

**Determination of the complete nucleotide sequence of a vaccine strain of porcine reproductive and respiratory syndrome virus and identification of the Nsp2 gene with a unique insertion**

S. Shen, J. Kwang, W. Liu, and D. X. Liu

Institute of Molecular Agrobiology,  
National University of Singapore, Singapore

Accepted January 11, 2000

**Summary.** The complete nucleotide sequence of the genomic RNA of a vaccine strain (SP) of porcine reproductive and respiratory syndrome virus (PRRSV) was determined. It shares approximately 94% identity of nucleotide sequence with two recently reported North American strains, 16244B and VR2332, but only 78% with a European strain, Lelystad virus (LV). Its genome is the longest among the four published complete sequences of PRRSV, due to an insertion in the Nsp2-encoding region. Compared to Nsp2 of the North American strains and the European strain, the predicted Nsp2 of strain SP contains 36 and 155 amino acid insertions, respectively, near the C-terminus, in addition to several highly variable regions. The insertion shows no homology with any equivalent arterivirus proteins. This high sequence disparity of Nsp2 among different PRRSV isolates suggested that it could be used as a marker to differentiate PRRSV genotypes. The 5' RACE and primer extension analysis of three North American strains demonstrated that the utmost 5'-end nucleotides are conserved among PRRSV strains isolated from two continents. The predicted polyprotein 1a/b contains conserved proteinase, polymerase and helicase domains responsible for polyprotein processing, RNA transcription and replication.

### **Introduction**

Porcine reproductive and respiratory syndrome virus (PRRSV) is an important pathogen that causes reproductive failure in breeding swine and respiratory problems in piglets, one of the most economically significant diseases in swine herds worldwide [15, 16]. PRRSV is a member of the family *Arteriviridae* in the order Nidovirales [4], together with equine arteritis virus (EAV), lactate dehydrogenase elevating virus (LDV) and simian hemorrhagic fever virus (SHFV). It is a small,

enveloped virus with a positive, single-stranded RNA genome [23]. The genomic RNA is 5'-capped and 3'-polyadenylated and contains two large open reading frames (ORF 1a and 1b) coding for two large polyproteins (1a and 1a/1b). In the case of EAV, these polyproteins are processed to mature, nonstructural proteins (Nsp1 to Nsp12) by three 1a-encoded proteinases [23]. The 1a and 1a/b polyproteins of PRRSV contain similar cleavage sites but only the cleavage of Nsp1 $\alpha$  and Nsp1 $\beta$  by two papain-like proteinases has been demonstrated [6]. The viral structural proteins are encoded by six subgenomic RNAs [23].

Sequence comparison between PRRSV isolates reveals that there are two major genotypes represented by North American and European prototypes LV [11] and VR2332 [13], respectively, and perhaps more minor genotypes among North American strains. This conclusion was drawn largely based on sequence comparison of the structural regions [2, 13, 26], as data on the nonstructural region of North American strains were not available until recently [1, 13, 14]. The sequence diversity in the structural regions is reflected in antigenic differences between North American and European isolates as well as within the major genotypes [5, 7, 12]. The genetic diversity in terms of gene length and sequence variation may also exist in the 1a region within PRRSV isolates as suggested by insertions and deletions in the 1a genes of EAV, LVD and SHFV [23]. Vaccines against PRRSV infection in pigs have been available through great efforts since the isolation of the virus. However, mechanisms of attenuation are still poorly understood, partly due to the relatively slow accumulation of sequence data in the 5'-NCR and the nonstructural protein-encoding region of the genome. Sequence analysis of these regions of more isolates may be of help to a better understanding of the PRRSV pathogenicity and to the assessment of reverse mutations of live vaccine strains [3]. The aim of this study was therefore to investigate the genetic variation in the 5'-NCR, 1a and 1b regions of the North American strains and to identify the most variable region of the viral genome.

In this study, we report the complete nucleotide sequence of a North American vaccine strain (SP) and the comparison of the SP sequence at both nucleotide and amino acid levels with other PRRSV isolates. Meanwhile, the utmost 5'-end sequence of strain SP was carefully analyzed. The results revealed a unique insertion in the 1a region and significant sequence diversity within the North American strains, especially the Nsp2 protein encoded by ORF 1a, suggesting that the molecular evolution of arteriviruses may involve mutations, insertions and deletions of the viral genome.

