The region between the M and S genes of porcine haemagglutinating encephalomyelitis virus is highly similar to human coronavirus OC43

Elke Vieler, T. Schlapp and W. Herbst

Institut für Hygiene und Infektionskrankheiten der Tiere der Justus-Liebig-Universität, Giessen, Germany

The nucleotide sequences of the regions between the membrane and spike protein genes of three strains of porcine haemagglutinating encephalomyelitis virus (HEV) were determined. A total of 739 (HEV strain 67N) and 751 (strains NT9 and VW572) nucleotides were sequenced. Two ORFs, potentially encoding proteins of 12·8 and 9·6 kDa, were identified. Pairwise comparisons with the corresponding ORFs in bovine coronavirus (BCV) and human coronavirus (HCV) OC43 revealed sequence similarities of greater than 88·5% at the nucleotide and 85·3% at the amino acid level for the 12·8 kDa ORF product. For the 9·6 kDa ORF product similarities were greater than 96·9% and 95·2%, respectively. An additional 12 nucleotide deletion upstream of the 12·8 kDa ORF start codon was found in HEV 67N compared to NT9 and VW572. These results reveal a genomic organization of HEV in the region analysed that is homologous to HCV OC43 but different from BCV.

Porcine haemagglutinating encephalomyelitis virus (HEV) is a member of the *Coronaviridae*. Coronaviruses are spherical, enveloped viruses with a capped and polyadenylated positive-sense ssRNA genome of 27 to 32 kb. Infected cells contain genomic size virus mRNA and six to eight subgenomic mRNA species (depending on the virus) that have a 3′-coterminal nested set structure. A common 5′ leader sequence, derived from the 5′ end of the genomic RNA, is present in each mRNA species. There are several different models for coronavirus mRNA synthesis. The leader-primed transcription model proposes interaction of the leader sequence with the full-length minus-strand RNA at the consensus intergenic sequence to initiate synthesis of subgenomic plus-strand RNAs. In contrast, the model of discontinuous transcription during negative-strand RNA synthesis suggests subgenomic minus-strand RNAs, containing an antileader sequence at the 3′ end, serve as templates for transcription of plus-strand subgenomic RNAs (Holmes & Lai, 1996).

HEV belongs to the antigenic group II of coronaviruses, together with bovine coronavirus (BCV), human coronavirus (HCV) strain OC43, turkey coronavirus (TCV) and mouse hepatitis virus (MHV). Members of this group possess five structural proteins, nucleocapsid (N), membrane (M), small membrane (sM), spike (S) and haemagglutinin–esterase (HE), which are encoded by mRNAs 7, 6, 5·1, 3 and 2·1, respectively. Additionally, ORFs potentially encoding non-structural (ns) proteins have been identified. Their number and sizes differ between coronavirus species (Spaan *et al.*, 1988). Comparing BCV and HCV OC43, the only major genomic difference was found in the region between the M and S genes. While in BCV four ORFs were located in this region, potentially encoding proteins of 4·9 and 4·8 (mRNA 4), 12·7 (mRNA 5) and 9·5 kDa (mRNA 5·1), in HCV OC43 only two ORFs, encoding proteins of 12·9 (mRNA 5) and 9·5 kDa (mRNA 5·1), were identified (Abraham *et al.*, 1990; Mounir & Talbot, 1993*a*). To date, it has not been determined if the 4·9 and 4·8 kDa ns proteins are expressed in BCV-infected cells. The 9·5 kDa proteins of BCV and HCV OC43, in contrast, have been shown to be expressed (Abraham *et al.*, 1990; Mounir & Talbot, 1993*a*). The functions of the small ns proteins are not known. However, it has been proposed that they may act as an anchoring region for structural proteins during virus assembly or that they may maintain membrane association of the virus polymerase during replication (Boursnell & Brown, 1984; Boursnell *et al.*, 1985; Leibowitz *et al.*, 1988; Skinner & Siddell, 1985; Skinner *et al.*, 1985).

Among the haemagglutinating coronaviruses the genomes of MHV, BCV and HCV OC43 have been sequenced to the most part. Comparisons of the predicted amino acid sequences

Author for correspondence: Elke Vieler. Present address: Retrovirus Laboratory, Division of Comparative Medicine, Johns Hopkins University School of Medicine, 720 Rutland Avenue/Traylor G-60, Baltimore, MD 21205, USA.
Fax +1 410 955 9823. e-mail VIELER@WELCHLINK.WELCH.JHU.EDU

The nucleotide sequence data reported in this paper have been deposited with the EMBL nucleotide sequence database under accession numbers X89861, X89862 and X89863.

