VIROLOGY 209, 637–642 (1995)

SHORT COMMUNICATION

Sequence of the Genome of Lactate Dehydrogenase-Elevating Virus: Heterogenicity between Strains P and C

G. A. PALMER, L. KUO,[1] Z. CHEN, K. S. FAABERG, and P. G. W. PLAGEMANN[2]

Department of Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota 55455

Received December 8, 1994; accepted March 1, 1995

The complete nucleotide sequence of genomic RNA (14104 nt) of one strain of lactate dehydrogenase-elevating virus (LDV), LDV-P, is reported. It exhibits only about 80% nucleotide identity with the sequence reported for another LDV strain, LDV-C (Godeny *et al.*, Virology 194, 585–596 (1993), and is 68 nucleotides shorter than the reported LDV-C sequence. The difference in length is largely due to the lack of a 59-nucleotide-long direct repeat in ORF 1a of the reported LDV-C sequence. Sequence analysis of a total of 1.4 kb of ORF 1a of LDV-C via reverse transcription/polymerase chain reaction (RT/PCR) technology failed to confirm the presence of this repeat in the LDV-C genome as well as of 24 deletions/insertions of single nucleotides that give rise to apparent transient reading frame differences between the LDV-P and LDV-C genomes and might have represented frameshift mutations. An additional 35 nucleotides in ORF 1a of the RT/PCR LDV-C products were the same as in the LDV-P rather than the reported LDV-C genome. The nucleotide sequences of the 5′ leader and the 3′ noncoding ends of the two genomes and the heptanucleotides involved in joining the 5′ leader to the bodies of the subgenomic mRNAs were highly conserved or identical. The predicted LDV-P proteins, however, differed from those predicted for the LDV-C proteins between 25% for the ORF 2 protein and 1% for the ORF 7 nucleocapsid protein. All functional motifs of the ORF 1a and ORF 1b proteins were conserved. The ORF 1a protein possesses 11 potential transmembrane segments that flank the serine protease domain. © 1995 Academic Press, Inc.

Lactate dehydrogenase-elevating virus (LDV), along with equine arteritis virus (EAV), simian hemorrhagic fever virus, and porcine reproductive and respiratory syndrome virus (PRRSV) comprise a new group of enveloped, positive-strand RNA viruses provisionally classified as *Genus arterivirus* (*1, 2*). The organization and replication of the genomes of these viruses is similar to those of coronaviruses. Genome expression involves the formation of a 3′ coterminal nested set of six or seven subgenomic mRNAs, each of which possesses a 5′ leader derived from the 5′ end of the genome and a 3′ end poly(A) tail. Morphologically and in genome size (13–15 kb), on the other hand, these viruses differ greatly from coronaviruses (*3, 4*). They are much smaller (50–60 nm in diameter) than coronaviruses, possess a spherical rather than a helical nucleocapsid, and have a relatively smooth surface (*1*). The genomes of single isolates of EAV (Bucyrus strain, 12.7 kb; *5*), PRRSV (Lelystad virus, LV, 15.1 kb; *6*), and LDV (LDV-C, 14.2 kb; *7*) have been sequenced. The organization of the genomes of all three viruses is very similar and they all encode nine open reading frames (ORFs; see Fig. 1). ORFs 1a and 1b specify proteins with protease, replicase, helicase, and zinc binding motifs. The predicted products of ORFs 7, 6, and 5 are the nucleocapsid protein (N/VP-1, 12–14 kDa), the nonglycosylated envelope protein (N/VP-2, 18–19 kDa), and the major envelope glycoprotein (VP-3, 25–42 kDa), respectively. The virion envelope possesses a second minor glycoprotein (≥27 kDa) encoded by ORF 2 (*8* and K. S. Faaberg and P. G. W. Plagemann, manuscript in preparation). The functions of the predicted glycoproteins encoded by ORFs 3 and 4 are unknown.

We have previously reported the sequence of the 3.7-kb 3′ end and of the 1.7-kb 5′ end of the genome and of sections of ORF 1a and ORF 1b (cDNAs 4–11 and 4–35; see Fig. 1) of LDV strain P (*9–12*). We have now completed the sequence analysis of the LDV-P genome by sequencing additional genomic cDNAs isolated previously (C56, A69, 4–6; *9, 10*) and polymerase chain reaction (PCR) products (G1–5) generated to segments bridging these and other cDNAs (Fig. 1). Various segments in ORF 1a were confirmed by generation of additional reverse transcription (RT)/PCR products (see later).

Genomic RNA was extracted by a hot phenol–chloroform method from virus partially purified from pools of plasma of 1-day-infected FVB mice as described pre-

[1] Present address: Laboratory of Infectious Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892.

[2] To whom correspondence and reprint requests should be addressed.

0042-6822/95 $6.00
Copyright © 1995 by Academic Press, Inc.
All rights of reproduction in any form reserved.

