Molecular Characterization of Porcine Reproductive and Respiratory Syndrome Virus, a Member of the Arterivirus Group[1]

KARL-KLAUS CONZELMANN,*,[2] NICO VISSER,† PETER VAN WOENSEL,† AND HEINZ-JÜRGEN THIEL*

*Federal Research Centre for Virus Diseases of Animals, D-7400 Tübingen, Paul-Ehrlich-Strasse 28, Federal Republic of Germany; and †Intervet International BV, PO Box 31, NL-5830 AA Boxmeer, The Netherlands

Received September 28, 1992; accepted October 27, 1992

Porcine reproductive and respiratory syndrome virus (PRRSV)-specific cDNA clones spanning the 3′ terminal 5 kb of the genomic RNA were isolated, sequenced, and used as probes for identification of PRRSV-specific RNAs. The PRRSV genome is a positive-stranded polyadenylated RNA of about 15 kb. In infected cells, a 3′ coterminal nested set of six major subgenomic mRNAs could be demonstrated. Within the 3′ terminal 3.5 kb of the PRRSV genome, six overlapping reading frames (ORFs) were identified, each most likely expressed by one of the subgenomic mRNAs. Amino acid sequence comparisons revealed that the most 3′ terminal ORF (ORF7) encodes the PRRSV nucleocapsid protein with a calculated molecular weight of 14 kDa. It displays 44.8% amino acid identity with the capsid protein of lactate dehydrogenase-elevating virus (LDV) and 23.6% with that of equine arteritis virus (EAV). The product of ORF6, the second 3′ terminal ORF, represents a putative membrane protein and exhibits 53.2 and 27.2% amino acid identity with the corresponding LDV and EAV polypeptides, respectively. Similar to EAV, ORFs 2 through 5 might encode glycosylated viral proteins. The polypeptide deduced from the most 5′ ORF (ORF1b) contains two conserved domains common to EAV and coronavirus polymerases. Genome organization, strategy of gene expression, and the sequence of deduced proteins show that PRRSV belongs to the Arterivirus group of viruses.

INTRODUCTION

Porcine reproductive and respiratory syndrome virus (PRRSV) represents the causative agent of a new porcine disease which emerged in northern Germany in 1990 and spread in form of an epidemic all over Europe in 1991. As indicated by the name of the virus, major clinical signs in swine are abortion and respiratory distress (Lindhaus and Lindhaus, 1991). A similar "mystery swine disease" has first been reported in 1987 in North America (Keffaber, 1989), which is now usually called "swine infertility and respiratory syndrome" (SIRS, Collins *et al.*, 1991).

Both the European PRRSV and the American SIRSV have recently been isolated and characterized (Ohlinger *et al.*, 1991; Terpstra *et al.*, 1991; Wensvoort *et al.*, 1991a,b; Benfield *et al.*, 1992; Collins *et al.*, 1992). Serological, ultrastructural, and biophysical data indicate that PRRSV and SIRSV represent the same virus. Serological variation, however, exists especially between the American and European isolates, but also among virus isolates from North America (Wensvoort *et al.*, 1992).

PRRSV has been described as a small enveloped RNA virus (Wensvoort *et al.*, 1991b, Benfield *et al.*, 1992) with morphological and morphogenetical similarities to members of the arterivirus group, including equine arteritis virus (EAV) and lactate dehydrogenase-elevating virus of mice (LDV). In addition, relationships between PRRSV and arteriviruses are suggested by the nature of permissive cells. The arteriviruses infect particular subpopulations of macrophages (Plagemann and Moennig, 1992) and PRRSV apparently grows exclusively in alveolar lung macrophages. However, serological crossreactions could so far not be demonstrated between PRRSV and any of the arteriviruses.

Members of the arterivirus group are currently classified within the Togaviridae family (Westaway *et al.*, 1985), but the need for reclassification has become obvious after cloning and molecular analysis of the total EAV genome (Den Boon *et al.*, 1991) and of parts of the LDV genome (Godeny *et al.*, 1990). In contrast to togaviruses, arterivirus gene expression does not occur by translation and subsequent processing of polyproteins, but by transcription of multiple subgenomic mRNAs, each encoding one protein. Similar to coronaviruses, arteriviral mRNAs form a 3′ coterminal nested set and possess common 5′ terminal leader sequences which are joined to the bodies of the mRNAs

[1] Sequence data from this article have been deposited with the GenBank Data Library under the accession number L04493.

[2] To whom reprint requests should be addressed.

during transcription. Moreover, the putative EAV polymerase gene is probably expressed by ribosomal frameshifting as in coronaviruses and possesses conserved domains also present in corona- and torovirus polymerases (Den Boon *et al.*, 1991).

Using purified PRRS virions from infected macrophages as starting material, molecular cDNA cloning and sequencing was performed. This approach should not only elucidate the relationship of PRRSV to other viruses, but also provide tools for diagnostic purposes and the development of vaccines. In this paper we provide the first molecular data on the PRRSV genome. Its organization, transcription features and also deduced viral protein sequences prove the close evolutionary relationship between PRRSV and arteriviruses.

MATERIALS AND METHODS

Cells and virus isolation

Porcine alveolar macrophages were harvested from the lungs of SPF pigs as described (Wensvoort *et al.*, 1991a,b). PRRSV isolate 10 was obtained from a stillborn piglet from a sow with PRRS. The lungs were homogenized and 0.3 ml of the homogenate were added to a flask with alveolar macrophages. Three days post-infection cell lysis was observed. The virus was further cultured on SPF macrophages and characterized by immunofluorescence, electron microscopy, and Western blot (Ohlinger *et al.*, 1991). Swine infected with isolate 10 showed symptoms and an immune response typical of PRRS. PRRSV could be reisolated from all infected pigs. Two pregnant SPF sows challenged with isolate 10 gave birth to weak or dead piglets with typical PRRS symptoms.

Virus purification

PRRSV was harvested 24 hr after infection of lung macrophages. The supernatant was first pelleted for 4 hr at 19,000 rpm in a Beckman R19 rotor. After an overnight incubation at 4° in TES (20 m*M* Tris, pH 7, 1 m*M* EDTA, 150 m*M* NaCl) the pellet was resuspended. The virus concentrate was extracted three times with arcton (1,1,2-tri-chloro-tri-fluoro-ethane). The aqueous phase was passed through a 5-μm filter and subsequently through a Sephacryl S400 superfine 55-ml column (C16/40). Chromatography was performed at a flow rate of 0.7 ml/min and fractions of 1 ml were collected. Virus-containing fractions were centrifuged through a sucrose step gradient (0.5 ml of 60%, 1 ml of 45%, and 1 ml of 25% sucrose) in a Beckman SW60 rotor at 50,000 rpm for 30 min. For RNA isolation, the 45/25% interphase was used.

RNA preparation, cDNA synthesis and cloning

Genomic RNA from virions and infected macrophages was isolated according to Chirgwin *et al.* (1979). Preparative cDNA synthesis according to Gubler and Hoffman (1983) was performed on 0.5 μg RNA (approximately 0.1 μg PRRSV genomic RNA) as described (Conzelmann *et al.*, 1991), using 0.2 ng oligo(dT) primer and the second-strand mix from the Pharmacia cDNA synthesis kit. *Eco*RI/*Not*I adaptor ligation and phosphorylation were performed as recommended by the supplier. The cDNA was size selected by preparative agarose gel electrophoresis and cloned in λZAPII (Stratagene) according to the supplier's instructions. Recombinant pBluescript were excised *in vivo* as recommended by the supplier. Labeled first-strand PRRSV cDNA was prepared by using 10 μCi [α-^{32}P] dCTP (3000 Ci/mMol; Amersham) together with 1 m*M* each of dGTP, dATP, and dTTP. The cDNA was separated from not incorporated nucleotides by Sephadex G50 (Pharmacia) chromatography and used directly for Northern hybridizations as described (Conzelmann *et al.*, 1991).

Oligonucleotide hybridization

Deoxyoligonucleotide 3′*M* (5′–TCGGTCACATGGTTCCTGCCT–3′) was synthesized on a Biosearch 8700 DNA Synthesizer and purified by denaturing polyacrylamide gel electrophoresis. Labeling with polynucleotide kinase and hybridization to Northern blots at 61° was done as described (Conzelmann *et al.*, 1991).