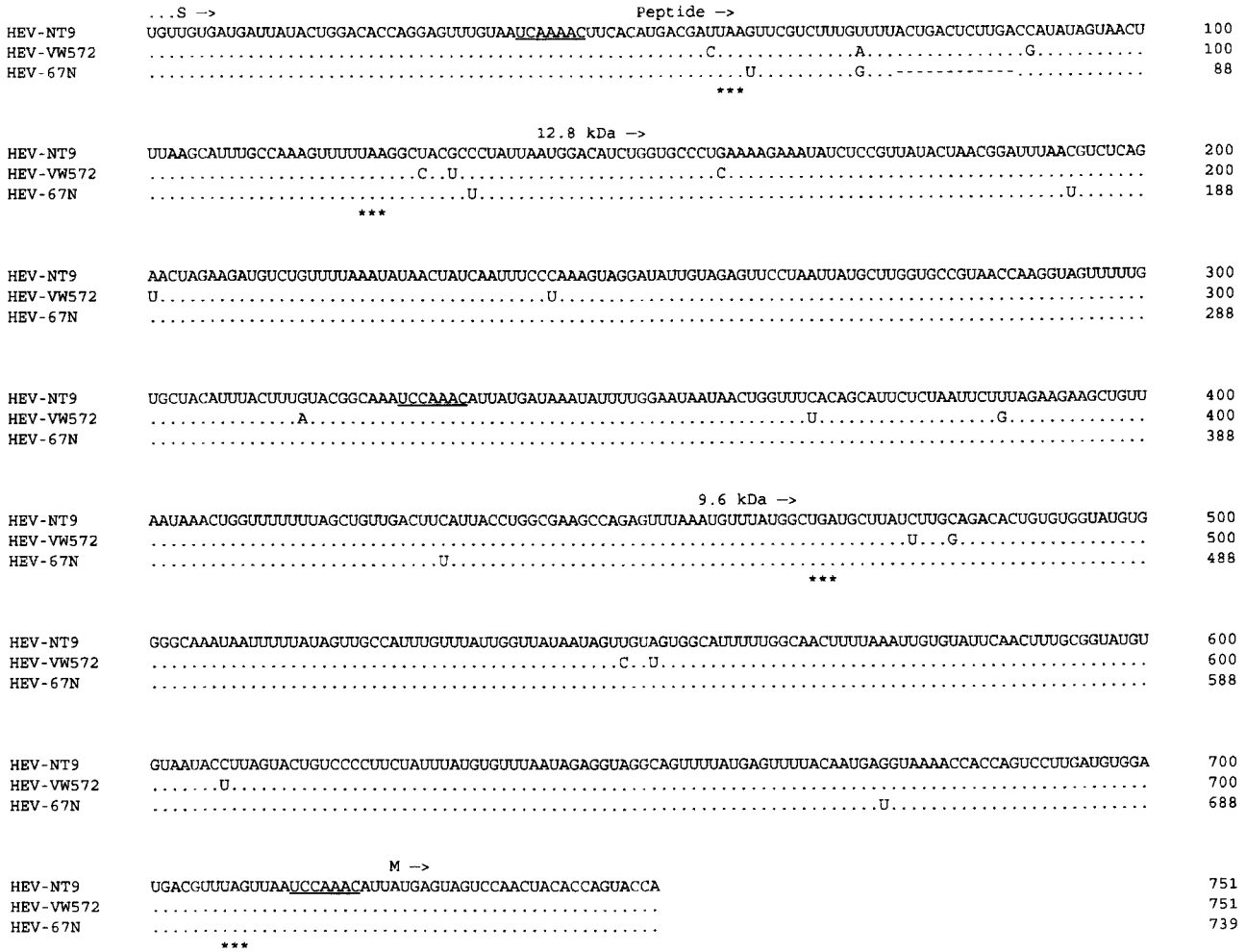

Fig. 1. Comparison of the nucleotide sequences between the M and S genes of HEV NT9, VW572 and 67N. The sequence of HEV NT9 is written in full. Positions having identical nucleotides are marked with dots, positions of deleted nucleotides are marked with dashes. Consensus intergenic sequences are underlined. Asterisks indicate stop codons.

of the structural and ns proteins of BCV and HCV OC43 revealed more than 91% identity (Cox *et al.*, 1989; Kamahora *et al.*, 1989; Künkel & Herrler, 1993; Labonté *et al.*, 1995; Lapps *et al.*, 1987; Mounir & Talbot, 1992, 1993*a*, *b*; Zhang *et al.*, 1991*a*, *b*, 1992). In contrast, no sequence data for HEV, except for a small portion of the S gene, have been published (Vieler *et al.*, 1995). This short nucleotide sequence as well as hybridization studies using BCV S gene-specific probes indicated a close genomic relationship to BCV and HCV OC43.

