




NeoCoV Is Closer to MERS-CoV than SARS-CoV

Mohamed M Hassan¹ , Mohamed A Hussain², Sumaya Kambal¹, Ahmed A Elshikh³, Osama R Gendeel⁴, Siddig A Ahmed⁴, Rami A Altayeb⁴, Abdelhafiz MA Muhajir⁴ and Sofia B Mohamed¹ 

¹Bioinformatics and Biostatistics Department, National University Biomedical Research Institute, National University, Khartoum, Sudan. ²Department of Pharmaceutical Microbiology, International University of Africa, Khartoum, Sudan. ³Department of Microbiology, Faculty of Pure and Applied Sciences, International University of Africa, Khartoum, Sudan. ⁴Faculty of Science and Technology, Omdurman Islamic University, Omdurman, Sudan.

Infectious Diseases: Research and Treatment
Volume 13: 1–21
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1178633720930711


ABSTRACT: Recently, Coronavirus has been given considerable attention from the biomedical community based on the emergence and isolation of a deadly coronavirus infecting human. To understand the behavior of the newly emerging MERS-CoV requires knowledge at different levels (epidemiologic, antigenic, and pathogenic), and this knowledge can be generated from the most related viruses. In this study, we aimed to compare between 3 species of Coronavirus, namely Middle East Respiratory Syndrome (MERS-CoV), Severe Acute Respiratory Syndrome (SARS-CoV), and NeoCoV regarding whole genomes and 6 similar proteins (E, M, N, S, ORF1a, and ORF1ab) using different bioinformatics tools to provide a better understanding of the relationship between the 3 viruses at the nucleotide and amino acids levels. All sequences have been retrieved from National Center for Biotechnology Information (NCBI). Regards to target genomes' phylogenetic analysis showed that MERS and SARS-CoVs were closer to each other compared with NeoCoV, and the last has the longest relative time. We found that all phylogenetic methods in addition to all parameters (physical and chemical properties of amino acids such as the number of amino acid, molecular weight, atomic composition, theoretical pI, and structural formula) indicated that NeoCoV proteins were the most related to MERS-CoV one. All phylogenetic trees (by both maximum-likelihood and neighbor-joining methods) indicated that NeoCoV proteins have less evolutionary changes except for ORF1a by just maximum-likelihood method. Our results indicated high similarity between viral structural proteins which are responsible for viral infectivity; therefore, we expect that NeoCoV sooner may appear in human-related infection.

KEYWORDS: MERS-CoV, SARS-CoV, NeoCoV, 6 proteins, bioinformatics analysis, evolutionary study, coronaviruses

RECEIVED: February 28, 2020. **ACCEPTED:** May 7, 2020.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Mohamed M. Hassan, Bioinformatics and Biostatistics Department, National University Biomedical Research Institute, National University, P.O. Box 3738 11111 Khartoum, Sudan. Email: mhassan0210@gmail.com.

Introduction

Coronaviruses (CoVs) have a large-scale spreading worldwide and infect human and various animal hosts, causing diseases which range from mostly upper respiratory tract infections in humans to gastrointestinal tract infections, encephalitis, and demyelination in animals and can be lethal.¹ Bat CoVs have been given extraordinary consideration as 2 emerged CoVs have been linked to unpredicted human disease outbreaks in the 21st century resulted in high mortality rate and extremely economic disruption.² These 2 viruses are the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) which were suggested to be originated from bats.² Based on genotypic and serological characterization, the International Committee for Taxonomy of Viruses (ICTV) has reported 4 CoVs genera, namely *Alphacoronaviruses*, *Betacoronaviruses*, *Gammacoronaviruses*, and *Deltacoronaviruses*.³ The CoVs are enveloped with positive-sense single-stranded RNA genomes ranging between 26.2 and 31.7 kilobyte (kb) length and they are the largest among known RNA viruses.⁴ The genome of most CoVs encodes 2 replicates and expressed in a form of 2

polyproteins. The first is open reading frame (ORF) 1a, measuring approximately 450 kDa, and the second is ORF1ab, measuring approximately 750 kDa. These polyproteins have processed into numbers of nonstructural (NS) proteins and 4 structural proteins, namely spike (S) protein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein.⁵ These structural proteins are important components as they play key roles in CoV infectivity in which the integral membrane protein “M” adapts a region of membrane for virus assembly and captures other structural proteins at the budding site, the “N” protein chaperones and protects the viral RNA genome, “S” protein which consists of 3 copies of the S glycoprotein promoting receptor-binding and membrane fusion, and the small membrane protein “E” which presents in sub-stoichiometric amounts and acts as budding enhancer.⁶

It has previously been identified that MERS-CoV is a beta CoV causing high morbidity and mortality in humans.⁷ In Saudi Arabia, it has been detected in a patient clinically diagnosed with a severe respiratory infection.⁸ In June 11, 2014, MERS-CoV infection has been diagnosed in 699 patients mainly from the Arabian Peninsula with a case fatality rate



potentially exceeding the rate reported during the SARS-CoV pandemic.⁹ A research has reported a high rate of neutralizing antibodies against MERS-CoV found in camels in the Arabian Peninsula showing high relationship at genetic level to those from human cases suggesting that these camels are constituting the source of human infections.¹⁰ At molecular level, the CoVs have a high frequency of recombination due to their unique replication mechanism which increases the propensity to result in high rates of mutation allowing the viruses to acclimatize to new hosts and ecological niches.¹¹ De Benedictis et al have characterized small genomic sequence fragments of bat CoVs (BtCoVs) that were closely related to MERS-CoV and suggested that MERS-CoV ancestors may have evolved in bats.¹² In China, since 2002, SARS-CoV was implicated as a causative agent of SARS and caused atypical pneumonia that spread rapidly throughout parts of Asia, North America, and Europe during 2002 to 2003 with cases having been reported in 30 countries.¹³ According to the World Health Organization report, the mortality rate of SARS-CoV was more than 10%.⁵ Close person-to-person contact has been shown to be the major transmission way of SARS-CoV principally via contact with aerosolized droplets or other bodily fluids.¹⁴ Shortly after SARS-CoV outbreak and the subsequent implication of bats as reservoir hosts of the causative agent, CoV drove numerous studies on bats and the viruses they harbor. A specimen from *Neoromicia cf. zuluensis* bat in 2011 yielded a novel betacoronavirus called NeoCoV.¹⁵

According to the Ndapewa Ithete and his colleagues results (in 2013), NeoCoV differed from MERS-CoV by only one amino acid (a.a) exchange (0.3%) in the translated 816-nt *RdRp* gene fragment and by only a 10.9% a.a sequence distance in the gene that encodes the glycoprotein responsible for CoV attachment and cellular entry. Thus, NeoCoV was much more related to MERS-CoV than any other known virus.¹⁶ Victor Max Corman et al¹⁰ reported that 85% of the NeoCoV genome was identical to MERS-CoV at the nucleotide level; therefore, NeoCoV shared essential details of genome architecture with MERS-CoV and thus they have suggested that NeoCoV and MERS-CoV belonged to one viral species. The presence of a genetically divergent S1 subunit within the NeoCoV spike gene indicated that intra-spike recombination events may have been involved in the emergence of MERS-CoV.⁹ Despite the clinical similarities between MERS and SARS, MERS-CoV is distinct from SARS-CoV in several biological aspects such as it uses a distinct receptor (DPP4) and was classified as a “generalist” CoV which enable it to infect a broad range of cells in culture.⁷

In this study, we have attempted to provide a better understanding of the relationship between MERS-CoV, SARS-CoV, and NeoCoV at the level of amino acids regarding 6 similar proteins, including E, M, N, S, ORF1a, and ORF1ab,

using different bioinformatics tools. The leading force for this study was the previous studies which constructed phylogenetic tree between different species of Coronaviridae based on either structural protein and nonstructural protein or whole genome, and they have found that there was some relationship between MERS-CoV and SARS-CoV, while others studied the relationship between MERS-CoV and NeoCoV but there was no study included MERS-CoV, SARS-CoV, and NeoCoV in the same study to know whose is the most related to whom. Bioinformatics tools and Phylogenetic analysis enables us to understand relationships between ancestral sequences and its descendants.

Materials and Methods

Bioinformatics processing and data analysis

In this study, genome sequences of the 3 target species of CoV were retrieved from the National Center for Biotechnology Information (NCBI; genome and nucleotide databases; <https://www.ncbi.nlm.nih.gov/genome>, <https://www.ncbi.nlm.nih.gov/nucleotide>), namely MERS-CoV (genome ID: 31360), SARS-CoV (genome ID: 10320), and NeoCoV (genome ID: KC869678). However, 4 structural proteins, E, S, N, and M, and 2 NS proteins, ORF1a and ORF1ab, of each species were obtained from the NCBI protein database (www.ncbi.nlm.nih.gov/Protein/). Table 1 presents general information about all retrieved both nucleotide and protein sequences. These Genome and protein sequences were then subjected for comparison using different bioinformatics prediction tools.

Compute nucleotide composition and pair-wise alignments

Nucleotide composition of the target genomes (MERS-CoV, SARS-CoV, and NeoCoV) was calculated as shown in Table 2 using Molecular Evolutionary Genetics Analysis Software Version 7.0 (MEGA7; <https://www.megasoftware.net/home>). Furthermore, pairwise alignment was done for each pair of target genomes using BLAST Needleman-Wunsch Global Align Nucleotide Sequences (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) as it is presented in Figure 1.

Multiple sequence alignment and phylogenetic tree

For the purpose of protein sequences comparison, first, the Multiple Sequence Alignments (MSA) was done using the Clustal method implemented in Clustal Omega tool (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Following the alignment, phylogenetic relationships were depicted in phylogram using distance matrix methods (Neighbor-Joining [NJ] and Unweighted Pair Group Method with Arithmetic mean [UPGMA]) in Phylogeny server (http://www.ebi.ac.uk/Tools/phylogeny/clustalw2_phylogeny/).^{17,18} Once trees were

constructed, they were viewed by TreeDyn viewer tool (<http://www.treedyn.org/>) as shown in boxes B and C in Figures 4, 6, 8, 10, 12, and 14. By the same token, second scenario was as follows, and MSA was done by Multiple Sequence Comparison by Log-Expectation (MUSCLE) method using Muscle online tool (<https://www.ebi.ac.uk/Tools/msa/muscle/>). After that, alignment results in Phylip or Clustal format were subjected to Gblocks program version 0.91b (alignment curation tool). Furthermore, PhyML 3.0 (using maximum-likelihood method) and Protpars (using Parsimony method) were used to generate Newick format tree files which have been viewed by TreeDyn viewer tool¹⁹ as shown in boxes A and D in Figures 4, 6, 8, 10, 12, and 14. Previous tools are available at Gblocks (http://molevol.cmima.csic.es/castrejana/Gblocks_server.html), PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>), and Protpars (<http://www.trex.uqam.ca/>). The third scenario was revolved around constructing ultrametric phylogenetic trees using Muscle method for multiple sequence alignments and RelTime method to generate the tree. Then, Time-trees were generated by the Molecular Evolutionary Genetics Analysis software Version 7.0 (MEGA7; <https://www.megasoftware.net/home>). Figures 2 and 3 show the constructed phylogenetic trees of this scenario that have been done by MEGA7.

