Analysis of a 9.6 kb sequence from the 3′ end of canine coronavirus genomic RNA

Brian C. Horsburgh, Ian Brierley and T. David K. Brown*

Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, U.K.

We have analysed the organization of the 3′ end of the genomic RNA of canine coronavirus (CCV), a virus which has a close antigenic relationship to transmissible gastroenteritis virus (TGEV), porcine respiratory coronavirus (PRCV) and feline infectious peritonitis virus (FIPV). Genomic RNA isolated from CCV strain Insavc-1-infected A72 cells was used to generate a cDNA library. Overlapping clones, spanning approximately 9.6 kb [from the 3′ end of the polymerase gene, 1b, to the poly(A) tail] were identified. Sequencing and subsequent analyses revealed 10 open reading frames (ORFs). Three of these code for the major coronavirus structural polypeptides S, M and N; a fourth codes for a small membrane protein, SM, a putative homologue of the IBV structural polypeptide 3c, and five code for polypeptides, designated 1b, 3a, 4, 7a and 7b, homologous to putative non-structural polypeptides encoded in the TGEV or FIPV genomes. An extra ORF which had not hitherto been identified in this antigenic group of coronaviruses was designated 3x. Pairwise alignment of these ORFs with their counterparts in TGEV, PRCV and FIPV revealed high levels of identity and highlighted the close relationship between the members of this group of viruses.

Introduction

Canine coronavirus (CCV), a causative agent of enteritis in neonatal dogs, was first identified in 1971 (Binn et al., 1974). The disease is characterized by infection of the absorptive epithelium of the villi and the onset of diarrhoea followed by villus atrophy (Keenan et al., 1976). CCV belongs to the Coronaviridae, a family of enveloped viruses possessing a ssRNA genome of positive polarity. In infected cells, a set of 3′-coterminal subgenomic RNAs are produced and, as a result, the 5′ end of each mRNA contains unique sequence information not present on smaller RNAs in the nested set. Only this unique region of each mRNA is translated (reviewed by Spaan et al., 1988), thus the mRNAs are, in principle, functionally monocistronic. Nevertheless, some mRNAs contain two or more coding regions within the unique sequence and thus may be functionally bi- or tricistronic (Brierley et al., 1987; Liu et al., 1991; Liu & Inglis, 1992). The CCV virion is known to contain at least four protein species: the 204K spike glycoprotein, S; the 32K membrane glycoprotein, M; the 9.2K small membrane protein, SM; and the 50K nucleocapsid protein, N (Garwes & Reynolds, 1981; Godet et al., 1992).

CCV belongs to one of the major antigenic groups of coronaviruses (Siddell et al., 1983; Spaan et al., 1988) and is serologically related to feline infectious peritonitis virus (FIPV), feline enteric coronavirus (FECV), transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus (PRCV) (Sanchez et al., 1990). These viruses have been distinguished mainly by their host species of origin. It has been reported, however, that some strains of CCV can also infect cats (Barlough et al., 1984; Stoddart et al., 1988) and swine without causing any apparent disease (Woods & Wesley, 1986). Likewise, TGEV can also infect other species (Woods & Pedersen 1979; Norman et al., 1970) and FIPV can infect swine (Woods et al., 1981). This close relationship indicates that the viruses may have a common ancestor (Horzinek et al., 1982; Sanchez et al., 1990).

