JOURNAL OF COMPLEMENTARY MEDICINE RESEARCH
Journal of Complementary Medicine Research, ISSN: 2146-8397 Vol. 13, No. 1, 2022 (pp. 92-97)
RESEARCH ARTICLE
WWW.JOCMR.COM

Detection of Missense and Frame Shift Mutations in Envelope Proteins using Molecular Phylogenetic SARS-COV-2

Noor Abady[1*], Zaytoon A. Alkhafaji[2], Ali S. Baay[3]

[1]*PhD student at Babylon University College of Medicine, Al-Qasim Green University*

[2]*Department of Microbiology, Babylon University College of Medicine*

[3]*Department of Medicine, Babylon University, Hammurabi University of Medicine*

ABSTRACT

The envelope (E) protein, which is 76–109 amino acids long, is a structural viroporin identified in coronaviruses. On March 1, 2020, ten distinct E-proteins were selected out of a total of 50 at the general Health laboratory (Babil), and the partial E gene of SARS-CoV-2 genomes was sequenced and recoded in the GenBank with accession number MW827729, MW827730, MW827731, MW827732, MW827733, MW827734, MW827735,MW827736, MW827737, MW827738. The study looked for missense and frame shift mutations in the envelope proteins of different Covid 19 patients to determine the lineage of the direct envelope protein (SARS-CoV-2). A phylogenetic analysis of envelope proteins, which looks at sequence homology and amino acid conservation, provides even more evidence to the evolutionary origin. Frame-shift mutations were found in both the N and C terminals of the 10 known partial sequences of human SARS- COV2. genomes.

Corresponding Author e-mail: Ilhamalsaedi2008@gmail.com

How to cite this article: Abady N, Alkhafaji ZA, Baay AS (2022). Detection of Missense and Frame Shift Mutations in Envelope Proteins using Molecular Phylogenetic SARS-COV-2. Journal of Complementary Medicine Research, Vol. 13, No. 1, 2022 (pp. 92-97).

INTRODUCTION

The continuing epidemic, which is unquestionably life-threatening, that our planet has been experiencing since December 2019 (Favalli *et al.*, 2020) has been caused by a new coronavirus. Coronaviruses (CoV) are viruses that contain RNA as a genetic substance with a positive sense that cause respiratory illnesses in humans. a wide variety of creatures. Several novel human coronaviruses, such as serious (SARS-CoV) MERSCoV and SARS-CoV-2, were found recently, which have piqued the interest of scientists thorough knowledge of viruses and the discovery of antiviraldrugs targets for therapeutic treatment development A CoV is made from of severalproteins (structural, non-structural, complementary, etc.) are found in the human body.

Which names are contained in the two major structural proteins (CoVs) Membrane and Spike(S) (M) Two forms of glycoproteins(Bartlam *et al.*, 2005). All Coronavirus in the genus has an envelope (E) protein, which has 76–109 amino acids and is required for virus assembly, budding, and replication morphogenesis, host cell entrance, and certain other cellular processes (Liao *et al.*, 2006) functions. This E protein is primarily a glycoprotein observed in the ERGIC "Endoplasmic Reticulum-Golgi Intermediate Compartment" of cells transfected with an E protein producing plasmid or the ERGIC of high expression with an E protein encoding genetic material infection with SARS-CoV (Nieto-Torres *et al.*, 2011). The SARS-CoV-2 envelope protein is 75 amino acids long and has three key domains: the (N)-terminus, which contains a 7–9 hydrophilic area, and the transmembrane domainTMD comprising 29 amino acid residues (hydrophobic region) with a high leucine/isoleucine/valine content, and (C)-terminus with (Nieto-Torres *et al.*, 2014)The hydrophilic area (Fig 2). Ion channels are formed by the envelope (E) protein of the coronavirus (CoV) of the genus famiy (Wilson *et al.*, 2004). The transmembrane domain (TMD) of theE protein isresponsible for theobserved ionchannel activity, which may signal that the virus's infectivity has been reduced. Missense mutations are caused by mutations in the E protein led to inhibited ion channel activity which is it responsible for the attenuation. (Parthasarathy *et al.*, 2008). TMD is said to form stable pentamers, which has been verified by molecular modeling and in vitro oligomerization (Han *et al.*, 2016). According to reports, the hydrophobic amino acid residues in the E's TMD have been mutated proteins containing charged amino acids have a major impact on migratory bacteria. The E protein's characteristics (BartlamYang and Rao, 2005). Y. Liao et al. (2006) discovered that the

