Original Article

Bioinformatics analysis and comparison of the S protein of canine coronavirus and feline coronavirus

Bo Dong[1,3,4†*] Xiaowei Zhong[1,2†] Jiaqi Zhuo[1] Renyi Hu[1] Xiaodong Zhang[1,2] Xiaolian Xu[1] Haiyan Deng[1] Xiaofang Yan[1] Weiming Lin[1,3,4*]

Abstract

The Coronavirus S protein induces neutralizing antibodies, specific receptor binding, and membrane fusion between the virus and host cells. These changes are crucial to understanding the mechanism, genetic correlates, and prophylactic treatment of viral infection. In this study, the spike (S) protein structure and phylogenetic association of CCoV and FcoV were analyzed using bioinformatics. The results showed that the S proteins of CCoV and FCoV contained transmembrane regions and signal peptides. The S proteins of these coronaviruses were soluble. The S proteins of these strains also contained two functional domains and two motifs, i.e., corona_S1 and corona_S2. Phylogenetic analysis showed that CCoV-I and FCoV-I were included in the same clade, while CCoV-IIb, FCoV-II, and TGEV formed another clade. Our findings can provide a theoretical basis for further understanding the function of coronavirus S proteins and the design of antiviral drugs and vaccines.

Keywords: canine coronavirus, feline coronavirus, porcine transmissible gastroenteritis virus, bioinformatics analysis, epitope, phylogenetic analysis

[1]*College of Life Science of Longyan University, Longyan, China*
[2]*College of Animal Sciences, Fujian Agriculture and Forestry University, Fuzhou, China*
[3]*Engineering Research Center for the Prevention and Control of Animal Original Zoonosis, Fujian Province University, College of Life Science, Longyan University, Longyan, China*
[4]*Fujian Provincial Key Laboratory for the Prevention and Control of Animal Infectious Diseases and Biotechnology, Longyan, China*

†They contributed equally to this work
****Correspondence**: wmlin925@126.com, bodong1983@qq.com* (W. Lin, B. Dong)
Received April 24, 2023
Accepted December 11, 2023

Introduction

Canine enteric coronavirus (CCoV) belongs to the alphacoronavirus family, along with human coronavirus (229E and NL63), feline coronavirus (FCoV), porcine transmissible gastroenteritis virus (TGEV), and porcine epidemic diarrhea virus (PEDV) (Adams *et al.* 2016). The main symptoms of CCoV infection are mild and include those typical of gastroenteritis, e.g., diarrhea (Pratelli 2008). However, coinfection of dogs with canine parvovirus (CPV) causes intense symptoms and is associated with high mortality risk (Pratelli *et al.* 1999). CCoV was initially thought to cause only self-limiting enteritis and mild diarrhea (Keenan *et al.* 1976). However, several strains of CCoV with enhanced virulence have been reported, which are associated with symptoms including lethargy, anorexia, vomiting, bloody diarrhea, ataxia, and epilepsy (Alfano *et al.* 2020,Evermann *et al.* 2005).

CCoV mainly infects dogs, wolves, foxes, and other canine species but can also infect other animals, such as cats, pigs, and pandas (Gao *et al.* 2009). The expansion of these hosts is related to the spike (S) protein, which binds the host cell receptor and triggers the fusion of the virus with the cell membrane, mainly determining the host species, tissue and cell tropism, pathogenicity, and spread (Regan *et al.* 2012). The recombination or mutation of the S protein may result in cross-species virulence. A novel canine-feline recombinant alphacoronavirus has been isolated from a human patient with pneumonia (Vlasova *et al.* 2022). If confirmed as pathogenic, it may become the eighth unique coronavirus to cause disease in humans, potentially posing a threat to public health.

CCoV possesses two different serotypes: type I and type II (Le Poder 2011). Although CCoV-I and CCoV-II are closely related, their S proteins differ (Pratelli *et al.* 2003). CCoV-I and feline coronavirus type I (FCoV-I) originated from a common ancestor, whereas CCoV-II arose from multiple recombination events of unknown genetic origin (Lorusso *et al.* 2008). Based on the amino acid sequence of the N-terminal domain (NTD) of the S protein, CCoV-II can be divided into two sub-genotypes: CCoV-IIa and CCoV-IIb. CCoV-IIa has an NTD consistent with the prototype CCoV, but CCoV-IIb is genetically distinct from CCoV-IIa, and the S protein of CCoV-IIb has an NTD similar to that of TGEV (Regan *et al.* 2012). The emergence of TGEV appears to be based on the recombination of serotype CCoV with other coronaviruses (Decaro *et al.* 2007). FCoV-II also occurs because of an independent recombination event between FCoV-I and CCoV-II, thus allowing FCoV-II to acquire the CCoV-II S protein (Herrewegh *et al.* 1998).

The coronavirus S protein is a major antigenic determinant responsible for host cell receptor binding and viral entry. Aminopeptidase N (APN) is a common receptor for FCoV and TGEV (Wentworth and Holmes 2001). Although each virus is thought to utilize the species-specific homolog of its respective host during infection, the feline homolog is a common receptor for type II FCoV, type II CCoV, and TGEV (Tresnan *et al.* 1996).

In this study, we used bioinformatics software to analyze and compare the biological functions of the S proteins of CCoV and FCoV. The physical and chemical properties, transmembrane region, signal peptides, functional domains, protein modifications, and antigenic epitopes of the S proteins of several viruses were analyzed using bioinformatics software. This study will help to elucidate the biological characteristics of several coronaviruses and provide data for designing targeted vaccines and antiviral drugs based on the S protein.

Materials and Methods

Virus information: The nucleotide and amino acid sequences of CCoV-I Elmo/02 (GenBank ID: AY307020), CCoV-IIa A76 (GenBank ID: JN856008), CCoV-IIb 1-71 (GenBank ID: JQ404409), FCoV-I Black (GenBank ID: EU186072) and FCoV-II 79-1146 (GenBank ID: DQ010921) S genes were downloaded from the National Center for Biotechnology Information (Table 1).

Software: The physical and general biological characteristics of the CCoV and FCoV S proteins were calculated using ProtPsram and ProtScale tools on the ExPASy server. The transmembrane region (transmembrane helix (TMH)), signal peptide, phosphorylation site, and glycosylation site of the S protein was predicted using TMHMM Server v.2.0, SignalP 4.0, NetPhos 3.1 Server, and NetNGlyc 4.0Server software. The amino acid sequences of the S protein was submitted to Predicting Antigenic Peptides, SMART, and PROSITE, which were used to predict the epitopes and functional domains of each S protein sequence. Multiple Sequence Alignment and sequence analysis were performed using the Multiple Sequence Alignment tool of DNAMAN 6.0 software (Lynnon BioSoft, Point-Claire, Quebec, Canada). Sequences across different viral strains were compared using pairwise distances using the untitled ClustalW (weighted) method. Phylogenetic trees derived from the nucleotide sequences were constructed using MEGA version 5.2 using the neighbor-joining method with the p-distance model and 1,000 bootstrap replicates.

Table 1 S-gene sequences of coronaviruses.

