LUPINE PUBLISHERS

Lupine Online Journal of Pharmacology & Clinical Research

DOI: 10.32474/LOJPCR.2021.02.000140

ISSN: 2644-1373

Research Article

Alignment of SARS-Cov-2 Spike Protein Compared with Ebola, and Herpes Simplex Viruses in Human Murine and Bats

Hend M Tag[1,2]*, Salwa S Bawazir[1] and Ibtihal H Badawi[1]

[1]*Department of Biology, College of Science and Arts at Khulais, University of Jeddah, Saudi Arabia*

[2]*Department of Zoology, Faculty of Sciences, Suez Canal University, Egypt*

***Corresponding author:** Hend MT, Department of Biology and zoology, College of Science and Arts at Khulais, University of Jeddah, Jeddah, Saudi Arabia

Received: February 08, 2021 **Published:** February 18, 2021

Abstract

Our analytical approach consisted in sequence alignment analysis of spike protein in different viruses, followed by construction of phylogenetic tree. Additionally, we investigated some commotional parameters on the protein sequence determining chemical composition as well as estimated PI .Our observation revealed significant difference in S protein between the 3 tested viruses in different species These differences may have significant implications on pathogenesis and entry to host cell.

Introduction

Viruses are inert outside the host cell which are unable to generate energy. As obligate intracellular parasites, during replication, they are fully reliant on the complex biochemical machinery of eukaryotic or prokaryotic cells. The central purpose of a virus is to transport its genome into the host cell to allow its expression by the host cell [1]. Binding of a single surface glycoprotein of virus to its host receptor promotes pH-dependent conformational variations once within the endosome acidic environment, thereby transporting the viral bilayer in closeness with the host cell membrane to promote fusion [2]. The structural physiognomies of virus coats (capsids) are extremely appropriate to virus propagation. The coat must enclose and protect the nucleic acid, flexible against interference be talented of broaching the outer wall of a target cell and provide a confident pathway for attending nucleic acid into the target cell. Hollow spikes on the capsid fulfill the latter two roles, from which it has been deduced that they must have unusually high strength and stiffness in axial compression [3]. The spike protein (S protein) is a type I transmembrane protein, a sequence of amino acids ranging from 1,160 for avian infectious bronchitis virus (IBV) and up to 1,400 amino acids for feline coronavirus [4]. Spike proteins assemble to create the distinctive "corona" or crown-like look in trimers on the surface of the virion. The ectodomain of all CoV spike proteins portions the same organization in two domains: a receptor-binding N-terminal domain called S1 and a fusion-responsible C-terminal S2 domain The variable spike proteins (S proteins) reflect CoV diversity, which have developed into forms that vary in their receptor interactions and their reaction to different environmental triggers of virus-cell membrane fusion [5]. A notable peculiarity between the spike proteins of diverse coronaviruses is whether it is cleaved or not during assembly and exocytosis of virions [6]. The entry Herpesvirus and membrane fusion to the host cell equire three virion glycoproteins, gB and a gH/gL heterodimer, that function as the "core fusion machinery" [7]. The viral envelope of Ebola virus contains spikes consisting of the glycoprotein (GP) trimer. These viral spike glycoprotein docks viral particles to the cell surface, endosomal entry, and membrane fusion [8]. Acute respiratory syndrome coronavirus (SARS-CoV) is a zoonotic pathogens that traversed the species barriers to infect humans. These coronaviruses hold a surface-located spike (S) protein that recruits infection by mediating receptor-recognition and membrane fusion [9]. Bioinformatics plays an important role in all aspects of protein analysis, including sequence analysis, and

evolution analysis. In sequence analysis, several bioinformatics techniques can be used to provide the sequence comparisons. In evolution analysis, we use the technique like phylogenetic trees to find homologous proteins and identified the most related taxa. With bioinformatics techniques and databases, function, structure, and evolutionary history of proteins can be easily identified [10]. Several bioinformatics methods can be used in sequence analysis to provide sequence comparisons.

