Commagene Journal of Biology

Kurt, (2022) *Comm. J. Biol.* 6(1), 68-74.
DOI: 10.31594/commagene.1079045

e-ISSN 2602-456X

Research Article/Araştırma Makalesi

Comparative Bioinformatics Analyses of SARS-CoV-2 Spike Glycoproteins in Different Countries

Barış KURT

Muş Alparslan University, Faculty of Education, Department of Math, Muş, TÜRKİYE
ORCID ID: Barış KURT: https://orcid.org/0000-0002-1406-0915

Received: 25.02.2022 **Accepted:** 01.05.2022 **Published online:** 20.06.2022 **Issue published:** 30.06.2022

Abstract: In this study, 27 protein sequences of SARS-CoV-2 spike (S) glycoprotein from 23 different countries were analyzed using bioinformatics approaches. In this context, post-translational modifications, sequence and domain analyses, phylogenetic analysis, and 3D structure analysis of the spike glycoprotein proteins were performed. Also, molecular docking analysis of the SARS-CoV-2 spike protein S1 receptor-binding domain (SS1) with human ACE2 protein was conducted. It was found that although all SARS-CoV-2s include Spike_rec_bind (PF09408) and Corona_S2 (PF01601) domain structures, the C-terminal S2 region was more diverse than the S1 region. The predicted N-glycosylation and phosphorylation sites were determined to be between 17 and 19 and 136 and 168, respectively. In phylogenetic analysis, SARS-CoV-2s were found to have more similarity with bat RaTG13 and pangolin CoV-2 than MERS CoV and bat SARS CoV. The predicted 3D protein structures of human SARS-CoV-2 and bat RaTG13 showed high similarity, ranging from 0.76 to 0.78. The docking analyses revealed that Asp30, Lys31, His34, Glu35, Glu37, Asp38, Asn330, and Gln325 residues were binding residues in the ACE2 protein for the N-terminal S1 subunit of SARS-CoV-2. The findings are particularly important for the studies of drug development and drug design.

Keywords: Covid-19, coronavirus, spike protein, molecular docking, in silico analyses.

SARS-CoV-2 Spike Glikoproteinlerinin Farklı Ülkelerde Karşılaştırmalı Biyoinformatik Analizleri

Öz: Bu çalışmada, biyoinformatik yaklaşımlar kullanılarak 23 farklı ülkeden SARS-CoV-2 spike (S) glikoproteininin 27 protein dizisi analiz edildi. Bu kapsamda saçak glikoproteinlerinin post-translasyonel modifikasyonları, sekans ve domain analizleri, filogenetik analizleri ve 3 boyutlu yapı analizleri gerçekleştirilmiştir. Ayrıca, insan ACE2 proteini ile SARS-CoV-2 saçak proteini S1 reseptör bağlama alanının (SS1) moleküler yerleştirme analizi yapıldı. Tüm SARS-CoV-2'lerin Spike_rec_bind (PF09408) ve Corona_S2 (PF01601) alan yapılarını içermesine rağmen, C-terminal S2 bölgesinin S1 bölgesinden daha çeşitli olduğu bulundu. Öngörülen N-glikosilasyon ve fosforilasyon bölgelerinin sırasıyla 17 ve 19, 136 ve 168 arasında olduğu belirlendi. Filogenetik analizde, SARS-CoV-2'lerin yarasa RaTG13 ve pangolin CoV-2 ile MERS CoV ve yarasa SARS CoV'den daha fazla benzerliğe sahip olduğu bulundu. İnsan SARS-CoV-2 ve yarasa RaTG13'ün tahmin edilen 3D protein yapıları, 0.76 ile 0.78 arasında değişen yüksek benzerlik gösterdi. Yerleştirme analizleri, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Asn330 ve Gln325 rezidülerinin, SARS-CoV-2'nin N-terminal S1 alt birimi için ACE2 proteininde bağlayıcı kalıntılar olduğunu ortaya çıkardı. Bulgular özellikle ilaç geliştirme ve ilaç tasarımı çalışmaları için önemlidir.

Anahtar kelimeler: Kovid-19, koronavirus, saçak proteini, moleküler kenetleme, in siliko analizler.