KEYWORDS:
E gene, Frame shift, Missense mutation, SARS-COV-2

ARTICLE HISTORY:
Received Dec 06,2021
Accepted Jan 13,2022
Published Mar 10,2022

DOI:
10.5455/jcmr.2022.13.01.17

TMD is required for the protein's membrane permeabilizing action and that missense mutations in the TMD of the E protein impair the protein's function (LiaoYuanTorresTam and Liu, 2006). It is correct discovered that the envelope protein of SARS-CoV and SARS-CoV-2 consists of three cysteine residuesat positions 40, 43, and 44(Lopez *et al.*, 2008). The first and thirdcysteine residues are located at amino acid positions 40 and 41, respectively and 44, respectively, have previously been implicated in the oligomerization of the E protein (Horwitz *et al.*, 1988)1988. Additionally, biochemical analysis revealed that all three cysteine residues are palmitoylated during translation (Yang *et al.*, 1995). Mutagenesis investigations have revealed that the transmembrane domain is responsible for the transmembrane domain's membrane permeabilizing action. (Liao *et al.*, 2004) Protein E of SARS-CoV. Envelope protein's (C)-terminal domain. Human PALS1, a tight junction-associated protein, binds to SARS-CoV-2 in SARS-CoV-2 Protein is required for the growth and maintenance of all living things.(Teoh *et al.*, 2010)Mammalian epithelial polarity. The translation of the overlapping but displaced ORF1a,b gene region, which codes for the unstructured polypeptides that initiate the viral protein production cascade, is one of the most important stages in viral replication. This is accomplished by the process of 1 programmed ribosomal frameshifting, a method used by many viruses to delay ribosomal translation and deal with overlapping reading frames (Brierley, 1995). Envelope is implicated in crucial elements of the viral life cycle, according to the data, and CoVs missing E constitute potential vaccine candidates. Because of the high mortality rate, The ease with which some CoVs can be transmitted underscores the necessity for greater study into CoV molecular biology.This may help in the development of efficient anti-coronaviral medicines against both human and enzootic CoVs.(Schoeman and Fielding, 2019) The ribosomal frameshifting needed for translation of overlapping open reading frames is a crucial stage in the viral replication cycle within host cells. One of three highly conserved sections of coronaviruses, the frameshifting element (FSE), contains an RNA pseudoknot that is required for ribosomal switching.

MATERIAL AND METHOD

50 COVID-19 patients were brought to Morgan Hospital on March 1, 2021, and 50 nasopharyngeal swaps were obtained by viral transfer mediaVTM and delivered to the general Health laboratory for examination, where they were placed in the Smarter equipment for RNA extraction for 38 minutes. Following that, RNA was put into a PCR tube containing PCR mix (Dna Rna extraction identification Kit). RNA Extraction and Reverse Transcription: As directed by the manufacturer, RNA was isolated from Nasopharyngeal swabs using a magnetic bead nucleic acid purification kit (Cat No. DA0630, DAAN Gene, China). A Promega (fluorometer, ng/ul) is used to measure the purity and amount of viral RNA. The following RT-PCR began with 2 l of total RNA, a random hexamer of 100 pmol, and a final reaction volume of 2 l at a final reaction volume of 20 l. The thermocycling was carried out using the BioRad Three thermal cycler (Biorun,). For 10 minutes, the initial state was adjusted to 15 C°, cDNA synthesis began at 32C° for 20 minutes, and heat inactivation was set at 95C° for 5 minutes. As previously stated, nasopharyngeal swab samples were collected and processed in special tubes containing viral transport medium before being submitted to the laboratory for SARS-CoV-2 RNA extraction and detection by real-time reverse transcription polymerase chain reaction (RT-PCR). A positive test result was defined as a period threshold value (Ctvalue) of 37, and a negative test result as a Ct-value of 40 or above. These diagnostic criteria were developed in accordance with the National Institute for Viral Disease Control and Prevention of China (Zhang *et al.*, 2020).

RNA to cDNA Conversion

Appropriate RNA for SARS-cov2 was collected for genetic testing, and the purity of the RNA was assessed using a single tube-format procedure (promega), in which 0.5 mix was added to the tube and incubated for 5 minutes before being examined using a QuantusTM Fluorometer (Ex 460nm, Em 515–575nm).