Materials and Methods

Sequences, alignment, and construction of phylogenetic tree

Amino acids sequences for the spike protein of Covid-19, Ebola, and herpes simplex viruses from bats, murine as well as Homo sapiens were taken from the National Center for Biotechnology Information database. The accession numbers of the corresponding database entries and species names are listed in Table 1 and Figure 1. Sequences were aligned with by CLUSTALW [11]. Selection of conserved blocks was performed using GBlocks (version 0.91 to eliminate divergent regions and poorly aligned positions using standard settings according to Castresana [12]. The Akaike information criterion (AIC) were performed, using a maximum-likelihood (ML) starting tree to estimates the quality of each model, relative to each of the other models (Table 2). The most suitable model was WAG+G+F [13]. Then phylogenetic tree was constructed with neighbor-joining and maximum likelihood algorithms using MEGA version 5 [14]. The stability of the topology of the phylogenetic tree was assessed using the bootstrap method with 100 repetitions [15].

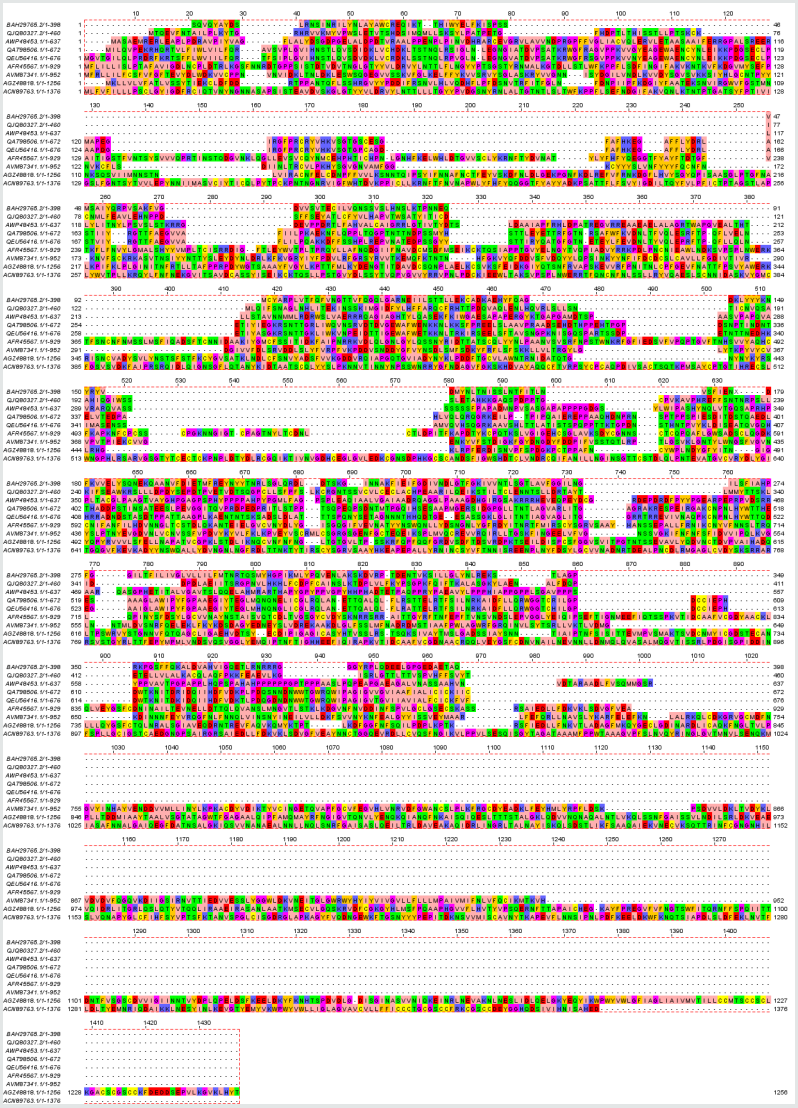

Figure 1: Multiple sequence alignment of Spike protein of coronaviruses (CoVs), Ebola, and herpes simplex viruses from human , bat and Murine.