Since the only major genomic difference between BCV and HCV OC43 identified so far was found in the region between the M and S genes it was of interest to study this genomic region in the closely related HEV. We report here the sequence between the S and M genes of three HEV strains. Only two ORFs, with the potential to encode proteins of 12·8 and 9·6 kDa, were identified in the HEV genome, a situation similar to HCV OC43 and different from BCV.

The origin and cultivation of HEV strains 67N, VW572 and NT9, BCV L9 and HCV OC43 as well as the HRT-18 cell line have been described previously (Vieler *et al.*, 1995). Total RNA for RT–PCR was isolated from virus-infected cells by phenol–chloroform extraction followed by ethanol precipitation. Single-stranded cDNA was synthesized at 47·5 °C using SuperScript Plus RNase H^- reverse transcriptase with primer 6341 (5′ AACGTCATCCACATCAAGAAC 3′) corresponding to nucleotides 6321–6341 of the BCV 9·5 kDa sM protein (EMBL accession number M30612). cDNA for preparation of the N gene probe used for Northern blot hybridization was synthesized basically under the same conditions but using primer 3′-N (5′ CTCCTGGTAAGCAATCCAGT 3′). The digoxigenin-labelled probe (DIG dUTP; Boehringer Mannheim), corresponding to nucleotides 827–1376 of the BCV N gene (EMBL M16620), was prepared by PCR from BCV L9 cDNA with primers 3′-N and 5′-N (5′ TAGTAACCCT GAGGGAGTAC 3′). For sequencing of the region between the M and S genes, DNA was prepared by PCR using primers 6341 and 5327 (5′ ATGTGGTGGTTGTTGTGATGA 3′).

(*a*)

```
HEV-NT9      MDIWCPEKKYLRYTNGFNVSELEDVCFKYNYQFPKVGYCRVPNYAWCRNQGSFCATFTLYGKSKHYDKYPGIITGFTAFSNSLEEAVNKLVFLAVDFITWRSQSLNVYGX  109
HEV-VW572    ......Q.........*..V.....................................................V.....................  109
HEV-67N      ................*.....................................................................................  109
HCV-OC43     ....R........I...*......A...F.............SH.......R..................V.N.....A.TV.D..............R.E......  109
BCV-Mebus    ....R..I.........*......A...F..K.........SH............L.............V.......A.TV.................R E......  109
```

(*b*)

```
HEV-NT9      MFMADAYLADTVWYVGQIIFIVAICLLVIIVVVAFLATFKLCIQLCGMCNTLVLSPSIYVFNRGRQFYEFYNEVKPPVLDVDDVX  84
HEV-VW572    .......FG.....................................................................  84
HEV-67N      .....................................................................D.....  84
HCV-OC43     ............................T.....................................D.....  84
BCV-Mebus    .......F.....................................G.....................  84
```

(*c*)

```
HEV-NT9      MTIKFVFVLLTLDHIVTLSICQSFX                  24
HEV-VW572    ..T....D.....D.........X                    24
HEV-67N      ...N..----GF..........X                     20
HCV-OC43     ..T....DS.HX                               11
BCV-Mebus    ..T....D..AP.D.LHPFNHVKLIIRPIEVEHIIIATTMPAVX  43
```

Fig. 2. Alignment of the predicted amino acid sequences of the 12·8 kDa ORF product (*a*), the 9·6 kDa ORF product (*b*) and the 24 and 20 aa peptide (*c*) of HEV strains NT9, VW572 and 67N compared to the corresponding sequences in HCV OC43 (Mounir & Talbot, 1993) and BCV Mebus (Abraham *et al.*, 1990). The sequence of HEV NT9 is written in full. Only amino acid differences are indicated for the other coronavirus strains. Identical amino acid residues are marked with dots. The potential *N*-glycosylation site in the 12·8 kDa ORF product is indicated by an asterisk.