Calculate physical and chemical parameters

To determine physical and chemical properties of the protein sequence, ProtParam tool <http://web.expasy.org/protparam/> has been used (which gives the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL databases or for a user entered sequence. The computed parameters are the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) as presented in Tables 3 to 14.²⁸

Pair-wise alignment for protein sequences (primary structure)

Same type of protein sequences of the CoV species of interest was compared using Basic Local Alignment Search Tool (BLAST; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Pairwise alignment was done to determine the matched regions and the number of identical/similar amino acids as described in Table 15.

Proteins secondary structure prediction

For the purpose of converting the primary protein structure to secondary protein structure, GOR IV Tool has been used

(version 4.0; https://npsa-prabi.ibcp.fr/cgi-bin/npsa_auto-mat.pl?page=npsa_gor4.html), which was based on the information theory which gives the 2 outputs. The first output comprised the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand, and C=coil. The second presents probability values for each secondary structure at each amino acid position. The program gives the predicted secondary structure with the highest probability compatible with a predicted helix segment of at least 4 residues and a predicted extended segment of at least 2 residues²⁹ as shown in Figures 16 to 21.

Proteins homology modeling prediction

The three-dimensional (3D) structure prediction of target structural proteins (E, M, N, and S) was obtained by using CPH models and RaptorX servers. In CPH server, the template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions (<http://www.cbs.dtu.dk/services/CPHmodels/>).³⁰ Proteins that do not have close 3D structures were subjected to RaptorX server, which was developed by Xu group. It is excelling at predicting 3D structures for protein sequences without close homologs in the Protein Data Bank (PDB). Additionally, it predicts secondary and tertiary structures, contacts, solvent accessibility, disordered regions, and binding sites with many confidence scores to indicate the quality of the predicted 3D model including *P* value for the relative global quality, global distance test (GDT) and un-normalized GDT (uGDT) for the absolute global quality, and modeling error at each residue.³¹ Then, for the purpose of protein 3D structures visualization, Chimera software v1.8 has been used (<http://www.cgl.ucsf.edu/chimera/>). It is a high-quality extensible molecular graphics program designed to maximize interactive visualization, analysis system, and related data³² as shown in Figures 22 to 25.

Measuring of template modeling score (TM-score) using Zhang-lab tool

The TM-score is defined to assess the topological similarity of 2 protein structures.³³ Zhang tool is designed to solve 2 major problems in the traditional metrics such as root mean square deviation (RMSD): (1) TM-score measures the global fold similarity and is less sensitive to the local structural variations and (2) magnitude of TM-score for random structure pairs is length-independent. TM-score has the value between 0 and 1, where 1 indicates a perfect match between 2 structures. Following strict statistics of structures in the PDB, scores below 0.17 correspond to randomly chosen unrelated proteins, whereas with a score higher than 0.5 assume generally the same fold in SCOP/CATH (<https://zhanglab.ccmb.med.umich.edu/TM-score/>)³⁴ (Table 16).

Results

Table 1. General information of retrieved genomes and protein sequences.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Genome ID	31360	10320	KC869678
Genome size	30.12 (kb)	29.75 (kb)	30.111 (kb)
Protein-coding gene	11	14	11
<i>Protein ID</i>			
Envelope (E)	YP_009047209.1	NP_828854.1	AIG13101.1
Membrane (M)	YP_009047210.1	NP_828855.1	AIG13102.1
Nucleocapsid (N)	YP_009047211.1	NP_828858.1	AIG13103.1
Spike (S)	YP_009047204.1	NP_828851.1	AGY29650.2
Open reading frame (ORF)1a	YP_009047203.1	NP_828850.1	AIG13097.1
Open reading frame (ORF)1ab	YP_009047202.1	NP_828849.2	AGR87639.3

All information obtained from NCBI (National Center for Biotechnology Information) database (<https://www.ncbi.nlm.nih.gov/>).

Table 2. Nucleotide composition of target genomes.

VIRUSES	T	C	A	G	TOTAL
SARS-CoV	30.7	20.0	28.5	20.8	29751
MERS-CoV	35.6	15.2	27.6	21.7	30738
NeoCoV	33.2	19.2	26.7	21.0	30108
Avg.	33.2	18.1	27.6	21.2	30199

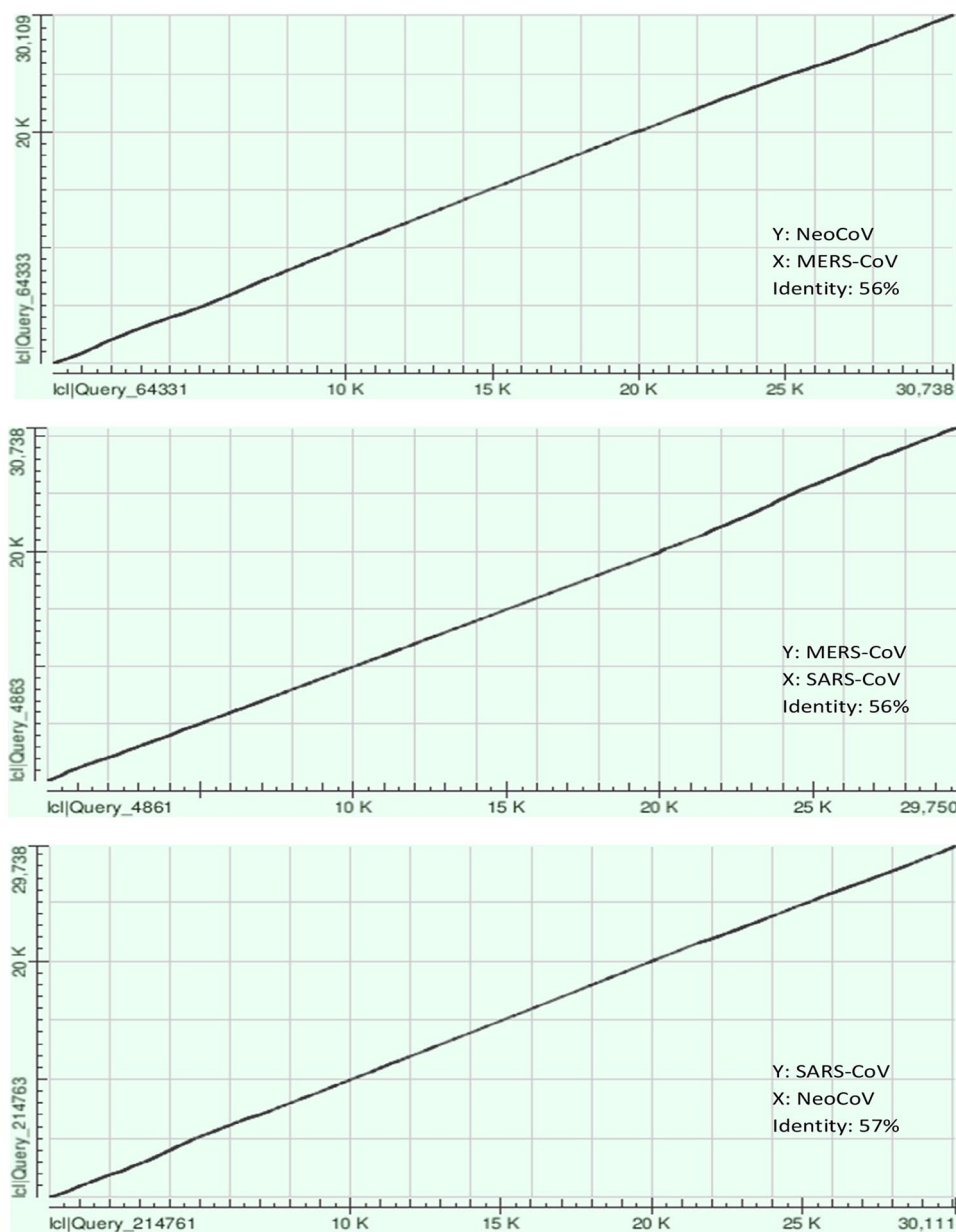


Figure 1. Global pairwise alignment results by Needleman-Wunsch method.

