Acta virologica **53**: 111 – 120, 2009 doi:10:4149/av_2009_02_111

Analysis of putative recombination hot sites in the S gene of canine coronaviruses

Y.Y. WANG[1,3], C.P. LU[1,2*]

[1]College of Agriculture and Biology, Shanghai Jiaotong University, 201101 Shanghai, P.R. China; [2]College of Veterinary Medicine, Nanjing Agriculture University, 210095 Nanjing, P.R. China; [3]Department of Medical Microbiology and Parasitology, College of Medicine, Fudan University, 200032 Shanghai, P.R. China

Received September 21, 2007; accepted May 13, 2009

Summary. – The S gene sequence of Canine coronavirus strain 1-71 (CCoV 1-71) was cloned, sequenced, and compared to those of other CCoVs, Transmissible gastroenteritis virus (TGEV), and Feline coronavirus (FCoV). The sequence analysis showed that CCoV 1-71 displayed a 98.8–99.8% identity with CCoVs strains V1, K378, and GP. Four putative recombination sites were found at the 5'-end of the S gene, namely at nt 53, 75, 425, 991. Both sequences flanking each site were significantly different. Three recombination hot regions were found on the S gene, namely at nt 337–437, 1545–3405, and 4203–4356, which shared a common recombination signal with Group 2 coronaviruses. The G/CTAAAAA/GT sequence downstream of the recombination site may represent a specific recombination signal in CCoVs. The CCoV 1-71 S protein sequence was found to be similar to those of other CCoVs except for several N-glycosylation sites at the N-terminus of the S protein, which could be related to the differences in virulence and cell tropism in individual CCoVs. This study indicated that the similarity of CCoVs in virulence and tropism was mostly acquired by the homologous RNA recombination and not only by simple mutation and selection.

Keywords: canine coronaviruses; S gene; recombination; hot sites

Introduction

Canine coronavirus (CCoV) was first identified in 1971 associated with moderate to severe enteritis in young pups (Binn *et al*., 1974). CCoV is a member of the genus *Coronavirus,* the family *Coronaviridae.* Coronaviruses are divided into 3 groups called 1, 2, and 3 according to the specific binding between the virus and a specific cellular receptor, what is required for the establishment of viral infection. CCoV belongs to the Group 1 along with TGEV, PEDV, FCoV, and Human coronavirus 229E (HCoV 229E) (Cavanagh, 1997). Today, CCoV appears to be enzootic in dogs worldwide and is able to infect also cats, pigs, pandas, foxes, and wolves (Herrewegh *et al.*, 1998; Wesley 1999; Zarnke *et al*., 2001; Wang *et al.*, 2006). FCoVs can be distinguished into two serotypes I and II, on the basis of a virus neutralization assay *in vitro* using both type-specific polyclonal sera and monoclonal antibodies directed against the S protein (Herrewegh *et al.*, 1998). Differences in the S gene of FCoVs type I and type II may also account for the different properties observed *in vitro.* Indeed, FCoVs type I grow poorly in tissue culture cells, while type II strains grow well. Typical CCoV specific sera neutralize type II FCoV, but fail to neutralize or neutralize poorly viruses of type I. Recently, new variants of CCoV have been discovered in the feces of infected dogs indicating that this virus has evolved (Wesley, 1999; Naylor *et al*., 2001; Pratelli *et al*., 2002). The variant Elmo/02 has phenotypic properties different from CCoVs, but similar to the type I FCoVs. On the basis of the significant genetic similarity to FCoV, CCoVs were classified into two genotypes. The

*Corresponding author. E-mail: lucp@njau.edu.cn; fax: +86-025-84396517.

Abbreviations: CCoV = Canine coronavirus; FCoV = Feline coronavirus; HCoV 229E = Human coronavirus 229E; IBV = Infectious bronchitis virus; MHV = Murine hepatitis virus, PEDV = Porcine epidemic diarrhea virus, RDP = recombination detection program, TGEV = Transmissible gastroenteritis virus

CCoV strains similar to the type I FCoV were designated as type I CCoVs. Typical CCoV strains were designated as type II CCoVs and are similar to the type II FCoV. Unlike the Group 1 coronaviruses, type I CCoVs have a potential cleavage site in the S protein that is shared by the members of Groups 2 and 3 (Pratelli *et al.*, 2003).

The coronaviruses have a high frequency of RNA recombination in both tissue culture and natural infection (Liao *et al.*, 1992). During replication of the coronavirus, RNA polymerase complex may jump from its original template to another template, what facilitates RNA homologous recombination. RNA recombinants of the coronaviruses are common and occur randomly at almost every potential recombination site in the RNA genome (Banner *et al.*, 1991). However, some regions of the genome have a high ratio for RNA recombination and these recombinant "hot-sites" contain characteristic nucleotide sequences such as TCTAA in most coronaviruses and ATGATG and CTTTATAAA in MHV (Liao *et al.*, 1992). Research by Wu *et al.* (2003) indicated that there were common replication-signaling elements in the coronaviruses of the Group 2, and pointed at a high potential for recombination among those viruses. Wang *et al*. (1993) discovered that IBV strain PP14 had vaccine Ark99 and Mass41 crossover-position sites in the S1 gene and the sequences CTT(A/T)(A/T)G could be found at every junction where the homology shifts were found. Lee and Jackwood (2000) demonstrated that IBV strain DE072 was a recombinant virus that contained IBV strain D1466-like sequence in the S gene and an intergenic consensus sequence (CTGAACAA or CTTAACAA) that may serve as a template-switching site for virus encoded polymerase. FCoV type II, strains 79-1683 and 79-1146 originated from a double recombination event occurring between FCoV type I and CCoV (Herrewegh *et al*., 1998). In our previous study, we found the recombination sites in M gene of some CCoVs. In the CCoV strain TN449, the first 50 amino acid residues at the N-terminus of the M protein shared 100% identity with the FCoV 79-1683 and 81.6% with TGEV. However, the downstream amino acid sequences shared a higher identity (96.7%) with TGEV than with FCoV (90.2%). These results indicated that a recombination event had occurred. These same recombination sites were also present in the CCoV wild isolate NJ17 demonstrating that a recombination event preceding those occurring in the CCoV strain TN449 had taken place in this isolate (Wang *et al*., 2005).

