



Brief Report

A Novel Synonymous Mutation of SARS-CoV-2: Is This Possible to Affect Their Antigenicity and Immunogenicity?

Sung-Jae Kim ^{1,†} , Van-Giap Nguyen ^{2,†}, Yong-Ho Park ^{3,*}, Bong-Kyun Park ¹ and Hee-Chun Chung ^{1,*} 

¹ Department of Veterinary Medicine Virology Lab, College of Veterinary Medicine and Research Institute for Veterinary Science, Seoul National University, Seoul 08826, Korea; kvirus7734@gmail.com (S.-J.K.); parkx026@snu.ac.kr (B.-K.P.)

² Department of Veterinary Microbiology and Infectious Diseases, Faculty of Veterinary Medicine, Vietnam National University of Agriculture, Hanoi 100000, Vietnam; nguyengiap83@gmail.com

³ Department of Veterinary Microbiology, College of Veterinary Medicine and Research Institute for Veterinary Science, Seoul National University, Seoul 151-742, Korea

* Correspondence: yhp@snu.ac.kr (Y.-H.P.); heeskyi@snu.ac.kr (H.-C.C.); Tel.: +82-2-880-1255 (Y.-H.P. & H.-C.C.); Fax: +82-2-885-0263 (Y.-H.P. & H.-C.C.)

† The authors contributed equally to this work.

Received: 15 April 2020; Accepted: 12 May 2020; Published: 14 May 2020



Abstract: The S glycoprotein of coronaviruses is important for viral entry and pathogenesis with most variable sequences. Therefore, we analyzed the S gene sequences of SARS-CoV-2 to better understand the antigenicity and immunogenicity of this virus in this study. In phylogenetic analysis, two subtypes (SARS-CoV-2a and -b) were confirmed within SARS-CoV-2 strains. These two subtypes were divided by a novel synonymous mutation of D614G. This may play a crucial role in the evolution of SARS-CoV-2 to evade the host immune system. The region containing this mutation point was confirmed as a B-cell epitope located in the S1 domain, and SARS-CoV-2b strains exhibited severe reduced antigenic indexes compared to SARS-CoV-2a in this area. This may allow these two subtypes to have different antigenicity. If the two subtypes have different serological characteristics, a vaccine for both subtypes will be more effective to prevent COVID-19. Thus, further study is urgently required to confirm the antigenicity of these two subtypes.

Keywords: COVID-19; SARS-CoV-2; spike protein; antigenicity

Coronavirus (CoV) is a class of genetically diverse RNA viruses found in a wide range of hosts including reptiles, birds, and mammals. Most pathogenic CoVs usually cause respiratory and intestinal symptoms in animals [1–5]. Over the past 20 years, a few novel beta-coronaviruses originated from bats have been transmitted to humans and caused severe respiratory syndrome. SARS-CoV and MERS-CoV were first introduced to humans in 2002 and 2012, respectively [3]. Recently, a new novel beta-coronavirus named SARS-CoV-2 first broke out in Wuhan city, China. The first case of this virus was reported in December 2019 and this virus has subsequently spread explosively worldwide and severely threatened human health [3,6].

CoV is generally composed of four major structural proteins: nucleocapsid protein (N), membrane (M), envelope (E), and spike glycoprotein (S). Among these proteins, the S glycoprotein plays crucial roles in viral entry and pathogenesis as its widely exposed structure forms large petal-shaped spikes on the surface of the virion [7]. Mutations in the spike glycoprotein can allow novel coronavirus strains to infect humans and spread pandemically [8]. Therefore, S gene encoding S glycoprotein has widely

been used for molecular analysis of coronaviruses due to the significant features of the S glycoprotein affecting the antigenicity and immunogenicity [2,4,9–13].

Thus far, there has been little data comparing and analyzing S gene sequences within SARS-CoV-2. Generally, several types of coronavirus are divided into subtypes depending on amino acid mutations in S gene sequences, and molecular analysis based on the S gene can provide insights into antigenicity, immunogenicity, or evolutionary trends [2,4,10,12,13]. Thus, we analyzed the S gene sequences of SARS-CoV-2 to better understand this virus in this study.

