

Comparison of the structural protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus

M. P. Murtaugh, M. R. Elam, and L. T. Kakach

Department of Veterinary Pathobiology, University of Minnesota, St. Paul, Minnesota, U.S.A.

Accepted March 15, 1995

Summary. The 3'-portion of the genome of a U.S. isolate of the porcine reproductive and respiratory syndrome (PRRS) virus, ATCC VR-2332, was cloned and sequenced. The resultant 3358 nucleotides contain 6 open reading frames (ORFs) with homologies to ORFs 2 through 7 of the European strain of the PRRS virus and other members of the free-standing genus of arteriviruses. Both VR-2332 and the European isolate (called the Lelystad virus) have been identified as infectious agents responsible for the swine disease called PRRS. Comparative sequence analysis indicates that there are degrees of amino acid identity to the Lelystad virus open reading frames ranging from 55% in ORF 5 to 79% in ORF 6. Hydropathy profiles indicate that the ORFs of VR-2332 and Lelystad virus correspond structurally despite significant sequence differences. These results are consistent with the biological similarities but distinct serological properties of North American and European isolates of the virus.

Introduction

A new viral disease of pigs, characterized by reproductive failure in sows and respiratory difficulties in piglets, was detected in North America in 1987 [9] and in Europe in 1990 [14]. The disease, known variously as porcine reproductive and respiratory syndrome (PRRS), swine infertility and respiratory syndrome (SIRS), porcine epidemic abortion and respiratory syndrome (PEARS), and mystery swine disease, among others, is characterized by late-term abortions and stillbirths in sows, and respiratory insufficiencies in nursery pigs [3, 16, 19]. The causative agent is a small, enveloped, positive-stranded RNA virus that is recovered primarily from alveolar macrophages of infected swine [1, 20]. It appears to be closely related to arteriviruses in morphology, genome organization, transcription strategy, and macrophage tropism [15].

The complete nucleotide sequence of an European isolate of the virus has been determined in 1993 by Meulenbergh et al. [13] and a partial sequence was determined by Conzelmann et al. [4]. The positive-strand genome encodes eight open reading

frames (ORFs) whose arrangement of genes is similar to that of coronaviruses [17] and arteriviruses [10, 15]. The two most 5' open reading frames (ORFs 1a and 1b) likely encode the viral RNA polymerase. ORFs 2, 5, and 6 appear to encode structural proteins associated with viral membranes, and ORF 7 is believed to encode a nucleocapsid protein. These proteins are expressed from a nested set of RNA transcripts with overlapping 3' ends. While this expression strategy is shared with the *Coronaviridae*, the physical properties of the arteriviruses originally placed them in the family *Togaviridae*. Plagemann and Moenning [15] have proposed a new group, the arteriviruses, to encompass viruses with these dual properties. This free standing group includes the PRRS virus, equine arteritis virus (EAV, [5]), lactate dehydrogenase-elevating virus (LDV, [15]), and simian hemorrhagic fever virus (SHFV, [8]).

The American and European isolates, VR-2332 and Lelystad virus (LV), respectively, display a high degree of antigenic variation considering that they cause the same disease. In a comparison of 24 field sera and seven viral isolates from Europe and North America, Wensvoort et al. [21] were unable to distinguish a single common antigen which was able to reliably diagnose infection. In order to determine the molecular differences between American and European isolates, ORFs 2 through 7 of a well characterized American isolate, VR-2332, were cloned and sequenced. The results indicate that the two strains are related but show a high level of amino acid sequence variation, ranging from 55% identity in ORF 5 to 79% identity in ORF 6.

Materials and methods

Viruses and cells

The virus isolate was a fourth cell culture passage of ATCC VR-2332 grown on 2621 cells [3]. Cells were cultured in Eagle's MEM supplemented with 4% fetal calf serum in a 5% humidified CO₂ atmosphere at 37 °C.

Construction of cDNA library

Total RNA from infected cell lysates was isolated by acid guanidine phenol extraction as described [2]. Poly A-containing RNA was selected by oligo dT column chromatography (Gibco BRL, Gaithersburg, MD). A cDNA library was constructed in the lambda unidirectional phage vector, UniZap XR, using Gigapack II Gold packaging extract and *E. coli* SURE cells as directed by the manufacturer (Stratagene, La Jolla, CA).