Strain name	Genome source	Genome accession	Position in genome	Protein accession
CCoV-I	Elmo/02	AY307020	-	AAP72149
CCoV-IIa	A76	JN856008	20324-24745	AEQ61968
CCoV-IIb	1-71	JQ404409	20364-24725	AFG19726
FCoV-I	Black	EU186072	20404-24798	ABX60145
FCoV-II	79-1146	DQ010921	20206-24564	AAY32596

Results

Physical and Chemical Properties: The amino acid sequences of CCoV-I, CCoV-IIa, and CCoV-IIb spike proteins were loaded into ProtParam online software. The results of this analysis are presented in Table 2. The CCoV-I S protein encoded 1,480 amino acids, and its molecular weight and isoelectric point were 166076.85 and 4.93, respectively. The protein contained 150 negatively and 104 positively charged residues. The instability index of the CCoV-I S protein was 30.94, the aliphatic index was 83.44, and the total average hydrophilicity was -0.109. Moreover, CCoV-IIa and CCoV-IIb S proteins encoded 1,473 and 1,453 amino acids, respectively, with a molecular weight of 163099.13 and 160643.62, respectively, and an isoelectric point of 4.95 and 5.02 respectively. The CCoV-IIa protein contained 127 negatively and 88 positively charged residues. The instability index of the S protein was 28.77, the aliphatic index was 90.21, and the total average hydrophilicity was 0.006. In addition, the CCoV-IIb protein contained 130 negatively charged residues and 90 positively charged residues. The instability index of the S protein was 29.80, the aliphatic index was 91.00, and the total average hydrophilicity was 0.026.

The FCoV-I and FCoV-II S proteins contained 1,464 and 1,452 amino acids, respectively. Their molecular weights were 164320.18 and 160471.3, respectively. The isoelectric points were 5.41 and 5.03, respectively. The FCoV-I S protein contained 129 negatively and 97 positively charged residues. FCoV-II had 126 and 86 cells. The instability indices of the FCoV-I and FCoV-II S proteins were 32.67 and 28.18, respectively. The aliphatic indices were 88.31 and 89.92, respectively. In addition, the grand averages of hydropathicity were -0.043 and 0.033, respectively.

Hydrophilicity and Hydrophobicity: The hydrophilic and hydrophobic amino acid sequences of the S protein were analyzed using ExPASy-ProtScale software. CCoV-I had the strongest hydrophilic asparagine at 120aa and the most hydrophobic leucine (Leu) at 1430aa. The asparagine at 976aa of CCoV-IIa was the most hydrophilic, while Leu at 1423aa was the most hydrophobic. The Lys at 958aa and Arg at 959aa of CCoV-IIb were the most hydrophilic, while Leu at 6aa was the most hydrophobic. Additionally, asparagine at 792aa of the S protein of FCoV-I was the most hydrophilic, and Leu at 1414aa was the most hydrophobic. The predicted results of FCoV-II showed that the 956aa serine (Ser) of S protein was the most hydrophilic, and 1402aa (Leu) was the most hydrophobic (Fig. 1). Together, these results indicated that the S proteins of these coronaviruses were soluble.

Transmembrane Region of S Proteins: The amino acid sequences of CCoV and FCoV S proteins were predicted using TMHMM Server v2.0. The results showed that all CCoV and FCoV strains of coronavirus S proteins had transmembrane regions. Slight differences were observed between CCoV-I and CCoV-IIa. In CCoV-I, 1-1419aa are outside the membrane, 1420-1442 sites can form a typical transmembrane helix region, and 1443-1480aa are inside the membrane. In CCoV-IIa, 1-1412aa are outside the membrane, and 1413-1435aa could form a typical transmembrane helix region, and 1436-1473aa were inside the membrane. In CCoV-IIb, the 1393-1415aa of S protein could form a typical transmembrane helix region, and 1-1392aa were outside the membrane, and 1416-1453aa were in the membrane. In FCoV-I, the 1406-1428aa aa S protein could form a typical transmembrane helix region, and 1-1405aa were outside the membrane, and 1429-1464aa were in the membrane. In FCoV-II, 1-1393aa are outside the membrane, and 1394-1416aa could form a typical transmembrane helix region, and 1417-1452aa were inside the membrane (Fig. 2). The prediction results showed that the distribution of the S protein was similar to that of other coronaviruses, and most of the S proteins of these viruses were outside the virus envelope.

Signal Peptide of S Proteins: SignalP 4.1 was used to predict the signal peptide in the amino acid sequence of the S protein of CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I, and FCoV-II strains using the Neural Network (NN) model. The results showed that there may be a signal peptide in residues 1-16 of the N-terminus of the S protein of CCoV-I. The signal peptide sequence used was MKIFLLSALLAIANCKDEAGP. In addition, the predicted signal peptides of the other strains were CCoV-IIa, 1-16aa, MKVLLFLALFSIARCD; CCoV-IIb, 1-19aa, MIVLILCLLLFSYNSVICT; FCoV-I, 1-15aa, MIVLIFALLSTARSE; FCoV-II, 1-20aa, MIVLVTCLLLLCSYHTVLST (Fig. 3).

Table 2 Physical and chemical properties of coronavirus spike proteins.

Characteristic	CCoV-I	CCoV-IIa	CCoV-IIb	FCoV-I
Number of amino acids	1480	1473	1453	1464
Molecular weight	166076.85	163099.13	160643.62	164320.18
Theoretical (PI)	4.93	4.95	5.02	5.41
Number of negatively charged residues	150	127	130	129
Number of positively charged residues	104	88	90	96
Instability index (II)	30.94	28.77	29.80	32.67
Aliphatic index	83.44	90.21	91.00	88.31
Grand average of hydropathicity	-0.109	0.006	0.026	-0.043

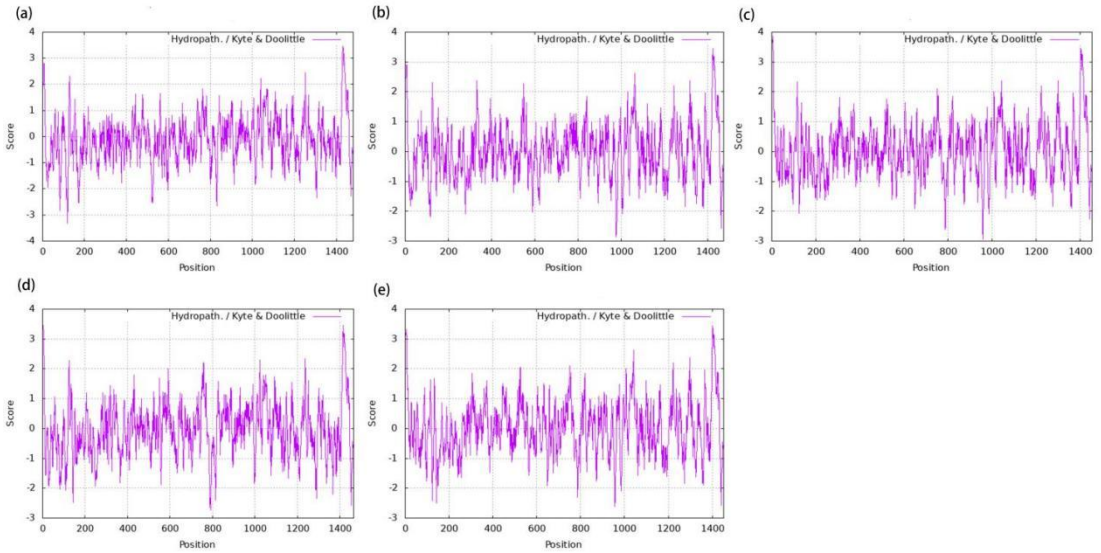

Figure 1 Hydrophilic and hydrophobic of spike proteins. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II. The abscissa represents the amino acid position, and the ordinate represents the amino acid value. >0 is hydrophobic; <0 is hydrophilic.

