Nucleotide and Predicted Amino Acid Sequences of All Genes Encoded by the 3′ Genomic Portion (9.5 kb) of Respiratory Bovine Coronaviruses and Comparisons Among Respiratory and Enteric Coronaviruses

VLADIMIR N. CHOULJENKO, KONSTANTIN G. KOUSOULAS,* XIAOQING LIN, & JOHANNES STORZ

Department of Veterinary Microbiology and Parasitology, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 70803

Abstract. The 3′-ends of the genomes (9538 bp) of two wild-type respiratory bovine coronavirus (RBCV) isolates LSU and OK were obtained by cDNA sequencing. In addition, the 3′-end of the genome (9545) of the wild-type enteric bovine coronavirus (EBCV) strain LY-138 was assembled from available sequences and by cDNA sequencing of unknown genomic regions. Comparative analyses of RBCV and EBCV nucleotide and deduced amino acid sequences revealed that RBCV-specific nucleotide and amino acid differences were disproportionally concentrated within the S gene and the genomic region between the S and E genes. Comparisons among virulent and avirulent BCV strains revealed that virulence-specific nucleotide and amino acid changes were located within the S and E genes, and the 32 kDa open reading frame.

Key words: cDNA, S, non-structural, structural, respiratory, coronavirus

Introduction

Coronaviruses are important etiological agents of human and animal diseases including respiratory infection, gastroenteritis, hepatic and neurological disorders as well as immune-mediated disease such as feline infectious peritonitis, and other persistent infections (1,2). Enteric bovine coronaviruses (EBCV) are generally associated with enteric disease of newborn calves and winter dysentery of adult cattle (2). Recently, numerous respiratory bovine coronaviruses (RBCV) were isolated in our laboratory from cattle arriving with fever and respiratory disease in feedlots or livestock shows of 8 different states in the USA. The cytopathogenic, cell fusion, and other phenotypic properties of these viruses were different from the known EBCV (3).

Coronaviruses contain a single stranded, capped, and polyadenylated positive-sense (infectious) RNA molecule of approximately 30 kb length, which directs the synthesis of a nested set of subgenomic mRNAs (4,5). The 3′-end of the genomic RNA consists of approximately 9.5 kb and contains the spike (S) glycoprotein, the hemagglutinin-esterase (HE) glycoprotein, the integral membrane (M) protein, the small membrane protein (E) and the phosphorylated nucleocapsid (N) protein and a number of ORFs potentially encoding non-structural proteins (N_s) (5). The 32 kDa non-structural protein is a phosphoprotein that accumulates in the cytoplasm of infected cells (6,7). It is not known whether the 12.7 and 4.8 kDa ORFs are expressed in infected cells, while the 4.9 kDa putative protein, most likely, is not translated (8).

BCV uses N-acetyl-9-O acetyl neuraminic acid as receptor determinant to initiate infection (9). Although the HE glycoprotein also has an affinity for 9-O-acetylated sialic acid, the S glycoprotein was identified as the major sialic acid binding protein of BCV (10). The S glycoprotein facilitates viral

attachment to susceptible cells, causes cell fusion after cell-surface expression (fusion from within), and induces viral infectivity neutralizing antibodies (1). Porcine transmissible gastroenteritis virus (TGEV) strains were isolated which exhibited respiratory tissue tropism. These viruses contained point mutations or deletions within the first 250 aa of TGEV S1 which were associated with reduced enteropathogenicity and loss of hemagglutinating activity (11–13).

To examine the genetic basis for the phenotypic differences between RBCV and EBCV, we cloned and sequenced the 3′-end of the viral genomes of two virus strains RBCV-LSU-94LSS-051-2(LSU) and RBCV-OK-0514-3(OK) that originated from Louisiana and Oklahoma cattle, respectively. We report here, the nucleotide and predicted amino acid sequences of all genes encoded by the 3′ genomic portion (9.5 kb) of two wild-type RBCV strains and comparisons among respiratory and enteric coronaviruses.

Materials and Methods

Viruses and cell line.   All RBCV and EBCV strains were propagated in the G clone of human rectal tumor cells (HRT-18G) developed recently through selection and medium modulation (3). Supernatant fluids from infected HRT-18G cells were collected and viruses were purified as described (14). RBCV OK and LSU virus stocks were tested at the third and fourth passages, respectively. EBCV LY-138(LY) virus stocks were prepared at the second passage, while the EBCV-L9 virus strain, derived from the EBCV-Mebus strain, had been propagated 80 times in cell cultures.

Strategy for cDNA construction and assembly of the 9.5 Kb cDNA sequence representing the 3′-end of different BCV strains.   TRI Reagent from the Molecular Research Center, Inc. (Cincinnati, OH, USA), was used for total RNA extraction. Ready-To-Go You-Prime First-Strand Beads from Pharmacia Biotech Inc. (Uppsala, Sweden) were used for cDNA library construction. All amplifications were performed using the Gene-Amp PCR system 9600 (Perkin-Elmer, Norwalk, CT, USA) with PCR reagents and AmpliTaq from Perkin-Elmer. The TA-cloning kit from Invitrogen Inc. (San Diego, CA) was used for cloning of RT-PCR products. Restriction

enzymes were obtained from New England Biolabs (Beverly, MA, USA).