For phylogenetic analysis based on the S gene, 144 sequences of SARS-CoV-2 that globally originated from several countries (China, USA, Italy, Spain, Japan, Vietnam, Taiwan, and Pakistan) were retrieved from GenBank. Using IQ-TREE v1.6.12 [14], the genetic relationships between SARS-CoV-2 were inferred by the maximum likelihood (ML) method. The “-m MFP” option was invoked to help select the data best-fit amino acid substitution model. The branch support values were estimated by ultrafast bootstrap approximation [15] implemented in IQ-TREE [14] via the “-bb 1000” option. The reconstructed phylogenies were displayed and midpoint rooted by FigTree v1.4.3. In the ML tree, completely divided clades were identified among the analyzed SARS-CoV-2 strains (Figure 1A).

Interestingly, only one reliable synonymous change was found to distinguish between subtypes A and B in this study. SARS-Cov-2a and -2b strains consistently exhibited Ala (D) and Gly (G) at the amino acid sequence position 614, respectively (Figure 1B). The virus consistently evolves to evade the host immune system with synonymous mutations that are so-called positive selection. More evolved viruses are better able to survive, thus such viruses will likely be dominant within the group [16]. SARS-CoV-2a includes the China strains confirmed in 2019, but SARS-CoV-2b only includes the USA strains confirmed after 2020. In addition, the USA was one of the latest countries to experience a COVID-19 outbreak. Although there was only a few months’ difference, SARS-CoV-2b may be a more evolved form. If the mutation of D614G plays a crucial role in the positive selection process, SARS-CoV-2b will be the dominant type of SARS-CoV-2 in the future. More long-term tracking will be required to validate this assumption.

The S glycoprotein is important for viral entry and pathogenesis with the most variable sequences in the coronavirus genomes. Human beta-coronavirus S proteins are cleaved into S1 and S2 subunits by host proteases [17]. The S1 subunit forming a globular shape is responsible for receptor binding [18] while the S2 subunit forming a rod shape mediates membrane fusion [19]. More specifically, the S1 subunit is composed of two major domains (S1-NTD and S1-CTD) and two sub-domains (SD-1 and SD-2). One or both of the major domains is potentially responsible for binding host-receptors, and the sub-domains that are complex folding of elements may allow receptor-induced conformational changes [11,20,21]. Thus, mutations within the S1 region are associated with changes in antigenicity and viral pathogenicity [22].

In fact, the S1 subunit contains numerous major and minor neutralizing antibody epitopes [21,23]; thus, it is difficult to investigate all putative epitopes to discover which can play a crucial role on the antigenicity and immunogenicity of viruses. In this situation, inferences considering both aspects of virus evolution and epitope analysis can be helpful to investigate which epitopes really play a crucial role. B-cell epitopes on the S1 subunit were predicted by BepiPred-2.0 [24], the Chou & Fasman method [25], the Kolaskar and Tongaonkar method [26] and Parker’s Hydrophilicity [27]. Regions between amino acids 614 and 621 were equally identified as a B cell epitope by all four methods (Figure 1C). The predicted B-cell epitope including amino acid 614 was located in a relatively well-exposed part of the S1 subunit in the 3D-view structure (Figure 1D); this B-cell epitope sequence was identified within the sequence corresponding to the SD-1/-2 domain (Figure 1E).