Molecular analysis has helped to elucidate some of the aspects of this phylogenetic relationship and some of the mechanisms involved in pathogenesis. TGEV, PRCV and FIPV have been characterized in some detail and the genes encoding the structural proteins have been cloned and sequenced (de Groot et al., 1987; Vennema et al., 1991; Britton et al., 1988a, b; Rasschaert & Laude, 1987; Rasschaert et al., 1990). A comparison of the available

---

The nucleotide sequence data reported in this paper have been submitted to GenBank and EMBL and assigned the accession number D13096.
FIPV amino acid sequences with the corresponding sequences of TGEV and PRCV has revealed that the structural genes are very closely related. For S the identities were 81.6% (TGEV) and 76% (PRCV), for M 84.4% and 85.9%, and for N 77% and 75.6%, respectively. This contrasts greatly with the relationship to murine hepatitis virus (MHV), a prototypic coronavirus from another antigenic group, where the identities for these polypeptides are 24%, 30%, and 27%, respectively (Schmidt et al., 1987; Skinner & Siddell, 1983; Armstrong et al., 1984). Despite this high degree of similarity amongst the structural proteins of these three viruses there are, nevertheless, differences at the 3' end of their viral genomes and in their subgenomic message organisation.

CCV is the least characterized virus from this antigenic group. Here we report the cloning and sequencing of 9.6 kb from the 3' end of the RNA of the avirulent CCV strain Insavc-1, subgenomic message analysis and comparison to available TGEV, PRCV and FIPV sequence data which illuminate the evolutionary relationship of this family of viruses. The presented sequence, which includes all of the CCV coding information except for the polymerase region, represents the first report of cloning and sequencing of a canine coronavirus.

Methods

Virus and cells. Canine A72 cells and CCV strain Insavc-1 were obtained from Dr W. Baxendale (Intervet UK, Houghton, U.K.). A72 cells were grown in Gibco's Welcome formula, a modified Eagle's medium supplemented with 10% foetal calf serum (FCS) containing penicillin (100 units/ml) and streptomycin (100 μg/ml) (MEM). Flasks (175 cm²) of A72 cells were washed with PBS and infected with CCV at an m.o.i. of 0-1 in 10 ml MEM. Virus adsorption was allowed to proceed for 60 min at 37°C and the inoculum was then replaced by MEM-10% FCS.

Preparation of CCV genomic and messenger RNAs. CCV genomic RNA was prepared as follows. At 48 h post-infection (p.i.) the culture supernatant was harvested, chilled to 4°C and the cell debris removed by low-speed centrifugation (3000 g for 15 min). Virus was pelleted from the supernatant at 30000 g for 2 h (Beckman type 19 rotor) and the pellet homogenized in 6 M-guanidinium isothiocyanate, 0.5% N-lauroyl sarcosinate, 5 mM-sodium citrate. The mixture was layered onto a 5.7 M-CsCl pad and viral RNA pelleted by centrifugation (108000 g for 12 h) and analysed on a 1% Tris-borate-EDTA agarose gel containing 0.1% SDS. A single species of high Mf RNA was identified with the characteristic mobility of coronavirus genomic RNA.

Subgenomic RNAs were prepared in a similar manner. Briefly, at 36 h p.i. the infected cells were chilled to 4°C, washed three times with ice cold PBS then pelleted at 30000 g for 10 min. The cell pellet was homogenized in 6 M-guanidinium isothiocyanate, 0.5% N-lauroyl sarcosinate, 5 mM-sodium citrate then treated as described above.

Cloning of CCV genomic RNA

(i) cdNA cloning. A cdNA library from CCV genomic RNA was prepared by reverse transcription after priming with oligo(dT) and random pentanucleotides using the instructions and contents of the Boehringer Mannheim Biochemica cdNA synthesis kit. The resulting cdNA was blunt-ended using T4 DNA polymerase and ligated into the Smal site of pUC119. Portions of the ligation mixture were transformed into Escherichia coli strain TG-1 and clones were identified by colour selection. Inserts of viral origin were confirmed by colony hybridization using cdNA prepared by random priming of CCV RNA as a probe. CCV-derived recombinant clones were analysed by restriction enzyme digestion and those containing inserts of 1-8 kb or greater in size were retained for further study.

