch with same level of the phylogeny. The phylogeny in Fig. 9 describes that the E proteins of QJR89447 and QJQ84210 of SARS-CoV2 are very close. It is obtained that the E proteins of AVP78044 (Bat-CoV), QIG55947 (Pangolin-CoV) and YP_009724392 (SARS-CoV-2) are close to that of QJR89447 (SARS-CoV-2) from the phylogeny based on amino acid conservations (Fig. 10). The E proteins of QKI36855, QJA42107, QKE45838, QKI36831 and QJI54124 of SARS-CoV2 are in the close proximity to that of the QJQ84210 (SARS-CoV-2) based on amino acid conservations. Again the E proteins of QKU31207 and YP_009724392 of SARS-CoV-2 are found to be near enough based on the homology based phylogeny.

It is observed that almost all the E proteins of SARS-CoV-2 as well as Bat and Pangolin-CoVs do not contain the amino acids tryptophan, glutamine and histidine. The E proteins of all the host CoVs are leucine and valine residues rich as observed in Table 10.

Based on the amino acid frequency vector for each proteins, the Shannon entropy (SE) is computed which is tabulated in Table 10. This SE of the amino acid conservation of the E protein suggests molecular level closeness of the E protein.

From the Table 10, it is quite evident that the conservation of amino acids over the E protein of Bat-CoV is highly diverse as SE value is in the

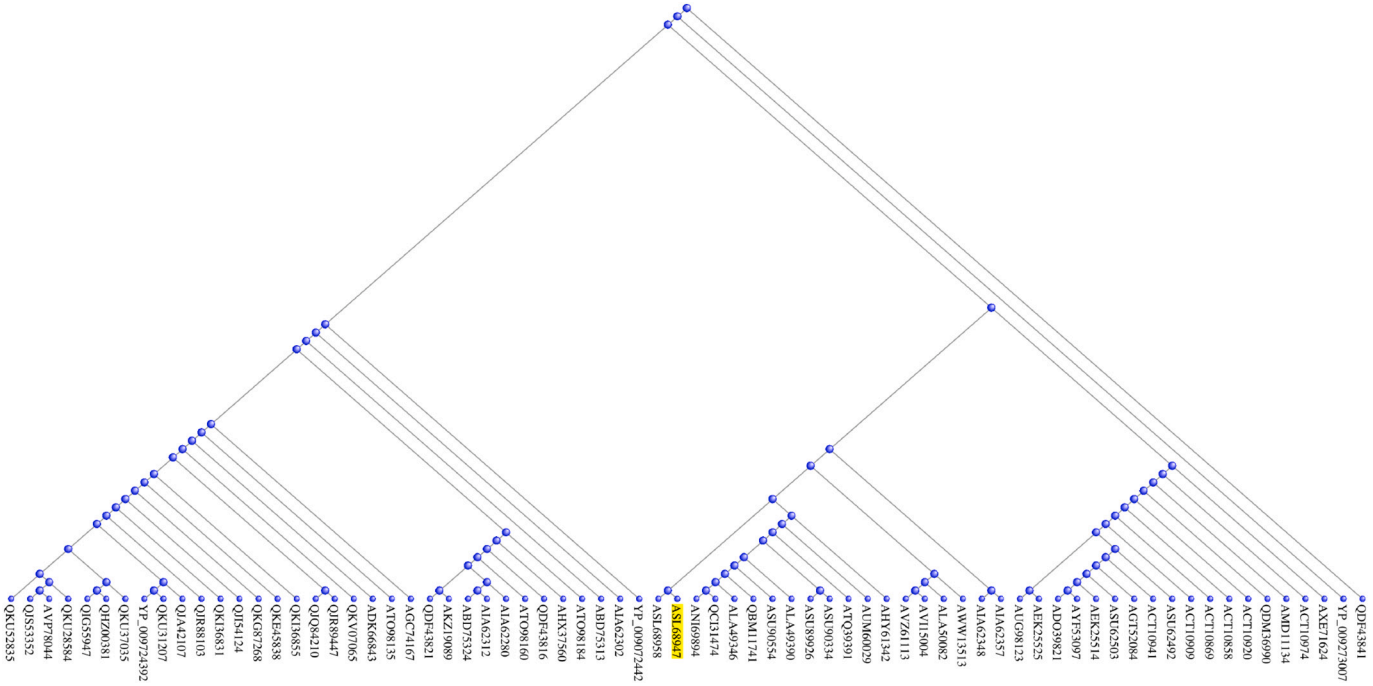

Fig. 9. Phylogenetic relationship among envelope proteins of the different host CoVs with respect to the sequence based homology.