Many coronavirus recombinant events occurred in the S gene especially in the 5'-end region (Wesley, 1999). The 5'-end variation of the coronavirus S gene may in fact be responsible for the variation in coronavirus antigenicity, virulence, recognition, and host adaptation (Liao *et al.*, 1992). Natural recombination events are likely to be common in wild viruses resulting in emergence of new variants with improved host adaptability and virulence. CCoV strain UCD-1 and TGEV share a high identity and have a similar host cell preference (Wesley, 1999) suggesting that a recombinant event might have happened.

In this study, we determined the sequence of CCoV 1-71 S gene and corresponding protein and compared them to those of other CCoVs in Group 1. In addition, we determined respective identities in comparison with the CCoV 1-71 S gene and found a putative recombination signal, hot regions, and hot sites. The obtained results indicated that a homologous recombination between the CCoVs has occurred in the S gene.

Materials and Methods

Viruses and cells. A72 cells and CCoV strain 1-71 were kindly provided by Prof. Bayer, Giessen University, Germany. The A72 cells were propagated in MEM with 10% FCS. Confluent A72 cells maintained in MEM with 2% FCS were inoculated with the CCoV 1-71 and harvested when 90% of the cells showed a cytopathic effect. Other CCoV isolate NJ17 was isolated from the feces of CCoV-infected dog.

Total RNA extraction. Total RNA was extracted from the cell cultures using Trizol reagent (MDBio) according to the manufacturer's instructions. Briefly, 250 µl of CCoV-infected cells were added to 750 µl Trizol and vortexed for 60 secs. Then, 200 µl of chloroform was added, vortexed for 15 secs, and incubated for 5 min at room temperature. The sample was centrifuged at 12,000 x g for 15 mins and the aqueous phase was transferred into a new tube, combined with 500 µl of isopropanol, incubated for 15 mins at room temperature, and centrifuged at the maximum speed for 10 mins at 4°C. The pellet was then washed twice with 500 µl of 70% ethanol and centrifuged at the maximum speed for 2 mins at 4°C. The pellet was air-dried for 10 mins and then dissolved in 30 µl distilled H_2O.

cDNA synthesis. The reverse transcription of cDNA was carried out according to the manufacturer's instructions with the M-MuLV reverse transcriptase (Promega). The reaction system was composed of 5 µl of RNA template, 1 µl of reverse primer 100 pmol/µl, 4 µl of 5x RT buffer, 1 µl of 10 mmol/l dNTPs, 20 U of RNase inhibitor, and 50 U of M-MuLV RT. Distilled H_2O was added to a total volume of 20 µl. The samples were incubated at 42°C for 1 hr.

PCR. Three fragments of the S gene were amplified with primers presented in Table 1. The reaction mixture contained 5 µl of cDNA, 2 mmol/l $MgCl_2$, 1 U Taq polymerase (Promega), 50 pmol of each primer and 1x PCR buffer added to volume 50 µl. PCR conditions consisted of an initial activation step of 95°C for 5 mins, followed by 35 cycles of denaturation at 94°C for 30 secs, annealing at 47°C for 30 secs, extension at 72°C for 1 min, and a final extension at 72°C for 10 mins. The final products were detected by gel electrophoresis, ethidium bromide staining, and UV light detection.

Cloning and sequencing. PCR products were purified from the agarose gels using a DNA fragment quick purification kit (Dingguo) and cloned into the pMD 18-T vector (Takara) according to

Table 1. Primers used for CCoV S gene PCR

Primers	Sequences	Amplicon size (bp)	Location (nt)
S1F	5'-ATG ATT GTG CTT ACA TTG TGC C-3'		
S1R	5'-CCT CTA CGC TTC ATA CCA AGA T-3'	1640	1~1640
S2F	5'-TCT TGG TAT GAA GCG TAG-3'		
S2R	5'-T ATC TGC TGG TTC TTG TTC-3'	1581	1620~3200
S3F	5'-GAA CTG TTAG ATC TAA CAC AT TGG AC-3'		
S3R	5'-CCG AAT TCA GTG AAC ATG C-3'	2080	2290~4370

the manufacturer's instructions. The Cassay Company performed the sequencing (Shanghai, P.R. China).

Phylogenic and recombination analysis. Firstly, the sequences of CCoV 1-71 and NJ-17 were analyzed by BLAST using http://blast.ncbi.nlm.nih.gov/Blast.cgi website. Then, the DNAStar MegAlign program was used for alignment and construction of phylogenetic trees by Jotun Hein method and TreeView was used for visualization of the trees. References for the S gene sequences included sequences of the CCoVs and the representative strains of TGEV and FCoVs available in the GenBank (Table 2). The recombinant signals found in the Group 2 and 3 coronaviruses were scanned for their presence in the CCoV S sequences. The putative recombinant "hot sites" and the CCoV S genes were also analyzed by recombination detection program (RDP) v2.0 software (Martin *et al.*, 2005). Nucleic acid AT content and T_m near the putative recombination sites were analyzed with the Primer Premier 5.0 program.

Analysis of CCoV strain 1-71 S protein structure. S protein signal peptides, N-glycosylation sites, and transmembrane site were analyzed using the Signal P 3.0 Server, NetGlyc, and TMHMM programs using internet site www.cbs.dtu.dk.

Results

CCoV 1-71 S gene sequence and phylogenetic analyses

The complete sequence of CCoV 1-71 S gene is 4362 nts long and encodes 1453 aa (Acc. No. AY7796289). The CCoV 1-71 S gene sequence was compared with the available CCoV strains S genes and with the representative strains of FCoV and TGEV by BLAST (Table 2). The results demonstrated that CCoV 1-71 S gene shares a 99% identity with