Polymer chains reaction for cDNA amplification (PCR).

The E gene primer was utilized for this study since it is essential for diagnosing Covid19 and identifying its genetic characteristics. Future-oriented primer Forward primer TTCGTTTCGGAAGAGACAGGT(NC045512.2), reverse primer CCA GAAGATCAGGAACTCTAGAAGA ((NC045512.2), 28274..29533) The 2014bp product duration PCR material was examined for bands indicating positive DNA using gel electrophoresis. Tm57.2C annealing master mix applied to primer 0.5 for each sample to the DNA ending volume to the 20 and put into thermo cycler.

Sequencing

The Pcr amplification was sequenced using the sanger dideoxy sequencing method, and the resulting Fasta file was aligned using blast, and ten samples were submitted into the Gene Bank database with accession numbers (MW827729, MW827730, MW827731, MW827732, MW827733, MW827734, MW827735,MW827736, MW827737, MW827738)

Phylogenetic investigation

PhylogenyphyML was employed for genetic study, and muscle aliment was conducted in the tree building for Iraqi E genes with ncbi genes, demonstrating the closed strain and similarity. Multiple Sequence Alignment was used to align protein sequences (Clustal Omega - EMBL-EBI)

RESULT AND DISCUSSION

E gene (envelope protein) detection that was employed for the molecular weight detection of 214 bp in the target primar and the results given in Fig 1 showing a comparison of 10 unique bands for E gene to the ladder of the (1000bp).

A short hydrophilic portion of 7–12 Amino acids, according to structural data, is the N,amino terminus of the envelope or E protein. Following it is a large hydrophobic middle region with 25 transmembrane acids, which is followed by an extended hydrophilic C-terminal which is responsible for the greater part of the protein. The E protein participates in the virus's generation and maturation phases by interacting with the host cell membrane protein(BakhshandehJahanafroozAbbasiGoliSadeghiMottaqi and Zamani, 2021)ÿÿþ.

SARS-Cov2 missense- mutations in the E protein. Only four of the ten known partial sequences of human CoV genomes

exhibited changes owing to frame-shifts, as shown in Fig. 3. The frame shifting element (FSE) is a short area in the center of the ORF1a,b gene region (less than 100 nucleotides); such processes are assumed to be dependent on specific structural modules and/or related structural- transitions. (Ritchie *et al.*, 2014). SARS-CoV-2 E has made the most progress, with researchers identifying structural requirements for its activities in the CoV-2 life cycle as well as pathogenic processes. The SARS-CoV-2 envelope proteins are highly conserved as illustrated by (Hassan *et al.*, 2020a). It is worth noting that the N-terminal sequence contains four frame-shifts as shown in the Fig.3 and one frame shift mutation at c terminal as shown in the Fig 3.

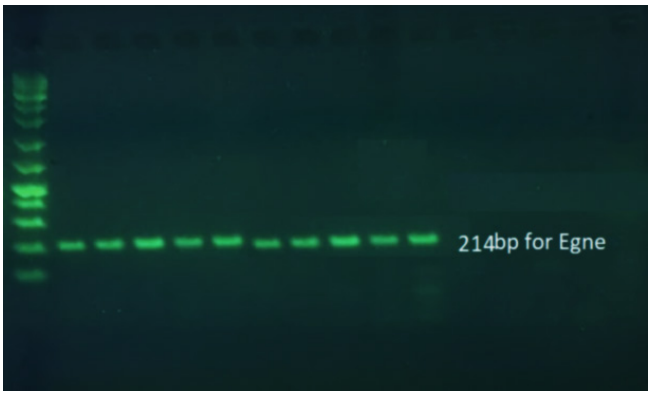

Fig. 1: Electrophoresis of genomic DNA and PCR products for the E gene on an agarose gel (SARS-cov2).

Multiple sequence alignment is used to determine the mutations of the E proteins (given in Fig.3). It should be mentioned that the major changes occure in the N- terminal region rather than in C-terminal region as mentioned in the recent researh for the E protein SARS-CoV-2 have already been documented in published manuscript(Hassan *et al.*, 2020b)2020barticle which record that the majority of the missense mutations occurred at the C-terminus. While one frame shift mutation occurred at C terminal with one V61D missense mutation. Another reseach have been mentioned that E protein of QKN20885 (USA) and QJQ84210 (USA: New Orleans, LA) has a mutation in the TMD at F26L this mutation in the TMD stops the ion channel from working, which could lead to in vivo attenuation.