The 3′ genomic end of BCV Mebus strain consisting of 8695 nucleotides was assembled from available sequences deposited in Genbank. The accession numbers of the cDNA sequences used to assemble the EBCV Mebus genome were M31053 for S and HE genes, M31054 for the 4.9, 4.8, 12.7 and 9.5 kDa (E) ORFs, and M16620 for the M, N genes and I ORF. The assembled Mebus genomic sequence did not contain the 32 kDa ORF. An 850 nt sequence containing the 32 kDa ORF of BCV-Quebec (accession number X15445) was used for comparisons with other BCV. The S and HE cDNA sequences specified by LY-138 were previously reported (15,16).

A series of overlapping cDNA clones representing the entire 3′-end of two RBCV isolates and unpublished sequences of LY-138 were constructed. Two cDNA libraries were produced, a library was made using the BCV3′ primer representing the 3′ terminus of the genomic RNA, and a second cDNA library was produced using an oligonucleotide (3B11) to prime cDNAs starting at nucleotide 6345 (counting from the 3′-end of the viral genome) (Fig. 1). The entire 9545 nt sequence representing the 3′-end of the BCV genome was divided into six overlapping cDNA regions. Each cDNA was amplified by PCR using specific primer pairs. Primer pair 5F6/BCV 3′ amplified a cDNA fragment containing the M and N genes. Primer pair 5F5/3B3 amplified a cDNA fragment containing the 3′-end of S, 4.9 kDa, 4.8 kDa, 12.7 kDa, E, M and the 5′-end of N genes. Primer pair B5′/B3′ amplified the 3′-end of the spike gene, primer pair 5F24/A3′ amplified a cDNA that coded for the carboxy-terminal portion of the S1 subunit, primer pair A5′/3B11 amplified a cDNA fragment that coded for the amino-terminus of S, and primer pair 5F16/3B10 amplified the 32 kDa and HE genes. B3′ and A5′ primers were designed to contain an extra *Bam*HI and *Eco*RI sites for cloning purposes, while a BstXI site was naturally present in the 5F24 primer. The actual primer sequences are:
BCV3′:5′GTGATTCTTCCAATTGGCC;
B5′:5′TCTTCGACAGCTGCAACC;
A3′:5′GGTTGCAGCTGTCGTGAAAGA;
B3′:5′AAC<u>GGATCC</u>AATATATCGTCAGGAGCC;
A5′-5′:GCT<u>GAATTC</u>GATAATGGTACTAGGCTG;
5F16-5′:AGAATGGCAGTTGCTTATG;
5F24-5′:TCT-<u>CC</u>ATTAAATT<u>GGG</u>AACGTAAG;
3B10-5′:GATTATGGTCTAAGCATCATG;
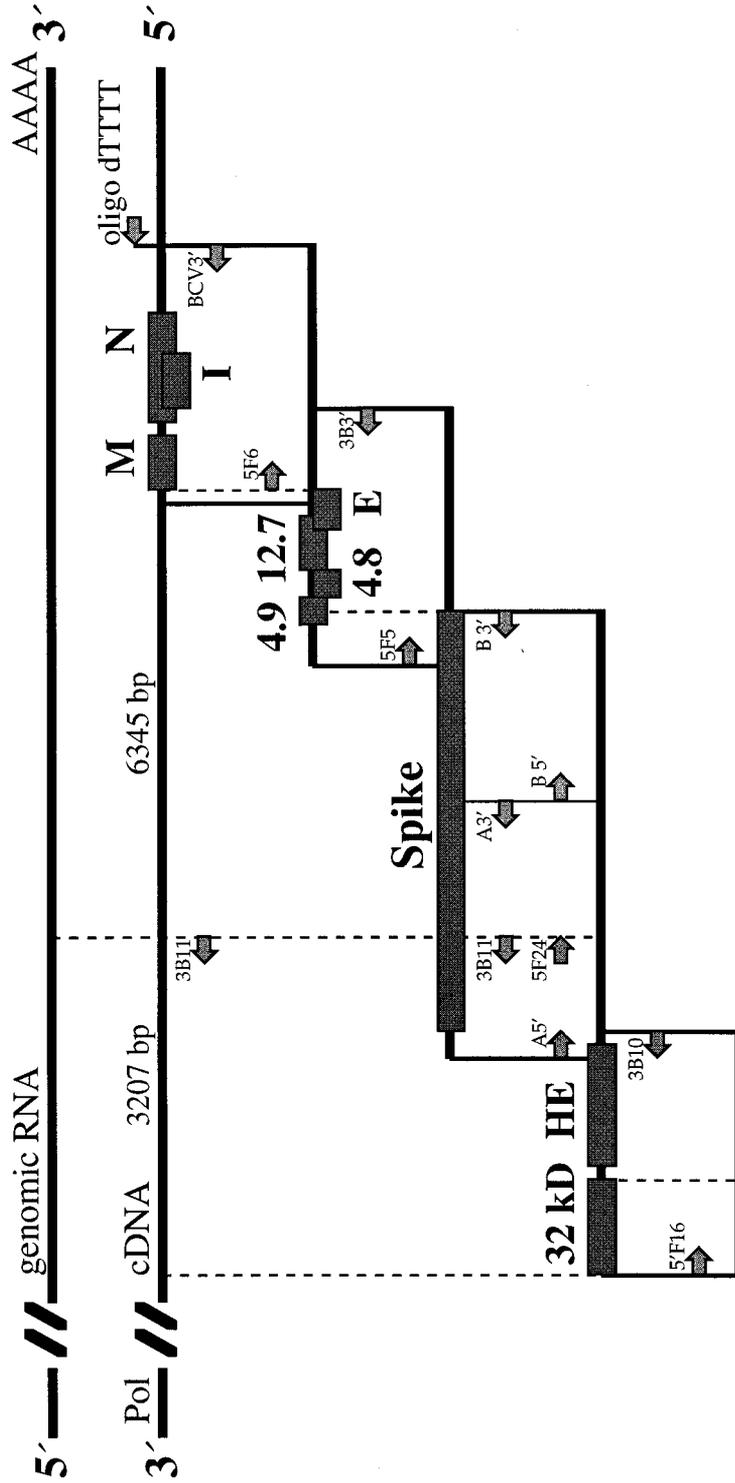
Fig. 1. **BCV genome arrangement and cDNA cloning and sequencing strategy.** Top line depicts the + stranded RNA genome of BCV with a truncation marking (//) separating the 5′ portion of the genome coding for the RNA-dependent RNA polymerase (Pol) from the remaining BCV genes. The second line diagrams two cDNA libraries produced by reverse transcription of the BCV genome using oligonucleotide primers BCV3′ and 3B11. Coding sequences for each gene are indicated as filled rectangular boxes. The approximate location of oligonucleotide primers used for RT-PCR and cloning of individual cDNAs are shown as arrows pointing towards the direction of cDNA synthesis.