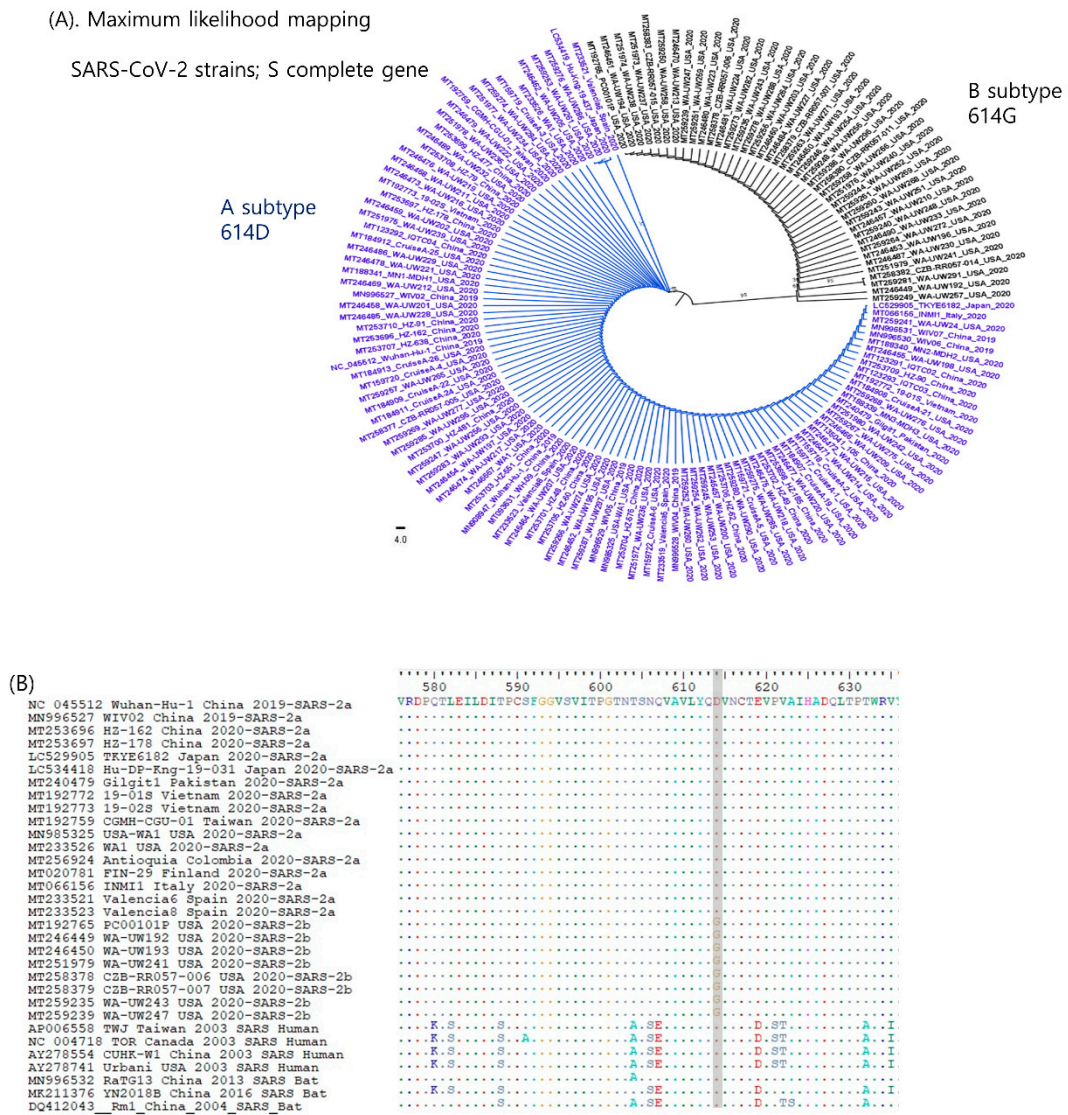


Figure 1. Cont.

(C). Predicted of B cell epitope;
Reference;GenBank: NC045512, Wuhan-Hu-1, China, 2019

BepiPred-2.0
Threshold: 0.5

```

...EEEE.....EEEEEE.....EEEEEEEEEEEEEEEEEEEE
PFGINTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFGT
-600-----610-----620-----630-----640-----
    
```

Predicted residue scores:

