

Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions

Denis Rasschaert, Mariela Duarte and Hubert Laude*

Laboratoire de Virologie et d'Immunologie Moléculaires, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy-en-Josas, Domaine de Vilvert, 78350 Jouy-en-Josas, France

The genome organization of porcine respiratory coronavirus (PRCV), a newly recognized agent which has a close antigenic relationship to the enteropathogenic transmissible gastroenteritis virus (TGEV), was studied. Genomic RNA from cell-cultured PRCV (French isolate RM4) was used to produce cDNA clones covering the genomic 3' end to the start of the spike (S) glycoprotein gene (7519 nucleotides). Six open reading frames (ORFs) were identified that allowed the translation of three coronavirus structural proteins and three putative non-structural (NS) poly-peptides, homologous to TGEV ORFs designated

NS3-1, NS4 and NS7. Pairwise alignment of PRCV nucleotide and amino acid sequences with sequence data available for three TGEV strains revealed a 96% overall homology. However, the genome of PRCV exhibited two important distinctive features. The first was that the S gene lacked 672 nucleotides in the 5' region and encoded a truncated form of the S polypeptide, and secondly, the first NS ORF downstream of the S gene was predicted to be non-functional as a consequence of a double deletion. The significance of genomic deletions with respect to tissue tropism and evolution of coronaviruses is discussed.

Introduction

A new porcine respiratory coronavirus, tentatively designated PRCV, which has a close antigenic relationship to the enteropathogenic transmissible gastroenteritis virus (TGEV), suddenly emerged in 1983 to 1984 and spread within less than 2 years in most if not all European countries, where it now persists enzootically (Pensaert *et al.*, 1986; Brown & Cartwright, 1986; Jestin *et al.*, 1987; Have, 1990). A cytopathic agent has been isolated in cell culture and pathogenesis studies have shown that it replicated at a high titre in the respiratory tract, but to a very low extent in the gut (Pensaert *et al.*, 1986). The significance of this virus as a pathogen is still unclear. Although it was first considered as non-pathogenic (Pensaert *et al.*, 1986), subsequent investigations have linked PRCV with field outbreaks of respiratory disease (Jestin *et al.*, 1987) and with pneumonia lesions following an experimental infection (Duret *et al.*, 1988; O'Toole *et al.*, 1989; Van Nieuwstadt & Pol, 1989). Conventional serological tests do not distinguish between PRCV- and TGEV-infected animals. However, antigenic dissimilarities were evident on examining the heterologous reactivity of monoclonal antibodies (MAbs). PRCV was found to be non-reactive towards several non-neutralizing anti-TGEV MAbs directed against the S (spike) or M (membrane) proteins (Callebaut *et al.*, 1988; Garwes *et*

al., 1988); neutralization-mediating epitopes unique to TGEV have been identified also (Laude *et al.*, 1988). Conversely, anti-PRCV S-specific MAbs have been shown to differentiate between the two viruses (O. Derriabine, R. L'Haridon, J. Gelfi & H. Laude, unpublished data). Examination of the structural polypeptide pattern of PRCV virions (TLM83 strain) shown by immunoblotting did not reveal any significant size discrepancy as compared to the TGEV Purdue strain (Callebaut *et al.*, 1988).

The molecular characterization of this newly recognized agent has been undertaken in the laboratory with the aim of gaining information with respect to two important questions: what is the phylogenetic origin of PRCV?, and which gene(s) control(s) the expression of the enteric tropism of TGEV? In the present study part of the genetic information of PRCV was determined and compared to available TGEV sequence data. We showed that although the known part of their genomes are nearly identical, one major difference exists in both a structural and a non-structural gene.

Methods

Virus and cells. The French isolate RM4 of PRCV at four passages in the swine testicle (ST) cell line (Duret *et al.*, 1988), supplied by Rhône-Mérieux Laboratoire IFFA, was used as a virus source. The virus was

adapted to the pig kidney (PDS) cell line (three passages), and then plaque-purified. Infected cultures were maintained in Dulbecco's MEM plus 5% foetal calf serum and incubated for 40 h at 38 °C. Infectivity titres measured in ST cells averaged 10⁶ p.f.u./ml.

RNA extraction. Clarified supernatant from the infected PDS cell culture medium was used as a starting material. The virions were semi-purified by ultracentrifugation through a 10 ml 25% glycerol cushion (Beckman rotor SW28, 60 min at 25000 r.p.m.). Pellets were treated with proteinase K plus 2% SDS and RNA was phenol-extracted as described (Rasschaert *et al.*, 1987). RNA material was precipitated overnight by 2 M-LiCl at -20 °C then washed twice with 80% ethanol and resuspended in double-distilled water.

cDNA cloning. The strategy and protocols were essentially as reported (Rasschaert *et al.*, 1987). Briefly, purified RNA was copied by reverse transcriptase (Moloney murine leukaemia virus, Bethesda Research Laboratories) using the TGEV-specific reverse primer pE2, located 47 nucleotides downstream of the S gene (5' CATCATCCT-TAACAAAATTCTCTAGCAGAA 3'). RNase T2-treated cDNA-RNA hybrids were dC-tailed and inserted into *Pst*I-cut dG-tailed pBR322. Transfection of *Escherichia coli* RR1 cells and screening of clones by colony hybridization with TGEV cDNA clones pTG2.18 and pTG2.26 (Rasschaert *et al.*, 1987), were performed using standard methods (Sambrook *et al.*, 1989). Two additional clones (designated pTVR) were obtained by insertion of PCR-amplified (see below) cDNA fragments into the *Sma*I site of the pTZ vector (Pharmacia) and transformation of *E. coli* NM522 cells.

Polymerase chain reaction (PCR). PCR-amplified fragments were obtained using cDNA-RNA heteroduplexes as templates and 20-mer oligonucleotides derived from the PRCV sequence. Taq DNA polymerase (Stratagene, 2 units) was employed in the conditions described by Saiki *et al.* (1988). Thirteen cycles (90 °C, 55 °C and 70 °C for 1 min each) were performed in a Hybaid apparatus (Ceralabo).

Supercoiled DNA sequencing. DNA matrices were prepared according to Lim & Péne (1988). Three µg of plasmid DNA was denatured by 0.15 M-NaOH plus 0.15 M-EDTA for 30 min at 37 °C, then precipitated overnight with ethanol and 0.2 M-ammonium acetate. Sequence reactions were performed using the Sequenase kit (USB) and synthetic primers (Biosearch 8600 apparatus).

Analysis of viral polypeptides. Radiolabelled cell extracts were prepared as described (Laude *et al.*, 1986) with the following modifications. ST cell monolayers were infected or mock-infected by PRCV at a multiplicity of 1 p.f.u. per cell and maintained in Eagle's MEM plus 2% foetal calf serum. At 3 h post-infection, [³⁵S]methionine was added at 100 µCi/ml. Fifteen h later the cells were scraped into RIPA lysis buffer and immunoprecipitation was performed using a mixture of three anti-TGEV MAbs: 51.13, 22.6 and 3.60 (Laude *et al.*, 1986). Resulting samples were analysed on a 9 to 20% polyacrylamide gradient SDS gel plus 5% 2-mercaptoethanol.

Results

Generation and mapping of the PRCV cDNA clones

PRCV genomic RNA extracted from semi-purified virions was reverse-transcribed after priming with a TGEV-specific sequence downstream of the S gene. Heteroduplex RNA-DNA material was then introduced into the pBR322 vector and cloned into *E. coli*. The resulting clones were probed against two TGEVcDNA

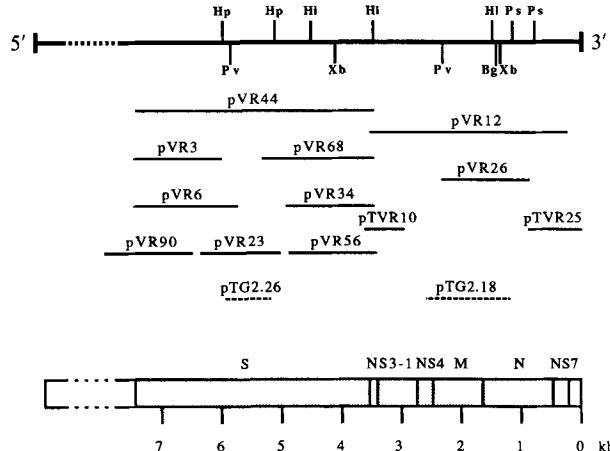


Fig. 1. Restriction and gene maps of PRCV genome 3' region. The positions of the cDNA clones used for sequencing are shown. pTG2.26 and pTG2.18 (dotted lines) are TGEV cDNA clones that were used for screening. Enzyme sites: *Bgl*II (*Bgl*), *Hind*III (*Hi*), *Hpa*I (*Hpa*), *Pst*I (*Ps*), *Pvu*II (*Pv*), *Xba*I (*Xb*).

clones to detect the S or the M and nucleocapsid (N) genes, respectively (Rasschaert *et al.*, 1987). The pVR12 and pVR26 clones arose from non-specific priming (Fig. 1). This appears to be the case also for the clones covering the S gene, as the TGEV reverse primer used (derived from Purdue virus sequence) was found to match imperfectly with the PRCV sequence, in particular at the 3' end. The clones pTVR10 and pTVR25 were obtained subsequently through PCR amplification of portions of the PRCV genome delimited using appropriate amplicons derived from partial sequencing. Restriction analysis of the pVR and pTVR clones permitted the construction of a continuous map covering approximately 7.5 kb of the PRCV genomic 3' region.