1. Introduction

Coronaviruses (CoVs) are reported to cause a wide variety of diseases in humans and animals (Masters & Perlman, 2013). Particularly, three types of CoV viruses have caused deadly pneumonia in humans in the past 20 years: severe acute respiratory syndrome coronavirus (SARS-CoV) (Drosten et al., 2003), Middle-East respiratory syndrome coronavirus (MERS-CoV) (Zaki et al., 2012), and SARS-CoV-2 (Zhu et al., 2020). CoVs are categorized into four main taxa that are alpha-CoV, beta-CoV, gamma-CoV, and delta-CoV (Lefkowitz et al., 2018). Two beta-CoVs, SARS-CoV and SARS-CoV-2, are closely related viruses. The origins of these viruses are suggested to be bats that seemed to be reservoir hosts for these two viruses (Zhou et al., 2020). Phylogenetical analyses showed that B beta-CoV and Bat SARS-like (SL) CoV, and RaTG13 have high nucleotide sequence identities with 96%, suggesting that SARS CoV-2 probably evolved from a Bat SL-CoV (Zhou et al., 2020; Ou et al., 2020). On the other hand, there are varieties of beta-CoV genus that have low-pathogenicity ability in humans such as HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoVHKU1 (Walls et al., 2020). SARS-CoV contains large positive-sense RNA genomes with about 30 kb in length and includes four structural proteins such as spike (S), nucleocapsid (N), envelope (E), and membrane (M) (Siddell & Ziebuhr, 2005). The spike glycoproteins (S) of CoVs bind to its receptor, angiotensin converting enzyme 2 (ACE2) and mediate membrane fusion and virus entry. S protein is trimeric and the weights of each monomer is about 180 kDa. The attachment and membrane fusion of the virus is controlled with two subunits: N-terminal S1 subunit and membrane-embedded C-terminal S2 region (Ou et al., 2020). The S1 region contains amino terminal domain (S1-NTD) and the carboxy-terminal domain (S1-CTD). The S1-CTD subunit - receptor interaction is a vital step in determining host range of CoVs and tissue tropism (Lu et al., 2015; Cui et al., 2019). After membrane fusion, the CoV's genome is

*Corresponding author: b.kurt@alparslan.edu.tr

delivered into the cytoplasm using host cell membrane or endosome membrane. Then, viral RNA starts translation of the two polyproteins (pp1a and pp1ab) encoding 16 non-structural proteins (nsp1–nsp16) and replication of the viral genome (Fehr & Perlman, 2015; Ou et al., 2020). The CoV polyproteins are cleaved by the main protease and papain-like protease to produce the nonstructural proteins (Kandeel et al., 2020). In this study, comparative bioinformatics analyses of 27 SARS-CoV-2 spike (S) protein sequences were performed. Later, the predicted 3D structures and docking of SARS-CoV-2 S proteins were generated and analyzed by *in silico* approaches.

2. Material and Methods

2.1. Sequence collections

Firstly, a total of 27 protein sequences of SARS-CoV-2 spike glycoprotein from 23 different countries were retrieved from the NCBI database (NCBI National Library of Medicine, 1988), including Australia: Victoria (QHR84449.1), Brazil (QJA41641.1), Türkiye (QIZ16509.1), Japan (BCB97901.1), Vietnam (QIK50448.1), Pakistan: Gilgit (QIQ22760.1), Colombia: Antioquia (QIS30054.2), Pakistan: KPK (QIS60276.1), Iran (QIU80900.1), South Korea (QIV15008.1), Iran (QIX12195.1), USA:MN (QJA17740.1), Taiwan (QJD20656.1), India: Kerala State (QIA98583.1), Hong Kong (QJC20993.1), Finland (QHU79173.2), India: Rajkot (QJC19491.1), Peru (QIS60288.1), Israel (QIT06999.1), France (QIX12148.2), South Africa (QIZ15537.1), Greece (QIZ16571.1), USA:OR (QJA17360.1), USA: New Orleans (QJC21041.1), Spain (QJC21017.1), Serbia (QJC21051.1), and Sweden (QIC53204.1). Later, collected sequences of SARS-CoV-2 were used for further bioinformatics analyses.

2.2. Sequence alignment and analyses

The retrieved 27 protein sequences were aligned using BioEdit software v7.2.5 (Hall, 1999). Domain analyses of SARS-CoV-2 spike glycoproteins were performed using Pfam 32.0 database (El-Gebali et al., 2019). Percent identities of SARS-CoV-2 spike glycoproteins were analyzed using blastp (protein-protein BLAST) suite (Johnson et al., 2008) in the NCBI database. For this analysis, SARS CoV BJ01 (AAP30030.1), bat SARS CoV Rs_672/2006 (ACU31032.1), bat coronavirus RaTG13 (QHR63300.2), pangolin CoV (QIQ54048.1), and MERS CoV (YP_009047204.1) were used for comparisons to SARS-CoV-2 spike glycoproteins. In addition, Tajima's test statistic (D) (Tajima, 1989), number of segregating sites (S), and nucleotide diversity (π) were calculated using MEGA-X software. The putative N-Glycosylation and serine-threonine or tyrosine phosphorylation sites of SARS-CoV S proteins were identified using NetNGlyc 1.0 (Blom et al., 2004) and NetPhos 3.1 (Blom et al., 2004) servers, respectively.