## Materials and methods

### *Viruses and cells*

A vaccine strain of PRRSV (PRIME PAC PRRS) was used for determination of the complete nucleotide sequence. It was purchased from Schering-Plough Animal Health and referred to as strain SP in this study. The reference strains VR2332 and 12068-96p were obtained from Dr. Fernando A. Osorio, Department of Veterinary and Biomedical Sciences, University of Nebraska-Lincoln, USA. The viruses were passaged in MARC-145 cells, a subclone of the monkey kidney cell line MA-104. MARC-145 cells were maintained in complete DMEM

medium (GIBCO BRL) supplemented with 10% newborn calf serum and infected with viruses at low multiplicity of infection (0.1) as described previously [9]. Viral stock of the passage 5 was used for sequence determination.

#### *Purification of virus*

Viruses were harvested by freezing and thawing virus-infected cells three times. Cell debris was removed by centrifugation at 6,000 rpm for 15 min (Beckmen JA-25.50). The supernatant was centrifuged through a 20% sucrose (in TNE buffer) cushion at 150,000 g (Beckmen SW28) for 2 h at 4 °C. Pellet was suspended in TNC buffer (100 mM NaCl, 2 mM CaCl<sub>2</sub> 20 mM Tris-Cl, pH 7.4) and stored at -80 °C.

#### *Amplification of viral RNA genome*

Viral RNA was extracted from partially purified viruses using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Reverse transcription and polymerase chain reaction (RT-PCR) were performed using the Expand Reverse Transcription and High Fidelity PCR Kits (Boehringer). Annealing and extension times of PCR were optimized for amplification of PCR products with different sizes using different primers. More than 100 specific primers were used for amplification, sequencing and cloning. Plasmid pKT0 [10] and pACYC177 (BioLab) were used for cloning of amplified PCR products.

#### *5' Rapid amplification of cDNA ends (5' RACE)*

Sequence at the 5'-end of the viral genomic RNA was amplified using a 5'/3' RACE Kit (Boehringer) according to the manufacturer's instructions. Briefly, first strand cDNA was synthesized using a viral specific primer (primer 1, CTTTCTCAAGCCTGGCC) and AMV reverse transcriptase. The incubation temperature was raised to 55 °C to help proceed through regions with difficult secondary RNA structures and to the 5'-end. The cDNA was purified and a poly (A) tail was added using terminal transferase. The tailed cDNA was then amplified by PCR using a viral specific primer (primer 2, CTAAATGGACCTATCGTCG) and the oligo dT-anchor primer, which is a mixture of primers with a non T (A, C, or G) nucleotide at the 3'-end. The oligo dT-anchor primer was forced to bind to the inner end of the poly(A) tail. The resulting cDNA was further amplified by a second round PCR using a nested, viral specific primer (primer 3, GTCACAGAAGGGTGTCTGTGCAGCAAG) and the PCR anchor primer. The PCR product was then used for sequencing and cloning.

#### *Primer extension experiment*

Primer extension was performed by standard procedures [18]. Briefly, the specific primer 3 was end-labeled with [ $\gamma$ -<sup>33</sup>P]-ATP using T4 polynucleotide kinase and was purified using NucleoTrap Probe Kit (Stratagene). The viral RNA (0.2  $\mu$ g) and the labeled primer (10,000 cpm) were incubated at 65 °C for 90 min, allowed to cool slowly to room temperature and applied for RT reaction with the Expand Reverse Transcriptase. After 1 h incubation at 55 °C, the sample was treated with DNase-free RNase (1 U/50  $\mu$ l) at 37 °C for 30 min, extracted with phenol/chloroform and precipitated with ethanol. The primer extension product was analyzed on sequencing gels together with sequencing samples of the 5' RACE products described above to determine the 5'-end sequence.

#### *Sequence analysis*

Automated sequencing of both strands was carried out using PCR products with specific primers as previously described [18]. Sequencing Kit (AmpliCycle, Perkin Elmer) was also

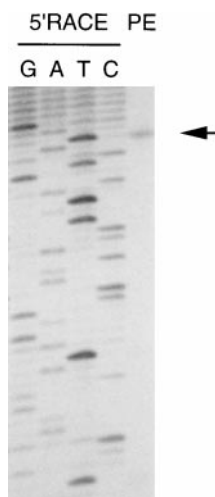
used for determination of the 5'-end sequence of the 5' RACE product. Sequence data from gel reading were assembled using the CLUSTAL W sequence analysis program [25]. Further analyses were carried out using the GCG suite of programs.

## Results

### *Determination of the utmost 5'-end sequence of the genomic RNA of strain SP*

The utmost 5'-end sequence of strain SP was determined by sequencing the 5' RACE products and by analyzing the primer extension product. As described, the oligo dT-anchor primer is a mixture of primers with a non-T (A, C, or G) nucleotide at the 3'-end so that the primer can be forced to bind to the inner end of the poly(A)-tailed cDNA. However, if there is one U at the 5'-end of RNA, the 3'-end of the first strand cDNA will end with an A. In this circumstance, it would be difficult to determine the first nucleotide of the genome just by sequencing the 5' RACE products. Primer extension experiments were carried out to confirm the utmost 5'-end nucleotide of the viral genome.

