Journal of General Virology (1991), 72, 579–587. *Printed in Great Britain*

Sequence comparison of the 5′ end of mRNA 3 from transmissible gastroenteritis virus and porcine respiratory coronavirus

Kevin W. Page,† Karen L. Mawditt and Paul Britton*

AFRC Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire RG16 0NN, U.K.

Analysis of porcine transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus (PRCV) mRNA species indicated a deletion in mRNA 3 of PRCV. Polymerase chain reaction (PCR) was used to clone the 5′ end of mRNA 3 from PRCV for comparison with the equivalent region in TGEV. Small deletions were observed within and around the PRCV sequence equivalent to the putative open reading frame (ORF) ORF-3a identified in TGEV. The potential RNA polymerase–leader complex binding site (leader RNA binding site), ACTAAAC, found upstream of ORF-3a in TGEV, was absent from the PRCV genome but a potential site was found in the PRCV genome upstream of a gene equivalent to TGEV ORF-3b. PCR analysis, using primers corresponding to sequences within the ORF-3b gene and the leader RNA sequence, confirmed that the leader RNA binding site was upstream of a gene equivalent to TGEV ORF-3b on PRCV mRNA 3 but upstream of ORF-3a on TGEV mRNA 3. The presence of the new leader RNA binding site would be responsible for generating the smaller mRNA 3 species found in PRCV-infected cells.

Introduction

Transmissible gastroenteritis virus (TGEV), first isolated by Doyle & Hutchings (1946), causes gastroenteritis in pigs resulting in high mortality and morbidity in neonates. The virus replicates in and destroys the enterocytes covering the tips of the villi in the small intestine, resulting in diarrhoea and dehydration. TGEV replication has also been shown in lung tissue by immune fluorescence and virus isolation (Underdahl *et al.*, 1974).

TGEV belongs to the *Coronaviridae* family, a group of enveloped viruses with a positive-strand RNA genome. A virus antigenically related to TGEV was detected recently, which spread rapidly through the pig population in several European countries between 1984 and 1986. The virus does not cause diarrhoea and essentially replicates only in the respiratory tract and not in the gastrointestinal tract, with little or no clinical signs (Pensaert *et al.*, 1986; Pensaert, 1989; Cox *et al.*, 1990). The causative agent, isolated in Belgium (Pensaert *et al.*, 1986) and in Britain (Brown & Cartwright, 1986), was identified as a coronavirus and named porcine respiratory coronavirus (PRCV). The serological response in pigs to infection with PRCV could not be distinguished from TGEV by available diagnostic tests.

TGEV and PRCV virions contain three major structural polypeptides; a surface glycoprotein [spike (S)] with a monomeric M_r of 200000, a glycosylated membrane protein (M), observed as a series of polypeptides between M_r 28000 and 31000 and a basic phosphorylated protein [the nucleoprotein (N)] of M_r 47000 associated with the viral genomic RNA (Garwes & Pocock, 1975; Britton *et al.*, 1990). TGEV-infected cells contain, in addition to the genomic RNA (mRNA 1), six species of subgenomic mRNA observed as a 'nested' set (mRNA 2 to 7; Britton *et al.*, 1986; Jacobs *et al.*, 1986). The TGEV genome, except for the polymerase gene, has been cloned and sequenced from a virulent British strain, FS772/70 (Britton *et al.*, 1987, 1988*a*, *b*, 1989; Britton & Page, 1990) and from an American avirulent strain, Purdue-115 (Kapke & Brian, 1986; Kapke *et al.*, 1987, 1988; Jacobs *et al.*, 1987; Laude *et al.*, 1987; Rasschaert & Laude, 1987; Rasschaert *et al.*, 1987). Both TGEV and PRCV have identical leader RNA sequences (Page *et al.*, 1990). In this paper, we have adopted the nomenclature for coronavirus subgenomic mRNAs and viral genes suggested by the coronavirus study group (Cavanagh *et al.*, 1990). Sequencing studies on TGEV FS772/70 showed that mRNA 2 encoded the S gene, mRNA 3 contained two potential open reading frames (ORFs) ORF-3a and ORF-3b

† Present address: MRC Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, U.K.

The nucleotide sequence data reported in this paper for PRCV have been submitted to the EMBL/GenBank/DDBJ nucleotide sequence databases and have been assigned the accession number D00658.

[formerly ORFs 1 and 2 (Britton *et al.*, 1989)], mRNA 4 contained one potential ORF, ORF-4 [formerly ORF 3 (Britton *et al.*, 1989)], mRNA 5 encoded the M gene, mRNA 6 encoded the N gene and mRNA 7 contained a single ORF, ORF-7 [formerly ORF 4 (Britton *et al.*, 1989)] whose product has been identified in TGEV- and PRCV-infected cells (Garwes *et al.*, 1989). Sequence analysis showed that the potential product of the ORF-3b gene from two virulent strains, FS772/70 (Britton *et al.*, 1989) and Miller (Wesley *et al.*, 1989), had an M_r of 27600. However, sequence data from two groups (Rasschaert *et al.*, 1987; Kapke *et al.*, 1988) on the avirulent Purdue-115 strain gave conflicting results. Rasschaert *et al.* (1987) found that the ORF-3b gene product had an M_r of 18800, due to a point mutation in the equivalent initiation codon found in the FS772/70 and Miller strains, but Kapke *et al.* (1988) found that the Purdue-115 ORF-3b gene product had an M_r of 27600.

In this paper we report the cloning and sequencing of the 5′ end of mRNA 3 from two British isolates of PRCV for comparison with the corresponding region of TGEV. These studies were undertaken to identify any sequence variation in this part of the PRCV genome which would account for the observed reduction in size of PRCV mRNA 3.