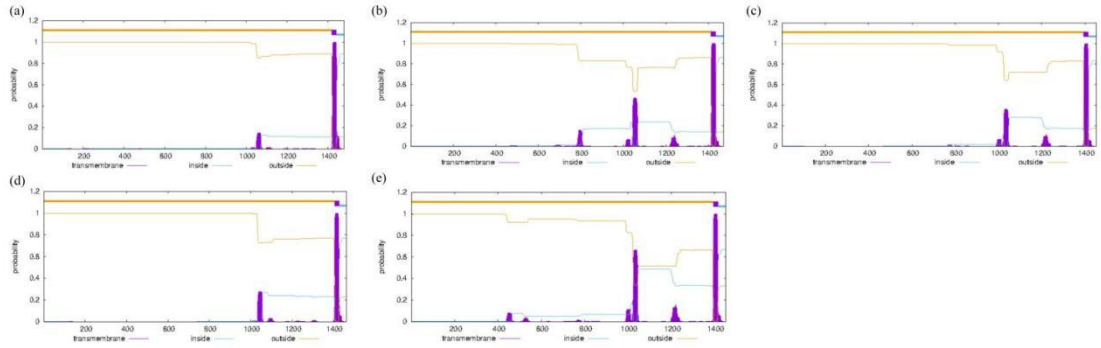

Figure 2 TMHMM was used to predict the transmembrane domain of coronavirus spike protein sequences. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

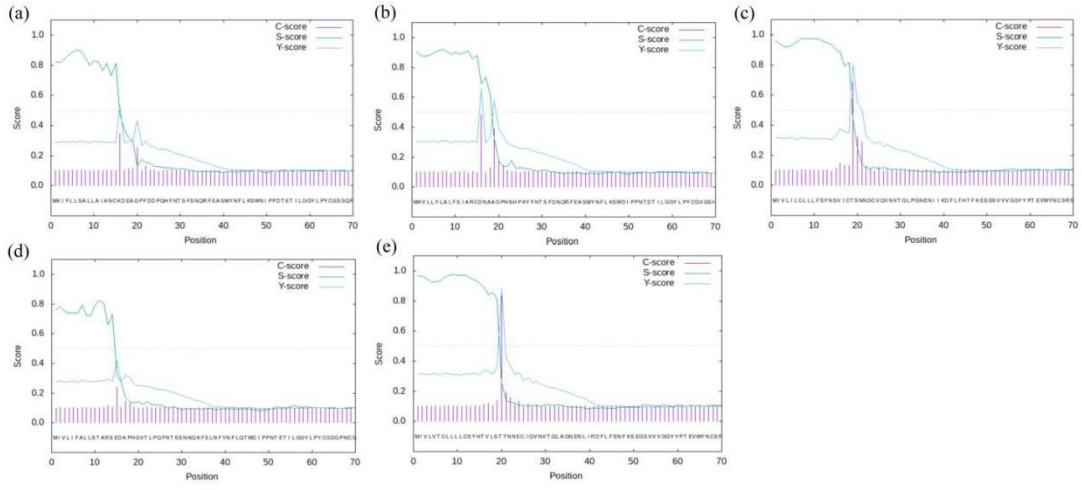

Figure 3 Prediction of coronavirus spike protein signal peptide. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

Phosphorylation Sites of Spike Proteins: Protein phosphorylation occurs mainly at the tyrosine (Tyr), Ser, and threonine (Thr) residues in the peptide chain. NetPhos 3.1 Server online software was used to predict the phosphorylation modification sites of CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I, and FCoV-II. The results showed that the CCoV-I S protein had 67 Ser, 48 Thr, and 24 Tyr modification sites. CCov-IIa contained 71 Ser, 42 Thr, and 31 Tyr modification sites. CCoV-IIb contained 62 Ser, 44 Thr, and 32 Tyr modification sites. FCoV-I contained 71 Ser, 45 Thr, and 29 Tyr-modified cation sites. FCoV-II had 62 Ser, 52 Thr, and 29 Tyr modification sites (Fig. 4).

Glycosylation Sites of S Proteins: Glycosylation can regulate protein functions, including N-linked and O-linked sugar chains. NetNGlyc4.0 Serve software was used to predict the glycosylation modification sites of

the amino acid sequences of the CCoV and FCoV S proteins. The results showed that none of these viral S proteins had O-glycosylation modification sites. CCoV-I contained 27 N-glycosylation sites. Twenty-one and 31 sites were identified on CCoV-IIa and CCoV-IIb, respectively. In addition, FCoV-I and FCoV-II had 24 and 33 sites, respectively (Table 3 and Fig. 5).

Predicted Epitopes of S Proteins: The amino acid epitopes of CCoV and FCoV S proteins were predicted using Predicting Antigenic Peptides online software. There were 63 CCoV-I, 61 CCoV-IIa, 64 CCoV-IIb, 64 FCoV-I, and 63 FCoV-II epitopes, respectively (Fig. 6).

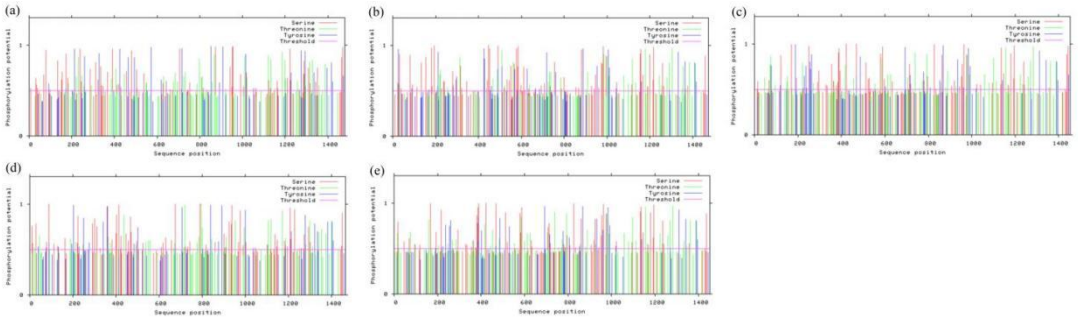

Figure 4 Prediction of coronavirus spike protein phosphorylation sites. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

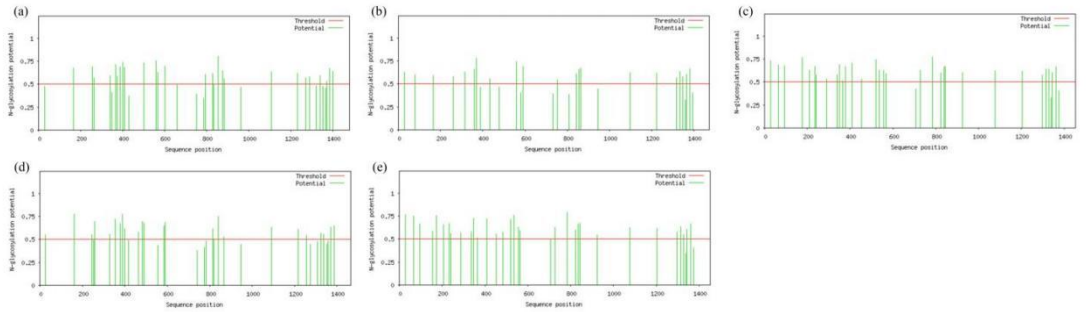

Figure 5 Prediction of coronavirus spike protein glycosylation sites. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