As shown in Fig. 1, sequencing samples of the 5' RACE products were run on a denaturing gel together with the primer extension product. The primer extension product of SP strain comigrated with a band terminated with a T. As the same primer (primer 3) was used for sequencing of the complementary strand and the primer extension experiments, we believe that it represents the 5'-end of the genome. The utmost 5'-end nucleotide of the viral genome is therefore an A. This conclusion was confirmed by sequencing of independently prepared 5' RACE products three times. In addition, the 5' RACE products of strains VR2332 and 12068-96p were analyzed in parallel. We found that the first 55 nucleotides (upstream of the sequencing primer) of the 5'-end of these three strains are identical. It is interesting to note that 10 out of the first 12 nucleotides at the 5'-end are conserved between the European strain LV and the North American strains (ATGATGTGTGTAGG vs ATGACGTATAGG) analyzed.



**Fig. 1.** Sequencing of the 5' RACE products and analysis of the primer extension product. The 5' RACE products of strain SP were sequenced with specific primer 3. The primer extension reaction was carried out using the same primer and RNA templates extracted from the purified SP virions. The sequencing samples and the primer extension product (PE) of SP were run as indicated on an 8M urea-6% polyacrylamide gel. The gel was dried and autoradiographed

*Amplification and cloning of the whole genome of strain SP*

At the onset of this study, no sequence data on the majority of the nonstructural region of North American strains were available. Therefore, the design of RT and PCR primers for this region was based on the sequence data of the European strain LV and other closely related arteriviruses. We obtained PCR fragments covering the 3' end 11 kilobase (kb) of the genome using one North American- and one European-strain specific primer for each PCR product. The 5' end 4.5 kb of the genome was amplified using more specific primers after the sequence data of 16244B was released [1]. These PCR fragments were used in the first round sequencing. Seven overlapping PCR fragments covering the whole genome of strain SP were then obtained using SP-specific primers. These PCR fragments were flanked with unique restriction enzyme sites and were cloned into vector pKT0 [10] and pACTY177. The cDNA clones were used for sequencing and construction of full-length cDNA clone.

*Determination of the complete nucleotide sequence of strain SP*

The complete nucleotide sequence of strain SP was determined first by direct sequencing of overlapped PCR fragments and then by sequencing different cDNA clones at least two times. About 90% of the genome were sequenced from both strands. The overlapped gel readings were assembled into consensus, full-length genomic RNA.

The complete nucleotide sequence of strain SP is 15520 nucleotides long excluding the poly(A) tail (Table 1). The 5'-NCR is 190 nucleotides in length while the 3'-NCR is 151 nucleotides long. Nine ORFs were identified, two (ORFs 1a and 1b) for nonstructural proteins, and seven (ORFs 2a, 2b to 7) for structural proteins, including the homologue of ORF 2a recently identified in the EAV virion [24]. The accession number of the complete nucleotide sequence in GenBank is AF184212.

**Table 1.** Comparison of the genome length (nucleotides) of strain SP with those of the North American strain 16244B and the European strain LV

	SP	16244B	LV
Total	15520	15411	15088
5'-NCR	190	189	221
ORF1a	7617	7509	7188
ORF1b	4389	4389	4389
ORF2-7	3187	3187	3185
3'-NCR	151	151	114

VR2332	<b>T</b> ATGACGTATAGGTGT TGGCTCTATGCCTTG GCATTTGTATTGTCA GGAGCTGTGACCATT
SP	<u>  </u> ATGACGTATAGGTGT TGGCTCTATGCCTTG ACATTTGTATTGTCA GGAGCTGTGAT <u>  </u> CATT
VR2332	GGCACAGCCCAAA <b>A</b> C TTGCTGCACAGAAAC ACCCTTCTGTGAT <b>A</b> G CCTCCTTCAGGGGAG
SP	<u>  </u> GCACAGCCCAAA <b>G</b> C TTGCTGCACAGAAAC ACCCTTCTGTGAC <b>A</b> G CCTCCTTCAGGGGAG
VR2332	<b>C</b> TTAGGG <b>T</b> T-TGTCC CTAGCACCTTGCTTC CGGAGTTGCACTGCT TTACGGTCTCTCCAC
SP	<u>  </u> <b>T</b> TTAGGG <b>G</b> <u>  </u> <b>T</b> <u>  </u> CTGTCC CTAGCACCTTGCTTC CGGAGTTGCACTGCT TTACGGTCTCTCCAC
VR2332	CCCTTTAACC      190
SP	CCCTTTAACC      190