interval of 0.84 and 0.94 whereas SE of most common E protein of SARS-CoV-2 and that of the Pangolin-CoV are found to be identical and it is close to 0.846. Note that, there is an E protein of AVP78044 (Bat-CoV) whose SE is also identical to 0.846. The remaining fifteen different E proteins of SARS-CoV-2 are close enough to other Bat-CoVs by accumulating various missense mutations. The SE of the E protein of SARS-CoV2 lies in between 0.824 and 0.862. There are E proteins of SARS-CoV-2 whose SE of amino acid conservation is tightly bounded by that of Pangolin and Bat-CoVs. It is found that SE of E proteins of ADK66843 (Bat-CoV) and QKU52835 (SARS-CoV-2) are turned out to be identical (0.852). There are other such examples too which are clearly observed in the Table 10. This phylogenetic relationship is endorsed by the amino acid conservation and their associated SE found in Table 10. We also observed phylogenetic relationship among E proteins from Bat-CoV, Pangolin-CoV and SARS-CoV-2 (Fig. 11). This relationship was drawn using amino acid conservation and their associated SE (Table 10).

4. Conclusions

Here, we performed phylogenetic analysis of E protein sequences of coronaviruses from different hosts although different investigators also performed phylogenetic analysis using the genomic and protein sequences of few coronaviruses from different hosts [21]. But the phylogenetic analysis, using E protein sequences from a large number of seuquences, may provide a better picture of the relationship among hosts coronaviruses so far as the intermediate host between human and bat is concerned since protein is the functional unit in the cell. So, this study, using protein sequence variations, may provide the clue why few hosts are resistant or sensitive to the disease Covid-19. We observed variations in protein sequences of E-protein in Human-SARS-CoV-2, Bat-CoV, Camel-CoV etc. Based on mutation characteristics and amino acid conservations over the E proteins across various host CoVs, this report predicts potential close kins of human SARS-CoV-2 as the Pangolin-CoV and Bat-CoV which was also reported in a recent study [21]. Pangolin, the closest kin of SARS-CoV-2, is also confirmed by the analysis made in this study. The missense mutations of the E protein across various host CoVs, may bar the usual functions of the envelope protein and consequently the virus may become weaker in infectivity. It is our belief that various missense mutations in the E protein could weaken the SARS-CoV-2 and would help us gets rid of COVID-19 in future since any virus does not like to destroy its host for its survival for a long to come.

Data availability

The protein sequences of the SARS-CoV-2 and other host CoVs used in this study are available in the NCBI virus database *https : //www. ncbi. nlm. nih. gov/labs/virus/vssi/*.

Author contributions

SH conceived the problem. SH determined the mutations. SH, PPC, BR analyzed the data and result. SH wrote the initial draft which was checked and edited by all other authors to generate the final version.

Declaration of Competing Interest

The authors do not have any conflicts of interest to declare.