Method	Position	Residue	Start	End	Peptide	Score
Chou & Fasman Threshold: 1.0	614	D	611	617	LYQDVNC	1.06
	615	V	612	618	YQDVNCT	1.113
	616	N	613	619	QDVNCTE	1.056
Kolaskar & Tongaonkar Threshold: 1.041	613	Q	610	616	VLYQDVN	1.119
	614	D	611	617	LYQDVNC	1.123
	615	V	612	618	YQDVNCT	1.075
	616	N	613	619	QDVNCTE	1.03
	617	C	614	620	DVNCTEV	1.083
	618	T	615	621	VNCTEVP	1.111
	619	E	616	622	NCTEVPV	1.111
Parker Hydrophilicity Threshold: 1.238	620	V	617	623	CTEVPVA	1.152
	621	P	618	624	TEVPVAI	1.115
	614	D	611	617	LYQDVNC	1.371
	615	V	612	618	YQDVNCT	3.429
	616	N	613	619	QDVNCTE	4.814
	617	C	614	620	DVNCTEV	3.429
618	T	615	621	VNCTEVP	2.3	
619	E	616	622	NCTEVPV	2.3	
620	V	617	623	CTEVPVA	1.6	

(D). COVID-19 3D protein
Reference; PDB 6VXX

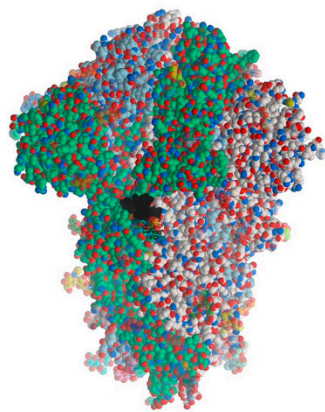
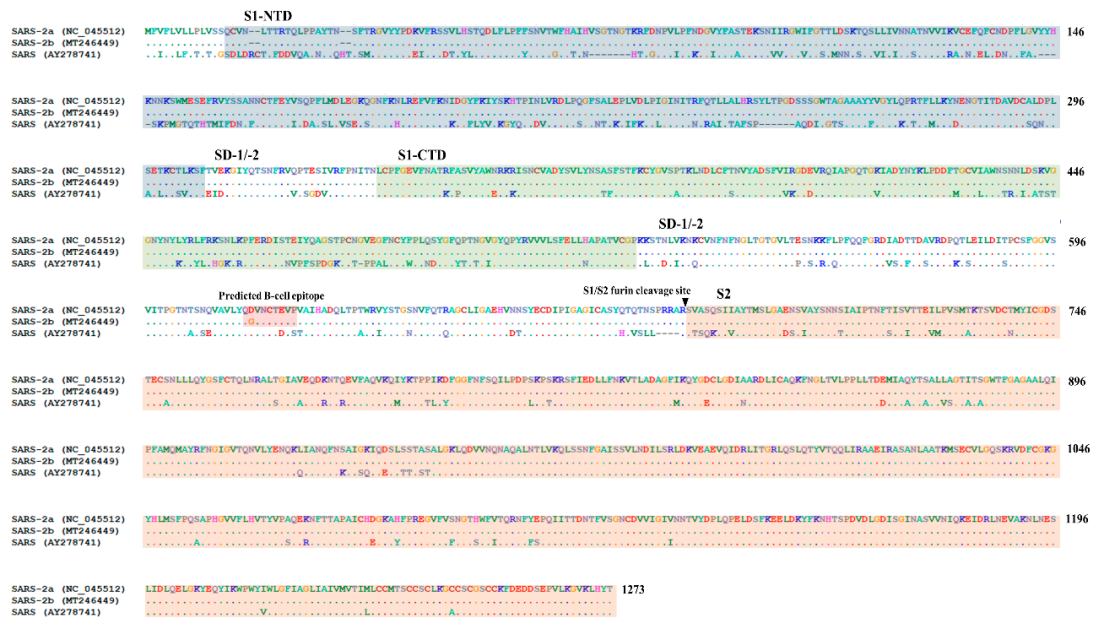


Figure 1. Cont.