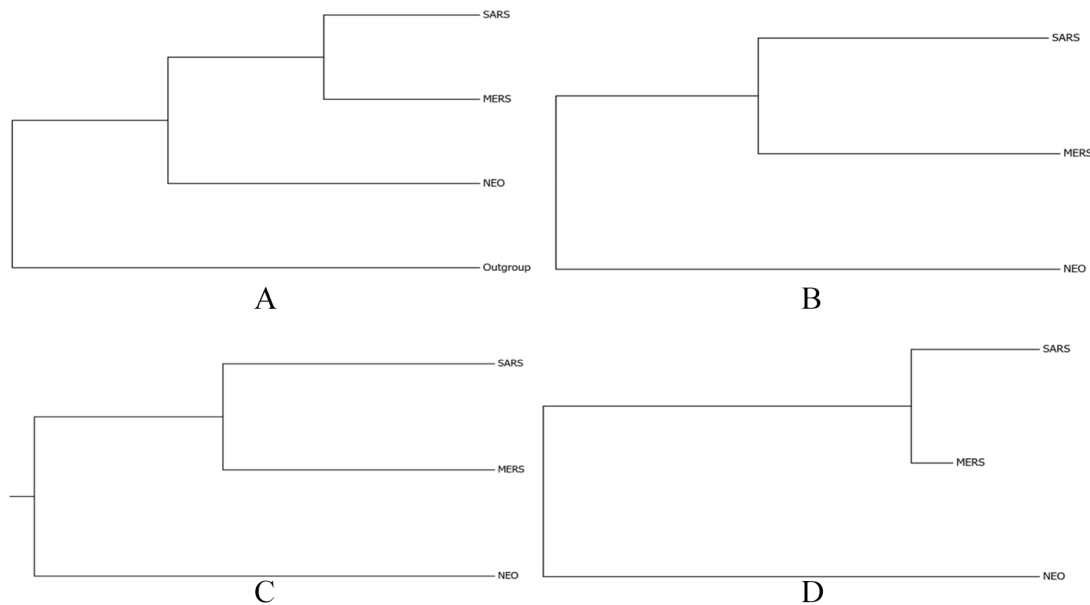


Figure 2. Phylogenetic trees comparing whole genomes of coronavirus species MERS-CoV, SARS-CoV, and NeoCoV, and trees “A-D” were built using different methods. (A) The evolutionary history was inferred using the Maximum Parsimony (MP) method. The most parsimonious tree with length=42663 is shown. The consistency index is 0.980475 (0.613278), the retention index is 0.369417 (0.369417), and the composite index is 0.362204 (0.226555) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm^{20(p126)} with search level 0 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 4 nucleotide sequences. There were a total of 29693 positions in the final dataset. (B) The evolutionary history was inferred using the Neighbor-Joining method.²¹ The optimal tree with the sum of branch length=18.91227594 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method²² and are in the units of the number of base substitutions per site. The analysis involved three nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 29690 positions in the final dataset. (C) The evolutionary history was inferred using the UPGMA method.²³ The optimal tree with the sum of branch length=18.91227594 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method²² and are in the units of the number of base substitutions per site. The analysis involved 3 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 29690 positions in the final dataset. (D) The evolutionary history was inferred by using the Maximum-Likelihood method based on the Tamura-Nei model.²⁴ The tree with the highest log likelihood (-121 024.68) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved three nucleotide sequences. There were a total of 29693 positions in the final dataset. All MSA of used sequences was curated by using Gblocks, and evolutionary analyses were conducted in MEGA7.²⁵ MSA indicates Multiple Sequence Alignments, UPGMA, Unweighted Pair Group Method with Arithmetic Mean.

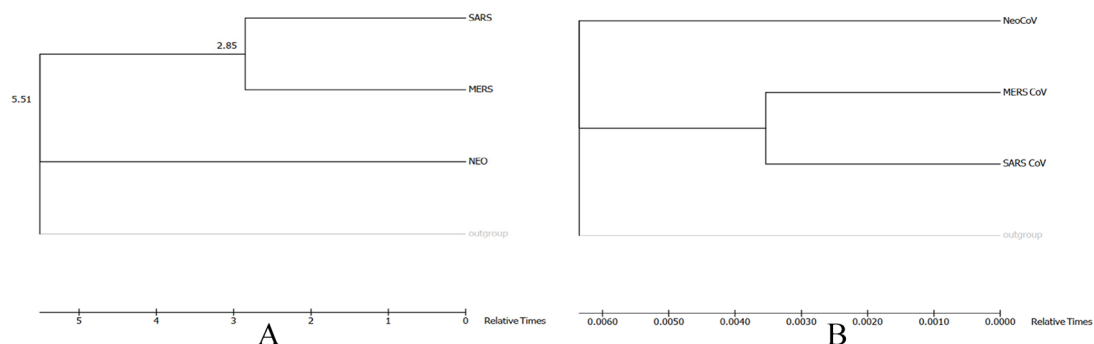


Figure 3. Molecular phylogenetic analysis of MERS-CoV, SARS-CoV, and NeoCoV genomes. The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Tamura-Nei model.²⁴ (A) The estimated log likelihood value of the topology shown is 135 729.24. The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 nucleotide sequences. There were a total of 30738 positions in the final dataset. (B) The estimated log likelihood value of the topology shown is 79 370.24. The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 nucleotide sequences. There were a total of 22620 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵ (A) Without MSA curation. (B) With MSA curation. MSA indicates Multiple Sequence Alignments.

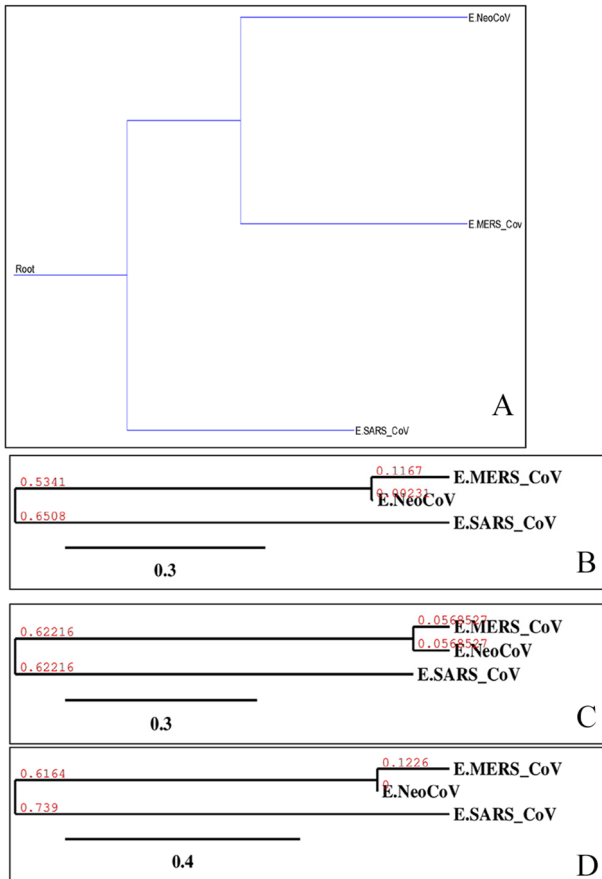


Figure 4. Phylogenetic of “E” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees (A)-(D) were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum-Likelihood. UPGMA indicates Unweighted Pair Group Method with Arithmetic Mean.

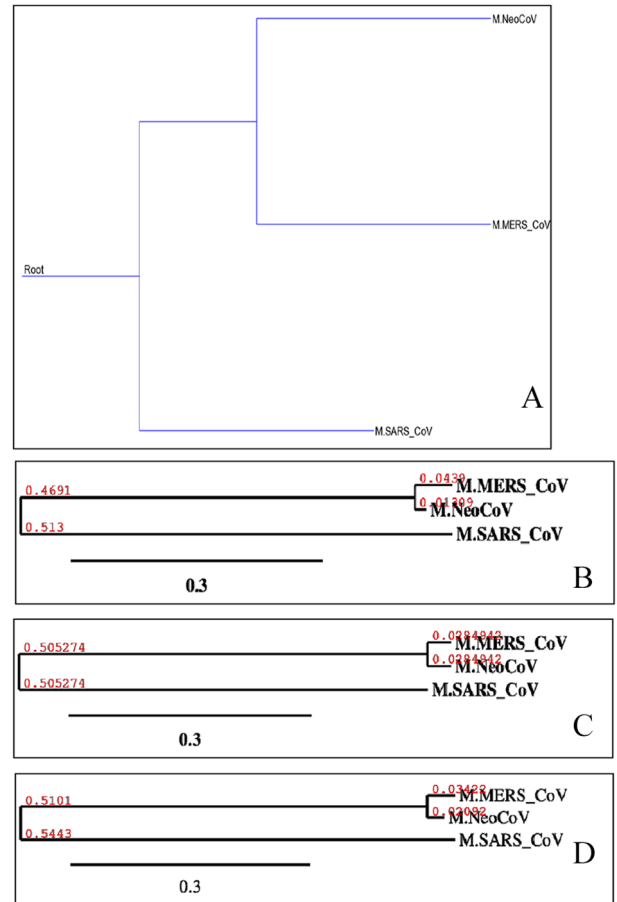


Figure 6. Shows phylogenetics of “M” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees “A-D” were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum Likelihood. UPGMA, Unweighted Pair Group Method with Arithmetic Mean.

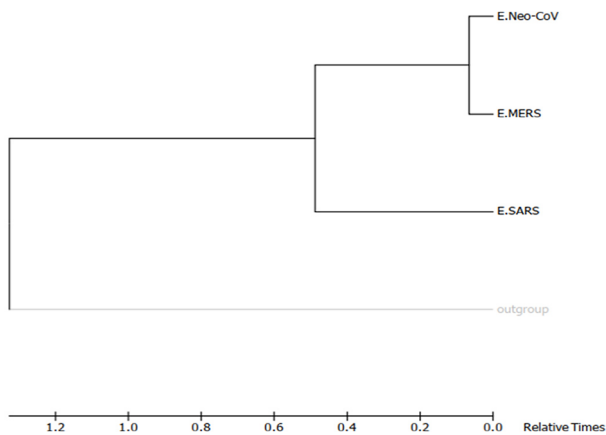


Figure 5. Molecular phylogenetic tree of target coronaviruses “E” proteins by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -703.32 . The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 86 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

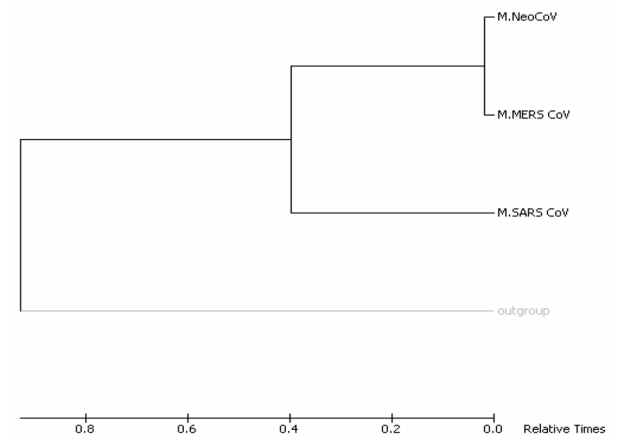


Figure 7. Molecular phylogenetic analysis by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -1802.98 . The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 244 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

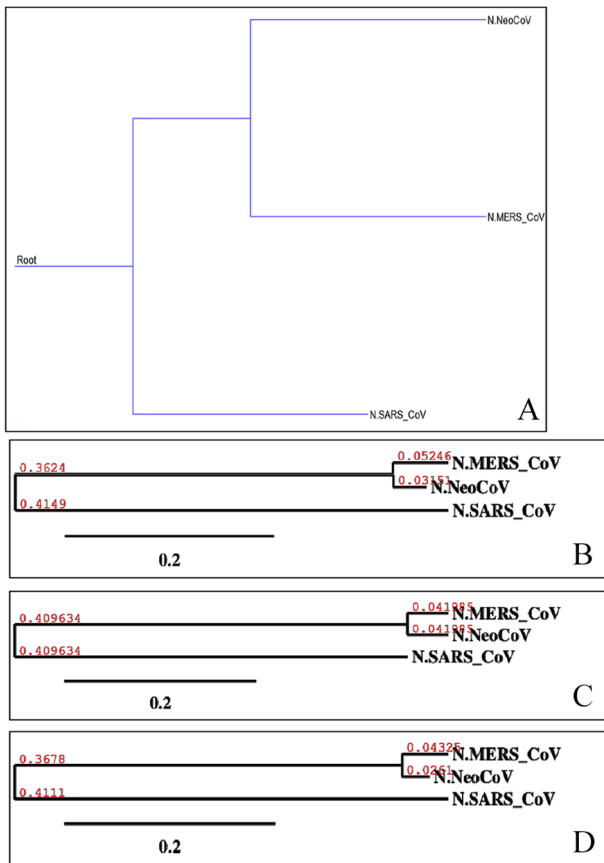


Figure 8. Phylogenetic of “N” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees (A)-(D) were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum Likelihood. UPGMA indicates Unweighted Pair Group Method with Arithmetic Mean.