2.3. Phylogenetic analysis

Phylogenetical analysis was conducted by including 42 amino acid sequences. For this purpose, bat SARS CoVs (ACU31032.1, ACU31051.1, ABD75332.1, ABG47069.1, and QHR63300.2), MERS CoVs (YP_009047204.1, AID50418.1, AVN89453.1, AVN89344.1, and AVN89291.1), and SARS CoVs (AAP30030.1, ABA02260.1, ACZ72065.1, AAV91631.1, and AAX16192.1) sequences were retrieved to be added into the data used in sequential analysis. The relationship among lineages were constructed based on the Neighbor-Joining (NJ) method. The bootstrap confidence values were calculated by 1000 replication for each clade (Felsenstein, 1985). Evolutionary analyses were conducted in MEGA-X (Kumar et al., 2018). A total of 1360 positions were found in the dataset as a result of computation of evolutionary distances according to Poisson correction method.

2.4. Predicted 3D structure and docking analyses

First, predicted 3D structures of selected SARS-CoV-2 S proteins such as Colombia, Pakistan-Gilgit, Sweden, and Vietnam were generated using Phyre (Protein Homology/analogY Recognition Engine V 2.0) online server at intensive mode (Kelley et al., 2015). Later, the similarity of protein structures was evaluated using template modeling score (TM-score) (Zhang & Skolnick, 2004). The structural evaluation and stereo-chemical analyses of the modeled proteins was checked using Ramachandran plot analysis on the Rampage server (Lovell et al., 2003).

The ACE2 protein was extracted from the file (6M0J) in PDB Bank (Berman, 2000) and Sweden SARS-CoV-2 spike protein S1 receptor-binding domain (SS1) was used for protein- protein docking analysis.ss HADDOCK 2.4 server was employed for the docking procedure (van Zundert et al., 2016). Since His228 in the file showed two distinct confirmations and these confirmations caused a problem for docking analysis to proceed, the file was corrected manually. The interacting residues for HADDOCK parameters were obtained from (Lan et al., 2020) and docking analysis was completed by default settings. The resulting docking file was opened in PyMOL and the best confirmation was used for alignment with the input proteins and RMSD value was calculated based on this alignment. The binding affinity was found using Prodigy server at 25°C (Xue et al., 2016). The contacting residues between proteins were visualized in Chimera 1.14rc (Pettersen et al., 2004). LigPlot+ was used for the identification of binding site interactions with their residue distances (Laskowski & Swindells, 2011). The detailed analysis of the docking such as number of contacting residues, salt bridges, and non-bonded interactions between the molecules were identified by PDBsum server (Laskowski et al., 2018).

3. Results and Discussion

3.1. Genetic variations in SARS-CoV-2 spike glycoproteins

All SARS-CoV-2 S proteins contained Corona_S2 (PF01601) domain between 662 and 1266 amino acid residues and Spike_rec_bind (PF09408) domain structure was between 330 and 583 amino acid residues by Pfam database (Fig. 1). According to sequence analyses of spike receptor binding domain (PF09408) structure (S1 subunit), there was just one amino acid substitution in the amino acid chain of 367 residues and only valine (V) in Hong Kong SARS-CoV2 (QJC20993.1) replaced with phenylalanine (F) (V367F). In C-terminal S2 region, A930V in India: Kerala State, V772I in Türkiye, T791I in Taiwan, and F797C in Sweden amino acid substitutions were identified. To understand the genetic diversity level of SARS-CoV-2 S proteins, Tajima's D was calculated and found as -1.805. In addition, the number of segregating

sites and nucleotide diversity (π) were identified as 10 and 0.0009, respectively. For a better understanding of human SARS-CoV 2 S protein sequences, four SARS-CoV-2 S protein sequences were selected from our study along with Bat coronavirus RaTG13 and Pangolin Coronavirus S protein sequences that are most similar to human S proteins. Then, they were examined by aligning (Fig. 2). In particular, when the amino acid changes in the spike receptor-binding domain were examined, it was found that there were changes in 33 and 21 regions between human and pangolin and human and bat SARS-CoV2 S proteins, respectively. Thus, pangolin SARS-CoV2 S protein sequences were found to show more variations in the region of receptor binding domain compared to human SARS-CoV2.