Methods

Isolation of viral RNA. Viral RNA was isolated from LLC-PK1 cells infected with either TGEV strain FS772/70 or PRCV strains 86/137004 or 86/135308 as described previously by Britton *et al.* (1987) and Page *et al.* (1990).

Northern blot analysis. Specific restriction fragments from TGEV cDNA clones were separated on agarose gels, purified using Geneclean (Stratech Scientific) and labelled with [α-^{32}P]dATP (Amersham) (Maniatis *et al.*, 1982). TGEV and PRCV subgenomic mRNA species were denatured with 6 M-glyoxal, electrophoresed into 1% agarose gels, Northern-blotted onto Biodyne membranes (P/N BNNG3R 1·2 μm; PAL) and hybridized to ^{32}P-labelled TGEV cDNA fragments (Britton *et al.*, 1988*a*, 1990).

Preparation of oligonucleotide primers. Oligonucleotides were synthesized by the phosphoramidite method on an Applied Biosystem 381A DNA synthesizer. Four oligonucleotides were synthesized for polymerase chain reaction (PCR) amplification: oligo 51 (5′ CTGTCCTTCCTAAATTGCAACACACCATGCATAGC 3′) was complementary to a region on the TGEV genome 121 bp downstream of the TGEV ORF-4 initiation codons; oligo 52 (5′ GGCCTTGGTATGTGTGGCTACTAATAGGC 3′) corresponded to a region on the TGEV genome 152 bp upstream of the TGEV S gene termination codon; oligo 74 (5′ AACTCAGAGTATTAAGAAAA 3′) was complementary to a region on the TGEV and PRCV genomes 14 bp downstream of the ORF-3b initiation codon and oligo 75 (5′ CCTTTTAAAGTAAAGTGAGT 3′) corresponded to the 5′ end of the TGEV/PRCV leader RNA (Page *et al.*, 1990). See Fig. 2 for the position of the oligonucleotides on the TGEV genome.

First-strand cDNA synthesis. Total RNA (5 μg) isolated from virus-infected cells was re-dissolved in water containing 40 units (U) of RNasin (Promega). Synthesis of first-strand cDNA was carried out in either a solution containing 50 mM-Tris–HCl pH 8·3, 3mM-$MgCl_2$, 75 mM-KCl, 10 mM-DTT, 2·5 mM-dNTPs and primed with 160 ng of oligo 51 using 1200 U of Moloney murine leukaemia virus reverse transcriptase (Gibco BRL) at 37 °C for 2 h or in a solution containing 50 mM-Tris–HCl pH 8·3, 10 mM-$MgCl_2$, 35 mM-KCl, 30 mM-2-mercaptoethanol, 10 mM-DTT, 2·5 mM-dNTPs and primed with 200 ng of oligo 74 using 23 U of avian myeloma virus reverse transcriptase (Super-RT; Anglian Biotech) at 45 °C for 90 min.

PCR amplification of first-strand cDNA. PCR amplification reactions (100 μl) contained 10 μl of the first-strand cDNA, 10 mM-Tris–HCl pH 8·3, 50 mM-KCl, 1·5 mM-$MgCl_2$, 0·01% gelatin, 250 μM-dNTPs, 5 U of Taq polymerase (AmpliTaq; Perkin-Elmer-Cetus) and 500 ng each of either oligos 51 and 52 or oligos 74 and 75. The reaction mixtures were overlaid with 100 μl of light mineral oil (Sigma) and the DNA was amplified during 35 cycles of 94 °C for 1 min, 45 °C for 2 min and 72 °C for 3 min with a final elongation step of 72 °C for 9 min in a Techne *PHC-1* programmable thermal cycler.

Cloning of PCR-generated fragments. The PCR fragments obtained using oligos 51 and 52 were ether-extracted three times to remove the mineral oil, electrophoresed on a 1% agarose gel using TAE buffer (Maniatis *et al.*, 1982), excised from the gel and purified with Geneclean. The cDNA was 5′-phosphorylated with T4 polynucleotide kinase (Gibco BRL), any incomplete ends were repaired using the Klenow fragment (Pharmacia) and *Bam*HI linkers (product no. 1065; New England Biolabs) were added. Following digestion with *Bam*HI the DNA fragment was electrophoresed into an agarose gel, purified by Geneclean, ligated into *Bam*HI-cut dephosphorylated pUC13 (Pharmacia) and transformed into *Escherichia coli* strain BRD342.

Sequencing of cloned PCR fragments. Plasmid DNA was purified by CsCl density gradient centrifugation and sequenced directly as described by Murphy & Kavanagh (1988) with Sequenase (United States Biochemical) for dideoxynucleotide chain-termination reactions and a variety of oligonucleotide primers generated from TGEV sequence data. Both strands of DNA were sequenced several times to eliminate any ambiguous data.

Direct sequencing of PCR-produced fragments. PCR fragments were purified by Geneclean and sequenced with Sequenase in the presence of DMSO (Sigma) according to the method of Winship (1989).

Primer extension of TGEV and PRCV mRNA 3. Gel-purified oligo 74 (500 ng) was 5′ end-labelled using 20 U of T4 polynucleotide kinase and 20 μCi of [γ-^{32}P]ATP (Amersham). Poly(A)-containing RNA (1·5 μg) isolated from TGEV- and PRCV-infected cells was resuspended in water, heated at 60 °C for 3 min and subjected to primer extension from mRNA 3 with 120 ng of ^{32}P-labelled oligo 74 as described by Page *et al.* (1990). The products were electrophoresed into a buffer gradient sequencing gel (Biggin *et al.*, 1983) and autoradiographed.

