Journal of General Virology (2013), **94**, 2036–2049 **DOI** 10.1099/vir.0.054940-0

Evolutionary dynamics of bovine coronaviruses: natural selection pattern of the spike gene implies adaptive evolution of the strains

Mehdi R. M. Bidokhti,[1] Madeleine Tråvén,[1] Neel K. Krishna,[2] Muhammad Munir,[3,4] Sándor Belák,[3,4] Stefan Alenius[1] and Martí Cortey[5]

Correspondence
Mehdi R. M. Bidokhti
Mehdi.bidokhti@slu.se

[1]Division of Ruminant Medicine and Veterinary Epidemiology, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Box 7054, SE-750 07 Uppsala, Sweden

[2]Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, 700 West Olney Road, Norfolk, VA 23507-1696, USA

[3]Department of Virology, Immunobiology and Parasitology, National Veterinary Institute, Box 7028, SE-751 89 Uppsala, Sweden

[4]Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

[5]Department of Immunology, The Pirbright Institute, Ash Road, Pirbright GU24 0NF, UK

Coronaviruses demonstrate great potential for interspecies transmission, including zoonotic outbreaks. Although bovine coronavirus (BCoV) strains are frequently circulating in cattle farms worldwide, causing both enteric and respiratory disease, little is known about their genomic evolution. We sequenced and analysed the full-length spike (S) protein gene of 33 BCoV strains from dairy and feedlot farms collected during outbreaks that occurred from 2002 to 2010 in Sweden and Denmark. Amino acid identities were >97 % for the BCoV strains analysed in this work. These strains formed a clade together with Italian BCoV strains and were highly similar to human enteric coronavirus HECV-4408/US/94. A high similarity was observed between BCoV, canine respiratory coronavirus (CRCoV) and human coronavirus OC43 (HCoV-OC43). Molecular clock analysis of the S gene sequences estimated BCoV and CRCoV diverged from a common ancestor in 1951, while the time of divergence from a common ancestor of BCoV and HCoV-OC43 was estimated to be 1899. BCoV strains showed the lowest similarity to equine coronavirus, placing the date of divergence at the end of the eighteenth century. Two strongly positive selection sites were detected along the receptor-binding subunit of the S protein gene: spanning amino acid residues 109–131 and 495–527. By contrast, the fusion subunit was observed to be under negative selection. The selection pattern along the S glycoprotein implies adaptive evolution of BCoVs, suggesting a successful mechanism for BCoV to continuously circulate among cattle and other ruminants without disappearance.

Received 30 April 2013
Accepted 25 June 2013

INTRODUCTION

Bovine coronavirus (BCoV) is a member of the family *Coronaviridae*, in the order *Nidovirales* (Cavanagh, 1997). Coronaviruses possess the largest viral RNA genome in nature. Recently, the International Committee on Taxonomy of Viruses has proposed two subfamilies for the family *Coronaviridae*: subfamily *Coronavirinae* and subfamily *Torovirinae*. The former comprises three groups renamed as genera *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus*, respectively (de Groot *et al.*, 2012) and with a novel (but yet to be approved) genus, provisionally named *Deltacoronavirus* (Woo *et al.*, 2012). Four separate lineages (A through D), some of them encompassing multiple virus species, are commonly recognized within the genus *Betacoronavirus*. BCoV, together with human coronavirus OC43 (HCoV-OC43), equine coronavirus (ECoV) and porcine haemagglutinating encephalomyelitis virus (PHEV) belong to the virus species *Betacoronavirus 1* of the lineage A of the genus *Betacoronavirus* (de Groot *et al.*, 2012). A recently isolated canine respiratory coronavirus (CRCoV) has also shown a high genetic similarity to *Betacoronavirus 1* (Erles *et al.*, 2007).

The GenBank/EMBL/DDBJ accession numbers for the sequences of the S gene of BCoV strains reported in this study are KF169908–KF169940.

One supplementary figure is available with the online version of this paper.

BCoV is an enveloped virus with a single-stranded, positive-sense, non-segmented RNA genome of approximately 31 kb (Clark, 1993). A 4092 nt fragment of the BCoV genome encodes the large petal-shaped surface spike (S) protein. This is a type 1 membrane glycoprotein of 1363 aa that comprises two hydrophobic regions, an N-terminal signal sequence and a C-terminal membrane anchor (Parker *et al.*, 1990). The S protein is cleaved by an intracellular protease between amino acids 768 and 769 to form two functionally distinct subunit domains, a variable S1 N-terminal domain (NTD) and the more conserved S2 C-terminal domain (Abraham *et al.*, 1990). The S1 subunit is a peripheral protein, mediating virus binding to host-cell receptors (Li, 2012; Peng *et al.*, 2012), haemagglutinating activity (Schultze *et al.*, 1991) and inducing neutralizing antibodies (Yoo & Deregt, 2001). The S2 subunit is a transmembrane protein which mediates fusion of viral and cellular membranes (Yoo *et al.*, 1991a).

BCoV is the causative agent of neonatal calf diarrhoea (CD), winter dysentery (WD) in adult cattle (Alenius *et al.*, 1991; Mebus *et al.*, 1973; Saif *et al.*, 1988) and respiratory tract disorders in cattle of all ages (Cho *et al.*, 2001; Decaro *et al.*, 2008a; Lathrop *et al.*, 2000). This infection is not effectively controlled in the herds by current commercial vaccines (Saif, 2010). BCoV negatively impacts the cattle industry as it results in reduced milk production, loss of body condition and the death of young animals (Clark, 1993; Saif, 2010). BCoV outbreaks most often happen during autumn and winter (Clark, 1993). However, studies from various climatic regions have also reported BCoV outbreaks in the warmer seasons (Bidokhti *et al.*, 2012; Decaro *et al.*, 2008b; Park *et al.*, 2006).

Studies have shown a high prevalence of BCoV infections in cattle farms in many countries (Fulton *et al.*, 2011; Paton *et al.*, 1998; Saif, 2010; Tråvén *et al.*, 2001). Also BCoV-like coronaviruses transmissible to gnotobiotic (Gn) calves have been found among various wild ruminants (Alekseev *et al.*, 2008; Tsunemitsu *et al.*, 1995). The public health impact of BCoVs has also been raised due to the isolation of a BCoV-like human enteric coronavirus-4408/US/94 (HECV-4408/US/94) from a child with acute diarrhoea (Zhang *et al.*, 1994), and also the outbreaks of severe acute respiratory syndrome coronavirus (SARS-CoV) (Groneberg *et al.*, 2003; Zhong & Wong, 2004). Molecular evolutionary analysis of HCoV-OC43 isolates suggests BCoV is their genetically closest counterpart compared with other coronavirus species (Vijgen *et al.*, 2006). Recently, a novel coronavirus, HCoV-EMC, was found that has been circulating in the Middle East and caused death with similar clinical signs to SARS-CoV (Al-Ahdal *et al.*, 2012; Zaki *et al.*, 2012). Such veterinary and public health concerns rationalize the study of the genetic diversity and evolution of BCoV strains and their relationship with the other betacoronaviruses.

The S gene sequence of BCoV has been exploited for epidemiological (Bidokhti *et al.*, 2012; Decaro *et al.*, 2008c; Hasoksuz *et al.*, 2002; Jeong *et al.*, 2005; Lathrop *et al.*, 2000; Liu *et al.*, 2006; Martínez *et al.*, 2012) and evolutionary (Vijgen *et al.*, 2005b; Woo *et al.*, 2012) studies. So far, no study has systematically defined the positive selection pattern of the S protein of BCoV strains which is probably important for BCoV adaptive evolution. In the present study, to better understand the epidemiological dynamics of BCoV and to investigate the adaptive evolutionary process of BCoVs, we sequenced the full-length S gene and analysed the molecular epidemiology, evolution and selective pressures of this virus in cattle herds in Sweden and Denmark. Reference strains from other hosts in virus species *Betacoronavirus 1* including human, wild ruminants, pig and horse and also CRCoV from dog were included in this analysis to estimate their time of divergence and update their genetic relationship.

RESULTS

Sequence data and genome analysis

Comparative analysis of the S gene (4092 nt) indicated that all 33 Swedish and Danish strains (GenBank accession numbers: KF169908–KF169940) shared a high degree of sequence identity both at the nucleotide level (>97.8 %) and deduced amino acid level (>97.4 %). Compared with the BCoV/Mebus/US/72 strain, 78–113 nt substitutions (97.2–97.9 % sequence identity) were found resulting in 37–54 aa changes (96–97.2 % sequence identity) within the entire S gene of the strains. The 100 % identical strains SWE/I/07-3, SWE/I/07-4 and SWE/I/07-5 from Sweden were found to be 99.7 % similar to the strain SWE/P/09-1. SWE/I/07-3 and SWE/I/07-4 were obtained from different cows with enteric disease in the same herd in Gotland island in south-eastern Sweden. SWE/I/07-5 was obtained from another herd in Gotland island during the same time. SWE/P/09-1 was obtained from a cow with respiratory disease in a herd in south-western Sweden.

SWE/N/05-1 and SWE/N/05-2 showing 8 nt substitutions (99.8 % identity) were sampled from different calves with enteric disorders at the same occasion in a large dairy herd. The oldest strain, SWE/C/92 showed the highest identity (nucleotide 98.7 %, amino acid 98.7 %) to an old strain, DEN/03-3, and the lowest identity (nucleotide 97.8 %, amino acid 97.4 %) to a recent strain, SWE/M/10-1. SWE/Y/10-3 from northern Sweden and SWE/P/10-4 from south-western Sweden showed 99.9 % nucleotide identity. These strains were obtained during the same year from different regions.