Sequence determination and analysis

Both strands of recombinant pBluescript were subjected to unidirectional deletion using Exonuclease III and S1-Nuclease according to Hennikoff (1984) and were sequenced (Sanger *et al.*, 1977) on double-strand plasmid templates according to Zhang *et al.* (1988) using T7-DNA-Polymerase. Computer analysis of the nucleotide and peptide sequences was performed using the UWGCG software (Devereux *et al.*, 1984) on a VAX 4000 (Digital). Protein homologies were calculated using the GAP program, the parameters were gap weight, 3; gap length weight, 0.1.

RESULTS

Synthesis and isolation of PRRSV-specific cDNA clones: Demonstration of PRRSV genomic and subgenomic mRNAs

As determined by denaturing agarose gel electrophoresis and subsequent staining with acridine orange, an RNA species of about 15 kb could be enriched in the course of virion purification from the supernatants of PRRSV-infected macrophages (not shown). For arteriviruses, similar sizes of genomic RNAs have been described, namely 13 kb for EAV and 13 to 14 kb for LDV (Den Boon *et al.*, 1991; Godeny *et al.*, 1990; Kuo *et al.*, 1991). It seemed likely that PRRSV

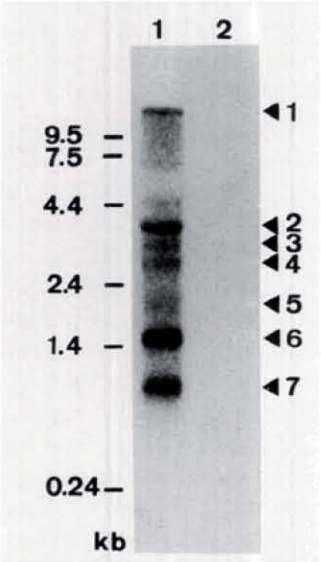

FIG. 1. Demonstration of PRRSV-specific RNAs (arrowheads) in infected macrophages. Oligo(dT)-primed radioactively labeled cDNA of RNA from purified virions was hybridized to total RNA from infected (lane 1) and noninfected macrophages (lane 2).

genomic RNA also terminates with a poly(A) tail. Accordingly, oligo(dT) was employed for priming of an analytical first cDNA strand synthesis, where efficient incorporation of [α-^{32}P]-labeled dCTP occurred (not shown). To investigate whether the labeled cDNA originated mainly from the viral RNA or from contaminating cellular RNAs, Northern hybridizations were performed with total RNA from PRRSV-infected and noninfected macrophages. Prominent specific bands appeared with the RNA of infected cells, while only faint hybridization signals were obtained with RNA from noninfected cells (Fig. 1). The largest hybridizing RNA population corresponded in size to the supposed 15-kb PRRSV genomic RNA. In addition, a set of subgenomic RNAs in the range of 0.9 to 3.7 kb were observed. Thus the bulk of cDNA primed with oligo(dT) originated from PRRSV RNA which is obviously polyadenylated. Using again oligo(dT) as a primer, preparative cDNA synthesis was performed. A cDNA bank was established in λZAPII phages and screened with the labeled single-stranded cDNA probe. Plaques showing strong hybridization signals were isolated and recombinant pBluescript were obtained by *in vivo* excision from λZAPII. The isolated inserts proved to be specific for PRRSV by hybridization experiments with infected and noninfected cells as described above. Northern hybridization showed in addition, that the RNA isolated from purified virions, which was used for cDNA synthesis, consisted mainly of the 15-kb population (Fig. 2).

PRRSV-specific cDNA clones were characterized by restriction mapping and terminal sequencing. Most clones from the oligo(dT) primed cDNA started within a short stretch of nucleotides, which was supposed to be located close to the genomic 3′ end. Clone pPRRSV-T1 had a size of 5 kb and possessed a terminal stretch of four A residues; most likely these represent the first nucleotides of the genomic poly(A) tail (see below).

PRRSV subgenomic mRNAs form a 3′ nested set

The set of subgenomic RNAs detected by the minus-stranded labeled cDNA was also recognized by the isolated and nick-translated cDNA inserts of clones pPRRSV-T1 and pPRRSV-T23 (size of 3.8 kb) both in total RNA (Fig. 2) and in poly(A)$^+$ enriched RNA from infected macrophages (not shown). After hybridization with a labeled 5′ terminal fragment of clone pPRRSV-T1, however, only the 15-kb RNA was detected (not shown). Thus, all prominent subgenomic plus-stranded mRNAs correspond to the 3′ terminal part of the PRRSV genome which is spanned entirely by clone pPRRSV-T1. In order to demonstrate that the mRNAs form a 3′ coterminal nested set, a minus-sense oligonucleotide located four residues from the putative PRRSV genomic 3′ end was synthesized, labeled, and used in Northern hybridization experiments (Fig. 3). The oligonucleotide recognized six major subgenomic RNA populations with sizes of 0.9, 1.4, 2.0, 2.6, 3.1, and 3.7 kb, respectively, along with the 15-kb genomic RNA. According to the EAV nomenclature, PRRSV RNAs were

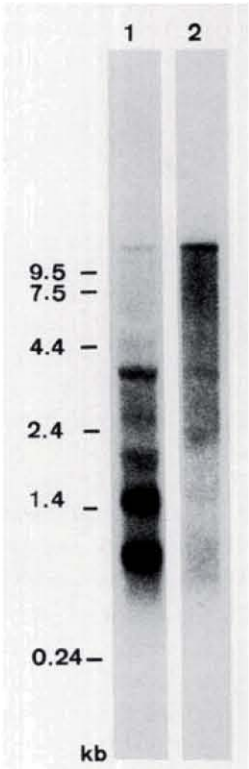

FIG. 2. A Northern blot of RNA from infected macrophages (lane 1) and from purified virions (lane 2) isolated 1 day p.i. was hybridized with the insert of pPRRSV-T1.

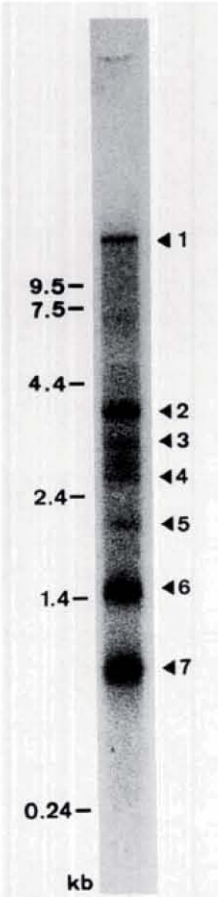

FIG. 3. Northern hybridization of total RNA from infected macrophages with a PRRSV 3′ terminal oligonucleotide. The six major subgenomic mRNAs were numbered according to size, RNA 1 representing genome size RNA.

numbered 1 (genome) through 7 (0.9 kb mRNA). Striking differences were observed in the relative amounts of the subgenomic mRNAs. RNAs 2, 6, and 7 represented the most abundant mRNAs, whereas mRNAs 3, 4, and 5 were present in much lower amounts.

Sequence determination

As mentioned above, pPRRSV-T1 completely encompassed the genomic region from which all prominent subgenomic PRRSV mRNAs are transcribed. In analogy to EAV, this stretch should contain all PRRSV structural protein genes. In addition to pPRRSV-T1, pPRRSV-T23 and a 1.2-kb PvuII fragment of pPRRSV-T19 (Fig. 4) were sequenced entirely to ensure that pPRRSV-T1 did not represent an aberrant PRRSV cDNA clone.

Open reading frames

The resulting sequence of 4920 nucleotides (Fig. 5) revealed a distribution of open reading frames strikingly similar to the EAV genome and to the published 3′ part of the LDV genome. Eight ORFs with coding capacities of more than 100 amino acids were identified on the plus strand (Fig. 4). With the exception of the incomplete ORF at the 5′ end of the determined sequence (nt 2–1603) all ORFs are partially or, in the case of one ORF (x) fully overlapping. Partially overlapping ORFs were numbered 2 through 7 from 5′ to 3′ direction, according to the EAV nomenclature (Den Boon *et al.*, 1991). ORFs 2 through 7 comprise nucleotides 1614–2363, 2222–3019, 2764–3315, 3312–3917, 3905–4426, and 4416–4802, respectively. On the minus-strand, only two ORFs with a coding capacity of more than 100 aa were detected (not shown).