Family	Description	Entry type	Clan	Envelope		Alignment	
				Start	End	Start	End
Corona_S2	Coronavirus S2 glycoprotein	Family	CL0595	662	1266	672	1266
Spike_rec_bind	Spike receptor binding domain	Domain	n/a	330	583	330	583

Figure 1. Domain analyses of SARS-CoV2 S protein sequences using Pfam database. These two domain structures identified all SARS-CoV2 S proteins.

Figure 2. Multiple sequence alignment of S protein sequences of selected four human SARS-CoV2 (Serbia, Spain, Taiwan, and India: Rajkot), bat coronavirus RaTG13, and pangolin coronavirus. Conserved regions, showing at least 80% similarity, are shown in black and red line represent the spike receptor binding domain (PF09408) structure (between 330 and 583 amino acid residues) based on Pfam database.

The percent identities of SARS-CoV-2 S proteins were identified by blastp tool in NCBI database (Table 1). Human SARS-CoV-2 S protein sequences were compared with SARS CoV BJ01, MERS CoV, Bat CoV, Bat COV RaTG13, and Pangolin CoV sequences. When the results are examined, SARS-CoV-2 & SARS CoV BJ01, SARS-CoV-2 & MERS CoV, SARS-CoV-2 & Bat CoV, SARS-CoV-2 & Bat COV RaTG13, and SARS-CoV-2 and Pangolin CoV percent identity (%) values were between 76.04 and 76.19, 34.82 and 34.91, 74.51 and 74.76, 97.33 and 97.41, and 91.99 and 92.14, respectively. Considering the general identity ratios, it was seen that SARS-CoV-2 S proteins gave the highest similarity values with bat CoV RaTG13 S protein (about 97%), followed by Pangolin CoV (about 92%) and SARS CoV BJ01 (about 76%). In addition, the lowest similarity values with SARS-CoV-2 S proteins were about 35% with MERS CoV.

Table 1. Percent identity (%) of SARS-CoV-2 S proteins using blastp tool in NCBI database

SARS-CoV2	SARS CoV BJ01	MERS CoV	Bat CoV	Bat COV RaTG13	Pangolin CoV
Australia: Victoria	76.12	34.91	74.67	97.33	92.07
Brazil	76.12	34.91	74.76	97.41	92.14
Türkiye	76.12	34.91	74.59	97.33	92.07
Japan	76.12	34.91	74.59	97.41	92.14
Vietnam	76.12	34.91	74.59	97.41	92.14
Pakistan: Gilgit	76.12	34.91	74.59	97.41	92.14
Colombia: Antioquia	76.12	34.91	74.59	97.41	92.07
Pakistan: KPK	76.12	34.91	74.59	97.41	92.14
Iran (QIU80900.1)	76.12	34.91	74.59	97.41	92.14
South Korea	76.12	34.91	74.59	97.41	92.14
Iran (QIX12195.1)	76.12	34.91	74.59	97.41	92.14
USA:MN	76.12	34.91	74.59	97.41	92.14
Taiwan	76.04	34.91	74.51	97.33	92.07
India: Kerala State	76.04	34.82	74.51	97.33	92.07
Hong Kong	76.04	34.91	74.51	97.33	92.07
Finland	76.19	34.91	74.51	97.33	92.07
India: Rajkot	76.04	34.82	74.51	97.25	91.99
Peru	76.04	34.91	74.51	97.33	92.07
Israel	76.04	34.91	74.51	97.33	92.07
France	76.04	34.91	74.51	97.33	92.07
South Africa	76.04	34.91	74.51	97.33	92.07
Greece	76.04	34.91	74.51	97.33	92.07
USA: OR	76.04	34.91	74.51	97.33	92.07
USA: New Orleans	76.04	34.91	74.51	97.33	92.07
Spain	76.04	34.91	74.51	97.33	92.07
Serbia	76.04	34.91	74.51	97.33	92.07
Sweden	76.04	34.82	74.51	97.33	92.07

3.2. Post-translational modifications of SARS-CoV S proteins

The putative N-glycosylation sites (Table 2) in CoV S protein sequences were ranged from 17 to 19 by NetNGlyc 1.0 server. In terms of the number of serine-threonine or tyrosine phosphorylation sites, they changed from 136 to 140 and were higher than the number of the N-glycosylation sites.

Table 2. The predicted N-Glycosylation and serine-threonine or tyrosine phosphorylation sites of SARS-CoV S proteins, respectively.