Data handling and analysis. A sonic digitizer (Graf/Bar; Science Accessories) was used to read sequence data into a Elonex PC-286 microcomputer and data were analysed on a MicroVAX 3600 by the computer programs of Staden (1982), the University of Wisconsin Genetics Computer Group (UWGCG) (Devereux *et al.*, 1984) and CLUSTAL (Higgins & Sharp, 1988).

Results

Viral mRNA analysis

Subgenomic mRNA from cells infected with either TGEV or two different strains of PRCV were blotted

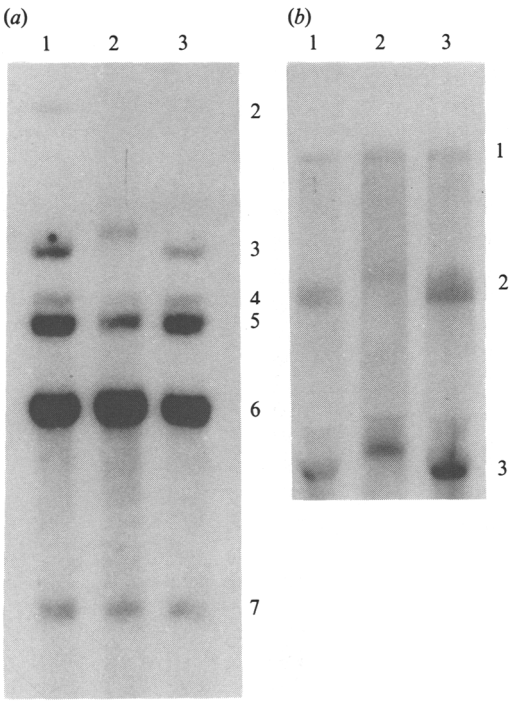


Fig. 1. Autoradiographs of TGEV and PRCV RNA Northern-blotted onto a Biodyne membrane and probed with ^{32}P-labelled TGEV cDNA. Lanes 1 are PRCV strain 86/135308, lanes 2 are TGEV strain FS772/70 and lanes 3 are PRCV strain 86/137004 RNA. (*a*) shows mRNA species 3 to 7 and (*b*) shows mRNA species 1 to 3. The mRNA species are numbered on the right hand side.

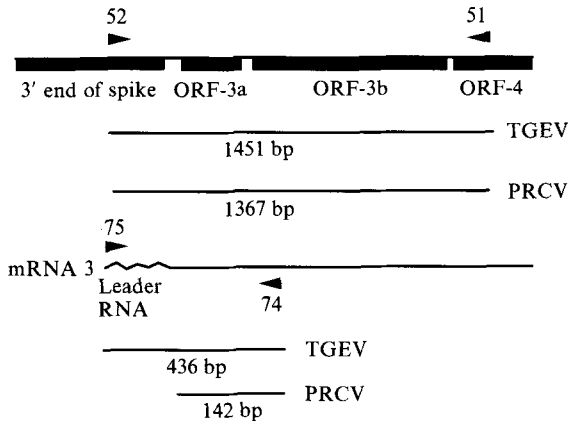


Fig. 2. Schematic diagram of the TGEV genome and mRNA 3 showing the position of the oligonucleotides used for PCR cloning. The thick line at the top shows the position of the TGEV genes and the single line with the wavy end represents the 5′ end of mRNA 3. Arrowheads show the position and orientation of the oligonucleotides. The single lines show the TGEV (upper) and PRCV (lower) fragments obtained by PCR.

