37

ORGANIZATION OF THE SIMIAN HEMORRHAGIC FEVER VIRUS GENOME AND IDENTIFICATION OF THE sgRNA JUNCTION SEQUENCES

X. C. Wang, S. L. Smith, and E. K. Godeny

Department of Veterinary Microbiology and Parasitology
School of Veterinary Medicine
Louisiana State University
Baton Rouge, Louisiana 70803

1. ABSTRACT

SHFV is a member of the *Arteriviridae* family. Viruses within this family encode eight open reading frames (ORFs), two of which are translated from the full-length genome RNA. The remaining six ORFs are translated from a nested set of six or seven 3′ co-terminal, subgenomic RNAs (sgRNAs). We have cloned and sequenced approximately 6000 nucleotides (nt) from the 3′ end of the SHFV genome. Eleven ORFs, numbered ORFs 1a, 1b, 2a, 2b, 3, 4, 5, 6, 7, 8, and 9, were identified, three more than the other arteriviruses. The characteristics of the peptides encoded by ORFs 2a through 9 were determined from their computer-generated amino acid sequences. We also amplified the junction sequences from each of the SHFV subgenomic RNAs (sgRNAs) using RT-PCR analysis . Eight separate junction sequences were found which suggests that SHFV produces eight sgRNAs during replication. ORFs 2a and 2b appear to be encoded on the same sgRNA implying that RNA 2 is polycistronic. Sequence analysis identified the conserved SHFV junction sequence as 5′-(U/C)(C/U)N(U/C)(U/C)(A/C/G)AC(C/U)-3′. Since SHFV encodes additional ORFs and produces additional sgRNAs during replication, these data suggest that SHFV may be more complex than the other arteriviruses.

2. INTRODUCTION

Simian hemorrhagic fever virus (SHFV), along with equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) and porcine reproductive and respiratory

Coronaviruses and Arteriviruses, edited by Enjuanes *et al.*
Plenum Press, New York, 1998

syndrome virus (PRRSV), was recently reclassified into a new virus family, the *Arteriviridae*. These viruses are morphologically similar to the togaviruses; however, their genome organization and replication strategy is similar to the coronaviruses. During replication, the arteriviruses produce six or seven subgenomic mRNAs (sgRNAs) which are nested at the 3′ end of the viral genome (Snijder and Spaan, 1995). The arterivirus genomes consists of eight overlapping open reading frames (ORFs; den Boon *et al.*, 1991; Godeny *et al.*, 1993; Meulenberg *et al.*, 1993b). Due to a frame-shifting mechanism, the first two ORFs, ORFs 1a and 1b, are translated from the full length genome RNA as one large polypeptide (Snijder and Spaan, 1995). The remaining ORFs are translated from the sgRNAs which are 3′ co-terminal. Only the 5′-most ORF is translated from each sgRNA.

Each of the sgRNAs contains a leader sequence at the 5′ end which is encoded at the 5′ terminus of the viral genome. The 5′ leader sequence is joined to the sgRNAs at junction or intergenic sequences. These junction sequences are highly conserved among the sgRNAs of a specific virus and are similar among all of the arteriviruses thus far sequenced. The conserved junction sequences of the other arteriviruses are: 5′-UCAAC-3′ (EAV; den Boon *et al.*, 1991); 5′-U(A/G)(U/A)A(A/-)CC-3′ (LDV; Chen *et al.*, 1993) and 5′-GNUNAAC-3′ (PRRSV; Meulenberg *et al.*, 1993a).

It had previously been reported that SHFV produces six sgRNAs during replication and that these RNAs are nested and 3′ co-terminal (Godeny *et al.*, 1995; Zeng *et al.*, 1995). The SHFV capsid (C) and nonglycosylated membrane (M) proteins map to the ultimate and penultimate 3′ ORFs, respectively, suggesting that these proteins are translated from the two smallest sgRNAs (Godeny *et al.*, 1995). The SHFV 5′ leader was sequenced and the conserved junction sequences of the two smallest SHFV sgRNAs had been determined to be 5′-U(U/C)AACC-3′ (Zeng *et al.*, 1995). We have recently obtained the 3′ sequence of the SHFV genome (Smith *et al.*, 1997). The genome sequence shows that SHFV encodes three additional ORFs at the 3′ end of the genome as compared to the other arteriviruses. The purpose of this study was to identify the characteristics of the peptides encoded by the SHFV 3′ ORFs and to sequence and identify the conserved junction sequence for the remaining SHFV sgRNAs.