Table 1: Accession numbers of NCBI entries for Spike protein in different species.

Name of Species	Accession No.	Length of Sequence [Amino Acids]
Coronaviruses (CoVs)		
Homo sapiens(covid-19)	AFR45567.1	929
Murine-	ACN89763.1	1370
Bat-	AGZ48818.1	1256
Herpesvirus		
Homo sapiens	AWP48453.1	637
Murine-	QJQ80327.2	460
Bat-	BAH29765.2	398
Ebola Virus		
Homo sapiens	AVM87341.1	952
Murine-	QEU56416.1	676
Bat-	QAT98506.1	672

Table 2: Maximum Likelihood fits of 56 different amino acid substitution models.

No.	Model	Parameters	BIC	AICc	lnL
1.	WAG+G+F	35	37502.06197	37260.8	-18595.22611
2.	WAG+F	34	37502.97665	37268.6	-18600.13502
3.	WAG+G+I+F	36	37510.9536	37262.8	-18595.22036
4.	WAG+I+F	35	37511.38078	37270.12	-18599.88552
5.	WAG+G	16	37598.57351	37488.2	-18728.06167
6.	WAG	15	37601.8263	37498.34	-18734.13963
7.	WAG+G+I	17	37607.47664	37490.21	-18728.06167
8.	WAG+I	16	37610.70715	37500.33	-18734.12849
9.	Dayhoff+G+F	35	37614.95526	37373.69	-18651.67276
10.	JTT+G+F	35	37620.10854	37378.84	-18654.2494
11.	Dayhoff+G+I+F	36	37623.86163	37375.71	-18651.67437
12.	LG+G+F	35	37626.68024	37385.41	-18657.53525
13.	JTT+F	34	37628.80767	37394.43	-18663.05053
14.	JTT+G+I+F	36	37629.01185	37380.86	-18654.24948
15.	LG+G+I+F	36	37635.065	37386.92	-18657.27606
16.	JTT+I+F	35	37637.49419	37396.23	-18662.94222
17.	JTT+G	16	37641.1497	37530.77	-18749.34977
18.	Dayhoff+F	34	37642.35268	37407.97	-18669.82304
19.	JTT+G+I	17	37650.06413	37532.79	-18749.35541
20.	Dayhoff+I+F	35	37650.54218	37409.28	-18669.46622
21.	JTT	15	37652.90999	37549.43	-18759.68148
22.	LG+F	34	37659.25529	37424.87	-18678.27434
23.	JTT+I	16	37661.81545	37551.44	-18759.68264
24.	LG+I+F	35	37664.26168	37423	-18676.32597
25.	rtREV+G+F	35	37697.96732	37456.7	-18693.17878
26.	rtREV+G+I+F	36	37705.34115	37457.19	-18692.41413
27.	rtREV+F	34	37734.78779	37500.41	-18716.04059
28.	rtREV+I+F	35	37736.91791	37495.65	-18712.65408
29.	LG+G	16	37759.71264	37649.34	-18808.63124
30.	LG+G+I	17	37768.51458	37651.24	-18808.58064