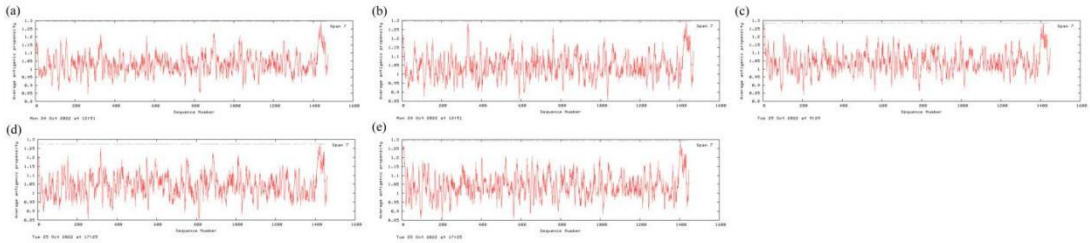

Figure 6 Prediction of coronavirus spike protein epitopes. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

Structure Domain of S Proteins: SMART software was used to predict the amino acid sequence domains of the CCoV and FCoV S proteins. The results showed that CCoV-I contained two domains: corona_S1 (283aa to 803aa) and corona_S2 (835aa to 1479aa). Corona _S1 was the spike_rec_binding domain, and corona_S2 was the highly conserved functional domain. CCoV-IIa contained two domains, corona_S1 (284aa to 803aa) and corona_S2 (821aa to 1472aa). CCoV-IIb contained two domains: corona_S1 (259aa to 779aa) and corona_S2 (802aa to 1452aa). FCoV-I has two domains: corona S1 (272aa–790aa) and corona_S2 (820aa–1463aa). FCoV-II has two domains: corona S1 (258aa–778aa) and corona_S2 (801aa–1451aa) (Fig. 7).

Prediction of Functional Motif of Coronavirus Spike Protein: Motifs reflect various biological functions of proteins. According to PROSITE database analysis, the functional motifs of the CCoV-I S protein included the COV_S2_HR1 region profile of 1070-1189 amino acids and the COV_S2_HR2 region profile of 1334-1433 amino acids. The functional motifs of the CCoV-IIa S protein included the COV_S2_HR1 region profile of 1061-1180 amino acids and the COV_S2_HR2 region profile of 1329-1426 amino acids. The functional motif of the CCoV-IIb S protein included the COV_S2_HR1 region profile of 1041-1160 amino acids and the COV_S2_HR2 region profile of 1309-1406 amino acids. The functional motifs of the FCoV-I S protein included the COV_S2_HR1 region profile of 1054-1173 amino acids and the COV_S2_HR2 region profile of 1318-1417 amino acids. The functional motifs of the FCoV-II S protein included the COV_S2_HR1 region profile of 1040-1159 amino acids and the COV_S2_HR2 region profile of 1308-1405 amino acids (Fig. 8).

Table 3 N-glycosylation sites of coronavirus spike protein sequences.

CCoV-I		CCoV-IIa		CCoV-IIb		FCoV-I		FCoV-II	
Position	**Seq**	**Position**	**Seq**	**Position**	**Seq**	**Position**	**Seq**	**Position**	**Seq**
167-170	NDTK	29-32	NTSF	28-31	NVTQ	27-30	NTSS	29-32	NVTQ
255-258	NRTT	80-83	NKSS	66-69	NCSR	161-164	NTSS	67-70	NCSR
263-266	NNTG	165-168	NNTV	94-97	NSTG	244-247	NRTT	95-98	NSTG
340-343	NNSA	261-264	NNTA	175-178	NGTK	258-261	NYTH	155-158	NSTC
366-369	NFSG	314-317	NKSS	209-212	NVTI	329-332	NNSA	174-177	NGTK
375-378	NATA	360-363	NDSV	235-238	NFTY	355-358	NFSD	208-211	NVTL
389-392	NSTA	371-374	NFTE	242-245	NNTN	378-381	NTTA	234-237	NFTY
401-404	NVTD	433-436	NNTA	289-292	NSST	390-393	NVTD	241-244	NNTN
410-413	NTTS	558-561	NVSV	338-341	NNTV	398-401	NSSV	288-291	NSST
502-505	NGTS	590-593	NETE	349-352	NFTT	464-467	NFSI	337-340	NNTV
558-561	NVTL	753-756	NVST	366-369	NTTG	482-485	NYTD	348-351	NFTA
566-569	NDTE	842-845	NITH	379-382	NDTV	491-494	NGTG	365-368	NTTG
602-605	NVTN	857-860	NVTI	409-412	NGTA	585-588	NGSV	408-411	NGTA
795-798	NHTH	863-866	NFTI	453-456	NLTT	590-593	NVTS	452-455	NLTT
829-832	NDTS	1098-1101	NITQ	520-523	NKSV	814-817	NDTS	483-486	NVTY
834-837	NCTS	1224-1227	NGTH	536-539	NITI	841-844	NVTH	519-522	NKSV
856-859	NVTK	1318-1321	NATV	558-561	NITL	867-870	NFTV	535-538	NITI
877-880	NITI	1335-1338	NQTV	566-569	NNTD	1091-1094	NITL	557-560	NITL
883-886	NFTV	1348-1351	NWTV	729-732	NVSD	1217-1220	NGTH	565-568	NNTD
1107-1110	NITL	1365-1368	NLTG	784-787	NYTN	1256-1259	NHTY	707-710	NSTL
1233-1236	NGTH	1382-1385	NTTV	823-826	NVTH	1324-1327	NKTI	728-731	NVSD
1272-1275	NDTY			838-841	NVTI	1337-1340	NYTT	783-786	NYTS
1290-1293	NGTY			844-847	NFTI	1370-1373	NLTT	822-825	NVTH
1340-1343	NKTI			925-928	NSTE	1386-1389	NKTL	837-840	NVTI
1372-1375	NLTA			1078-1081	NITQ			843-846	NFTI
1386-1389	NLTV			1204-1207	NGTH			924-927	NSTE
1402-1405	NKTL			1298-1301	NATV			1077-1080	NITQ
				1315-1318	NQTV			1203-1206	NGTH
				1328-1331	NWTV			1297-1300	NATV
				1345-1348	NLTG			1314-1317	NQTV
				1362-1365	NTTV			1327-1330	NWTV
								1344-1347	NLTG
								1361-1364	NTTV