3. MATERIALS AND METHODS

3.1. Computer Analysis

The 3′ end of the genome sequence from the prototype strain, LVR 42–0/M6941, of SHFV was reported previously (Smith *et al.*, 1997) and can be accessed through GenBank (accession number U63121). Translation of the SHFV ORFs and the characteristics of the deduced peptides were determined by the Translation and Protein Analysis programs, respectively, supplied in the University of Wisconsin Genetics Computer Group (GCG) software (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, WI, 53711).

3.2. Isolation, Cloning, and Sequencing of the SHFV sgRNA Junction Sequences

Using the method of Sawicki *et al.* (1981), intracellular RNA was isolated from MA104 cells twenty hours post-infection with the prototype strain of SHFV. The SHFV

sgRNA junction sequences were reverse transcribed from the intracellular RNA using cDNA primers complementary to specific SHFV genome ORF sequences. The resulting SHFV cDNA was amplified in a thermocycler; the forward primer used was identical to nucleotides 60 through 77 of the SHFV 5′ leader sequence (Zeng *et al.*, 1995) and the reverse primers were those used to reverse transcribe the viral RNA. The amplified products were cloned into the pCRII plasmid vector supplied in the TA Cloning Kit (Invitrogen Corp., San Diego, CA). The resulting clones were sequenced by the dideoxy chain-termination method using the Sequenase™ DNA Sequencing Kit (U.S. Biochemical Corp., Cleveland. OH).

4. RESULTS

4.1. Organization of the SHFV Genome

The 3′ end sequence of the SHFV genome, beginning at the 3′ poly(A) tract and continuing into the helicase domain in ORF 1b, has been reported (Smith *et al.*, 1997). As shown in Figure 1, this sequence contains 9 complete ORFs. Although similar in organization to the other arteriviruses, SHFV has three additional ORFs at the 3′ end (Fig. 1). With the exception of the beginning of ORFs 4 and 7, all of the SHFV ORFs overlap adjacent ORFs. Interestingly, the 5′ end of ORF 2a overlaps the 3′ end of ORF 1b. This is a unique property among the arteriviruses as the 3′ ends of ORFs 1b of EAV, LDV, and PRRSV do not overlap their adjacent ORFs.

ORF 9, which encodes a peptide 111 amino acids in length, is the smallest SHFV ORF and ORF 2a is the largest ORF encoding a 281 amino acid peptide. The deduced peptides encoded by the SHFV ORFs have slightly acidic to very basic pI values ranging from 6.2 for peptides 3 and 5 through 11.7 for peptide 9. All of the SHFV ORFs encode peptides with at least one potential N-linked glycosylation site. However, the utilization of these potential glycosylation sites during viral protein processing has yet to be determined.

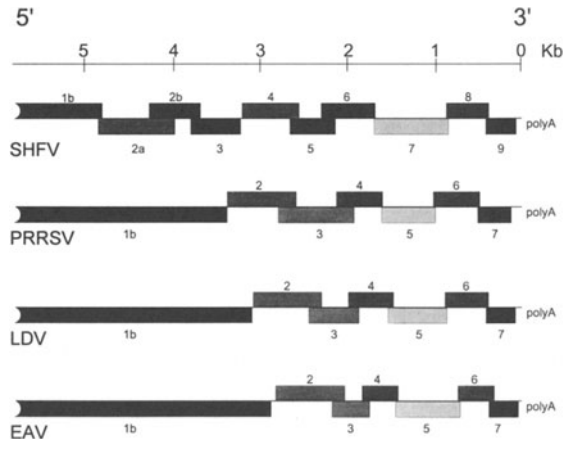

Figure 1. Comparison of the genome organization of the 3′ SHFV ORFs with those of the other arteriviruses. The size and location of each of the SHFV ORFs was determined from the SHFV genome sequence using the GCG Translation program. ORFs are drawn approximately to scale.