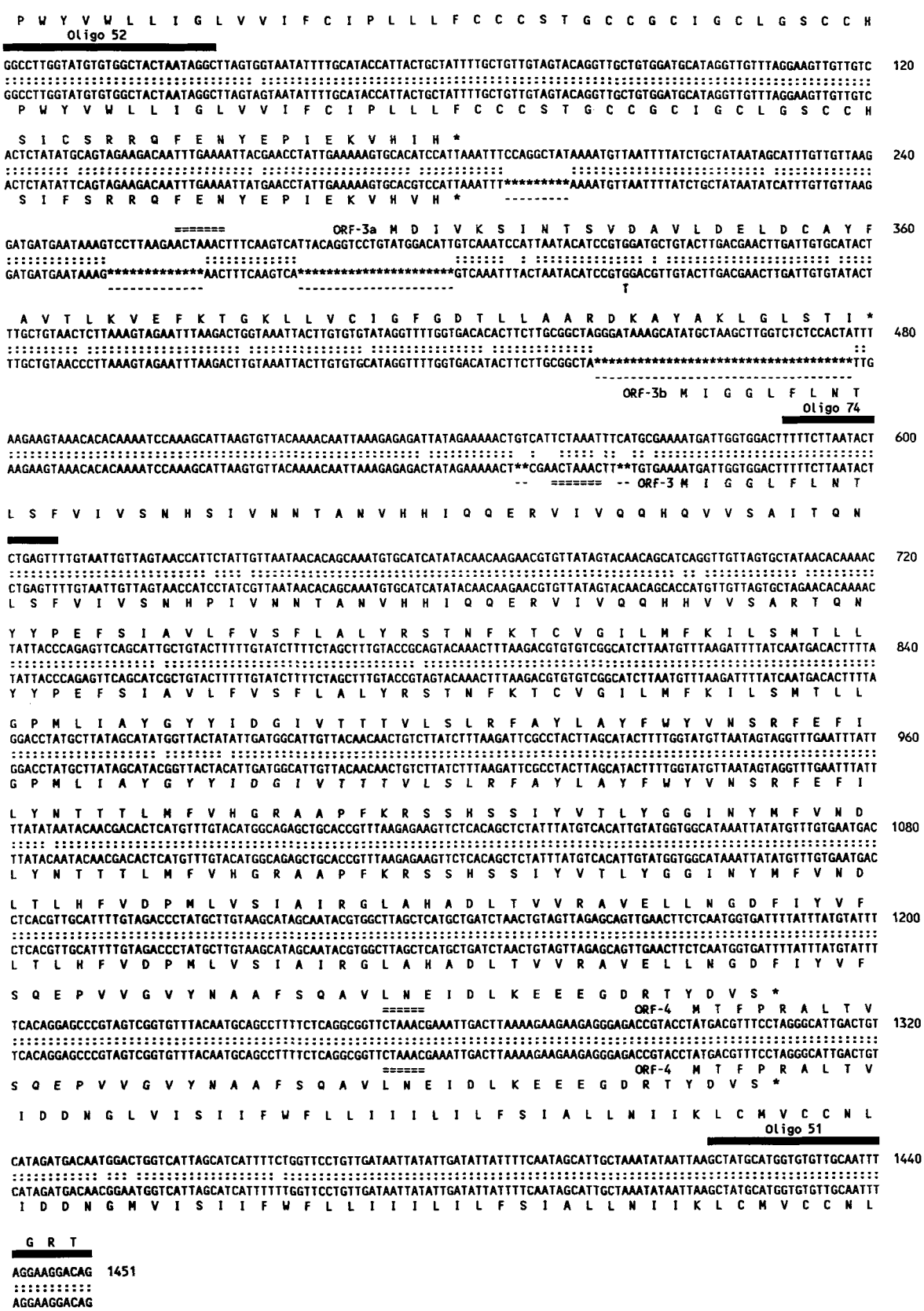
onto membranes and probed with ^{32}P-labelled TGEV cDNA. Seven mRNA species were identified for PRCV (Fig. 1) as previously identified for TGEV (Britton *et al.*, 1986; Jacobs *et al.*, 1986). The pattern of mRNA species from PRCV-infected cells was very similar to that for TGEV, with identical sizes for mRNAs 4 to 7. However, mRNA 3 from PRCV-infected cells migrated faster than the equivalent TGEV species corresponding to a size of about 3·6 kb as compared to the 3·9 kb mRNA 3 in TGEV-infected cells. The mRNA 2 species from PRCV also migrated faster than the equivalent TGEV species. This indicated that the 5′ end of PRCV mRNA 3 contained a potential deletion of about 300 bases. There were no observable differences in the sizes of mRNA 1 due to the resolution of the gel system used.

Cloning of the PRCV genome between oligonucleotides 51 and 52

To investigate the potential deletion observed in PRCV mRNA 3 the genome region from the 3′ end of the S gene and the 5′ end of ORF-4, corresponding to the 5′ coding region of mRNA 3, was cloned using PCR (Saiki *et al.*, 1988) from PRCV strain 86/137004. Following PCR amplification, using oligos 51 and 52, a cDNA fragment of about 1350 bp was obtained from PRCV RNA. The equivalent region from TGEV has 1451 bases (Fig. 2) indicating that the PRCV-derived PCR fragment was 200 bp larger than expected from the mRNA analysis. The 1350 bp PRCV cDNA fragment was cloned into pUC13 and positive clones were identified using ^{32}P-labelled TGEV cDNA from ORF-3b as described by Britton *et al.* (1988*a*). The plasmid DNA from a positive clone, pKP-1, was used for DNA sequencing.

Sequence comparison of the TGEV and PRCV genomes between oligonucleotides 51 and 52

The PRCV cDNA cloned into plasmid pKP-1 was sequenced directly, in both directions, using several oligonucleotide primers corresponding to sequence data determined from the TGEV genome. The cDNA insert in pKP-1 was found to be 1367 bp in length and its sequence was compared to the corresponding region of the TGEV, strain FS772/70, genome (Fig. 3). The region between oligos 51 and 52 on the TGEV genome corresponds to a sequence of 1451 bases (Britton *et al.*, 1989) indicating that PRCV had 84 nucleotides deleted. Comparison of the sequences, between oligos 51 and 52, from the PRCV and TGEV (FS772/70) genomes using the GAP program revealed six deletion sites in the PRCV cDNA (Fig. 3). The first two deletions of 9 and 13 bases, nucleotide positions 190 and 255 on the TGEV sequence, were downstream of the S gene termination codon and within the non-coding region between the S and ORF-3a genes on the TGEV genome. The third


Oligo 52
ORF-3a
ORF-3b
Oligo 74
Oligo 51

deletion of 22 bases at nucleotide position 281 on the TGEV sequence spanned the 5′ end of ORF-3a and resulted in the loss of the initiation codon. The fourth deletion of 36 bases at nucleotide position 442 on the TGEV sequence was 149 bases downstream of the ORF-3a initiation codon and within the coding region of ORF-3a. The fifth and sixth deletions consisted of two bases each, at nucleotide positions 551 and 565 on the TGEV sequence, and were seven and 21 bases, within a non-coding region, upstream of the initiation codon of a gene, ORF-3, equivalent to TGEV ORF-3b. To confirm that the sequence identified from the PRCV genome was a true reflection of the PCR fragment population, the PCR mixture used for cloning was directly sequenced and compared to the sequence generated from plasmid pKP-1. No differences were found indicating that all the deletions identified were representative of the PCR-generated fragments.

In order to check the validity of the deletions in the PRCV genome the region between oligos 51 and 52 was cloned by PCR from another strain of PRCV, strain 86/135308. The PCR fragment was identical in size to that obtained from PRCV strain 86/137004 and again was about 200 bases larger than the size expected from the Northern blot (Fig. 1). A region of the genome from PRCV strain 86/135308, corresponding to nucleotides 121 to 511 on the PRCV 86/137004 sequence, was sequenced directly from PCR-generated fragments and found to be identical to PRCV strain 86/137004 except for one nucleotide at position 282 (Fig. 3).