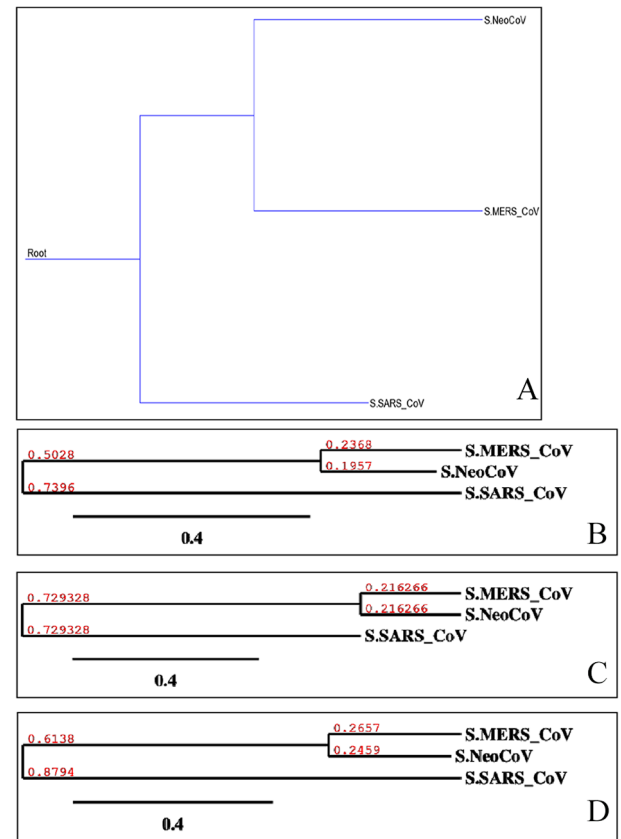


Figure 10. Shows phylogenetic of “S” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees (A)-(D) were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum Likelihood. UPGMA indicates Unweighted Pair Group Method with Arithmetic Mean.

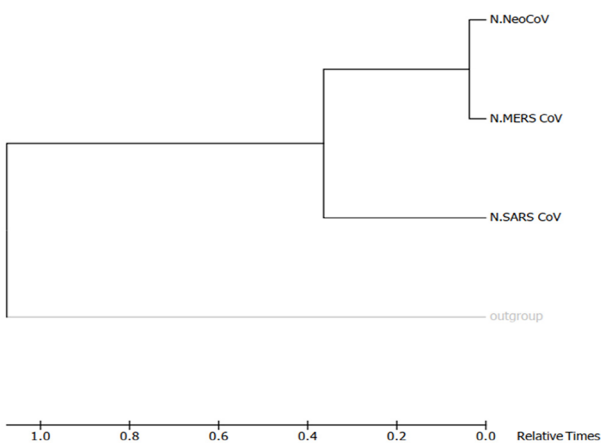


Figure 9. Molecular phylogenetic analysis by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -3425.67 . The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 460 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

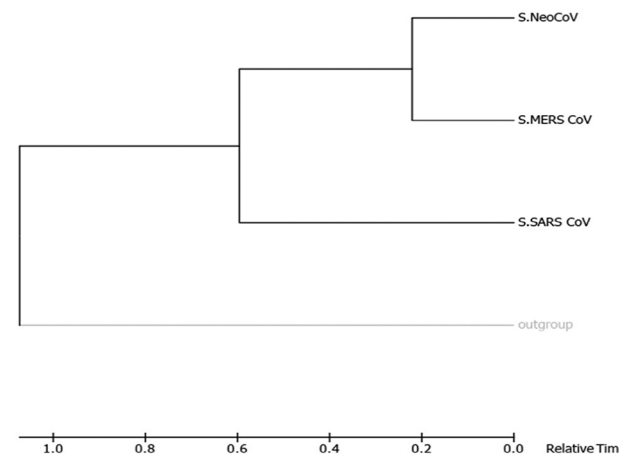


Figure 11. Molecular phylogenetic analysis by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -13052.98 . The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 1544 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

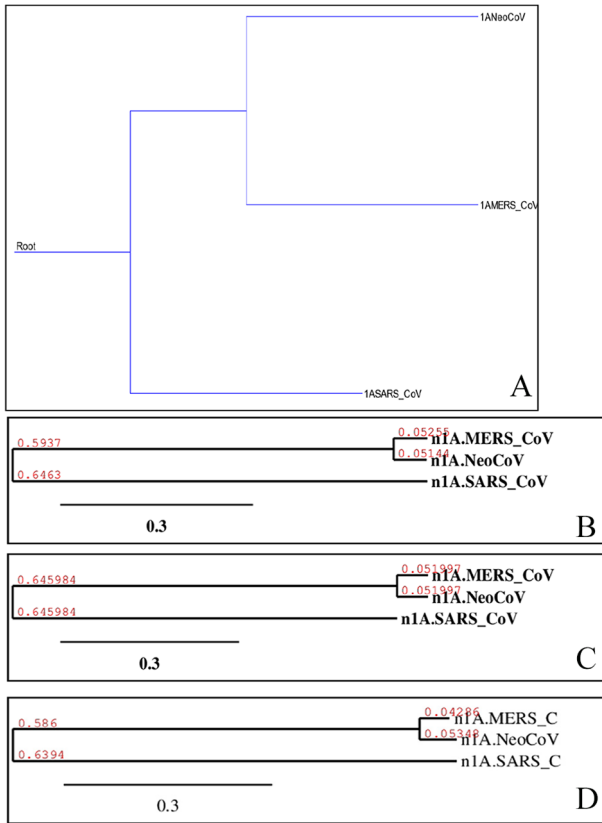


Figure 12. Shows phylogenetic of “1A” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees (A)-(D) were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum Likelihood. UPGMA indicates Unweighted Pair Group Method with Arithmetic Mean.

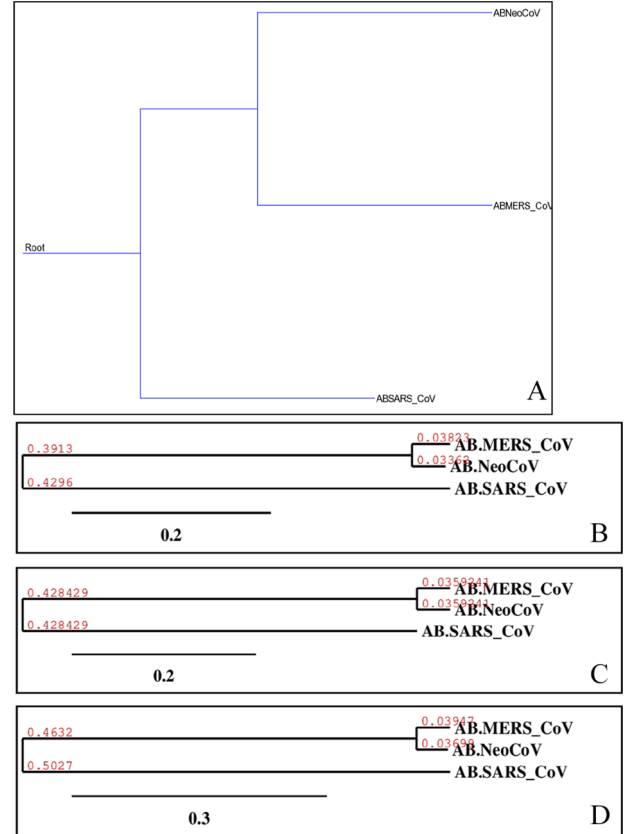


Figure 14. Phylogenetic of “1AB” proteins of target coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV). Trees (A)-(D) were built using different methods, and they are, respectively, Maximum Parsimony, Neighbor-Joining, UPGMA, and Maximum Likelihood. UPGMA indicates Unweighted Pair Group Method with Arithmetic Mean.

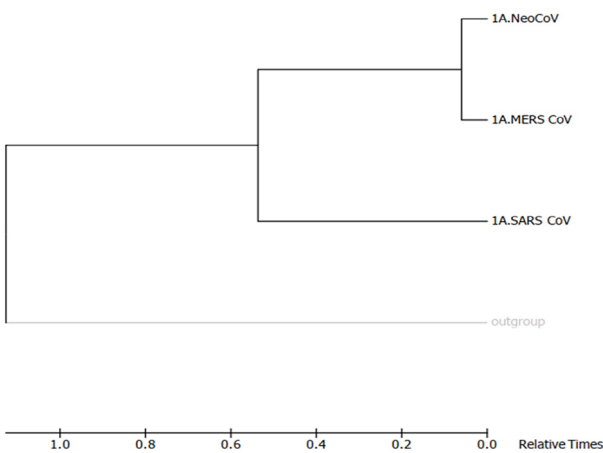


Figure 13. Molecular phylogenetic analysis by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -38685.85. The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 4988 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

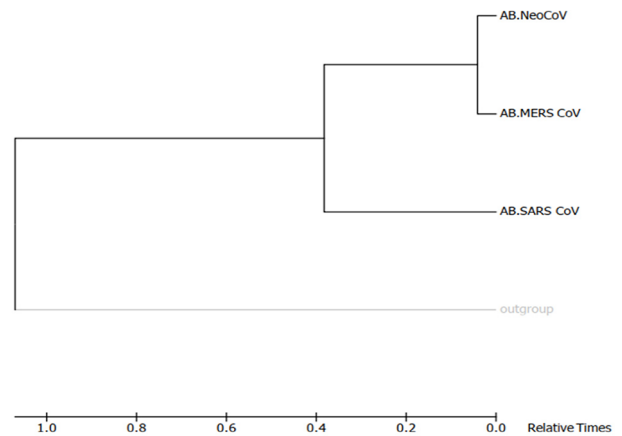


Figure 15. Molecular phylogenetic analysis by Maximum-Likelihood method (timetree). The timetree shown was generated using the RelTime method.²⁶ Divergence times for all branching points in the topology were calculated using the Maximum-Likelihood method based on the Equal Input model.²⁷ The estimated log likelihood value of the topology shown is -59576.47. The tree is drawn to scale, with branch lengths measured in the relative number of substitutions per site. The analysis involved 4 amino acid sequences. There were a total of 8041 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.²⁵

Table 3. Physical and chemical parameters of “E” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	82	76	82
Molecular weight	9354.2	8361.0	9265.1
Theoretical pI	7.64	6.01	7.64
<i>Atomic composition</i>			
Carbon (C)	439	388	432
Hydrogen (H)	677	625	670
Nitrogen (N)	101	89	100
Oxygen (O)	110	106	111
Sulfur (S)	7	4	7
<i>Formula</i>	$C_{439}H_{677}N_{101}O_{110}S_7$	$C_{388}H_{625}N_{89}O_{106}S_4$	$C_{432}H_{670}N_{100}O_{111}S_7$
Total number of atoms	1334	1212	1320
Total number of negatively charged residues (Asp + Glu)	4	4	3
Total number of positively charged residues (Arg + Lys)	5	4	4
Extinction coefficients	-10220 (Abs. 1.093) -9970 (Abs. 1.066)	-6085 (Abs. 0.728) -5960 (Abs. 0.713)	-10220 (Abs. 1.103) -9970 (Abs. 1.076)
Instability index	33.00 (stable)	30.48 (stable)	46.07 (unstable)
Aliphatic index	111.59	145.92	109.15
Grand average of hydropathicity (GRAVY)	0.795	1.141	0.782

Table 4. Amino acid composition of “E” protein.