(E). SARS-CoV-2 S gene sequence



(F). Jameson–Wolf antigenic index

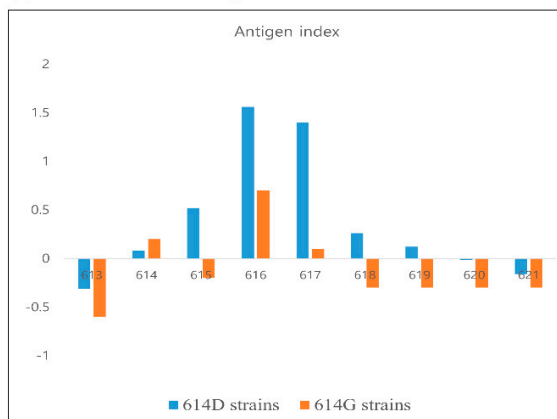


Figure 1. Characterization of the complete S gene in SARS-CoV-2. (A) Phylogenetic analysis of SARS-CoV-2 strains based on the S gene. The phylogenetic trees were reconstructed from 144 sequences of SARS-CoV-2 collected globally. Thus, the two subtypes (SARS-CoV-2a and -2b) were completely divided. (B) Alignment of SARS-CoV sequences including the aa 614 position are highlighted in gray. A novel reliable synonymous mutation was identified to distinguish the A and B subtypes. SARS-Cov-2a and -2b strains consistently exhibited Ala (D) and Gly (G) at the amino acid sequence position 614, respectively. (C) Identification of B-cell epitopes in the adjacent area with aa 614. The B-cell epitope was predicted by BepiPred-2.0 [24], the Chou & Fasman method [25], the Kolaskar and Tongaonkar method [26], and Parker’s Hydrophilicity [27]. The 614–621 region was predicted to consist of epitopes. (D) The 3D-structure of SARS-CoV-2 spike protein by Mol soft Mol Browser 3.8–5 according to the original publication from the National Center for Biotechnology Information (NCBI): PDB;6VXX. The predicted B-cell epitope (aa 613–620) highlighted in black color was located at a relatively well-exposed part. (E) Sequence alignment of SARS-CoV-1 and -2. The S1 subunit is responsible for receptor binding and the S2 subunit mediates membrane fusion. The S1 subunit consists of two major domains capable of binding to host receptors: an amino (N)-terminal domain (NTD) and a carboxy (C)-terminal domain (CTD) and two sub-domains that may allow receptor-induced conformational changes: SD-1 and SD-2. (F) The antigenic index of each amino acid constituting this region (amino acids 613–621) by the Jameson–Wolf method [28].

The antigenic indexes of each amino acid in this region (amino acids 613–621) were calculated by the Jameson–Wolf method. When an antigen index was >0.5 , it was believed to be a reliable position as an epitope [28]. The results of the antigenic index analysis showed severely reduced indexes of amino acids 615–617 in the SARS-CoV-2b strains compared to SARS-CoV-2a; it is predicted that the change of D614G affects the antigenicity of this region (Figure 1F). Since no amino acid changes were found in this area other than the change of D614G, it is believed that this amino acid change alters the conformation of these immunogenic determinants; consequently, this region is expected to no longer act as a B-cell epitope in SARS-CoV-2b. B-cell plays a major role in recognizing pathogens and stimulating adaptive immunity in the immune response against virus infection. Thus, the elimination of B cell epitopes will likely reduce immunogenicity by hampering the immune cell recognition of the virus [29,30].

When reflecting the above results, SARS-CoV-2b may have reduced immunogenicity compared to SARS-CoV-2a. If so, it can permit persistent or recurrent infection of SARS-CoV-2b while evading immune cell recognition. In addition, the mutation of the S1 domain may induce different antigenicity and viral pathogenicity between the two subtypes. Different virus subtypes will likely have somewhat different serological features depending on their antigenicity, although there may be some cross-reactivity. In addition, a certain subtype can serologically cover other serotypes [12,31]. Thus, this point should be considered to create a new SARS-CoV-2 vaccine. Indeed, if the two serotypes have different serological characteristics, a vaccine that includes both subtypes will be more effective at preventing COVID-19, particularly when developing a killed vaccine that has a narrow protection range compared to a live vaccine. This study was confined to investigating the phylogenetic and genetic features of SARS-CoV-2 due to limited information. Therefore, further study on the cross-reactivity between these two subtypes is required to validate our assumption. Viruses have continually evolved through genetic mutations to evade host immune systems in the long history of the fight between humans and viruses. SARS-CoV-2, which has only recently been introduced in humans, will continue to evolve for survival in the current situation, in which this virus has already become a pandemic. To respond properly against this virus, continuous surveillance of this virus' adaptation to evade host immune systems is important in the future.