31.	LG	15	37795.82397	37692.34	-18831.13847
32.	LG+I	16	37802.23609	37691.86	-18829.89296
33.	cpREV+G+I+F	36	37817.94486	37569.8	-18748.71599
34.	cpREV+G+F	35	37840.278	37599.01	-18764.33413
35.	cpREV+I+F	35	37865.57802	37624.31	-18776.98414
36.	Dayhoff+G	16	37867.26086	37756.88	-18862.40535
37.	Dayhoff+G+I	17	37876.16537	37758.9	-18862.40603
38.	mtREV24+G+F	35	37909.62484	37668.36	-18799.00755
39.	Dayhoff	15	37909.97971	37806.5	-18888.21634
40.	mtREV24+G+I+F	36	37917.51344	37669.36	-18798.50028
41.	Dayhoff+I	16	37918.26135	37807.89	-18887.90559
42.	cpREV+F	34	37923.6936	37689.31	-18810.49349
43.	cpREV+G+I	17	37949.02605	37831.76	-18898.83637
44.	mtREV24+F	34	37976.53989	37742.16	-18836.91664
45.	mtREV24+I+F	35	37980.86198	37739.6	-18834.62612
46.	cpREV+G	16	37990.42602	37880.05	-18923.98793
47.	rtREV+G	16	38038.61068	37928.23	-18948.08026
48.	cpREV+I	16	38042.84828	37932.47	-18950.19906
49.	rtREV+G+I	17	38046.00616	37928.74	-18947.32643
50.	cpREV	15	38046.54766	37943.07	-18956.50031
51.	rtREV	15	38082.28692	37978.81	-18974.36994
52.	rtREV+I	16	38083.99992	37973.62	-18970.77488
53.	mtREV24+G	16	39271.828	39161.45	-19564.68892
54.	mtREV24+G+I	17	39280.71825	39163.45	-19564.68247
55.	mtREV24	15	39342.21581	39238.73	-19604.33439
56.	mtREV24+I	16	39348.93782	39238.56	-19603.24382

Computation of the theoretical pI (isoelectric point) of protein sequences

Estimation of the isoelectric point (pI) based on the amino acid sequence was determined using isoelectric Point Calculator (IPC), a web service and a standalone program for the accurate estimation of protein and peptide pI using different sets of dissociation constant (pKa) values [16]. Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites is evolutionarily invariable (+I). Abbreviations: TR: General Time Reversible; JTT: Jones-Taylor-Thornton; rtREV: General Reverse Transcriptase; cpREV: General Reversible Chloroplast; mtREV24: General Reversible Mitochondrial.

Results

In order to unravel the phylogenetic relationship of Spike protein between the different taxa, a phylogenetic consensus tree was constructed using the Bayesian Inference (BI) and Maximum Likelihood (ML) methods (Figure 1). The present results revealed that the identity between different taxa was nonsignificant, However alignment of human Coronavirus (Covid-19) and Bat Coronavirus revealed identity equal to 57.98% followed by Murine Coronavirus which displayed 27.61% when compared by human Coronavirus. The chemical composition of the tested protein is illustrated in Table 3. Figures 2 & 3 show the correlation plots between the theoretical isoelectric points for spike protein of different corona virus in different species. The current results displayed that estimated pI of spike protein sequence in three investigated viruses in different species ranges from 5.59 to 8.08. With highest value for spike protein in herpesvirus of bat. The Estimated charge over pH range of the investigated were listed in Table 4. The current results revealed that the behavior of s protein in different species exhibited different estimated charge at different pH ranges. Some S protein reveled low negative charge as pH increases such as Ebola Virus of bat while S protein of Corona virus in murine showed high negative charge as pH elevated.