DNA sequencing and analysis

The insert of each of the clones specified in Fig. 1 was sequenced directly on plasmid supercoiled DNA and in both directions. Translation of the resulting 7519 nucleotide long sequence allowed the identification of six ORFs in the viral sense strand (Fig. 1; nucleotide and amino acid sequences displayed in Fig. 2). The largest three were predicted to encode the typical coronavirus structural polypeptides S, M and N. The other three were found to be homologous to TGEV ORFs encoding putative non-structural polypeptides, and were designated NS3-1, NS4 and NS7 as discussed later. The repeated sequence 5' CTAAAC 3' found upstream of each ORF, assumed to act as a signal for the transcription of subgenomic RNA in coronaviruses (review: Spaan *et al.*, 1988), is identical to that of TGEV (Rasschaert *et al.*, 1987). The extreme 3' non-coding

region comprised an 8 nucleotide sequence conserved in coronaviruses, and ended in a poly(A) stretch.

Pairwise alignment of the PRCV sequence data with those available for the three different TGEV strains Purdue 115 (Laude *et al.*, 1987; Rasschaert & Laude, 1987; Rasschaert *et al.*, 1987), FS772/70 (Britton *et al.*, 1988a, b, 1989) and Miller (Wesley *et al.*, 1989; Wesley, 1990) revealed a very high level of homology. Thus, 96% overall homology was found between PRCV and TGEV Purdue strains at both the nucleotide and amino acid level. However, several features unique to the PRCV genome were apparent. The most striking are three deletions affecting ORFs: (i) a 672 base in-frame deletion (position marked 1 in Fig. 2), located 59 bases downstream of the initiation codon of the S gene and (ii) two deletions which result in the alteration of the ORF situated immediately downstream of the S gene in the TGEV genome, a deletion of 38 bases (at position 2 in Fig. 2) removing both the CTAAAC sequence and the ATG initiation codon, and a deletion of 36 bases (at position 3) removing part of the body sequence. Moreover, two short deletions of two and six bases (positions 4 and 5) were seen in the PRCV 3' non-coding region by comparison to the Purdue and FS772/70 strain sequences (not yet available for the Miller strain). To summarize, the observed differences between PRCV and TGEV genomes consisted of deletions and point mutations only; no sequence unique to PRCV was found.

Estimation of M_r for PRCV S protein

The above sequence data predicted an M_r of 135K for the PRCV S apoprotein instead of the 158K for TGEV S. An experiment was designed to compare the relative size of PRCV and Purdue virus polypeptides by SDS-PAGE (Fig. 3). Both the PRCV S species and its intracellular precursor S' exhibited an increased mobility as compared to TGEV species S (220K) and S' (175K), formerly named E2 and E'2 (Laude *et al.*, 1986). An M_r of 190K was estimated for the PRCV S polypeptide. The two viruses showed no significant difference in M_r for either polypeptide N or M.

Discussion

In this study we cloned and analysed the 3' region of the genomic RNA of a French isolate of PRCV, a newly recognized porcine pneumotropic coronavirus which is antigenically closely related to TGEV but which does not induce intestinal lesions. The sequenced region, about 7500 nucleotides long, stretched from the beginning of the S glycoprotein gene up to the 3' poly(A) tail and probably encompasses all the viral genes except for the

polymerase gene. Thus a substantial part of the genetic information of PRCV was available for comparison to TGEV. It appeared that the genomes of the two viruses had not only a closely related organization, but exhibited also a very low level of nucleotide divergence (4%). However, as schematized in Fig. 4, PRCV and TGEV differed by two striking features.

First, the region homologous to the TGEV ORF NS3a contained no functional ORF (see Fig. 2 and Results). Additional detailed sequence data are needed to ascertain whether this is a common feature of PRCV isolates. However this view is supported by the recent report that the British isolate had the same ORF altered (P. Britton, *et al.*, IV International Coronavirus Symposium, Cambridge, July 1989). The NS3a ORF (NS3 for the Miller strain) has been shown to be transcribed for three different TGEV strains and has the potential to encode a non-structural polypeptide which remains to be characterized (Rasschaert *et al.*, 1987; Britton *et al.*, 1989; Wesley *et al.*, 1989). The efficient propagation of PRCV both *in vitro* and *in vivo* suggests that the expression of this gene is not an absolute requirement. Incidentally, alteration of an ORF encoding a putative NS polypeptide is not unprecedented in TGEV; the NS3b ORF from the Purdue strain propagated in our laboratory was reported to have a mutated first ATG codon and, at least in some clones, a 13 base frameshift deletion in the 3' region (Rasschaert *et al.*, 1987) when compared to the sequence from another laboratory (Kapke *et al.*, 1988).

Secondly, the genome of PRCV was found to encode a truncated S protein. Analysis of the S gene nucleotide sequence established from four cDNA clones predicted a 1225 long amino acid precursor polypeptide (see Fig. 2) instead of 1447 (Purdue virus; Rasschaert *et al.*, 1987) and 1449 amino acids (Miller and FS772/70 strains; Wesley, 1990; P. Britton, personal communication). The difference involves the omission in the PRCV S sequence of a continuous stretch of 224 residues starting at position 5 from the N terminus (Fig. 5). The sequence unique to TGEV S includes the 180 residue long segment where the antigenic site D has been mapped (Delmas *et al.*, 1990). This finding is consistent with the observed lack of reactivity of Purdue site D-specific MAbs towards three different PRCV isolates (Laude *et al.*, 1988). Further, it strengthens the view that the truncation of S protein, directly confirmed for the French isolate (Fig. 3), is a typical feature of this virus. Also of interest is the observation that one of the four MAbs defining site D (78.17) retained a significant reactivity towards PRCV, suggesting that the relevant epitope involves additional residues outside the TGEV-specific domain. Alignment of PRCV and TGEV S sequences revealed in addition 33 to 43 positions of non-identity, depending upon the PRCV-TGEV strain pair (Fig. 5).