The analysis of the predicted S proteins of the present 33 BCoV strains revealed a potential N-terminal signal peptide of about 14 aa by SignalP-HMM and SignalP-NN, respectively. A potential S1/S2 cleavage site located after RRSRR, identical for BCoV (Abraham *et al.*, 1990) and some HCoV-OC43 (Lau *et al.*, 2011), was identified in the S proteins of all strains excluding the 2010 strains. The

R-to-K amino acid change in the 764 position, leading to a KRSRR motif, was observed in the S proteins of SWE/Y/10-3 and SWE/P/10-4. The A-to-E amino acid change in the 769 position, leading to a RRSRRE motif, was observed downstream of the potential cleavage site in the S proteins of SWE/M/10-1 and SWE/M/10-2. It has been suggested that changes in the last position of the motif affect the S protein cleavability (Vijgen *et al.*, 2005a). This cleavage process is believed to play an important role in the fusion activity and viral infectivity of BCoV (Storz *et al.*, 1981; Vijgen *et al.*, 2005a). More sequence data and experimental studies are required to clarify the important role of these changes in the cleavage site of BCoV. The analysis of the S protein showed 20 potential *N*-linked glycosylation sites in all Swedish and Danish BCoV strains, with nine NXS (T133, M359, V437, P444, S696, D788, F895, I1234, Q1288) and 11 NXT (T59, F198, A649, R676, N714, S739, C937, N1194, Y1224, Q1253, V1267) sites.

Phylogenetic tree

The analysed samples showed low variability. Within the 4092 nt of the complete sequences of the S protein gene, 340 nt were variable (8.3 %). At the amino acid level the variation was slightly larger (147 variable amino acid residues, 10.8 %). Nucleotide proportion (p) distances among strains ranged between 0.1 and 2.7 %. This high degree of sequence identity is reflected in the neighbour-joining tree (Fig. 1): all Swedish and Danish strains from 2002 to 2010 clustered together as a unique clade with the Italian strains, BuCoV/ITA/179-07-11, BCoV/438/06-2/ITA and BCoV/ITA/339/06. The oldest Swedish strain SWE/C/92 was on a different branch from this clade and clustered into a separate clade with BCoV/GER/M80844/89 and human isolate HECV-4408/US/94. The remaining reference strains derived from cattle and wild ruminants clustered irrespective of the host. The CRCoV clade was most closely related to the BCoV and BCoV-like coronavirus clade; while HCoV-OC43, PHEV and ECoV clusters were more distant (Fig. 1).

There were 53 nucleotide differences found between strains SWE/M/06-3 and SWE/M/06-4 (98.7 % nucleotide similarity, 98.1 % amino acid similarity). These strains were obtained from two dairy herds with CD symptoms sampled at the same time in southern Sweden. SWE/M/06-3 clustered with SWE/AC/08-1, SWE/C/08-2, SWE/Z/07-1, SWE/C/07-2, SWE/C/07-6 and SWE/U/09-3 (Fig. 1), sharing more than 99.4 % sequence similarity.

Evolutionary rate and estimation of divergence dates

Using a Bayesian coalescent approach, molecular clock analysis of the S gene sequences of Swedish and Danish BCoV strains and reference strains of *Betacoronavirus 1* was performed to estimate the mean rate of evolution and their time to the most recent common ancestor (TMRCA) (Table 1). TMRCA of CRCoV and BCoV was estimated to be the year 1951. The mean evolution rate of Swedish and Danish BCoV strains compared with CRCoV was also estimated to be 4.4×10^{-4} substitutions per site year^{-1}. TMRCA analysis estimated earlier divergence of BCoV strains from HCoV-OC43 (1899), PHEV (1847) and ECoV (1797). The mean evolution rate of Swedish and Danish BCoV strains compared with HCoV-OC43 was 4.1×10^{-4} substitutions per site year^{-1}, 7.6×10^{-4} compared with PHEV and 7.9×10^{-4} substitutions per site year^{-1} compared with ECoV. TMRCA of BCoV compared with coronaviruses from wild ruminants was dated to 1963 and the mean rate of evolution was estimated to be 4.4×10^{-4} substitutions per site year^{-1}. Swedish and Danish BCoV strains sequenced in this study showed the highest mean rate of evolution to BCoV reference strains and HECV-4408/US/94: 8.7×10^{-4} and 8.3×10^{-4} substitutions per site year^{-1}, respectively. This resulted in estimating almost the same year for TMRCA, 1978 and 1977, respectively (Table 1).

Results from bootscan analysis were in line with the observations described above and with the phylogenetic tree (Fig. 1). Bootscan analysis showed a number of possible recombination sites when the S gene of BCoV strains were used as the query. Most of the region exhibits higher bootstrap support for the clustering of BCoV strains with CRCoV, except upstream of position 500, where higher bootstrap support for clustering with HCoV-OC43 strains was observed. Similar results were obtained when CRCoV strains were subjected to bootscan analysis (Fig. S1, available in JGV Online). When the S gene of HCoV-OC43 strains were used as the query, downstream of position 1800 exhibits higher bootstrap support for the clustering of HCoV-OC43 strains with PHEV. Similar results were obtained when PHEV strains were subjected to bootscan analysis (Fig. S1).

Selective pressure sites

The selection profiles of the amino acid sequence of all 33 Swedish and Danish BCoV strains showed two general patterns within the S protein. The cumulative difference between the non-synonymous substitution rate (d_N) and the synonymous substitution rate (d_S) (i.e. $d_N - d_S$) revealed that amino acid residues 109–131 and 495–527 of the S1 subunit were under strong positive selection (Fig. 2a). Amino acid residues 36–97, 315–420, 498–713, 910–1032, 1059–1234 and 1245–1279 were under negative selection. They covered most of the S2 subunit, indicating that S2 is relatively stable in BCoV (Fig. 2a).

SNAP analysis identified 133 positively selected sites. Of these 89 are in S1 and 44 in the S2 domain (Fig. 2b). Several of these sites were also identified by the random effects likelihood (REL) method at the posterior probability $P>90$ % level. The following positive selection sites were identified by SNAP and REL methods: 35, 112, 113, 115, 143, 147, 151, 157, 188, 257, 447, 458, 471, 482, 499, 501, 503,

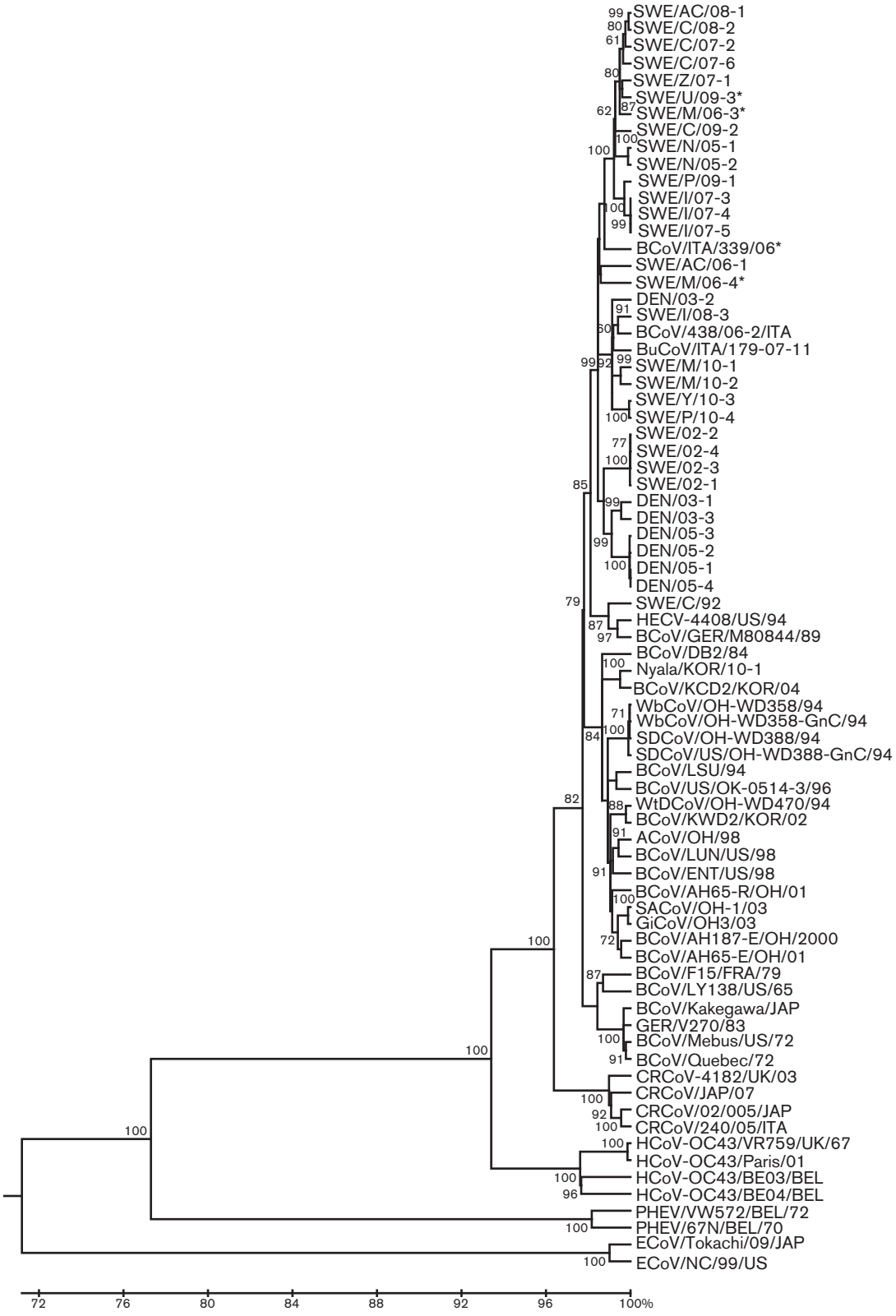

Fig. 1. Neighbour-joining tree based on the p distance of the complete nucleotide S sequences of virus species *Betacoronavirus 1* containing BCoV strains from Sweden (SWE) and Denmark (DEN) sequenced in this study. Bootstrap values above 70 % for 1000 iterations are shown at each branch. Strains marked with an asterisk are strains that were collected during the warm season.