(ii) Polymerase chain reaction (PCR). PCR-amplified fragments were obtained using cdNA:RNA heteroduplexes as template and oligonucleotides 7 and 8 (each of which contains a NotI site; Fig. 1) as primers. Tag DNA polymerase (Promega) was used to amplify the region of interest according to the recommendations of Sambrook et al. (1989) and 25 cycles (95°C, 1 min; 60°C, 1 min; and 72°C, 2 min) were performed in a Techne PHC-1 machine. The generated DNA fragment was cleaved with NotI, gel-purified, ligated into the NotI site of pKL1 and transformed into E. coli strain TG-1. (pKL1 is a PUC-based vector with a modified polylinker and was a gift from Dr K. Law, University of Cambridge, U.K.)

Sequencing

(i) M13 DNA sequencing. DNA sequencing was performed by Sanger's dideoxynucleotide chain termination method as described by Bankier et al. (1987). Briefly, insert DNA was excised from vector sequences, self-ligated and sonicated in a cup-horn sonicator (Heat Systems, Ultrasonics). The sonicated DNA fragments were end-repaired with the Klenow fragment of E. coli DNA polymerase I and T4 DNA polymerase prior to size selection on a 1-2% agarose gel. Fragments in the size range 300 to 500 bp were purified and cloned into Smal-digested, phosphatase-treated M13mp8. Shotgun sequence data were assembled using the SAP programs of Staden (1982) on a VAX 8350 and microVAX 3100 (Digital Equipment Corporation).

(ii) Superoiled DNA sequencing. DNA templates were prepared as described by Lim & Pène (1988). CsCl-purified plasmid DNA (3 μg) was denatured with 0.15 M-NaOH and 0.15 mM-EDTA for 30 min at 37°C, then centrifuged through a Sepharose CL-6-B column equilibrated in TE. Sequencing reactions were carried out on the eluate as described using the pUC forward and reverse primers.

(iii) RNA sequencing. Primer (50 pmol) was annealed to either 1 μg genomic RNA or 10 μg total infected cell RNA at room temperature for 15 min. Sequencing reactions were performed as described by Fichot & Girad (1990).

Northern blot hybridization. Total RNA extracted from CCV-infected cells was denatured for 15 min at 56°C in 50% deionized formamide, 2:2 M-formaldehyde and 0:5 mM-EDTA. The samples were cooled on ice after the addition of loading buffer containing 0.5% SDS, 0.025% bromophenol blue and 25% glycerol. The samples were electrophoresed overnight in a horizontal submerged gel containing 1:1 M-formaldehyde and 0.8% agarose. RNA was blotted from the gel to a nitrocellulose filter (Schleicher and Schuell). Prehybridization was carried out in 5 x SSC (1 x SSC is 150 mM-sodium chloride and 15 mM-sodium citrate), 10 x Denhardt's solution (1 x Denhardt's solution is 0.02% polyvinylpyrrolidone, 0.02% Ficoll and 0.02% bovine serum albumin), 100 μg/ml sonicated salmon sperm DNA and 0.1% SDS for 2 h at 65°C. Hybridization was carried out at 65°C overnight after addition of a 32P-radiolabelled DNA probe prepared by random priming the CCV-specific insert purified from pBH5 (Sambrook et al., 1989). Following hybridization, the filter was washed twice at 65°C with 2 x SSC, then washed three times at 42°C with 0 x SSC, prior to exposure to X-ray film.
Sequence of 3' end of canine coronavirus RNA

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Approximate size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBH5</td>
<td>1.9</td>
</tr>
<tr>
<td>pBH6</td>
<td>1.7</td>
</tr>
<tr>
<td>pBH7</td>
<td>2.6</td>
</tr>
<tr>
<td>pBH8</td>
<td>2.0</td>
</tr>
<tr>
<td>pBH9</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**Southern blotting and other cloning procedures.** These were carried out according to the protocols of Sambrook et al. (1989). Enzymes were used according to the manufacturers’ specifications (Boehringer Mannheim and New England Biolabs).

