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

Mehdi R. M. Bidokhti,1 Madeleine Trävén,1 Neel K. Krishna,2 Muhammad Munir,3,4 Sándor Belák,3,4 Stefan Alenius1 and Martí Cortey5

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

1Division of Ruminant Medicine and Veterinary Epidemiology, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Box 7054, SE-750 07 Uppsala, Sweden
2Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, 700 West Olney Road, Norfolk, VA 23507-1696, USA
3Department of Virology, Immunobiology and Parasitology, National Veterinary Institute, Box 7028, SE-751 89 Uppsala, Sweden
4Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden
5Department 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.

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 Alphacoronavirus and subfamily Betacoronavirus. 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).

Received 30 April 2013
Accepted 25 June 2013

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, 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; Trävén 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 RRSSR, 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
KRSSR motif, was observed in the S protein of SWE/Y/10-
3 and SWE/P/10-4. The A-to-E amino acid change in the
769 position, leading to a KRSRRE 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, H1234,
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/M8044/89
and human isolate HECV-4408/US/94: 8.7 \times 10^{-4}
and 8.3 \times 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 was observed. Similar results were obtained when
HECV 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 (dN)
and the synonymous substitution rate (dS) (i.e. dN − dS)
revealed that amino acid residues 109–131 and 495–527
of the S1 subunit were under strong positive selection (Fig.
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

<table>
<thead>
<tr>
<th>Reference strains</th>
<th>BCoV strains</th>
<th>TMRCA (year)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean rate of evolution substitutions per site year(^{-1}) (\times 10^{-4})*</td>
<td></td>
</tr>
<tr>
<td>Human (HEC-4408)</td>
<td>8.3 (6.7–9.9)</td>
<td>1977 (1975–1980)</td>
</tr>
<tr>
<td>Canine (CRCoV)</td>
<td>4.4 (3.2–5.5)</td>
<td>1951 (1939–1961)</td>
</tr>
<tr>
<td>Human (HCoV-OC43)</td>
<td>4.1 (3.2–4.7)</td>
<td>1899 (1884–1915)</td>
</tr>
<tr>
<td>Porcine (PHEV)</td>
<td>7.6 (6.0–9.3)</td>
<td>1847 (1815–1875)</td>
</tr>
<tr>
<td>Equine (ECoV)</td>
<td>7.9 (6.2–9.9)</td>
<td>1797 (1752–1844)</td>
</tr>
</tbody>
</table>

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

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%)
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).
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 (Shangin 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