Table 2. Examined strains of Group 1 coronaviruses

Virus	Acc. No.	S gene size (bp)	Country and year of virus isolation	Reference
CCoV 1-71	AY7796289*	4362	Germany, 1971	Ma *et al*. (2008)
CCoV NJ17	AY973171*	1581P	China, 2004	Ma *et al.* (2008)
CCoV V1	AY390342	4362	China, 2003	Qiao *et al*. (2005)
CCoV GP	AY436637	4362	China, 2003	Qiao *et al*. (2005)
CCoV K378	X77047	4359	Great Britain, 1993	Horsburgh *et al.* (1995)
CCoV 5821	AB 017789	4362	Japan, 1998	Kokubu *et al.* (1998)
CCoV Insavc-1	D13096	4362	Great Britain, 1980s	Horsburgh *et al.* (1992)
CCoV C54	A22886	4362	Unknown, 1992	Unpublished
CCoV-6	A22882	4335	Unknown, 1992	Unpublished
CCoV TN449	AF116245	2446P	USA, 1980s	Wesley (1999)
CCoV UCD-2	AF116247	2438P	USA, 1980s	Wesley (1999)
CCoV UCD-1	AF116248	2327P	USA, 1980s	Wesley (1999)
CCoV UWSMN-1	AF516907	751P	Australia, 2001	Naylor *et al*. (2002)
CCoV UWSMN-1	AF327928	468P	Australia, 2001	Naylor *et al*. (2002)
CCoV BGF10	AY342160	4362	Great Britain, 2004	Sanchez-Morgado *et al.* (2004)
CCoV Elmo/02	AY307020	4443	Italy, 2002	Pratelli *et al*. (2004)
FCoV 79-1683	X80799	4365	USA, 1970s	Herrewegh *et al.* (1998)
FCoV UCD1	AB088222	4374	USA, 1970s	Motokawa *et al*. (1995)
TGEV	AJ271965	4344	Spain, 2000	Almazan *et al*. (2000)

*Submitted by authors; Ppartial sequences.

Table 3. Sequence comparison of CCoV 1-71 and CCoV NJ17 S gene with Group 1 coronaviruses

Virus	Identity with CCoV 1-71 (%) (length of examined strain/ CCoV 1-71)	Compared region	Identity with CCoV NJ17 (%) (length of examined strain/CCoV NJ17)
CCoV V1	99(4345/4362)	Total length	92(1458/1581)
CCoV GP	99(4340/4362)	Total length	92(1458/1581)
CCoV K378	99(4326/4362)	Total length	92(1458/1581)
CCoV Insavc-1	93(4076/4362)	Total length	91(1436/1575)
CCoV 5821	91(3942/4308)	nt 53–4360	95(1508/1581)
CCoV BGF10	92(3617/3901) 87(373/426)	nt 460–4360 nt 1–425	92(1431/1581)
CCoV Tn449	89(2131/2369)	nt 75–2443	91(754/824)
CCoV UCD-1	91(1333/1449)	nt 991–2438	92(757/819)
CCoV UCD-2	90(1977/2181) 90(40/44)	nt 144–2324 nt 75–118	91(645/705)
CCoV UWSMN-1	92(698/751) 90(259/286) 86(99/114)	nt 1347–2097 nt 488–773 nt 311–424	90(433/477)
FCoV 79-1683	91(3918/4300)	nt 53–4352	92(1467/1581)
TGEV	92(3099/3363)	nt 991–4352	92(1463/1581)

the S gene of CCoV strains V1, GP, and K378, and a 93% identity with the S gene of CCoV strain Insavc-1. However, in some strains the identity was disrupted at certain sites (Table 3) and moreover, we found lower sequence identities at the two flanking sequences. The breakage sites were concentrated in the same region. For instance, the first 53 nts of the 5'-end of the S gene of CCoV 1-71 share a 91% identity with CCoV 5281 and FCoV 79-1683. We found 89% identity with CCoV TN449 and 90% identity with CCoV UCD-2 from the 75th nucleotide on. However, the CCoV 1-71 S gene shared incoherent homology with strains BGF10 and UWSMN-1 that was broken at nt 425~460 and nt 424~488, respectively. At the S gene 5'-end nt 1–991, CCoV UCD-1 shared a low homology with CCoV 1-71 and other CCoVs, but shared a high homology with TGEV (Wesley, 1999).

The DNAStar program was used to analyze all of the known S gene sequences. The phylogenetic trees of CCoV 1-71, CCoVs, FCoV, and TGEV S genes are shown in Fig. 1. Based on the homology breakage sites, the first 425 nts of the S gene 5'-end of the CCoV strains and reference strains (Table 2) were analyzed using DNAStar MegAlign (Fig. 2). The conventional sequences (1620~3200 nt) were amplified using the S2F and S2R primers (1581 bp) and analyzed (Fig. 3).

The results of the complete S gene sequence analysis showed that all CCoV strains were closer to the type II FCoVs than to TGEV, except for the type I CCoV strain Elmo/02 that was closer to the type I FCoV strain UCD1. However, there was a major difference in the first 425 nt sequences even among the CCoVs. TGEV shares only a 22.3~27.1% identity with the other strains, while the type II FCoV strain 79-1683 shares a 77.8~84.9% identity with the CCoVs representing higher numbers than the identity with CCoV type I strain UCD1 (31.3~58.3%.). Based on the phylogenetic trees, the strains were divided into three groups, the type II FCoVs and the type II FCoV on a branch, the type I CCoV and the type I FCoV on a branch, while CCoV UCD-1and TGEV on another branch. In the conserved region, there was no obvious difference between TGEV, FCoV 79-1683, and the type II CCoVs sharing a 90.2~92.9% identity. FCoV UCD1 and CCoV Elmo/02 shared 55.4~75.9% identity with the other strains. CCoV 1-71, CCoV-6, V1, GP, and K378 strains were always at the same branch, when analyzed either as complete genes or in 425 nt segments. However, the CCoV strains C54, BGF10, and 5281 showed some evolutionary independence.

Based on the sequence scan, we found that all of the CCoV strains except for CCoV-6 and FCoV 79-1683 contained ATGATG hexanucleotide repeat at nt 371 (or 374) and at nt 554 (or 557) that flanked both RNA recombination sites in MHV (Banner *et al*., 1991). A breakage site (nt 425) was located between two hexanucleotide repeats. Recombination events using RDP software in the CCoV strains C54 and 5821 were also found close to the two hexanucleotides at nt 3044 and 3263. Another signal sequence, TCTAA (Liao *et al*., 1992), which regularly appeared at the recombination sites was found in the CCoV S gene at nts 1723~1727, 2301~2305,