The calculated distance from the start of ORFs 2 through 7 to the putative 3′ end of the PRRSV genome correlates with the observed lengths of the 3′ nested subgenomic mRNAs 2 through 7. The differences of about 400 nucleotides in each case may be explained by the presence of a poly(A) tail at the 3′ ends of the mRNAs and, characteristic for the arterivirus group, the addition of a 5′ leader sequence of about 200 nucleotides to the bodies of the mRNAs (see below). The start codon regions of ORFs 2 through 7 are similar to the consensus sequence for translation initiation (Kozak, 1984).

The short ORF x mentioned above is unique in several aspects. While the other ORFs are partially overlapping, ORF x is located entirely within ORF5. No specific subgenomic mRNA could be correlated to this ORF and, finally, a T is located at position −3 from its putative start ATG codon instead of a purine. It seems likely that ORF x does not encode a PRRSV protein.

Deduced PRRSV proteins

A sequence of 533 aa was deduced from the 5′ terminal incomplete ORF. Amino acid comparisons with EAV proteins revealed an overall homology of 38.0% with the polypeptide encoded by EAV ORF1b, and of 60.0% when similar aa are considered. Two aa stretches (positions 35 to 88 and 297 to 371) were identified in which the amino acid identity reached 61.1 and 57.3%, respectively, without introduction of any gaps. These conserved stretches are located within the described conserved domains 3 and 4, shared by EAV, coronavirus and torovirus 1b polypeptides (Den Boon *et al.*, 1991). Thus, the identified partial PRRSV 5′ ORF apparently represents part of the viral polymerase gene and was denominated ORF1b.

The overlapping ORFs, most likely expressed from the six subgenomic mRNAs, encode a set of small polypeptides (Table 1). In both EAV and LDV, the nucleocapsid protein genes (EAV ORF7, LDV ORF1) are located at the genomic 3′ end. The 3′ terminal PRRSV ORF7 codes for a highly charged protein of 128 aa length. After the introduction of two gaps it displayed significant homology to the LDV capsid protein with

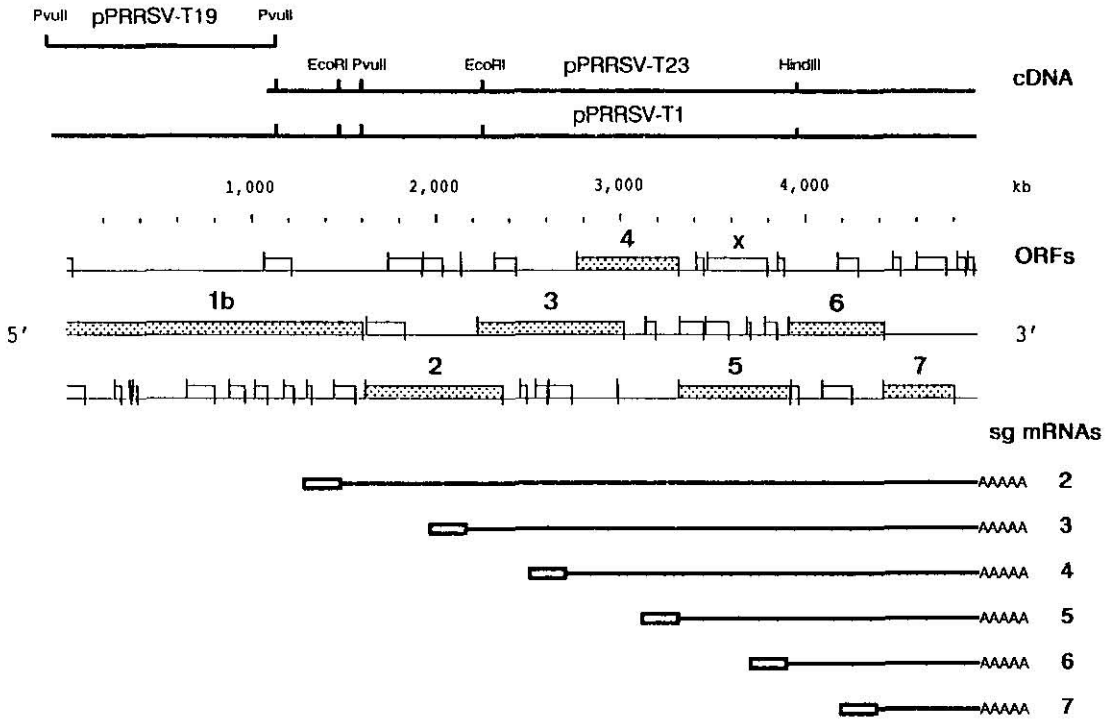

FIG. 4. Distribution of open reading frames in the determined PRRSV sequence, location of subgenomic mRNAs (sg mRNAs, boxes represent the leader RNA), and cDNA clones used for sequencing.

44.0% amino acid identity and 56.9% amino acid similarity. Obviously ORF7 encodes the PRRSV capsid protein gene. In comparison with the EAV nucleocapsid protein only poor homology was found with 23.6% aa identity and 42.5% similarity. However, the three arterivirus capsid proteins display common features; all represent small, highly charged proteins with calculated isoelectric points of 11.06, 12.27, and 11.37 for PRRSV, EAV and LDV, respectively, and show highly hydrophilic stretches especially within the aminoterminal half of the protein. Thus, the genome of PRRSV with the polymerase gene located in the 5′ region and the nucleocapsid gene at the 3′ terminus exhibits a typical arterivirus genome organization.

The products of the ORFs located between polymerase and capsid protein genes are assumed to encode arteriviral envelope proteins. The deduced PRRSV proteins showed only low homology to the respective EAV encoded proteins. In comparison to the published LDV sequences, however, striking similarities were found with regard to the ORF 5 and 6 products. The product of the second 3′ terminal ORF (for LDV called ORF2) is assumed to represent a triple membrane-spanning envelope protein characterized by three hydrophobic domains within the amino-terminal half of the protein (Kuo *et al.*, 1992). PRRSV ORF6 encodes a very similar polypeptide which is three aa longer and shows 53.2% aa identity and 70.8% aa similarity to LDV ORF2 protein, in contrast to 27.2% aa identity and 46.9% aa similarity, respectively, to the corresponding EAV ORF6 protein. These putative membrane proteins show a higher conservation of sequence and also of length (173, 171, and 162 aa, respectively) than the arteriviral capsid proteins. The carboxyterminal part of PRRSV ORF5 and the published corresponding fragment of the LDV ORF3 protein (131 aa) exhibit an aa identity of 47.0% and a similarity of 68.9%. PRRSV and EAV ORF5 proteins show only 20% sequence identity. However, a similar hydropathy plot with a stretch of about 70 hydrophobic residues in the middle of the polypeptides, indicates, that they represent corresponding proteins (not shown).

Due to the lack of sequence information from LDV, the deduced proteins of PRRSV ORFs 2, 3, and 4 could only be compared to the EAV proteins. Of these, only the ORF2 proteins showed a low degree of aa sequence homology (26.3% identity, 52.0% similarity) and similarities in the distribution of hydropathy. Virtually no sequence and hydropathy similarity could be demonstrated between the ORF4 and ORF3 proteins. Common to these proteins is, however, a high number of potential N-glycosylation sites (Table 1; Den Boon *et al.*, 1991).