**Fig. 2.** Sequence alignment of the 5' NCR of strains SP and VR2332. The variable nucleotides were shown in bold and their positions were underlined

#### *Sequence comparison with other PRRSV isolates*

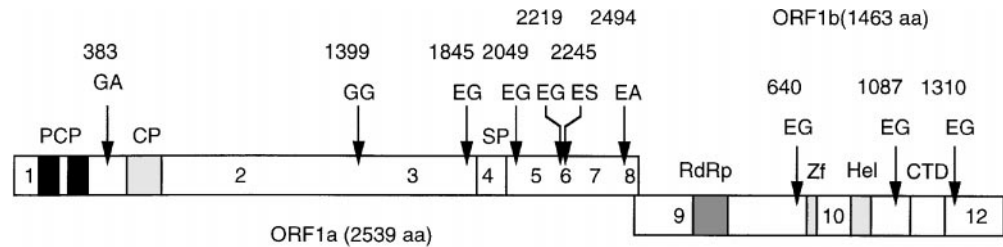
The 5'-end NCR of SP was compared with that of VR2332 recently published by Olekiewicz et al. [14]. The utmost 5'-end nucleotide of strain SP was defined as an A in this report (Figs. 1 and 2); but an additional T for strain VR2332 was reported by Olekiewicz et al. [14]. However, as mentioned above, the first nucleotide of VR2332 (as well as strain 12068-96p) is also an A based on our results. In addition, a C at position 130 of SP is missing for VR2332 (Fig. 2). The 5' NCR for both strains is 190 nucleotides in length. The percentage identity is 95.8 between them. Five stem-loops were predicted to be formed in the 5' NCR of strain SP. It was noted that three nucleotide changes in positions 121, 128 and 129

**Table 2.** Comparison of the nucleotide and amino acid sequences of strain SP with those of strains 16244B and LV

	SP vs LV		SP vs 16244B	
	Nt (%) <sup>a</sup>	Aa (%) <sup>b</sup>	Nt (%) <sup>a</sup>	Aa (%) <sup>b</sup>
Overall	63.4		94.0	
5'-NCR	40.0		92.1	
ORF1a	56.3	37.5–68.0	89.3	81.0–100
ORF1b	64.6	40.0–75.0	95.0	96.0–99.0
ORF2–7	67.0	58.0–79.0	93.0	89.0–95.0
3'-NCR	76.0		98.0	

<sup>a</sup>Nt (%), percentage of identity of nucleotide sequences

<sup>b</sup>Aa (%), percentage of identity of amino acid sequences



**Fig. 3.** Diagram showing the predicted cleavage products of polyproteins encoded by ORFs 1a and 1b of PRRSV strain SP. Arrows and numbers indicate predicted cleavage sites of the polyproteins 1a and 1b and boxes indicate predicted functional domains conserved among arteriviruses. Abbreviations: *PCP* papain-like cysteine proteinase; *CP* cysteine proteinase; *SP* serine proteinase; *RdRp* RNA dependent RNA polymerase; *Hel* helicase/NTPase; *Zf* zinc-finger; *CTD* conserved C-terminal domain specific for nidoviruses

were located in the stem-loop 4 (data not shown), suggesting that these changes could destabilize this putative stem-loop structure.

The sequence of strain SP was compared with those of the European strain LV and the North American strain 16244B. Table 1 shows that SP has the longest genome among the different strains sequenced to date, largely because of the longer ORF 1a. It is 109 and 432 nucleotides longer than 16244B and LV, respectively. As shown in Table 2, SP shares a high degree of homology with strain 16244B at both nucleotide and amino acid sequence levels, while considerable divergence exists between SP and strain LV in both structural and nonstructural regions. We noted that the 1a region is the most variable among the nine ORFs, even between the North American isolates.

For PRRSV, only two nonstructural proteins (Nsp1 $\alpha$  and Nsp1 $\beta$ ), containing two papain-like cysteine proteinase domains [6], have been shown to be auto-cleaved from the N-terminus of the 1a and 1b polyproteins. The other cleavage sites of PRRSV strains were predicted (Fig. 3) mainly based on the proteolytic processing data of EAV and sequence comparison [20–22, 27, 28]. In addition, an extra E/G site was found in Nsp6 of strain SP, but not in the equivalent region of VR2332 or 16244B. Furthermore, Nsp6 of PRRSV is 4 amino acids shorter than that of EAV; it is only 16 amino acids in length.