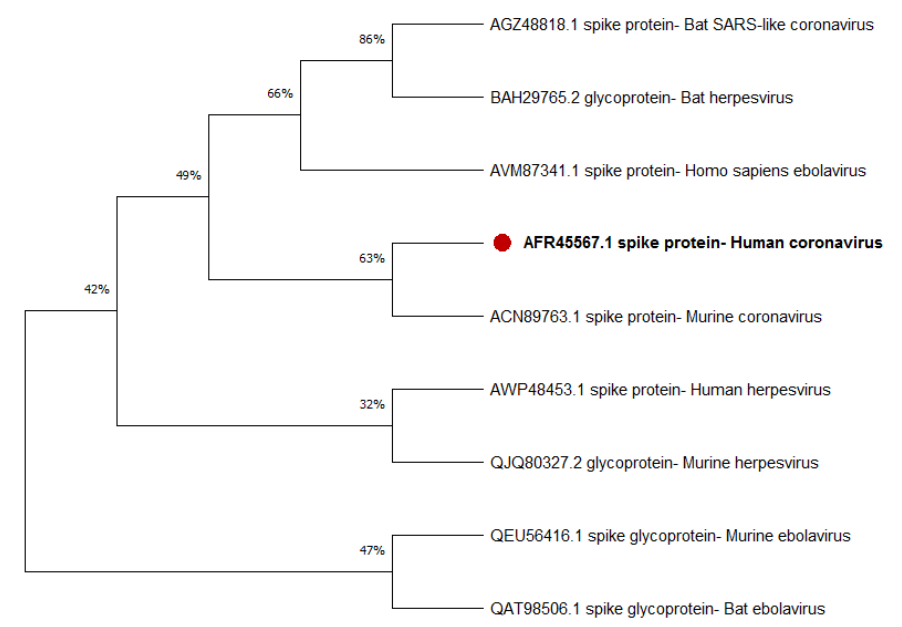

Figure 2: The phylogenetic tree shows the relationship of the Spike protein of coronaviruses (CoVs)with protein sequences from other species. The maximum likelihood tree is based on complete coding sequences.

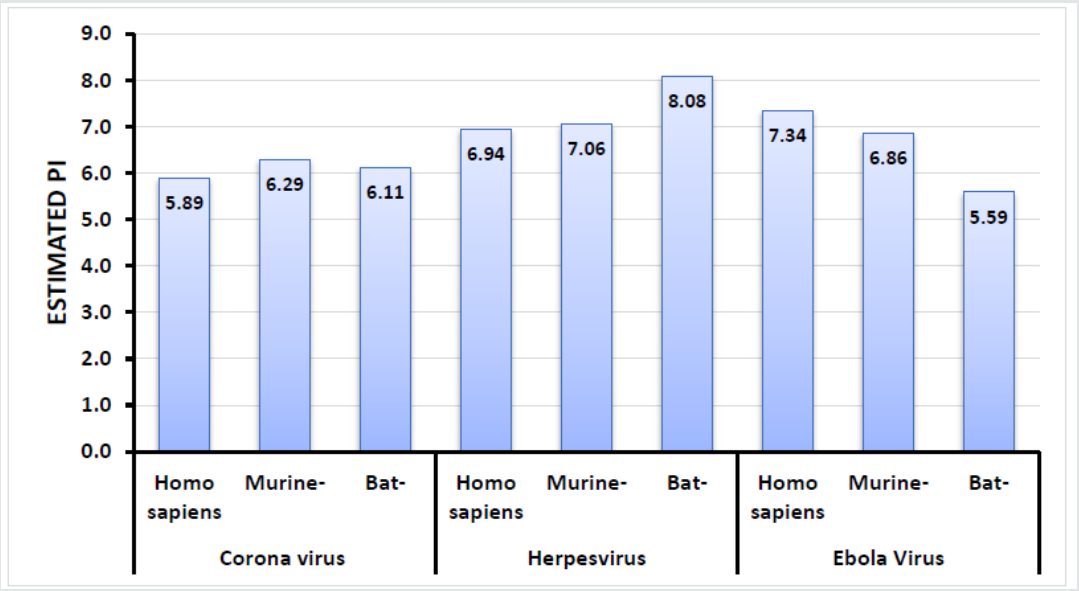

Figure 3: Estimated pI (isoelectric point) of spike protein sequences.

Table 3: Chemical composition of the Spike protein.