<table>
<thead>
<tr>
<th>Strain/isolate name</th>
<th>Sampling year</th>
<th>Sample origin</th>
<th>Sample type</th>
<th>Country</th>
<th>Previous label name*</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWE/C/92</td>
<td>1992</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>C1-9202</td>
<td>JN795143†</td>
</tr>
<tr>
<td>SWE/02-1</td>
<td>2002</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>Nc1N-02a</td>
<td>DQ121634†</td>
</tr>
<tr>
<td>SWE/02-2</td>
<td>2002</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>Nc2N-02</td>
<td>DQ121635†</td>
</tr>
<tr>
<td>SWE/02-3</td>
<td>2002</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>Nc3N-02</td>
<td>DQ121637†</td>
</tr>
<tr>
<td>SWE/02-4</td>
<td>2002</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>Nc4N-02</td>
<td>DQ121638†</td>
</tr>
<tr>
<td>DEN/03-1</td>
<td>2003</td>
<td>Calf</td>
<td>Faecal</td>
<td>Denmark</td>
<td>Kg1F-03</td>
<td>DQ121631†</td>
</tr>
<tr>
<td>DEN/03-2</td>
<td>2003</td>
<td>Calf</td>
<td>Faecal</td>
<td>Denmark</td>
<td>Ac1F-03</td>
<td>DQ121619†</td>
</tr>
<tr>
<td>DEN/03-3</td>
<td>2003</td>
<td>Calf</td>
<td>Faecal</td>
<td>Denmark</td>
<td>Dc1F-03</td>
<td>DQ121622†</td>
</tr>
<tr>
<td>DEN/05-1</td>
<td>2005</td>
<td>Cattle</td>
<td>Faecal</td>
<td>Denmark</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>DEN/05-2</td>
<td>2005</td>
<td>Cattle</td>
<td>Faecal</td>
<td>Denmark</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>DEN/05-3</td>
<td>2005</td>
<td>Cattle</td>
<td>Nasal</td>
<td>Denmark</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>DEN/05-4</td>
<td>2005</td>
<td>Cattle</td>
<td>Faecal</td>
<td>Denmark</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/N/05-1</td>
<td>2005†</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>N1-0511</td>
<td>JN795155†</td>
</tr>
<tr>
<td>SWE/N/05-2</td>
<td>2005†</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/02-1</td>
<td>2006</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>AC1-0611</td>
<td>JN795141†</td>
</tr>
<tr>
<td>SWE/M/06-3†</td>
<td>2006</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/M/06-4†</td>
<td>2006</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>M2-0605</td>
<td>JN795154†</td>
</tr>
<tr>
<td>SWE/Z/07-1</td>
<td>2007</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>Z2-0711</td>
<td>JN795163†</td>
</tr>
<tr>
<td>SWE/C/07-2</td>
<td>2007</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>C4-0712</td>
<td>JN795146†</td>
</tr>
<tr>
<td>SWE/I/07-3</td>
<td>2007</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>I3-0703</td>
<td>JN795151†</td>
</tr>
<tr>
<td>SWE/I/07-4</td>
<td>2007†</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/I/07-5</td>
<td>2007</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/C/07-6</td>
<td>2007†</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>C3-0711</td>
<td>JN795145†</td>
</tr>
<tr>
<td>SWE/AC/08-1</td>
<td>2008</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>Y1-0801</td>
<td>JN795161†</td>
</tr>
<tr>
<td>SWE/C/08-2</td>
<td>2008</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>C5-0801</td>
<td>JN795147†</td>
</tr>
<tr>
<td>SWE/I/08-3</td>
<td>2008</td>
<td>Adult cattle</td>
<td>Faecal</td>
<td>Sweden</td>
<td>I4-0810</td>
<td>JN795152†</td>
</tr>
<tr>
<td>SWE/P/09-1</td>
<td>2009</td>
<td>Adult cattle</td>
<td>Nasal</td>
<td>Sweden</td>
<td>P1-0902</td>
<td>JN795159†</td>
</tr>
<tr>
<td>SWE/C/09-2</td>
<td>2009†</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>C6-0903</td>
<td>JN795148†</td>
</tr>
<tr>
<td>SWE/U/09-3§</td>
<td>2009</td>
<td>Calf</td>
<td>Nasal</td>
<td>Sweden</td>
<td>U1-0907</td>
<td>JN795160†</td>
</tr>
<tr>
<td>SWE/M/10-1</td>
<td>2010</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/M/10-2</td>
<td>2010</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/Y/10-3</td>
<td>2010</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>SWE/P/10-4</td>
<td>2010</td>
<td>Calf</td>
<td>Faecal</td>
<td>Sweden</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>GER/V270/83</td>