The amino acid sequences of each putative cleavage product of the 1a polyprotein (Nsp1 to Nsp8) and the 1b polyprotein (Nsp9 to Nsp12) were compared with those of the equivalent, putative proteins of strains LV and 16244B. The results revealed that the Nsp2 protein is 36 and 155 amino acids longer than those of 16244B and LV, respectively, and shares lowest homology among the eight 1a-encoded proteins (Table 3). No homology was found between the 36 amino acid insertion and any other proteins in a blast search of the existing genetic databases. Compared to the equivalent Nsp2 of strains 16244b and VR2332, the 36 amino acid insertion is located between G<sup>1a1195</sup> and T<sup>1a1232</sup> of the SP Nsp2 (Fig. 4). It is flanked by one of three highly variable amino acid stretches near the C-terminus of the protein (Fig. 4). The N-terminal region, containing the putative catalytic

**Table 3.** Comparison of the amino acid sequences of NSP1-NSP12 of strain SP with those of strains 16244B and LV (percentage identity)

	SP vs 16244B	SP vs LV
Nsp1	93.2	52.0
Nsp2	81.0	<37.5 <sup>a</sup>
Nsp3	95.0	57.3
Nsp4	93.6	62.2
Nsp5	93.0	70.0
Nsp6	100	81.0
Nsp7	97.7	48.5
Nsp8	97.0	68.0
Nsp9	98.0	75.0
Nsp10	99.0	64.0
Nsp11	96.0	75.0
Nsp12	98.0	40.0

<sup>a</sup>The highest homology between the most conserved region of SP NSP2 and the equivalent region of LV

residues Cys<sup>1a437</sup> and His<sup>1a506</sup>, and C-terminal region were conserved. The putative cleavage products of the 1b polyprotein are relatively more conserved than those of the 1a polyprotein (except the small Nsp6) and structural proteins encoded by ORF2a, 2b to ORF7 (Tables 2 and 3).

Amino acid substitutes for the corresponding residues of the pathogenic strain 16244B in nonstructural proteins (except Nsp2) are listed in Table 4. It should be mentioned that 158 substitutes and 36 residues in the insertion in the Nsp2 (Fig. 4) exceed the total substitutes found in other nonstructural proteins (94 residues), with only 22 (11.3%) in 1b proteins. Some of these substitutes may be responsible for the attenuation of this vaccine strain of PRRSV.

## Discussion

In this communication, we present the complete nucleotide sequence of a vaccine strain, SP, attenuated from a North American PRRSV isolate. It was interesting to find that strain SP contains a long insertion (36 amino acids) in the putative Nsp2 protein, which turns out to be the most variable among PRRSV viral proteins, as compared with strains VR2332 and 16244B.

There are only two complete sequences of North American strains, VR2332 and 16244B, available [1, 12, 14]. These reports showed significant differences in the 5'-end NCR of the two North American strains. First, the first 12 nucleotides at the 5'-end of strain VR2332 are more similar to the European strain LV than to strain 16244B, despite the fact that strains VR2332 and LV represent two major genotypes of PRRSV emerged from two continents, while VR2332 and 16244B