Analysis of the PRCV sequences indicated that the 22-base deletion covering the initiation codon of the ORF-3a gene and the 36-base deletion within the ORF-3a gene would result in the loss of the ORF-3a gene from PRCV. However, the total number of bases lost by the PRCV deletions was 84 and did not account for the decrease in the size of the mRNA 3 present in PRCV-infected cells. Further analysis of the PRCV sequence showed that the 13-base deletion at nucleotide position 255 on the TGEV sequence resulted in the loss of the ACTA motif from the polymerase–leader complex binding site (leader RNA binding site) ACTAAAC. However, a new leader RNA binding site was created nine bases upstream of ORF-3 in the PRCV sequence as a result of the base substitutions, of two Ts in TGEV to A and C in PRCV (Fig. 3). In fact a further base substitution upstream of the first residue in the leader RNA binding site, T in TGEV to A in PRCV, resulted in the octameric sequence AACTAAAC found upstream of some TGEV genes.

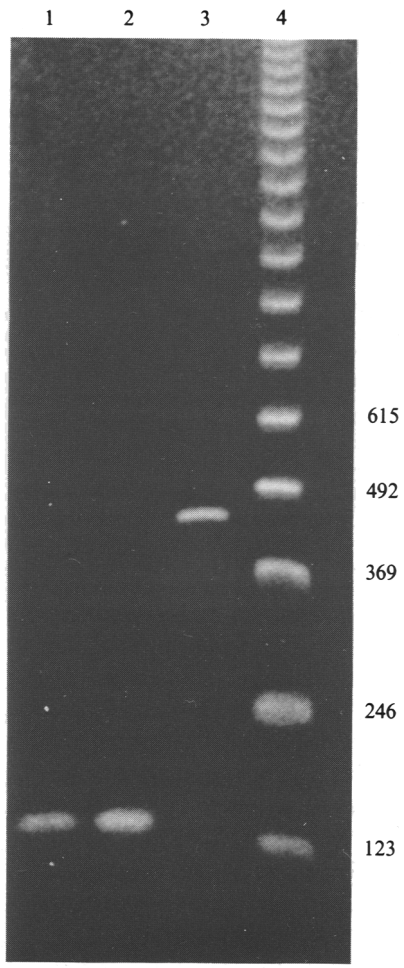


Fig. 4. Analysis of PCR-amplified products, on a 2% agarose gel, generated from: lane 1, PRCV strain 86/137004; lane 2, PRCV strain 86/135308; lane 3, TGEV strain FS772/70 mRNA 3 from virus-infected cells using oligos 74 and 75. Lane 4 corresponds to a 123 bp ladder (BRL) whose sizes are indicated on the right. The sizes of the PRCV and TGEV cDNA fragments were calculated from analysis on a 7 to 12% polyacrylamide gel.

Analysis of PRCV mRNA 3

Two approaches were used to investigate whether PRCV produced an mRNA 3 from the new leader RNA binding

Fig. 3. The nucleotide sequence of the TGEV strain FS772/70 (Britton *et al.*, 1989) (upper) and PRCV strain 86/137004 (lower) genomes between oligonucleotides 52 and 51. The amino acid sequences corresponding to the C terminus of the S protein, ORF-3a and ORF-3b (TGEV), ORF-3 (PRCV) and the N-terminal end of ORF-4 are shown above and below the nucleotide sequences. Positions of the leader RNA binding sites upstream of genes are marked with (= = =). The thick lines show the positions of the oligonucleotides 52, 74 and 51. The colons between the two sequences show identical bases and the asterisks within the PRVC nucleotide sequence show the positions of the deletions which are underlined (---). Differences in the nucleotide sequences between PRCV strains 86/137004 and 86/135308 determined between nucleotides 121 and 511 are shown below the PRCV sequence.

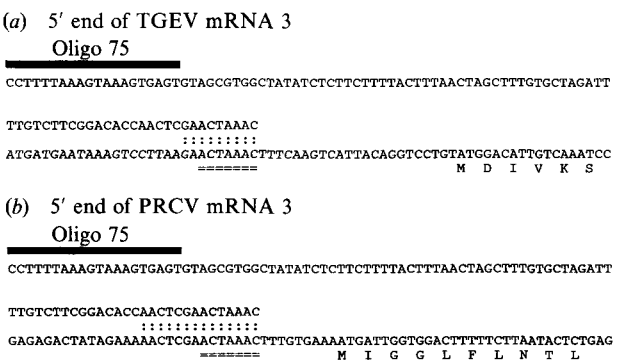


Fig. 5. The nucleotide sequence of the 5′ ends of the mRNA 3 species from TGEV (*a*) and PRCV (*b*) generated by PCR amplification using oligos 74 and 75. The upper sequence in each case represents the leader RNA with the position of oligonucleotide 75 shown by a thick line and the lower sequences are derived from the viral genomes. The position of the leader RNA binding site (ACTAAAC) is underlined (= = =). The deduced amino acid sequences represent the N termini of ORF-3a (TGEV) and ORF-3 (PRCV). The colons show identical bases between the 3′ end of the leader RNA, upstream and including the leader RNA binding site, and the TGEV and PRCV genomes.