<td>1983</td>
<td>–</td>
<td>–</td>
<td>Germany</td>
<td>–</td>
<td>EF193075</td>
</tr>
<tr>
<td>BCoV/GER/M80844/89</td>
<td>1989</td>
<td>Calf</td>
<td>Faecal</td>
<td>Germany</td>
<td>–</td>
<td>M80844.1</td>
</tr>
<tr>
<td>BCoV/ITA/339/06§</td>
<td>2006</td>
<td>Cattle</td>
<td>Faecal</td>
<td>Italy</td>
<td>–</td>
<td>EF445634</td>
</tr>
<tr>
<td>BuCoV/ITA/179-07-11</td>
<td>2007</td>
<td>Buffalo calf</td>
<td>Faecal</td>
<td>Italy</td>
<td>–</td>
<td>EU019216</td>
</tr>
<tr>
<td>WtDCoV/OH-WD470/94</td>
<td>1994</td>
<td>White-tailed deer</td>
<td>Faecal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>FJ425187.1</td>
</tr>
<tr>
<td>BCoV/KWD2/KOR/02</td>
<td>2002</td>
<td>Cattle</td>
<td>Faecal</td>
<td>South Korea</td>
<td>–</td>
<td>AY935638.1</td>
</tr>
<tr>
<td>Nyla/KOR/10-1</td>
<td>2010</td>
<td>Nyla</td>
<td>Faecal</td>
<td>South Korea</td>
<td>–</td>
<td>HM573330.1</td>
</tr>
<tr>
<td>BCoV/KCD2/KOR/04</td>
<td>2004</td>
<td>Calf</td>
<td>Faecal</td>
<td>South Korea</td>
<td>–</td>
<td>DQ389633</td>
</tr>
<tr>
<td>BCoV/LSU/94</td>
<td>1994</td>
<td>Cattle</td>
<td>Nasal</td>
<td>USA, Louisiana</td>
<td>–</td>
<td>AF058943</td>
</tr>
<tr>
<td>BCoV/US/OK-0514-3/96</td>
<td>1996</td>
<td>Cattle</td>
<td>Nasal</td>
<td>USA, Louisiana</td>
<td>–</td>
<td>AF058944</td>
</tr>
<tr>
<td>WbCoV/OH-WD358/94</td>
<td>1994</td>
<td>Waterbuck</td>
<td>Faecal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>FJ425186.1</td>
</tr>
<tr>
<td>SDGCoV/OH-WD388-GnC/94</td>
<td>1994</td>
<td>Sambar deer</td>
<td>–</td>
<td>USA, Ohio</td>
<td>–</td>
<td>FJ425190.1</td>
</tr>
<tr>
<td>WbCoV/OH-WD358-GnC/94</td>
<td>1994</td>
<td>Waterbuck</td>
<td>Gn calf</td>
<td>USA, Ohio</td>
<td>–</td>
<td>FJ425185.1</td>
</tr>
<tr>
<td>SDGCoV/OH-WD388/94</td>
<td>1994</td>
<td>Sambar deer</td>
<td>Faecal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>FJ425189.1</td>
</tr>
<tr>
<td>SACoV/OH-1/0-3</td>
<td>2003</td>
<td>Sable antelope</td>
<td>Faecal</td>
<td>USA, Texas</td>
<td>–</td>
<td>EF424621.1</td>
</tr>
<tr>
<td>BCoV/AH65-E/OH/01</td>
<td>2001</td>
<td>Feedlot calf</td>
<td>Faecal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>EF424615.1</td>
</tr>
<tr>
<td>BCoV/AH65-R/OH/01</td>
<td>2001</td>
<td>Feedlot calf</td>
<td>Nasal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>EF424617.1</td>
</tr>
<tr>
<td>BCoV/ENT/US/98</td>
<td>1998</td>
<td>Cattle</td>
<td>Faecal</td>
<td>USA, Texas</td>
<td>–</td>
<td>AF391541</td>
</tr>
<tr>
<td>GiCoV/OH3/03</td>
<td>2003</td>
<td>Giraffe</td>
<td>Faecal</td>
<td>USA, Ohio</td>
<td>–</td>
<td>EF424623.1</td>
</tr>
<tr>
<td>ACoV/OH/98</td>
<td>1998</td>
<td>Alpaca</td>
<td>Faecal</td>
<td>USA, Oregon</td>
<td>–</td>
<td>DQ915164.2</td>
</tr>
<tr>
<td>BCoV/LUN/US/98</td>
<td>1998</td>
<td>Cattle</td>
<td>Nasal</td>
<td>USA, Texas</td>
<td>–</td>
<td>AF391542</td>
</tr>
</tbody>
</table>
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., 1991b). 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 restrictively 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., 2013).
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.

METHODOLOGY

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 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 (dS) and non-synonymous (dN) 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

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

*F, Forward primer.
†R, Reverse primer.
Evolution of Betacoronavirus 1

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 + G) with partitions into codon positions, and the remaining default parameters in the prior’s panel. For the S gene, the MCMC run was 3 × 10^9 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 estimates 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). 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. Li Hong Liu, Mr. Sven-Åke Bergkvist, Mrs. Karin Ullman and Behdad Zarnegar for valuable assistance.

REFERENCES


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