TYPES OF AMINO ACID	VIRUSES					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	4	4.9%	4	5.3%	4	4.9%
Arg (R)	3	3.7%	2	2.6%	2	2.4%
Asn (N)	3	3.7%	5	6.6%	3	3.7%
Asp (D)	2	2.4%	1	1.3%	0	0.0%
Cys (C)	4	4.9%	3	3.9%	4	4.9%
Gln (Q)	4	4.9%	0	0.0%	6	7.3%
Glu (E)	2	2.4%	3	3.9%	3	3.7%
Gly (G)	3	3.7%	2	2.6%	3	3.7%
His (H)	0	0.0%	0	0.0%	0	0.0%
Ile (I)	4	4.9%	3	3.9%	5	6.1%
Leu (L)	11	13.4%	14	18.2%	8	9.8%
Lys (K)	2	2.4%	2	2.6%	2	2.4%
Met (M)	3	3.7%	1	1.3%	3	3.7%
Phe (F)	8	9.8%	4	5.3%	7	8.5%
Pro (P)	6	7.3%	2	2.6%	6	7.3%
Ser (S)	2	2.4%	7	9.2%	3	3.7%
Thr (T)	7	8.5%	5	6.6%	7	8.5%
Trp (W)	1	1.2%	0	0.0%	1	1.2%
Tyr (Y)	3	3.7%	4	5.3%	3	3.7%
Val (V)	10	12.2%	14	18.2%	12	14.6%

Table 5. Physical and chemical parameters of “M” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	219	221	219
Molecular weight	24536.8	25060.5	24527.7
Theoretical pI	9.27	9.63	9.25
<i>Atomic composition</i>			
Carbon (C)	1130	1155	1129
Hydrogen (H)	1752	1809	1753
Nitrogen (N)	282	303	281
Oxygen (O)	302	300	305
Sulfur (S)	13	10	12
<i>Formula</i>	$C_{1130}H_{1752}N_{282}O_{302}S_{13}$	$C_{1155}H_{1809}N_{303}O_{300}S_{10}$	$C_{1129}H_{1753}N_{281}O_{305}S_{12}$
Total number of atoms	3479	3577	3480
Total number of negatively charged residues (Asp + Glu)	11	13	11
Total number of positively charged residues (Arg + Lys)	16	21	16
Extinction coefficients	-53525 (Abs. 2.181) -53400 (Abs. 2.176)	-52035 (Abs. 2.076) -51910 (Abs. 2.071)	-55015 (Abs. 2.243) -54890 (Abs. 2.238)
Instability index	43.67 (unstable)	30.44 (stable)	42.75 (unstable)
Aliphatic index	104.61	116.06	105.94
Grand average of hydropathicity (GRAVY)	0.436	0.417	0.407

Table 6. Amino acid composition of “M” protein.

TYPES OF AMINO ACID	VIRUS					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	19	8.7%	19	8.6%	17	7.8%
Arg (R)	9	4.1%	15	6.8%	9	4.1%
Asn (N)	10	4.6%	13	5.9%	10	4.6%
Asp (D)	6	2.7%	6	2.7%	6	2.7%
Cys (C)	2	0.9%	3	1.4%	2	0.9%
Gln (Q)	6	2.7%	5	2.3%	7	3.2%
Glu (E)	5	2.3%	7	3.2%	5	2.3%
Gly (G)	11	5.0%	15	6.8%	12	5.5%
His (H)	3	1.4%	3	1.4%	2	0.9%
Ile (I)	18	8.2%	18	8.1%	20	9.1%
Leu (L)	21	9.6%	31	14.0%	21	9.6%
Lys (K)	7	3.2%	6	2.7%	7	3.2%
Met (M)	11	5.0%	7	3.2%	10	4.6%
Phe (F)	10	4.6%	11	5.0%	9	4.1%
Pro (P)	10	4.6%	5	2.3%	10	4.6%
Ser (S)	20	9.1%	12	5.4%	22	10.0%
Thr (T)	14	6.4%	13	5.9%	13	5.9%
Trp (W)	7	3.2%	7	3.2%	7	3.2%
Tyr (Y)	10	4.6%	9	4.1%	11	5.0%
Val (V)	20	9.1%	16	7.2%	19	8.7%

Table 7. Physical and chemical parameters of “N” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	413	422	414
Molecular weight	45062.3	46025.0	44863.0
Theoretical pI	10.05	10.11	10.07
<i>Atomic composition</i>			
Carbon (C)	1966	1985	1955
Hydrogen (H)	3104	3150	3083
Nitrogen (N)	594	618	593
Oxygen (O)	611	633	609
Sulfur (S)	7	7	7
<i>Formula</i>	$C_{1966}H_{3104}N_{594}O_{611}S_7$	$C_{1985}H_{3150}N_{618}O_{633}S_7$	$C_{1955}H_{3083}N_{593}O_{609}S_7$
Total number of atoms	6282	6393	6247
Total number of negatively charged residues (Asp + Glu)	33	36	33
Total number of positively charged residues (Arg + Lys)	55	60	56
Extinction coefficients	-47900 (Abs. 1.063)	-43890 (Abs. 0.954)	-47900 (Abs. 1.068)
Instability index	48.62 (unstable)	52.28 (unstable)	50.56 (unstable)
Aliphatic index	57.00	49.81	53.60
Grand average of hydropathicity (GRAVY)	-0.866	-1.027	-0.883

Table 8. The amino acid composition of “N” protein.

TYPES OF AMINO ACID	VIRUSES					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	33	8.0%	34	8.1%	40	9.7%
Arg (R)	26	6.3%	31	7.3%	26	6.3%
Asn (N)	32	7.7%	25	5.9%	32	7.7%
Asp (D)	20	4.8%	22	5.2%	17	4.1%
Cys (C)	0	0.0%	0	0.0%	0	0.0%
Gln (Q)	24	5.8%	34	8.1%	23	5.6%
Glu (E)	13	3.1%	14	3.3%	16	3.9%
Gly (G)	38	9.2%	45	10.7%	38	9.2%
His (H)	6	1.5%	5	1.2%	5	1.2%
Ile (I)	13	3.1%	11	2.6%	9	2.2%
Leu (L)	27	6.5%	26	6.2%	25	6.0%
Lys (K)	29	7.0%	29	6.9%	30	7.2%
Met (M)	7	1.7%	7	1.7%	7	1.7%
Phe (F)	14	3.4%	13	3.1%	14	3.4%
Pro (P)	34	8.2%	31	7.3%	36	8.7%
Ser (S)	35	8.5%	35	8.3%	36	8.7%
Thr (T)	30	7.3%	33	7.8%	27	6.5%
Trp (W)	6	1.5%	5	1.2%	6	1.4%
Tyr (Y)		2.4%	11	2.6%	10	2.4%
Val (V)	16	3.9%	11	2.6%	17	4.1%

Table 9. Physical and chemical parameters of “S” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	1353	1255	1344
Molecular weight	149368.0	139109.1	148690.7
Theoretical pI	5.70	5.56	5.68
<i>Atomic composition</i>			
Carbon (C)	6682	6252	6642
Hydrogen (H)	10245	9593	10137
Nitrogen (N)	1735	1609	1739
Oxygen (O)	2029	1870	2022
Sulfur (S)	63	59	62
<i>Formula</i>	$C_{6682}H_{10245}N_{1735}O_{2029}S_{63}$	$C_{6252}H_{9593}N_{1609}O_{1870}S_{59}$	$C_{6642}H_{10137}N_{1739}O_{2022}S_{62}$
Total number of atoms	20754	19383	20602
Total number of negatively charged residues (Asp + Glu)	112	115	112
Total number of positively charged residues (Arg + Lys)	95	99	94
Extinction coefficients	-170865 (Abs. 1.144) -168240 (Abs. 1.126)	-143335 (Abs. 1.030) -140960 (Abs. 1.013)	-183815 (Abs. 1.236) -181190 (Abs. 1.219)
Instability index	36.60 (stable)	32.42 (stable)	31.01 (stable)
Aliphatic index	82.71	82.80	80.48
Grand average of hydropathicity (GRAVY)	-0.074	-0.043	-0.137

Table 10. Amino acid composition of “S” protein.