The sequencing of both strands was performed as previously described (Vieler *et al.*, 1995) using the additional primers HEV-NS1 (5′ CTTGGTGCCGTAACCAA 3′), HEV-NS2 (5′ CTCTACAATATCCTACTTT 3′), HEV-NS3 (5′ GGTTTCA CAGCATTCTCT 3′), HEV-NS4 (5′ AGAGAATGCTGTG AAACC 3′), HEV-NS5 (5′ GACACTGTGTGGTATGT 3′), HEV-NS6 (5′ ACATACCAACAGTGTC 3′), HEV-NS7 (5′ ATAGCTTCATCAGCAGT 3′) and HEV-NS8 (5′ GTGTAT TCAACTTTGCG 3′). The DNASTAR software package was used for DNA sequence analysis. For Northern blot analysis, poly(A)–RNA was isolated from virus-infected and mock-infected cells 16 h post-infection using the QuickPrep Micro mRNA Purification Kit (Pharmacia) according to the manufacturer's instructions. Subgenomic RNAs were separated on a 0·8% agarose gel containing 2% formaldehyde and were transferred to a nylon membrane (Biodyne B; Pall). Prehybridization and hybridization were done at 50 °C using the DIG Luminescent Detection kit for nucleic acids (Boehringer Mannheim) as recommended by the manufacturer. After hybridization, blots were washed twice for 5 min in 2 × SSC and 0·1% SDS at room temperature and then twice for 15 min at 75 °C in 0·5 × SSC and 0·1% SDS. The hybridization signals were immunologically detected according to the manufacturer's protocol.

As shown in Fig. 1, the region between the S and M genes of HEV 67N, NT9 and VW572 contained two ORFs potentially encoding proteins of 12·8 and 9·6 kDa. The predicted amino acid (aa) sequences of these ORFs are shown in Fig. 2. Additionally, an ORF coding for a 24 aa (HEV NT9 and VW572) and 20 aa (67N) protein was observed upstream of the 12·8 kDa ORF beginning at nucleotide (nt) 50.

The 12·8 kDa ORF (nt 140–469 in NT9 and VW572, 128–457 in 67N) potentially encodes a protein of 109 aa with a molecular mass of 12·82 (NT9 and VW572) or 12·77 kDa (67N). Nucleotide and amino acid sequence alignments with the corresponding sequences in BCV Mebus (Abraham *et al.*, 1990) and HCV OC43 (Mounir & Talbot, 1993*a*) revealed identities of more than 88·5% at the nucleotide and more than 85·3% at the amino acid level. As in BCV and HCV OC43 a potential *N*-linked glycosylation site was located at aa position 18. The putative initiation codon with its surrounding context (UUAAUGG) and the consensus intergenic sequence (UCAA AAC) were identical to those found for HCV OC43. The distance between the initiation codon and the consensus intergenic sequence was 101 nt in HEV NT9 and VW572, and 89 nt in 67N, while in HCV OC43 this distance is 109 nt. The shorter distance in HEV 67N compared to NT9 and VW572 results from an additional 12 nt deletion upstream of the

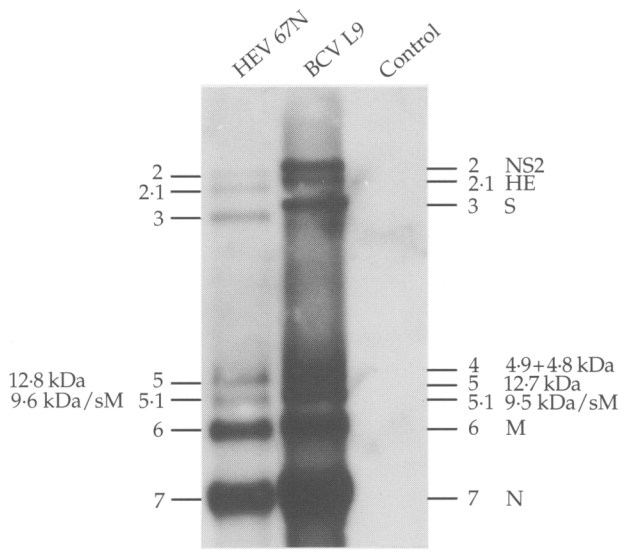

Fig. 3. Northern blot analysis of HEV 67N and BCV L9 subgenomic mRNAs. Poly(A)–mRNA from HEV 67N- and BCV L9-infected cells and from uninfected cells (control) was hybridized with a DNA probe corresponding to nucleotides 827–1376 of the BCV N gene (Lapps *et al.*, 1987). The subgenomic mRNA species are indicated.