3B11-5′:GGTCTTACGTTCCCAATTTAATGG;
5F5-5′:GGTATAGCCCCTAAGAGTG;
5F6-5′:GTTTAATAGAGGTAGGCAG;
3B3-5′:GGTAGCTGAGAAGTAGCAG.

DNA sequencing and analyses. DNA sequencing
was carried out with the modified dideoxynucleotide
chain termination procedure (17) using Sequenase[TM]
from Amersham Life Science (Arlington Heights, IL).
All sequencing data were based on sequences
obtained from at least two independently cloned
cDNA fragments. DNA sequences were analyzed
with the aid of the Sequence Analysis Software
Package of the Genetics Computer Group of the
University of Wisconsin and the MacVector Software
(IBI, New Haven, CT). Sequence data obtained from
all overlapping cDNA clones were compiled using the
program AssemblLine of the MacVector[TM] software
package running on a Power Macintosh Computer.
Consensus sequences of 9538 bp were produced for
LSU and OK strains, and 9545 bp for the LY-138
strain (Genbank accession numbers AF058942,
AF058943, AF058944 respectively). Analysis of
the cDNA sequence data using the MacVector
software package revealed 10 major ORFs: ORF1
(32 kDa, 15–851), ORF2 (HE, 863–2137), ORF3 (S,
2152–6243), ORF4 (4.9 kDa, 6233–6364), ORF5
(4.8 kDa, 6400–6537), ORF6 (12.7 kDa, 6617–
6946), ORF7 (E, 6933–7187), ORF8 (M, 7202–
7894), ORF9 (N, 7904–9250), and ORF10 (I, 7965–
8588) for LSU and OK strains. The corresponding
locations of ORFs starting from 4.9 kDa ORF were
different for LY-138, because of a codon insertion in
the 4.9 kDa ORF of LY-138 and the four nucleotide
deletion in the intergenic region between 4.8 and

12.7 kDa ORFs of RBCV. The locations for these
ORFs were: ORF4 (4.9 kDa, 6233–6367), ORF5
(4.8 kDa, 6403–6540), ORF6 (12.7 kDa, 6624–
6953), ORF7 (E, 6940–7194), ORF8 (M, 7209–
7901), ORF9 (N, 7911–9257), and ORF10 (I, 7972–
8595).

Results

*Overall comparisons of genes and predicted proteins
specified by RBCV (LSU, OK) and EBCV (LY, Mebus)
strains.* To establish the close evolutionary relation-
ship between LSU and OK strains and to ascertain
RBCV-specific amino acid changes (conserved in
LSU and OK but different in other strains), a pairwise
comparison of nucleotide and amino acid differences
among BCV strains for all ORFs, except for the ORF
coding for the RNA-dependent-RNA-polymerase and
the 32 kDa protein was performed (Table 1). In
general, the nucleotide and amino acid sequences of
RBCV strains LSU and OK were more conserved to
each other than to EBCV strains LY-138 and Mebus,
and they were more divergent to the Mebus strain than
to the LY-138. Specifically, the amino acid sequence
of M specified by LSU and OK were identical, while
they were different by one, and two aa from that of
Mebus and LY-138, respectively. The S glycoprotein
specified by LSU differed by only 4 amino acids from
that of OK, while S glycoproteins of LSU and OK
differed by 22 and 33 amino acids in comparison with
the LY and Mebus S sequences, respectively.
Furthermore, LSU and OK sequences of the N and I
ORF (located within N) were more conserved to each
other than to any other strain compared. Most amino

Table 1. The nucleotide and amino acid changes for each ORF of different BCV strains

Strain	Compared to Strain	HE 1272 (424)*	S 4089 (1363)	4.9 kDA 129 (43)	4.8 kDa 135 (45)	12.7 kDa 327 (45)	E 252 (84)	M 690 (230)	N 1344 (448)	I-ORF 621 (207)
OK	LSU	11 (4)[#]	29 (4)	2 (2)	5 (4)	3 (2)	0	1 (0)	4 (2)	3 (2)
LY	LSU	23 (5)	68 (22)	7 (4)	8 (6)	4 (1)	1 (0)	6 (2)	16 (4)	9 (6)
	OK	20 (5)	64 (22)	7 (4)	8 (5)	5 (3)	1 (0)	7 (2)	12 (4)	6 (4)
Mebus	LSU	24 (6)	70 (33)	3 (3)	8 (5)	5 (1)	2 (1)	4 (1)	17 (5)	10 (8)
	OK	23 (6)	66 (33)	3 (3)	7 (4)	6 (3)	2 (1)	5 (1)	13 (3)	7 (6)
	LY	13 (3)	57 (26)	4 (1)	5 (2)	1 (0)	1 (1)	2 (1)	9 (3)	5 (4)

* indicate the total number of nucleotides and amino acids for each gene.
# comparisons show the total number of nucleotide and amino acid substitutions (in parenthesis) for each ORF.

acid changes within HE, 4.9, 4.8, 12.7 kDa ORFs, E, N and the I ORF were strain-specific. HE and M contained one RBCV-specific aa change, and N and I ORF contained two RBCV-specific aa changes each.