SP	AGKRARKARSSATAT	VAGRALPVRRETRQVE	EHEVAGANKAEHLKH	YSPPAEGNCGWHCIS	<b>AIGNRMLNSKFETTL</b>	PERVRPPDDWATDED	473
VR2332	-----C-----	-----S-----	AK-----N-----	-----	---A---V---K-----	-----E-----	
16244B	-----C-----	-----S-----	AK-----D-----	-----	---A---V---I-----	-----D-----	
			*				
SP	LVNATQIILRLPAALD	RNGACASAKYVLKLE	GEHWTVVTVTPGMSPS	LLPLECVQGCCEHKG	GLGSPDAVEVFGFDP	ACLDWLAEVMHLPSN	563
VR2332	-V-----	-----T-----	-----	-----G---	-----V---S---	-----R-----	S
16244B	-A-----	-----T-----	-----	-----E---	-----I---S---	-----W-----	S
			*				
SP	ATPAALAEMSGDSNR	PASPVTTVTVSQQFL	ARHNGGNHPDQIRLG	KIISLCQVIEDCCCS	QNKTRVTPPEVAAK	IDLYLRGATNLEECL	653
VR2332	-----D-	S-----F	---S-----V---	-----	-----	-----	
16244B	-----D-	S-----F	---S-----V---	-----	-----	-----	
SP	ARLEKARPPRVMDS	FDWDVVLPGVEAATQ	TTELPQVNQCRALVP	VVTQKSLDNNVPLT	AFSLANYYYRAQGDE	VRHRERLTAVLSKLE	743
VR2332	-----I---	-----	-IK-----	-----	-----	-----	
16244B	-----I---	-----	-NK-----	-----	-----	-----	
SP	GVVREEYGLMPTGPG	PRPTLPRGLDELKDQ	MEVDLLKLANAQMTS	DMMAWAVEQVDLKTW	VKNYPRWTPPPPPPI	VQPRKTKLVKSLPES	833
VR2332	K-----E--	-----	--E---K---T--	-----	-----	-----K-----	P-----R
16244B	E-----E--	-----	--E---R---A--	-----	-----	-----K-----	P-----R
SP	KVPVAPRRKVRSDCD	CPTLSGNLPLDSWED	LAVGCPDLPSTPEP	VTPLESPASVSAPRR	SFRPVKPLSEVPVVP	APRKTVSRPATPLSE	923
VR2332	-----GS--G	S-VSL-GDV-N---	---SS-F---P---	A-PS--LVI--S-QC	I---AT-----A-I-	---G-----V-----	
16244B	-----GP--G	S-VSL-GDV-N---	---SS-L---P---	A-LS--LVI--S-QC	I---AT-----A-I-	---G-----V-----	
SP	PIVVPAPRRKQKQVE	KVNPAAAATLGCQDEF	PDLSSASHTYEYASP	LVLQNGDVLEVEER	EAEELSGISDILDA	IKPASASSSSSLSSV	1013
VR2332	-----K	RLSS---IPPY-D--	-----Q-----	PAP--S-G--G--GH	---T--E---MSGN	-----V-----	
16244B	-----K	RLSS--VPLH-N-P	L-----Q-----	SAP--S-G--G--GH	---T--E---MSGN	-----V-----	
SP	AITRPKYSAQAIIDS	GGPYSGHLQEVKETC	LSIMSEACDVTKLDD	PATQEWLSRMWDRVD	MLTWRNTSVHQASRT	LDDRKFKFLPKMILET	1103
VR2332	R-----	---C---E-----	--V-R---A-----	-----	-----Y---IC-	--G-L-----	
16244B	E-----	---C---G-----	--V-R---A-----	-----	-----C---IR-	--G-L-----	
SP	PPPYPCGFVMMPTP	APSVGAESDLTIGSV	ATEDVPRIFGKVNVD	CKMIDQRPLVLFENE	LADDQPARDPRTSSQ	RFDGSTPAPPAGTDG	1193
VR2332	-----E---H--	-----	-----LE-IEN-	GE-AN-G-LAFS-DK	PV---LVN---I--R	-P-E--S--S--G-	
16244B	-----E---H--	-----	-----LE-IEN-	GE-AN-E-SAFS-DK	PV---LVN---I--R	-P-E--A--S--G-	
SP	TGLASGPGVREVDSC	<b>EASSTEKIEQPFVLN</b>	<b>GGASTQASTFTNLPP</b>	<b>PGGIDAGSGPLQTV</b>	<b>RKKAERFFDLLSRQV</b>	<b>FNLVSHLPVFFSRLF</b>	1283
VR2332	A-		S--D--P	SD-A--D-G--FR-V	KR---L--Q-----	-D-----	1247
16244B	A-		S--D--S	SD-A--D-G--FR-A	KR---L--Q-----	-D-----	1247
SP	KPGGDYSPGDWGFAP	FTLLCLFLCYSYPAF	GAVPLLGVFSGSRR	VRMGFFGCWLAFAVS	LFPKPVSDPVGACEF	DSPECRNIIHSEFEL	1373
VR2332	Y--G--P-----	-----IA-----	-----S---	-----V-----	-----G-----	-----H-----	1337
16244B	H--G--T-----	-----IA-----	-----T---	-----V-----	-----G-----	-----L-----	1337
SP	KPWDVVRGLVVGVPV	LSLAIFGRLLG	1399				
VR2332	-----S-----	-G---L-----	1363				
16244B	-----S-----	-G---L-----	1363				

**Fig. 4.** Alignment of the predicted Nsp2 sequences of PRRSV strains SP, VR2332 and 16244B. Conserved residues are indicated by dashes. The 36 amino acid insertion in the Nsp2 of strain SP is shown in bold. Putative catalytic residues Cys<sup>1a437</sup> and His<sup>1a506</sup> are indicated by asterisks. Positions of amino acid residues are given to the right of the alignment

belong to closely related North American genotype. Second, the utmost 5'-end nucleotide is a C for 16244B [1], but a T for VR2332 [14]. Finally, seven nucleotides among the first ten of 16244B are different from those of VR2332. We were interested to investigate the genetic variations in the utmost 5'-end sequence of more North American strains. Using both the 5' RACE and the primer extension techniques, the 5'-ends of three viruses (SP, VR2332, and 12068-94p) were shown to be identical: the utmost 5'-end nucleotide is an A for each isolate