Signals for leader-joining

Short sequence motifs, TCAAC in EAV and TAACCA or TAAAACC in LDV, have been described which might serve as parts of recognition signals or junction sites for the addition of leaders during transcription of

```
   1  AGCCATCCAGCCTTGTTACAGGGAGAAACTTGAATCTAAGGCTAGGAACACCAGGGTGGTTTTTACCACCCGGCCTGTGGCCTTTGGTCAGGTGCTGACACCATACCATA
       AlaIleGlnProCysTyrArgGluLysLeuGluSerLysAlaArgAsnThrArgValValPheThrThrArgProValAlaPheGlyGlnValLeuThrProTyrHisL   37(ORF1b)

 111  AAGATCGCATCGGCTCTGCGATAACCATAGATTCATCCCAGGGGGCCACCTTTGATATTGTGACATTGCATCTACCATCGCCAAAGTCCCTAAATAAATCCCGAGCACTT
      ysAspArgIleGlySerAlaIleThrIleAspSerSerGlnGlyAlaThrPheAspIleValThrLeuHisLeuProSerProLysSerLeuAsnLysSerArgAlaLeu   73

 221  GTAGCCATCACTCGGGCAAGACACGGGTTGTTCATTTATGACCCTCATAACCAGCTCCAGGAGTTTTTCAACTTAACCCCTGAGCGCACTGATTGTAACCTTGTGTTCAG
      ValAlaIleThrArgAlaArgHisGlyLeuPheIleTyrAspProHisAsnGlnLeuGlnGluPhePheAsnLeuThrProGluArgThrAspCysAsnLeuValPheSe  110

 331  CCGTGGGGATGAGCTGGTAGTTCTGAATGCGGATAATGCAGTCACAACTGTAGCGAAGGCCCTAGAGACAGGTCCATCTCGATTTCGAGTATCAGACCCGAGGTGCAAGT
      rArgGlyAspGluLeuValValLeuAsnAlaAspAsnAlaValThrThrValAlaLysAlaLeuGluThrGlyProSerArgPheArgValSerAspProArgCysLysS  147

 441  CTCTCTTAGCCGCTTGTTCGGCCAGTCTGGAAGGGAGCTGTATGCCACTACCGCAAGTGGCACATAACCTGGGGTTTTACTTTTCCCCGGACAGTCCAGTATTTGCACCT
      erLeuLeuAlaAlaCysSerAlaSerLeuGluGlySerCysMetProLeuProGlnValAlaHisAsnLeuGlyPheTyrPheSerProAspSerProValPheAlaPro  183

 551  CTGCCAAAAGAGTTGGCGCCACATTGGCCAGTGGTTACCCACCAGAATAATCGGGCGTGGCCTGATCGACTTGTCGCTAGTATGCGCCCAATTGATGCCCGCTACAGCAA
      LeuProLysGluLeuAlaProHisTrpProValValThrHisGlnAsnAsnArgAlaTrpProAspArgLeuValAlaSerMetArgProIleAspAlaArgTyrSerLy  220

 661  GCCAATGGTCGGTGCAGGGTATGTGGTCGGGCCGTCCACCTTTCTTGGTACTCCTGGTGTGGTGTCATACTATCTCACACTATACATCAGGGGTGAGCCCCAGGCCTTGC
      sProMetValGlyAlaGlyTyrValValGlyProSerThrPheLeuGlyThrProGlyValValSerTyrTyrLeuThrLeuTyrIleArgGlyGluProGlnAlaLeuP  257

                G
 771  CAGAAACACTCGTTTCAACAGGACGTATAGCCACAGATTGTCGGGAGTATCTCGACGCGGCTGAGGAAGAGGCAGCAAAAGAACTCCCCCACGCATTCATTGGCGATGTC
      roGluThrLeuValSerThrGlyArgIleAlaThrAspCysArgGluTyrLeuAspAlaAlaGluGluGluAlaAlaLysGluLeuProHisAlaPheIleGlyAspVal  293

                                                                A
 881  AAAGGTACCACGGTTGGGGGGTGTCATCACATTACATCAAAATACCTACCTAGGTCCCTGCCTAAGGACTCTGTTGCCGTAGTTGGAGTAAGTTCGCCCGGCAGGGCTGC
      LysGlyThrThrValGlyGlyCysHisHisIleThrSerLysTyrLeuProArgSerLeuProLysAspSerValAlaValValGlyValSerSerProGlyArgAlaAl  330
                                                                Lys
                                  C
 991  TAAAGCCGTGTGCACTCTCACCGATGTGTACCTCCCGGAACTCCGGCCATATCTGCAACCTGAGACGGCATCAAAATGCTGGAAACTCAAATTAGACTTCAGGGACGTCC
      aLysAlaValCysThrLeuThrAspValTyrLeuProGluLeuArgProTyrLeuGlnProGluThrAlaSerLysCysTrpLysLeuLysLeuAspPheArgAspValA  367

1101  GACTAATGGTCTGGAAAGGAGCCACCGCCTATTTCCAGTTGGAAGGGCTTACATGGTCGGCGCTGCCCGACTATGCCAGGTTTATTCAGCTGCCCAAGGATGCCGTTGTA
      rgLeuMetValTrpLysGlyAlaThrAlaTyrPheGlnLeuGluGlyLeuThrTrpSerAlaLeuProAspTyrAlaArgPheIleGlnLeuProLysAspAlaValVal  403

1211  TACATTGATCCGTGTATAGGACCGGCGACAGCCAACCGTAAGATCGTGCGAACCACAGACTGGCGGGCCGACCTGGCAGTGACACCGTATGATTACGGTGCCCAGAACAT
      TyrIleAspProCysIleGlyProAlaThrAlaAsnArgLysIleValArgThrThrAspTrpArgAlaAspLeuAlaValThrProTyrAspTyrGlyAlaGlnAsnIl  440

1321  TTTGACAACAGCCTGGTTCGAGGACCTCGGGCCGCAGTGGAAGATTTTGGGGCTGCAGCCCTTTAGGCGAGCATTTGGCTTTGAAAACACTGAGGATTGGGCAATCCTTG
      eLeuThrThrAlaTrpPheGluAspLeuGlyProGlnTrpLysIleLeuGlyLeuGlnProPheArgArgAlaPheGlyPheGluAsnThrGluAspTrpAlaIleLeuA  477

1431  CACGCCGTATGAATGACGGCAAGGACTACACTGACTATAACTGGAACTGTGTTCGAGAACGCCCACACGCCATCTACGGGCGTGCTCGTGACCATACGTATCATTTTGCC
      laArgArgMetAsnAspGlyLysAspTyrThrAspTyrAsnTrpAsnCysValArgGluArgProHisAlaIleTyrGlyArgAlaArgAspHisThrTyrHisPheAla  513

                                  ****
1541  CCTGGCACAGAATTGCAGGTAGAGCTAGGTAAACCCCGGCTGCCGCCTGGGCAAGTGCCGTGAATTCGGAGTGATGCAATGGGGTCACTGTGGAGTAAAATCAGCCAGCT
      ProGlyThrGluLeuGlnValGluLeuGlyLysProArgLeuProProGlyGlnValProEnd          MetGlnTrpGlyHisCysGlyValLysSerAlaSerC   13(ORF2)

1651  GTTCGTGGACGCCTTCACTGAGTTCCTTGTTAGTGTGGTTGATATTGTCATTTTCCTTGCCATACTGTTTGGGTTCACCGTCGCAGGATGGTTACTGGTCTTTCTTCTCA
      ysSerTrpThrProSerLeuSerSerLeuLeuValTrpLeuIleLeuSerPheSerLeuProTyrCysLeuGlySerProSerGlnAspGlyTyrTrpSerPhePheSer   49

1761  GAGTGGTTTGCTCCGCGCTTCTCCGTTCGCGCTCTGCCATTCACTCTCCCGAACTATCGAAGGTCCTATGAAGGCTTGTTGCCCAACTGCAGACCGGATGTCCCACAATT
      GluTrpPheAlaProArgPheSerValArgAlaLeuProPheThrLeuProAsnTyrArgArgSerTyrGluGlyLeuLeuProAsnCysArgProAspValProGlnPh   86

1871  TGCAGTCAAGCACCCATTGGGTATGTTTTGGCACATGCGAGTTTCCCACCTAATTGATGAGATGGTCTCTCGTCGCATTTACCAGACCATGGAACATTCAGGTCAAGCGG
      eAlaValLysHisProLeuGlyMetPheTrpHisMetArgValSerHisLeuIleAspGluMetValSerArgArgIleTyrGlnThrMetGluHisSerGlyGlnAlaA  123

1981  CCTGGAAGCAGGTGGTTGGTGAGGCCACTCTCACGAAGCTGTCAGGGCTCGATATAGTTACTCATTTCCAACACCTGGCCGCAGTGGAGGCGGATTCTTGCCGCTTTCTC
      laTrpLysGlnValValGlyGluAlaThrLeuThrLysLeuSerGlyLeuAspIleValThrHisPheGlnHisLeuAlaAlaValGluAlaAspSerCysArgPheLeu  159

1091  AGCTCACGACTCGTGATGCTAAAAAATCTTGCCGTTGGCAATGTGAGCCTACAGTACAACACCACGTTGGACCGCGTTGAGCTCATCTTCCCCACGCCAGGTACGAGGCC
      SerSerArgLeuValMetLeuLysAsnLeuAlaValGlyAsnValSerLeuGlnTyrAsnThrThrLeuAspArgValGluLeuIlePheProThrProGlyThrArgPr  196

2201  CAAGTTGACCGACTTCAGACAATGGCTCGTCAGCGTGCACGCTTCCATTTTTTCCTCTGTGGCTTCATCTGTTACCTTGTTCATAGTGCTTTGGCTTCGAATTCCAGCTC
      oLysLeuThrAspPheArgGlnTrpLeuValSerValHisAlaSerIlePheSerSerValAlaSerSerValThrLeuPheIleValLeuTrpLeuArgIleProAlaL  233
                          MetAlaArgGlnArgAlaArgPheHisPhePheLeuCysGlyPheIleCysTyrLeuValHisSerAlaLeuAlaSerAsnSerSerSe   30(ORF3)

                                                                                T
2311  TACGCTATGTTTTTGGTTTCCATTGGCCCACGGCAACACATCATTCGAGCTGACCATCAACTACACCATATGCATGCCCTGTTCTACCAGTCAAGCGGCTCGCCAAAGGC
      euArgTyrValPheGlyPheHisTrpProThrAlaThrHisHisSerSerEnd                                                            249
      rThrLeuCysPheTrpPheProLeuAlaHisGlyAsnThrSerPheGluLeuThrIleAsnTyrThrIleCysMetProCysSerThrSerGlnAlaAlaArgGlnArgL   67
                                                                                Ala
2421  TCGAGCCCGGTCGTAACATGTGGTGCAAAATAGGGCATGACAGGTGTGAGGAGCGTGACCATGATGAGTTGTTAATGTCCATCCCGTCCGGGTACGACAACCTCAAACTT
      euGluProGlyArgAsnMetTrpCysLysIleGlyHisAspArgCysGluGluArgAspHisAspGluLeuLeuMetSerIleProSerGlyTyrAspAsnLeuLysLeu  103

                                                T
2531  GAGGGTTATTATGCTTGGCTGGCTTTTTTGTCCTTTTCCTACGCGGCCCAATTCCATCCGGAGTTGTTCGGGATAGGGAATGTGTCGCGCGTCTTCGTGGACAAGCGACA
      GluGlyTyrTyrAlaTrpLeuAlaPheLeuSerPheSerTyrAlaAlaGlnPheHisProGluLeuPheGlyIleGlyAsnValSerArgValPheValAspLysArgHi  140

2641  CCAGTTCATTTGTGCCGAGCATGATGGACCCAATTCAACCGTATCTACCGGACACAACATCTCCGCATTATATGCGGCATATTACCACCACCAAATAGACGGGGGCAATT
      sGlnPheIleCysAlaGluHisAspGlyProAsnSerThrValSerThrGlyHisAsnIleSerAlaLeuTyrAlaAlaTyrTyrHisHisGlnIleAspGlyGlyAsnT  177
```