Table 1. Mean estimations for the rate of evolution and TMRCA of the Swedish and Danish BCoV strains in comparison with the reference strains in species *Betacoronavirus 1*

Reference strains	BCoV strains	
	Mean rate of evolution substitutions per site year^{-1} ($\times 10^{-4}$)*	TMRCA (year)*
Human (HEC-4408)	8.3 (6.7–9.9)	1977 (1975–1980)
BCoV reference strains	8.7 (7.0–10.5)	1978 (1974–1981)
Wild ruminants	4.4 (3.2–5.7)	1963 (1954–1970)
Canine (CRCoV)	4.4 (3.2–5.5)	1951 (1939–1961)
Human (HCoV-OC43)	4.1 (3.2–4.7)	1899 (1884–1915)
Porcine (PHEV)	7.6 (6.0–9.3)	1847 (1815–1875)
Equine (ECoV)	7.9 (6.2–9.9)	1797 (1752–1844)

*95 % confidence interval (CI) values are in parentheses.

510, 523, 525, 543, 546, 573, 578, 590, 596, 718, 722, 888 and 1239 (Table 2).

Protein modelling comparisons

To determine if a homology model of the S protein for HECV-4408/US/94, SWE/C/92, DEN/03-3, SWE/M/10-1 and GER/V270/83 could be generated, each of these five sequences were searched individually against the Protein Data Bank (PDB) entries (http://www.rcsb.org/pdb/home/home.do) using default parameters. Based on the Z-score, all of these S protein sequences of BCoVs had the highest structural similarity to the crystal structure of murine hepatitis virus (PDB ID: 3R4D). Notably, the S1 sequences of the 33 BCoV strains contain a putative receptor-binding domain (amino acid residues 326–540, Fig. 2) with 94.8–97.6 % amino acid identities to sequences of BCoV/Mebus/US/72 and GER/V270/83. This part of the BCoV S proteins had the highest sequence similarity of the SARS-CoV receptor-binding domain-like superfamily (Scop ID: 143587), spanning amino acid residues 328–493 of the S protein of SARS-CoV; the so called C-domain (Wong *et al.*, 2004). Sialic acid is known to be the receptor for S protein binding in BCoV, although the receptor-binding domain is not well defined (Schultze *et al.*, 1991). The BCoV S protein also contains an NTD spanning amino acid residues 15–298, as recently defined in detail by Peng *et al.* (2012), with 92.9–95 % amino acid identities to sequences of BCoV/Mebus/US/72 and GER/V270/83.

Default parameters were used in server I-TASSER to predict structures of these proteins as explained in Methods. Results indicated that the NTD and putative C-domain of S1 were structurally similar for HECV-4408/US/94 and SWE/C/92 (Fig. 3a, b). This similarity is clearly illustrated when the two structures are aligned (Fig. 3c). In contrast, the predicted structures for SWE/M/10-1 and GER/V270/83 were substantially divergent while DEN/03-3 shows an intermediate conformation (Fig. 3d–f). Also in the S2 region HECV-4408/US/94 and SWE/C/92 differed in conformation compared with the other strains. The residues primarily predicted as potential receptor-binding sites based on homology with the S protein of SARS-CoV were used in the generation of structural models. Notably, parts of the putative receptor-binding domain and of the NTD were found to be in the strong positively selected regions on the surface of the S1 subunit (Fig. 3g, residues coloured green and red in SWE/C/92).

DISCUSSION

Circulation patterns of BCoV strains

This is the first evolution study to include full-length S gene sequences of BCoV strains obtained from European countries. The 26 Swedish and seven Danish BCoV strains sequenced in this study show low genetic diversity that result in their clustering as a unique clade in the phylogenetic tree (Fig. 1). We show based on the full-length S gene that there are no consistent differences between BCoV strains obtained from respiratory and enteric disease. This is in accordance with our previous study of partial S sequences (Bidokhti *et al.*, 2012). In two herds, identical sequences (e.g. SWE/02-1 and SWE/I/07-3) were found in different cattle sampled at the same occasion supporting previous findings that a herd disease outbreak is caused by a dominant strain (Bidokhti *et al.*, 2012; Liu *et al.*, 2006). However, in a large dairy herd (>200 cows) we found two slightly different (99.8 %) CD strains, SWE/N/05-1 and SWE/N/05-2, which were circulating at the same time. This finding indicates that strains with genetic diversity, though limited, can circulate in such herds. Large dairy herds were previously found to have a higher incidence of BCoV infection (Ohlson *et al.*, 2010; Smith *et al.*, 1998) which is consistent with the concept that large herds may foster a favourable environment for virus introduction and circulation of the strains.

A high similarity was observed between Italian and Swedish strains. We also identified a high similarity (99.4 %)