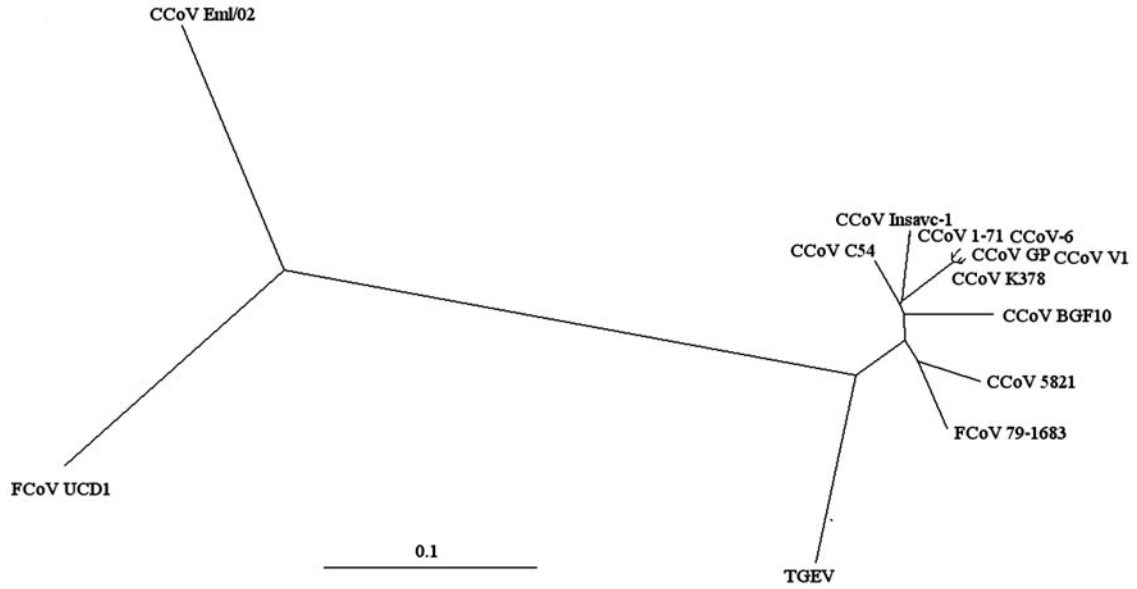

Fig. 1

Phylogenetic tree of Group 1 coronaviruses based on complete sequence of the S gene

Bar 0.1 represents evolutionary distance.

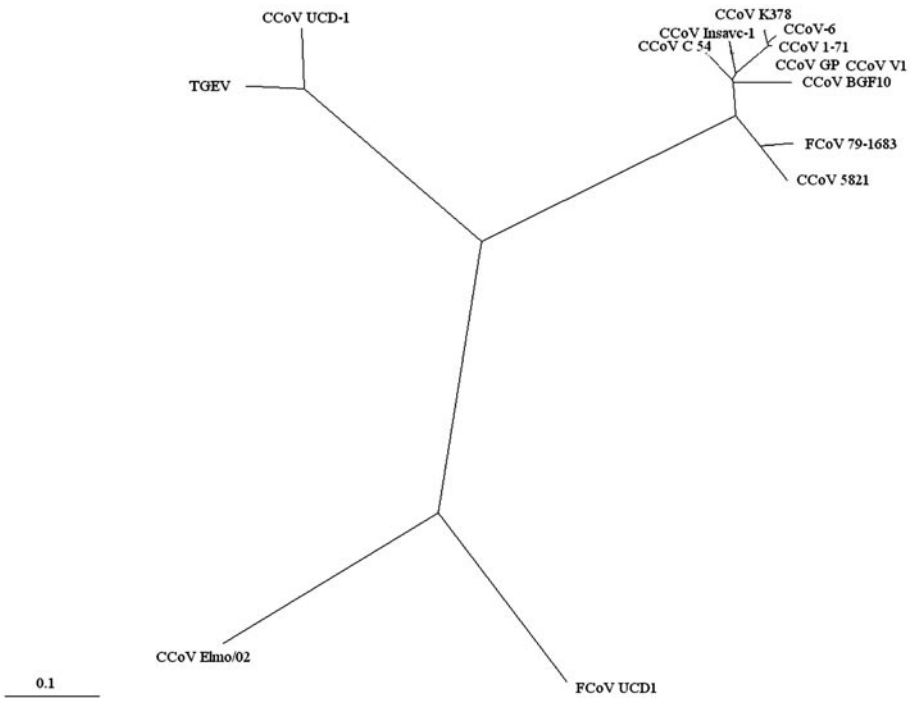

Fig. 2

Phylogenetic tree of Group 1 coronaviruses based on the S gene region of nt 1–425

Bar 0.1 represents evolutionary distance.

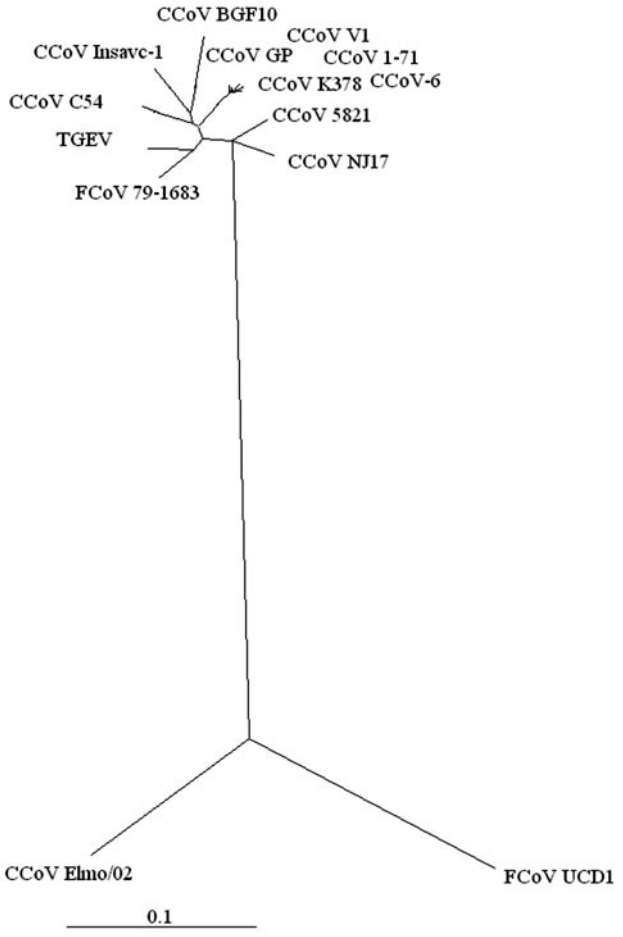

Fig. 3

Phylogenetic tree of Group 1 coronaviruses based on the S gene region of nt 1620–3200