Protein name	NetNGlyc 1.0 Server	NetPhos 3.1 Server
Colombia SARS-CoV2	17	142
Pakistan-Gilgit SARS-CoV2	17	142
Sweden SARS-CoV2	17	140
Vietnam SARS-CoV2	17	140
SARS CoV BJ01	18	137
MERS CoV	19	168
Bat COV RaTG13	17	137
Pangolin CoV	18	136

3.3. Phylogenetic analysis of SARS-CoV-2 S proteins

For phylogenetic analysis, a total of 43 S protein sequences were used, 27 of which are SARS-CoV-2, five of which are MERS CoV, five of which are Bat SARS CoV, one of which is pangolin CoV, and five of them are SARS CoV (Fig. 3). The phylogeny shows that MERS CoVs are outgroup to SARS CoVs and Bat SARS CoVs. Although MERS CoV, Bat SARS CoV, and SARS CoV S proteins split into different clades with 100% bootstrap value, SARS-CoV-2 S proteins did not split from other clusters. The pangolin CoV-2 and RaTG13 split from the MERS CoV, Bat SARS CoV, and SARS CoV S corona viruses as a result of mutations and they constituted paraphyletic group containing some SARS-CoV 2 lineages. Moreover, bat coronavirus RaTG13 separated from other bat sequences and clustered with the SARS-CoV-2 S sequences. When the distribution of the SARS-CoV-2 sequences were examined, although most of them were clustered together, Iran, Pakistan: KPK, Pakistan: Gilgit, Vietnam, and Sweden SARS-CoV-2 sequences were sister groups under a different sub-cluster.

3.4. Comparative 3D structure and docking analyses

The predicted tertiary (3D) structures of Bat CoV, MERS CoV, SARS CoV BJ01, and RaTG13 proteins were generated using Phyre server. The number of the residues in favored and allowed regions were identified as 98.9% for SARS CoV BJ01, 96.3% for RaTG13, 96.1% for MERS CoV, and 98.6% for Bat CoV according to Ramachandran plot analysis using Rampage online server, suggesting reliable 3D predictions. To offer more insights about the structural similarities of corona virus proteins in the study, the SARS-CoV-2 S protein structure and different CoV S protein were compared using the TM-score server. For this analysis, four human SARS-CoV-2 were selected such as Colombia, Pakistan-Gilgit, Sweden, and Vietnam (Table 3). The SARS-CoV-2 S proteins were compared with the four different CoV S proteins modeled in this study, including Bat CoV (ACU31032.1), MERS CoV (YP_009047204.1), SARS CoV BJ01 (AAP30030.1), and RaTG13 (QHR63300.2). As a result of the analysis, SARS-CoV-2 showed a high level of structural similarity with RaTG13 S protein, followed by SARS CoV BJ01 and Bat CoV.

The docking results were shown in Fig 4. The RMSD values for the docking were found to be in turn 0.169 and 0.160 for SS1 and ACE2 proteins, suggesting that the model qualities of docking were high since RMSDs were lower than one (Karaca et al., 2010). The binding affinity between SS1 and ACE2 was -10.4 kcal/mol, calculated on Prodigy server, indicating high attraction force between two proteins. A total of 43 interface residues were found between ACE2 and SS1 (23 for ACE2 and 20 for SS1). The total numbers of hydrogen bonds were nine whereas only one salt bridge was observed across interface. Of these hydrogen bonds, only one had double bonds. As a result of the docking analyses, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Asn330, and Gln325 were found as binding

residues of ACE2 to Arg403, Tyr473, Gln474, Tyr489, Gly496, Gln498, Thr500, and Gly502 residues of SS1. The stabilizing effect of salt bridge between the proteins was observed between Asp38 of ACE2 and Arg403 of SS1 along with double hydrogen bonds. The salt bridge distance between ACE2 and SS1 was 3.27 Å. Furthermore, the docking analysis showed that except Lys31 and His34 residues in ACE2 and Arg403, Tyr473, and Gln474 residues in SS1, all other residues in ACE2 and SS1 were found as contacting residues reported by Lan et al. (2020).

Table 3. Predicted 3D structure overlap of some selected spike proteins between SARS-CoV-2 and bat, MERS, and SARS CoVs using TM-score online server for measuring the similarity.