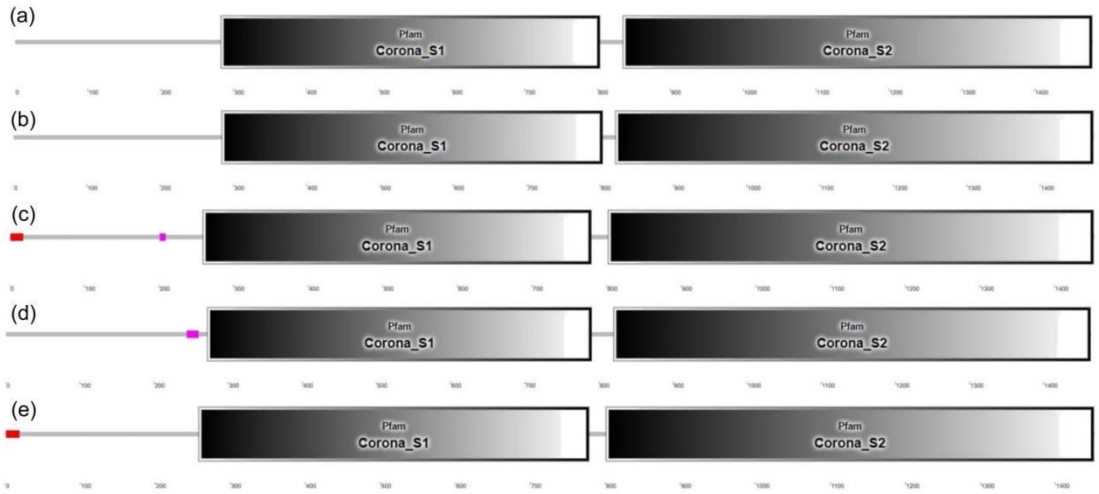

Figure 7 Structure domain of coronavirus spike proteins. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

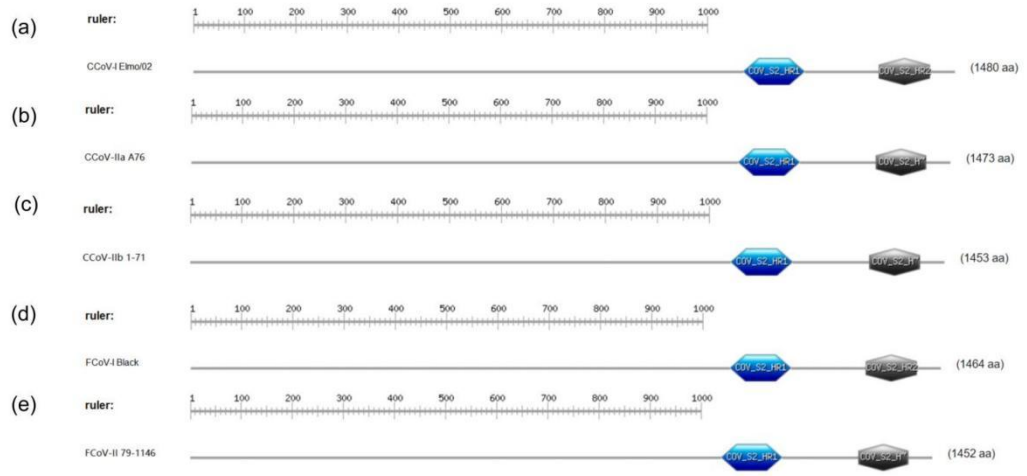

Figure 8 Prediction of coronavirus spike protein functional motif. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II.

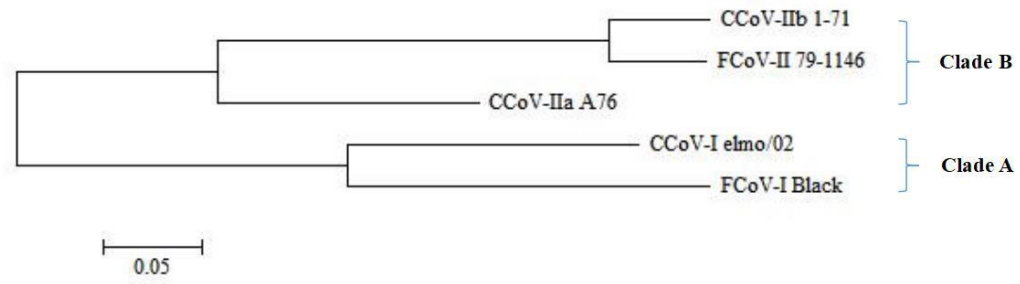

Figure 9 Phylogenetic analysis. The phylogenetic tree was generated using the neighbor-joining method and supported by 1,000 bootstraps.

Identity and Phylogenetic Analyses: S-gene nucleotide multi-sequence alignment analysis identified 60.8%, 56.5%, 73.2%, and 55.6%, as CCoV-I and CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II, respectively. In addition, the nucleotide identities of CCoV-IIb and FCoV-II were the highest (90.6%). The identity between FCoV-I and CCoV-I was higher than that between FCoV-II and CCoV-I (73.3% vs. 55.0%). The identity between CCoV-IIa and CCoV-IIb was 71.4%, which was lower than that between FCoV-II (90.6%). Phylogenetic analysis of the S genes showed that CCoV-I and FCoV-I were included in the same clade, while CCoV-IIb and FCoV-II formed another clade. Our results also indicated that the CCoV-IIa A76 S gene was in an intermediate position between the two serotypes (Fig. 9).

Discussion

All *Alphacoronavirus* are similar in genomic compositions and protein structures, including four structural proteins, i.e., N, S, E, and M, located at the 3' end (Lin *et al.* 2022). The S protein is divided into S1 and S2 subunits. The receptor-binding domain (RBD) in the S1 subunit can bind to receptors on the host cell membrane, and the S2 subunit enables the virus to carry out membrane fusion with the host cell (Licitra *et al.* 2014). The S protein can also induce a host immune response and is a key protein in vaccine development (Patel *et al.* 2022). Additionally, changes in the structure of the S protein directly affect the virulence of the virus (Guo *et al.* 2019,Vennema *et al.* 1998). Therefore, bioinformatics predictive analysis of coronavirus S protein can lay a foundation for further study of proteins. The comparative analysis of several proteins can also be helpful in exploring the fundamental differences between different viruses of the same genus or different genotypes of the same virus. The overall goal is to help clarify the causes of different infection mechanisms and host ranges.

The structure of a protein determines its biological functions. Changes in the structure of proteins or chemical modifications can affect their properties and functions. Bioinformatics software was used to predict the physicochemical properties, structure, and antigenic epitopes of the S proteins, which helped us to understand the mechanism of action of these proteins. Physicochemical property prediction showed that the PI of FCoV-I S protein was the highest (5.41), and that of CCoV-I protein was the lowest (4.93). The coronavirus S proteins selected in this study were thus identified as acidic proteins with better solubility than alkaline proteins (Scheller *et al.* 2020). Moreover, hydrophilic analysis using ExPASy-ProtScale software revealed that the surface of the S proteins of these viruses was rich in hydrophilic amino acids. Our observations further suggested that these proteins were soluble.