Bar 0.1 represents evolutionary distance.

and 2425~2429. Results of the sequence analysis showed that all of the CCoV strains were conserved and related to FCoV 79-1683 and TGEV at the TCTAA (nt 1723~1727) pentanucleotide upstream region. CCoV 1-71, GP, V1, K378, and CCoV-6 had the same mutation sites between this pentanucleotide (nt 1723~1727) and other pentanucleotide TCTAA (nt 2425~2429), while the other CCoV strains shared the same mutation sites with FCoV and TGEV. This finding was an obvious recombination event. Liao and Lai, 1992 found that a nonanucleotide, CTTTATAAA has always appeared downstream of the RNA recombination signal TCTAA of the S gene. However, we did not find this nonanucleotide at the given position. On the other hand, a similar AACTAAAAAT decanucleotide was found near 425 nt breakage site (Fig. 4a). Another similar nonanucleotide TTGTAAAAA/GT was also found downstream of TCTAA nt 1723~1727 and 2425~2429 (Fig 4b). In addition, we found a number of CTT(A/T) (A/T)G (Wang *et al.*, 1993) nucleotide repeats in the S gene at nts 1237, 1323, 3125, 3609, 3743, and 4166. Results of the RDP analysis showed that nt 1545–3405 region and the 3'-end region (nt 4203–4356) of the CCoV S genes were recombinant hot regions and that recombinant events in these regions were found in many strains (Fig. 5). The sites of recombination were near the recombinant signal sequences TCTAA or CTT(A/T) (A/T)G.

The AT content was higher near the breakage site or the signal sequence in comparison with entire S gene as determined by nucleic acid analysis. The AT content in these regions was 67.74~80% and in the complete S gene sequence 62.68%. The highest value (80%) was found near the 425 nt breakage site.