FIG. 5. Nucleotide sequence of pPRRSV-T1 representing the 3′ terminal 4.9 kb of the PRRSV genome and deduced amino acid sequences of PRRSV proteins. Putative leader/RNA junction sites are indicated by stars. Nucleotide and amino acid sequence ambiguities in pPRRSV-T19, and pPRRSV-T23 are indicated above and below the pPRRSV-T1 sequences, respectively.

```
2751  GGTTCCATTTGGAATGGCTGCGGCCACTCTTTTCCTCCTGGCTGGTGCTCAACATATCATGGTTTCTGAGGCGTTCGCCTGTAAGCCCTGTTTCTCGACGCATCTATCAG
      rpPheHisLeuGluTrpLeuArgProLeuPheSerSerTrpLeuValLeuAsnIleSerTrpPheLeuArgArgSerProValSerProValSerArgArgIleTyrGln  213
                  MetAlaAlaAlaThrLeuPheLeuLeuAlaGlyAlaGlnHisIleMetValSerGluAlaPheAlaCysLysProCysPheSerThrHisLeuSerA   33(ORF4)

2861  ATATTGAGACCAACACGACCGCGGCTGCCGGTTTCATGGTCCTTCAGGACATCAATTGTTTCCGACCTCACGGGGTCTCAGCAGCGCAAGAGAAAATTTCCTTCGGAAAG
      IleLeuArgProThrArgProArgLeuProValSerTrpSerPheArgThrSerIleValSerAspLeuThrGlySerGlnGlnArgLysArgLysPheProSerGluSe  250
      spIleGluThrAsnThrThrAlaAlaAlaGlyPheMetValLeuGlnAspIleAsnCysPheArgProHisGlyValSerAlaAlaGlnGluLysIleSerPheGlyLys   69

2971  TCGTCCCAATGTCGTGAAGCCGTCGGTACTCCCCAGTACATCACGATAACGGCTAACGTGACCGACGAATCATACTTGTACAACGCGGACTTGCTGATGCTTTCTGCGTG
      rArgProAsnValValLysProSerValLeuProSerThrSerArgEnd                                                              265
      SerSerGlnCysArgGluAlaValGlyThrProGlnTyrIleThrIleThrAlaAsnValThrAspGluSerTyrLeuTyrAsnAlaAspLeuLeuMetLeuSerAlaCy  106

3081  CCTTTTCTACGCCTCAGAAATGAGCGAGAAAGGCTTCAAAGTTATCTTTGGGAATGTCTCTGGCGTTGTTTCCGCTTGTGTCAATTTCACAGATTATGTGGCCCATGTGA
      sLeuPheTyrAlaSerGluMetSerGluLysGlyPheLysValIlePheGlyAsnValSerGlyValValSerAlaCysValAsnPheThrAspTyrValAlaHisValT  143

                                                                                                  ****
3191  CCCAACATACCCAGCAGCATCATCTGGTAATTGATCACATTCGGTTGCTGCATTTCCTGACACCATCTGCAATGAGGTGGGCTACAACCATTGCTTGTTTGTTCGCCATT
      hrGlnHisThrGlnGlnHisHisLeuValIleAspHisIleArgLeuLeuHisPheLeuThrProSerAlaMetArgTrpAlaThrThrIleAlaCysLeuPheAlaIle  179

3301  CTCTTGGCGATATGAGATGTTCTCACAAATTGGGGCGTTTCTTGACTCCGCACTCTTGCTTCTGGTGGCTTTTTTTGCTGCCTACCGGCTTGTCCTGGTCCTTTGCCGAT
      LeuLeuAlaIleEnd                                                                                                 183
                MetArgCysSerHisLysLeuGlyArgPheLeuThrProHisSerCysPheTrpTrpLeuPheLeuLeuProThrGlyLeuSerTrpSerPheAlaAsp   33(ORF5)

3411  GGCAACGGCGACAGCTCGACATACCAATACATATATAACTTGACGATATGCGAGCTGAATGGGACCGACTGGTTGTCCAGCCATTTTGGTTGGGCAGTCGAGACCTTTGT
      GlyAsnGlyAspSerSerThrTyrGlnTyrIleTyrAsnLeuThrIleCysGluLeuAsnGlyThrAspTrpLeuSerSerHisPheGlyTrpAlaValGluThrPheVa   70

3521  GCTTTACCCAGTTGCCACTCATATCCTCTCACTGGGTTTTCTCACAACAAGCCATTTTTTTGACGCGCTCGGTCTCGGCGTTGTATCCACTGCAGGACTTGTTGGCGGGC
      lLeuTyrProValAlaThrHisIleLeuSerLeuGlyPheLeuThrThrSerHisPhePheAspAlaLeuGlyLeuGlyValValSerThrAlaGlyLeuValGlyGlyA  107

3631  GGTACGTACTCTGCAGCGTCTACGGCGCTTGTGCTTTCGCAGCGTTCGTATGTTTTGTCATCCGTGCTGCTAAAAATTGCATGGCCTGCCGCTATGCCCGTACCCGGTTT
      rgTyrValLeuCysSerValTyrGlyAlaCysAlaPheAlaAlaPheValCysPheValIleArgAlaAlaLysAsnCysMetAlaCysArgTyrAlaArgThrArgPhe  143

                                                  A
3741  ACCAACTTCATTGTGGACGACCGGGGGAGAGTCCATCGATGGAGGTCTCCAATAGTGGTAGAAAAATTGGGCAAAGCCGAAGTCGACGGCAACCTCGTCACCATTAAACA
      ThrAsnPheIleValAspAspArgGlyArgValHisArgTrpArgSerProIleValValGluLysLeuGlyLysAlaGluValAspGlyAsnLeuValThrIleLysHi  180
                                                  Lys
                                  ****
3851  TGTCGTCCTCGAAGGGGTTAAAGCTCAACCCTTGACGAGGACTTCGGCTGAGCAATGGGAGGCCTAGATGATTTTTGCAACGATCCTATCGCCGCACAAAAGCTTGTGCT
      sValValLeuGluGlyValLysAlaGlnProLeuThrArgThrSerAlaGluGlnTrpGluAlaEnd                                            201
                                                                  MetGlyGlyLeuAspAspPheCysAsnAspProIleAlaAlaGlnLysLeuValLe   19(ORF6)

3961  AGCCTTTAGCATCACATACACACCTATAATGATATACGCCCTTAAGGTGTCACGCGGCCGACTCCTGGGGCTGTTGCACATCCTAATATTTCTGAATTGTTCCTTTACAT
      uAlaPheSerIleThrTyrThrProIleMetIleTyrAlaLeuLysValSerArgGlyArgLeuLeuGlyLeuLeuHisIleLeuIlePheLeuAsnCysSerPheThrP   56

4071  TCGGATACATGACATATGTGCATTTTCAATCCACCAACCGTGTCGCACTTACCCTGGGGGCTGTTGTCGCCCTTCTGTGGGGTGTTTACAGCTTCACAGAGTCATGGAAG
      heGlyTyrMetThrTyrValHisPheGlnSerThrAsnArgValAlaLeuThrLeuGlyAlaValValAlaLeuLeuTrpGlyValTyrSerPheThrGluSerTrpLys   92

4181  TTTATCACTTCCAGATGCAGATTGTGTTGCCTTGGCCGGCGATACATTCTGGCCCCTGCCCATCACGTAGAAAGTGCTGCAGGTCTCCATTCAATCTCAGCGTCTGGTAA
      PheIleThrSerArgCysArgLeuCysCysLeuGlyArgArgTyrIleLeuAlaProAlaHisHisValGluSerAlaAlaGlyLeuHisSerIleSerAlaSerGlyAs  129

4291  CCGAGCATACGCTGTGAGAAAGCCCGGACTAACATCAGTGAACGGCACTCTAGTACCAGGACTTCGGAGCCTCGTGCTGGGCGGCAAACGAGCTGTTAAACGAGGAGTGG
      nArgAlaTyrAlaValArgLysProGlyLeuThrSerValAsnGlyThrLeuValProGlyLeuArgSerLeuValLeuGlyGlyLysArgAlaValLysArgGlyValV  166

        ****
4401  TTAACCTCGTCAAGTATGGCCGGTAAAAACCAGAGCCAGAAGAAAAAGAAAAGTACAGCTCCGATGGGGAATGGCCAGCCAGTCAATCAACTGTGCCAGTTGCTGGGTGC
      alAsnLeuValLysTyrGlyArgEnd                                                                                      173
                    MetAlaGlyLysAsnGlnSerGlnLysLysLysLysSerThrAlaProMetGlyAsnGlyGlnProValAsnGlnLeuCysGlnLeuLeuGlyAl   32(ORF7)

4511  AATGATAAAGTCCCAGCGCCAGCAACCTAGGGGAGGACAGGCCAAAAAGAAAAAGCCTGAGAAGCCACATTTTCCCCTGGCTGCTGAAGATGACATCCGGCACCACCTCA
      aMetIleLysSerGlnArgGlnGlnProArgGlyGlyGlnAlaLysLysLysLysProGluLysProHisPheProLeuAlaAlaGluAspAspIleArgHisHisLeuT   69

4621  CCCAGACTGAACGCTCCCTCTGCTTGCAATCGATCCAGACGGCTTTCAATCAAGGCGCAGGAACTGCGTCGCTTTCATCCAGCGGGAAGGTCAGTTTTCAGGTTGAGTTT
      hrGlnThrGluArgSerLeuCysLeuGlnSerIleGlnThrAlaPheAsnGlnGlyAlaGlyThrAlaSerLeuSerSerSerGlyLysValSerPheGlnValGluPhe  105

4731  ATGCTGCCGGTTGCTCATACAGTGCGCCTGATTCGCGTGACTTCTACATCCGCCAGTCAGGGTGCAAGTTAATTTGACAGTCAGGTGAATGGCCGCGATTGGCGTGTGGC
      MetLeuProValAlaHisThrValArgLeuIleArgValThrSerThrSerAlaSerGlnGlyAlaSerEnd                                       128

4841  CTCTGAGTCACCTATTCAATTAGGGCGATCACATGGGGGTCATACTTAATCAGGCAGGAACCATGTGACCGAAATTAAAA(A)n
```