12·8 kDa start codon (between nt 75 and 88 in NT9 and VW572; Fig. 1).

The 9·6 kDa ORF (nt 456–710 in HEV NT9 and VW572, 444–698 in 67N) is predicted to encode an 84 aa protein with a molecular mass of 9·55 kDa in NT9, 9·56 kDa in VW572 and 9·58 kDa in 67N. Similarities with the corresponding regions and their products in BCV and HCV OC43 are more than 96·9% at the nucleotide and more than 95·2% at the amino acid level. The UCCAAAC consensus intergenic sequence begins 129 nt upstream of the first potential start codon, which is in the context UAAAUGU. These features correspond to those reported for BCV and HCV OC43 (Abraham *et al.*, 1990; Mounir & Talbot, 1993*a*).

Detailed analysis of the amino acid sequences of the predicted proteins using the Clustal method (data not shown; DNASTAR software package) revealed a closer relationship between BCV and HCV OC43 than between either of them and the three HEV strains. The only exception was the 9·6 kDa sM protein of HEV 67N, which was more closely related to HCV OC43 than to all other coronaviruses tested.

The 24 and 20 aa-encoding ORFs (nt 50–121 and 50–109), respectively, encode predicted peptides with molecular masses of 2·76 (HEV NT9), 2·75 (VW572) and 2·30 kDa (67N). Since only 645 nt of the HEV S2 gene were sequenced (Vieler *et al.*, 1995) the search for the consensus intergenic sequence of these peptides was restricted to this fragment. However, at the position where the consensus intergenic sequence UCAAAC (nt 3760–3765, Mounir & Talbot, 1993*b*) was found for the HCV OC43 peptide and the corresponding consensus sequence for the BCV Mebus 4·9 kDa ns protein gene (CCA AAC; nt 1–6, Abraham *et al.*, 1990), the sequence CCAAUC (nt 394–399; Vieler *et al.*, 1995) was found for HEV 67N. No other potential consensus sequence was found within the HEV 67N S2 gene fragment. Alignments of the first 10 aa of the predicted peptides revealed a high similarity with the 11 codon ORF in HCV OC43 and with the 3′ end of the ORF encoding the 4·9 kDa protein in BCV Mebus (Fig. 2).

Using Northern blot analysis (Fig. 3), HEV 67N and BCV L9 mRNAs 7, 6, 5·1 and 5, encoding the N , M , sM and 12·8 kDa ns proteins, showed similar size. mRNA 3, 2·1 and 2, encoding the S, HE and NS2 proteins, migrated more slowly in BCV L9 than the corresponding mRNA species in HEV 67N. The subgenomic RNA potentially encoding the 4·9 and 4·8 kDa ns proteins in BCV (mRNA 4) was visible as a minor band but was absent in HEV 67N RNA. Similar results were obtained for HEV NT9 and VW572 (data not shown). Mounir & Talbot (1993*a*) reported that despite the lack of these two ORFs in HCV OC43, small amounts of mRNA 4 were present in HCV OC43-infected cells. In this study, however, mRNA 4 could not be detected in any of the three HEV strains tested. It seems possible that mRNA 4 in HEV might be obscured by mRNA 5, since it is at least 290 nt smaller than BCV mRNA 4. Another reason may be that only very small amounts of this mRNA species were produced in the HEV-infected HRT-18 cells.

Since BCV, HEV and HCV OC43 all replicate well in HRT-18 cells, although the 4·9 and 4·8 kDa ns protein-encoding genes were absent in HEV and HCV OC43, it appears that these putative proteins are not essential for virus replication.

It has been suggested that the absence of these two genes in HCV OC43, together with differences in the S protein, might be involved in the apparently preferential respiratory tropism of this virus (Mounir & Talbot, 1993*a*). However, HEV has a strong tropism for neurons in experimentally infected mice and rats and in naturally infected piglets, but actively replicates in the respiratory tract before spreading to the nervous system (Fenner *et al.*, 1993; Hirano *et al.*, 1993; Yagami *et al.*, 1993), indicating that other mechanisms than the absence of these genes are involved in tissue tropism.

References

Abraham, S., Kienzle, T. E., Lapps, W. E. & Brian, D. A. (1990). Sequence and expression analysis of potential nonstructural proteins of 4·9, 4·8, 12·7 and 9·5 kDa encoded between the spike and the membrane protein genes of the bovine coronavirus. *Virology* **177**, 488–495.