ATGGATTACTAAGGAAGGTAAGTGTCTATTAGAAATAATGGTAAGTTA <u>C</u> <u>TAAAC</u> TTGGTAACCACCTTATTAAACACCATGAAAAAATTATTTGGTCTTGCGTAAATGCCATT	S	M	K	K	L	F	V	V	V	M	P	L	120
<u>1</u>													13
GATTATGGAGACAAGTT <u>C</u> <u>T</u> ACTTCGGTAGTTCCAATTGCACTGATCAATGTGCTAGTTATGGCTAATGTTTACTATACCTACAGGAGGCTTATACCATCAGATTTACTTT	I	Y	G	D	K	P	T	S	V	S	N	C	240
<u>C</u> <u>T</u> DQCAASVYVANVFTIILPGGFIPSPDFSF													53
TAATAATTGGTCTCTTAACATAATAGTCCACGGTGTAAATGGTAATTAGTACCAAAACAGCCTTATTAGTTAATTGCTTATGGCAGTCCCTAGCTTGAAGAAGTAGCTTAC	N	N	W	F	L	L	T	N	S	S	T	L	360
<u>N</u> <u>N</u> <u>W</u> <u>F</u> <u>L</u> <u>L</u> <u>T</u> <u>N</u> <u>S</u> <u>S</u> <u>T</u> <u>L</u> <u>V</u> <u>N</u> <u>G</u> <u>K</u> <u>L</u> <u>V</u> <u>T</u> <u>K</u> <u>Q</u> <u>P</u> <u>L</u> <u>L</u> <u>V</u> <u>N</u> <u>C</u> <u>L</u> <u>W</u> <u>P</u> <u>V</u> <u>P</u> <u>S</u> <u>F</u> <u>E</u> <u>E</u> <u>V</u> <u>A</u> <u>S</u> <u>T</u>													93
ATTTGGTTGAAGGTGTACTTGTATGCAATGTAATGGTGTAAATGGTAATTAGTACAGCTAGTGTGATTGGAGTTTCCAGTTACCGTGAATTCCGTCGGCTAACATGGAC	F	C	F	E	G	A	D	F	D	Q	C	N	480
<u>F</u> <u>C</u> <u>F</u> <u>E</u> <u>G</u> <u>A</u> <u>D</u> <u>F</u> <u>D</u> <u>Q</u> <u>C</u> <u>N</u> <u>G</u> <u>A</u> <u>V</u> <u>L</u> <u>N</u> <u>N</u> <u>T</u> <u>V</u> <u>D</u> <u>V</u> <u>I</u> <u>R</u> <u>F</u> <u>N</u> <u>L</u> <u>N</u> <u>E</u> <u>T</u> <u>T</u> <u>N</u> <u>V</u> <u>Q</u> <u>S</u> <u>G</u> <u>K</u> <u>G</u> <u>A</u> <u>T</u>													133
AGTGTTCATTGAACACAACGGGGGTGTCACTCTGAAATCTCATGTTAATGATACAGTGTGATTGGAGTTTCCAGTTACCGTGAATTCCGTCGGCTAACATGGAC	V	F	S	L	N	T	T	G	G	V	T	L	600
<u>V</u> <u>F</u> <u>S</u> <u>L</u> <u>N</u> <u>T</u> <u>T</u> <u>G</u> <u>G</u> <u>V</u> <u>T</u> <u>L</u> <u>E</u> <u>I</u> <u>S</u> <u>C</u> <u>Y</u> <u>N</u> <u>D</u> <u>T</u> <u>V</u> <u>S</u> <u>D</u> <u>S</u> <u>S</u> <u>F</u> <u>S</u> <u>S</u> <u>Y</u> <u>G</u> <u>E</u> <u>I</u> <u>P</u> <u>F</u> <u>G</u> <u>V</u> <u>T</u> <u>N</u> <u>G</u> <u>P</u>													173
ACGGTACTGTACGCTACTCTATAATGGCACAGCTTAAAGTATGTTACAGCTAGTGTGATTGGAGTTGCTTACCGTGAATTCCGTCGGCTAACATGGAC	R	Y	C	Y	V	L	Y	N	G	T	A	L	720
<u>R</u> <u>Y</u> <u>C</u> <u>Y</u> <u>V</u> <u>L</u> <u>Y</u> <u>N</u> <u>G</u> <u>T</u> <u>A</u> <u>L</u> <u>K</u> <u>Y</u> <u>L</u> <u>G</u> <u>T</u> <u>L</u> <u>P</u> <u>P</u> <u>S</u> <u>V</u> <u>K</u> <u>E</u> <u>I</u> <u>A</u> <u>I</u> <u>S</u> <u>K</u> <u>W</u> <u>G</u> <u>H</u> <u>F</u> <u>Y</u> <u>I</u> <u>N</u> <u>G</u> <u>Y</u> <u>N</u> <u>F</u>													213
CTTTCAGCACATTCTTATTGATGTATCTTTAATGGACTACTGGTGTAGTGTGCTCTGGCAATAGCTTACACATGCTACACTGAACGCTTACAGTGAAGTGGAAACACAGC	F	S	T	F	P	I	D	C	I	S	F	N	840
<u>F</u> <u>S</u> <u>T</u> <u>F</u> <u>P</u> <u>I</u> <u>D</u> <u>C</u> <u>I</u> <u>S</u> <u>F</u> <u>N</u> <u>L</u> <u>T</u> <u>T</u> <u>G</u> <u>D</u> <u>S</u> <u>D</u> <u>V</u> <u>F</u> <u>W</u> <u>T</u> <u>I</u> <u>A</u> <u>Y</u> <u>T</u> <u>S</u> <u>Y</u> <u>T</u> <u>E</u> <u>A</u> <u>L</u> <u>V</u> <u>Q</u> <u>V</u> <u>E</u> <u>N</u> <u>T</u> <u>A</u>													253
TATACAAATGTGACCTATTGTAATGTTAATACATTAATGCTCTAACACTGCTAATTGATAATGGATTATCTCTGTTCTCAAGTGAAGTGGTCTGTCAATAAA	I	T	N	V	T	Y	C	N	S	Y	V	N	960
<u>I</u> <u>T</u> <u>N</u> <u>V</u> <u>T</u> <u>Y</u> <u>C</u> <u>N</u> <u>S</u> <u>Y</u> <u>V</u> <u>N</u> <u>N</u> <u>I</u> <u>K</u> <u>C</u> <u>S</u> <u>Q</u> <u>L</u> <u>T</u> <u>A</u> <u>N</u> <u>L</u> <u>N</u> <u>N</u> <u>G</u> <u>F</u> <u>Y</u> <u>P</u> <u>V</u> <u>S</u> <u>S</u> <u>S</u> <u>E</u> <u>V</u> <u>G</u> <u>S</u> <u>V</u> <u>N</u> <u>K</u>													293
GAGTGTGTGTTACTACCTAGCTTCTGACACATACCCATTGTAACATAACTATGGCTTGTGATGAACGGCTAGTGGTTATGGTCAACCCATAGCCTAACGCTAACATGACT	S	V	V	L	L	P	S	F	L	T	H	T	1080
<u>S</u> <u>V</u> <u>V</u> <u>L</u> <u>L</u> <u>P</u> <u>S</u> <u>F</u> <u>L</u> <u>T</u> <u>H</u> <u>T</u> <u>I</u> <u>V</u> <u>N</u> <u>I</u> <u>T</u> <u>I</u> <u>G</u> <u>L</u> <u>G</u> <u>M</u> <u>K</u> <u>R</u> <u>S</u> <u>G</u> <u>Y</u> <u>G</u> <u>Q</u> <u>P</u> <u>I</u> <u>A</u> <u>S</u> <u>T</u> <u>L</u> <u>S</u> <u>N</u> <u>I</u> <u>T</u> <u>L</u>													333
ACCAATGCAGGATAACAAACACGATGTGTACTGTGTTCTGACCAATTTCAGTTATGTTACCTGACTTGTGAAAGTGGTAAAGCGAACATGTTTAAAGCGAACACTGACCGA	F	M	Q	D	N	N	N	D	V	Y	C	V	1200
<u>F</u> <u>M</u> <u>Q</u> <u>D</u> <u>N</u> <u>N</u> <u>N</u> <u>D</u> <u>V</u> <u>Y</u> <u>C</u> <u>V</u> <u>R</u> <u>S</u> <u>D</u> <u>Q</u> <u>F</u> <u>S</u> <u>V</u> <u>V</u> <u>H</u> <u>S</u> <u>T</u> <u>C</u> <u>K</u> <u>S</u> <u>V</u> <u>L</u> <u>W</u> <u>D</u> <u>N</u> <u>V</u> <u>F</u> <u>K</u> <u>R</u> <u>N</u> <u>C</u> <u>T</u> <u>D</u>													373
CGTTTATGATGCCACAGCTGTTATAAAACTGGTACTCTCTTCTCATTTGATAATTGAAACATTACTTAAACAATGGTCTGTTGCTGTTGAGTCCGTTGGCTAATTG	V	L	D	A	T	A	V	I	K	T	G	T	1320
<u>V</u> <u>L</u> <u>D</u> <u>A</u> <u>T</u> <u>A</u> <u>V</u> <u>I</u> <u>K</u> <u>T</u> <u>G</u> <u>T</u> <u>C</u> <u>P</u> <u>F</u> <u>S</u> <u>F</u> <u>D</u> <u>K</u> <u>L</u> <u>N</u> <u>N</u> <u>Y</u> <u>L</u> <u>T</u> <u>F</u> <u>N</u> <u>K</u> <u>F</u> <u>C</u> <u>L</u> <u>S</u> <u>L</u> <u>S</u> <u>P</u> <u>V</u> <u>G</u> <u>A</u> <u>N</u> <u>C</u>													413
TAAGTGTGATGTAGCTGCCGTAACGAAACCAATGAGTGTGTTGATGAAGAAGGAGAGACACGATAGTGTGTTGACCGTGTACACATGAGTGGTTGCACAGA	K	F	D	V	A	R	T	R	T	N	D	Q	440
<u>K</u> <u>F</u> <u>D</u> <u>V</u> <u>A</u> <u>R</u> <u>T</u> <u>R</u> <td></td> <td>453</td>													453
TTTGTGACTGCTACACCTAGTCGTCACAGATTACATATATGGTAACTGGTGTGTTATTAGACAAACTAACAGGACGATACTTAGTGGCTTATATTACACATCACTATC	L	S	V	L	H	L	D	S	C	T	D	Y	1560
<u>L</u> <u>S</u> <u>V</u> <u>L</u> <u>H</u> <u>L</u> <u>D</u> <u>S</u> <u>C</u> <u>T</u> <u>D</u> <u>Y</u> <u>N</u> <u>I</u> <u>Y</u> <u>G</u> <u>R</u> <u>T</u> <u>G</u> <u>V</u> <u>G</u> <u>I</u> <u>I</u> <u>R</u> <u>Q</u> <u>T</u> <u>N</u> <u>R</u> <u>T</u> <u>I</u> <u>L</u> <u>S</u> <u>G</u> <u>L</u> <u>Y</u> <u>T</u> <u>S</u> <u>L</u> <u>S</u>													493
TGGTGTAGTTACAAATGTTAGGTGTTACTCTGTAACCCAGTGTGATTTAGGGCAGACAGCTTACATGTTGACATGTTGTTGCTATCAGTGGCTTACCTTCAT	G	D	L	L	G	F	T	N	V	S	D	G	1680
<u>G</u> <u>D</u> <u>L</u> <u>L</u> <u>G</u> <u>F</u> <u>T</u> <u>N</u> <u>V</u> <u>S</u> <u>D</u> <u>G</u> <u>V</u> <u>I</u> <u>Y</u> <u>S</u> <u>V</u> <u>T</u> <u>P</u> <u>C</u> <u>D</u> <u>V</u> <u>S</u> <u>A</u> <u>Q</u> <u>A</u> <u>I</u> <u>D</u> <u>G</u> <u>T</u> <u>I</u> <u>V</u> <u>G</u> <u>A</u> <u>I</u> <u>T</u> <u>S</u> <u>I</u>													533