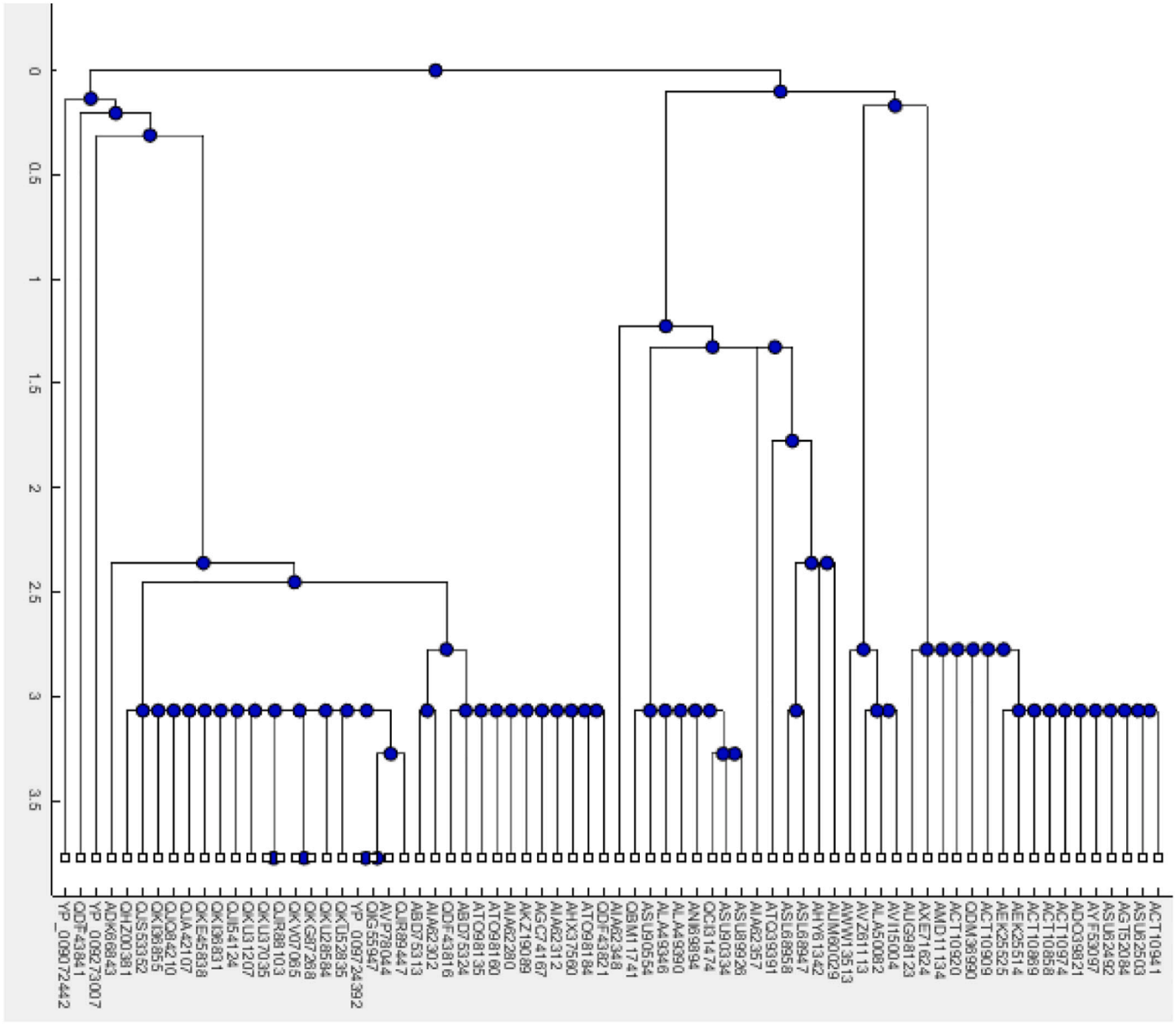

Fig. 10. Phylogenetic relationship among envelope proteins of the different host CoVs with respect to the amino acids conservation.