Boursnell, M. E. G. & Brown, T. D. K. (1984). Sequencing of coronavirus IBV genomic RNA: a 195-base open reading frame encoded by mRNA B. *Gene* **29**, 87–92.

Boursnell, M. E. G., Binns, M. M. & Brown, T. D. K. (1985). Sequencing of coronavirus IBV genomic RNA: three open reading frames in the 5′ 'unique' region of mRNA D. *Journal of General Virology* **66**, 2253–2258.

Cox, G. J., Parker, M. D. & Babiuk, L. A. (1989). The sequence of cDNA of bovine coronavirus 32K nonstructural gene. *Nucleic Acids Research* **17**, 5847.

Fenner, F. J., Gibbs, E. P. J., Murphy, F. A., Rott, R., Studdert, M. J. & White, D. O. (editors) (1993). *Coronaviridae.* In *Veterinary Virology,* 2nd edn, pp. 457–469. San Diego: Academic Press.

Hirano, N., Haga, S. & Fujiwara, K. (1993). The route of transmission of hemagglutinating encephalomyelitis virus (HEV) 67N strain in 4-week-old rats. *Advances in Experimental and Medical Biology* **342**, 333–338.

Holmes, K. V. & Lai, M. M. C. (1996). *Coronaviridae*: the viruses and their replication. In *Fields Virology,* 3rd edn, pp. 1075–1093. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott-Raven.

Kamahora, T., Soe, L. H. & Lai, M. M. C. (1989). Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Research* **12**, 1–9.

Künkel, F. & Herrler, G. (1993). Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **195**, 195–202.

Labonté, P., Mounir, S. & Talbot, P. J. (1995). Sequence and expression of the ns2 protein gene of human coronavirus OC43. *Journal of General Virology* **76**, 431–435.

Lapps, W., Hogue, B. G. & Brian, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein gene. *Virology* **157**, 47–57.

Leibowitz, J. L., Perlman, S., Weinstock, G., Devries, J. R., Budzilowicz, C., Weissemann, J. M. & Weiss, S. R. (1988). Detection of a murine coronavirus nonstructural protein encoded in a downstream open reading frame. *Virology* **164**, 156–164.

Mounir, S. & Talbot, P. J. (1992). Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for O-glycosylation. *Journal of General Virology* **73**, 2731–2736.

Mounir, S. & Talbot, P. J. (1993*a*). Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus. *Virology* **192**, 355–360.

Mounir, S. & Talbot, P. J. (1993*b*). Molecular characterization of the S protein gene of human coronavirus OC43. *Journal of General Virology* **74**, 1981–1987.

Skinner, M. A. & Siddell, S. G. (1985). Coding sequence of coronavirus MHV-JHM mRNA 4. *Journal of General Virology* **66**, 593–596.

Skinner, M. A., Ebner, D. & Siddell, S. G. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *Journal of General Virology* **66**, 581–592.

Spaan, W., Cavanagh, D. & Horzinek, M. C. (1988). Coronaviruses: structure and genome expression. *Journal of General Virology* **69**, 2939–2952.

Vieler, E., Schlapp, T., Anders, C. & Herbst, W. (1995). Genomic relationship of porcine hemagglutinating encephalomyelitis virus to bovine coronavirus and human coronavirus OC43 as studied by the use of bovine coronavirus S gene-specific probes. *Archives of Virology* **140**, 1215–1223.

Yagami, K., Izumi, Y., Kajiwara, N., Sugiyama, F. & Sugiyama, Y. (1993). Neurotropism of mouse-adapted hemagglutinating encephalomyelitis virus. *Journal of Comparative Pathology* **109**, 21–27.

Zhang, X., Kousoulas, K. G. & Storz, J. (1991*a*). Comparison of the nucleotide and deduced amino acid sequences of the S genes specified by virulent and avirulent strains of bovine coronaviruses. *Virology* **183**, 397–404.

Zhang, X., Kousoulas, K. G. & Storz, J. (1991*b*). The hemagglutinin/esterase glycoprotein of bovine coronaviruses: sequence and functional comparisons between virulent and avirulent strains. *Virology* **185**, 847–852.

Zhang, X., Kousoulas, K. G. & Storz, J. (1992). The hemagglutinin/esterase gene of human coronavirus strain OC43: phylogenetic relationships to bovine and murine coronavirus and influenza C virus. *Virology* **186**, 318–323.

Received 12 January 1996; Accepted 23 February 1996