TMHMM is a program used to predict transmembrane helices based on the Markov model, combining the hydrophobicity of the transmembrane region, length of the helices, charge bias, and topological constraints of the membrane proteins. TMHMM Server v.2.0, online software was used to predict the transmembrane area of the CCoV and FCoV coronavirus S proteins. The predicted results indicated that the S proteins of these viruses existed outside the viral capsule, some existed in the viral capsule, and some resided inside the viral capsule. These parts of the amino acid sequence outside the viral capsule may contain an RBD binds to host cell receptors. Some coronavirus RBDs have been screened and verified, and these RBDs are located in the S1 subunit of the S protein (Ma *et al.* 2014,Tai *et al.* 2020,Tai *et al.* 2016). According to the SMART and PROSITE databases, the S proteins of the viruses selected in this study have two functional domains, spike_rec_binding and corona_S2. spike_rec_binding functional domain and corona_S2 functional domain. For most coronaviruses, one characteristic of the S protein is that it has a proteolytic site that distinguishes the S1 RBD from the S2 fusion domain (Dong *et al.* 2022,Follis *et al.* 2006,Whittaker *et al.* 2018). Compared to the published S1 subunit at the S protein position, the CCoV and FCoV viruses predicted in this study all contained their own S1 subunit in the extracapsular region. In addition, the S1 and S2 subunits were not significantly different from the predicted results of SMART and PROSITE. As each software uses different software and parameters, the predicted results may be different.

Signal peptides mainly exist in secreted proteins, transmembrane proteins, and eukaryotic organelles and promote the secretion of extracellular proteins (Bos *et al.* 2020). The prediction results of SignalP4.0 showed that the signal peptides of the CCoV and FCoV coronavirus S proteins in this study were all located in the top 20 amino acids of the N-terminal. This was the first hydrophobic peptide synthesized during mRNA translation, which could be ignored when choosing to express the S protein in vitro. Phosphorylation can regulate protein activity and enhance protein interaction (Lu *et al.* 2021). In this study, we predicted that the S proteins of the CCoV and FCoV viruses would be phosphorylated. Specific glycosylation has an important effect on the physicochemical properties and physiological functions of proteins (Watanabe *et al.* 2020). Glycosylation is the process of a non-enzymatic reaction with glucose after protein translation to produce glucose, glucose chains, and protein sites through covalent bonds to form a glycosylation site (Eichler 2019). Glycosylation can alter the conformation of peptides in the S protein, rendering amino acids disabled and epitopes unrecognized by cells and antibodies, thus enabling the coronavirus to escape innate and adaptive immune responses (Walls *et al.* 2019). The predicted results of this study revealed that none of the CCoV or FCoV viruses contained O-glycosylation modification sites, and all of them contained more N-glycosylation modification sites.

Antigenic epitopes are chemical groups that stimulate the production of antibodies or sensitized lymphocytes and can aid in the development of diagnostic methods and novel vaccines as well as the design of antiviral immunization strategies (Chernyavtseva *et al.* 2019,Gong *et al.* 2019,Li *et al.* 2017). They are crucial in determining the structure and function of antigen molecules and the mechanism of antigen-antibody reactions and can be used in the research and development of polypeptide vaccines and novel anti-coronavirus drugs. The software-predicted epitope is suitable for software analysis of known proteins or peptides to predict potential epitope regions. The basis of the software prediction method is that epitopes are related to the amino acid sequences or higher structural features of proteins. In this study, online antigen-peptide prediction software based on the Kolas-kar and Tongaonkar methods was used to predict the epitopes of CCoV and FCoV coronavirus S proteins. The results revealed 63 epitopes for CCoV-I, 61 epitopes for CCoV-IIa, 64 epitopes for CCoV-IIb, 64 epitopes for FCoV-I, and 63 epitopes for FCoV-II. Some of the possible antigenic epitopes predicted by software have been validated by synthesizing peptides containing these epitopes, reacting with antibodies in vitro, and measuring the immunogenicity of these peptides in vivo (Pan *et al.* 2017,Sun *et al.* 2022,Zheng *et al.* 2021). Therefore, the accuracy of the epitopes

predicted in this study needs to be verified by subsequent in vitro experiments.

According to identity and phylogenetic analyses, we found a higher identity between CCoV-IIb and FCoV-II S proteins. FCoV-II viruses may be generated by dual recombination of FCoV-I and CCoV (Jaimes and Whittaker 2018). Although there is some debate as to where the recombination event occurs in the viral genome, similar S proteins provide further evidence supporting this hypothesis. In addition, we found that the identity between FCoV-I and CCoV-I was significantly higher than that between FCoV-I and FCoV-II S proteins or CCoV-I and CCoV-II S proteins. This is also consistent with the proposed novel classification of two clades within the Alphacoronavirus 1 species: clade A viruses encompassing FCoV-I and CCoV-I, and clade B viruses including CCoV-IIa, CCoV-IIb, FCoV-II, and TGEV viruses (Whittaker *et al.* 2018).

Recombination plays a crucial role in the genetic evolution of coronaviruses, and novel variants arising from recombination can pose a potential epidemic threat (Wesley 1999). CCoV-II is a novel recombinant variant derived from an unknown coronavirus, and CCoV-I is closely related to TGEV and FCoV-II (Duijvestijn *et al.* 2016). This hypothesis was indirectly validated in the present study, where phylogenetic analysis showed that CCoV-II, TGEV, and FCoV-II strains were in closer developmental clusters. In addition, CCoV-I and FCoV-I were also in the same developmental cluster, which was consistent with previous reports that they both originated from a common ancestor (Jaimes *et al.* 2020). Moreover, it has been reported that CCoV-IIa viruses have an NTD consistent with the prototype CCoV, but CCoV-IIb is genetically distinct from CCoV-IIa, with the CCoV-IIb spike gene having a TGEV-like NTD (Regan *et al.* 2012). The phylogenetic analysis of S proteins in this study indicated that CCoV-IIb was located in an intermediate position between clades A and B, which may also support this concept.

For nearly 20 years, the coronaviruses circulating in some animals have posed a threat to humans (Azhar *et al.* 2014,Bolles *et al.* 2011). Recently, a novel canine-feline recombinant alphacoronavirus was isolated from a human patient with pneumonia (Vlasova *et al.* 2022). Therefore, the genetic characteristics of coronaviruses currently circulating in companion animals must be identified. The standards and parameters used by different software programs are often different, and the predicted results may also differ. To obtain reliable prediction results, multiple types of software must be used simultaneously for the analysis. In this study, the biological functions and structures of the CCoV and FCoV S proteins were analyzed, providing a reference for the subsequent study of these coronaviruses and their potential functions (Fig 10 and 11). Furthermore, research on the biological and immune properties of the S protein is conducive to the application of the S protein as a target in vaccine design.