site. The first approach consisted of primer extension studies using oligo 74 which was complementary to a region at the start of ORF-3b in TGEV and ORF-3 in PRCV. If the leader RNA was upstream of ORF-3 on PRCV mRNA 3, a primer-extended product of about 142 bases would be observed in contrast to a product of 436 bases from the leader RNA upstream of ORF-3a on TGEV mRNA 3. Following primer extension, a fragment of about 140 bases was generated from RNA isolated from PRCV-infected cells which was consistent with the leader RNA being upstream of ORF-3 on PRCV mRNA 3. No fragment was detected from RNA isolated from TGEV-infected cells possibly as a result of secondary structure preventing either the annealing of the primer or incomplete primer extension. The second approach consisted of generating PCR fragments from TGEV and PRCV mRNA 3 using oligo 74 and an oligonucleotide, oligo 75, which corresponded to the extreme 5′ end of the TGEV and PRCV leader RNA determined by Page *et al.* (1990). The PCR fragments should have corresponded to about 436 bp if the leader RNA was joined upstream of ORF-3a or 142 bp if it was joined upstream of ORF-3. Following PCR amplification of the 5′-terminal regions of TGEV and PRCV mRNA 3 a 148 bp fragment was generated from PRCV RNA and a 440 bp fragment from TGEV RNA (Fig. 4). The sizes of the fragments were consistent with the leader RNA being present upstream of ORF-3a in TGEV and ORF-3 in PRCV. The PCR fragments were sequenced directly and the sequence data (Fig. 5) confirmed that the leader RNA was upstream of ORF-3a in TGEV but upstream of ORF-3 in PRCV. The latter result explained the observation of the smaller mRNA 3 (3·6 kb) species in PRCV-infected cells. The difference between the two priming sites on the TGEV genome is 285 bases, consistent with the difference in the observed size of the mRNA 3 molecules.

Discussion

Analysis of TGEV and PRCV subgenomic mRNA species showed that mRNAs 4 to 7 were identical in size and that PRCV mRNAs 2 and 3 were smaller than the equivalent TGEV mRNAs (Fig. 1) indicating that there was a deletion of about 300 bases in PRCV mRNA 3. Comparison of PRCV 86/137004 cDNA derived from the genome corresponding to mRNA 3 of the equivalent region from TGEV FS772/70 indicated that the two sequences were highly conserved with several deletions within the PRCV sequence. However, the deletions corresponded to only 84 bases and were confirmed in the genome of a second PRCV isolate, 86/135308 (Fig. 3). Two base substitutions in the PRCV genome resulted in the creation of a new leader RNA binding site upstream of ORF-3, equivalent to ORF-3b in TGEV, that resulted in the synthesis of a smaller mRNA in PRCV-infected cells.

TGEV strains FS772/70 and Purdue-115 produce a 3·9 kb mRNA 3 with two potential genes at the 5′ end with the sequence CTAAAT 11 bases upstream of the second ORF (Britton *et al.*, 1989; Rasschaert *et al.*, 1987). However, the Miller strain has the hexameric leader RNA binding site, CTAAAC, upstream of ORF-3b and was shown to produce an extra mRNA species of 3·7 kb (Wesley *et al.*, 1989). The CTAAAC motif is the potential leader RNA binding site for the synthesis of mRNA 4 in all strains of TGEV sequenced so far as well as the two PRCV strains described in this paper. It is interesting to note that mRNA 4 in TGEV- and PRCV-infected cells and the 3·7 kb mRNA species from the Miller strain of TGEV, all of which have the CTAAAC leader RNA binding site, are produced in lower amounts than mRNA species with the heptameric ACTAAAC leader RNA binding site.

Previous work comparing the ORF-3a/3b region of TGEV between the Purdue-115 (Rasschaert *et al.*, 1987) and FS772/70 (Britton *et al.*, 1989) or Miller (Wesley *et al.*, 1989) strains identified several small deletions within a non-coding region upstream of ORF-3a and at the 3′ end of ORF-3a. Comparison of this region from TGEV strains FS772/70, Miller and Purdue-115, and the two PRCV strains using the CLUSTAL program is shown in Fig. 6 and indicates that this particular region of the TGEV/PRCV genome is hypervariable in contrast to other regions of the genomes sequenced to date. The