Ethics Approval and Consent to Participate: This article does not contain any studies with live animals performed by any of the authors.

Author Contributions: H.-C.C., Y.-H.P., S.-J.K. and B.-K.P. designed the study; H.-C.C. and V.-G.N. analyzed the data; H.-C.C., S.-J.K. and V.-G.N. interpreted the results and drew up the manuscript. H.-C.C., Y.-H.P. and B.-K.P. reviewed the manuscript. All authors read and approved the final manuscript.

Funding: This study was supported by the BioGreen 21 Program, Rural Development Administration (grant no. PJ009015). The funder had no role in this study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments: The authors would like to thank Jung Ah Kim and Eun Ok Kim for excellent technical assistance.

Conflicts of Interest: The authors declare that they have no competing interests.

Consent for Publication: Not applicable.

Availability of Data and Material: The full-length genomes of SARS-CoV 2 strains used are registered within GenBank.

References

1. Woo, P.C.; Huang, Y.; Lau, S.K.; Yuen, K.-Y. Coronavirus genomics and bioinformatics analysis. *Viruses* **2010**, *2*, 1804–1820. [[CrossRef](#)]
2. Martínez, N.; Brandão, P.E.; de Souza, S.P.; Barrera, M.; Santana, N.; de Arce, H.D.; Pérez, L.J. Molecular and phylogenetic analysis of bovine coronavirus based on the spike glycoprotein gene. *Infect. Genet. Evol.* **2012**, *12*, 1870–1878. [[CrossRef](#)] [[PubMed](#)]