TYPE OF AMINO ACID	VIRUSES					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	88	6.5%	85	6.8%	98	7.3%
Arg (R)	44	3.3%	39	3.1%	43	3.2%
Asn (N)	77	5.7%	81	6.5%	96	7.1%
Asp (D)	66	4.9%	73	5.8%	63	4.7%
Cys (C)	42	3.1%	39	3.1%	43	3.2%
Gln (Q)	72	5.3%	55	4.4%	66	4.9%
Glu (E)	46	3.4%	42	3.3%	49	3.6%
Gly (G)	92	6.8%	79	6.3%	93	6.9%
His (H)	20	1.5%	15	1.2%	21	1.6%
Ile (I)	73	5.4%	78	6.2%	75	5.6%
Leu (L)	120	8.9%	99	7.9%	114	8.5%
Lys (K)	51	3.8%	60	4.8%	51	3.8%
Met (M)	21	1.6%	20	1.6%	19	1.4%
Phe (F)	71	5.2%	83	6.6%	68	5.1%
Pro (P)	62	4.6%	57	4.5%	58	4.3%
Ser (S)	134	9.9%	95	7.6%	114	8.5%
Thr (T)	92	6.8%	99	7.9%	96	7.1%
Trp (W)	10	0.7%	11	0.9%	11	0.8%
Tyr (Y)	76	5.6%	54	4.3%	81	6.0%
Val (V)	96	7.1%	91	7.3%	85	6.3%

Table 11. Physical and chemical parameters of “ORF1a” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	4391	4382	4394
Molecular weight	485956.4	486372.7	486923.0
Theoretical pI	6.28	5.91	6.19
<i>Atomic composition</i>			
Carbon (C)	21877	21746	21935
Hydrogen (H)	34022	34030	34125
Nitrogen (N)	5638	5656	5619
Oxygen (O)	6412	6468	6431
Sulfur (S)	229	255	233
<i>Formula</i>	$C_{21877}H_{34022}N_{5638}O_{6412}S_{229}$	$C_{21746}H_{34030}N_{5656}O_{6468}S_{255}$	$C_{21935}H_{34125}N_{5619}O_{6431}S_{233}$
Total number of atoms	68178	68155	68343
Total number of negatively charged residues (Asp + Glu)	416	461	421
Total number of positively charged residues (Arg + Lys)	385	404	387
Extinction coefficients	-575415 (Abs. 1.184) -567040 (Abs. 1.167)	-530660 (Abs. 1.091) -521660 (Abs. 1.073)	-583115 (Abs. 1.198) -574490 (Abs. 1.180)
Instability index	34.07 (stable)	35.51 (stable)	32.61 (stable)
Aliphatic index	91.54	89.43	92.12
Grand average of hydropathicity (GRAVY)	0.081	-0.020	0.089

Table 12. Amino acid composition of “ORF1a” proteins.

TYPE OF AMINO ACID	VIRUSES					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	343	7.8%	325	7.4%	330	7.5%
Arg (R)	146	3.3%	146	3.3%	141	3.2%
Asn (N)	209	4.8%	214	4.9%	200	4.6%
Asp (D)	239	5.4%	221	5.0%	240	5.5%
Cys (C)	134	3.1%	144	3.3%	138	3.1%
Gln (Q)	145	3.3%	147	3.4%	152	3.5%
Glu (E)	177	4.0%	240	5.5%	181	4.1%
Gly (G)	255	5.8%	269	6.1%	255	5.8%
His (H)	83	1.9%	86	2.0%	77	1.8%
Ile (I)	197	4.5%	212	4.8%	203	4.6%
Leu (L)	426	9.7%	444	10.1%	429	9.8%
Lys (K)	239	5.4%	248	5.9%	246	5.6%
Met (M)	95	2.2%	111	2.5%	95	2.2%
Phe (F)	219	5.0%	195	4.5%	219	5.0%
Pro (P)	170	3.9%	166	3.8%	164	3.7%
Ser (S)	324	7.4%	298	6.8%	329	7.5%
Thr (T)	314	7.2%	320	7.3%	312	7.1%
Trp (W)	50	1.1%	45	1.0%	50	1.1%
Tyr (Y)	196	4.5%	184	4.2%	201	4.6%
Val (V)	430	9.8%	357	8.1%	432	9.8%

Table 13. Physical and chemical parameters of “ORF1ab” proteins.

DESCRIPTIONS	VIRUSES		
	MERS-COV	SARS-COV	NEOCOV
Number of amino acids	7078	7073	7082
Molecular weight	789461.2	790248.3	790641.1
Theoretical pI	6.47	6.19	6.39
<i>Atomic composition</i>			
Carbon (C)	35532	35380	–
Hydrogen (H)	54937	55002	–
Nitrogen (N)	9183	9262	–
Oxygen (O)	10371	10437	–
Sulfur (S)	398	410	–
<i>Formula</i>	$C_{35532}H_{54937}N_{9183}O_{10371}S_{398}$	$C_{35380}H_{55002}N_{9262}O_{10437}S_{410}$	–
Total number of atoms	110421	110491	–
Total number of negatively charged residues (Asp + Glu)	687	743	692
Total number of positively charged residues (Arg + Lys)	647	674	648
Extinction coefficients	–978520 (Abs. 1.239) –964020 (Abs. 1.221)	–920760 (Abs. 1.165) –906260 (Abs. 1.147)	–986220 (Abs. 1.247) –971470 (Abs. 1.229)
Instability index	34.24 (stable)	33.65 (stable)	33.06 (stable)
Aliphatic index	88.03	87.08	88.06
Grand average of hydropathicity (GRAVY)	0.013	–0.071	0.014

Table 14. Amino acid composition of “ORF1ab” proteins.

TYPES OF AMINO ACID	VIRUS					
	MERS-COV		SARS-COV		NEOCOV	
Ala (A)	524	7.4%	511	7.2%	513	7.2%
Arg (R)	248	3.5%	259	3.7%	244	3.4%
Asn (N)	355	5.0%	366	5.2%	355	5.0%
Asp (D)	400	5.7%	396	5.6%	400	5.6%
Cys (C)	233	3.3%	233	3.3%	237	3.3%
Gln (Q)	226	3.2%	234	3.3%	233	3.3%
Glu (E)	287	4.1%	347	4.9%	292	4.1%
Gly (G)	404	5.7%	419	5.9%	406	5.7%
His (H)	150	2.1%	160	2.3%	143	2.0%
Ile (I)	335	4.7%	343	4.8%	338	4.8%
Leu (L)	645	9.1%	674	9.5%	644	9.1%
Lys (K)	399	5.6%	415	5.9%	404	5.7%
Met (M)	165	2.3%	177	2.5%	165	2.3%
Phe (F)	365	5.2%	331	4.7%	368	5.2%
Pro (P)	274	3.9%	274	3.9%	266	3.8%
Ser (S)	503	7.1%	458	6.5%	502	7.1%
Thr (T)	486	6.9%	495	7.0%	483	6.8%
Trp (W)	81	1.1%	77	1.1%	81	1.1%
Tyr (Y)	348	4.9%	324	4.6%	353	5.0%
Val (V)	650	9.2%		8.2%	653	9.2%

Table 15. Pairwise alignment for protein sequences (primary structure) of coronavirus species.

DESCRIPTIONS	VIRUSES		
	MERS VS SARS	NEO VS SARS	MERS VS NEO
<i>E protein</i>			
Number of matches regions	1	1	1
Identical amino acids	29 (37.7%)	29 (37.7%)	73 (89%)
Similarity amino acids	10 (13%)	13 (16.9%)	7 (8.5%)
Gaps	5 (6.5%)	2 (2.6%)	0 (0%)
<i>M protein</i>			
Number of matched regions	1	1	1
Identical amino acids	89 (43.6%)	93 (42.5%)	207 (94.5%)
Similar amino acids	38 (18.6%)	44 (20.1%)	8 (3.7%)
Gaps	0 (0%)	2 (0.9%)	0 (0%)
<i>N protein</i>			
Number of matched regions	1	1	1
Identical amino acids	187 (51.4%)	199 (49.9%)	378 (91%)
Similarity amino acids	52 (14.3%)	54 (13.5%)	17 (4.1%)
Gaps	16 (4.4%)	25 (6.3%)	3 (0.7%)
<i>S protein</i>			
Number of matched regions	4	3	1
Identical amino acids	408 (33.9%)	417 (34.3%)	868 (63.4%)
Similarity amino acids	200 (16.6%)	192 (15.8%)	179 (13.1%)
Gaps	94 (7.8%)	91 (7.5%)	41 (3%)
<i>ORF1a protein</i>			
Number of matched regions	6	4	1
Identical amino acids	1576 (36%)	1487 (33.9%)	3935 (89.6%)
Similarity amino acids	840 (19.2%)	809 (18.5%)	207 (4.7%)
Gaps	286 (6.5%)	230 (5.2%)	13 (0.3%)
<i>ORF1ab protein</i>			
Number of matched regions	6	6	1
Identical amino acids	3357 (46.8%)	3359 (46.2%)	6559 (92.5%)
Similarity amino acids	1206 (16.8%)	1235 (17%)	245 (3.5%)
Gaps	299 (4.2%)	313 (4.3%)	14 (0.2%)

Abbreviations: MERS, Middle East Respiratory Syndrome; SARS, Severe Acute Respiratory Syndrome.

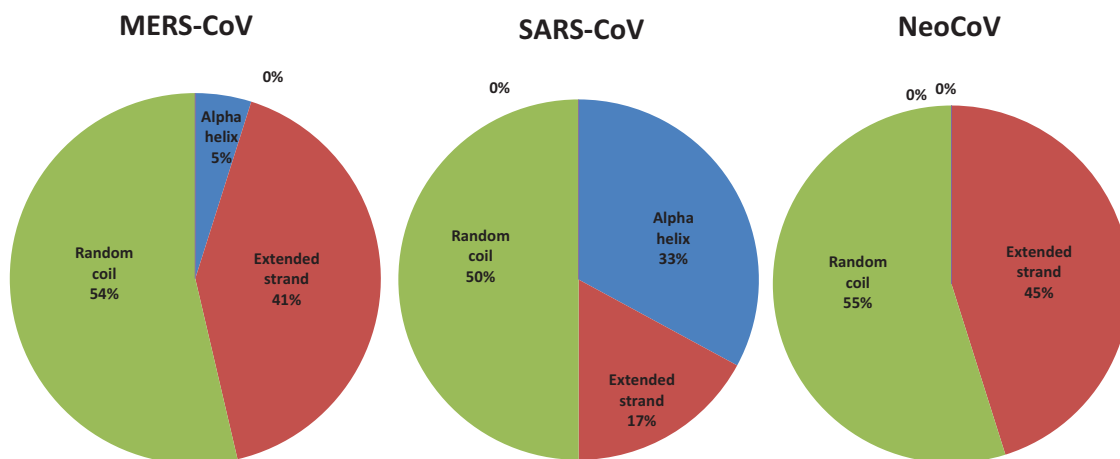


Figure 16. Percent of secondary structure component of E proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.