SARS-CoV-2	Bat CoV (ACU31032.1)	MERS CoV (YP_009047204.1)	SARS CoV BJ01 (AAP30030.1)	RaTG13 (QHR63300.2)
Colombia	0.2470	0.1326	0.2665	0.7783
Pakistan-Gilgit	0.2484	0.1330	0.2642	0.7740
Sweden	0.2492	0.1314	0.2710	0.7659
Vietnam	0.2484	0.1306	0.2676	0.7563

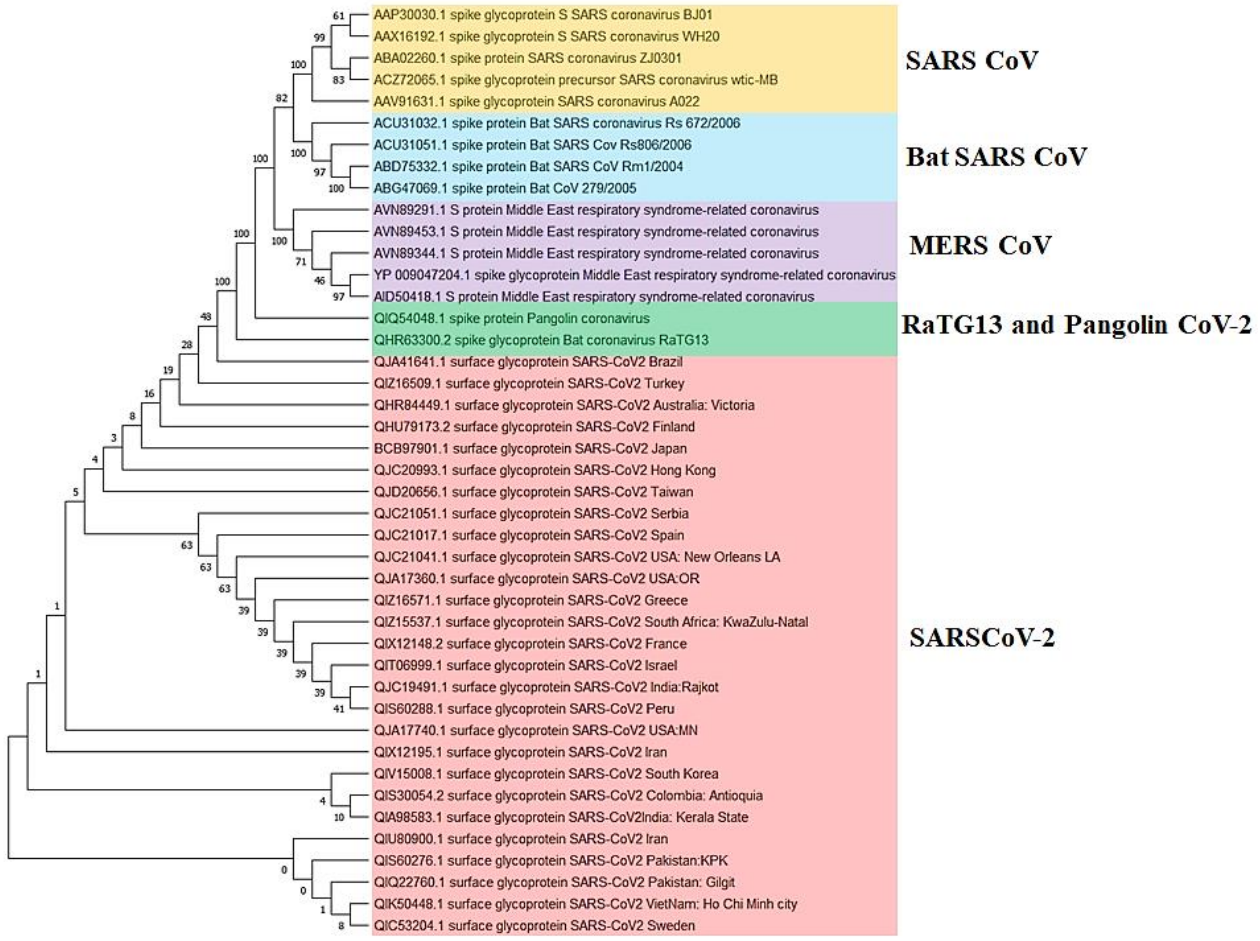

Figure 3. Phylogenetic distributions of SARS-CoV-2, MERS CoV, bat SARS CoV, Pangolin CoV, and SARS CoV S protein sequences by NJ tree method using MEGA X with 1000 bootstrap replicates.