RBCV-specific amino acid substitutions within S. The S1 subunit contained most of the RBCV-specific aa substitutions and included an amino acid change within the signal sequence as well as two clusters of amino acid substitutions within the amino-terminus and the hypervariable region (Fig. 2). The proteolytic cleavage site that separates S1 and S2 subunits was conserved among RBCV, LY-138 and Mebus strains. In contrast to the high number of RBCV-specific substitutions within S1, S2 contained only two RBCV-specific amino acid changes, an Ala 769 to Ser change immediately adjacent to the proteolytic cleavage site and an Asp 1026 to Gly located within the heptad repeat sequence. The RBCV-G95 strain was isolated from a nasal sample of a calf that had diarrhea and signs of respiratory distress (18). The nucleotide and predicted primary structures of S and HE glycoproteins specified by RBCV-G95 were reported previously (19,20). LSU and OK had ten unique aa substitutions within S in comparison to all other BCV strains, while G95, LSU, and OK shared only three aa substitutions at aa 100, aa 465, and aa 1026 (Fig. 2).

RBCV-specific nucleotide and amino acid substitutions within the 4.9, 4.8, and 12.7 kDa ORFs. The human coronavirus strain OC43 (HCV-OC43) lacks two ORFs which potentially encode two non-structural proteins of 4.9 and 4.8 kDa (21). Furthermore, the same genomic areas are deleted in three Hemagglutinating Encephalomyelitis Virus (HEV) strains of swine (22). The fact that respiratory HCV-OC43 and EBCV strains show remarkable genomic and protein similarities as well as immunological cross-reactivities, prompted us to compare the nucleotide sequences specified by the genomic region between the S and the 12.7 kDa ORF of EBCV, RBCV, HCV-OC43, and three porcine HEV strains (Fig. 3). LSU and OK contained a deletion (TAAT) in comparison with Mebus as well as LY-138. This deletion was located 31–34 bp upstream from the AUG of the 12.7 kDa ORF in the genomic region between the 4.8 kDa and 12.7 kDa ORFs. The nucleotide sequences downstream (3′) to this deletion

were highly conserved among all strains. The 12.7 kDa ORF of LSU and OK contained the aa substitution of Lys 5 to Arg, while the aa substitutions of Ile 8 to Thr, and Glu 21 to Asp were unique for OK.

Comparison of the Mebus and L9 strains against virulent (LSU, OK, LY, F15) strains. To examine the evolutionary relationship of virulent strains to the avirulent, cell adapted strains, and to delineate specific amino acids which may cause loss of virulence, a comparison was performed between L9, Mebus and all other strains. The E protein which is encoded by the 9.5 kDa ORF contained one aa substitution (Gly 53 to Val) which was conserved for all virulent strains (F15, LY-138, OK, LSU) in comparison to the avirulent Mebus and L9 strains. Two other RBCV strains from Texas and Arizona cattle also contained the same aa substitution (data not shown). Four aa changes in S1 at aa 33, aa 40, aa 278 and aa 470 and three aa substitutions in S2 at aa 965, aa 1241 and aa 1341 were conserved in all virulent strains in comparison to Mebus and L9 (Fig. 2) as well as Norden Vaccine, and EBCV-Quebec strains (data not shown). The 32 kDa ORF codes for a non-structural phosphoprotein that accumulates in the cytoplasm of infected cells (6,7). The EBCV-Quebec sequence of the 32 kDa ORF was used for comparison, because the 32 kDa nucleotide sequence of the Mebus strain was not available. DNA sequencing of three independently derived cDNA clones for each strain revealed the presence of a single nucleotide (A) insertion at position 757 (aa 253) specified by both LSU and OK altering the translational reading frame for the next 9 aa until a double nucleotide insertion (AA) at nucleotides 784 and 785 restored the original frame. LY-138 specified a similar 9 aa peptide which contained one aa change, Ser to Phe at aa 254 in comparison to RBCV. In contrast, the corresponding genomic region of the avirulent EBCV-Quebec strain specified an 8 aa sequence which was different than the 9 aa sequence specified by virulent EBCV and RBCV strains. Two additional amino acid changes, Ser (192) to Ile and Pro (aa 273) to Ser (274), located at the carboxy terminal portion of the 32 kDa ORF, were specific for all virulent strains (Fig. 4). Identical aa changes were also found in the 32 kDa protein of two more RBCV strains isolated from Texas and Arizona cattle (data not shown).