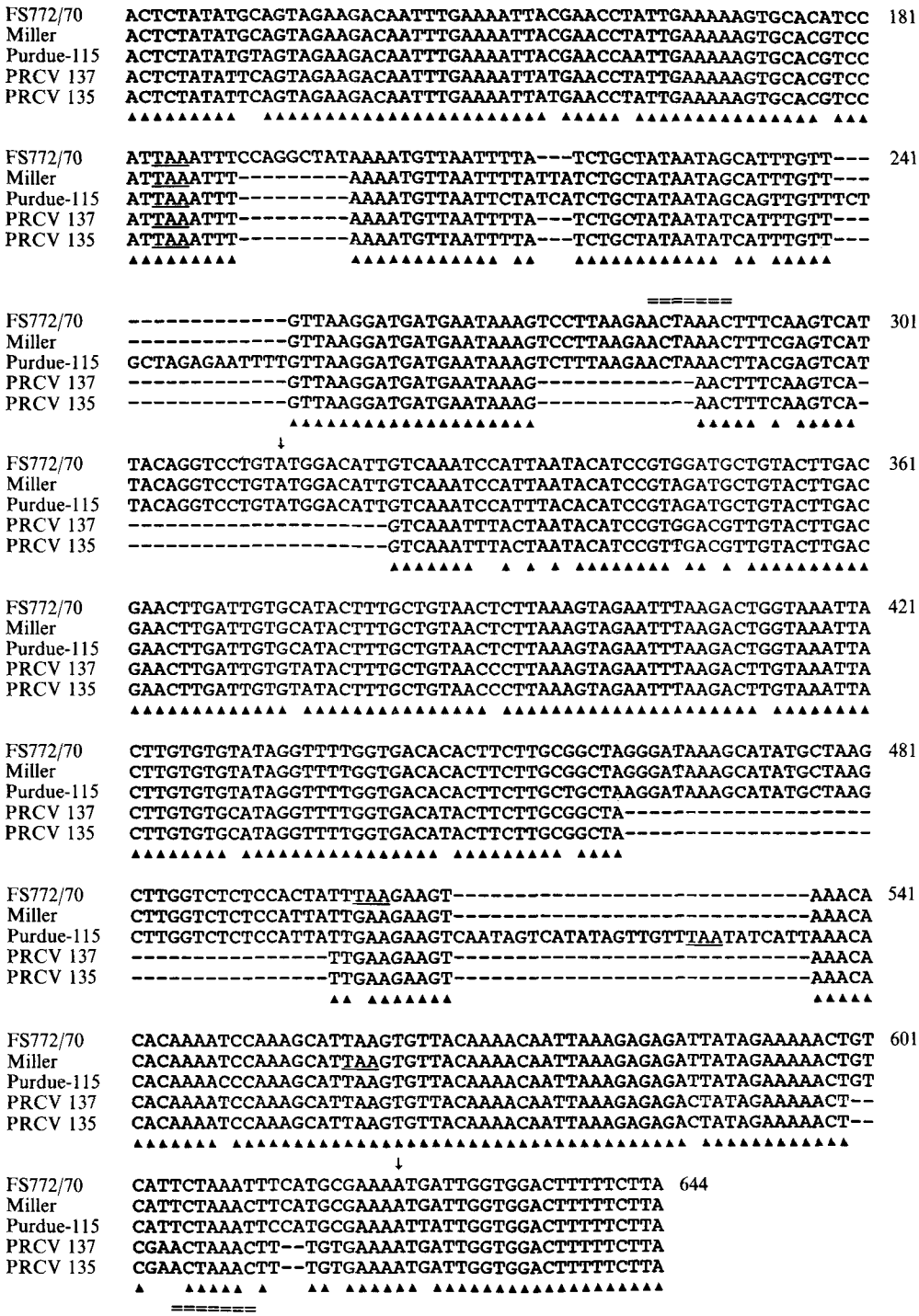


Fig. 6. Nucleotides 121 to 595 (TGEV FS772/70 in Fig. 3) from three TGEV and two PRCV strains aligned by the CLUSTAL program. (▲), Identical bases; (---), deletions; (= = =), leader RNA binding sites; (⇩), the translational start of genes; stop codons are underlined. The numbering of the nucleotides differs from Fig. 3 to allow for the insertions present in the various sequences. The ATT sequence at nucleotide 623 for Purdue-115 (Rasschaert *et al.*, 1987) was found to be ATG for ORF-3b by Kapke *et al.* (1988).

deletions are summarized in Table 1. The PRCV deletions of 13, 22, 36, two and two bases, at nucleotides 274 to 286, 300 to 321, 461 to 496, 599 to 600 and 613 to 614 respectively (Fig. 6), were unique to the PRCV sequences and resulted in the loss of the potential ORF-3a gene. The PRCV deletions of nine, three, 16 and 29 bases, at nucleotides 190 to 198, 214 to 216, 238 to 253 and 507 to 535 respectively (Fig. 6), were consistent with

Table 1. *Length of deletions observed in the different strains of TGEV and their comparison to PRCV*

Strain	Length of deletions in bases								
FS772/70*	–‡	3	16	–	–	–	29	–	–
Purdue*	9	–	–	–	–	–	–	–	–
Miller*	9	–	16	–	–	–	29	–	–
86/135308†	9	3	16	13	22	36	29	2	2
86/137004†	9	3	16	13	22	36	29	2	2
Bases	190–198	214–216	238–253	274–286	300–321	461–496	507–535	599–600	613–614

* TGEV strains.
† PRCV strains.
‡ –, No deletion.

at least one strain of TGEV. The differences observed at the 3′ end of TGEV ORF-3a result in differences at the C-terminal end of the potential ORF-3a products. The position of the termination codons are shown in Fig. 6. It should be noted that if the FS772/70 ORF-3a did not prematurely terminate, the extra amino acids would be the same as in the Miller sequence and that the nucleotide sequences of the two PRCV strains are identical to the Miller strain over this region (Fig. 6). The Purdue-115 nucleotide sequence is also very similar, except for a single base substitution and the 29-base insertion. These observations indicate that the virus genomes have mutated within this region resulting in changes to the C terminus of the potential ORF-3a gene product. It will be interesting to compare the equivalent regions of the genomes from the two other coronaviruses belonging to the TGEV subgroup, feline infectious peritonitis virus and canine coronavirus, once their sequences are available. This should provide information on the evolution of the TGEV coronavirus subgroup.

The deduced amino acid sequences of ORF-3b are very similar between TGEV (FS772/70) and PRCV, with only three single amino acid substitutions: serine (position 19) to proline, glutamine (position 41) to histidine and isoleucine (position 46) to arginine (Fig. 3). The amino acid substitution at residue 19 resulted in the loss of one of the three potential *N*-glycosylation sites found in TGEV ORF-3b; however, the other two potential *N*-glycosylation sites, residues 22 and 132, were conserved. The amino acid sequence of the Miller strain is also very similar with serine and glutamine at positions 19 and 41 and arginine at position 46, but the Miller strain had lysine at position 31 and glutamic acid at position 237.

Sequence data from part of the genome of two British isolates of PRCV presented in this paper indicate that the nucleotide and derived amino acid sequences are very similar to those of TGEV. The differences, apart from the deletions, are consistent with those observed between different strains of TGEV. These observations indicate that PRCV may be a variant of TGEV and that one or more of the observed deletions may be responsible for the altered pathogenicity of PRCV and suggest that deletion events may play a role in the evolution of coronaviruses.