4. Discussion

The goal of this research was to evaluate 27 SARS-CoV-2 S protein sequences obtained from 23 countries throughout the world using various bioinformatics methods in order to contribute to the knowledge of SARS-CoV-2 evolution. Positive selection or population expansions lead low frequency alleles to increase, resulting in a negative Tajima's D. Under a standard neutral model, positive selection may cause genetic variance to fluctuate (Biswas & Akey, 2006). Tajima's D was determined to be -1.805 for this present review indicating positive selection of S genes in SARS-CoV-2. The receptor binding domain (RBD) of 2019-nCoV contains 348A, 354N, 417K, 430N, 438S, and 519H residues according to Wu et al. (2020). In this study, 348A, 354N, 417K except from pangolin, and 519H except from bat and pangolin were present in alignment of human, bat RaTG13, and pangolin SARS-CoV-2 (Fig. 2). These variations in RBD region may support the mutation potential of SARS-CoV-2. In RBD region, it was revealed that SARS-CoV-2 and Pangolin-CoV were highly conserved and only one amino acid change (500H/500Q) was identified. Besides, this change does not belong to the interaction with human ACE2 (Zhang & Skolnick, 2004). In this investigation, just one amino acid substitution (367V/367F) was discovered and was not detected in the catalytic residues of ACE2 binding. Zhu et al. (2020) stated that Pangolin-CoV, RaTG13, and SARS-CoV-2 were clustered in "SARS-CoV-2 group" in the phylogenetic tree suggesting that these data proved our results (Table 1 and Fig. 3). Post-translational modifications (PTMs) aim to contribute to the temporal and spatial regulation of protein functions by making some covalent modifications after the polypeptide is synthesized. Coronavirus (CoV) is RNA

virus with an envelope causing diseases in both human and animals and many CoV proteins undergo PTMs such as palmitoylation, N- or O-linked glycosylation, phosphorylation, and other PTMs. The N-linked glycosylation support the conformation changes of S protein affecting the receptor binding of S protein (Ritchie et al., 2010; Fung & Liu, 2018). Consequently, the existence of a large number of PTMs may be related to the functional regulation of proteins in SARS-CoV metabolism.

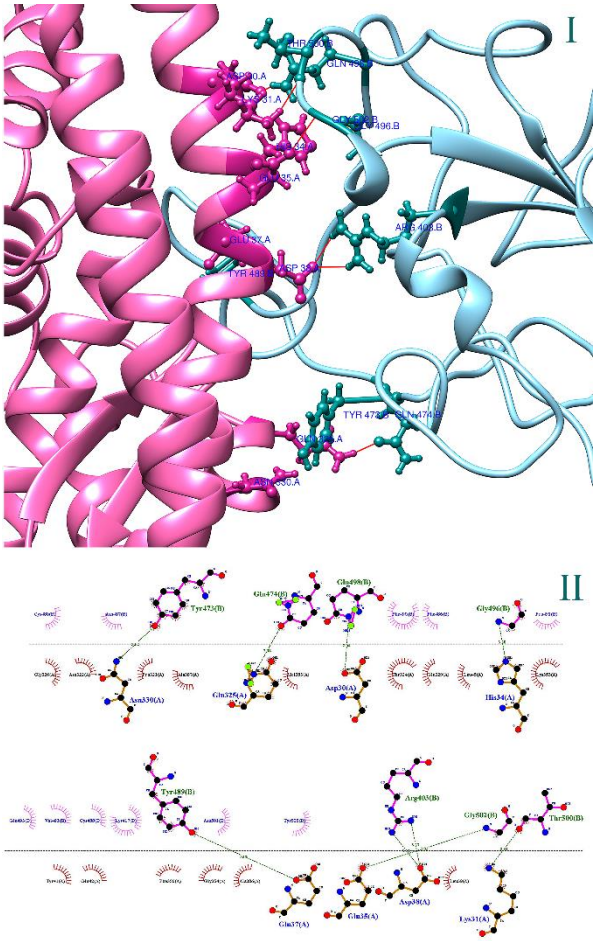

Figure 4. The interacting residues of ACE2 (A) and Sweden SARS-CoV-2 S1 protein (SS1) (B). (I) ACE2 was shown in hot pink with binding residues in violet red whereas SARS-CoV-2 was displayed in sky blue with interacting residues in dark cyan. The hydrogen bonds among residues were shown in red with threshold of three Å. (II) The schematic representation of interacting residues. Hydrogen bonds were shown in green with interacting residues and their distances. Residue numbers shown with A and B represent ACE2 and SS, respectively.

The template modeling score (TM-score) can be used to assess the similarity between two protein structures. The TM score has values ranging from 0 to 1 in which one indicates the perfect match whereas 0 implies that no similarity exists between binary structural comparisons of the proteins. If the TM-score falls below 0.17, two proteins are decided to be unrelated. By contrast, if the TM-score exceeds 0.50, it is assumed that the proteins are in the same fold in SCOP/CATH (Zhang & Skolnick, 2004; Xu & Zhang, 2010). In this study, MERS CoV & SARS-CoV-2 S protein structures had about 0.13 TM-score; therefore, the protein structures appear to be not very close. On the other hand, the SARS-CoV-2 & RaTG13 S protein sequences were found to yield TM-score over 0.75 indicating high similarity of two protein structures. These findings were in parallel with phylogenetic and sequence similarity results.