While the missense mutations occurred at the N-terminus for the E envelope protein at N1Y as shown in the Fig.3.The SARS-CoV-2 S protein is made up of an N-terminal S1 subunit that is important for virus-receptor affinity and a C-terminal S2 subunit that facilitates virus-cell endocytosis(Walls *et al.*,

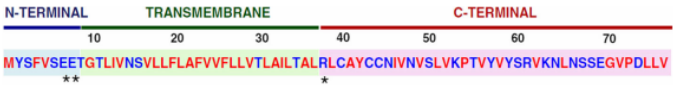

Fig. 2: Amino Acid Sequence and Domains SARS-CoV E Protein The three domains that comprise the SARS-CoV E protein are the amino (N)-terminal domain, the transmembrane (TMD) domain and the carboxy (C)-terminal domain. It shows hydrophobic -red, hydrophilic blue, polar and loaded amino acids (asterisks)*(Bakhshandeh et al., 2021)2021

```
CLUSTAL O(1.2.4) multiple sequence alignment

MZ540294.1    NSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MZ540292      NSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MZ540290      NSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MT875582      NSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
QNA42613.1    NSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827737      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827735      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827734      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827732      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827731      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827730      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827729      YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827738      ----LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827736      ----LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
MW827733      ----LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
                ********************************************

MZ540294.1        RVKNLNSSRVPDLLV
MZ540292          RVKNLNSSRVPDLLV
MZ540290          RVKNLNSSRVPDLLV
MT875582          RVKNLNSSRVPDLLV
QNA42613.1        RVKNLNSSRVPDLLV
MW827737          RVKNLNSSRVPDLL-
MW827735          RVKNLNSSRVPDLL-
MW827734          RVKNLNSSRVPDLL-
MW827732          RVKNLNSSRVPDLL-
MW827731          RVKNLNSSRVPDLL-
MW827730          RVKNLNSSRVPDLL-
MW827729          RVKNLNSSRVPDLL-
MW827738          RVKNLNSSRVPDLLD
MW827736          RVKNLNSSRVPDLLD
MW827733          RVKNLNSSRVPDLLD
                  ***************
```

Fig. 3: Using clustal omega, we aligned the human CoV2-E protein sequence with five internationaly recoded proteins in the NCBI gene bank data set They are denoted by a square.

2020, Wrapp *et al.*, 2020). An N-terminal domain (NTD) and a receptor-binding domain (RBD), sometimes known as domain A and B, respectively, make up the S1 subunit(Tortorici and Veesler, 2019). Using the Clustal-Omega server (Garriga *et al.*, 2019, Madeira *et al.*, 2019), we created a multiplesequence alignment of the Eprotein sequences (Table-1) to identify missense mutations. (Table- 3) shows the amino acid residues with their descriptions and characteristic of each are stated. Madeira Park Lee Buso Gur Madhu soodanan Basutkar Tivey Potter Finn and Lopez, 2019). Significantly, genome sequence findings show that the transition from C to T is the most prevalent nucleotide alteration, resulting in considerably greater rates of synonymous and missense mutations than deletion and insertion mutations(BakhshandehJahanafroozAbbasiGoliSadeghiMottaqi and Zamani, 2021)ÿÿþ.Using of amino acid frequency based phylogeny to establish a more detailed phylogenetic connection between the E proteins of the SARS COV-2. The amino acid frequencies for each of the common E proteins were calculated from each sequence of the host, as shown in (Table-1) . The distance has been computed for E protein, and the phylogeny has been generated as a result Fig. 4and 5.

Sequence-based homology supports this connection. Based on sequence homology, we show the evolutionary relationship among the E proteins (Table -1) spanning 10 SARS-Cov2 sequences in Fig. 3. The E proteins of MW827737 SARS-CoV2 and MW827735 SARS-CoV-2 are fairly close to each other among all E proteins, according to the phylogeny in Fig. 4-5. Using amino acid frequency based phylogeny to offer a more comprehensive evolutionary relationship among the E proteins for all sequences obtained.