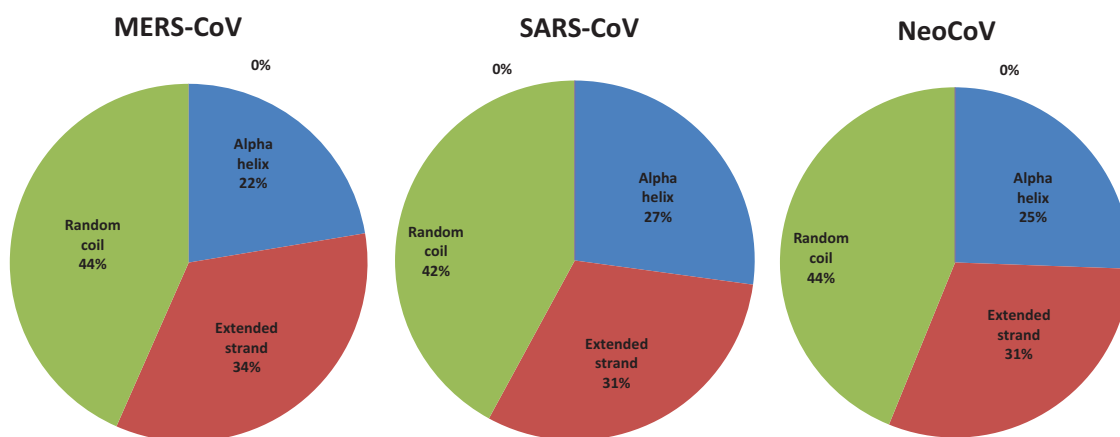


Figure 17. Percent of secondary structure component of M proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.

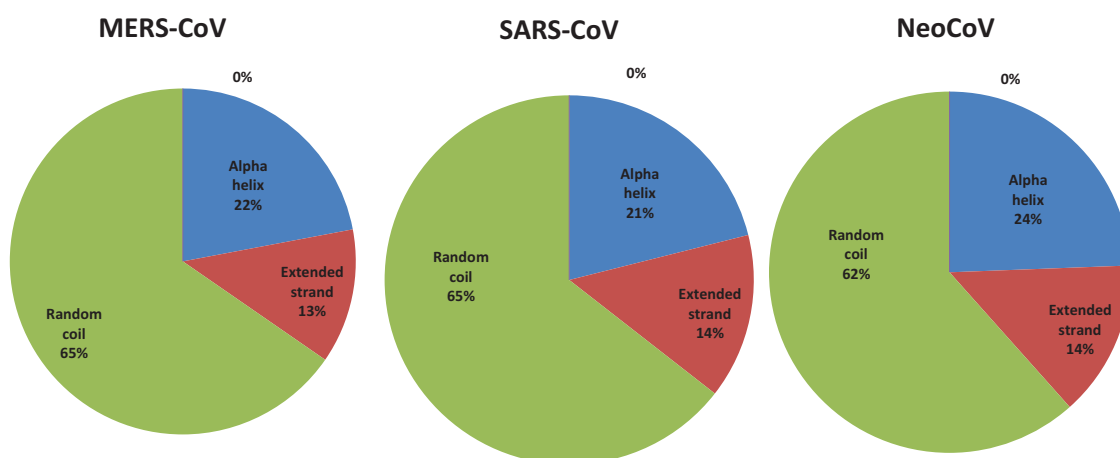


Figure 18. Percent of secondary structure component of N proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.

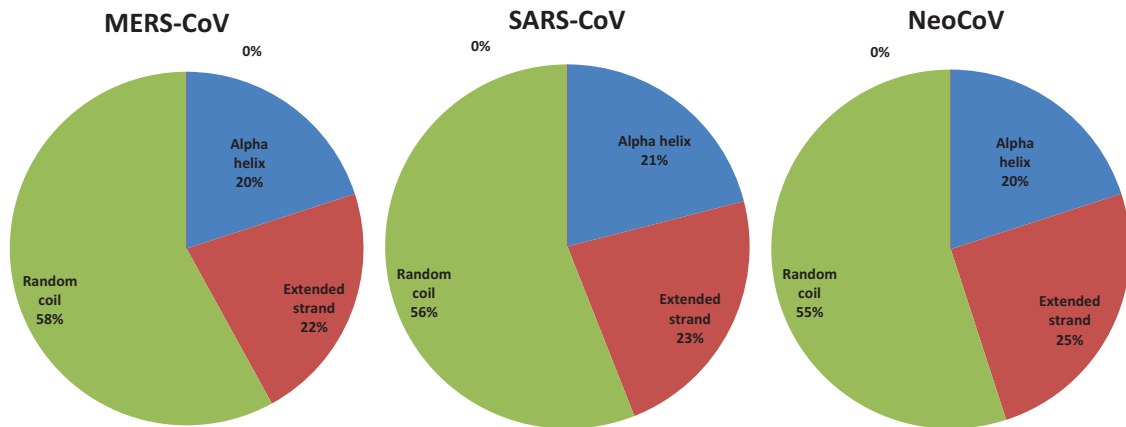


Figure 19. Percent of secondary structure component of S proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.

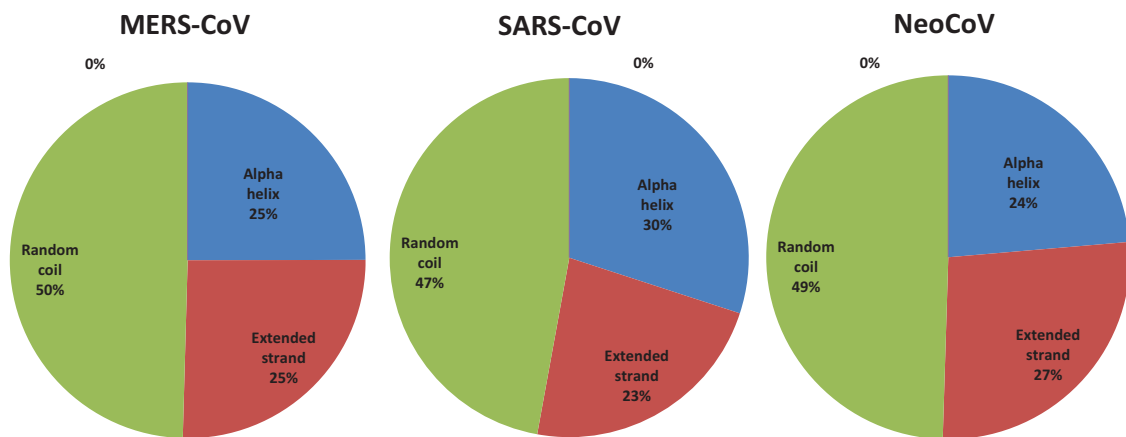


Figure 20. Percent of secondary structure component of ORF1a proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.

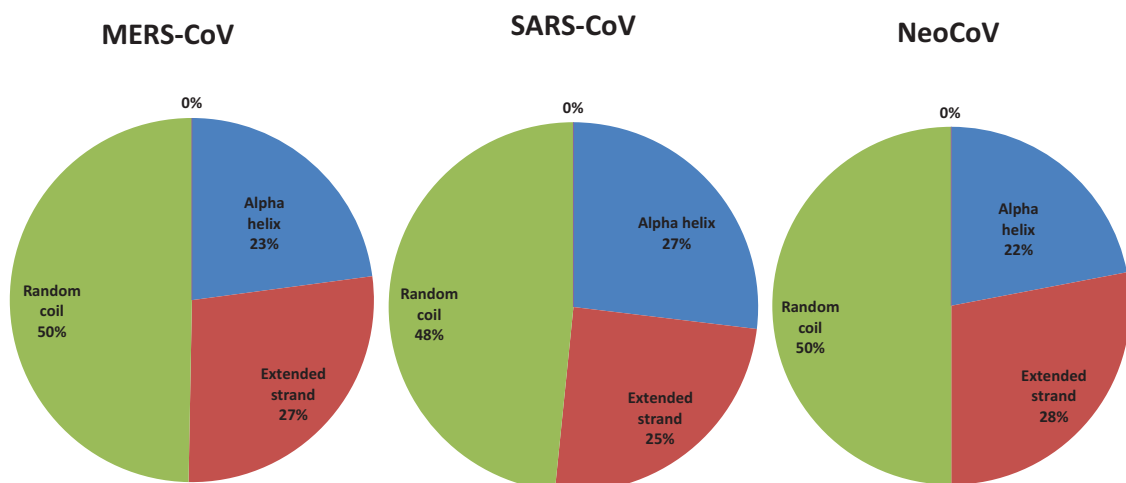


Figure 21. Percent of secondary structure component of ORF1ab proteins. Blue color for alpha helix, brown for extended strand, and green color for the random coil.



Figure 22. Three-dimensional (3D) structures of “E” proteins of 3 coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV).

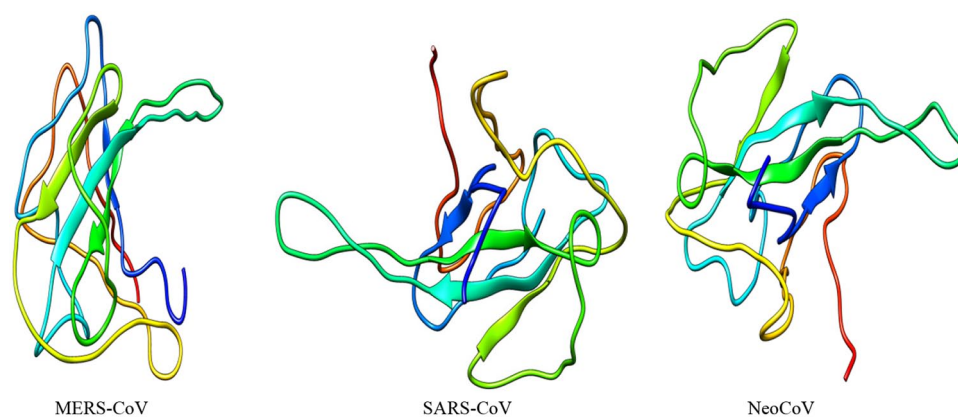


Figure 23. Three-dimensional (3D) structures of “M” proteins of 3 coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV).