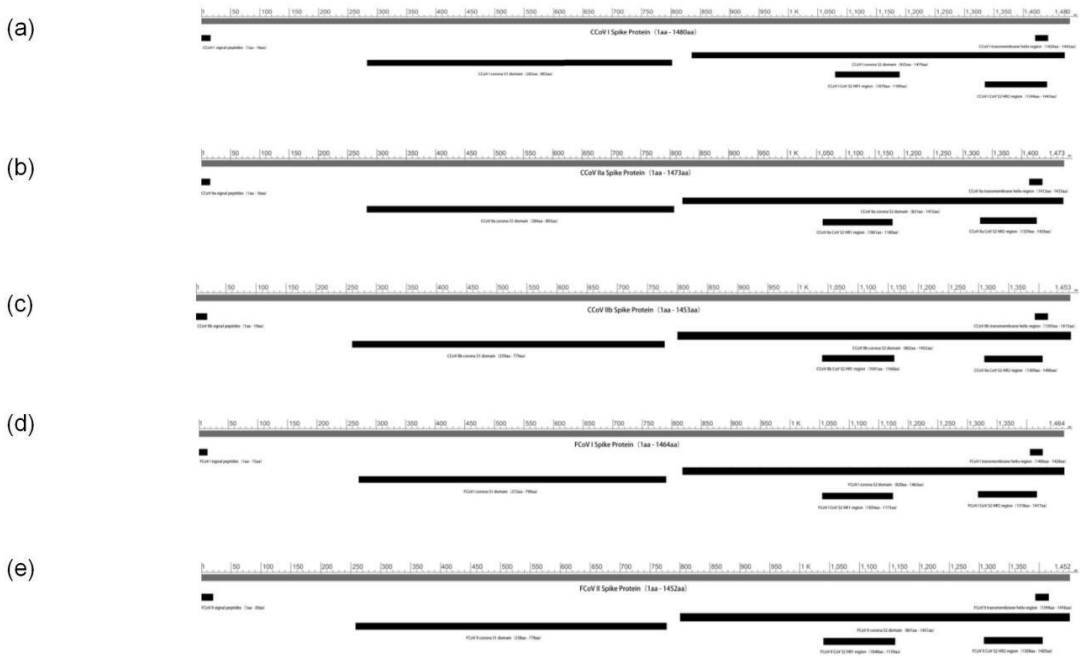

Figure 10 Structure of CCoV and FCoV S proteins. Both CCoV and FCoV S proteins contain a signal peptide, a transmembrane domain, two domains (S1 and S2), and functional motifs located in S2. (a-e): CCoV-I, CCoV-IIa, CCoV-IIb, FCoV-I and FCoV-II

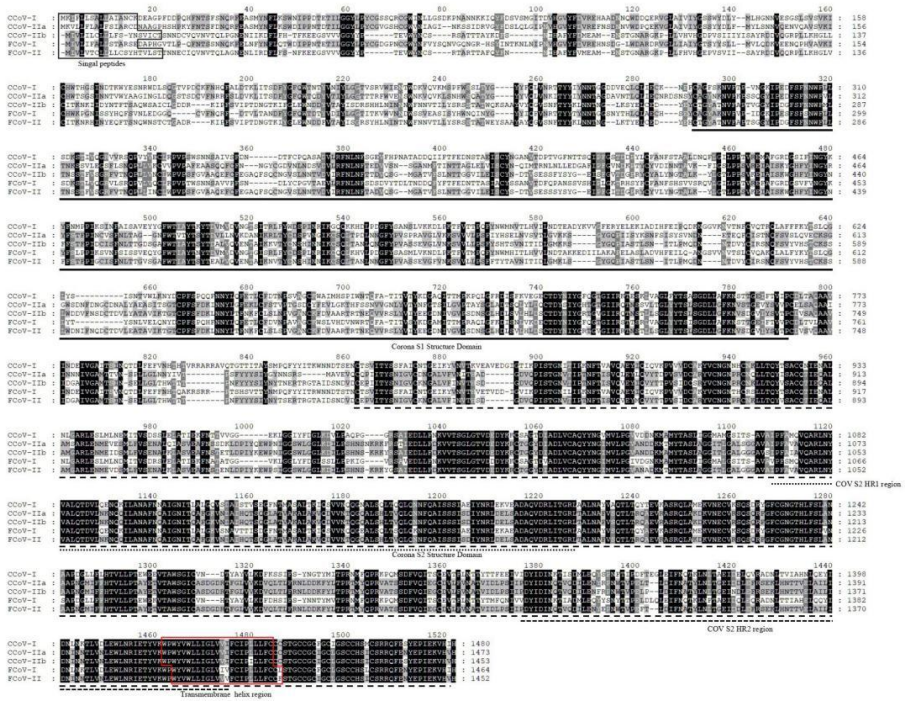

Figure 11 Amino acid sequence alignment of the S proteins from CCoV-I, CCoV-II, FeCoV-I, and FeCoV-II

Acknowledgments

This study was supported by Fujian Natural Science Foundation general project (2022J011155 and 2023J01991). Project support was provided by the Research Foundation of Engineering Research Center for the Prevention and Control of Animal Original Zoonosis, Fujian Province University (Grant No.2022K009 and 2023K02).

References

Adams MJ, Lefkowitz EJ, King AM, Harrach B, Harrison RL, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Mushegian AR, Nibert M, Sabanadzovic S, Sanfaçon H, Siddell SG, Simmonds P, Varsani A, Zerbini FM, Gorbalenya AE and Davison AJ 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. Arch Virol. 161: 2921-2949.

Alfano F, Fusco G, Mari V, Occhiogrosso L, Miletti G, Brunetti R, Galiero G, Desario C, Cirilli M and Decaro N 2020. Circulation of pantropic canine coronavirus in autochthonous and imported dogs, Italy. Transbound Emerg Dis. 67: 1991-1999.

Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM and Madani TA 2014. Evidence for camel-to-human transmission of MERS coronavirus. N Engl J Med. 370: 2499-2505.

Bolles M, Donaldson E and Baric R 2011. SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. Curr Opin Virol. 1: 624-634.

Bos R, Rutten L, van der Lubbe JEM, Bakkers MJG, Hardenberg G, Wegmann F, Zuijdgeest D, de Wilde AH, Koornneef A, Verwilligen A, van Manen D, Kwaks T, Vogels R, Dalebout TJ, Myeni SK, Kikkert M, Snijder EJ, Li Z, Barouch DH, Vellinga J, Langedijk JPM, Zahn RC, Custers J and Schuitemaker H 2020. Ad26 vector-based COVID-19 vaccine encoding a prefusion-stabilized SARS-CoV-2 Spike immunogen induces potent humoral and cellular immune responses. NPJ Vaccines 5. 5: 91.

Chernyavtseva A, Cave NJ, Munday JS and Dunowska M 2019. Differential recognition of peptides within feline coronavirus polyprotein 1 ab by sera from healthy cats and cats with feline infectious peritonitis. Virology. 532: 88-96.

Decaro N, Martella V, Elia G, Campolo M, Desario C, Cirone F, Tempesta M and Buonavoglia C 2007. Molecular characterisation of the virulent canine coronavirus CB/05 strain. Virus Res. 125: 54-60.

Dong B, Zhang X, Bai J, Zhang G, Li C and Lin W 2022. Epidemiological investigation of canine coronavirus infection in Chinese domestic dogs: A systematic review and data synthesis. Prev Vet Med. 209: 105792.

Duijvestijn M, Mughini-Gras L, Schuurman N, Schijf W, Wagenaar JA and Egberink H 2016. Enteropathogen infections in canine puppies: (Co-)occurrence, clinical relevance and risk factors. Vet Microbiol. 195: 115-122.

Eichler J 2019. Protein glycosylation. Curr Biol. 29: R229-R231.

Evermann JF, Abbott JR and Han S 2005. Canine coronavirus-associated puppy mortality without evidence of concurrent canine parvovirus infection. J Vet Diagn Invest. 17: 610-614.

Follis KE, York J and Nunberg JH 2006. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. Virology. 350: 358-369.

Gao FS, Hu GX, Xia XZ, Gao YW, Bai YD and Zou XH 2009. Isolation and identification of a canine coronavirus strain from giant pandas (*Ailuropoda melanoleuca*). J Vet Sci. 10: 261-263.