Docking results were in agreement with the experimental results of (Lan et al., 2020). Also, it was found that molecular interacting forces between SS1 and ACE2 show high binding affinity. Particularly, the salt bridge and double hydrogen bonded form between Asp38 of ACE2 and Arg403 of SS1 stabilizes the protein complex; therefore, the common interacting residues involved in binding can be used as potential targets for the development of new drugs against Covid19 outbreak.

Ethics committee approval: Ethics committee approval is not required for this study.

Conflict of interest: The author declares that there is no conflict of interest.

References

Biswas, S., & Akey, J.M. (2006). Genomic insights into positive selection. *Trends in Genetics*, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6), 1633–1649. https://doi.org/10.1002/pmic.200300771

Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. https://doi.org/10.1038/s41579-018-0118-9

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., … & Doerr, H. W. (2003). Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England Journal of Medicine*, 348(20), 1967–1976. https://doi.org/10.1056/NEJMoa030747

Fehr, A.R., & Perlman, S. (2015). Coronaviruses: An Overview of Their Replication and Pathogenesis. *Nature Public Health Emergency Collection*, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, 39(4), 783. https://doi.org/10.2307/2408678

Fung, T.S., & Liu, D.X. (2018). Post-translational modifications of coronavirus proteins: roles and function. *Future Virology*, 13(6), 405–430. https://doi.org/10.2217/fvl-2018-0008

Hall, T.A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., ... & Finn, R.D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, 427-432. https://doi.org/10.1093/nar/gky995

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242. https://doi.org/10.1093/nar/28.1.235 and Retrieved From: https://www.rcsb.org

Kandeel, M., Ibrahim, A., Fayez, M., & Al-Nazawi, M. (2020). From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. *Journal of Medical Virology*, 92(6), 660–666. https://doi.org/10.1002/jmv.25754

Karaca, E., Melquiond, A.S.J., de Vries, S.J., Kastritis, P.L., & Bonvin, A.M.J.J. (2010). Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Molecular & Cellular Proteomics : MCP*, 9(8), 1784–1794. https://doi.org/10.1074/mcp.M000051-MCP201

Kelley, L. A., Mezulis, S., Yates, C.M., Wass, M. N., & Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., … & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5

Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S., & Thornton, J.M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science: A Publication of the Protein Society*, 27(1), 129–134. https://doi.org/10.1002/pro.3289

Laskowski, R.A., & Swindells, M.B. (2011). LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778–2786. https://doi.org/10.1021/ci200227u

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., & Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, 46, 708-717. https://doi.org/10.1093/nar/gkx932

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., … & Richardson, D.C. (2003). Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3), 437–450.

Lu, G., Wang, Q., & Gao, G.F. (2015). Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in Microbiology*, 23(8), 468–478. https://doi.org/10.1016/j.tim.2015.06.003

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36, 5–9. https://doi.org/10.1093/nar/gkn201

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., … & Qian, Z. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, 11(1), 1620. https://doi.org/10.1038/s41467-020-15562-9

Masters, P.S., & Perlman, S. (2013). in Fields Virology Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 825–858.

NCBI National Library of Medicine (US). (1988). National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., & Ferrin, T.E. (2004). UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Ritchie, G., Harvey, D.J., Feldmann, F., Stroeher, U., Feldmann, H., Royle, L., … & Rudd, P.M. (2010). Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology*, 399(2), 257–269. https://doi.org/10.1016/j.virol.2009.12.020

Siddell, S.G., & Ziebuhr, J,S.E. (2005). Coronaviruses, Toroviruses, and Arteriviruses. In V. T. M. B.W.J. Mahy (Ed.), *Topley and Wilson's microbiology and microbial infections* (pp. 823–856). Hodder Arnold, London.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–95.

van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., … & Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, 428(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 183(6), 1735. https://doi.org/10.1016/j.cell.2020.11.032

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., … & Jiang, T. (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe*, 27(3), 325–328. https://doi.org/10.1016/j.chom.2020.02.001

Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, 26(7), 889–895. https://doi.org/10.1093/bioinformatics/btq066

Xue, L.C., Rodrigues, J.P., Kastritis, P.L., Bonvin, A.M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics (Oxford, England)*, 32(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., & Fouchier, R.A.M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine*, 367(19), 1814–1820. https://doi.org/10.1056/NEJMoa1211721

Zhang, Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins*, 57(4), 702–710. https://doi.org/10.1002/prot.20264

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., … & Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., … & Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727–733. https://doi.org/10.1056/NEJMoa2001017