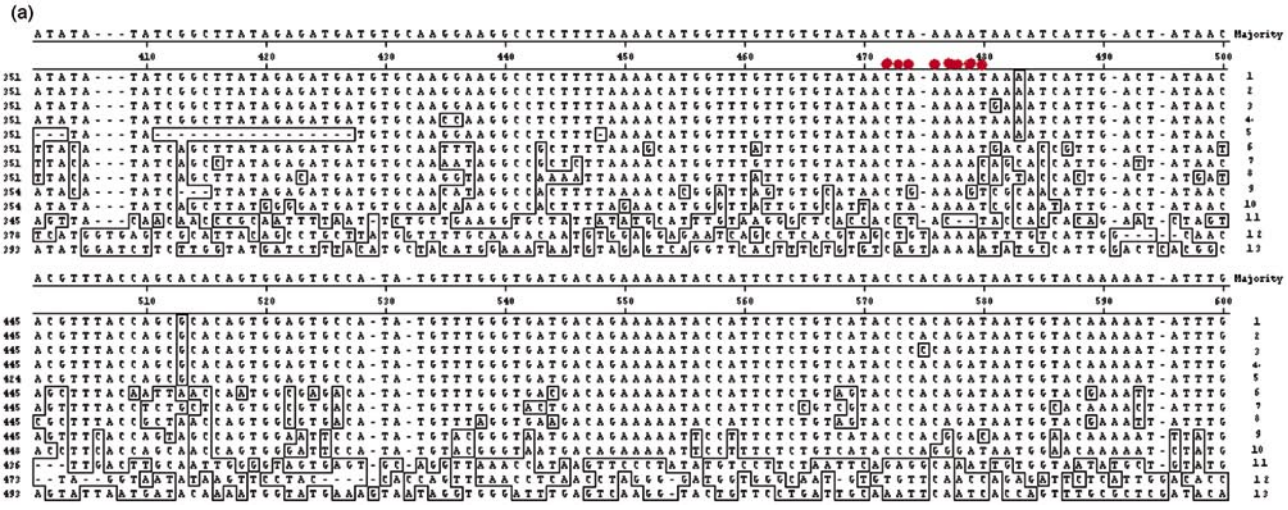

Fig. 4

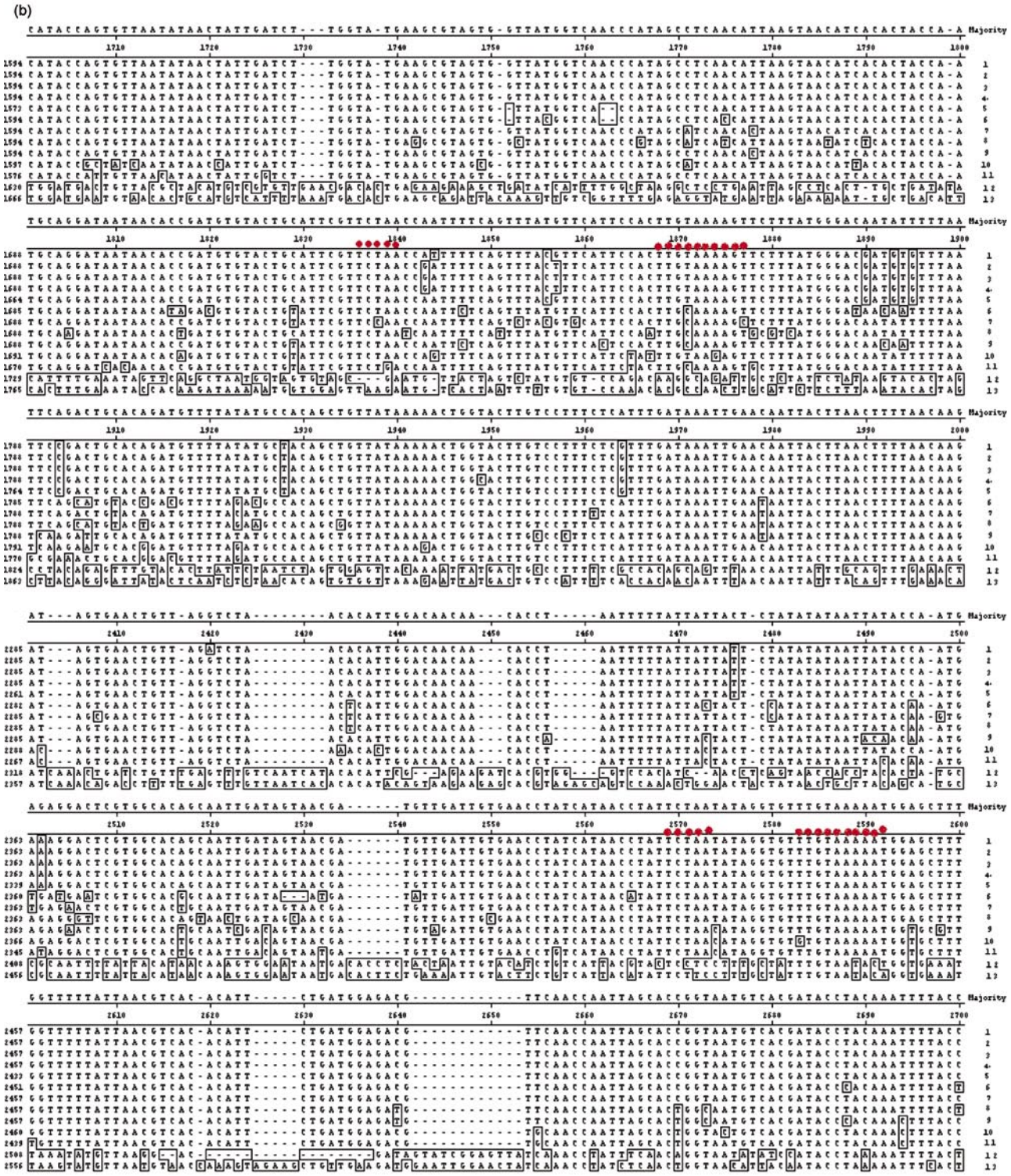

Fig. 4

Putative recombination "hot sites" in the CCoVs S genes

1. CCoV 1-71; 2. CCoV V1; 3. CCoV GP; 4. CCoV K378; 5. CCoV-6; 6. CCoV Insavc-1; 7. CCoV-C54; 8. CCoV BGF10; 9. CCoV 5821; 10. FCoV 79-1683; 11. TGEV; 12. FCoV UCD1; 13 CCoV Elmo/02. Red dots represent putative signal sequences. (a) Putative recombination "hot site" AACTAAAAAT close to the nt 425 breakage site at the 5'-end of the S gene. (b) Recombination signal sequence TCTAA found in the CCoV S gene at nt 1723~1727, 2301~2305, and 2425~2429.

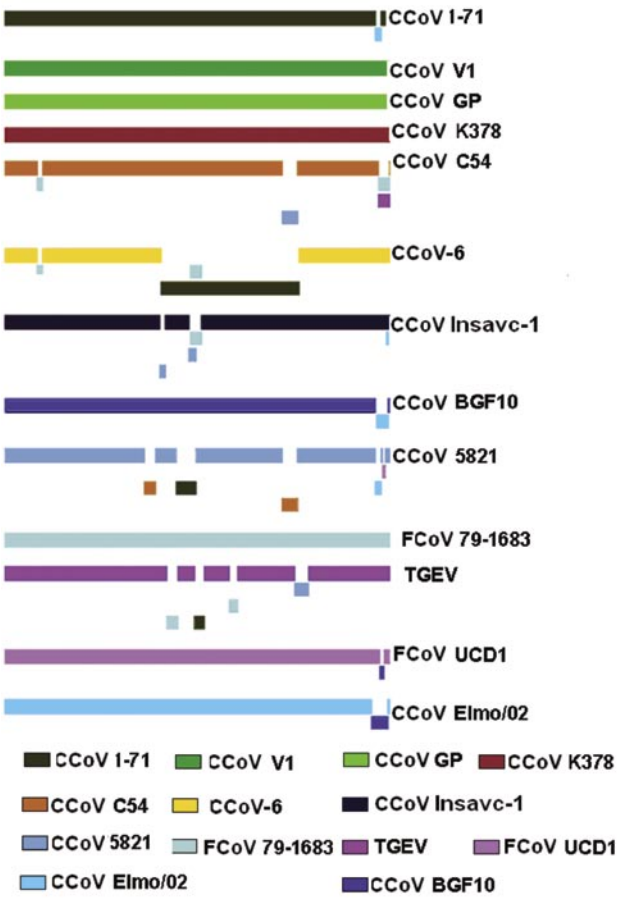

Fig. 5

Putative recombination regions in the S genes of Group 1 coronaviruses

CCoV 1-71 S protein structure and its predicted function

Protein prediction analysis showed that the CCoV 1-71 S protein consisted of aa 1453 with 31 N-glycosylation sites. The signal peptide was most likely cleaved between aa at positions 18 and 19 and the transmembrane fragment was found between aa 1393 and 1415. The RRXRR enzyme cleavage site was not found suggesting that the CCoV 1-71 S protein cannot be cleaved into S1 and S2 fragments. All of the type II CCoV strains and the type II FCoV 79-1683 had similar signal peptide cleavage sites (VIC-TS, VSC-TS or ASS-TS) with the exception of type I CCoV Elmo/02 that has the same cleavage site as type I FCoV UCD1 (ANC-KD) and TGEV that has a different cleavage site (IYG-DN). The S protein of the examined strains had more variations in the N-terminal 700 aa. N-glycosylation sites were also different in this region, although relatively conserved between some strains. The NetGlyc program prediction showed that CCoV 1-71 shared the same N-glycosylation sites with CCoV strains V1, GP, and K378, and only one N-glycosylation site was different from that of FCoV 79-1683, which lacked an N-838 glycosylation site. CCoV Insavc-1 and C54 had the same N-glycosylation sites, although one glycosylation site at N-142 was different from CCoV 1-71 and lacked an N-566 glycosylation site, when compared to CCoV 1-71. Compared with CCoV 1-71, CCoV-6 lacked N-glycosylation sites at N-94, N-379, N-409, and N-905. CCoV strains BGF10 and 5821 had a specific N-glycosylation site at N-635 that was a pro-X1 site. The CCV BGF10 also had N-142 site and CCoV 5821 lacked N-379 site when compared CCoV 1-71. Most of the N-glycosylation sites of the FCoV UCD1 S protein were the same as in CCoV Elmo/02 and were located downstream the aa 800. This similarity was also present in the S protein of CCoV1-71 and TGEV downstream the aa 400. All of the CCoV strains had the same N-glycosylation sites at the N-terminus. Completely different aa sequence was found at TGEV, FCoV UCD1, and CCoV Elmo/02. There were no N-glycosylation sites at the N-terminal 160 aa of the S protein at FCoV UCD1 and CCoV Elmo/02, while 3 different N-glycosylation sites were found on TGEV. In all examined coronaviruses the transmembrane sites were similar and conserved with a leucine zipper structure at aa 1346–1367 and a cysteine rich region between aa 1408–1436. This finding was consistent with the previous report (Sanchez *et al.*, 1999). Four amino acid mutations E-521, H-577, D-767, and S-1141 were unique to CCoV 1-71 replacing K, A, G, and N, respectively.