Screening of cDNA library

Positive plaques were selected by plaque lifts onto NitroPlus nitrocellulose membranes (Micron Separations Inc., Westboro, MA) and subsequent hybridization to ³²P-labeled PCR-derived fragments corresponding to full length Lelystad virus ORFs 2, 6, and 7. Hybridization-positive clones were further analyzed by EcoRI and XhoI restriction endonuclease digestion and the sizes of the VR-2332-specific inserts were estimated by electrophoresis in agarose gels. Next, the nucleotide sequence of 23 clones was determined at the 3' end. Twenty of 23 clones had identical 3' sequences, suggesting these clones were coterminaly nested. Six of these 20 clones of various sizes containing the same 3' end were selected for further DNA sequencing.

Sequence analyses

Sequence data were obtained by manual dideoxynucleotide sequencing with Sequenase (US Biochemicals, Cleveland, OH) and automated fluorescence sequencing (Applied Biosystems, Foster City, CA). Primers were derived from newly obtained sequence approximately every 250–300 bases. Analysis was performed using GCG (University of Wisconsin, Madison, WI) and Intelligenetics, Inc. (Mountain View, CA) software.

Sequence accession number

The 3' 3442 bases of VR-2332 sequence are deposited in Genbank with the accession number U00153. The accession includes 84 bases 5' to the start of ORF 2.

Results*Cloning and sequencing of VR-2332 genome*

In preliminary studies, LV ORF 7 DNA generated by PCR was used to probe Northern blots of RNA from 2621 cells infected with VR-2332. Radiographic bands were obtained with infected cells, but not with uninfected cells, indicating that LV and VR-2332 shared similar sequences (data not shown). Primers specific for LV ORF 7 then were used to amplify cDNA from VR-2332 as a rapid means of obtaining VR-2332 genomic information. However, no amplified products were obtained under a variety of conditions. Since it appeared that the two viruses might have considerable sequence differences, a cloning strategy using LV sequences for screening a cDNA library was devised to obtain the nucleotide sequence corresponding to the putative structural protein genes of the VR-2332 strain of the PRRS virus.

To isolate the genes for the structural proteins of the VR-2332 strain of the PRRS virus, total cellular RNA was obtained from infected 2621 cells and used to construct an oligo dT-primed cDNA library. VR-2332-specific cDNA clones were identified in plaque hybridizations using Lelystad virus ORF 2, 6, or 7 DNA fragments as probes. The hybridization-positive clones were analyzed by restriction endonuclease digestion and the sizes of the VR-2332-specific inserts were estimated in agarose gels with molecular weight standards.

The nucleotide sequence of the 3' end of the 23 independent clones of various sizes was determined. Twenty of 23 clones contained identical 3' sequences, suggesting this virus, like Lelystad virus, possessed a coterminal nested set of RNA transcripts. Figure 1 shows the six clones chosen for further sequence analysis. Clones 431, 412, and 416 were sequenced from their 5' ends to overlap with the sequence generated from the next smaller clone. The gap between the 5' end of clone 416 and the beginning of ORF 2 (contained in both clones 712 and 761) was sequenced from both ends by using VR-2332-specific primers. Additional sequencing was performed to confirm the sequence on the opposite strand (Fig. 1). With this strategy, a sequence of 3358 nucleotides was obtained on both strands from two to six independent clones. In addition, the region contained in clone 431 was confirmed by sequencing a PCR amplified cDNA fragment obtained from infected cell RNA.