4.2. SHFV sgRNA Junction Sequences

The junction sequences on the SHFV sgRNAs between the leader sequence and the 5′-most ORF on that sgRNA were determined and are shown in Figure 2. Although SHFV encodes nine ORFs at the 3′ end of its genome, only eight SHFV sgRNA junction sequences were found. ORFs 2a and 2b appear to share the same junction sequence. There is a junction sequence located upstream of each of the remaining seven SHFV ORFs. The consensus junction sequence among the sgRNAs is 5′-(U/C)(C/U)N(U/C)(U/C)(A/C/G)AC(C/U)-3′ (Fig.

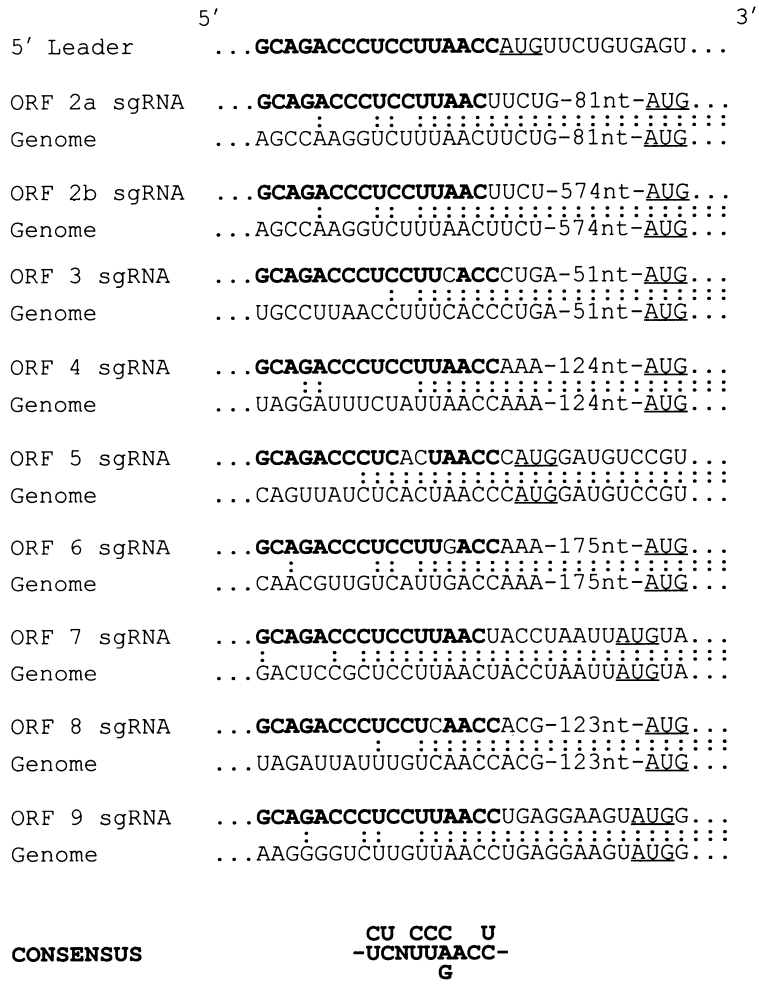

Figure 2. The 5′ nucleotide sequences of the SHFV sgRNAs encoding ORFs 2a, 2b, 3, 4, 5, 6, 7, 8 and 9. Nucleotides in bold-type represent the 5′ leader sequence which is encoded upstream of ORF 1a. ":" indicates nucleotide identity between the sgRNA and the genome sequences. Initiation codons for the respective ORFs are underlined. The consensus junction sequence is shown at the bottom.

2). The distance between the junction sequence and the initiation codon for the respective ORF varies from 1 nt (ORF 5) to 177 nts (ORF 6).