Discussion

There are many sub-genomes of the coronavirus in the infected cell for the reason of special transcription events (Cavanagh, 1997). The replicase complex, which is involved in the synthesis of new RNA strand along with a template, may dissociate from the original genomic template and use another RNA strand as a template to complete synthesis of the genomic RNA (Liao *et al.*, 1992). This mechanism of replication suggests that potential recombination events occur between different original RNAs and as a consequence they encourage the evolution of the virus.

In this study, we found that the CCoV 1-71 S gene shared a high identity with the CCoV strains K378, CCoV-6, V1, and GP, what suggested that the CCoV 1-71-like CCoVs are widely distributed throughout the world indicating that these coronavirus strains may have the same ancestor. The CCoV strains Insavc-1, C54, BGF10, and 5281 shared lower (91.2~94.1%) identity with CCoV 1-71 suggesting that they evolved independently. The differences in the identity at each site of the breakage suggested that recombinant events had occurred. The identity differences were distinct and occurred

near nt 425. The differences also appeared between the two hexanucleotide repeats ATGATG sequences that served as recombinant signals in the Group 2 coronaviruses (Banner *et al.*, 1991). However, in the conserved nt 1723–2429 region, there was another obvious variation between the two pentanucleotides TCTAA. The S gene sequence of the CCoV 1-71, GP, V1, K378, and CCoV-6 strains was more likely to be replaced by a different sequence than in the other strains. A characteristic sequence G/CTAAAAA/GT was always present downstream of the two pentanucleotides TCTAA and nt 425 breakage site. The identity differences adjacent to the breakage sites and the signal-like sequences revealed that some recombinant events had occurred. Thus, we hypothesized that CCoV had a recombination mechanism similar to that of the Group 2 coronaviruses and several hot sites in the S gene. No special sequences were found at nt 991 breakage and the upstream CCoV UCD-1 sequence shared a high identity and the same host cell specificity as that of TGEV (Wesley, 1999). Near the nt 991 site there might be another recombinant signal sequence or pattern even only as a coincidence.

The recombination sites were also found downstream of the pentanucleotide TCTAA. However, when a nonanucleotide CTTTATAAA merged with the TCTAA, the recombination site was always retro-positioned (Liao *et al.*, 1992). Here, a similar nonanucleotide G/CTAAAAA/GT was found in the CCoV S gene near the putative recombination sites and likely served as the recombinant signal sequence. The pentanucleotide TCTAA was supposed to be similar to the sequence CTTAAG upstream of the 5'-end of the coronavirus S gene ORF. The sequence CTTTATAAA was supposed to be similar to the sequence CTTAACAA imitating binding sites for the RNA polymerase (Liao *et al.*, 1992).

In this study, we analyzed the putative recombination sites and found that the AT contents were higher than the average level that is typical of the S genes. In the CCoV S gene, the G/CTAAAAA/GT downstream of TCTAA contains five adenosines that may allow the RNA polymerase complex to dissociate easily and enhance the frequency of the RNA recombinants. Near other breakage sites at nts 53 and 75, we did not identify known signals or other characteristic sequences, although the AT content was higher than average. When AT content is high, the RNA polymerase complex easily dissociates from the template strand and then re-associates with another template strand that is similar in nucleotide sequence.

The 5'-end mutation in the S gene caused variation in virulence, antigenicity, and cell tropism (Sanchez-Morgado *et al.*, 2004). Differences of the signal peptide cleavage sites indicated a virus genotype. The cleavage sites occurred in the same place of the identity breakage sites at nt 53 in the Blast results, which showed that the first 53 nt was correlated with the genotype of Group 1 coronaviruses.

Prediction analysis of the S proteins showed that most of the N-glycosylation sites and secondary structures were conserved within the CCoV strains, but the aa sequences were dramatically different and several N-glycosylation sites changed near the 5'-end. The function of the S protein is mainly based on its glycosylation. For instance, most of the type II CCoVs have a cell tropism similar to those of type II FCoV, and this may be due to the similar N-glycosylation sites that were identified in this study (Mostl, 1990). Sanchez-Morgado *et al.* (2004) found that the protein region encoded by nt 217–665 of the S gene determined the virulence and cell tropism. In our study, a breakage site in CCoV BGF10 and recombinant events in CCoV C54 and CCoV-6 were found in the region of nt 425. The protein region encoded by nt 1518–2184 was previously found to be recognized by the CCoV-receptor aminopeptidase N (APN) (Benbacer *et al.*, 1997). The CCoV strain 5821 can infect canine respiratory system and in this receptor-recognized region has a special N-glycosylation site, N-635. Other putative recombination sites and events in CCoV-6, Insavc-1, 5821, and TGEV were also found in this area. Therefore, the recombination events in this region may have an impact on the ability of CCoV binding to its receptor.

Analysis of the CCoV S gene sequences in this study demonstrated that RNA recombination had occurred in the S gene. In addition, a similarity of N-glycosylation sites consistent with the nucleotide changes of the S gene suggested that the variation in the cell tropism and virulence of the CCoV strains were mostly acquired by the RNA recombination and not simply by nucleotide mutation or selective advantage.

References

Almazan F, Gonzalez JM, Penzes Z, Izeta A, Calvo E, Plana-Duran J, Enjuanes L (2000): Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. Proc. Natl. Acad. Sci. USA 97, 5516–5621. doi:10.1073/pnas.97.10.5516

Banner LR, Lai MM (1991): Random nature of coronavirus RNA recombination in the absence of selection pressure. Virology 185, 441–445. doi:10.1016/0042-6822(91)90795-D

Benbacer L, Kut E, Besnardeau L, Laude H, Delmas B (1997): Interspecies aminopeptidase-N chimeras reveal species-specific receptor recognition by canine coronavirus, feline infectious peritonitis virus, and transmissible gastroenteritis virus. J. Virol. 71, 734–737.