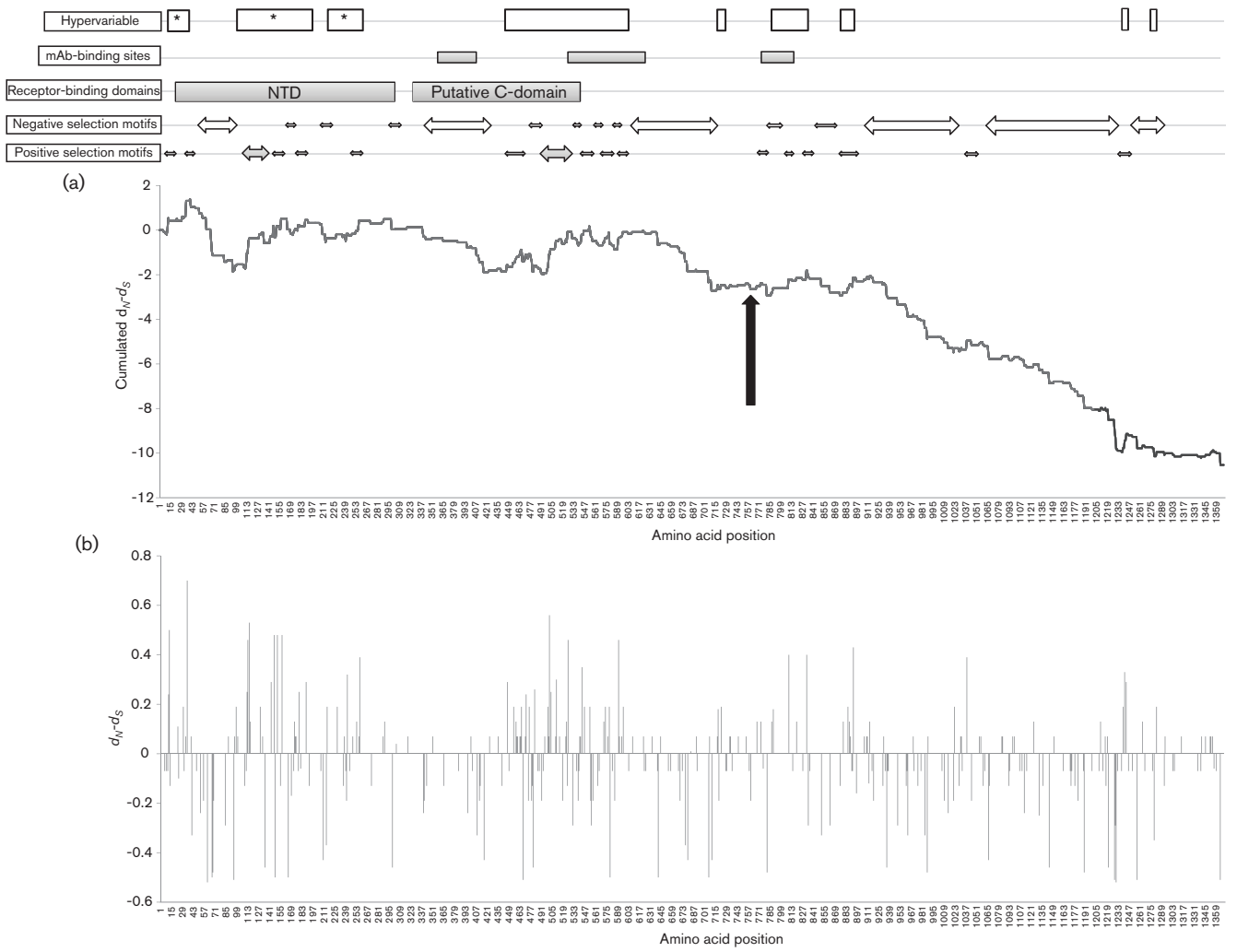

Fig. 2. The distribution of accumulated (a) and per codon (b) positive selection sites, identified using SNAP , along the S protein of the BCoV strains sequenced in this study. The two functionally distinct domains S1 and S2 are marked together with the cleavage site (vertical arrow, amino acid residues 768–769). The first upper line represents the hypervariable regions. The regions labelled with an asterisk were previously described (Bidokhti *et al.*, 2012) and the rest were found in the study; spanning amino acid residues 447–596, 718–722, 785–828, 875–888, 1235–1239 and 1275–1278. The second upper line represents the mAb-binding sites previously described for the S1 subunit (Yoo & Deregt, 2001) and for the S2 subunit (Vautherot *et al.*, 1992b) of BCoV, spanning amino acid residues 351–403, 517–621 and 769–798. The third upper line represents receptor-binding domains previously described: the NTD spanning amino acid residues 15–298 of BCoV (Peng *et al.*, 2012) and the C-domain spanning amino acid residues 318–510 of SARS-CoV (Wong *et al.*, 2004). The putative C-domain of the BCoV strains was predicted to span amino acid residues 326–540 using InterProScan. The last two lines represent the negative and positive selection motifs based on accumulated $d_N - d_S$. Thicker arrows show the strong selection motifs.

between the strain SWE/M/06-3 and six other strains that circulated in 2007–2009 in distant regions of Sweden, implying that certain strains may have the potential to spread directly or indirectly to distant regions or to other countries. No identical strains obtained from different epidemic seasons have been identified, but some strains were highly similar. High stability of certain BCoV strains was shown by the finding of identical strains in Gotland island in 2007 (e.g. SWE/I/07-3) and a highly similar strain obtained from another region in 2009 (SWE/P/09-1). Highly similar strains were also found in different regions in 2010 (SWE/Y/10-3, SWE/P/10-4). This suggests that these BCoV strains were part of common transmission chains. These data support previous findings that S gene sequences can provide data to clarify the transmission routes of BCoV strains (Bidokhti *et al.*, 2012; Kanno *et al.*, 2013).

Table 2. REL analysis results for the S protein of the BCoV sequence strains

No. of sequences	Mean $d_N - d_S$*	No. of positively selected sites	Posterior probability	Positively selected sites†
30‡	2.04	39	>90 %	**35**, **112, 113, 115**, 143, 147, 151, 157, 188, **257**, 447, 458, 471, 482, **499, 501, 503**, **510**, 523, 525, 543, 546, 573, 578, 590, 596, **718**, 722, 805, 828, 881, 883, 888, 1034, 1120, 1206, **1237**, 1239, 1278

*Datamonkey reports $d_N - d_S$ in place of d_N/d_S because d_S could be 0 for some sites.
†Positively selected sites identified with posterior probability P>95 % are in boldface. The underlined values are also reported by SNAP.
‡Three identical sequences were excluded from analysis.

Rate of evolution of BCoV strains

This evolutionary analysis encompassed a large dataset of species *Betacoronavirus 1* sequences of full-length S gene obtained over 45 years (1965–2010), including newly sequenced Swedish and Danish BCoV strains from the last decade and one strain from 1992. Sampling over time provides us with heterochronous data to calculate an evolutionary rate and to estimate the time of divergence of the recent BCoV sequences. The estimated rate of nucleotide substitution in the S gene of BCoV (8.7×10^{-4} substitution per site year^{-1}) is comparable to that observed as the standard range (orders of 10^{-3}–10^{-5}) in other rapidly evolving RNA viruses, such as the non-structural protein 2 (NSP2) of rotavirus A (Donker & Kirkwood, 2012) and the E gene of Dengue virus 3 (Sall *et al.*, 2010). The TMRCA estimate for BCoV strains in this study compared to published BCoV S gene sequences from other countries was 1978 (95 %CI: 1974–1981). This time period is even shorter than expected results reported previously (Vijgen *et al.*, 2006), showing a recent divergence during the last 60 years for BCoVs: 1944 (95 %CI: 1910–1963). This implies the strong ability of BCoV to adapt to cattle populations and spread over a large geographical region in a relatively short period of time.

Molecular clock analysis of the S gene of the recent BCoV strains and HCoV-OC43 strains estimated an evolutionary rate in the order of 4.1×10^{-4} substitutions per site year^{-1}, which is similar to a previous estimate of 4.7×10^{-4} substitutions per site year^{-1} (Vijgen *et al.*, 2005b). The Bayesian coalescent approach dated the TMRCA to around 1899, highly similar to the previous estimate of around 1890 (Vijgen *et al.*, 2005b). Evolutionary analysis of our BCoV strains along with other virus species in *Betacoronavirus 1* demonstrated a closer relationship of BCoV to canine and human coronaviruses than to porcine and equine coronaviruses. The TMRCA of coronaviruses is in accordance with their clustering in the phylogenetic tree (Fig. 1). The time of divergence of BCoV and CRCoV strains was estimated to have occurred five decades after that of BCoV and HCoV-OC43 strains, suggesting a closer common ancestor of the former. The S protein of CRCoV-4182/UK/03 has been shown to have a higher genetic similarity to BCoV/Mebus/US/72 and BCoV/LY138/US/65 than to HCoV-OC43/VR759/UK/67 (Erles *et al.*, 2007). In that study, the cross-reactivity of CRCoV-4182/UK/03 with polyclonal antisera against BCoV was also shown (Erles *et al.*, 2007). This corresponds to what is illustrated in the phylogenetic tree (Fig. 1); the clade of ruminant coronaviruses is clustered closer to the clade of CRCoV strains than to the other virus species in *Betacoronavirus 1*. At the tree level, coronaviruses from bovines and several wild ruminant species clustered closely together, implying that such interspecies transmission of coronaviruses may occur as suggested previously (Alekseev *et al.*, 2008).

In this study, we reported a close genetic relationship (98.9 % nucleotide identity, 98.6 % amino acid identity) and high simulated structural similarity of the S protein of HECV-4408/US/94 with a BCoV field strain, SWE/C/92. The infectivity of HECV-4408/US/94 for Gn calves and complete cross-protection against BCoV/DB2/84 isolate showing 98.2 % amino acid identity (98.6 % nucleotide identity) to HECV-4408/US/94 in the S protein has been experimentally confirmed (Han *et al.*, 2006). Thus, the similarity between SWE/C/92 and HECV-4408/US/94 S protein conformation further supports the hypothesis of possible interspecies transmission of these viruses. Future studies to find novel strains of *Betacoronavirus 1* and determination of the structure of the S protein would greatly assist in determining how such interspecies transmissions occur.

Positive selection on the S protein

The selection profiles identified two main patterns within the subunit domains S1 and S2 of the S protein. The S1 subunit is exposed on the surface of the viral particle, and is the target of neutralizing antibodies (Deregt & Babiuk, 1987; Yoo & Deregt, 2001; Yoo *et al.*, 1991b). The S1 subunit has two domains with a clear positive selection pattern (Fig. 2). Positively selected fragments of genes encoding viral proteins exposed on the surface of the capsid have been documented in other viruses, such as in porcine circovirus type 2 (Olvera *et al.*, 2007) and porcine parvovirus (Shangjin *et al.*, 2009). There is an association between positively selected sites along the S1 subunit identified in this study and mapped neutralizing epitopes. Epitopic fragments spanning amino acid residues 324–720