TAACAGTGAATTGTTAGGTCTAACATGGACACAAACACCTAATTTCATGTTGCTTACACTACTGCTTACATACCCAGTGTGAAAGACTCTGCTGACTCCATTGGCAGTATGACGTTGA	N	S	E	L	L	G	L	T	H	W	T	T	1800
<u>N</u> <u>S</u> <u>E</u> <u>L</u> <u>L</u> <u>G</u> <u>L</u> <u>T</u> <u>H</u> <u>W</u> <u>T</u> <u>T</u> <td></td> <td>573</td>													573
TTGTGAACCTGCTATAACCTTACACATGGTGTGTTGGTTTATTAACGTCACACATTCTGTGATGGACAGCTGCAACCAATTAGCAGTGGTAAACGCTACAGAT	C	E	P	V	I	T	Y	S	N	I	G	V	1920
<u>C</u> <u>E</u> <u>P</u> <u>V</u> <u>I</u> <u>T</u> <u>Y</u> <u>S</u> <u>N</u> <u>I</u> <u>G</u> <u>V</u> <u>C</u> <u>K</u> <u>N</u> <u>G</u> <u>A</u> <u>L</u> <u>V</u> <u>F</u> <u>I</u> <u>N</u> <u>V</u> <u>T</u> <u>H</u> <u>S</u> <u>D</u> <u>G</u> <u>D</u> <u>V</u> <u>Q</u> <u>P</u> <u>I</u> <u>S</u> <u>T</u> <u>G</u> <u>N</u> <u>V</u> <u>T</u> <u>I</u>													613
ACCTACTAACCTTACATATCCGTCGAAGTCAACATTAGGTTACACTACACCCAGTGTCAAGACTGTTCAAGATGTTGTAATGGCAACCCCTAGGTGTAACAAACTGTTAC	P	T	N	E	T	I	S	V	Q	V	E	Y	2040
<u>P</u> <u>T</u> <u>N</u> <u>E</u> <u>T</u> <u>I</u> <u>S</u> <u>V</u> <u>Q</u> <u>V</u> <u>E</u> <u>Y</u> <u>I</u> <u>Q</u> <u>V</u> <u>Y</u> <u>T</u> <td></td> <td>653</td>													653
ACAATACGTTCTGCAATGCTAACACTTGGCAAGCAGCAGCTTGGCAATGGGCCAGACTGAAAACATGGAAGTGTGATTGCTTGTGAAATTGCCCCTAAATTGGCTTCTGT	Q	Y	V	S	A	C	Q	T	I	E	Q	A	2160
<u>Q</u> <u>Y</u> <u>V</u> <u>S</u> <u>A</u> <u>C</u> <u>Q</u> <u>T</u> <u>I</u> <u>E</u> <u>Q</u> <u>A</u> <u>L</u> <u>A</u> <u>M</u> <u>G</u> <u>A</u> <u>R</u> <u>L</u> <u>E</u> <u>N</u> <u>M</u> <u>E</u> <u>V</u> <u>D</u> <u>S</u> <u>M</u> <u>L</u> <u>F</u> <u>V</u> <u>S</u> <u>E</u> <u>N</u> <u>A</u> <u>L</u> <u>K</u> <u>L</u> <u>A</u> <u>S</u> <u>V</u>													693
CGAACGATTCATAGTTCGAGAAACCTTGGATCTTACACATGGCAACAGGATGGCTTACATGGCTGAGCTGCAACACATGGCTGAGTGGCTTACAGTGGTAAACGCTACAG	E	A	F	N	S	S	E	T	L	D	P	I	2280
<u>E</u> <u>A</u> <u>F</u> <u>N</u> <u>S</u> <u>S</u> <u>E</u> <u>T</u> <u>L</u> <u>D</u> <u>P</u> <u>I</u> <td></td> <td>733</td>													733
TGCTTACGCTATAGGGACTTGTGTTCTGATGTTGCTTACACTGCTGTTAGCTGTTAGGGCTTACAGTGTGTTAGCTGCTGACTTGTGCTGTTGCTGTTGCTA	R	S	A	I	E	D	L	L	F	S	K	V	2400
<u>R</u> <u>S</u> <u>A</u> <u>I</u> <u>E</u> <u>D</u> <u>L</u> <u>L</u> <u>F</u> <td></td> <td>773</td>													773
ATACATAATGGCATTAGGTCTGGCTTAATGGCTGACAAACATGACTGATCACAGCATCCCTCGCAGGGCTTAAACATTAGGTGCACTTGTGTTGAGGCGCCCTGGCTAT	Y	N	G	I	M	V	L	P	G	V	A	N	2520
<u>Y</u> <u>N</u> <u>G</u> <u>I</u> <u>M</u> <u>V</u> <u>L</u> <u>P</u> <u>G</u> <u>V</u> <u>A</u> <u>N</u> <u>D</u> <u>K</u> <u>M</u> <u>T</u> <u>M</u> <u>Y</u> <u>T</u> <u>A</u> <u>S</u> <u>L</u> <u>A</u> <u>G</u> <u>G</u> <u>I</u> <u>T</u> <u>G</u> <u>G</u> <u>A</u> <u>V</u> <u>A</u> <u>I</u>													813
ACCTTTGGCACTGGCTAGCTTGGCTAGCTTAAACTGGTGTGTTGATGAACAAAACAGCAGATCTGGCTAGTGTGTTTAATCAAGGTTACATTGGTAACATACACAC	P	F	A	V	A	V	Q	R	L	N	Y	V	2640
<u>P</u> <u>F</u> <u>A</u> <u>V</u> <u>A</u> <u>V</u> <u>Q</u> <u>R</u> <u>L</u> <u>N</u> <u>Y</u> <u>V</u> <u>A</u> <u>L</u> <u>Q</u> <u>T</u> <td></td> <td>853</td>													853
GTCTTGGTAACGTTATGATGCTATACATCAACATCAGGGCTTACAACTGTTGCTAAAGGCTTACAGTGTGATGTCAGACACAGGCTGATCACAGGAAGACTTAC	S	F	G	K	V	N	D	A	I	H	Q	T	2760
<u>S</u> <u>F</u> <u>G</u> <u>K</u> <u>V</u> <u>N</u> <u>D</u> <u>A</u> <u>I</u> <u>H</u> <u>Q</u> <u>T</u> <u>S</u> <u>R</u> <u>G</u> <u>L</u> <u>T</u> <u>T</u> <u>V</u> <u>A</u> <u>K</u> <u>L</u> <u>V</u> <u>Q</u> <u>D</u> <u>V</u> <u>V</u> <u>N</u> <u>T</u> <u>Q</u> <u>G</u> <u>Q</u> <u>A</u> <u>L</u> <u>R</u> <u>H</u> <u>L</u>													893
AAACAGTACAATTGCAAAATAATTCCGACCTAGTGTCTATTAGTACATTATAATAGGTTGATGAATGAGTGTGATGTCAGACAGGCTGATCACAGGAAGACTTAC	T	V	Q	L	Q	N	F	Q	A	I	S	S	2880
<u>T</u> <u>V</u> <u>Q</u> <u>L</u> <u>Q</u> <u>N</u> <u>F</u> <u>Q</u> <u>A</u> <u>I</u> <u>S</u> <u>S</u> <u>I</u> <u>S</u> <u>D</u> <u>I</u> <u>Y</u> <u>N</u> <u>R</u> <u>L</u> <u>D</u> <u>E</u> <u>L</u> <u>S</u> <u>A</u> <u>D</u> <u>A</u> <u>Q</u> <u>V</u> <u>D</u> <u>R</u> <u>L</u> <u>I</u> <u>T</u> <u>G</u> <u>R</u> <u>L</u> <u>T</u>													933
ACCACTTATGCAATTGGCTACACTCTGCAACAGAGCCAGGGTTAGGGCTTAGACAGCAACTTGGCTAAAGGCTTAATGAATGGCTTAGGGCTAGCTGCTAGAGATTGGCTT	A	L	N	A	F	V	S	Q	T	L	R	Q	973
<u>A</u> <u>L</u> <u>N</u> <u>A</u> <u>F</u> <u>V</u> <u>S</u> <u>Q</u>													
CTGGTGTGTTAGGTACACATTGGCTTACCTGGCAATGGCAGGACAAATGGCTGATGTTCTCTTACACAGTGTGCTTACCAACGGCTGATGAAACTGTGACTGCTGGTCAAGCTT	C	G	N	G	T	H	L	F	S	L	A	N	3120
<u>C</u> <u>G</u> <u>N</u> <u>G</u> <u>T</u> <u>H</u> <u>L</u> <u>F</u>													1013
TGCTTGTAGTTGACTGGCACTTTGGCTTAAAGATGGCTAGTGTGTTAGTCTGCTAATCTAGTACAGTGTGTTAGTCTGCTTACCTGGCTAGTGTGTTAGTGTGCTAG	A	L	D	V	D	R	T	F	G	L	V	V	3240
<u>A</u> <u>L</u> <u>D</u> <u>V</u> <u>D</u> <u>R</u> <u>T</u> <u>F</u>													1053
AACTAGTCTGTTGCAATTGGCTGATGGCTGTTGTTAAACACTGTAACAGTGTGTTAGTCTGCTAGTATTACCTGTTGATGTTGCTGTTAGTCTGTTGCTGTTAGTGTGCTAG	T	S	S	D	F	V	Q	I	E	G	C	D	3360
<u>T</u> <u>S</u> <u>S</u> <u>D</u> <u>F</u> <u>V</u> <u>Q</u> <u>I</u> <u>E</u> <u>G</u> <u>C</u> <u>D</u> <u>V</u> <u>L</u> <u>F</u>													1093
ATTAGAAAATTTAGGCAAAATTGGACTGACTCTGAGCTGACATGGACGTTTAAACGCAACCTTAAACCTGACTGGTGAATGAGTTAGGTGCAAGAAAGCTACA	L	E	N	F	R	P	N	W	T	V	P	E	3480
<u>L</u> <u>E</u> <u>N</u> <u>F</u> <u>R</u> <u>P</u> <u>N</u> <u>W</u> <u>T</u> <u>V</u> <u>P</u> <u>E</u>													1133
AAACACTACTGAGAACTTGGCCATTCTGCAACATTAACAACTAGTGTGACTGATGGCTTAATAGGCTTAATAGGAAACTTATGTTGCTAGGCTTACATGGCTAG	N	T	T	V	E	L	T	D	V	F	N	A	3600
<u>N</u> <u>T</u> <u>T</u> <u>V</u> <u>E</u> <u>L</u> <u>T</u> <u>D</u> <u>V</u> <u>F</u>													1173
AGGCTTACTGAGTAATTTGCTACCTGAGCTGACTGTTGCTGTTAGTACAGGTTGCTGAGCTGATGGCTAGTGTGACTGCTTACATGGCTAGGCTTACATGGCTAG	G	L	V	V	I	F	C	I	P	L	F	C	3720
<u>G</u> <u>L</u> <u>V</u> <u>V</u> <u>I</u> <u>F</u> <u>C</u> <u>I</u> <u>P</u> <td></td> <td>1213</td>													1213
TGAAAGATTAACGTTGAGGAGTGTGCACTGGCTTACGGCTTACATGGCTGTTAGTGTGACTGCTTACATGGCTAGGCTTACATGGCTAGGCTTACATGGCTAG	A	E	N	Y	E	P	I	E	K	V	H	V	3840
<u>A</u> <u>E</u> <u>N</u> <u>Y</u> <u>E</u> <u>P</u> <td></td> <td>1225</td>													1225
TTACTAATACATGGCTGAGCTGACTGGCTGTTGCTGTTAGTGTGACTGCTTACGGCTTACGGCTTACATGGCTAGGCTTACATGGCTAGGCTTACATGGCTAG	T	T	S	S	D	F	V	Q	I	E	G	D	3960