	1	11	13	33	36	38	40	88	100	113	115	118	169	173	179	181	248	253	256	331	362	440	447
1. F15	M	M	L	V	I	T	T	T	I	I	H	M	N	H	R	V	M	S	M	C	S	L*	I
2. LY	M	M	L	V	V*	T	T	T	I	I	N	M	H	H	R	I*	M	S	M	C	S	V	T
3. G95	M	M	F	V	I	R*	T	R	T	T*	H	M	N	H	R	V	M	S	M	G*	S	V	T
4. LSU	M	T	F	V	I	T	T	T	T	I	D	K	N	N	Q	V	M	S	L*	C	S	V	T
5. OK	M	T	F	V	I	T	T	T	T	I	D	K	N	N	Q	V	M	S	M	C	S	V	I
6. L9	M	M	F	A	I	T	I	R	I	I	K	M	H	H	K	V	L	S	M	C	C*	V	T
7. Mebus	M	M	F	A	I	T	I	R	I	I	K	M	H	H	K	V	L	N*	M	C	S	V	T

	458	465	470	484	499	501	509	510	531	543	571	578	580	591	608	648	651	707	742	744	767	768	769
1. F15	S	V	D	S	N	P	N	S	N	S	Y	T	Q	S	D	A*	T	I	A*	V	V*	R	A
2. LY	S	V	D	S	S	S	N	S	N	S	Y	T	Q	S	D	V	T	I	S	V	R	R	A
3. G95	S	A	D	S	T	P	T*	S	D	S	H*	T	Q	A*	G*	V	T	I	S	A	R	R	S
4. LSU	S	A	D	S	S	S	N	T	G	A	Y	S	Q	S	D	V	T	F*	S	A	R	R	S
5. OK	S	A	D	S	S	S	N	T	G	A	Y	S	R*	S	D	V	T	I	S	A	R	R	A
6. L9	S	V	H	S	N	P	N	S	D	S	Y	T	Q	S	D	V	P*	S	S	A	R	R	A
7. Mebus	P*	V	H	T*	N	P	N	S	N	S	Y	T	Q	S	D	V	T	I	S	A	R	R	A

	778	817	834	906	909	927	936	965	984	988	1026	1085	1095	1100	1223	1241	1269	1286	1316	1341	1363
1. F15	N	T	C*	E*	K	A	N	E	W	A	D	R*	F*	A	V	P	T	L*	F	K	D
2. LY	N	I*	Y	D	K	A	N	E	W	A	D	A	L	A	V	P	T	V	L*	K	D
3. G95	N	T	Y	D	R*	S*	Y*	E	W	A	G	A	L	A	V	P	N*	V	F	K	D
4. LSU	N	T	Y	D	K	A	N	E	W	A	G	A	L	A	V	P	T	V	F	K	D
5. OK	N	T	Y	D	K	A	N	E	W	A	G	A	L	A	V	P	T	V	F	K	D
6. L9	N	T	Y	D	K	A	N	V	W	A	D	A	A	A	A*	H	T	V	F	I	D
7. Mebus	T*	T	Y	D	K	A	N	V	L*	V*	D	A	A	V*	V	H	T	V	F	I	D

Discussion

Genetic comparisons among different BCVs revealed substantial differences between RBCV and EBCV strains principally within the S gene and within ORFs located between the S and E genes. Furthermore, genetic differences between virulent and avirulent strains were identified within the S gene, the E gene and the 32 kDa ORF. The salient features of genetic differences between RBCV and EBCV strains are discussed below:

RBCV-specific genetic alterations in the S gene. A pairwise alignment of TGEV and MHV S aa sequences revealed that the N-terminal portion of S1 which is deleted in the porcine respiratory coronaviruses (PRCV) and HCV-229E, in comparison with TGEV, is the region corresponding to the MHV receptor binding-site (aa 1–330) (23). The TGEV receptor-binding site is in a different location (aa 500–700) and aligns with the S polymorphic region of the MHV strain. Recently, it was shown that only two aa changes at the N-terminus of TGEV S resulted in the loss of enteric tropism (13). The S1 amino terminus specified by RBCV strains LSU and OK contained aa changes at aa11, aa115, aa118, aa173 and aa179 which may affect S1-mediated receptor binding. Hemagglutination of chicken red blood cells (RBC) was shown to be mediated by the S glycoprotein, because purified S of the EBCV Mebus strain agglutinated chicken RBC, while purified HE did not (10,24). RBCV strains LSU and OK agglutinated mouse and rat, but not adult chicken RBC (3). Therefore, aa changes within S specified by RBCV may be responsible for the inability to hemagglutinate chicken RBC.

The S1A virus neutralizing (VN) immunoreactive epitope (aa 351–403) (25) was identical for all viral strains, except for a single aa change at aa 362 specified by the avirulent, cell culture-adapted strain EBCV-L9 (Fig. 2). Furthermore, the S1A epitopes of HCV-OC43 and the BCV Mebus strain were identical (21). Monoclonal antibodies (MABS) against the EBCV Mebus cross-reacted with different animal and human coronaviruses (25). Therefore, it is likely that these antibodies react with the S1A epitope. The hypervariable region of the S glycoprotein contains the S1B immunoreactive epitope which is the target for virus neutralizing Mabs (25). Four RBCV-specific aa substitutions at aa 510, aa 531, aa 543 and aa 578 were located within or proximal to this epitope. Based on the observed aa changes, it can be predicted that Mabs specific for this region may be able to distinguish between respiratory, enteric and vaccine BCV strains.