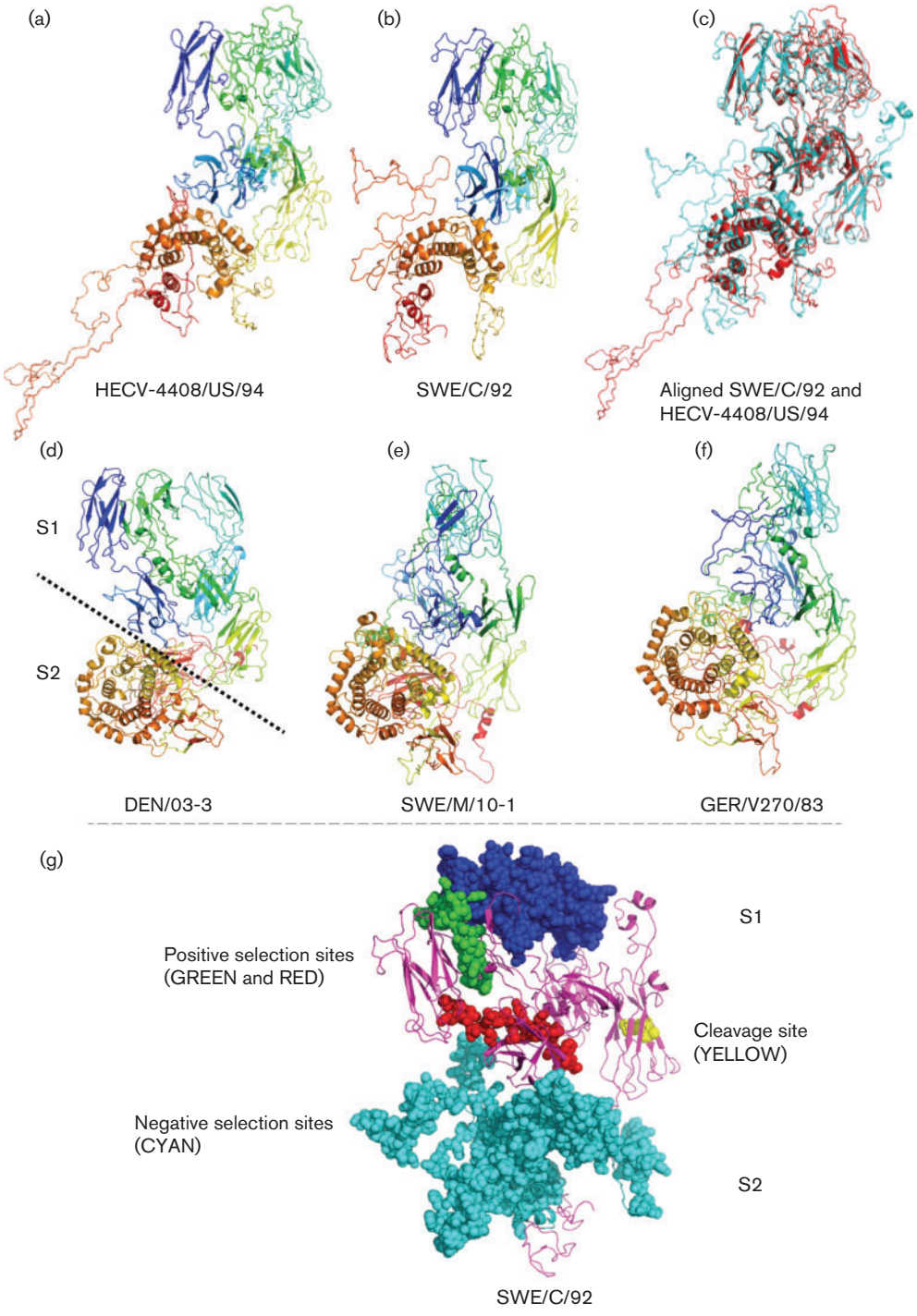

Fig. 3. Predicted 3D structures of S proteins belonging to several strains of coronaviruses including HECV-4408/US/94 (a), SWE/C/92 (b), DEN/03-3 (d), SWE/M/10-1 (e) and GER/V270/83 (f). In (c) the first two S proteins were aligned using MacPymole, HECV-4408/US/94 (red) and SWE/C/92 (cyan). In (g) the cleavage site of the S protein of SWE/C/92 is labelled yellow (amino acid residues 768–769), as well as regions of the S protein under positive selection (amino acid residues 109–131 in red and 495–527 in green). The regions (910–1032, 1059–1234 and 1245–1279) of the S2 subunit under negative selection are marked cyan. The putative receptor-binding domain (so called C-domain spanning amino acid residues 326–540) is coloured blue and green.

Table 3. BCoV strains utilized in this study

Strain/isolate name	Sampling year	Sample origin	Sample type	Country	Previous label name*	GenBank accession no.
SWE/C/92	1992	Adult cattle	Faecal	Sweden	C1-9202	JN795143†
SWE/02-1	2002	Calf	Nasal	Sweden	Nc1N-02a	DQ121634†
SWE/02-2	2002	Calf	Nasal	Sweden	Nc2N-02	DQ121635†
SWE/02-3	2002	Calf	Nasal	Sweden	Nc3N-02	DQ121637†
SWE/02-4	2002	Calf	Nasal	Sweden	Nc4N-02	DQ121638†
DEN/03-1	2003	Calf	Faecal	Denmark	Kc1F-03	DQ121631†
DEN/03-2	2003	Calf	Faecal	Denmark	Ac1F-03	DQ121619†
DEN/03-3	2003	Calf	Faecal	Denmark	Dc1F-03	DQ121622†
DEN/05-1	2005	Cattle	Faecal	Denmark		This study
DEN/05-2	2005	Cattle	Faecal	Denmark		This study
DEN/05-3	2005	Cattle	Nasal	Denmark		This study
DEN/05-4	2005	Cattle	Faecal	Denmark		This study
SWE/N/05-1	2005‡	Calf	Faecal	Sweden	N1-0511	JN795155†
SWE/N/05-2	2005‡	Calf	Faecal	Sweden		This study
SWE/AC/06-1	2006	Adult cattle	Faecal	Sweden	AC1-0611	JN795141†
SWE/M/06-3§	2006	Calf	Faecal	Sweden		This study
SWE/M/06-4§	2006	Calf	Faecal	Sweden	M2-0605	JN795154†
SWE/Z/07-1	2007	Adult cattle	Faecal	Sweden	Z2-0711	JN795163†
SWE/C/07-2	2007	Adult cattle	Faecal	Sweden	C4-0712	JN795146†
SWE/I/07-3	2007‖	Adult cattle	Faecal	Sweden	I3-0703	JN795151†
SWE/I/07-4	2007‖	Adult cattle	Faecal	Sweden		This study
SWE/I/07-5	2007	Adult cattle	Faecal	Sweden		This study
SWE/C/07-6	2007	Adult cattle	Faecal	Sweden	C3-0711	JN795145†
SWE/AC/08-1	2008	Adult cattle	Faecal	Sweden	Y1-0801	JN795161†
SWE/C/08-2	2008	Adult cattle	Faecal	Sweden	C5-0801	JN795147†
SWE/I/08-3	2008	Adult cattle	Faecal	Sweden	I4-0810	JN795152†
SWE/P/09-1	2009	Adult cattle	Nasal	Sweden	P1-0902	JN795159†
SWE/C/09-2	2009	Calf	Nasal	Sweden	C6-0903	JN795148†
SWE/U/09-3§	2009	Calf	Nasal	Sweden	U1-0907	JN795160†
SWE/M/10-1	2010	Calf	Faecal	Sweden		This study
SWE/M/10-2	2010	Calf	Faecal	Sweden		This study
SWE/Y/10-3	2010	Calf	Faecal	Sweden		This study
SWE/P/10-4	2010	Calf	Faecal	Sweden		This study
GER/V270/83	1983	–	–	Germany		EF193075
BCoV/GER/M80844/89	1989	Calf	Nasal	Germany		M80844.1
BCoV/ITA/339/06§	2006	Cattle	Faecal	Italy		EF445634
BuCoV/ITA/179-07-11	2007	Buffalo calf	Faecal	Italy		EU019216
WtDCoV/OH-WD470/94	1994	White-tailed deer	Faecal	USA, Ohio		FJ425187.1
BCoV/KWD2/KOR/02	2002	Cattle	Faecal	South Korea		AY935638.1
Nyala/KOR/10-1	2010	Nyala	Faecal	South Korea		HM573330.1
BCoV/KCD2/KOR/04	2004	Calf	Faecal	South Korea		DQ389633
BCoV/LSU/94	1994	Cattle	Nasal	USA, Louisiana		AF058943
BCoV/US/OK-0514-3/96	1996	Cattle	Nasal	USA, Louisiana		AF058944
WbCoV/OH-WD358/94	1994	Waterbuck	Faecal	USA, Ohio		FJ425186.1
SDCoV/US/OH-WD388-GnC/94	1994	Sambar deer	–	USA, Ohio		FJ425190.1
WbCoV/OH-WD358-GnC/94	1994	Waterbuck	Gn calf	USA, Ohio		FJ425185.1
SDCoV/OH-WD388/94	1994	Sambar deer	Faecal	USA, Ohio		FJ425189.1
SACoV/OH-1/03	2003	Sable antelope	Faecal	USA, Ohio		EF424621.1
BCoV/AH65-E/OH/01	2001	Feedlot calf	Faecal	USA, Ohio		EF424615.1
BCoV/AH65-R/OH/01	2001	Feedlot calf	Nasal	USA, Ohio		EF424617.1
BCoV/ENT/US/98	1998	Cattle	Faecal	USA, Texas		AF391541
GiCoV/OH3/03	2003	Giraffe	Faecal	USA, Ohio		EF424623.1
BCoV/AH187-E/OH/2000	2000	Feedlot calf	Faecal	USA, Ohio		EF424619.1
ACoV/OH/98	1998	Alpaca	Faecal	USA, Oregon		DQ915164.2
BCoV/LUN/US/98	1998	Cattle	Nasal	USA, Texas		AF391542

Table 3. cont.