**Results**

**Generation and mapping of CCV clones**

To clone the 3' end of the CCV genome we prepared a cDNA library from CCV genomic RNA. Inserts from recombinant clones of 1.8 kb or greater were selected for further analyses. In order to map the clones, we took advantage of the suspected nucleotide sequence homology between the genomes of CCV and TGEV. Partial sequencing of recombinant clones revealed identity in excess of 95%. This permitted initial alignment of the CCV clones with respect to the TGEV genome. This approach proved fruitful in that four clones were identified which spanned some 8.5 kb at the 3' end (Fig. 1). A region at the 3' end for which large clones were not represented in the library was prepared by PCR amplification (BH6; Fig. 1). The relationships between putative overlapping clones were confirmed by Southern hybridization. Therefore, partial sequencing and Southern blotting identified five overlapping clones which covered approximately 9.6 kb from the 3' end of the CCV genome.

**Shotgun DNA sequencing and sequence analyses**

The inserts from the plasmids detailed in Fig. 1 were sequenced using the shotgun methods of Bankier et al. (1987). The consensus nucleotide sequence of 9624 bp presented in Fig. 2 was analysed using the SAP programs of Staden (1982). Analysis revealed the presence of 10 open reading frames (ORFs) (Fig. 3). Pairwise alignment of these ORFs with their likely counterparts from other members of this coronavirus group disclosed very high levels of identity (Table 1) and indicated that the CCV structural proteins S, M and N are encoded by ORFs 2, 4 and 5, respectively. Each of the 10 ORFs is described in more detail below.

With respect to subgenomic mRNA synthesis, it is known that the minimal conserved signal for transcription in this coronavirus group, CTAAAC, is identical in TGEV, PRCV and FIPV and is therefore likely to be conserved in CCV (as reviewed by Spaan et al., 1988). Indeed, analysis of the CCV sequence revealed that this sequence was present upstream of all the ORFs with the exception of the first and last. As ORF 1 is incomplete (see below), an additional CTAAAC sequence is presumably located at the 5' end of the genomic RNA. When we analysed intracellular RNAs produced during CCV infection of canine A72 cells, eight species of RNA were observed (Fig. 4); the species observed between species 5 and 6 could not be accounted for in terms of the Y-coterminal nested arrangement of coronavirus subgenomic RNAs and the observed positions of consensus transcription initiation signals. Taking into account the predicted size of each mRNA and the known location of the CTAAAC sequences, we predict a subgenomic message organization as depicted in Fig. 3. The ORFs encoded by each mRNA are described below.

The numbering of CCV RNAs used here is based on that currently employed by workers studying the Purdue-115 and FS772/70 strains of TGEV. The RNA organization of CCV strain Insavc-I is most closely related to that described for these TGEV strains. This numbering scheme is not, however, applicable in a straightforward fashion to all members of the antigenic group. In the case of the Miller strain of TGEV, an RNA originally designated 4b (Wesley et al., 1989) may be involved in the expression of ORF 3b and ORF 4; no additional RNA was detected between this RNA and the RNA coding for the membrane protein (RNA 5), but the
Sequence of 3' end of canine coronavirus RNA
Fig. 2. Sequence of the extreme 3' 9624 nucleotides of the CCV genome and the deduced amino acid sequences encoded by the ORFs. The consensus intergenic sequences are underlined. The octanucleotide sequence conserved in the 3' non-coding region of all coronaviruses is shown in bold. The predicted ORFs are translated into the single-letter amino acid code. The putative signal peptides for S and M proteins are underlined.

Fig. 3. Gene and subgenomic message organization predicted from the sequence data and Northern blot analyses cited in Results. Genes are designated according to the recommendations of the coronavirus study group (Cavanagh et al., 1990a, b). ORFs are represented by boxes. The vertical line in ORF 3b represents a stop codon and the black boxes represent leader sequences. Numbers represent ORFs encoded by that message.