Residue	Coronaviruses (CoVs)			Herpesvirus			Ebola Virus		
	Homo sapiens	Murine-	Bat-	Homo sapiens	Murine-	Bat-	Homo sapiens	Murine-	Bat-
Alanine	46	92	82	103	29	21	23	50	46
Arginine	29	49	42	49	13	19	38	33	38
Asparagine	78	110	87	10	17	29	65	39	34
Aspartate	51	65	71	26	18	16	64	34	34
Glutamine	30	50	54	17	16	19	21	27	29
Glutamate	28	50	43	34	23	21	36	35	48
Glycine	60	96	77	49	14	27	61	53	46
Histidine	12	16	14	23	17	6	8	17	17
Isoleucine	49	80	77	16	24	28	44	41	48
Leucine	81	124	102	48	57	42	97	52	54
Lysine	38	55	60	5	25	20	63	31	24
Methionine	14	16	17	10	9	7	17	4	6
Phenylalanine	58	64	83	11	26	21	72	29	25
Proline	41	66	62	98	31	12	25	38	54
Serine	77	108	98	40	41	27	66	45	40
Threonine	73	92	93	28	42	26	33	71	58
Tyrosine	52	72	51	21	13	22	70	16	11
Valine	64	95	93	40	24	28	114	35	32
Tryptophan	8	16	11	4	4	2	9	14	15
Cysteine	40	60	39	5	17	4	26	12	13

Table 4: Estimated charge over pH range.

Name of Species	pH range												
	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Corona virus													
Homo sapiens	56.6	34.9	15.5	4.7	-1.1	-5.8	-10.3	-15	-22.6	-35.4	-51.5	-71	-97
Murine-	87.3	55.8	27.6	11.9	3.8	-2.8	-8.9	-15.6	-26.7	-45.5	-69	-97.2	-134.1
Bat-	83.6	52.4	24.5	9.1	1.3	-4.6	-9.7	-14.8	-22.5	-35.5	-52.6	-75.2	-106.7
Herpesvirus													
Homo sapiens	60	43.4	28.4	19.3	12.9	5.9	-0.6	-4.6	-7.2	-9.8	-13	-17.9	-25.3
Murine-	43.4	32	21.7	15.5	10.9	5.6	0.5	-3.3	-7.4	-13.3	-20.2	-28.6	-39.6
Bat-	34.6	24.4	15.3	10.2	7.4	5.2	3.3	1.8	0.3	-2	-5.8	-12.7	-24.1
Ebola Virus													
Homo sapiens	80.6	53.2	28.8	15.6	9.4	5.5	2.2	-1.2	-6.8	-16.8	-31.8	-55.7	-92.1
Murine-	61.4	42.4	25.3	15.5	9.6	3.9	-1.2	-4.9	-8.3	-13	-19.3	-28.3	-41.4
Bat-	55.7	33.2	12.9	1.5	-5.1	-11	-6.2	-9.9	-3.4	-8.2	-3.9	-1.3	-1.5

Discussion

One aspect that may provide some insight into the interactions of the S protein is the electrostatic potential it generates. the affinity constant for the receptor-binding domain (RBD) of viron protein potentially contributing to its transmission efficiency [17]. The S protein remains predominantly in the closed conformation to mask its receptor-binding domains (RBDs), thereby impeding their binding. To bind with ACE2, the S protein transforms into its open conformation, revealing its binding interface [18]. The present study relies on a comparative investigation, regarding the identity of spike protein of 3 viruses (Covid-19, Ebola, and herpes simplex) to identify the most related taxa. Our investigation revealed high

similarities of Spike protein of coronaviruses in human and bats [19]. The computed amino acid composition of spike protein. Several residues showed a significant difference between the compositions in spike proteins in the three investigated virus in different species. This result reveals the importance of specific residues in these classes of proteins. The polar uncharged residues, especially Serine, Asparagine, and Glycine, have higher occurrence in spike protein of murine, which are important for the folding, stability, and function of such class of proteins [20]. Our analysis revealed that the Coronavirus spike protein of murine may be more efficient in discovering suitable vaccine for Coronavirus. The S protein amino acids variations among different coronaviruses such as (SARS, herpes and ebola). The SARS-CoV-2 virus shares 57.98% of its genome with the other bat coronaviruses. The sequence identity in the S protein bat coronavirus appears to be the closest relative of SARS-CoV-2. Our results in accordance with Rothan [21]. The isoelectric point (pI) is the pH value at which a molecule's net charge is zero [22]. Information about the isoelectric point is important because the solubility and electrical repulsion are lowest at the pI. Hence, the tendency to aggregation and precipitation is highest. In the case of viruses, the value thus provides information about the viral surface charge in a specific environment [23]. In polar media, such as water, viruses possess a pH-dependent surface charge [23]. This electrostatic charge governs the movement of the soft particle in an electrical field and thus manages its colloidal behavior, which plays a major role in the processes of virus entry. The pH value at which the net surface charge changes its sign is called the isoelectric point and is a characteristic parameter of the virion in equilibrium with its atmosphere in water chemistry [24]. For some viruses the attachment influences that encourage binding to accommodating cells are extremely definite, but the arrangement of actions that activates viral entry is only now establishment to be understood. The charge of attachment protein may play an important role in attachment and entry of virus [25]. The current results revealed that the pH affect the net charge of S protein of different taxa with different behavior. All the investigated taxa exhibit increases in negative charge as the pH increased except for the Ebola virus form Bats which showed unstable behavior regarding the S protein charge.