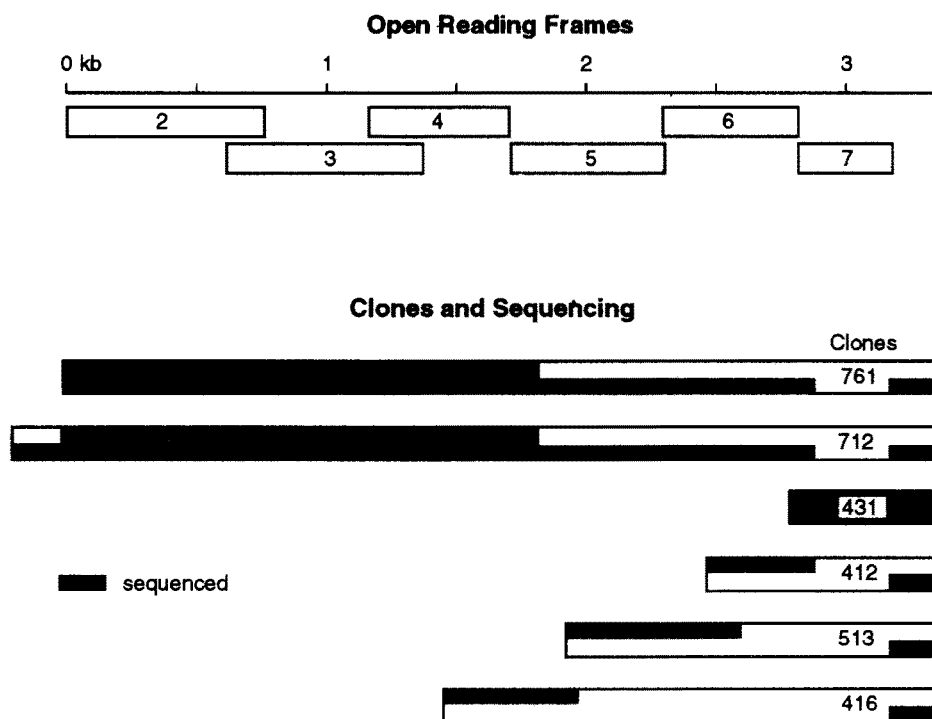


Fig. 1. Organization of VR-2332 ORFs 2 through 7. The cDNA clones used to determine the nucleotide sequence of the VR-2332 strain of the PRRS virus are shown. The regions sequenced are in black

Comparison of the ORFs of VR-2332, Lelystad virus, LDV, and EAV

The deduced amino acid sequence of the 3' end of VR-2332 revealed a genome organization characteristic of the arteriviruses, with overlapping coding regions in different reading frames. The six open reading frames (ORFs) corresponded to ORFs 2 through 7 of Lelystad virus, LDV, and EAV. The VR-2332 ORFs were most like those of Lelystad virus (Table 1). Comparison of VR-2332 to LDV and EAV also indicated that it shares an evolutionary history with these viruses. ORF 6

Table 1. Amino acid similarities and identities of VR-2332 with Lelystad virus, LDV, and EAV

ORF	Lelystad virus		LDV		EAV	
	% similarity	% identity	% similarity	% identity	% similarity	% identity
2	76	63	59	39	45	20
3	72	58	62	34	49	22
4	80	68	59	36	43	20
5	80	59	66	49	42	20
6	91	78	70	52	44	26
7	74	65	63	51	51	28

Homologies were determined using the Needleman-Wunsch algorithm to align sequences in the GAP program of GCG. The LDV sequence is from Genbank accession number L06811

Table 2. Comparison of VR-2332 and Lelystad virus ORFs 2 through 7

ORF	Amino acids		Predicted kD		pI		Glycosylation sites	
	2332	LV	2332	LV	2332	LV	2332	LV
2	256	249	29.4	28.4	11.0	10.2	2	2
3	254	265	29.0	30.6	8.1	9.4	7 ^a	7
4	178	183	19.5	20.0	7.9	6.1	4	4
5	200	201	22.4	22.4	8.3	8.2	3	2
6	174	173	19.0	18.9	11.3	11.9	1	2
7	123	128	13.5	13.8	10.4	11.2	1 ^a	1

^aNot all sites are identical

is the most highly conserved amino acid sequence between VR-2332 and LV, and between VR 2332 and LDV, with respect to similarity and identity. ORF 7 is the most similar ORF between VR-2332 and EAV.