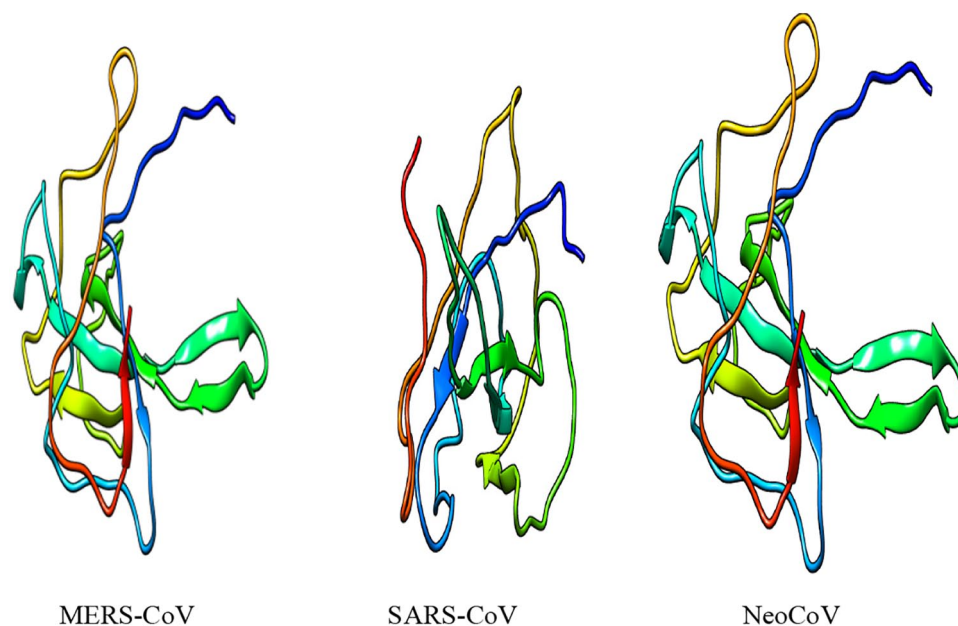


Figure 24. Three-dimensional (3D) structures of “N” proteins of 3 coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV).

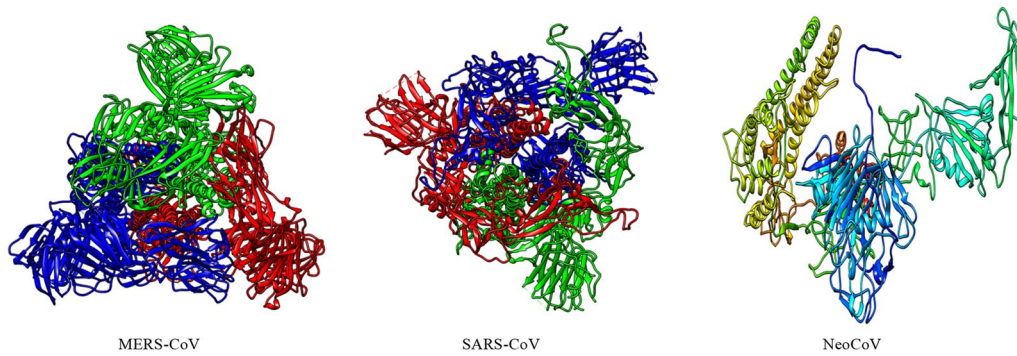


Figure 25. Three-dimensional (3D) structures of “S” proteins of 3 coronaviruses (MERS-CoV, SARS-CoV, and NeoCoV).

Table 16. TM-score to the measure of similarity between couple 3D protein structures.

PROTEIN TYPE	VIRUSES COMPARISONS		
	MERS VS SARS	NEO VS SARS	MERS VS NEO
E proteins	0.35661	0.37293	0.94734
M proteins	0.37210	0.35990	0.95354
N proteins	0.72849	0.72986	0.99963
S proteins	0.1369	0.1533	0.2245

Abbreviations: MERS, Middle East Respiratory Syndrome; SARS, Severe Acute Respiratory Syndrome; 3D, three-dimensional.

Discussion

In this study, we have endeavored to provide a deep understanding of the relationship between MERS-CoV, SARS-CoV, and NeoCoV at the amino acids level as the proteins are representing the functional unit of the genome and are directly involved in chemical processes essential for life. The proteins are species and organ-specific in which the proteins of one species or organs differ from those of another species or organs. However, proteins of similar function have similar amino acid composition and sequence. Despite the difficulties in explaining functions of protein from its amino acid sequence, understanding the correlations between structure and function is the key role of protein function.

With respect to the aim of determining the properties of amino acids that compose proteins of this study, Table 1 shows the physical and chemical properties of these proteins' amino acids that are present in all CoV species of interest.³⁵ We have found that the number of amino acids of E, M, N, and S proteins in addition to all other parameters including molecular weight, atomic composition, theoretical pI, and structural formula of MERS-CoV and NeoCoV were close to each other if not identical, and this has supported the previous finding that NeoCoV was closely related to MERS-CoV and suggested that MERS-CoV's ancestors may have evolved in bats.¹² This finding is in contrary to Victor Max Corman et al¹⁰ results which have reported that NeoCoV and MERS-CoV belonged to one viral species and that the presence of a genetically

divergent **S1** subunit within the NeoCoV spike gene indicated that intra-spike recombination events may have been involved in the emergence of MERS-CoV, because there were some differences regarding all 6 proteins and not **S** protein only.⁹ In accordance with our results, Agnihothram et al¹ have demonstrated that NeoCoV shared essential details of genome architecture with MERS-CoV. But, however, disagreement in that 85% of the NeoCoV genome is identical to MERS-CoV at the nucleotide level.

In this study, we used 5 different methods of phylogenetic tree construction including Maximum Parsimony (MP), Neighbor-Joining (NJ), Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Maximum Likelihood (ML), and RelTime (RT) to depict the relatedness, evolution change, and relative time between the viruses of interest (in the level of genome and protein). According to Phylogenetic results of the whole genomes which had relied on MUSCLE alignment, results have shown that joining of MERS-CoV and SARS-CoV with the nearest common ancestor and MERS-CoV has the lowest evolutionary change (Genetic distances). The RelTime method showed that NeoCoV was the oldest while MERS-CoV and SARS-CoV were belonged to the same time, based on the relative time. Furthermore, according to phylogenetic results of protein sequences which had relied on MUSCLE and CLUSTALW alignment methods, in general, trees have shown that joining NeoCoV and MERS-CoV proteins in same clades indicates that they are

closest on the basis of all used methods. Furthermore, according to the horizontal branch length through used methods, most NeoCoV proteins have the shortest branch length comparing to others.

Regarding protein's primary and secondary structures, most of the comparison results showed the most similarity between NeoCoV and MERS-CoV. Another comparison tool has template modeling score (TM-score), which is used to measure the topological similarity between the structure of proteins, and this method is insensitive to local structural variation. The TM results confirmed that NeoCoV was more close to MERS-CoV than SARS-CoV. Generally, phylogenetic analysis of the 6 proteins (E, S, M, N, ORF1a, and ORF1ab) revealed that there were high similarities between the 3 viruses although NeoCoV appeared close to MERS-CoV. This result indicated that they have the same common ancestor and NeoCoV may implicate in human-related infection sooner because of high similarity in portions involved in viral infectivity.

Author Contributions

MMH conceived the idea and designed the methodology. AAE, ORG, SAA, RAA, and AMAM performed the initial draft analyses; in addition, MMH carried the final analysis. MMH and SBM interpreted the results. MAH, KS, and MMH wrote the manuscript and developed the final draft. All authors read and approved the final manuscript.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID iDs

Mohamed M Hassan  <https://orcid.org/0000-0003-1544-7932>

Sofia B Mohamed  <https://orcid.org/0000-0001-6718-3540>

REFERENCES

1. Agnihothram S, Gopal R, Yount BL, et al. Evaluation of serologic and antigenic relationships between Middle Eastern respiratory syndrome coronavirus and other coronaviruses to develop vaccine platforms for the rapid response to emerging coronaviruses. *J Infect Dis.* 2014;209:995-1006.
2. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multi-functional protein. *Viruses.* 2014;6:2991-3018.
3. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. *Virol J.* 2015;22:221.
4. The International Committee for Taxonomy of Viruses (ICTV). http://talk.ictvonline.org/files/ictv_documents/m/msl/4090.aspx. Accessed June 27, 2014.
5. Jain A, Mittal N, Sharma PC. Genome wide survey of microsatellites in ssDNA viruses infecting vertebrates. *Gene.* 2014;552:209-218.
6. Li SW, Lin CW. Human coronaviruses: clinical features and phylogenetic analysis. *Biomedicine (Taipei).* 2013;3:43-50.
7. Neuman BW, Kiss G, Kunding AH, et al. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol.* 2011;174:11-22.
8. Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A.* 2014;111:15214-15219.
9. Maltezou HC, Tsiodras S. Middle East respiratory syndrome coronavirus: implications for health care facilities. *Am J Infect Control.* 2014;42:1261-1265.
10. Corman VM, Ithete NL, Richards LR, et al. Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J Virol.* 2014;88:11297-11303.
11. Institute of Medicine. *Emerging Viral Diseases: The One Health Connection: Workshop Summary.* Washington, DC: The National Academies Press; 2015.
12. De Benedictis P, Marciano S, Scaravelli D, et al. Alpha and lineage C betaCoV infections in Italian bats. *Virus Genes.* 2014;48:366-371.
13. Zhong NS, Zheng BJ, Li YM, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet.* 2003;362:1353-1358.
14. Wu Z, Yang L, Ren X, et al. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISMEJ.* 2016;10:609-620.
15. Suwantarant N, Apisarnthanarak A. Risks to healthcare workers with emerging diseases: lessons from MERS-CoV, Ebola, SARS, and avian flu. *Curr Opin Infect Dis.* 2015;28:349-361.
16. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev.* 2006;19:531-545.
17. Ithete NL, Stoffberg S, Corman VM, et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg Infect Dis.* 2013;19:1697-1699.
18. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539.
19. McWilliam H, Li W, Uludag M, et al. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* 2013;41:W597-W600.
20. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870-1874.
21. Nei M, Kumar S. *Molecular Evolution and Phylogenetics.* Oxford, UK: Oxford University Press; 2000.
22. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4:406-425.
23. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A.* 2004;101:11030-11035.
24. Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A.* 2012;109:19333-19338.
25. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993;10:512-526.
26. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics.* 2010;26:889-895.
27. Sneath PHA, Sokal RR. *Numerical Taxonomy.* San Francisco, CA: Freeman; 1973.
28. Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008;36:W465-W469.
29. Hassan MM, Mohamed SB, Hussain MA, Dowd AA. Deleterious nonsynonymous SNP found within HLA-DRB1 gene involved in allograft rejection in Sudanese family: using DNA sequencing and bioinformatics methods. *Open J Immunol.* 2015;5:222-232.
30. Garnier J, Gibrat JF, Robson B, Doolittle R. GOR secondary structure prediction method version IV. *Methods Enzymol.* 1996;266:540-553.
31. Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels 3.2—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* 2010;38:576-581.
32. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.* 2012;7:1511-1522.
33. Couch GS, Donna K, Hendrix DK, Ferrin TE. Nucleic acid visualization with UCSF Chimera. *Nucleic Acids Res.* 2006;34:e29.
34. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins.* 2004;57:702-710.
35. Tajima F, Nei M. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol.* 1984;1:269-285.