3. Sun, J.; He, W.-T.; Wang, L.; Lai, A.; Ji, X.; Zhai, X.; Li, G.; Suchard, M.A.; Tian, J.; Zhou, J. COVID-19: Epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol. Med.* **2020**. [[CrossRef](#)] [[PubMed](#)]
4. Chung, H.-C.; Nguyen, V.G.; Le Huynh, T.M.; Moon, H.-J.; Kang, B.-K.; Kim, S.-J.; Kim, H.-K.; Park, S.-J.; Park, K.-T.; Park, Y.-H. Molecular characterization of a Korean porcine epidemic diarrhea virus strain NB1. *Can. J. Vet. Res.* **2019**, *83*, 97–103. [[PubMed](#)]
5. Li, J.; Li, J.J.; Xie, X.; Cai, X.; Huang, J.; Tian, X.; Zhu, H. Game consumption and the 2019 novel coronavirus. *Lancet Infect. Dis.* **2020**, *20*, 275–276. [[CrossRef](#)]
6. Zehender, G.; Lai, A.; Bergna, A.; Meroni, L.; Riva, A.; Balotta, C.; Tarkowski, M.; Gabrieli, A.; Bernacchia, D.; Rusconi, S. Genomic Characterisation and Phylogenetic Analysis of Sars-Cov-2 in Italy. *J. Med. Virol.* **2020**. [[CrossRef](#)]
7. Abraham, S.; Kienzle, T.E.; Lapps, W.; Brian, D.A. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology* **1990**, *176*, 296–301. [[CrossRef](#)]
8. Menachery, V.D.; Yount, B.L., Jr.; Debbink, K.; Agnihothram, S.; Gralinski, L.E.; Plante, J.A.; Graham, R.L.; Scobey, T.; Ge, X.-Y.; Donaldson, E.F. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **2015**, *21*, 1508. [[CrossRef](#)]
9. Du, L.; Zhao, G.; Li, L.; He, Y.; Zhou, Y.; Zheng, B.-J.; Jiang, S. Antigenicity and immunogenicity of SARS-CoV S protein receptor-binding domain stably expressed in CHO cells. *Biochem. Biophys. Res. Commun.* **2009**, *384*, 486–490. [[CrossRef](#)]
10. Sánchez, C.M.; Gebauer, F.; Suñé, C.; Mendez, A.; Dopazo, J.; Enjuanes, L. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* **1992**, *190*, 92–105. [[CrossRef](#)]
11. Kirchdoerfer, R.N.; Cottrell, C.A.; Wang, N.; Pallesen, J.; Yassine, H.M.; Turner, H.L.; Corbett, K.S.; Graham, B.S.; McLellan, J.S.; Ward, A.B. Pre-fusion structure of a human coronavirus spike protein. *Nature* **2016**, *531*, 118–121. [[CrossRef](#)] [[PubMed](#)]
12. Wang, X.; Chen, J.; Shi, D.; Shi, H.; Zhang, X.; Yuan, J.; Jiang, S.; Feng, L. Immunogenicity and antigenic relationships among spike proteins of porcine epidemic diarrhea virus subtypes G1 and G2. *Arch. Virol.* **2016**, *161*, 537–547. [[CrossRef](#)] [[PubMed](#)]
13. Ntafis, V.; Mari, V.; Decaro, N.; Papanastassopoulou, M.; Pardali, D.; Rallis, T.S.; Kanellos, T.; Buonavoglia, C.; Xylouri, E. Canine coronavirus, Greece. Molecular analysis and genetic diversity characterization. *Infect. Genet. Evol.* **2013**, *16*, 129–136. [[CrossRef](#)] [[PubMed](#)]
14. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
15. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)] [[PubMed](#)]
16. Pereira, C.A.; Leal, E.S.; Durigon, E.L. Selective regimen shift and demographic growth increase associated with the emergence of high-fitness variants of canine parvovirus. *Infect. Genet. Evol.* **2007**, *7*, 399–409. [[CrossRef](#)] [[PubMed](#)]
17. Zaki, A.M.; Van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.; Fouchier, R.A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [[CrossRef](#)]
18. Kubo, H.; Yamada, Y.K.; Taguchi, F. Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol.* **1994**, *68*, 5403–5410. [[CrossRef](#)]
19. Luo, Z.; Weiss, S.R. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. *Virology* **1998**, *244*, 483–494. [[CrossRef](#)]
20. Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **2016**, *3*, 237–261. [[CrossRef](#)]
21. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veelsler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**. [[CrossRef](#)] [[PubMed](#)]
22. Ballesteros, M.; Sanchez, C.; Enjuanes, L. Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism. *Virology* **1997**, *227*, 378–388. [[CrossRef](#)]
23. Jiang, S.; He, Y.; Liu, S. SARS vaccine development. *Emerg. Infect. Dis.* **2005**, *11*, 1016. [[CrossRef](#)] [[PubMed](#)]
24. Jespersen, M.C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **2017**, *45*, W24–W29. [[CrossRef](#)] [[PubMed](#)]

25. Chou, P.Y.; Fasman, G.D. Prediction of protein conformation. *Biochemistry* **1974**, *13*, 222–245. [[CrossRef](#)] [[PubMed](#)]
26. Kolaskar, A.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* **1990**, *276*, 172–174. [[CrossRef](#)]
27. Parker, J.; Guo, D.; Hodges, R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **1986**, *25*, 5425–5432. [[CrossRef](#)]
28. Jameson, B.; Wolf, H. The antigenic index: A novel algorithm for predicting antigenic determinants. *Bioinformatics* **1988**, *4*, 181–186. [[CrossRef](#)]
29. Lin, Y.-M.; Jow, G.-M.; Mu, S.-C.; Chen, B.-F. Naturally occurring hepatitis B virus B-cell and T-cell epitope mutants in hepatitis B vaccinated children. *Sci. World J.* **2013**, *2013*. [[CrossRef](#)]
30. Nagata, S.; Pastan, I. Removal of B cell epitopes as a practical approach for reducing the immunogenicity of foreign protein-based therapeutics. *Adv. Drug Deliv. Rev.* **2009**, *61*, 977–985. [[CrossRef](#)]
31. Wilson, S.; Illambas, J.; Siedek, E.; Stirling, C.; Thomas, A.; Plevová, E.; Sture, G.; Salt, J. Vaccination of dogs with canine parvovirus type 2b (CPV-2b) induces neutralising antibody responses to CPV-2a and CPV-2c. *Vaccine* **2014**, *32*, 5420–5424. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).