Strain/isolate name	Sampling year	Sample origin	Sample type	Country	Previous label name*	GenBank accession no.
BCoV/DB2/84	1984	Cattle	–	USA, Maryland		DQ811784
BCoV/F15/FRA/79	1979	–	Faecal	France		D00731
BCoV/LY138/US/65	1965	Cattle	Faecal	USA, Utah		AF058942
BCoV/Mebus/US/72	1972	Cattle	–	USA		U00735
BCoV/Quebec/ 72	1972	Cattle	–	Canada		AF220295
HCoV-OC43/VR759 /UK/67	1967	Human	Nasal	England		AY391777
PHEv/vW572/BEL/72	1972	Pig	Tonsil	Belgium		DQ011855.1
PHEV/67N/BEL/70	1970	Piglet	–	Canada		AY078417
HECV-4408/US/94	1994	Human infant	Faecal	USA, Louisiana		L07748.1
BCoV/438/06-2/ITA	2006	Feedlot calf	Nasal	Italy		EU814647.1
BCoV/Kakegawa/JAP	1976	Cattle	Faecal	Japan		AB354579
HCoV-OC43/BE03/BEL	2003	Human infant	Nasal	Belgium		AY903454
HCoV-OC43/BE04/BEL	2004	Human infant	Nasal	Belgium		AY903455
HCoV-OC43/Paris/01	2001	Adult human	Nasal	France		AY585229
CRCoV/02/005/JAP	2002	Puppy	Nasal	Japan		AB242262.1
CRCoV/JAP/07	2007	Dog	Nasal	Japan		AB370269.1
CRCoV-4182/UK/03	2003	Puppy	Nasal	England		DQ682406
CRCoV/240/05/ITA	2005	Dog	Nasal	Italy		EU999954
ECoV/Tokachi/09/JAP	2009	Horse	Faecal	Japan		BAJ52885.1
ECoV/NC/99/US	1999	Foal	Faecal	USA, North Carolina		EF446615.1

*The label names of strains partially sequenced in our previous studies (Bidokhti *et al.*, 2012; Liu *et al.*, 2006) are designated here.
†Strains were partially sequenced previously and their fragments A and B are available in databases. Other fragments of these strains were sequenced in this study.
‡Samples were collected from the same farm in November 2005.
§Samples were collected during the warm season.
||Samples were collected from the same farm in March 2007.

of the S1 subunit of BCoV and the N-terminus of the S2 subunit spanning amino acid residues 769–798 have been previously recognized using mAbs (Vautherot *et al.*, 1992a; Yoo *et al.*, 1991b). A polymorphic region spanning amino acid residues 456–592 has also been shown by sequence analysis of BCoV strains (Rekik & Dea, 1994). It has been reported that mutations in the S1 and the N-terminus of the S2 sequence often result in changes in antigenicity (Kanno *et al.*, 2013; Vautherot *et al.*, 1992b; Yoo & Deregt, 2001). Likewise, parts of the putative receptor-binding domain defined in this study and the NTD defined in detail in a previous study (Peng *et al.*, 2012) were shown to be under strong positive selection in the BCoV strains. Taken together, the strong positively selected motifs among the S protein may thus be associated with the immune response and receptor binding and would thus be important in future BCoV vaccine development. The negative selection pattern of the S2 subunit is also reported (Fig. 2). Negative selection is usually reported in genome fragments with essential functions in the viral life cycle (Yang, 2005). For example, extensive syncytia formation was observed in cells infected with an S2 recombinant of BCoV (Yoo *et al.*, 1991a). The structure of the SARS-CoV S2 fusion protein core has been shown to provide a framework for the design of entry inhibitors that could be used in the therapeutic intervention against this virus (Supekar *et al.*, 2004). Thus, we speculate that the S2 subunit, except its N-terminus, would mostly interact with cellular compartments rather than immune system elements of the host.

Vaccination with an inactivated vaccine against BCoV has been used very restrictedly in Swedish cattle herds. Thus, we conclude that selective pressure sites observed in the receptor-binding subunit of the S protein gene of BCoV strains indicate a natural mode of evolution that is mainly due to exposure to the host immune system. Currently available vaccines are based on old enteric BCoV strains, genetically and antigenically different from currently circulating BCoV strains (Fulton *et al.*, 2013). Thus, continuous monitoring of sequence changes in positive selection sites may provide potentially useful data for identifying future dominant epidemic strains. This can then help to update the vaccine strains.

Studies are also warranted to detect the emergence of new genotypes and recombinants of BCoV as well as other betacoronaviruses and to assess their significance and potential in causing future epidemics. Nevertheless, it should be noted that the sequencing of a single gene may not be sufficient to define the genotypes of BCoV, as previously shown for human betacoronaviruses (Lau *et al.*,

2011; Woo *et al.*, 2006). Based on the lessons from HCoV-OC43 genotyping (Lau *et al.*, 2011) and recent evolutionary evaluation of the diverse genetic BCoV population through pioneering in-depth sequencing analysis (Borucki *et al.*, 2013), the deep sequencing of BCoV should therefore be performed to better understand the molecular epidemiology of BCoV, to determine genotypes and to reveal possible recombination events.

METHODS

Clinical samples. In total, 33 field samples, 25 faecal and 8 nasal, were sequenced from cattle in 29 herds (Table 3) from Sweden and Denmark. Sampled animals in all herds were showing clinical signs of BCoV infection. The samples were collected during outbreaks that occurred from 2002 to 2010. All seven Danish samples (one nasal and six faecal) were from 2003 and 2005. The oldest Swedish strain, which was from a WD outbreak in Uppland in 1992, was also sequenced. In this study, no cell-culture-passaged virus was utilized. Samples were kept frozen at −70 °C until analysed.

RNA extraction, cDNA synthesis, primer pairs and PCR. RNA extraction with TRIzol LS reagent (Invitrogen) and cDNA synthesis with random priming were performed as described previously (Liu *et al.*, 2006). In order to amplify and sequence the S gene (4092 nt), seven pairs of primers (Table 4) were used to generate a set of overlapping PCR products encompassing the entire S gene. Among these primers, six pairs (AF/AR, BF/BR, CF/CR, DF/DR, GF/GR, HF/HR) were already published (Hasoksuz *et al.*, 2002; Jeong *et al.*, 2005), while one pair (EF/ER) was designed by our group.

Amplification of the full-length S gene was performed in a DNA Thermal Cycler (Perkin-Elmer) using PfuUltra DNA polymerase (Stratagene). Briefly, 1 µl cDNA was amplified in a 50 µl reaction containing 5 µl 10 × PfuUltra buffer, 1 µl 10 mM dNTP, 1 µl AF and 1 µl HR primers (10 µM), 2.5U PfuUltra DNA polymerase and 40 µl double-distilled water. The cycling profiles consisted of 2 min of denaturation at 95 °C followed by 35 cycles of 95 °C for 30 s, 50 °C for 60 s, 72 °C for 4 min and a final extension step for 7 min at 72 °C.

In order to increase the sensitivity of the PCR detection method, nested and semi-nested PCR assays were developed as described previously (Bidokhti *et al.*, 2012). Briefly, the first PCR product (5 µl) was added to a tube with 45 µl PCR mixture, comprising 5 µl 10 × PCR buffer, 1 µl 10 mM dNTPs mixture, 5 µl 1 mg BSA ml^{-1}, 1.5 µl each primer (10 µM), 5 µl 25 mM $MgCl_2$, 1U *Taq* DNA polymerase (AmpliTaq; Perkin-Elmer) and 24 µl water. The thermocycling profile included 35 cycles of denaturation at 94 °C for 45 s, annealing at 50 °C for 60 s, extension at 72 °C for 3 min and a final extension at 72 °C for 7 min. For each strain, all seven fragments (A, B, C, D, E, G and H) were amplified by the corresponding primer pairs.

DNA sequencing and genome analysis. All seven PCR products of each strain were purified and sequenced in both directions using the same primers as for PCR and an ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) as described previously (Liu *et al.*, 2006). Capillary electrophoresis was performed in an ABI 3100 genetic analyser (Applied Biosystems). Sequence chromatograms were aligned and assembled into a final 4092 nt fragment of S gene, stretching from nucleotide positions 23 641 to 27 733 (amino acid residues 1–1363 of the S glycoprotein) of the BCoV strain *Mebus*.

Sequences were aligned with the CLUSTAL W program available in the BioEdit Sequence Alignment Editor (Hall, 1999). Phylogenetic tree construction was performed from the nucleotide sequences using a neighbour-joining algorithm with bootstrap values calculated from 1000 replicates in the program MEGA 5 (Tamura *et al.*, 2011). The prediction of the receptor-binding domain of the S protein was performed using InterProScan (Apweiler *et al.*, 2001). The prediction of potential *N*-glycosylation sites in the S proteins was performed using the CBS NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). Reference sequences of virus species of *Betacoronavirus 1* including BCoV, HCoV-OC43, PHEV, ECoV and BCoV-like coronaviruses in wild ruminants and also CRCoV were retrieved from GenBank and included in this analysis (Table 3).

Selective pressure analysis. To explore the potential overall differences in selective pressure on complete S gene sequences of the Swedish and Danish BCoV strains, we analysed the occurrences of synonymous (d_S) and non-synonymous (d_N) substitutions using SNAP (available at http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.

Table 4. S gene and reference of primer pairs used in this study

Primer name	Primer sequence (5′→3′)	Primer location	Primer reference
AF*	5′-ATG TTT TTG ATA CTT TTA ATT-3′	1–21	Hasoksuz *et al.* (2002)
AR†	5′-AGT ACC ACC TTC TTG ATA AA-3′	654–635	Hasoksuz *et al.* (2002)
BF	5′-ATG GCA TTG GGA TAC AG-3′	549–565	Hasoksuz *et al.* (2002)
BR	5′-TAA TGG AGA GGG CAC CGA CTT-3′	1039–1018	Hasoksuz *et al.* (2002)
CF	5′-GGG TTA CAC CTC TCA CTT CT-3′	782–801	Hasoksuz *et al.* (2002)
CR	5′-GCA GGA CAA GTG CCT ATA CC-3′	1550–1531	Hasoksuz *et al.* (2002)
DF	5′-GTC CGT GTA AAT TGG ATG GG-3′	1460–1479	Hasoksuz *et al.* (2002)
DR	5′-TGT AGA GTA ATC CAC ACA GT-3′	2286–2267	Hasoksuz *et al.* (2002)
EF	5′-GAA CCA GCA TTG CTA TTT CGG A-3′	2109–2131	This study
ER	5′-TTA TAA CTT TGC ACA CAA ATG AGG TC-3′	2876–2851	This study
GF	5′-CCC TGT ATT AGG TTG TTT AG-3′	2691–2710	Jeong *et al.* (2005)
GR	5′-ACC ACT ACC AGT GAA CAT CC-3′	3606–3587	Jeong *et al.* (2005)
HF	5′-GTG CAG AAT GCT CCA TAT GGT-3′	3439–3459	Jeong *et al.* (2005)
HR	5′-TTA GTC GTC ATG TGA TGT TT-3′	4092–4073	Jeong *et al.* (2005)

*F, Forward primer.