Gong L, Gong Y, Lin Y, Qin J, Liu Y, Qian X, Xue C and Cao Y 2019. Identification and characterization of a

novel linear epitope in the spike protein of the porcine epidemic diarrhea virus. Acta Virol. 63: 88-95.

Guo J, Fang L, Ye X, Chen J, Xu S, Zhu X, Miao Y, Wang D and Xiao S 2019. Evolutionary and genotypic analyses of global porcine epidemic diarrhea virus strains. Transbound Emerg Dis. 66: 111-118.

Herrewegh AA, Smeenk I, Horzinek, MC, Rottier PJ and de Groot RJ 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. J Virol. 72: 4508-4514.

Jaimes JA, Millet JK, Stout AE, André NM and Whittaker GR 2020. A Tale of Two Viruses: The Distinct Spike Glycoproteins of Feline Coronaviruses. Viruses. 12: 83.

Jaimes JA, Whittaker GR 2018. Feline coronavirus: Insights into viral pathogenesis based on the spike protein structure and function. Virology. 517: 108-121.

Keenan KP, Jervis HR, Marchwicki RH and Binn LN 1976. Intestinal infection of neonatal dogs with canine coronavirus 1-71: studies by virologic, histologic, histochemical, and immunofluorescent techniques. Am J Vet Res. 37: 247-256.

Le Poder S 2011. Feline and canine coronaviruses: common genetic and pathobiological features. Adv Virol. 609465.

Li C, Li W, Lucio de Esesarte E, Guo H, van den Elzen P, Aarts E, van den Born E, Rottier PJM and Bosch BJ 2017. Cell Attachment Domains of the Porcine Epidemic Diarrhea Virus Spike Protein Are Key Targets of Neutralizing Antibodies. J Virol. 91: 00273-00217.

Licitra BN, Duhamel GE and Whittaker GR 2014. Canine enteric coronaviruses: emerging viral pathogens with distinct recombinant spike proteins. Viruses. 6: 3363-3376.

Lin F, Zhang H, Li L, Yang Y, Zou X, Chen J and Tang X 2022. PEDV: Insights and Advances into Types, Function, Structure, and Receptor Recognition. Viruses. 14: 1744.

Lorusso A, Decaro N, Schellen P, Rottier PJ, Buonavoglia C, Haijema BJ and de Groot RJ 2008. Gain, preservation, and loss of a group 1a coronavirus accessory glycoprotein. J Virol. 82: 10312-10317.

Lu S, Ye Q, Singh D, Cao Y, Diedrich JK, Yates JR 3rd, Villa E, Cleveland DW and Corbett KD 2021. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. Nat Commun. 12: 502.

Ma C, Wang L, Tao X, Zhang N, Yang Y, Tseng CK, Li F, Zhou Y, Jiang S and Du L 2014. Searching for an ideal vaccine candidate among different MERS coronavirus receptor-binding fragments--the importance of immunofocusing in subunit vaccine design. Vaccine. 32: 6170-6176.

Pan, W, Chen DS, Lu YJ, Sun FF, Xu HW, Zhang YW, Yan C, Fu LL, Zheng KY and Tang RX 2017. Bioinformatic prediction of the epitopes of *Echinococcus granulosus* antigen 5. Biomed Rep. 6: 181-187.

Patel R, Kaki M, Potluri VS, Kahar P and Khanna D 2022. A comprehensive review of SARS-CoV-2 vaccines: Pfizer, Moderna & Johnson & Johnson. Hum Vaccin Immunother. 18: 2002083.

Pratelli A 2008. Canine coronavirus inactivation with physical and chemical agents. Vet J. 177: 71-79.

Pratelli A, Martella V, Pistello M, Elia G, Decaro N, Buonavoglia D, Camero M, Tempesta M and Buonavoglia C 2003. Identification of coronaviruses in dogs that segregate separately from the canine coronavirus genotype. J Virol Methods. 107: 213-222.

Pratelli A, Tempesta M, Roperto FP, Sagazio P, Carmichael L and Buonavoglia C 1999. Fatal coronavirus infection in puppies following canine parvovirus 2b infection. J Vet Diagn Invest. 11: 550-553.

Regan AD, Millet JK, Tse LP, Chillag Z, Rinaldi VD, Licitra BN, Dubovi EJ, Town CD and Whittaker GR 2012. Characterization of a recombinant canine coronavirus with a distinct receptor-binding (S1) domain. Virology. 430: 90-99.

Scheller C, Krebs F, Minkner R, Astner I, Gil-Moles M and Wätzig H 2020. Physicochemical properties of SARS-CoV-2 for drug targeting, virus inactivation and attenuation, vaccine formulation and quality control. Electrophoresis. 41: 1137-1151.

Sun Q, Huang Z, Yang S, Li Y, Ma Y, Yang F, Zhang Y and Xu F 2022. Bioinformatics-based SARS-CoV-2 epitopes design and the impact of spike protein mutants on epitope humoral immunities. Immunobiology. 227: 152287.

Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y and Du L 2020. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol. 17: 613-620.

Tai W, Wang Y, Fett CA, Zhao G, Li F, Perlman S, Jiang S, Zhou Y and Du L 2016. Recombinant Receptor-Binding Domains of Multiple Middle East Respiratory Syndrome Coronaviruses (MERS-CoVs) Induce Cross-Neutralizing Antibodies against Divergent Human and Camel MERS-CoVs and Antibody Escape Mutants. J Virol. 91: 01651-01616.

Tresnan DB, Levis R and Holmes KV 1996. Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. J Virol. 70: 8669-8674.

Vennema H, Poland A, Foley J and Pedersen NC 1998. Feline infectious peritonitis viruses arise by mutation from endemic feline enteric coronaviruses. Virology. 243: 150-157.

Vlasova AN, Diaz A, Damtie D, Xiu L, Toh TH, Lee JS, Saif LJ and Gray GC 2022. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. Clin Infect Dis. 74: 446-454.

Walls AC, Xiong X, Park YJ, Tortorici MA, Snijder J, Quispe J, Cameroni E, Gopal R, Dai M, Lanzavecchia A, Zambon M, Rey FA, Corti D, Veesler D 2019. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. Cell. 176: 1026-1039.

Watanabe Y, Allen JD, Wrapp D, McLellan JS and Crispin M 2020. Site-specific analysis of the SARS-CoV-2 glycan shield. bioRxiv. 2020.03.26.010322. Preprint.

Wentworth DE, Holmes KV 2001. Molecular determinants of species specificity in the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. J Virol. 75: 9741-9752.

Wesley RD 1999. The S gene of canine coronavirus, strain UCD-1, is more closely related to the S gene of transmissible gastroenteritis virus than to that of feline infectious peritonitis virus. Virus Res. 61: 145-152.

Whittaker GR, André NM and Millet JK 2018. Improving Virus Taxonomy by Recontextualizing Sequence-Based Classification with Biologically Relevant Data: the Case of the Alphacoronavirus 1 Species. mSphere. 3: 00463-00417.

Zheng J, Xia Z, Xu Y, Ou Z, Lin X, Jin S, Liu Y and Wu J 2021. Identification of novel hepatitis B virus therapeutic vaccine candidates derived from polymerase protein. Aging. 13: 14372-14384.