**Table 4.** Amino acid substitutes in Nsp1 to Nsp12 between PRRSV strains SP and 16244B

Position <sup>a</sup>	SP	16244B	Position <sup>a</sup>	SP	16244B	Position <sup>a</sup>	SP	16244B
Nsp1 $\alpha$			Nsp3			Nsp7		
18	M	V	1635	V	T	2254	T	M
35	P	L	1693	F	I	2300	T	V
42	P	S	1700	I	V	2339	N	S
105	L	I	1724	F	L	2369	L	P
118	T	A	1774	L	M	2400	R	K
127	H	R	1778	I	V	2466	D	N
141	Y	F	Nsp4			2467	R	K
Nsp1 $\beta$			1861	V	I	2473	T	I
171	D	E	1872	I	T	Nsp8		
197	E	R	1877	I		2500	V	M
199	V	I	1917	D	N	Nsp9		
213	P	A	1925	T	A	67	V	I
241	T	A	1929	E	A	74	M	L
247	R	C	1930	E	D	141	E	K
276	N	D	1975	I	V	151	I	V
284	N	D	1998	S	N	163	R	Q
291	G	N	2002	V	I	192	T	A
321	L	S	2029	E	K	208	S	P
329	C	Y	2031	V	I	389	P	S
344	E	G	2032	G	S	480	C	Y
349	P	S	Nsp5			sNsp10		
368	D	N	2079	G	S	704	W	G
369	K	T	2113	V	I	957	I	V
370	D	E	2133	Y	F	1025	S	G
371	E	G	2138	M	V	Nsp11		
378	N	S	2151	Y	H	1126	A	V
Nsp2			2158	S	N	1146	E	K
See Fig. 4			2168	I	M	1164	I	L
Nsp3			2178	V	I	1184	V	L
1403	H	C	2179	A	T	1187	K	E
1411	F	L	2197	C	G	12141	F	V
1465	N	D	2200	N	H	1301	K	R
1478	F	L				Nsp12		
1487	T	A	Nsp6			1399	K	R
1537	V	A	No difference			1446	M	L
1568	E	D				1449	Y	H
1571	I	V						

<sup>a</sup>The amino acid residue is numbered based on the SP

and conserved with that of strain LV. To confirm these results, the 5' RACE was repeated at least three times for each isolate. We believe that incubation of the TMV reverse transcriptase at high temperature (55 °C) is critical for the enzyme to overcome the secondary structures and reach the end of its template. The result may be helpful to clarify the controversy on the 5'-end sequence reported recently

[1, 12, 14] and may be crucial for making a full-length infectious cDNA originated from North American strains.

Sequence comparison revealed that the most variable region of PRRSV located within the 1a gene, especially the Nsp2-encoding region, in which an insertion and several highly variable regions were identified. In addition to the insertion, this protein contains 158 amino acid residue substitutes (Fig. 4). The variable regions may be associated with cell and tissue tropism for PRRSV and may be involved in species-specific functions for arteriviruses. The finding is the first experiment evidence showing that considerable variations exist within the closely related PRRSV isolates, suggesting that further classification of genotypes within the major North American serotype (genotype) may be required. Since the first report of PRRSV, isolates with different genetic background have been documented [2, 8, 17]. It is suggested that rapid molecular evolution of PRRSV may occur, perhaps adding difficulty to the control of PRRSV infection of swine herds worldwide. The Nsp2 gene may be an ideal marker for monitoring the genetic variation and for developing differential diagnostic means. As this gene seems to be more tolerable to mutations and insertions, it might be capable of harboring foreign genes if an arterivirus vector is to be developed. On the other hand, it would be interesting to investigate if the insertion would affect the biological function(s) of Nsp2, for example, the proteolytic activity at the Nsp2/Nsp3 junction of the 1a and 1a/b polyproteins.

### Acknowledgement

This work was supported by a grant from the National Science and Technology Board (NSTB), Singapore.

### References

1. Allende R, Lewis TL, Lu Z, Rock DL, Kutish GF, Ali A, Doster AR, Osorio FA (1999) North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions. *J Gen Virol* 80: 307–315
2. Andreyev VG, Wesley RD, Mengeling WL, Vorwald AC, Lager KM (1997) Genetic variation and phylogenetic relationships of 22 porcine reproductive and respiratory syndrome virus (PRRSV) field strains based on sequence analysis of open reading frame 5. *Arch Virol* 142: 993–1001
3. Bøtner A, Strandbygaard B, Sørensen KJ, Have P, Madsen KG, Smegegaard ME, Alexandersen S (1997) Appearance of acute PRRS-like symptoms in sow herds after vaccination with modified live PRRS vaccine. *Vet Rec* 141: 497–499
4. Cavanagh D (1997) *Nidovirales*: a new order comprising *Coronaviridae* and *Arteriviridae*. *Arch Virol* 142: 629–633
5. Dea S, Gagon CA, Mardassi H, Milane G (1996) Antigenic variability among North American and European strains of porcine reproductive and respiratory syndrome virus as defined by monoclonal antibodies to the matrix protein. *J Clin Microbiol* 34: 188–1493
6. den Boon JA, Faaberg KS, Meulenberg JJM, Wassenaar ALM, Plagemann PGW, Gorbalenya AE, Snijder EJ (1995) Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *J Virol* 69: 4 500–4 505