5. DISCUSSION

SHFV is one of four members of the new virus family, *Arteriviridae*. The genomes of the other arteriviruses, EAV, LDV and PRRSV, encode six ORFs in the region between ORF 1b and the 3′ terminus (Godeny *et al.*, 1993; Meulenberg *et al.*, 1993b; den Boon *et al.*, 1991). The ultimate and penultimate ORFs at the 3′ end of the arterivirus genomes encode the capsid and membrane proteins, respectively (Meulenberg *et al.*, 1995; de Vries *et al.*, 1992; Godeny *et al.*, 1990). For EAV and LDV, ORFs 2 and 5 have been shown to encode the small and large envelope glycoproteins (de Vries *et al.*, 1992; Faaberg and Plagemann, 1995). The PRRSV ORF 5 product was also shown to encode an envelope glycoprotein but the ORF 2 product could not be detected in purified virions (Meulenberg *et al.*, 1995). The PRRSV ORFs 3 and 4 gene products were also reported as viral envelope glycoproteins (van Nieuwstadt *et al.*, 1996); however, these gene products in EAV and LDV have not been identified.

The 3′ end of the SHFV genome RNA, beginning in the helicase domain within ORF 1b and ending at the 3′ poly(A) tract, has been cloned and sequenced (Smith *et al.*, 1997). Nine complete ORFs were found in this genome region, which is three additional ORFs as compared to the other arteriviruses. Although the gene products of ORFs 8 and 9 have been identified as the membrane and capsid proteins, respectively (Godeny *et al.*, 1995), the remaining SHFV ORF products remain to be identified.

Previously, SHFV has been shown to produce six sgRNAs during replication (Godeny *et al.*, 1995; Zeng *et al.*, 1995). However, this study suggests that SHFV produces at least eight sgRNAs, since separate junction sequences have been obtained for all of the 3′ ORFs except ORF 2b. These sgRNAs are diagramed in Figure 3. Interestingly, sgRNA 2 contains two complete ORFs at the 5′ end and therefore has the potential to en-

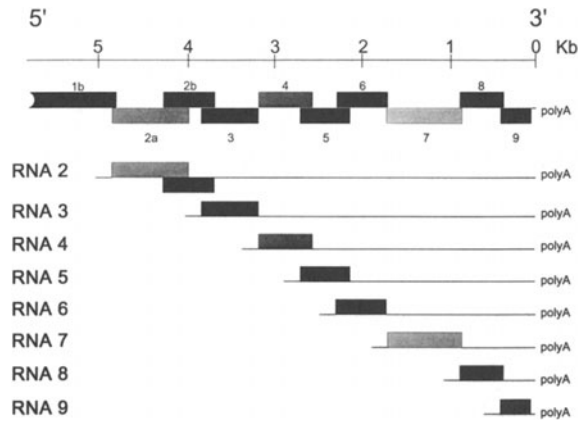

Figure 3. Replication strategy of SHFV. The deduced SHFV subgenomic mRNAs and the ORFs translated from each RNA species are shown.

code two peptides. Although this characteristic has been observed for a few of the coronavirus sgRNAs (Liu and Inglis, 1992; Senanayake *et al.*, 1992; Liu *et al.*, 1991), it is a unique feature among the arteriviruses.

From the eight junction sequences, we determined the conserved junction sequence on the SHFV sgRNAs as 5′-(U/C)(C/U)N(U/C)(U/C)(A/C/G)AC(C/U)-3′. This sequence is similar to the conserved junction sequences of EAV (den Boon *et al.*, 1991), LDV (Chen *et al.*, 1993) and PRRSV (Meulenberg *et al.*, 1993b). Each of the arterivirus junction sequences contains a form of the nucleotide sequence -AAC- towards the 3′ end of this sequence. The SHFV junction sequence is longer and more diverse than the junction sequences of the other arteriviruses. The SHFV sequence is also rich in uridine and cytidine residues; whereas, the EAV, LDV and PRRSV junction sequences are rich in adenine and uridine (Snijder and Spaan, 1995). Interestingly, the junction sequence found on the sgRNAs after the 5′ leader sequence was attached is not the same for all of the sgRNAs. The first two nucleotides at the 5′ end of the junction sequence were derived from the leader sequence whereas the last six nucleotides were derived from the genome RNA sequence. The third nucleotide is the most variable in the sequence and may be derived from either the leader or genome sequences. These data suggests that a switching event occurs between the 5′ leader sequence and position 3 of the junction sequence during the replication of the SHFV sgRNAs.