3	B	
TTCTTGCGGTTATTGAAGAAGTAAACACACAAAATCCAAGCATTAAAGTGTACAAAACAATAAAGAGAGACTATAGAAAAACTCGA <u>CTAAAC</u> <u>TTTG</u> TGAAAATGATTGGTGGACTTT NS3-1 M I G G L 5		
TTCTTAATACTCTGAGTTTGTAAATGTTAGTAAACATCCTATCGTTAAACACAGCAAATGTGCATCATACAAACAAGAACGCTTATAGTACAAACACCACATGTTGTTAGTGTCA F L N T L S F V I V S N H P I V N N T A N V H H I Q Q E R V I V Q Q H H V V S A 45		
GAACACAAAACATTACCCAGACTTCAGCATCGCTGACTTTTGATCTTCTAGCTTGTACCGTACTACAAACCTTAAGACGTGTGCGCATCTAAATGTTAAGATTTATCAA R T Q N Y Y P E F S I A V L F V S F L A L Y R S T N F K T C V G I L M F K I L S 85		
TGACACTTTAGGACCTATGCTTACGGTTACTACATTGATGGCATTGTTACACAAACTGCTTAAAGATTGCCACTTAGCATACTTTGTTATGTTAAGTAGGT M T L L G P M L I A Y G Y I D G I V T T T V L S L R F A Y L A Y F W Y V N S R 125		
TTGAAATTATTTATACAAACGACACTCATGGTGTACATGGCAGAGCTGCACCGTTAAAGAGAAGTCTCACAGCTCTATTATGTCACATTGTCATGGTCATAAAATTATGTT F E F I L Y N T T T L M F V H G R A A P F K R S S H S S I Y V T L Y G G I N Y M 165		
TTGTAATGACCTCACGTTGATTTGTAGACGCCATGCTGTAAGCATGCAACATGTTGCTAGCTCATGCTATCTAAGTAGGAGCTTCATGGTCATAAATGTCATGGTCATAAA F V N D L T L H F V D P M L V S I A I R G L A H A D L T V V R A V E L L N G D F 205		
TTTATGTTATTCACAGGCCGTTAGCTGGTTAACATGCAGCCTTCTCAGCGGTT <u>CTAAAC</u> GAATTGACTTAAAGAAGAGGGAGACCTTATGCGTTCTAGG I Y V F S Q E P V V G V Y N A A F S Q A V L N E I D L K E E E G D R T Y D V S 244		
GCATTGACTGTCATAGATGACAACGGATGGTCAATTAGCATTATTTGGTTCTCTGTTGATAATTATATTGATATTCTAACATGCTTAAAGTCTAACATTAAAGCTATGCTG A L T V I D D N G M V I S I I F W F L L I I I L L S I A L L N I I K L C M V 45		
TGTCGAATTAGGAAGGAGCACATTATTGTCAGTCACATGTCAGCTTAAAGATTTATGCAATTAAAGCATAACCCGTATGGGACTCCCTGTTGAA <u>CTAAAC</u> C C N L G R T V I V P V Q H A Y D A Y K N F M R I K A Y N P D G A L L V 82		
<u>M</u> AAATGAAGATTGTTGATATTAGCGTGTGCGATTGCAATGTCAGCAGAACGCTATTGCTATGAAAGACGATACAGGTTGCTATGCGCAATGGCACGGCTGTGTTGA M K I L L I L A C A I A C T C G E R Y C A M K D D T G L S C R N G T A S D C E 39		
GTCATGCTCAACAGAGGCATCTATTGGCTCTGGCAACTGGACTCATGGCTATAATTGATCTTTTATTACTGTCATGAAATGGAGACCTCAATTAGCTGGTT S C F N R G D L I W L L A N W N F S W S I I L I I F I T V L Q Y G R P Q F S W F 5280		
CGTGTATGGCATTAATGCTTAAATGTCGTTATATGGCGATTGTTGGCTTACGATTTTAATGCAATCTGGAAATACCAAGTGTCCAGGTATGTAATGTCGGTTAGTAT V Y G I K M L I M W L L W P I V L A L T I F N A Y S E Y Q V S R Y V M F G F S I 119		
TGCAGGTGCAATTGTCATTTGACTCTGGATTATGTTGTAAGATCCATTGCTAGCAGAGACTAATGTTGCTTCACCCCTGAAACTAACCGAATTCTTGGCT A G A I V T F V L W I M Y F V R S I Q L Y R R T K S W W S F N P E T N A I L C V 159		
TAGTGCATTAGGAAGAACGATGTGCTTCTCGAAGGTGCGCAACTGGTGTACTCTAACATTGCTTCAGGGATTGTCAGCTGAAGGGTCAAAATTGCAAGGTGATGACCAT S A L G R S Y V L P L E G V P T G V T L T L L S G N L Y A E G F K I A G G M T I 199		
CGACAATTGCAAAATACGTAATGGTGTACCTACCCAGCAGGACTATTGTTACACACTGGCAAGAAGTGAAGCAACTAGTGTGACTGGATGGCTTACTATGAA D N L P K Y V M V A L P S R T I V Y T L V G K K L K A S S A T G W A Y Y V K S K 5760		
AGCTGGTGAATCTAACAGAGGCAAGAACTGATAATTGAGTGAAGAAAAATTATTACATATGGTATA <u>CTAAAC</u> TTCTAAATGGCAACAGGACAAGTGTCA A G D Y S T E A R T D N L S E Q E K L L B M V N M A N Q G Q R V S W G D 5880		
TGAATCCACCAAAATACGTGGTGCCTCAATTCCCGTGGCGAGATTAAACATACCTCTTCATTCTCAACCCATAACCCCTCAGCAAGGTCAAATTGGAAACTCATGTC E S T K I R G R S N S R G R K I N N N I P L S F F N P I T L Q Q G A K F W N S C P 6000		
GAGAGATTGTCACCAAAAGGAATAGGTAATACGGATCACAGATGGTTATTGGAATAGCACAAACTCGCTATCCCATGGTCAAGGGCCAACGTAACAGGCTCTGAAAGGTGCTT R D F V P K G I G N R D Q Q I G W N R Q T R Y R M V K G Q R K E L P E R W F F 6120		
TTACTACTGGACTGGACCTCATGCGATGCCAAATTAAAGATAATTAGATGGAGTTGCTGGCAAGGATGGTGCATGACAAACCAACACGGCTTGGTAGCTGG Y L G T T G P H A D A K F D K L D G V V W V V A K D G A M N K P T T L G S R G A 6240		
TATAATGAATCCAAGGTTGAAATTGATGTTAAAGTGCAGGCCAATTCAACTTGAAGTTAACAGCTTGGGACAACCTAACGGTCAAGCTCAACTTAGATGGCTCTAGAAA N N E S K A L K F D G K V P G E F Q L E V N Q S R D N S R S R S Q S R S R S R N 6360		
CAGATCTCAATCTAGGGTAGGCACAACTCCAATAACAGAAGGGATGACAGTGTAGAACAGCTTGGCCACTAAAAAGTGGTTAACACAGAAAAACACGCAACGCTC R S Q S R G R Q Q S N N K D D S V E Q A V L A A L K K L G V Y T E K Q Q Q R S 6480		
TCGTTCTAAATCTAAAGAACGTAACCTAACAAAGAGATACTACGCCAACGAAATGAAACACACCTGGAGAGAACACTGCAAGGTAAAGGTGATGTGACAAGATT R S K S K E R S N S K T R D T T P K N E N K H T W K R T A G K G D V T R F Y G A 252		
TAGAAGCAGCTGCCAATTGGTGCAGTGCCTCGTCCGCAATGGGAGCAGTGCAGCAAGCATTACCCACAATTGGCTGAATGCTCCATCTGTCAGCATTTGTTGGAA R S S S A N F G D S D L V A N G S S A K H Y P Q L A E C V P S V S S I L F G S Y 6720		
TTGGACTTCAAAGGAAGATGGCAGCACAGATAAGTCAGTTACACACAAATACCAACTGGCAAGGATCATCTAAACTGAAACAATTCTTCAGCAGATTATGCCATGCTAGCCC W T S K E D G D Q I E V T F T H K Y H L P K D H P K T E Q F L Q Q I N A Y A S P 6840		
ATCAGAATTGGCAAAAGAACAGAGAAAAGAGAAGTCTCGTTCTAAATCTGCAGAAAGGTCAAGGAGATGGTGTACCTGATTAATAGAAAATACAGATGTTGATGACAC S E L A K E Q R K R K S R S K S A E R S E Q E V V P D S L I E N Y T D V F D D T 6960		
ACAGGGTTGAGATGATTGACGAGGTAACGA <u>CTAAAC</u> AGATGCTCGCTCCCTCCAGGCTGTATGTTACAGTTAACCTTACTAAATTGGTAGACTCCAA Q V E M I D E V T N NS3-1 M L V L L Q A V C I T V L T L L L I G R L Q L L E R L 7080		
TTACTTAATCACTCTTCAATCTAAACTGTTGATTTTAATCTATAGGAGTTAGCAGAAACTAGATTACTAAAGTGGCTCTGATTAATCTTTTAGTCTTATTAGGA L L N H S F N L K T V D D F N I L Y R S L A E T R L L K V V L R L I F L V L L G 7200		
TTTGGCTGCTATGAGTTGTTAGCTATTAATGAGCAAGGCAACCCGATG <u>CTAAAC</u> TTGGTTCTCGAGGAAACTCTGGTACATGGCTACTCTGTCAGAGTGTAAAGC F C C Y R L L I M 7320		
ATAGGAGGTACAAGCAACCTATTGCATATTGAGTTAGATGCAATGCTAGATTAGTAATTAGACAAGTTAAAGATCCGCTACGACGAGCAACATGGAAGGCTAACGT 5 7460		