The sequence homology (Fig 4 and 5.) shows that the E proteins of USA-CoV MZ540294.1, Mz540292, MZ540290, and MT875582 are highly similar. The phylogeny (Fig.5) further indicated that the E proteins of Iraqi -CoV MW827737, MW827735, MW827734, MW827732, MW827731, MW827730, and MW827729 are on the same branch with the same degree of phylogeny based on amino acid conservation. The E proteins of SARS-CoV2 MW827738 and MW827736 are highly similar, as seen in the phylogeny . According to the phylogeny based on amino acidconservations, the E proteins MW827733 and MW827736 are related (SARS-CoV-2) The effect of mutation on the E protein has been agreed upon (Loney *et al.*, 2021, Melegari *et al.*, 1997).

In addition, PhyloML was used to create a phylogenetic tree based on nucleotide alignment. MW827737, MW827735, MW827734, MW827731 sharing identical information while, MW827732, MW827730, MW827729 identical and MW827738, MW827736 were separate but still grouped with them from phylogeny revealing the effect of mutation clearly seen on the evolution tree the same result were mentioned by the recent study In 15 (0.414 percent) of the 3617 SARS-CoV2 genomes studied, the envelope (E) protein has multiple non-synonymous mutations in the transmembrane and C-terminus domains. More specifically, missense mutations over the E-protein of SARS-CoV2 genomes were found in 10 (0.386 percent) of 2588 genomes from the United States, 3 (0.806 percent) from Asia, 1 (0.348 percent) from Europe, and 1 (0.274 percent) from Oceania(HassanChoudhury and Roy, 2020b, Chai *et al.*, 2021).

CONCLUSION

Coronavirus phylogenetic analysis for E-protein sequences in SARS-COV2 patients was performed in conjunction with genomic

Table 1: Envelope proteins differ amongst patients CoV-2.

Protein ID(SARS-Cov2)	*E proteinsequence (N to C terminal of protein)*	*Length*
QTH36184.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36182.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36181.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36179.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36178.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36177.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36176.1	YSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLL	60
QTH36185.1	LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLD	57
QTH36183.1	LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLD	57
QTH36180.1	LFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLD	57
QNA42613.1	MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV	76

Table 2: Amino acid counts were taken over the envelope proteins of the human host CoV-2.

Amino acid counts over the envelope proteins over the Human SARS- CoV-2s														
Protein ID	A	R	N	D	C	I	L	K	F	P	S	T	Y	V
QTH36184.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36182.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36181.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36179.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36178.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36177.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36176.1	4	3	4	1	3	2	11	2	3	2	6	2	2	10
QTH36185.1	3	3	4	2	3	2	12	2	4	2	5	2	2	9
QTH36183.1	3	3	4	2	3	2	12	2	4	2	5	2	2	9
QTH36180.1	3	3	4	2	3	2	12	2	4	2	5	2	2	9

Table 3: Residues of amino acids, as well as their color and properties

Amino acid residuesas well as their color and properties, are depicted in Figure 3		
Residue Color Property		
V,L,F,A,V,I and P	RED	hydrophobic (incl.aromatic eY)
D	BLUE	Acidic
Rand K	MAGENTA	Basic - H
S,T,Y,Cand N	GREEN	Hydroxyl + sulfhydryl + amine + G

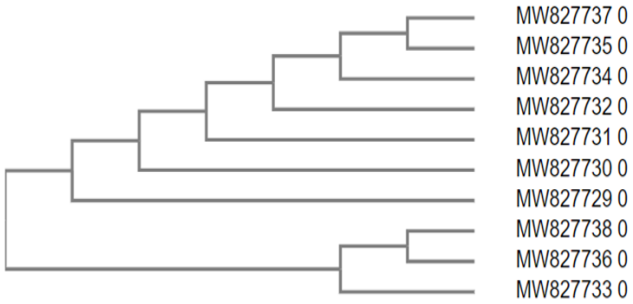

Fig. 4: Based on sequence homology, phylogeny of the envelope protein of distinct host-CoVs (Iraqi- isolate)

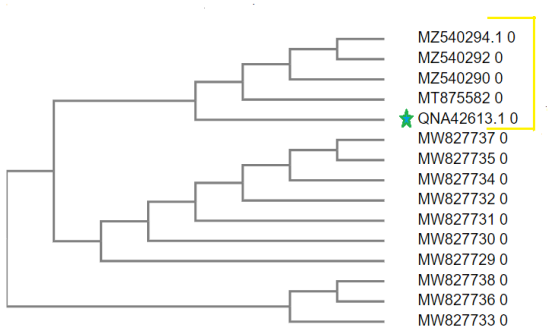

Fig. 5: Phylogeny of the envelope protein ofb SARS-CoV-2 based on sequence homology (iraqi isolate vs. near international isolate). The yellow line marked the internationally with one E proteins Iraqi QNA42613.1 is denoted by a star (Erbil) and the remaining ten E genes utilized in this investigation are likewise from the Iraqi strain.