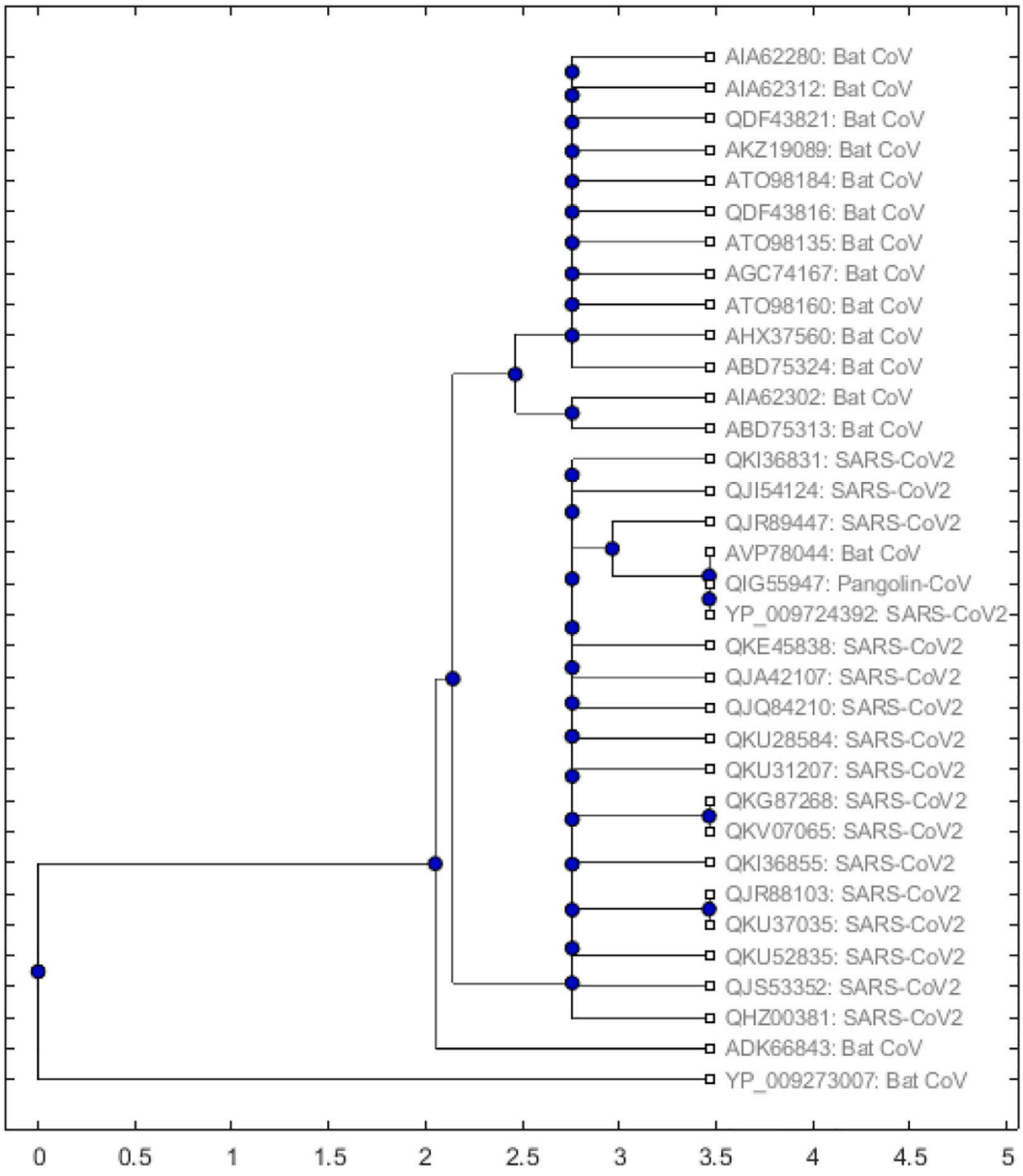

Fig. 11. Phylogenetic relationship among envelope proteins of the SARS-CoV2, Bat and Pangolin CoVs with respect to the amino acids conservation.

References

[1] E.G. Favalli, F. Ingegnoli, O. De Lucia, G. Cincinelli, R. Cimaz, R. Caporali, Covid-19 infection and rheumatoid arthritis: faraway, so close!, Autoimmun. Rev. 102523 (2020).

[2] M. Bartlam, H. Yang, Z. Rao, Structural insights into sars coronavirus proteins, Curr. Opin. Struct. Biol. 15 (6) (2005) 664–672.

[3] Y. Liao, Q. Yuan, J. Torres, J. Tam, D. Liu, Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein, Virology 349 (2) (2006) 264–275.

[4] J.L. Nieto-Torres, M.L. DeDiego, E. Álvarez, J.M. Jiménez-Guardeño, J.A. Regla-Nava, M. Llorente, L. Kremer, S. Shuo, L. Enjuanes, Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, Virology 415 (2) (2011) 69–82.

[5] J.L. Nieto-Torres, M.L. DeDiego, C. Verdia-Baguena, J.M. Jimenez-Guardeno, J.A. Regla-Nava, R. Fernandez-Delgado, C. Castano-Rodriguez, A. Alcaraz, J. Torres, V.M. Aguilella, et al., Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis, PLoS Pathog. 10 (5) (2014).

[6] L. Wilson, C. Mckinlay, P. Gage, G. Ewart, Sars coronavirus e protein forms cation-selective ion channels, Virology 330 (1) (2004) 322–331.

[7] K. Parthasarathy, L. Ng, X. Lin, D.X. Liu, K. Pervushin, X. Gong, J. Torres, Structural flexibility of the pentameric sars coronavirus envelope protein ion channel, Biophys. J. 95 (6) (2008) L39–L41.

[8] J. To, W. Surya, T.S. Fung, Y. Li, C. Verdia-Baguena, M. Queralt-Martin, V.M. Aguilella, D.X. Liu, J. Torres, Channel-inactivating mutations and their revertant mutants in the envelope protein of infectious bronchitis virus, J. Virol. 91 (5) (2017) e02158–16.

[9] J. Han, K. Pluhackova, R.A. Böckmann, Exploring the formation and the structure of synaptobrevin oligomers in a model membrane, Biophys. J. 110 (9) (2016) 2004–2015.

[10] L.A. Lopez, A.J. Riffle, S.L. Pike, D. Gardner, B.G. Hogue, Importance of conserved cysteine residues in the coronavirus envelope protein, J. Virol. 82 (6) (2008) 3000–3010.

[11] B. Horwitz, A. Burkhardt, R. Schlegel, D. DiMaio, 44-amino-acid e5 transforming protein of bovine papillomavirus requires a hydrophobic core and specific carboxyl-terminal amino acids, Mol. Cell. Biol. 8 (10) (1988) 4071–4078.