Table 1. Pairwise sequence homology between CCV and FIPV, TGEV, PRCV and MHV ORFs

<table>
<thead>
<tr>
<th>CCV ORF</th>
<th>No. of amino acids</th>
<th>Pairwise identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FIPV</td>
<td>TGEV</td>
</tr>
<tr>
<td>1b</td>
<td>168</td>
<td>95-2</td>
</tr>
<tr>
<td>2 (S)</td>
<td>160</td>
<td>91-1</td>
</tr>
<tr>
<td>3a</td>
<td>6-7</td>
<td>83-5</td>
</tr>
<tr>
<td>3b</td>
<td>28-4</td>
<td>251</td>
</tr>
<tr>
<td>4 (SM)</td>
<td>9-3</td>
<td>82</td>
</tr>
<tr>
<td>5 (M)</td>
<td>29-5</td>
<td>262</td>
</tr>
<tr>
<td>6 (N)</td>
<td>43-4</td>
<td>401</td>
</tr>
<tr>
<td>7a</td>
<td>11-5</td>
<td>101</td>
</tr>
<tr>
<td>7b</td>
<td>29-4</td>
<td>213</td>
</tr>
</tbody>
</table>

* ORF 3x is not included (see Results).
† NK, Not known.
‡ Incomplete.
§ Disregarding terminator, otherwise $M_i = 4000$.

possibility that such an RNA may be synthesized at a low level must be considered because a CTAAAC signal was observed. Similarly, in the case of FIPV strain 79-1146 no RNA has been detected between RNA 3 and the membrane polypeptide RNA (de Groot et al., 1987) but the possibility of an equivalent of the TGEV RNA 4 has been alluded to (de Groot, 1989). Thus, the numbering conventions employed do not deal adequately with the variations in expression strategy observed in this region of genome within this group of closely related viruses.

**ORFs encoded by mRNAs 1 and 2**

ORF 1 is incomplete, has no AUG start codon, encodes 168 amino acids and terminates in a UGA stop codon at position 510 (Fig. 2). A comparison of this ORF
with TGEV strain FS772/70 shows 99.2% similarity to 1b and 47
and 52.7% identity to genes 1b of avian infectious bronchitis virus (IBV) and MHV, respectively (Britton & Page, 1990; Boursnell et al., 1987; Bredenbeek et al., 1990). Thus, this ORF represents the 3' end of the putative polymerase-encoding region of genome mRNA 1.

ORF 2 located immediately downstream of the polymerase gene would be translated from the 9-1 kb subgenomic message 2. This ORF is 4356 nucleotides long representing 1452 amino acids with a calculated $M_r$ of 160K. Comparison of this ORF with sequences held in the EMBL database reveals remarkably high identity to the FIPV spike glycoprotein-encoding sequences (91.1%) and, to a lesser degree, the porcine virus S genes (Table 1), indicating that this is the CCV S gene. In some strains of MHV, the haemagglutinin-esterase glycoprotein gene (HE) is found downstream of the polymerase gene (Luytjes et al., 1988) but it is clear that CCV, like TGEV and IBV, encodes only the polymerase gene upstream of the S gene (Britton & Page, 1990; Boursnell et al., 1987).

The CCV S protein shows features characteristic of a type I membrane protein, i.e. a putative signal sequence (Von Heijne, 1986; positions 506 to 563; Fig. 2) and transmembrane domain (positions 4682 to 4742; Fig. 2). There are also 30 potential N-glycosylation sites which probably account for the increased size of the S protein found in the virion (Garwes & Reynolds, 1981).