We would like to thank Dr D. H. Pocock for helpful comments on the manuscript. This research was supported by the Biotechnology Action Programme of the Commission of the European Communities Contract Number [BAP-0235-UK(HI)].

References

BIGGIN, M. D., GIBSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 3963–3965.

BRITTON, P. & PAGE, K. W. (1990). Sequence of the S gene from a virulent British field isolate of transmissible gastroenteritis virus. *Virus Research* (in press).

BRITTON, P., GARWES, D. J., MILLSON, G. C., PAGE, K., BOUNTIFF, L., STEWART, F. & WALMSLEY, J. (1986). Towards a genetically-engineered vaccine against porcine transmissible gastroenteritis virus. In *Biomolecular Engineering in the European Community (Final report)*, pp. 301–313. Edited by E. Magnien. The Hague: Martinus Nijhoff.

BRITTON, P., GARWES, D. J., PAGE, K. & WALMSLEY, J. (1987). Expression of porcine transmissible gastroenteritis virus genes in *E. coli* as β-galactosidase chimaeric proteins. In *Coronaviruses, Advances in Experimental Medicine and Biology*, vol. 218, pp. 55–64. Edited by M. M. C. Lai & S. A. Stohlman. New York & London: Plenum Press.

BRITTON, P., CARMENES, R. S., PAGE, K. W., GARWES, D. J. & PARRA, F. (1988*a*). 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. (1988*b*). 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.

BRITTON, P., GARWES, D. J., PAGE, K. & STEWART, F. (1990). Molecular aspects of the relationship of transmissible gastroenteritis virus (TGEV) with porcine respiratory coronavirus (PRCV). In *Coronaviruses and Their Diseases, Advances in Experimental Medicine and Biology*, vol. 276, pp. 441–446. Edited by D. Cavanagh & T. D. K. Brown. New York & London: Plenum Press.

BROWN, I. & CARTWRIGHT, S. (1986). New porcine coronavirus? *Veterinary Record* **119**, 282–283.

CAVANAGH, D., BRIAN, D. A., ENJUANES, L., HOLMES. K. V., LAI, M. M. C., LAUDE, H., SIDDELL, S. G., SPAAN, W., TAGUCHI, F. & TALBOT, P. J. (1990). Recommendations of the coronavirus study group for the nomenclature of the structural proteins, mRNAs and genes of coronaviruses. *Virology* **176**, 306–307.

COX, E., PENSAERT, M. B., CALLEBAUT. P. & VAN DEUN, K. (1990). Intestinal implication of a porcine respiratory coronavirus closely related antigenically to the enteric transmissible gastroenteritis virus. *Veterinary Microbiology* **23**, 237–243.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DOYLE, L. P. & HUTCHINGS, L. M. (1946). A transmissible gastroenteritis in pigs. *Journal of the American Veterinary Medical Association* **108**, 257–259.

GARWES, D. J. & POCOCK, D. H. (1975). The polypeptide structure of transmissible gastroenteritis virus. *Journal of General Virology* **29**, 25–34.

GARWES, D. J., STEWART, F. & BRITTON, P. (1989). The polypeptide of M_r 14000 of porcine transmissible gastroenteritis virus: gene assignment and intracellular location. *Journal of General Virology* **70**, 2495–2499.

HIGGINS, D. G. & SHARP, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237–244.

JACOBS, L., VAN DER ZEIJST, B. A. M. & HORZINEK, M. C. (1986). Characterization and translation of transmissible gastroenteritis virus mRNAs. *Journal of Virology* **57**, 1010–1015.

JACOBS, L., DE GROOT, R., VAN DER ZEIJST, B. A. M., HORZINEK, M. C. & SPAAN, W. (1987). The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV). *Virus Research* **8**, 363–371.

KAPKE, P. A. & BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41–49.

KAPKE, P. A., TUNG, F. Y. C., BRIAN, D. A., WOODS, R. D. & WESLEY, R. (1987). Nucleotide sequence of the porcine transmissible gastroenteritis coronavirus matrix protein. In *Coronaviruses, Advances in Experimental Medicine and Biology*, vol. 218, pp. 117–122. Edited by M. M. C. Lai & S. A. Stohlman. New York & London: Plenum Press.

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., 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.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

MURPHY, G. & KAVANAGH, T. (1988). Speeding-up the sequencing of double-stranded DNA. *Nucleic Acids Research* **16**, 5198.

PAGE, K. W., BRITTON, P. & BOURSNELL, M. E. G. (1990). Sequence analysis of the leader RNA of two porcine coronaviruses: transmissible gastroenteritis virus and porcine respiratory coronavirus. *Virus Genes* **4**, 289–301.

PENSAERT, M. B. (1989). Transmissible gastroenteritis virus (respiratory variant). In *Virus Infections of Vertebrates*, vol. 2, pp. 154–165. Edited by M. C. Horzinek & M. B. Pensaert. Amsterdam: Elsevier.

PENSAERT, M. B., CALLEBAUT, P. E. & VERGOTE, J. (1986). Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Veterinary Quarterly* **8**, 257–260.

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 organisation and expression. *Biochimie* **69**, 591–600.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, H., 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.

STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731–4751.

UNDERDAHL, N. R., MEBUS, C. A., STAIR, E. L., RHODES, M. B., MCGILL, L. D. & TWIEHAUS, M. J. (1974). Isolation of transmissible gastroenteritis virus from lungs of market-weight swine. *American Journal of Veterinary Research* **35**, 1209–1216.

WESLEY, R. D., CHEUNG, A. K., MICHAEL, D. D. & WOODS, R. A. (1989). Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNA species between the peplomer and matrix protein genes. *Virus Research* **13**, 87–100.

WINSHIP, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Research* **17**, 1266.

(*Received 15 August 1990; Accepted 23 November 1990*)