†R, Reverse primer.

html) (Korber, 2000), which plots the cumulative and per codon occurrence of each type of substitution from the start to the end of the S gene.

In order to examine the robustness of the positive selections identified by SNAP, we also analysed our datasets using HYPHY, a package accessed through the Datamonkey facility (http://www.datamonkey.org) (Pond & Frost, 2005). Datamonkey includes REL for detecting sites under selection. To detect positively selected sites, the default significant level of Bayes factor >50 was used for REL. A REL method is often the only method that can infer selection from small (5–15 sequences) or low divergence alignments and tends to be the most powerful test. This method was run using the GTR (general time reversible) substitution model on a neighbour-joining phylogenetic tree by the Datamonkey web server in order to investigate selective pressure along the S protein of BCoV strains sequenced in this study. Bootscan analysis was also used to detect possible recombination using the nucleotide alignment of the S gene sequences of virus species in *Betacoronaviruses 1* and also CRCoV. Bootscan analysis was performed using Simplot version 3.5.1 as described previously (Lau *et al.*, 2011; Woo *et al.*, 2006), with BCoV, HCoV, ECoV, PHEV and CRCoV strains as the query.

Evolutionary rate and estimation of divergence dates. Estimations of the rate of evolution and divergence times were calculated based on S gene sequence data using a Bayesian Markov chain Monte Carlo (MCMC) approach implemented in BEAST v.1.6.2 package (Drummond & Rambaut, 2007). Three independent runs of MCMC per dataset were performed under a strict molecular clock model, using the Hasegawa–Kishino–Yano model of sequence evolution with a proportion of invariant sites and gamma distributed rate heterogeneity (HKY+I+Γ) with partitions into codon positions, and the remaining default parameters in the prior's panel. For the S gene, the MCMC run was 3×10^7 steps long and the posterior probability distribution of the chains was sampled every 1000 steps. Convergence was assessed on the basis of an effective sampling size after a 10 % burn-in using Tracer software, version 1.5 (Rambaut & Drummond, 2007). The estimations are the mean values obtained for the three runs. The mean time of the most recent common ancestor (TMRCA) and the 95 % CI were calculated, and the best-fitting models were selected by a Bayes factor using marginal likelihoods implemented in Tracer (Suchard *et al.*, 2001).

***In silico* model analysis.** Based on strain sequence identity and phylogenetic analysis, the amino acid sequences of the S protein of five coronaviruses were chosen: HECV-4408/US/94 (the human isolate most closely related to BCoV), SWE/C/92 (the oldest Swedish strain clustered with HECV-4408/US/94), DEN/03-3 (the strain with highest identity to SWE/C/92), SWE/M/10-1 (the strain with lowest identity to SWE/C/92) and GER/V270/83 (a bovine reference isolate from Germany). Initially, a metathreading approach was applied in I-TASSER (Zhang, 2008; Zhang & Skolnick, 2004a), to identify templates for the subjected sequences in a non-redundant protein database structure library. From the generated consensus threading templates, the fragments of the sequences were assembled using modified replica-exchange Monte-Carlo simulations into three-dimensional models. In order to refine overall topology, models were clustered in SPICKER (Zhang & Skolnick, 2004b). A C-score was defined based on the quality of the threading alignments and the convergence of parameters of the structure assembly simulations. The structures were visualized and annotated in MacPyMol v1.3 (Schrödinger).

ACKNOWLEDGEMENTS

This work was supported by grant from The Swedish Farmers' Foundation for Agricultural Research (V0830402). The authors would like to thank Dr Lihong Liu, Mr Sven-Åke Bergkvist, Mrs Karin Ullman and Behdad Zarnegar for valuable assistance.

REFERENCES

Abraham, S., Kienzle, T. E., Lapps, W. & Brian, D. A. (1990). Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology* **176**, 296–301.

Al-Ahdal, M. N., Al-Qahtani, A. A. & Rubino, S. (2012). Coronavirus respiratory illness in Saudi Arabia. *J Infect Dev Ctries* **6**, 692–694.

Alekseev, K. P., Vlasova, A. N., Jung, K., Hasoksuz, M., Zhang, X., Halpin, R., Wang, S., Ghedin, E., Spiro, D. & Saif, L. J. (2008). Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. *J Virol* **82**, 12422–12431.

Alenius, S., Niskanen, R., Juntti, N. & Larsson, B. (1991). Bovine coronavirus as the causative agent of winter dysentery: serological evidence. *Acta Vet Scand* **32**, 163–170.

Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F. & other authors (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* **29**, 37–40.

Bidokhti, M. R., Tråvén, M., Ohlson, A., Baule, C., Hakhverdyan, M., Belák, S., Liu, L. & Alenius, S. (2012). Tracing the transmission of bovine coronavirus infections in cattle herds based on S gene diversity. *Vet J* **193**, 386–390.

Borucki, M. K., Allen, J. E., Chen-Harris, H., Zemla, A., Vanier, G., Mabery, S., Torres, C., Hullinger, P. & Slezak, T. (2013). The role of viral population diversity in adaptation of bovine coronavirus to new host environments. *PLoS ONE* **8**, e52752.

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

Cho, K. O., Hoet, A. E., Loerch, S. C., Wittum, T. E. & Saif, L. J. (2001). Evaluation of concurrent shedding of bovine coronavirus via the respiratory tract and enteric route in feedlot cattle. *Am J Vet Res* **62**, 1436–1441.

Clark, M. A. (1993). Bovine coronavirus. *Br Vet J* **149**, 51–70.

de Groot, R. J., Baker, S. C., Baric, R., Enjuanes, L., Gorbalenya, A. E., Holmes, K. V., Perlman, S., Poon, L. L., Rottier, P. J. M. & other authors (2012). *Coronaviridae*. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 806–828. Edited by A. M. King, E. Lefkowitz, M. J. Adams & E. B. Carstens. Oxford, UK: Elsevier Inc.

Decaro, N., Campolo, M., Desario, C., Cirone, F., D'Abramo, M., Lorusso, E., Greco, G., Mari, V., Colaianni, M. L. & other authors (2008a). Respiratory disease associated with bovine coronavirus infection in cattle herds in Southern Italy. *J Vet Diagn Invest* **20**, 28–32.

Decaro, N., Mari, V., Desario, C., Campolo, M., Elia, G., Martella, V., Greco, G., Cirone, F., Colaianni, M. L. & other authors (2008b). Severe outbreak of bovine coronavirus infection in dairy cattle during the warmer season. *Vet Microbiol* **126**, 30–39.

Decaro, N., Martella, V., Elia, G., Campolo, M., Mari, V., Desario, C., Lucente, M. S., Lorusso, A., Greco, G. & other authors (2008c). Biological and genetic analysis of a bovine-like coronavirus isolated from water buffalo (*Bubalus bubalis*) calves. *Virology* **370**, 213–222.

Deregt, D. & Babiuk, L. A. (1987). Monoclonal antibodies to bovine coronavirus: characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology* **161**, 410–420.

Donker, N. C. & Kirkwood, C. D. (2012). Selection and evolutionary analysis in the nonstructural protein NSP2 of rotavirus A. *Infect Genet Evol* **12**, 1355–1361.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Erles, K., Shiu, K. B. & Brownlie, J. (2007). Isolation and sequence analysis of canine respiratory coronavirus. *Virus Res* **124**, 78–87.

Fulton, R. W., Step, D. L., Wahrmund, J., Burge, L. J., Payton, M. E., Cook, B. J., Burken, D., Richards, C. J. & Confer, A. W. (2011). Bovine coronavirus (BCV) infections in transported commingled beef cattle and sole-source ranch calves. *Can J Vet Res* **75**, 191–199.

Fulton, R. W., Ridpath, J. F. & Burge, L. J. (2013). Bovine coronaviruses from the respiratory tract: antigenic and genetic diversity. *Vaccine* **31**, 886–892.

Groneberg, D. A., Zhang, L., Welte, T., Zabel, P. & Chung, K. F. (2003). Severe acute respiratory syndrome: global initiatives for disease diagnosis. *QJM* **96**, 845–852.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Han, M. G., Cheon, D. S., Zhang, X. & Saif, L. J. (2006). Cross-protection against a human enteric coronavirus and a virulent bovine enteric coronavirus in gnotobiotic calves. *J Virol* **80**, 12350–12356.

Hasoksuz, M., Sreevatsan, S., Cho, K. O., Hoet, A. E. & Saif, L. J. (2002). Molecular analysis of the S1 subunit of the spike glycoprotein of respiratory and enteric bovine coronavirus isolates. *Virus Res* **84**, 101–109.