FIG. 5—*Continued*

mRNAs (De Vries *et al.*, 1990; Kuo *et al.*, 1992). A similar sequence motif, AACC, is common to the regions preceding four of the identified PRRSV ORF translational start codons (ORF2 ATG −42; ORF5 ATG −36, ORF6 ATG −28, ORF7 ATG −13). Interestingly, an extended common sequence, GNTNAACC, (N for any nucleotide) precedes ORFs 2, 6, and 7 (Fig. 5); the corresponding mRNAs 2, 6, and 7 represent the most abundant PRRSV specific RNAs (Figs. 1–3).

3′ Terminal noncoding sequence

The PRRSV noncoding sequence following the ORF7 stop codon down to the putative 3′ end is 114

TABLE 1

CHARACTERISTICS OF DEDUCED PRRSV PROTEINS

mRNAs	kb	ORF No.	Amino acids encoded	Mr (kDa)	Potential N-glycosylation sites
2	3.7	ORF2	249	28.4	2
3	3.1	ORF3	265	30.6	7
4	2.6	ORF4	183	20.0	4
5	2.0	ORF5	201	22.4	2
6	1.4	ORF6	173	18.9	2
7	0.9	ORF7	128	13.9	1

nucleotides in length, thus exceeding LDV and EAV 3′ noncoding sequences by 34 and 55 residues, respectively. Only few nucleotides at the polyadenylation sites are conserved in PRRSV and LDV showing the consensus sequence C–C–G–G/A–A–A–T–T–poly(A). The polyadenylation site of the EAV genome is similar, with three deviations from the PRRSV/LDV consensus sequence (Fig. 6a). Interestingly, a sequence identical to the EAV polyadenylation site motif is found upstream of the PRRSV polyadenylation site. In addition, 13 of the preceding 26 nucleotides are identical (Fig. 6b).

DISCUSSION

The recent epidemic "porcine respiratory and reproductive syndrome" (PRRS) is caused by a previously unknown virus. The presented molecular data clearly demonstrate that PRRSV is a member of the "arterivirus" group, as was already suspected because of similar ultrastructure. Thus far, the arterivirus group included EAV, representing the only member of the Arterivirus genus within the Togaviridae family, as well as the unclassified togaviruses LDV and simian hemorrhagic fever virus (SHFV).