ORFs encoded by mRNAs 3 and 4

There are four ORFs distal to the S gene coding sequence which are likely to be encoded by messages 3 and 4 (Fig. 3). Three of these have close similarity to their porcine virus counterparts and have been named 3a (8-6K), 3b (28.4K) and 4 (9.3K) (Table 1). The fourth ORF, which to date has not been detected in this group of viruses, could potentially encode a 71 amino acid protein with a predicted $M_r$ of 10K and overlaps ORFs 3a and 3b (Fig. 2 and 3). This ORF has been designated 3x. The CCV 3b ORF was expected to encode a 28K protein like its TGEV counterpart (Jacobs et al., 1986). However, this strain of CCV has acquired a termination codon, UAA (at position 5515; Fig. 2), which would result in a truncated polypeptide of only 33 amino acids. Direct sequencing of the viral genomic and mRNAs has confirmed the authenticity of this stop codon (data not shown). The CCV 4 ORF encodes a small membrane protein that is related to the 3c product of IBV (Fig. 5).

Message 4, as predicted from our sequence data, was detected in Northern blots (see Fig. 4). This message could only express ORF 4, as the proposed signal for transcription, CTAAAC, is found 43 nucleotides upstream of the predicted ORF 4 start codon. This arrangement is found in a number of strains of TGEV.

ORFs encoded by mRNAs 5, 6 and 7

 Messenger RNA species 5 and 6 encode ORFs which resemble the coding sequences for the other coronavirus structural proteins, M and N, respectively (Table 1). Translation of poly(A)-selected CCV intracellular RNA in the rabbit reticulocyte lysate system produced pro-
IBV-Beaudette  MMNLNLKSLENSGFITALYIIVGFIALYLLGRLAOGFVQADACLFWTYYWVVI
CCV-Insavc-i  MTFRPALTVDNGMVIIIIFWPLIIIILLSSIIILNIIKLCMVCNLGRTVIV
TGEV-Miller  MTFRPALTVDNGMVIIIIFWPLIIIILLSSIIILNIIKLCMVCNLGRTVIV
MHV-JHM  MFNLFLTDTWVYVQIIFIVAVCMTIIVVAFASLXIRCIQLGCLNLLLS
BCV-Mebus  MFMAAYFADTVWYVQIIFIVACLLVIIVVVAFLATFNLQICLGMCNLGLS

Fig. 5. Alignment of the putative small membrane protein amino acid sequences from five different strains of coronaviruses. The hydrophobic core is shown in bold. Asterisks represent conserved features.

Discussion

In this study approximately 9.6 kb of the 3' end of the CCV strain Insavc-1 genome was cloned and sequenced. This region is likely to include all of the viral genes excluding the polymerase gene for which only the 3'-terminal 168 amino acids have been determined. Therefore, a substantial part of the virus' genetic information was available for comparison with other antigenically related coronaviruses, namely TGEV, PRCV and FIPV. The deduced sequence and genetic organization of CCV are shown in Fig. 2 and 3, respectively.