Jeong, J. H., Kim, G. Y., Yoon, S. S., Park, S. J., Kim, Y. J., Sung, C. M., Shin, S. S., Lee, B. J., Kang, M. I. & other authors (2005). Molecular analysis of S gene of spike glycoprotein of winter dysentery bovine coronavirus circulated in Korea during 2002–2003. *Virus Res* **108**, 207–212.

Kanno, T., Ishihara, R., Hatama, S. & Uchida, I. (2013). Antigenic variation among recent Japanese isolates of bovine coronaviruses belonging to phylogenetically distinct genetic groups. *Arch Virol* **158**, 1047–1053.

Korber, B. (2000). HIV signature and sequence variation analysis. In *Computational Analysis of HIV Molecular Sequences*, pp. 55–72. Edited by A. G. Rodrigo & G. H. Learn. Dordrecht, Netherlands: Kluwer Academic Publishers.

Lathrop, S. L., Wittum, T. E., Loerch, S. C., Perino, L. J. & Saif, L. J. (2000). Antibody titers against bovine coronavirus and shedding of the virus via the respiratory tract in feedlot cattle. *Am J Vet Res* **61**, 1057–1061.

Lau, S. K., Lee, P., Tsang, A. K., Yip, C. C., Tse, H., Lee, R. A., So, L. Y., Lau, Y. L., Chan, K. H. & other authors (2011). Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol* **85**, 11325–11337.

Li, F. (2012). Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol* **86**, 2856–2858.

Liu, L., Hägglund, S., Hakhverdyan, M., Alenius, S., Larsen, L. E. & Belák, S. (2006). Molecular epidemiology of bovine coronavirus on the basis of comparative analyses of the S gene. *J Clin Microbiol* **44**, 957–960.

Martínez, N., Brandão, P. E., de Souza, S. P., Barrera, M., Santana, N., de Arce, H. D. & Pérez, L. J. (2012). Molecular and phylogenetic analysis of bovine coronavirus based on the spike glycoprotein gene. *Infect Genet Evol* **12**, 1870–1878.

Mebus, C. A., Stair, E. L., Rhodes, M. B. & Twiehaus, M. J. (1973). Neonatal calf diarrhea: propagation, attenuation, and characteristics of a coronavirus-like agent. *Am J Vet Res* **34**, 145–150.

Ohlson, A., Heuer, C., Lockhart, C., Tråvén, M., Emanuelson, U. & Alenius, S. (2010). Risk factors for seropositivity to bovine coronavirus and bovine respiratory syncytial virus in dairy herds. *Vet Rec* **167**, 201–206.

Olvera, A., Cortey, M. & Segalés, J. (2007). Molecular evolution of porcine circovirus type 2 genomes: phylogeny and clonality. *Virology* **357**, 175–185.

Park, S. J., Jeong, C., Yoon, S. S., Choy, H. E., Saif, L. J., Park, S. H., Kim, Y. J., Jeong, J. H., Park, S. I. & other authors (2006). Detection and characterization of bovine coronaviruses in fecal specimens of adult cattle with diarrhea during the warmer seasons. *J Clin Microbiol* **44**, 3178–3188.

Parker, M. D., Yoo, D., Cox, G. J. & Babiuk, L. A. (1990). Primary structure of the S peplomer gene of bovine coronavirus and surface expression in insect cells. *J Gen Virol* **71**, 263–270.

Paton, D. J., Christiansen, K. H., Alenius, S., Cranwell, M. P., Pritchard, G. C. & Drew, T. W. (1998). Prevalence of antibodies to bovine virus diarrhoea virus and other viruses in bulk tank milk in England and Wales. *Vet Rec* **142**, 385–391.

Peng, G., Xu, L., Lin, Y. L., Chen, L., Pasquarella, J. R., Holmes, K. V. & Li, F. (2012). Crystal structure of bovine coronavirus spike protein lectin domain. *J Biol Chem* **287**, 41931–41938.

Pond, S. L. K. & Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533.

Rambaut, A. & Drummond, A. J. (2007). Tracer v1.4: MCMC trace analyses tool. Available: http://beast.bio.ed.ac.uk/Tracer. Accessed 20 June 2008.

Rekik, M. R. & Dea, S. (1994). Comparative sequence analysis of a polymorphic region of the spike glycoprotein S1 subunit of enteric bovine coronavirus isolates. *Arch Virol* **135**, 319–331.

Saif, L. J. (2010). Bovine respiratory coronavirus. *Vet Clin North Am Food Anim Pract* **26**, 349–364.

Saif, L. J., Redman, D. R., Brock, K. V., Kohler, E. M. & Heckert, R. A. (1988). Winter dysentery in adult dairy cattle: detection of coronavirus in the faeces. *Vet Rec* **123**, 300–301.

Sall, A. A., Faye, O., Diallo, M., Firth, C., Kitchen, A. & Holmes, E. C. (2010). Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. *J Virol* **84**, 765–772.

Schultze, B., Gross, H. J., Brossmer, R. & Herrler, G. (1991). The S protein of bovine coronavirus is a hemagglutinin recognizing 9-*O*-acetylated sialic acid as a receptor determinant. *J Virol* **65**, 6232–6237.

Shangjin, C., Cortey, M. & Segalés, J. (2009). Phylogeny and evolution of the NS1 and VP1/VP2 gene sequences from porcine parvovirus. *Virus Res* **140**, 209–215.

Smith, D. R., Fedorka-Cray, P. J., Mohan, R., Brock, K. V., Wittum, T. E., Morley, P. S., Hoblet, K. H. & Saif, L. J. (1998). Epidemiologic herd-level assessment of causative agents and risk factors for winter dysentery in dairy cattle. *Am J Vet Res* **59**, 994–1001.

Storz, J., Rott, R. & Kaluza, G. (1981). Enhancement of plaque formation and cell fusion of an enteropathogenic coronavirus by trypsin treatment. *Infect Immun* **31**, 1214–1222.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* **18**, 1001–1013.

Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A. & Carfi, A. (2004). Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc Natl Acad Sci U S A* **101**, 17958–17963.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA 5: molecular evolutionary genetics analysis using

maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.

Tråvén, M., Näslund, K., Linde, N., Linde, B., Silván, A., Fossum, C., Hedlund, K. O. & Larsson, B. (2001). Experimental reproduction of winter dysentery in lactating cows using BCV – comparison with BCV infection in milk-fed calves. *Vet Microbiol* **81**, 127–151.

Tsunemitsu, H., el-Kanawati, Z. R., Smith, D. R., Reed, H. H. & Saif, L. J. (1995). Isolation of coronaviruses antigenically indistinguishable from bovine coronavirus from wild ruminants with diarrhea. *J Clin Microbiol* **33**, 3264–3269.

Vautherot, J. F., Laporte, J. & Boireau, P. (1992a). Bovine coronavirus spike glycoprotein: localization of an immunodominant region at the amino-terminal end of S2. *J Gen Virol* **73**, 3289–3294.

Vautherot, J. F., Madelaine, M. F., Boireau, P. & Laporte, J. (1992b). Bovine coronavirus peplomer glycoproteins: detailed antigenic analyses of S1, S2 and HE. *J Gen Virol* **73**, 1725–1737.

Vijgen, L., Keyaerts, E., Lemey, P., Moës, E., Li, S., Vandamme, A. M. & Van Ranst, M. (2005a). Circulation of genetically distinct contemporary human coronavirus OC43 strains. *Virology* **337**, 85–92.

Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A. M. & Van Ranst, M. (2005b). Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* **79**, 1595–1604.

Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M. & Van Ranst, M. (2006). Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* **80**, 7270–7274.

Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. (2004). A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* **279**, 3197–3201.

Woo, P. C., Lau, S. K., Yip, C. C., Huang, Y., Tsoi, H. W., Chan, K. H. & Yuen, K. Y. (2006). Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J Virol* **80**, 7136–7145.

Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., Bai, R., Teng, J. L., Tsang, C. C. & other authors (2012). Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *J Virol* **86**, 3995–4008.

Yang, Z. (2005). The power of phylogenetic comparison in revealing protein function. *Proc Natl Acad Sci U S A* **102**, 3179–3180.

Yoo, D. & Deregt, D. (2001). A single amino acid change within antigenic domain II of the spike protein of bovine coronavirus confers resistance to virus neutralization. *Clin Diagn Lab Immunol* **8**, 297–302.

Yoo, D. W., Parker, M. D. & Babiuk, L. A. (1991a). The S2 subunit of the spike glycoprotein of bovine coronavirus mediates membrane fusion in insect cells. *Virology* **180**, 395–399.

Yoo, D. W., Parker, M. D., Song, J., Cox, G. J., Deregt, D. & Babiuk, L. A. (1991b). Structural analysis of the conformational domains involved in neutralization of bovine coronavirus using deletion mutants of the spike glycoprotein S1 subunit expressed by recombinant baculoviruses. *Virology* **183**, 91–98.

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814–1820.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.

Zhang, Y. & Skolnick, J. (2004a). Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc Natl Acad Sci U S A* **101**, 7594–7599.

Zhang, Y. & Skolnick, J. (2004b). SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem* **25**, 865–871.

Zhang, X. M., Herbst, W., Kousoulas, K. G. & Storz, J. (1994). Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child. *J Med Virol* **44**, 152–161.

Zhong, N. S. & Wong, G. W. (2004). Epidemiology of severe acute respiratory syndrome (SARS): adults and children. *Paediatr Respir Rev* **5**, 270–274.