The PRRSV genomic RNA, with an estimated size of nearly 15 kb, is the largest of the arterivirus group; viral RNAs from LDV, EAV, and SHFV range between 12 and 13 kb (Plagemann and Moennig, 1992). Sequence analysis of the 3′ terminal 5 kb of the PRRSV genome revealed that its organization is identical to that of EAV. As in EAV, six small, partially overlapping ORFs are expressed via transcription of a set of 3′ coterminal subgenomic mRNAs. There is evidence that PRRSV mRNAs also possess a common 5′ leader sequence (not shown). As suggested by the comparison of deduced protein sequences, the gene order is similar to the one of EAV. The PRRSV nucleocapsid gene is located at the 3′ end of the genome and is preceded by an ORF encoding a well-conserved putative membrane-spanning protein. The four upstream ORFs probably encode glycosylated envelope proteins of lower conservation. PRRSV and EAV ORF2 proteins and ORF5 polypeptides display some similarity in hydropathy, indicating that they might represent corresponding proteins, but such similarities could not be observed for ORF3 or ORF4 polypeptides. While the ORF3 polypeptides show the most striking length difference of the deduced PRRSV and EAV proteins (265 and 163 aa, respectively), they resemble each other by possessing the highest number of potential N-glycosylation

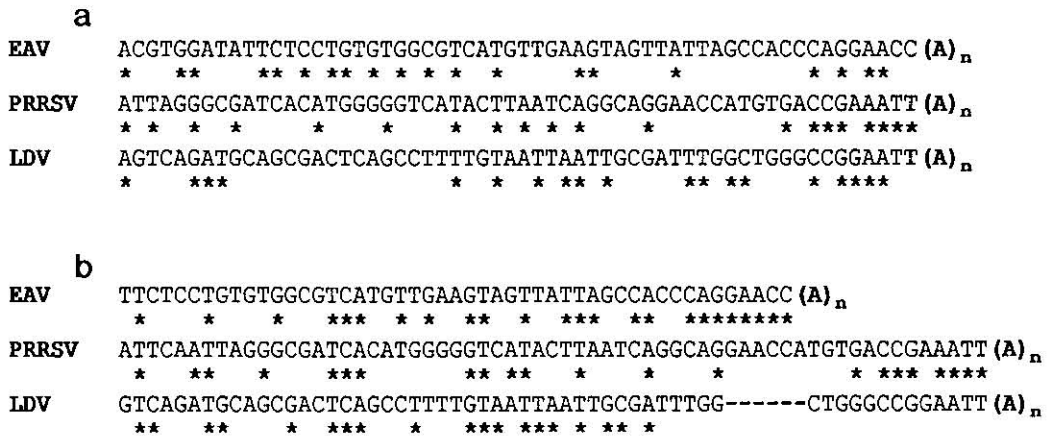

FIG. 6. Alignment of 3′ terminal sequences of PRRSV, LDV, and EAV. Common nucleotides are marked by stars. (a) 3′ coterminal alignment; (b) alignment of the EAV 3′ terminal sequence to an upstream PRRSV sequence.

sites (6 and 7, respectively). Thus, they may represent highly divergent viral glycoproteins.

Clear homologies exist between PRRSV ORF7, 6, and 5 polypeptides and the published LDV protein sequences (LDV ORF 1, 2, and part of ORF3, respectively) with 44.8, 53.2, and 47.0% amino acid identity. Thus, as indicated already by the data obtained from the comparison of PRRSV and EAV proteins, the putative membrane protein (ORF6) is apparently the most conserved structural protein among members of the arterivirus group. Interestingly, a slightly higher variation of the LDV membrane proteins in contrast to the capsid proteins was observed when LDV isolates were compared, which differ in neurovirulence (Kuo *et al.*, 1992).

The comparative protein sequence analyses demonstrate, that PRRSV is much closer related to LDV than to EAV. Surprisingly, seven instead of six subgenomic RNAs could be identified in LDV-infected cells (Kuo *et al.*, 1991, 1992). It remains to be determined, whether LDV actually possesses an additional ORF encoding a protein.

The general mechanisms of transcription are apparently highly similar for all arteriviruses. In EAV, a leader RNA of 207 nucleotides derived from the genome 5′ end is linked to defined sites of the subgenomic RNAs, in a way that the respective ORF translational start codon represents the first AUG of the mRNA (De Vries *et al.*, 1990). The length of the PRRSV leader has not yet been determined exactly, but it exceeds 200 nucleotides (unpublished results). While in EAV every ORF is preceded by the junction site motif TCAAC (Den Boon *et al.*, 1991), for LDV, different motifs have been suggested, TAACCA in RNAs 5, 6, and 8, and TAAAACC in RNA 7 (Kuo *et al.*, 1992). Common putative junction motifs exceeding three nucleotides could not be identified in the PRRSV sequence, however, ORFs 7, 6, 5, and 2 are preceded by AACC, which represents part of the LDV motif. As in EAV, the PRRSV mRNAs 7, 6, and 2 represent the predominant subgenomic RNAs. Interestingly, the corresponding ORFs show a common extended "junction motif" GNTNAACC.

The terminal noncoding regions of many positive strand virus genera contain areas of high sequence conservation (Strauss and Strauss, 1988). The 3′ terminal noncoding regions of PRRSV, LDV, and EAV are heterogeneous in length (114, 80, and 59 nucleotides, respectively). Of the 10 nucleotides adjacent to the poly(A) tail, 8 are identical in PRRSV and LDV, and only five are conserved in the EAV genome. Notably, an octanucleotide sequence identical to the EAV polyadenylation site, is found in the PRRSV genome 22 nucleotides upstream. Aligned accordingly, the terminal PRRSV noncoding region shows more similarity to the 3′ noncoding region of EAV than to that of LDV, which lacks this conserved octanucleotide. In addition, after introduction of a gap in the LDV sequence some residues identical in all three virus sequences are observed (Fig. 6). This might be important with regard to polymerase recognition or replication initiation and it will be interesting, to compare the respective sequences from other strains and the fourth member of the arterivirus group, SHFV.

Despite a smaller size, the polymerase gene of EAV is organized similar to the ones of coronaviruses and toroviruses. All three consist of two overlapping ORFs, 1a and 1b, which are expressed most likely by a mechanism of ribosomal −1 frame-shifting (Brierly *et al.*, 1987; Snijder *et al.*, 1990; Den Boon *et al.*, 1991). In the 1b ORFs four conserved domains were identified in corresponding positions (Snijder *et al.*, 1990) which include motifs conserved in many RNA viruses, such as the GDD motif (domain 1) (Argos, 1988; Gorbalenya and Koonin, 1988) and the nucleotide triphosphate-binding helicase (GKS/T) motif (domain 3) (Gorbalenya *et al.*, 1988; Hodgman, 1988). In the determined PRRSV ORF1b protein sequence, domain 4 and part of domain 3 were identified. It is assumed that the remaining 5′ terminal 10 kb of the PRRSV genome represent the PRRSV polymerase gene, most likely in the form of two overlapping reading frames. According to the observed difference in genome size it should be at least 2 kb larger than the EAV polymerase gene.

It was shown here for an additional member of the arterivirus group, that these viruses share highly similar genome organization, expression strategy, and transcription mechanisms. Both the organization of the polymerase genes which supposedly possess a common ancestor (Den Boon *et al.*, 1991) and the transcription of multiple 3′ terminal mRNAs suggest relationships of arteriviruses, coronaviruses, and toroviruses. As a consequence the establishment of a new "coronavirus-like" superfamily was proposed (Den Boon *et al.*, 1990; Snijder *et al.*, 1990; Spaan *et al.*, 1990). However, the smaller genome size (about half of coronavirus genome), the extent of the leader (approximately triple of coronavirus leader), the overlapping of reading frames and also morphological data clearly distinguish arteriviruses from coronaviruses. It remains to be determined, whether the mode of arteri- and coronavirus transcription is identical. According to the current data coronavirus RNAs are generated independently by leader primed transcription (review; Lai, 1986; Spaan *et al.*, 1988), and in addition, minus sense subgenomic RNAs might function as replicons (Sethna *et al.*, 1989; Sawicki and Sawicki, 1990). In contrast, EAV UV transcription mapping had indicated the presence of genome length precursors (Van Berlo *et al.*, 1982) and thus alternative RNA splicing mechanisms were assumed to be responsible for EAV leader/mRNA linking (De Vries *et al.*, 1990). However, coronavirus-like mechanisms are not excluded (Den Boon *et al.*,

1991). Further detailed analysis of arterivirus transcription mechanisms will elucidate their evolutionary relationship with coronaviruses.

The common molecular features of the so-far four "arteriviruses" of simian (SHFV), porcine (PRRSV), murine (LDV), and equine origin (EAV) not only separate them from Togaviridae, but may justify the establishment of a new virus family. The term "Arteriviridae" refers only to the disease caused by EAV, while "Multiviridae" (Plagemann and Moennig, 1992) stresses the multiple subgenomic mRNAs. As one possible alternative, we would like to suggest as the new virus family name "Mamurnaviridae" (*ma*crophage *mu*ltiple *RNA* viruses) which also points toward the common group of target cells.

ACKNOWLEDGMENT

We thank Veronika Schlatt for perfect technical assistance.