The BCV S2 subunit of S induced cell fusion when it was expressed in insect cells, indicating that S2 contained membrane fusion domains (26). The hydrophobic and heptad repeat regions of S2 are believed to form the coiled-coil structure of the oligomeric S protein that have been associated with fusion activity. Specifically, three aa changes within a predicted heptad region of the MHV S2 subunit were shown to be responsible for pH-dependent cell fusion (27). RBCV strains LSU and OK are highly fusogenic in cell culture (data not shown). Additional experimentation is required to assess whether the aa change of Ala 769 to Ser immediately after the proteolytic cleavage site and the aa change of Asp 1026 to Gly within the heptad repeat are responsible for the extensive cell fusion induced by RBCV.

RBCV-specific genetic alterations between the S and E genes. The RBCV genomic regions between the S and E genes contained many nucleotide substitutions, deletions and insertions. HCV-OC43 and three porcine HEV strains specified deletions within the 4.9 and 4.8 kDa ORFs, indicating that they are not essential for virus replication. Similarly, the high number of mutations within the RBCV 4.8 and 4.9 kDa ORFs suggests that these ORFs are not essential for virus replication in cell culture (Table 1). The 65 nt leader of a cloned BCV defective interfering (DI) RNA when mapped by mutations, could be converted rapidly to the wild-type leader of a helper virus following DI RNA transfection into helper

Fig. 2. **Comparison of the predicted amino acid sequences of the BCV S glycoprotein specified by different strains.** Amino acids that are different in at least one strain are shown, except aa 1, aa 768 and aa 1363 which are included as reference points. * indicates unique amino acid changes for each strain. Boxed amino acids are common among different strains. Light-gray boxes contain RBCV-specific, dark boxes contain virulent-specific, and clear boxes contain EBCV-specific aa changes. aa 1–17 is the putative signal peptide; aa 351–403 is the S1A immunoreactive domain; aa 517–621 is the S1B immunoreactive domain; aa 452–593 is the hypervariable region; aa 955–992 is the hydrophobic region; aa 993–1032 is the heptad repeat sequence; aa 1312–1325 is the carboxy-terminal anchor sequence.

```
      1   4.9 kDa  Carboxy-end of Spike                                    70
Mebus  ATGACGACTA AGTTCGTCTT TGATTTATTG GCTCCTGACG ATATATTACA TCCCTTCAAT CATGTGAAGC
   LY  ATGACGACTA AGTTCGTCTT TGATTTATTG GCTCCTGACG ATATATTACA TCCCTTCAAT CATGTGAAGC
  LSU  ATGACGACTA AGTTCGTCTT TGATTTATTG GCTCCTGACG ATATATTACA TCCCTcCAAT CATGTGAAtC
   OK  ATGACGACTA AGTTCGTCTT TGATTTATTG GCTCCTGACG ATATATTACA TCCCTcCAAT CATGTGAAtC
HCVOC43 peptide 11  ATGACGACTA AGTTCGTCTT TGATTcATTG cactgatctc tTgTtagA.. .......... ..........
HEVNT9 peptide 24   ATGACGACTA AGTTCGTCTT TGtTTTAcTG aCTCtTGACc ATATAgTA.. .......... ..........
HEVVW572 peptide 24 ATGACGAttA AGTTCGTCTT TGATTTAcTG aCTCtTGACG ATATAgTA.. .......... ..........
HEV67N peptide 20   ATGACGACTA AtTTCGTCTT TGgTTT.... ........Cc ATATAgTA.. .......... ..........


      71                                                                    140
Mebus  TA...ATTAT AAGACCCATT GAGGTCGAGC ATATTATAAT AGCTACCACA ATGCCTGCTG TTTAGTGGGT
   LY  TAattATTAT AAGACCCATa GAGGTCGAGC ATATTATAAT AGCTACCACA ATGCCTGCTG TTTAGTGGGT
  LSU  TA...ATTAT AAGACCCATT tAGGTCGAGC ATATTATAAT AGCTACCACA ATGCCTGCTG TTTAGTGGGT
   OK  TA...ATTAT AAGACtCATT GAGGTCGAGC ATATTATAAT AGCTACCACA ATGCCTGCTG TTTAGTGGGT
HCVOC43 peptide 11  .......... .......... .......... .......... .......... .......... ..........
HEVNT9 peptide 24   .......... .......... .......... .......... .......... .......... ..........
HEVVW572 peptide 24 .......... .......... .......... .......... .......... .......... ..........
HEV67N peptide 20   .......... .......... .......... .......... .......... .......... ..........


      141                                          4.8 kDa                  210
Mebus  ACTGTGTCTT ATATAACTAG TAAACCTGTA ATGCCAATGG CTACAACCAT TGACGGTACA GATTATACTA
   LY  ACTGTGTCTT ATATAACTAG TAAACCTGTA ATGCCAATGG CTACAACCAT TGAtGGTACA GATTATACTA
  LSU  ACTGTGTCTT ATATAACTAG TAAACCTGcA ATGCCAATGG CTACAACCAT TGAaGGTgCA GATTATACTA
   OK  ACTGTGTCTT ATATAACTAG TAAACCTGTA ATGCCAATGG CTACAACCAT TGAtGGTACA GATTATACTA
HCVOC43 peptide 11  .......... .......... .......... .......... .......... .......... ..........
HEVNT9 peptide 24   .......... .......... .......... .......... .......... .......... ..........
HEVVW572 peptide 24 .......... .......... .......... .......... .......... .......... ..........
HEV67N peptide 20   .......... .......... .......... .......... .......... .......... ..........


      211                                                                   280
Mebus  ATATTATGCC TAGTACTGTT TCTACAACAG TTTATTTAGG CTGTTCTATA GGTATTGACA CTAGCACCAC
   LY  AcATTATGCC TAGTACTGTT TCTACAACAG TTTATTTAGG CgGTTCTATA GGTATTGAtA CTAGCACCAC
  LSU  AcATTATGCC TAtTACTGTT ctTACAACAG TTTATTTAGG CgtTTCTATA GGTATTGACA CTAGCACCAC
   OK  ATATTATGCC TAtTACTGTT TtTACAACAG TTTATTTAGG CgtTTtTATA GGTATTGACA CTAGCACCAC
HCVOC43 peptide 11  .......... .......... .......... .......... .......... .......... ..........
HEVNT9 peptide 24   .......... .......... .......... .......... .......... .......... ..........
HEVVW572 peptide 24 .......... .......... .......... .......... .......... .......... ..........
HEV67N peptide 20   .......... .......... .......... .......... .......... .......... ..........


      281                                                                   350
Mebus  TGGTTTTACC TGTTTTTCAC GGTACTAGTT CCAAACCATA TTATAATTTA GGTAGACCTT ATAACTTTAA
   LY  TGGTTTTACC TGTTTTTCAt GGTACTAGTT CCAAACCATA TTATAATTcA GGTAGACCTT ATAACTTTAA
  LSU  TGGTTTTACC TGTTTTTCAC GGTACTAGTT CCAAACCATA TTATAATTTA GGTAGACCTT ATAACTTTAA
   OK  TGGTTTTACC TGTTTTTCAC GGTACTAGTT CCAAACCATA TTATAATTcA GGTAGACCTT ATAACTTTAA
HCVOC43 peptide 11  .......... .......... .......... .......... .......... .....tCtTT tTgcaaTctA
HEVNT9 peptide 24   .......... .......... .......... .......... .......... .....ACtTT A........A
HEVVW572 peptide 24 .......... .......... .......... .......... .......... .....ACtTT A........A
HEV67N peptide 20   .......... .......... .......... .......... .......... .....ACtTT A........A


      351                                  394   12.7 kDa
Mebus  GCATTATTAA TTGCCAAAGT TTCTAAGGTC ACGCCCTAGT AATG
   LY  GCATTATTAA TTGCCAAAGT TcCTAAGGcC ACGCCCTAGT AATG
  LSU  GCATTAT... .TGCCAAAGT TTtTAAGGcC ACGCCCTAGT AATG
   OK  GCATTAT... .TGCCAAAGT TTtTAAGGcC ACGCCCTAGT AATG
HCVOC43 peptide 11  GCATT..... .TGttAAAGT TctTAAGGcC ACGCCCTAtT AATG
HEVNT9 peptide 24   GCATT..... .TGCCAAAGT TTtTAAGGct ACGCCCTAtT AATG
HEVVW572 peptide 24 GCATT..... .TGCCAAAGT TTtTAAGGcC ACtCCCTAgT AATG
HEV67N peptide 20   GCATT..... .TGCCAAAGT TTtTAAGGct ACGCtCTAtT AATG
```