Conclusion

In our study, we have investigated the variation of pH-dependent changes in charges of a protein. The current results revealed that the pH affect the net charge of S protein of different taxa with different behavior.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Gelderblom HR (1996) Structure and classification of viruses. In: Medical Microbiology 4th edition. University of Texas Medical Branch at Galveston, USA.
2. Arvin A, Campadelli Fiume G, Mocarski E, Moore PS, Roizman B, et al. (2007) Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, UK.
3. Feughelman M (2002) Natural protein fibers. Journal of Applied Polymer Science 83(3): 489-507.
4. Belouzard S, Millet JK, Licitra BN, Whittaker GR (2012) Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4(6): 1011-1033.
5. Li F (2016) Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol 3(1): 237-261.
6. Fehr AR, Perlman S (2015) Coronaviruses: an overview of their replication and pathogenesis. In: Coronaviruses. Springer p. 1-23.
7. Eisenberg RJ, Atanasiu D, Cairns TM, Gallagher JR, Krummenacher C, et al. (2012) Herpes virus fusion and entry: a story with many characters. Viruses 4(5): 800-832.
8. Beniac DR, Booth TF (2017) Structure of the Ebola virus glycoprotein spike within the virion envelope at 11 Åresolution. Sci Rep 7: 46374.
9. Lu G, Wang Q, Gao GF (2015) Bat-to-human: spike features determining 'host jump'of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol 23(8): 468-478.
10. Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5(4):276-287.
11. Thompson JD, Gibson TJ, Higgins DG (2003) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinforma 2: 3.
12. Castresana J (7000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4): 540-552.
13. Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B (Statistical Methodology 73(1): 3-36.
14. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10): 2731-2739.
15. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution (NY) 39(4): 783-791.
16. Kozlowski LP (2016) IPC--isoelectric point calculator. Biol Direct 11(1): 1-16.
17. Wang Y, Liu M, Gao J (2020) Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. Proc Natl Acad Sci 117(25): 13967-13974.
18. Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, et al. (2020) Citizen scientists create an exascale computer to combat COVID-19. BioRxiv.
19. Lau SKP, Li KSM, Tsang AKL, Lam CSF, Ahmed S, et al. (2013) Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory sy. J Virol 87(15): 8638-8650.

20. Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. Bioinforma Genet 317: 280-289.

21. Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 109: 102433.

22. Kozlowski LP (2017) Proteome-pI: proteome isoelectric point database. Nucleic Acids Res 45(D1): D1112-D1116.

23. Michen B, Graule T (2020) Isoelectric points of viruses. J Appl Microbiol 109(2): 388-397.

24. Langlet J, Gaboriaud F, Gantzer C, Duval JFL (2008) Impact of chemical and structural anisotropy on the electrophoretic mobility of spherical soft multilayer particles: the case of bacteriophage MS2. Biophys J 94(8): 3293-3312.

25. Connell BJ, Lortat-Jacob H (2013) Human immunodeficiency virus and heparan sulfate: from attachment to entry inhibition. Front Immunol 4: 385.

DOI: 10.32474/LOJPCR.2021.02.000140