7. Drew TW, Meulenbergh JJM, Sands JJ, Paton DJ (1995) Production, characterization and reactivity of monoclonal antibodies to porcine reproductive and respiratory syndrome virus. *J Gen Virol* 76: 1 361–1 369
8. Kapur V, Elam MR, Pawlovich TM, Murtaugh MP (1996) Genetic variation in porcine reproductive and respiratory syndrome virus isolates in the Midwestern United States. *J Gen Virol* 77: 1 271–1 276
9. Kim HS, Kwang J, Yoon IJ, Joo HS, Frey ML (1993) Enhanced replication of porcine reproductive and respiratory syndrome (PRRS) virus in a homogeneous subpopulation of MA-104 cell line. *Arch Virol* 133: 477–483
10. Liu DX, Xu HY, Brown TDK (1997) Proteolytic processing of the coronavirus infectious bronchitis virus 1a polyprotein: identification of a 10 kilodalton polypeptide and determination of its cleavage sites. *J Virol* 71: 48–57
11. 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* 19: 62–72
12. Nelsen EA, Christopher-Hennings J, Drew T, Wensvoort G, Collins JE, Benfield DA (1993) Differentiation of US and European isolates of porcine reproductive and respiratory syndrome virus by monoclonal antibodies. *J Clin Microbiol* 31: 3 184–3 189
13. Nelsen CJ, Murtaugh MP, Faaberg KS (1999) Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents. *J Virol* 73: 270–280
14. Oleksiewicz MB, Bøtner A, Nielsen J, Storgaard T (1999) Determination of 5'-leader sequences from radically disparate strains of porcine reproductive and respiratory syndrome virus reveals the presence of highly conserved sequence motifs. *Arch Virol* 144: 981–987
15. Rossow KD, Collin JE, Goyal SM, Nelson EA, Christopher-Hennings J, Benfield DA (1995) Pathogenesis of porcine reproductive and respiratory syndrome virus infection in gnotobiotic pigs. *Vet Pathol* 32: 361–373
16. Rossow KD (1998) Porcine reproductive and respiratory syndrome. *Vet Pathol* 35: 1–20
17. Rowland RRR, Steffen M, Ackerman T, Benfield DA (1999) The evolution of porcine reproductive and respiratory syndrome virus: quasispecies and emergence of a virus subpopulation during infection of pigs with VR233. *Virology* 259: 262–266
18. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
19. Shen S, McKee TA, Wang ZD, Desselberger U, Liu DX (1999) Sequence analysis and in vitro expression of genes 6 and 11 of an ovine group B rotavirus isolates, Kb63: evidence for a non-defective, C-terminally truncated NSP1 and a phosphorylated NSP5. *J Gen Virol* 80: 2 077–2 085
20. Snijder EJ, Wassenaar AL, Spaan WJM (1994) Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J Virol* 68: 5 755–5 764
21. Snijder EJ, Wassenaar AL, Spaan WJM, Garbalenya AE (1995) The Arterivirus Nsp2 protease. *J Biol Chem* 270: 16 671–16 676
22. Snijder EJ, Wassenaar AL, van Dinten LC, Spaan WJM, Garbalenya AE (1996) The Arterivirus Nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J Biol Chem* 271: 4 864–4 871
23. Snijder EJ, Meulenbergh JJM (1998) The molecular biology of arteriviruses. *J Gen Virol* 79: 961–979
24. Snijder EJ, van Tol H, Pedersen KW, Raamsman MJB, De Vries AAF (1999) Identification of a novel structural protein of arteriviruses. *J Virol* 73: 6 335–6 345
25. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of

- progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4 673–4 680
26. Yang SX, Kwang J, Laegreid W (1998) Comparative sequence analysis of open reading frames 2 to 7 of the modified live vaccine virus and other North American isolates of the porcine reproductive and respiratory syndrome virus. *Arch Virol* 143 3: 601–612
  27. van Dinten LC, Wassenaar ALM, Gorbalenya AE, Spaan WJM, Snijder EJ (1996) Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. *J Virol* 70: 6 625–6 633
  28. van Dinten LC, Rensen S, Gorbalenya AE, Snijder EJ (1999) Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *J Virol* 73: 2 027–2 037

Authors' address: Dr. S. Shen, Institute of Molecular Agrobiolgy, 1 Research Link, National University of Singapore, Singapore.

Received September 16, 1999