[12] C. Yang, C.P. Spies, R.W. Compans, The human and simian immunodeficiency virus envelope glycoprotein transmembrane subunits are palmitoylated, Proc. Natl. Acad. Sci. 92 (21) (1995) 9871–9875.

[13] Y. Liao, J. Lescar, J. Tam, D. Liu, Expression of sars-coronavirus envelope protein in escherichia coli cells alters membrane permeability, Biochem. Biophys. Res. Commun. 325 (1) (2004) 374–380.

[14] K.-T. Teoh, Y.-L. Siu, W.-L. Chan, M.A. Schlüter, C.-J. Liu, J.M. Peiris, R. Bruzzone, B. Margolis, B. Nal, The sars coronavirus e protein interacts with pals1 and alters tight junction formation and epithelial morphogenesis, Mol. Biol. Cell 21 (22) (2010) 3838–3852.

[15] Z. Zhao, H. Li, X. Wu, Y. Zhong, K. Zhang, Y.-P. Zhang, E. Boerwinkle, Y.-X. Fu,

Moderate mutation rate in the sars coronavirus genome and its implications, BMC Evol. Biol. 4 (1) (2004) 21.
[16] S.S. Hassan, P.P. Choudhury, P. Basu, S.S. Jana, Molecular conservation and differential mutation on orf3a gene in indian sars-cov2 genomes, Genomics 112 (5) (2020) 3226–3237.
[17] S.S. Hassan, P.P. Choudhury, B. Roy, Rare Mutations in the Accessory Proteins orf6, orf7b and orf10 of the Sars-cov2 Genomes, (2020).
[18] K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of sars-cov-2, Nat. Med. 26 (4) (2020) 450–452.
[19] Y.-Z. Zhang, E.C. Holmes, A genomic perspective on the origin and emergence of sars-cov-2, Cell 181 (2) (2020) 223–227.
[20] T. Zhang, Q. Wu, Z. Zhang, Probable pangolin origin of sars-cov-2 associated with the covid-19 outbreak, Curr. Biol. 30 (7) (2020) 1346–1351.
[21] P. Liu, J.-Z. Jiang, X.-F. Wan, Y. Hua, L. Li, J. Zhou, X. Wang, F. Hou, J. Chen, J. Zou, et al., Are pangolins the intermediate host of the 2019 novel coronavirus (sars-cov-2)? PLoS Pathog. 16 (5) (2020) e1008421.
[22] The Mathworks, Inc, Natick, Massachusetts, MATLAB version (R2020a), (2020).
[23] F. Johansson, H. Toh, Relative von neumann entropy for evaluating amino acid conservation, J. Bioinforma. Comput. Biol. 8 (05) (2010) 809–823.
[24] E. Garriga, P. Di Tommaso, C. Magis, I. Erb, L. Mansouri, A. Baltzis, H. Laayouni, F. Kondrashov, E. Floden, C. Notredame, Large multiple sequence alignments with a root-to-leaf regressive method, Nat. Biotechnol. 37 (12) (2019) 1466–1470.
[25] F. Madeira, Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R. Tivey, S.C. Potter, R.D. Finn, et al., The embl-ebi search and sequence analysis tools apis in 2019, Nucleic Acids Res. 47 (W1) (2019) W636–W641.
[26] J. Weako, A. Gursoy, O. Keskin, Mutational effects on protein–protein interactions, Protein Interactions: Computational Methods, Analysis And Applications 109 (2020).
[27] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, Virol. J. 16 (1) (2019) 69.
[28] M.K. Gupta, S. Vemula, R. Donde, G. Gouda, L. Behera, R. Vadde, In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel, J. Biomol. Struct. Dyn. (2020) 1–11.
[29] J.W. Westerbeck, C.E. Machamer, The infectious bronchitis coronavirus envelope protein alters golgi ph to protect the spike protein and promote the release of infectious virus, J. Virol. 93 (11) (2019) e00015–e00019.
[30] F. De Maio, E.L. Cascio, G. Babini, M. Sali, S. Della Longa, B. Tilocca, P. Roncada, A. Arcovito, M. Sanguinetti, G. Scambia, et al., Enhanced Binding of Sars-Cov-2 Envelope Protein to Tight Junction-Associated pals1 Could Play a Key Role in Covid-19 Pathogenesis, (2020).
[31] S.S. Hassan, P.P. Choudhury, B. Roy, Sars-cov2 Envelope Protein: Non-synonymous Mutations and its Consequences, (2020).