Fig. 3. **Alignment of BCV nucleotide sequences spanning the region demarkated by the 3′ end of the S gene and the 5′ end of the 12.7 kDa ORF.** Alignment was performed using the multiple sequence alignment with the hierarchical clustering software MultAlin (33). The leader-mRNA junction sequences are underlined. Start and stop codons for each ORF are represented by bolded letters. Nucleotide deletions are noted as dotted lines. Nucleotide differences among strains are noted as small letters, while conserved nucleotides are capital letters.

Strain	1	21	89	137	186	192	252	253	frame-shift segment	262 / 261	274 / 273	279 / 278
G RBCV LSU-94LSS-051-2	ATG	Q*	E*	N	A*	I*	K	S*	**S* Y* F* R* K* S* C* Q***	N	S*	TAA
G RBCV OK-0514-3	ATG	Q*	E*	N	A*	I*	K	S*	**S* Y* F* R* K* S* C* Q***	N	S*	TAA
G EBCV LY138	ATG	E	D*	V	I*	K		S*	**F* Y* F* R* K* S* C* Q***	N	S*	TAA
G EBCV Quebec	ATG	E	D	N	V	S	K	V	L I S V S L V	N	P	TAA

amino acids: 1 21 89 137 186 192 252 253 … 262 274 279 (EBCV Quebec: 261 273 278)

Fig. 4. **Comparison of the predicted amino acid sequences of the 32 kDa ORF specified by different BCVs.** Amino acids which are different in each BCV strain in comparison to BCV-Quebec are shown. * represent amino acid differences from BCV-Quebec. Amino acids in each BCV strain which are different from the BCV-Quebec as the result of the double-frame shift mutations are represented by bolded letters.

virus-infected cells (28). Nucleotide substitutions mapped the crossover region to a 24-nucleotide segment that starts from the last nt of the leader-mRNA junction sequence and extends further downstream. The RBCV isolates LSU, OK, as well as RBCV AZ-26649-2 (AZ) and TX-671-2 (TX) isolated from California and Texas cattle, respectively (data not shown), contained a four nucleotide deletion located within this 24 nucleotide segment (Fig. 3). This deletion may alter the recombination frequency between the leader and the leader-mRNA junction sequence immediately upstream of the 12.7 kDa subgenomic mRNA, and cause either inhibition or enhancement of the putative 12.7 kDa transcription and subsequent protein expression.