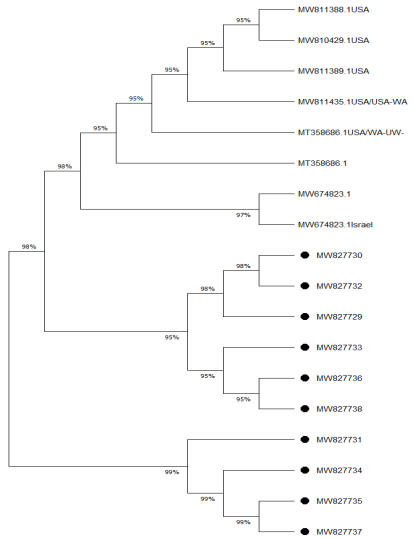

Fig. 6: Phylogenetic tree construction using the Maximum Likelihood approach for 10 E gene with NCBI genetic based genes from various countries using http://www.phylogeny.fr/ PhyloML (Dereeper et al., 2008)2008 black circle for E genes are Iraqi and other from NCBI (blast) Genebank.

and protein sequence phylogenetic analysis. As changed genes are a concern, the study suggested that the relationship between the patients may be further proved by phylogenetic research employing E Protein sequences for various sequences. Based on differences in the protein sequence, this discovery would explain Why are just a few hosts resistant or susceptible to the disease? Covid19. Human-SARS-CoV-2 has been revealed to contain changes in its protein sequences.

REFERENCES

1. BAKHSHANDEH, B., JAHANAFROOZ, Z., ABBASI, A., GOLI, M. B., SADEGHI, M., MOTTAQI, M. S. andZAMANI, M. 2021. Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus. *Microbial pathogenesis,* 154, 104831-104831.
2. BARTLAM, M., YANG, H. andRAO, Z. 2005. Structural insights into SARS coronavirus proteins. *Current Opinion in Structural Biology,* 15, 664-672.
3. BRIERLEY, I. 1995. Ribosomal frameshifting on viral RNAs. *Journal of General Virology,* 76, 1885-1892.
4. CHAI, J., CAI, Y., PANG, C., WANG, L., MCSWEENEY, S., SHANKLIN, J. andLIU, Q. 2021. Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein PALS1. *Nature Communications,* 12, 1-6.
5. DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J. F., GUINDON, S., LEFORT, V., LESCOT, M., CLAVERIE, J. M. andGASCUEL, O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res,* 36, W465-9.
6. FAVALLI, E. G., INGEGNOLI, F., DE LUCIA, O., CINCINELLI, G., CIMAZ, R. andCAPORALI, R. 2020. COVID-19 infection and rheumatoid arthritis: Faraway, so close! *Autoimmunity reviews,* 19, 102523.
7. GARRIGA, E., DI TOMMASO, P., MAGIS, C., ERB, I., MANSOURI, L., BALTZIS, A., LAAYOUNI, H., KONDRASHOV, F., FLODEN, E. andNOTREDAME, C. 2019. Large multiple sequence alignments with a root-to-leaf regressive method. *Nature biotechnology,* 37, 1466-1470.
8. HAN, J., PLUHACKOVA, K. andBÖCKMANN, RAINER A. 2016. Exploring the Formation and the Structure of Synaptobrevin Oligomers in a Model Membrane. *Biophysical Journal,* 110, 2004-2015.
9. HASSAN, S. S., CHOUDHURY, P. P. andROY, B. 2020a. Molecular phylogeny and missense mutations at envelope proteins across coronaviruses. *Genomics,* 112, 4993-5004.
10. HASSAN, S. S., CHOUDHURY, P. P. andROY, B. 2020b. SARS-CoV2 envelope protein: non-synonymous mutations and its consequences. *Genomics,* 112, 3890-3892.
11. HORWITZ, B., BURKHARDT, A., SCHLEGEL, R. andDIMAIO, D. 1988. 44-amino-acid E5 transforming protein of bovine papillomavirus requires a hydrophobic core and specific carboxyl-terminal amino acids. *Molecular and Cellular Biology,* 8, 4071-4078.
12. LIAO, Y., LESCAR, J., TAM, J. P. andLIU, D. X. 2004. Expression of SARS-coronavirus envelope protein in Escherichia coli cells alters membrane permeability. *Biochemical and Biophysical Research Communications,* 325, 374-380.
13. LIAO, Y., YUAN, Q., TORRES, J., TAM, J. P. andLIU, D. X. 2006. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology,* 349, 264-275.
14. LONEY, T., KHANSAHEB, H., RAMASWAMY, S., HARILAL, D., DEESI, Z. O., VARGHESE, R. M., BELAL AL ALI, A., KHADEEJA, A., AL SUWAIDI, H. andALKHAJEH, A. 2021. Genotype-phenotype correlation identified a novel SARS-CoV-2 variant possibly linked to severe disease. *Transboundary and emerging diseases*.
15. LOPEZ, L. A., RIFFLE, A. J., PIKE, S. L., GARDNER, D. andHOGUE, B. G. 2008. Importance of conserved cysteine residues in the coronavirus envelope protein. *Journal of virology,* 82, 3000-3010.

16. MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R. N., POTTER, S. C., FINN, R. D. andLOPEZ, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research,* 47, W636-W641.

17. MELEGARI, M., SCAGLIONI, P. P. andWANDS, J. R. 1997. The small envelope protein is required for secretion of a naturally occurring hepatitis B virus mutant with pre-S1 deleted. *Journal of virology,* 71, 5449-5454.

18. NIETO-TORRES, J. L., DEDIEGO, M. L., ÁLVAREZ, E., JIMÉNEZ-GUARDEÑO, J. M., REGLA-NAVA, J. A., LLORENTE, M., KREMER, L., SHUO, S. andENJUANES, L. 2011. Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology,* 415, 69-82.

19. NIETO-TORRES, J. L., DEDIEGO, M. L., VERDIÁ-BÁGUENA, C., JIMENEZ-GUARDEÑO, J. M., REGLA-NAVA, J. A., FERNANDEZ-DELGADO, R., CASTAÑO-RODRIGUEZ, C., ALCARAZ, A., TORRES, J. andAGUILELLA, V. M. 2014. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS pathogens,* 10, e1004077.

20. PARTHASARATHY, K., NG, L., LIN, X., LIU, D. X., PERVUSHIN, K., GONG, X. andTORRES, J. 2008. Structural Flexibility of the Pentameric SARS Coronavirus Envelope Protein Ion Channel. *Biophysical Journal,* 95, L39-L41.

21. RITCHIE, D. B., SOONG, J., SIKKEMA, W. K. andWOODSIDE, M. T. 2014. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *Journal of the American Chemical Society,* 136, 2196-2199.

22. SCHOEMAN, D. andFIELDING, B. C. 2019. Coronavirus envelope protein: current knowledge. *Virology journal,* 16, 1-22.

23. TEOH, K.-T., SIU, Y.-L., CHAN, W.-L., SCHLÜTER, M. A., LIU, C.-J., PEIRIS, J. M., BRUZZONE, R., MARGOLIS, B. andNAL, B. 2010. The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Molecular biology of the cell,* 21, 3838-3852.

24. TORTORICI, M. A. andVEESLER, D. 2019. Structural insights into coronavirus entry. *Advances in virus research,* 105, 93-116.

25. WALLS, A. C., PARK, Y.-J., TORTORICI, M. A., WALL, A., MCGUIRE, A. T. andVEESLER, D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell,* 181, 281-292. e6.

26. WILSON, L., MCKINLAY, C., GAGE, P. andEWART, G. 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology,* 330, 322-331.

27. WRAPP, D., WANG, N., CORBETT, K. S., GOLDSMITH, J. A., HSIEH, C.-L., ABIONA, O., GRAHAM, B. S. andMCLELLAN, J. S. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science,* 367, 1260-1263.

28. YANG, C., SPIES, C. P. andCOMPANS, R. W. 1995. The human and simian immunodeficiency virus envelope glycoprotein transmembrane subunits are palmitoylated. *Proceedings of the National Academy of Sciences,* 92, 9871-9875.

29. ZHANG, X., LI, M., ZHANG, B., CHEN, T., LV, D., XIA, P., SUN, Z., SHENTU, X., CHEN, H. andLI, L. 2020. The N gene of SARS-CoV-2 was the main positive component in repositive samples from a cohort of COVID-19 patients in Wuhan, China. *Clinica Chimica Acta,* 511, 291-297.