While VR-2332 is more related to the Lelystad virus than to other arteriviruses, the homologies were lower than expected for two viruses that cause the same disease. Substitutions, deletions, and additions occurred throughout the sequences. The predicted proteins were different in molecular weights, isoelectric points, and predicted N-linked glycosylation sites (Table 2). When other characteristics of the predicted proteins, such as the hydropathy profiles and percent basic character were compared the two viruses appeared more similar than was indicated by amino acid sequence. In addition, dot matrix analysis utilizing a sliding window of 21 amino acids with a requirement of 13 identical residues at each location showed clearly that all of the ORFs were nearly colinear between VR-2332 and LV (data not shown).

Figure 2 shows the hydropathy profiles for ORFs 2 through 7 of VR-2332 and LV. Although the overall patterns are conserved, as predicted by the high degree of amino acid similarity, differences are apparent, particularly in ORF 3.

ORF 3, which displays only a 58% sequence identity between VR-2332 and Lelystad virus, is the least conserved ORF. The predicted protein contains a carboxy-terminal deletion of twelve amino acids in VR-2332. As a result of the differences, the protein in Lelystad virus shows a strongly hydrophilic region centred on residue 240, whereas in VR-2332 it appears amphipathic in this region. The predicted molecular weight of the translation product of ORF 3 is approximately 30 kD, but it contains seven potential N-linked glycosylation sites in each virus, so that the apparent size may be greater. Three out of the seven sites are conserved between the two viruses. The other four sites are in close proximity.

ORF 4 is the only ORF which encodes a putative hydrophobic amino terminal signal sequence. However the carboxy terminus of this protein is also exceptionally hydrophobic in both viruses. Five putative membrane spanning domains are more pronounced in VR-2332 than in Lelystad virus. All four N-linked glycosylation sites are conserved between the two viruses.

ORF 5 may encode an envelope protein in the arteriviruses because of its hydropathy profile and putative glycosylation sites. In EAV, the G_L or ORF 5 protein is glycosylated [6]. VR-2332 ORF 5 contains three potential glycosylation