Genetic differences between virulent and avirulent BCV strains. The S glycoprotein contained 7 aa substitutions which were common for all virulent strains (aa 33, 40, 248, 470, 965, 1241 and 1341). Three mutations within the SI portion of S caused conservative aa changes, while one non-conservative aa change of His 470 to Asp was located within the S1 hypervariable region. All three mutations within S2 caused non-conservative amino acid changes. Amino changes within S1 and S2 may affect the structure and function of the S glycoprotein and alter the pathogenetic potential of these viruses. The 32 kDa ORF specified by RBCV virulent strains LSU and OK as well as EBCV LY-138 contained two frame-shift mutations which resulted in a 9 aa segment near the carboxy-terminus which was different from the corresponding amino acid sequence specified by the avirulent EBCV-Quebec strain (Fig. 4). A similar double frame shift mutation was found in the 32 kDa ORF specified by HCV-OC43 (29). These frame shift mutations increased substantially the hydrophilicity of the carboxy terminal portion of the 32 kDa protein in virulent versus avirulent strains (data not shown). Sequencing of the corresponding region of the 32 kDa protein of the avirulent EBCV L9 and Mebus strains as well as the 32 kDa of other virulent strains will substantiate these differences between virulent and avirulent strains. The aa substitution of Gly 53 to Val in the E protein was conserved for all BCV virulent strains (F15, LY-138, OK, LSU), HCV-OC43 (30), and three different porcine HEV strains (22). This mutation may affect the ability of these viruses to invade different tissues, because E is part of the virion and is expressed at infected cell-surfaces (31,32).

Acknowledgments

References

1. Spaan W., Cavanagh D., and Horzinek M.C., J Gen Virol *69*, 2939–2952, 1988.
2. Wege H., Siddell S., and ter Meulen V., Curr Top Microb & Immunol *99*, 165–200, 1982.
3. Storz J., Stine L., Liem A., and Anderson G.A., J Am Vet Med Assoc *208*, 1452–1455, 1996.
4. Lai M.M.C., Annu Rev Microbiol *44*, 303–333, 1990.
5. Lai M.M. and Cavanagh D., Adv Virus Res *48*, 1–100, 1997.
6. Cox G.J., Parker M.D., and Babiuk L.A., Nuc Acids Res *17*, 5847, 1989.
7. Cox G.J., Parker M.D., and Babiuk L.A., Virology *185*, 509–512, 1991.
8. Hofmann M.A., Chang R.Y., Ku S., and Brian D.A., Virology *196*, 163–171, 1993.
9. Schultze B. and Herrler G., J Gen Virol *73*, 901–906, 1992.
10. Schultze B., Gross H.J., Brossmer R., and Herrler G., J Virol *65*, 6232–6237, 1991.
11. Rasschaert D., Duarte M., and Laude H., J Gen Virol *71*, 2599–2607, 1990.
12. Krempl C., Schultze B., Laude H., and Herrler G., J Virol *71*, 3285–3287, 1997.
13. Ballesteros M.L., Sanchez C.M., and Enjuanes L., Virology *227*, 378–388, 1997.
14. Storz J., Zhang X.M., and Rott R., Arch Virol *125*, 193–204, 1992.
15. Zhang X.M., Kousoulas K.G., and Storz J., Virology *185*, 847–852, 1991.
16. Zhang X.M., Kousoulas K.G., and Storz J., Virology *183*, 397–404, 1991.
17. Sanger F., Nicklen S., and Coulson A.R., Proc Nattl Acad Sci USA *74*, 5463–5467, 1977.
18. Jimenez C., Herbst W., Biermann U., Muller J.M., and Schliesser T., Zentralblatt Fur Veterinarmedizin. Reihe B *36*, 635–638, 1989.
19. Zhang X.M., Kousoulas K.G., and Storz J., Virology *186*, 318–323, 1992.
20. Zhang X., Herbst W., Kousoulas K.G., and Storz J., Arch Virol *134*, 421–426, 1994.
21. Mounir S. and Talbot P.J., J Gen Virol *74*, 1981–1987, 1993.
22. Vieler E., Schlapp T., and Herbst W., J Gen Virol *77*, 1443–1447, 1996.
23. Laude H., et al. in Talbot P.J. and Levy G.A. (eds). *Corona and related viruses: Functional domains in the spike protein of transmissible gastroenteritis virus.* Plenum Press, New York, pp. 299–394, 1995.
24. Schultze B., Wahn K., Klenk H.D., and Herrler G., Virology *180*, 221–228, 1991.
25. Cavanagh D. in Siddell S.G. (ed.). *The Coronaviridae: The coronavirus surface glycoprotein.* Plenum Press, New York and London, pp. 73–113, 1995.
26. Yoo D.W., Parker M.D., and Babiuk L.A., Virology *180*, 395–399, 1991.
27. Gallagher T.M., Escarmis C., and Buchmeier M.J., J Virol *65*, 1916–1928, 1991.
28. Chang R.Y., Krishnan R., and Brian D.A., J Virol *70*, 2720–2729, 1996.
29. Labonte P., Mounir S., and Talbot P.J., J Gen Virol *76*, 431–435, 1995.
30. Mounir S., Labonte P., and Talbot P.J., Adv Exp Med & Biol *342*, 61–67, 1993.
31. Brown T.D.K. and Brierly I., in Siddell S. (ed.) *The Coronaviridae: The Coronavirus Non-Structural Proteins.* Plenum Press, New York and London, pp. 191–217, 1995.
32. Godet M., L'Haridon R., Vautherot J.F., and Laude H., Virology *188*, 666–675, 1992.
33. Corpet F., Nucl Acids Res *16*, 10881–10890, 1988.