CTGGATCTAGTGTATTGTTAAATGTAATTGTTGAAATTTCCTTGTAGTGTACACAAAAA

7519

Fig. 2. Sequence of the extreme 3' 7519 nucleotides of the PRCV genome and deduced amino acid sequences of the major ORFs. The positions of the deletions relative to three TGEV strains (see Results) are indicated by the framed numbers 1 to 5 above the nucleotide sequence. The framed letters A and B show the positions of two sequence segments unique to the Purdue strain (see Discussion). The consensus intergenic sequences are boxed. The octanucleotide sequence conserved in the 3' non-coding region of coronaviruses is overlined. The predicted ORFs are translated into the one-letter amino acid code. The 29 potential Asn-linked glycosylation sites found in protein S are underlined. Protein S and M signal peptides are indicated by bold lines.

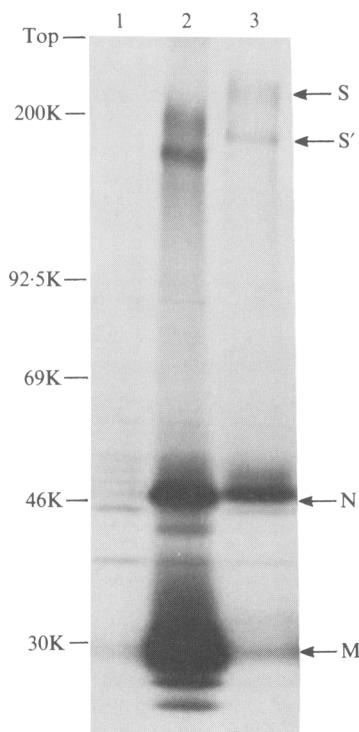


Fig. 3. Comparative gel electrophoresis of PRCV and TGEV intracellular polypeptides. Cytosols from [³⁵S]methionine-labelled infected ST cells were subjected to immunoprecipitation using a mixture of MAbs against Purdue strain (TGEV) structural polypeptides. Immunoprecipitated material was resolved in a 9 to 20% acrylamide gradient SDS-PAGE gel. Lane 1, mock-infected cells; lane 2, PRCV-infected cells (10^5 c.p.m.); lane 3, TGEV-infected cells (10^4 c.p.m.). M_r of marker proteins are indicated in the left margin.

In particular, one change, Ser to Phe, was present within the nine residue stretch expressing C epitopes (Delmas *et al.*, 1990); this correlates with the failure of two site C-specific MAbs to recognize the PRCV S antigen (Laude *et al.*, 1988). By contrast to site D- and C-specific MAbs, all site A- and B-specific MAbs exhibited a strong cross-reactivity towards PRCV S, despite the existence of 12 amino acid changes in the region stretching from positions 300 to 512, where the relevant epitopes have been mapped (Delmas *et al.*, 1990). This further supports our earlier assumption that biologically crucial determinants are expressed in this domain of the S molecule (Delmas *et al.*, 1986).