Binn LN, Lazar EC, Keenan KP, Huxsoll DL, Marchwicki RH, Strano AJ (1974): Recovery and characterization of a coronavirus from military dogs with diarrhea. Proc. Annu. Meet. US Anim. Health Assoc. 78, 359–366.

Cavanagh D (1997): Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch. Virol 142, 629–633.

Herrewegh AA, Smeenk I, Horzinek MC, Rottier PJ, 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.

Horsburgh BC, Brierley I, Brown TD (1992): Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA. J. Gen. Virol. 73, 2849–2862. doi:10.1099/0022-1317-73-11-2849

Horsburgh BC, Brown TD (1995): Cloning, sequencing and expression of the S protein gene from two geographically distinct strains of canine coronavirus. Virus Res. 39, 63–74. doi:10.1016/S0168-1702(95)00068-2

Kokubu T, Taharaguchi S, Katayama S, Hatano M, Takahasi T, Iwamoto K, Masubuchi K, Yamanaka M, Sasaki O, Inaba Y (1998): Nucleotide sequence of the spike protein of a canine coronavirus, strain 5821. J. Vet. Med. Sci. 51, 251–255.

Lee C W, Jackwood MW (2000): Evidence of genetic diversity generated by recombination among avian coronavirus IBV. Arch. Virol. 145, 2135–2148. doi:10.1007/s007050070044

Liao CL, Lai MM (1992): RNA recombination in a coronavirus: recombination between viral genomic RNA and transfected RNA fragments. J. Virol. 66, 6117–6124.

Ma GG, Wang YY, Lu CP (2008): Molecular characterization of the 9.36 kb C-terminal region of canine coronavirus 1-71 strain. Virus Genes 36, 491–497. doi:10.1007/s11262-008-0214-4

Martin DP, Williamson C, Posada D (2005): RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262. doi:10.1093/bioinformatics/bth490

Mostl K (1990): Coronaviridae, pathogenetic and clinical aspects: an update. Comp. Immunol. Microbiol. Infect. Dis. 13, 169–180. doi:10.1016/0147-9571(90)90085-8

Motokawa K, Hohdatsu T, Aizawa C, Koyama H, Hashimoto H (1995): Molecular cloning and sequence determination of the peplomer protein gene of feline infectious peritonitis virus type I. Arch. Virol. 140, 469–480. doi:10.1007/BF01718424

Naylor MJ, Harrison GA, Monckton RP, McOrist S, Lehrbach PR, Deane EM (2001): Identification of canine coronavirus strains from feces by S gene nested PCR and molecular characterization of a new Australian isolate. J. Clin. Microbiol. 39, 1036–1041. doi:10.1128/JCM.39.3.1036-1041.2001

Naylor MJ, Walia CS, McOrist S, Lehrbach PR, Deane EM, Harrison GA (2002): Molecular characterization confirms the presence of a divergent strain of canine coronavirus (UWSMN-1) in Australia. J. Clin. Microbiol. 40, 3518–3522. doi:10.1128/JCM.40.9.3518-3522.2002

Pratelli A, Elia G, Decaro N, Tola S, Tinelli A, Martella V, Rocca S, Tempesta M, Buonavoglia C (2004): Cloning and expression of two fragments of the S gene of canine coronavirus type I. J. Virol. Methods 117, 61–65. doi:10.1016/j.jviromet.2003.12.003

Pratelli A, Martella V, Decaro N, Tinelli A, Camero M, Cirone F, Elia G, Cavalli A, Corrente M, Greco G, Buonavoglia D, Gentile M, Tempesta M, Buonavoglia C (2003): Genetic diversity of a canine coronavirus detected in pups with diarrhoea in Italy. J. Virol. Methods 110, 9–17. doi:10.1016/S0166-0934(03)00081-8

Pratelli A, Tinelli A, Decaro N, Camero M, Elia G, Gentile A, Buonavoglia C (2002): PCR assay for the detection and the identification of atypical canine coronavirus in dogs. J. Virol. Methods 106, 209–213. doi:10.1016/S0166-0934(02)00165-9

Qiao J, Xia X, Meng Q, Hu G, Xie ZH, Yan F, Yang S, Huang G (2004a): Cloning and Sequence Analysis of Spike Protein Gene of Canine Coronavirus Strain V1. J. Shanghai Jiaotong University (Agricultural Science) 22, 53–58.

Qiao J, Xia X, Meng Q, Hu G, Xie, ZH, Yan, F, Yang, S, Huang G (2004b): Cloning and SequenceAnalysis of Spike Protein Gene of Canine Coronavirus Giant Panda's Isolate. Virol. Sin. 19, 48–50.

Sanchez-Morgado JM, Poynter S, Morris TH (2004): Molecular characterization of a virulent canine coronavirus BGF strain. Virus Res. 104, 27–31. doi:10.1016/j.virusres.2004.02.038

Sanchez CM, Izeta A, Sanchez-Morgado JM, Alonso S, Sola I, Balasch M, Plana-Duran J, Enjuanes L (1999): Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. J. Virol. 73, 7607–7618.

Wang L, Junker D, Collisson EW (1993): Evidence of natural recombination within the S1 gene of infectious bronchitis virus. Virology 192, 710–716. doi:10.1006/viro.1993.1093

Wang YY, Lu CP (2005): Recombination Analysis of M Protein of a Canine Coronavirus Wild Strain Isolated From Nanjing City. Virol. Sin. 20, 262–267.

Wang YY, Ma GG, Lu CP, Wen H (2006): Detection of canine coronaviruses genotype and in raised canidae animals in China Berl. Muench. Tieraerztl. Wochenschr. 119, 35–39.

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. doi:10.1016/S0168-1702(99)00032-5

Wu HY, Guy JS, Yoo D, Vlasak R, Urbach E, Brian DA (2003): Common RNA replication signals exist among group 2 coronaviruses: evidence for in vivo recombination between animal and human coronavirus molecules. Virology 315, 174–183. doi:10.1016/S0042-6822(03)00511-7

Zarnke RL, Evermann J, Ver Hoef JM, McNay ME, Boertje RD, Gardner CL, Adams LG, Dale BW, Burch J (2001): Serologic survey for canine coronavirus in wolves from Alaska. J. Wildl. Dis. 37, 740–745.