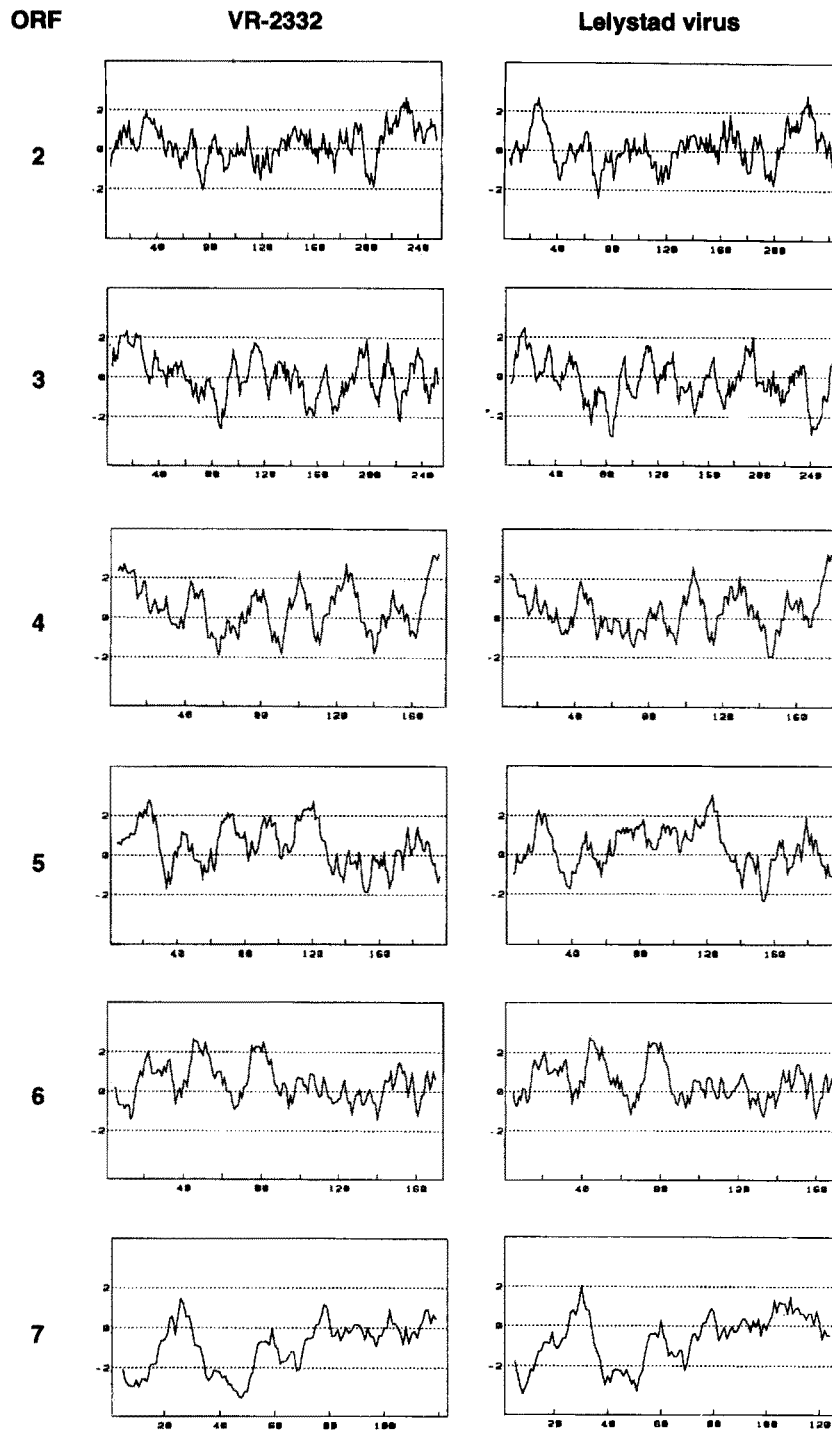


Fig. 2. Hydropathy profiles of ORFs 2 through 7 from VR-2332 and Lelystad virus. The amino acid residue numbers are on the horizontal axis. The vertical axis shows the average hydropathy for a window of seven amino acids. Hydrophobic regions appear above the middle line and hydrophilic regions appear below the line

sites, two of which are shared with LV. Their hydropathy profiles are similar even though only seven residues in the amino terminal 40 amino acids are the same. Potential membrane spanning domains between residues 65 and 130 are more pronounced in VR-2332.

ORF 6 showed three highly hydrophobic regions in the N-terminal half of the protein that are presumed to be membrane spanning domains. They are observed in VR-2332, LV, LDV, and EAV and thus appear to be a conserved characteristic of all arteriviruses. The VR-2332 protein shares one predicted N-linked glycosylation site with the Lelystad virus, however these sites may not be glycosylated in either virus since the M or ORF 6 protein in EAV is not glycosylated [6].

ORF 7 maps to the location of the nucleocapsid protein gene in LDV and EAV [6, 7]. The protein is 74% similar between VR-2332 and Lelystad virus, although the VR-2332 ORF 7 is smaller by five amino acids. Nevertheless, the N-terminal half of both proteins encoded by ORF 7 are 26 to 28% basic and the hydrophobicity plots were nearly identical. The basic residues presumably facilitate interactions with the RNA genome.

VR-2332 has a 3' untranslated sequence following ORF 7. This region consists of 151 nucleotides and a poly A tail of 19 to 20 bases in the cDNA clones sequenced. Lelystad virus has a noncoding region of 114 bases. Bases 50–151 of the non-coding region of VR-2332 share a strong homology to bases 13–114 of the non-coding region of Lelystad virus. The function of this sequence is unknown.

Partial nucleotide sequence has been reported from two North American PRRS virus isolates [11, 12]. The ORF 6 and 7 sequences are 98% identical between VR-2332 and both VR-2385 (Iowa) and IAF-exp91 (Quebec). The ORF 5 sequence is 89% identical between VR-2332 and VR-2385.

Discussion

The structural protein coding genes of VR-2332 were sequenced and compared to Lelystad virus. A high level of nucleotide sequence diversity between both PRRSV isolates was observed, which was reflected in comparative Northern blotting in which signal intensities were approximately 100-fold greater for homologous compared to heterologous hybridizations. Although the amino acid homologies are substantially less than expected for viruses that appear to cause an identical disease, the findings were consistent with the striking antigenic diversity reported from serological studies [21]. These studies showed that there is little serological cross-reactivity between European and North American isolates of the PRRS virus. We infer that the antigenic differences between VR-2332 and Lelystad virus are due to immunological responses of swine to the dissimilar regions of the viruses. In contrast to the differences noted between the European and U.S. isolates, preliminary observations indicate the North American isolates are highly conserved [11, 12].