The nature of the events responsible for the genomic diversity between PRCV and TGEV remains an open question. Both homologous and heterologous RNA recombinations are proposed to be a driving force for coronavirus evolution (see review by Spaan *et al.*, 1988). Thus the possibility cannot be formally excluded that the sequences unique to TGEV result from insertional events. However, we believe that the idea that PRCV is derived from TGEV through successive deletion events

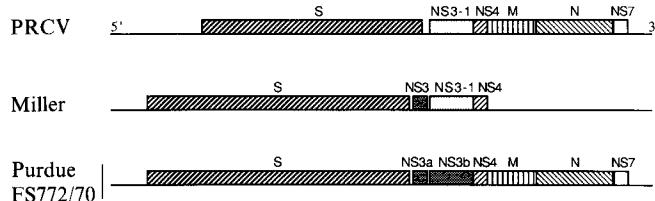


Fig. 4. Compared gene organization of PRCV and of three TGEV strains. The genes are designated according to the latest recommendations of the coronavirus study group (Cavanagh *et al.*, 1990). In particular, the second NS ORF downstream of the S gene is referred to as NS3-1 in the case of the Miller strain to acknowledge the fact that its transcription involves a distinct subgenomic RNA species (Wesley *et al.*, 1989), in contrast to NS3b ORFs of Purdue and FS772/70 viruses. The PRCV ORF named NS3-1 is preceded by an intergenic sequence homologous to that of the Miller ORF NS3-1. The TGEV ORFs NS3a, NS3b, NS4 and NS7 were formerly named X2a, X2b, X1 and X3, respectively (Rasschaert *et al.*, 1987). The diagram was drawn according to the sequence data cited in the Results section.

is the most plausible. Recent data indicate that deletions might be responsible for the observed polymorphism of the S protein of mouse hepatitis virus (MHV). In particular, the S genes of three selected MAb-resistant variants of MHV type 4 have been shown to encode truncated S proteins with a large deletion of 142 to 159 amino acids mapping approximately in the N-terminal two-thirds of the S1 subunit of S protein (Parker *et al.*, 1989). Besides, the S protein genes of two MHV strains, JHM (Schmidt *et al.*, 1987) and A59 (Luytjes *et al.*, 1987), harbour a sequence deleted in the same region relative to that of MHV-4. Examination of the TGEV nucleotide sequence did not reveal any feature indicative of a jumping of the RNA polymerase complex such as a site of strong secondary structure, or repeated sequences in the vicinity of the deletion site. The only noticeable feature was the presence of a heptanucleotide sequence AGTTTCC immediately adjacent to the unique sequence of the TGEV S gene which is repeated 10 nucleotides downstream; however, no simple mechanism underlying a possible deletion process could be reconstructed on this basis.

Another conclusion drawn from the comparison of PRCV and of TGEV strains sequences was a closer relationship of PRCV to the fully virulent Miller and FS772/70 strains than to the high passage Purdue strain, of which several sequence features were not seen in PRCV: (i) a six nucleotide deletion in the S gene, encoding the residues Asn and Asp at positions 151 and 152 (asterisks in Fig. 5) and (ii) two short insertions of 16 and 29 nucleotides located upstream and at the 3' end of the NS3a ORF, respectively (corresponding positions in PRCV sequence noted A and B in Fig. 2). In addition, the PRCV ORF NS3-1 is immediately preceded by a sequence CTAAAC, as in the Miller strain (Wesley *et al.*, 1989). Finally, pairwise comparisons of the different

224 aa unique to TGEV

MKKLFVVLLVMPLIYGDKFPTSVVSNCQASVANVFTILPGGFIPSDFSFNNWFLLTNSSTLVNGKLVTQPLLVNC	80
N T	
TQ	
S	
LWPVPSFEEVASTFCFEGADFDQCNMGAVLNNTVDVIRFNLNFTTNVQSGKATVFSLNTTGGVTLEISCYNDTVSDSSFS	160
A L G P	
A G	
** F	
W D	
SYGEIPFGVTNGPRYCYVLYNGTALKLGLPPSVKEIAISKWGHFYINGYNFFSTFP IDCISFNLTGDSDFWTIAYT	240
D H	
SYTEALVQVENTAITNVTCNSYVNNIKCSQLTANLNNGFYPVSSSEVGSVNKSVVLLPSFLTHTIVNITIGLGMKRSGY	320
K H I L Y	
GQPIASTLSNITLPMODNNNDVYCVRSDQFSVYVHSTCKSVLWDNVEKRNCTDVL DATAVIKTGTCPFPSFDKLNNYLTFN	400
H T I A I	
KFCLSLSPVGANCKFDVAARTRTNDOVVRSLYVIYEEGDSIVGVPSDNSGLHDLSVLHLDSCTDYNIYGRTGVCILRQTN	480
E N V	
RTILSGLYYTSLSGDLGFTNVSDGVIYSVTPCDVSAQAAIIDGTIVGAITSINSELLGLTHWTTPNFYYSIYNYTND	560
L K V	
R A D	
KTRGTPIGSNVDCEPVITYSNIGVCKNGALVFINVTHSDGDVQPISTGNVTIPTNFTISVQVEYIQVYTPVSIDCSRY	640
R A D F	
VCNGNPRCNKLLTQYVSACQTIEQALAMGARLENMEVDSMLFVSENALKLASVEAFNSSETLDPIYKEWPNIIGFWLEG	720
G S	
S	
KYILPSDNSKRKYRSAIEDLLFSKVVTSGLGTVDDEDYKRCTGGYDIADLVCAQYYNGIMVLPGVANADKMTMYTASLAGG	800
H D	
A	
ITLGALGGGAVAIPFAAVQARLNVALQTDVLNKQQILASAFNQAIGNITQSFGKVNDAIHQTSRGTLTVAKALAKVQ	880
I S	
DVVNTQGQALRHLTQQLQNNFQAISSISDIYNRLDELSADAQVDRLLITGRLTALNAFVSQTLTRQAEVRASRQLAKDKV	960
I S	
AP	
NECVRSQSQRFGFCGNGLHLFLANAAPNGMIFHTVLLPTAYETVTAWSGICALDVDRTEGLVVKDVQLTLFRNLDDKF	1040
P S G	
A F I	
YLTPRTMYQPRVATSSDFVQIEGCDVLFVNTPSDLPSIIPDYIDINQTVQDILENFRPNWTVPELTLDVFNATYLNLTG	1120
A F I	
L	
EIDDLEFRSEKLHNTTVELAILIDNINNTVNLWLNRIETYVKWPWYWLLIGLVVIFCIPLLLFCCTGCCGICGCL	1200
L	
C	
GSCCHSIFSRRQFENYEPIEKVVHV	1225
C	

Fig. 5. Alignment of the S protein sequences of PRCV and two TGEV strains. The position of the segment omitted in the PRCV sequence is shown. Amino acid changes relative to Miller and to Purdue strains are indicated above and beneath the PRCV sequence, respectively. The asterisks mark two amino acids absent in the Purdue S protein. The location of the major TGEV antigenic sites C (open box) and A and B (filled box) is given (Delmas *et al.*, 1990).

Table 1. Pairwise sequence homology between PRCV and TGEV polypeptides

PRCV predicted polypeptides	Number of amino acids	TGEV strain					
		Purdue		Miller		FS772/70	
Designation		aa*	%†	aa	%	aa	%
S	1225	44	97	34	97.3	-	-
M	262	13	95	-‡	-	11	95.8
N	382	14	96.3	-	-	9	97.6
NS3-1	244	7§	97.1	4	98.4	3	98.8
NS4	82	3	96.3	0	100	4	95.1
NS7	78	7	91	-	-	4	94.8

* Number of divergent amino acids (aa).

† Percentage homology.

‡ Sequence data not available.

§ Compared with full-length NS3b ORF (Kapke *et al.*, 1988).

translated ORFs revealed a greater overall homology between PRCV and Miller or FS772/70 strains than between PRCV and the Purdue strain (Table 1). A further point to mention with respect to PRCV/TGEV sequence comparison is that the TGEV strains for which sequence data are available were isolated a long time before the PRCV outbreak. Therefore, one cannot exclude the possibility that the observed divergence in terms of point mutations reflects, at least in part, the natural drift of TGEV.

No conclusion can be reached about the genome modification(s) controlling the phenotypic difference observed between PRCV and TGEV, namely the impaired multiplication of the former in the digestive tract. It is tempting to speculate that the truncation of the S gene is relevant as it is well known that the spike glycoproteins play a pivotal role in conferring tissue specificity to enveloped RNA viruses. In this respect, mutants of MHV encoding a point-mutated or a truncated S protein have been shown to be neuroattenuated for mice, thus indicating that at least one virulence determinant is associated with this protein (Dalziel *et al.*, 1986; Fleming *et al.*, 1986; Gallagher *et al.*, 1990). The omission of protein (about 30K) in the proximal region of PRCV S might affect, directly or by induced distant conformational change, a putative domain involved in the recognition of TGEV receptors on the enterocytes. Studies aiming to examine the pathogenicity for piglets of TGEV mutants in the region unique to the TGEV S gene are currently being developed as an approach to answer this question. In any case, the recognition that TGEV and MHV S proteins can sustain a large deletion each in a different region should help attempts to dissect the molecule functionally.