ACKNOWLEDGMENTS

This work was supported by a Public Health Service grant from the NCRR, RR06841.

REFERENCES

Chen, Z., Kuo, L., Rowland, R. R. R., Even, C., Faaberg, K. S. and Plagemann, P. G. W., 1993, Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of seven subgenomic mRNAs, *J. Gen. Virol.* **74:** 643–660.

de Vries, A. A. F., Chirnside, E. D., Horzinek, M. C. and Rottier, P. J. M., 1992, Structural proteins of equine arteritis virus, *J. Virol.* **66:** 6294–6303.

den Boon, J., 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-like family, *J. Virol.* **65:** 2910–2920.

Faaberg, K. S. and Plagemann, P. G. W., 1995, The envelope proteins of lactate dehydrogenase-elevating virus and their membrane topography, *Virology* **212:** 512–525.

Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V. and Brinton, M. A., 1993, Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV), *Virology* **194:** 585–596.

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.

Godeny, E. K., Zeng, L., Smith, S. L. and Brinton, M. A., 1995, Molecular characterization of the 3' terminus of the simian hemorrhagic fever virus genome, *J. Virol.* **69:** 2679–2683.

Liu, D. X., Cavanagh, D., Green, P. and Inglis, S. C., 1991, A polycistonic mRNA specified by the coronavirus infectious bronchitis virus, *Virology* **184:** 531–544.

Liu, D. X. and Inglis, S. C., 1992, Identification of two new polypeptides encoded by mRNA 5 of the coronavirus infectious bronchitis virus, *Virology* **186:** 342–347.

Meulenberg, J. J. M., de Meijer, E. J. and Moormann, R. J. M., 1993a, Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence, *J. Gen Virol.* **74:** 1693–1701.

Meulenberg, J. J. M., Hulst, M. M., de Meijer, E. J., Moonen, P. L. J. M., den Besten, A., de Kluyver, E. P., Wensvoort, G. and Moormann, R. J. M., 1993b, Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS) is related to LDV and EAV, *Virology* **192:** 62–72.

Meulenberg, J. J. M., Petersen-Den Besten, A., De Kluyver, E. P., Moormann, R. J. M., Schaaper, W. M. M. and Wensvoort, G., 1995, Characterization of proteins encoded by ORFs 2 to 7 of Lelystad virus, *Virology* **206:** 155–163.

Sawicki, S. G., Sawicki, D. L., Kääriänen, L. and Keranen, S., 1981, A sindbis virus mutant temperature sensitive in the regulation of minus strand RNA synthesis, *Virology* **115:** 161–172.

Senanayake, S. D., Hofmann, M. A., Maki, J. L. and Brian, D. A., 1992, The nucleocapsid gene of bovine coronavirus is bicistronic, *J. Virol.* **66:** 5277–5283.

Smith, S., Wang, X. and Godeny, E., 1997, Sequence of the 3' end of the simian hemorrhagic fever virus genome, *Gene* (In Press).

Snijder, E. J. and Spaan, W. J. M., 1995, The coronaviruslike superfamily, in: *The Coronaviridae* (S. G. Siddell, ed.), Plenum Press, New York, pp. 239–255.

van Nieuwstadt, A. P., Meulenberg, J. J. M., van Essen-Zandbergen, A., Petersen-den Besten, A., Bende, R. J., Moormann, R. J. M. and Wensvoort, G., 1996, Proteins encoded by open reading frames 3 and 4 of the genome of Lelystad virus (Arteriviridae) are structural proteins of the virion, *J. Virol.* **70:** 4767–4772.

Zeng, L., Godeny, E. K., Methven, S. L. and Brinton, M. A., 1995, Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5' leader, *Virology* **207:** 543–548.