The differences between the two viruses occurred throughout the 3' genomic sequences encoding ORFs 2 through 7 and the 3' untranslated region, and were due to nucleotide substitutions, base deletions, and base additions. The sequence divergence may have arisen from error-prone replication and suggests that the viral

replicase might lack proofreading activity. Whether the extensive differences were exacerbated by selection of variant strains to escape immune surveillance could not be determined in this study.

Structural and sequence analyses showed that, with few exceptions, sequence differences were widely distributed. The principal differences were in the 5' ends of the ORFs, many of which encode hydrophobic domains, and in ORF 4 in the region of amino acid residues 50–70. The hydropathy profiles comparing each ORF indicated that protein structures were conserved despite the extensive sequence differences. The algorithm clearly showed similar regions of uncharged and charged amino acids, in each ORF of VR-2332 and LV and may predict regions which span or do not span the membrane. We conclude that VR-2332 proteins are similar in structure and function to those of LV, but that the extensive amino acid differences in the proteins account for the extensive differences in serological cross-reactivity.

The similarity of both VR-2332 and LV isolates of the PRRS virus to LDV and EAV in nucleotide sequence, genome structure and organization, replication strategies, pathogenic mechanisms and host cell preferences suggests that these viruses are closely related and may have arisen from a common ancestor. Figure 3, showing a nucleotide sequence phylogeny comparison of VR-2332, Lelystad virus, LDV, and EAV ORF 5, indicates that the PRRS viruses are related and different from LDV and EAV. It also indicates that the PRRS viruses are more related to LDV than to EAV. Analyses of other ORFs including ORFs 6 and 7, which are more conserved than ORF 5, gave the identical result. Also, the amino acid similarity of LDV to the PRRS viruses is 59 to 70% (Table 1), whereas EAV is only 29–49% similar to LDV (data not shown). Hence, the close similarities of PRRS virus isolates and LDV raise the intriguing possibility that the PRRS virus is an LDV variant which adapted to a new host species and evolved rapidly away from LDV in its new environment. VR-2332 and LV may have diverged shortly thereafter, or the two PRRS virus isolates may have evolved independently from different LDV variants.

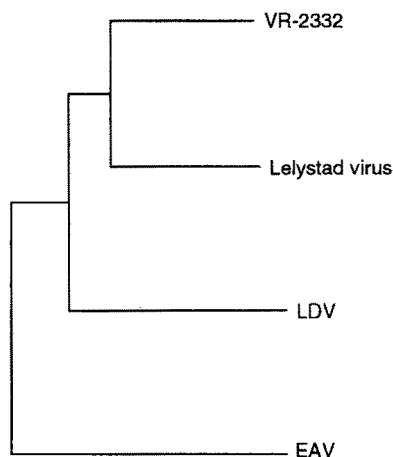


Fig. 3. Phylogram of ORF 5 of the *Arteriviridae*. The unrooted tree was derived from an analysis of nucleotide sequence data for VR-2332, Lelystad virus, LDV and EAV using Lake's method of phylogenetic invariants in the PAUP program. The LDV sequence used in the analysis was Genbank accession number L06811

Acknowledgements

We thank the University of Minnesota Agricultural Experiment Station and Boehringer Ingelheim Animal Health, Inc. for support of the research.