To conclude, the present findings about PRCV together with those reported on MHV allow the

proposition that genomic deletion as well as recombination might be a major source of diversity among the coronavirus family. The studies on PRCV provide the first indication that a deleted coronavirus may emerge not only as the result of experimental selection but also spontaneously in the field and, furthermore, may have great epizootiological impact.

We wish to thank Drs C. Duret and J. P. Soulebot for kindly providing the Rhône-Mérieux PRCV isolate and J. Gelfi for her expert assistance in producing PRCV stocks. Thanks are also due to A. M. Spite for typing the manuscript and to K. Rerat for revising the English style. M. Duarte (C.I.V.E.T., Pando, Uruguay) holds a fellowship from the French Government.

References

- BRITTON, P., CARMENES, R. S., PAGE, K. W., GARWES, D. J. & PARRA, F. (1988a). Sequence of the nucleoprotein from a virulent British field isolate of transmissible gastroenteritis virus and its expression in *Saccharomyces cerevisiae*. *Molecular Microbiology* **2**, 89–99.
- BRITTON, P., CARMENES, R. S., PAGE, K. W. & GARWES, D. J. (1988b). The integral membrane protein from a virulent isolate of transmissible gastroenteritis virus: molecular characterization, sequence and expression in *Escherichia coli*. *Molecular Microbiology* **2**, 497–505.
- BRITTON, P., LOPEZ OTIN, C., MARTIN ALONSO, J. M. & PARRA, F. (1989). Sequence of the coding regions from the 3.0 kb and 3.9 kb mRNA subgenomic species from a virulent isolate of transmissible gastroenteritis virus. *Archives of Virology* **105**, 165–178.
- BROWN, I. & CARTWRIGHT, S. (1986). New porcine coronavirus? *Veterinary Record* **119**, 282.
- CALLEBAUT, P., CORREA, I., PENSAERT, M., JIMÉNEZ, G. & ENJUANES, L. (1988). Antigenic differentiation between transmissible gastroenteritis virus of swine and a related porcine respiratory coronavirus. *Journal of General Virology* **69**, 1725–1730.
- CAVANAGH, D., BRIAN, D. A., ENJUANES, L., HOLMES, K. V., LAI, M. M. C., LAUDE, H., SIDDELL, S. G., SPAAN, W., TAGUSHI, F. & TALBOT, P. J. (1990). Recommendations of the coronavirus study group for the nomenclature of the structural proteins, mRNAs, and genes of coronavirus. *Virology* **175**, 306–307.
- DALZIEL, R. G., LAMPERT, P. W., TALBOT, P. J. & BUCHMEIER, M. J. (1986). Site-specific alteration of murine hepatitis virus type 4 peplomer glycoprotein E2 results in reduced neurovirulence. *Journal of Virology* **59**, 463–471.
- DELMAS, B., GELFI, J. & LAUDE, H. (1986). Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. *Journal of General Virology* **67**, 1405–1418.
- DELMAS, B., RASSCHAERT, D., GODET, M., GELFI, J. & LAUDE, H. (1990). Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein S. *Journal of General Virology* **71**, 1313–1323.
- DURET, C., BRUN, A., GUILMOTO, H. & DAUVERGNE, M. (1988). Isolement, identification et pouvoir pathogène chez le porc d'un coronavirus apparenté au virus de la gastro-entérite transmissible. *Recueil de médecine vétérinaire* **164**, 221–226.
- FLEMING, J. O., TROUSDALE, M. D., EL-ZAATAR, F. A. K., STOHLMAN, S. A. & WEINER, L. P. (1986). Pathogenicity of antigenic variants of murine coronavirus JHM selected with monoclonal antibodies. *Journal of Virology* **58**, 869–875.
- GALLAGHER, T. M., PARKER, S. E. & BUCHMEIER, M. J. (1990). Neutralization-resistant variants of a neurotropic coronavirus are generated by deletions within the amino-terminal half of the spike glycoprotein. *Journal of Virology* **64**, 731–741.
- GARWES, J., STEWART, F., CARTWRIGHT, S. F. & BROWN, I. (1988). Differentiation of porcine coronavirus from transmissible gastroenteritis virus. *Veterinary Record* **122**, 86–87.

- HAVE, P. (1990). Infection with a new porcine respiratory coronavirus in Denmark. Serologic differentiation from transmissible gastroenteritis virus using monoclonal antibodies. In *Coronaviruses and Their Diseases*. Edited by D. Cavanagh & T. D. K. Brown. New York & London: Plenum Press (in press).
- DESTIN, A., LE FORBAN, Y., VANNIER, P., MADEC, F. & GOURREAU, J.-M. (1987). Un nouveau coronavirus porcin. Etudes séro-épidémiologiques rétrospectives dans les élevages de Bretagne. *Recueil de médecine vétérinaire* **163**, 567-571.
- KAPKE, P. A., TUNG, F. Y. T. & BRIAN, D. A. (1988). Nucleotide sequence between the peplomer and matrix protein genes of the porcine transmissible gastroenteritis coronavirus identifies three large open reading frames. *Virus Genes* **2**, 293-294.
- LAUDE, H., CHAPSAL, J.-M., GELFI, J., LABIAU, S. & GROSCLAUDE, J. (1986). Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. *Journal of General Virology* **67**, 119-130.
- LAUDE, H., RASSCHAERT, D. & HUET, J.-C. (1987). Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1687-1693.
- LAUDE, H., GELFI, J., RASSCHAERT, D. & DELMAS, B. (1988). Caractérisation antigénique du coronavirus respiratoire porcin à l'aide d'anticorps monoclonaux dirigés contre le virus de la gastro-entérite transmissible. *Journées de la Recherche Porcine en France* **20**, 89-94.
- LIM, H. M. & PÈNE, J. J. (1988). Optimal conditions for supercoil DNA sequencing with the *Escherichia coli* DNA polymerase I large fragment. *Gene Analysis Techniques* **5**, 32-39.
- LUYTJES, W., STURMAN, L. S., BREDENBEEK, P. J., CHARITE, J., VAN DER ZEUST, B. A., HORZINEK, M. C. & SPAAN, W. J. M. (1987). Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site. *Virology* **161**, 479-487.
- O'TOOLE, D., BROWN, I., BRIDGES, A. & CARTWRIGHT, S. F. (1989). Pathogenicity of experimental infection with 'pneumotrophic' porcine coronavirus. *Research in Veterinary Science* **47**, 23-29.
- PARKER, S. E., GALLAGHER, T. M. & BUCHMEIER, M. J. (1989). Sequence analysis reveals extensive polymorphism and evidence of deletions within the E2 glycoprotein gene of several strains of murine hepatitis virus. *Virology* **173**, 664-673.
- PENSAERT, M., CALLEBAUT, P. & VERGOTE, J. (1986). Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Veterinary Quarterly* **8**, 257-261.
- RASSCHAERT, D. & LAUDE, H. (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1883-1890.
- RASSCHAERT, D., GELFI, J. & LAUDE, H. (1987). Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. *Biochimie* **69**, 591-600.
- SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGICHI, R., HORN, G. T., MULLIS, K. B. & ERLICH, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. New York: Cold Spring Harbor Laboratory.
- SCHMIDT, I., SKINNER, M. & SIDDELL, S. (1987). Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *Journal of General Virology* **68**, 47-56.
- SPAAN, W., CAVANAGH, D. & HORZINEK, M. C. (1988). Coronaviruses: structure and genome expression. *Journal of General Virology* **69**, 2939-2952.
- VAN NIEUWSTADT, A. P. & POL, J. M. A. (1989). Isolation of a TGE virus-related respiratory coronavirus causing fatal pneumonia in pigs. *Veterinary Record* **124**, 43-44.
- WESLEY, R. D. (1990). Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequence of the upstream polymerase gene of transmissible gastroenteritis virus (Miller strain). In *Coronaviruses and Their Diseases*. Edited by D. Cavanagh & T. D. K. Brown. New York & London: Plenum Press (in press).
- WESLEY, R. D., CHEUNG, A. K., MICHAEL D. D., & WOODS, R. D. (1989). Nucleotide sequence of coronavirus TG_EV genomic RNA; evidence for 3 mRNA species between the peplomer and matrix protein genes. *Virus Research* **13**, 87-100.

(Received 20 April 1990; Accepted 9 July 1990)