REFERENCES

ARGOS, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* **16**, 9909–9916.

BENFIELD, D. A., NELSON, E., COLLINS, J. E., HARRIS, L., GOYAL, S. M., ROBISON, D., CHRISTIANSON, W. T., MORRISON, R. B., GORCYCA, D., and CHLADEK, D. (1992). Characterization of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332). *J. Vet. Diagn. Invest.* **4**, 127–133.

BRIERLEY, I., BOURSNELL, E. G., BINNS, M. M., BILIMORIA, B., BLOCK, V. C., BROWN, T. D. K., and INGLIS, S. C. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**, 3779–3785.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J., and RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **24**, 5294–5299.

COLLINS, J. E., BENFIELD, D. A., and CHRISTIANSON, W. T. (1991). Swine infertility and respiratory syndrome. *Proc. Minn. Swine Conf. Vet.* **1**, 200–205.

COLLINS, J. E., BENFIELD, D. A., CHRISTIANSON, W. T., HARRIS, L., HENNINGS, J. C., SHAW, D. P., GOYAL, S. M., MCCULLOUGH, S., MORRISON, R. B., JOO, H. S., GORCYCA, D., and CHLADEK, D. (1992). Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. *J. Vet. Diagn. Invest.* **4**, 117–126.

CONZELMANN, K. K., COX, J. H., and THIEL, H.-J. (1991). An L (polymerase) deficient rabies virus defective interfering particle RNA is replicated and transcribed by heterologous helper virus proteins. *Virology* **184**, 655–663.

DEN BOON, J. A., SNIJDER, E. J., CHIRNSIDE, E. D., DE VRIES, A. A. F., HORZINEK, M. C., and SPAAN, W. J. M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronavirus superfamily. *J. Virol.* **65**, 2910–2920.

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

DE VRIES, A. A. F., CHIRNSIDE, E. D., BREDENBEEK, P. J., GRAVESTEIN, L. A., HORZINEK, M. C., and SPAAN, W. J. M. (1990). All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res.* **18**, 3241–3247.

GODENY, E. K., SPEICHER, D. W., and BRINTON, M. A. (1990). Map location of lactate dehydrogenase-elevating virus (LDV) capsid protein (Vp1) gene. *Virology* **177**, 768–771.

GORBALENYA, A. E., and KOONIN, E. V. (1988). Birnavirus RNA polymerase is related to polymerases of positive strand RNA viruses. *Nucleic Acids Res.* **16**, 7735.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1988). A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* **235**, 16–24.

GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HENNIKOFF, S. (1984). Unidirectional digestion of exonuclease III in DNA sequence analysis. *In* "Methods in Enzymology" (R. Wu, Ed.), Vol. 155, pp. 156–165. Academic Press, New York.

HODGMAN, T. C. (1988). A new superfamily of replicative proteins. *Nature (London)* **335**, 22–23.

KEFFABER, K. K. (1989). Reproductive failure of unknown etiology. *Am. Assoc. Swine Pract. Newslett.* **1**, 1–9.

KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**, 857–872.

KUO, L., HARTY, J. T., ERICKSON, L., PALMER, G. A., and PLAGEMANN, P. G. W. (1991). A nested set of eight RNAs is formed in macrophages infected with lactate dehydrogenase-elevating virus. *J. Virol.* **65**, 5118–5123.

KUO, L., CHEN, Z., ROWLAND, R. R. R., FAABERG, K. S., and PLAGEMANN, P. G. W. (1992). Lactate dehydrogenase-elevating virus (LDV): Subgenomic mRNAs, mRNA leader, and comparison of 3′ terminal sequences of two LDV isolates. *Virus Res.* **23**, 55–72.

LAI, M. M. C. (1986). Coronavirus leader-RNA primed transcription: An alternative mechanism to RNA splicing. *Bioessays* **5**, 257–260.

LINDHAUS, W., and LINDHAUS, B. (1991). Rätselhafte Schweinekrankheit. Der prakt. *Tierarzt* **5**, 423–425.

OHLINGER, V., WEILAND, F., HAAS, B., VISSER, N., AHL, R., METTENLEITER, T. C., WEILAND, E., RZIHA, H. J., SAALMÜLLER, A., and STRAUB, O. C. (1991). Der "Seuchenhafte Spätabort beim Schwein"—Ein Beitrag zur Ätiologie des "porcine reproductive and respiratory syndrome (PRRS)". *Tierärztl. Umsch.* **46**, 703–708.

PLAGEMANN, P. G. W., and MOENNIG, V. (1992). Lactate dehydrogenase-elevating virus, equine arteritis virus, and simian hemorrhagic fever virus. *Adv. Virus Res.* **41**, 99–192.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

SAWICKI, S. G., and SAWICKI, D. L. (1990). Coronavirus transcription: Subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* **64**, 1050–1056.

SETHNA, P. B., HUNG, S. L., and BRIAN, D. A. (1989). Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626–5630.

SNIJDER, E. J., DEN BOON, J. A., BREDENBEEK, P. J., HORZINEK, M. C., RIJNBRAND, R., and SPAAN, W. J. M. (1990). The carboxy-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionary related. *Nucleic Acids Res.* **18**, 4535–4542.

SPAAN, W. J. M., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **42**, 755–759.

SPAAN, W. J. M., DEN BOON, J. A., BREEDENBEEK, P. J., CHIRNSIDE, E. D., NOTEN, A. F. H., SNIJDER, E. J., DE VRIES, A. A. F., and HORZINEK, M. C. (1990). Comparative and evolutionary aspects of coronaviral, arteriviral, and toroviral genome structure and expression. *In* "New Aspects of Positive Strand RNA Viruses" (M. A. Brinton and F. X. Heinz, Eds.), pp. 12–16. Am. Soc. Microbiol., Washington, DC.

STRAUSS, J. H., and STRAUSS, E. G. (1988). Evolution of RNA viruses. *Annu. Rev. Biochem.* **42**, 657–683.

TERPSTRA, C., WENSVOORT, G., and POL, J. M. A. (1991). Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad virus: Koch's postulate fulfilled. *Vet. Q.* **13**, 131–136.

VAN BERLO, M. F., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1982). Equine arteritis virus-infected cells contain six polyadenylated virus-specific RNAs. *Virology* **118**, 3345–3352.

WENSVOORT, G., TERPSTRA, C., POL, J. M. A., TER LAAK, E. A., BLOEMRAD, M., DE KLUYVER, E. P., KRAGTEN, C., VAN BUITEN, L., DEN BESTEN, A., WAGENAAR, F., BROEKHUIJSEN, J. M., MOONEN, P. L. J. M., ZETSTRA, T., DE BOER, E. A., TIBBEN, H. J., DE JONG, M. F., VAN'T VELT, P., GROENLAND, G. J. R., VAN GENNEP, J. A., VOETS, M. T., VERHEIJDEN, J. H. M., and BRAAMSKAMP, J. (1991a). Mystery swine disease in the Netherlands: The isolation of Lelystad virus. *Vet. Q.* **13**, 121–130.

WENSVOORT, G., TERPSTRA, C., POL, J. M. A., and WAGENAAR, F. (1991b). Lelystad virus, the cause of porcine epidemic abortion and respiratory syndrome (mystery swine disease). *In* "Proc. Porcine Reprod. Respir. Syndrome Semin.," November 4–5, pp. 27–35. European Commission VI/B/II.2, Brussels, Belgium.

WENSVOORT, G., DE KLUYVER, E. P., LUIJTZE, E. A., DEN BESTEN, A., HARRIS, L., COLLINS, J. E., CHRISTIANSON, W. T., and CHLADEK, D. (1992). Antigenic comparison of Lelystad virus and swine infertility and respiratory syndrome (SIRS) virus. *J. Vet. Diagn. Invest.* **4**, 134–138.

WESTAWAY, E. G., BRINTON, M. A., GAIDAMOVICH, S. Y., HORZINEK, M. C., IDARASHI, A., KÄÄRIÄINEN, L., LVOV, D. K., PORTERFIELD, J. S., RUSSEL, P. K., and TRENT, D. W. (1985). Togaviridae. *Intervirology* **24**, 125–139.

ZHANG, H., SCHOLL, R., BROWSE, J., and SOMERVILLE, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Res.* **16**, 1220.