References

1. Benfield DA, Nelson E, Collins JE, Harris L, Goyal SM, Robison D, Christianson WT, Morrison RB, Gorcyca DE, Chladek DW (1992) Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). *J Vet Diagn Invest* 4: 127–133
2. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156–159
3. Collins JE, Benfield DA, Christianson WT, Harris L, Hennings JC, Shaw DP, Goyal SM, McCullough S, Morrison RB, Joo HS, Gorcyca DE, Chladek DW (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
4. Conzelmann K, Visser N, van Woensel P, Thiel H (1993) Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group. *Virology* 193: 329–339
5. den Boon JA, Snijder EJ, Chirnside ED, de Vries AAF, Horzinek MC, Spaan WJM (1991) Equine arteritis virus is not a togavirus but belongs to the coronavirus superfamily. *J Virol* 65: 2910–2920
6. de Vries AAF, Chirnside ED, Horzinek MC, Rottier PJM (1992) Structural proteins of equine arteritis virus. *J Virol* 66: 6294–6303
7. Godeny EK, Speicher DW, Brinton MA (1990) Map location of lactate dehydrogenase-elevating virus (LDV) capsid protein (Vp1) gene. *Virology* 177: 768–771
8. Godeny EK, Zeng L, Smith SL, Brinton MA (1993) In: *Proceedings of the 9th International Congress of Virology, Glasgow, Scotland 1993*, p 22
9. Hill H (1990) Overview and history of mystery swine disease (swine infertility and respiratory syndrome). In: *Proceedings of the Mystery Swine Disease Committee Meeting, Denver CO 1990*. Livestock Conservation Institute, Madison, WI, pp 29–30
10. Kuo HL, Harty JT, Erickson L, Palmer GA, Plagemann PGW (1991) A nested set of eight mRNAs is formed in macrophages infected with lactate dehydrogenase-elevating virus. *J Virol* 65: 5118–5123
11. Mardassi H, Mounir S, Dea S (1994) Identification of major differences in the nucleocapsid protein genes of a Quebec strain and European strains of porcine reproductive and respiratory syndrome virus. *J Gen Virol* 75: 681–685
12. Meng X-J, Paul PS, Halbur PG (1994) Molecular cloning and nucleotide sequences of the 3'-terminal genomic RNA of the porcine reproductive and respiratory syndrome virus. *J Gen Virol* 75: 1795–1801
13. Meulenbergh JJM, Hulst MM, de Meijer EJ, Moonen PLJM, den Besten A, De Kluyver EP, Wensvoort G, Moormann RJM (1993) Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* 192: 62–72
14. Paton DJ, Brown IH, Edwards S, Wensvoort G (1991) Blue ear disease of pigs. *Vet Rec* 128: 617
15. Plagemann PGW, Moenning V (1992) Lactate dehydrogenase-elevating virus, equine arteritis virus and simian hemorrhagic fever virus: a new group of positive-strand RNA viruses. *Adv Virus Res* 41: 99–192
16. Pol JMA, van Dijk JE, Wensvoort G, Terpstra C (1991) Pathological, ultrastructural, and immunohistochemical changes caused by Lelystad virus in experimentally induced infections of mystery swine disease (synonym: porcine epidemic abortion and respiratory syndrome (PEARS)). *Vet Q* 13: 137–143

17. Spaan WJM, Cavanagh D, Horzinek MC (1988) Coronaviruses: structure and genome expression. *J Gen Virol* 69: 2939–2952
18. Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
19. Wensvoort G, Terpstra C, Pol JMA, ter Laak EA, Bloemraad M, de Kluyver EP, Kragten C, van Buiten L, den Besten A, Wagenaar F, Broekhuijsen JM, Moonen PLJM, Zetstra T, de Boer EA, Tibben HJ, de Jong MF, van't Veld P, Groenland GJR, van Gennep JA, Voets MT, Verheijden JHM, Braamskamp J (1991) Mystery swine disease in the Netherlands: the isolation of Lelystad virus. *Vet Q* 13: 121–130
20. Wensvoort G, de Kluyver EP, Pol JMA, Wagenaar F, Moormann RJM, Hulst MM, Bloemraad R, den Besten A, Zetstra T, Terpstra C (1992) Lelystad virus, the cause of porcine epidemic abortion and respiratory syndrome: a review of mystery swine disease research at Lelystad. *Vet Microbiol* 33: 185–193.
21. Wensvoort G, de Kluyver EP, Luitze EA, den Besten A, Harris L, Collins JE, Christianson WT, Chladek D (1992) Antigenic comparison of Lelystad virus and swine infertility and respiratory syndrome (SIRS) virus. *J Vet Diagn Invest* 4: 134–138

Authors' address: Dr. M. P. Murtaugh, Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Avenue, St. Paul, MN 55108, U.S.A.

Received November 28, 1994