From antigenic data and cross-infectivity studies, the viruses within this group have been termed 'host range mutants' (Horzinek et al., 1982). This close evolutionary relationship is emphasized by our analyses of the CCV sequence data. The CCV spike protein is closely related to the other spikes and has the features typical of coronavirus peplomer glycoproteins. Any variation in the sequence of this protein within the group presumably reflects changes in cell tropism, drift as a result of polymerase errors and selection by the host's immune system. Similarly, interspecies comparison of the other structural proteins, M and N, revealed very high levels of identity (Table 1). Alignment of the M gene product amino acid sequences revealed that any variation was primarily found on what would be the exposed amino terminus of the protein (amino acids 22 to 44; Fig. 2), i.e. between the putative signal sequence (Von Heijne, 1986) and the single potential N-glycosylation site. The variation in this region is again probably a result of selection pressure from the host's immune system. Interestingly, alignment of the N gene amino acid sequences indicated that FIPV N has diverged to a greater extent than those of both CCV and TGEV (Fig. 6). This is unusual as N proteins are normally highly conserved; alignment of N gene amino acid sequences from five isolates of MHV showed at least 90% identity (Masters et al., 1990). Nevertheless, variation was mainly clustered in two regions of the N molecule, between positions 204 and 210, and 352 and 359 (Fig. 6). It has been proposed that these two loci represent spacers, which have little sequence specificity but connect conserved domains of the molecule involved in interaction with the RNA genome (Masters et al., 1990).
The ORFs that lie between the S and M genes have, like the other ORFs so far analysed, a high degree of identity to their porcine virus counterparts (Table 1) and presumably perform similar functions. A previously undetected ORF, 3x, was identified which could potentially encode a 10K polypeptide. However, codon usage and base preference programs of Staden (1982) suggest that this ORF does not encode a functional viral protein. Furthermore, the proximal AUG is in a poor context for translation initiation (Kozak, 1986) and the only other AUG is found at the very 3' end of the coding sequence. Therefore, it is very unlikely that this ORF is expressed in this strain of CCV and it probably represents an evolutionarily redundant sequence which is no longer required by the virus. Analysis of TGEV genomic sequence in this region revealed a counterpart for this canine virus pseudogene; 92 nucleotides have, however, been deleted. This deletion also results in a frameshift in the sequence which explains why this ORF has not hitherto been noticed (Fig. 7). In addition to the likely non-functionality of ORF 3x, it is also unlikely that ORF 3b is expressed in this strain of CCV. Although a transcription signal, CTAAAC, is present upstream of ORF 3b (Fig. 2, position 5213), we were unable to detect an mRNA of the predicted size on Northern blots. Even if low-level transcription occurs from this site, it is unlikely that ORF 3b is expressed as there is a termination codon (UAA) some 93 nucleotides downstream of the first AUG and subsequent AUG codons are in poor contexts for ribosome binding (Kozak, 1986). In fact, in vitro transcription and translation of this ORF did not yield any discernible products by SDS-PAGE analysis (data not shown). Alignment of ORF 4 amino acid sequences disclosed features in common with the
passage histories is mainly due to deletions in that gene which can lead to deletions of up to 159 amino acids. Consequently, this has an effect on pathogenicity, as deletions in the MHV-4 S coding sequence apparently result in a loss of ability to induce fatal encephalitis and the acquisition of a non-fatal demyelinating disease in mice (Parker et al., 1989). Polymorphism has also been observed in the S gene and in the region between the S and M genes for different strains of TGEV and the respiratory tract mutant, PRCV (Wesley et al., 1990; Rasschaert et al., 1990). In fact, an IBV strain (Port/322/85) has been reported which appears to have arisen as a result of recombination between the M and S genes from two other strains of IBV (Cavanagh et al., 1990b). The third ‘hot spot region’ is found downstream of the N gene. The porcine coronaviruses have a 69 nucleotide deletion in ORF 7a and ORF 7b is not present (de Groot et al., 1988). This phenomenon is not unique to the coronaviruses from this antigenic group. Deletions of up to 170 nucleotides are found downstream of the N gene in some strains of IBV (Collisson et al., 1990).

CCV ORF 7b has 57% identity to FIPV 6b. This ORF is the least conserved between the two viruses. Whether the protein produced from this ORF plays an important role in the immune-mediated disease seen in felines remains to be seen as all the viruses from this antigenic group can infect cats but only FIPV will produce this disease.

In conclusion, sequencing and subsequent analyses stress the very close relationship CCV has to the other viruses within its antigenic group. We must, however, be careful when generalizing about the CCV sequence data from this limited information. Coronavirus genomes are dynamic, subject to recombination, insertion and deletion, and as a consequence strains may show significant genetic differences. Clearly, there is a need to clone and sequence other strains in order to build a consensus picture of the CCV genome.

We wish to thank Dr W. Baxendale for supplying cells and virus stocks. This work was supported by a research grant from Intervet BV, Boxmeer, The Netherlands.

References


(Received 30 March 1992; Accepted 7 July 1992)