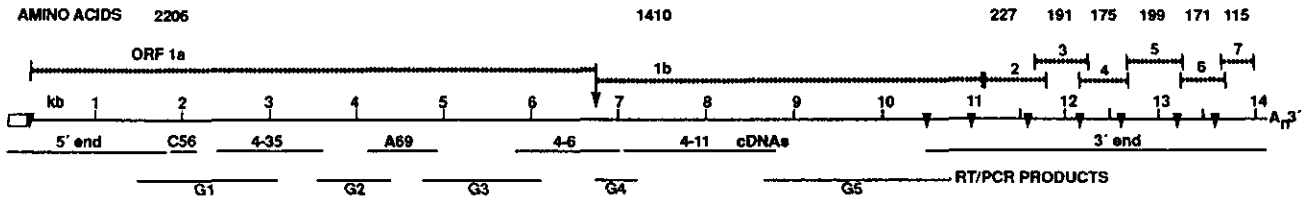

FIG. 1. Alignment of the ORFs (1a, 1b, 2–7) and of the 5′ and 3′ ends and various cDNAs and RT/PCR generated products (G1–G5) with the LDV-P genome. □ indicates the 5′ leader, ▼, the positions of the heptanucleotides involved in the joining of the 5′ leader to the bodies of the seven subgenomic LDV mRNAs (*11*), and A_n, the 3′ end poly(A) tail. The arrow indicates the frame shift location for the translation of the ORF 1b protein. The uppermost set of numbers indicates the numbers of amino acids of the predicted protein products.

viously (*13*). Samples of RNA were denatured with 100 m*M* methylmercury hydroxide and reverse transcribed using a cDNA cycle kit from Invitrogen (San Diego, CA) with only one round of reaction using random hexanucleotides or the antisense oligonucleotides listed in Table 1 as primers. Segments of the first-strand products were amplified by PCR as described previously (*11*), using the sets of antisense and sense oligonucleotides indicated in Table 1 as primers or additional sets of primers indicated in appropriate experiments. The PCR products were cloned into the *Sma*I site of pBluescript II KS(+) (Stratagene, La Jolla, CA), which had been modified for TA cloning (*11*). All cloned cDNAs and appropriate subclones thereof were sequenced in both orientations using the Sequenase Version 2.0 (USBC, Cleveland, OH). At least two independent clones of each PCR product were sequenced and each clone was sequenced at least twice.

The LDV-P genome consists of 14104 nt (Table 2), excluding the 40- to 80-nt-long poly(A) tail (*12*). The nucleotide sequence data have been submitted to the GenBank nucleotide sequence data bank and have been assigned Accession No. U15146. The base composition of LDV-P genomic RNA is 23.0% A, 27.6% U, 23.6% C, and 25.9% G and the CpG content is relatively low (ρ = 0.61) as also observed for the genomic RNAs of many other small RNA-containing eukaryotic viruses (*14*).

TABLE 1

Primers Used for RT/PCR Amplification

Segment[a]	RT primer[b] antisense	PCR primers[b] Sense	Antisense
G1	Random	D1001/(1407–1430)	/D1004 (3032–3053)
G2	Random	D1005/(3595–3615)	/D1007 (4428–4449)
G3	/A886 (6453–6472)	D1008/(4780–4802)	/D1010 (5926–5995)
G4	/D1013 (7834–7855)	D1011/(6657–6778)	/A932 (7080–7102)
G5	/PI-1 (10821–10841)	PI-3/(8589–8610)	/A1419 (10757–10780)

[a] For location in genome, see Fig. 1.

[b] The oligonucleotides were 20–25 nt in length. The slashes before and after the oligonucleotide designation indicate antisense and sense, respectively. The numbers in parentheses indicate the nucleotide segment in the LDV-P genome each represents. The expected PCR products ranged from 423 to 2200 nt.

Completion of the LDV-P genome sequence permits, for the first time, a detailed sequence comparison of the complete genomes of two independent isolates of one of the viruses in this virus group. LDV-P was isolated from a tumor-bearing C3H mouse (*15*) and LDV-C from a C58 mouse bearing the Ib tumor (*16*). The transplantable tumors of these mice probably became contaminated by passage in an LDV-infected laboratory mouse, but when and in which strain of mouse is unknown. However, both LDV-P and LDV-C are probably derived from LDVs endemic in wild house mouse populations (*Mus musculus domesticus; 1, 2, 17*). LDV-C and LDV-P replicate equally well in all strains of laboratory mice examined and in primary cultures of peritoneal macrophages from these mice, but LDV-C is more neurovirulent for C58 and AKR mice than LDV-P (*1, 16*).

The LDV-P genome is colinear with that reported for LDV-C but exhibits only ~80% nucleotide identity with that of LDV-C and is 68 nucleotides shorter (Table 2). However, in spite of this considerable sequence diver-

TABLE 2

Comparison of Nucleotide Sequences of Various Segments of the Genomes of LDV-P and LDV-C

Genome segment	Length (nt) LDV-P	LDV-C[a]	nt Identity (%)[b]
Total w/o poly(A)	14104	14172	~80
5′ Leader	156	>161	97
ORF 1a	6618	6678	76
ORF 1b	4230	4233	77
ORFs 2 to 7	2979	2979	84
Noncoding 3′ end	80	80	100
5′ Leader junctions	7	7	100
Slippery sequence	UUUAAAC	UUUAAAC	100

[a] Sequence reported by Godeny *et al.* (*7*; GenBank Accession No. L13298).

[b] Nucleotide (nt) identities were determined by the algorithm of Needleman and Wunsch (*18*) to align sequences and divide the number of identical nucleotides by the total number of nucleotides of the smaller of the indicated segments.

gence, certain segments that probably play a role in viral RNA transcription and replication are highly conserved. These include the 5′ leader, the heptanucleotide (UAUAACC) at its 3′ end, the sequence-related heptanucleotides preceding each of ORFs 2–7, where the 5′ leader becomes joined to the bodies of the corresponding subgenomic mRNAs (*11*), the slippery sequence that is presumably involved in the expression of ORF1b via frameshifting (*5*), and the 80-nt-long noncoding segment at the 3′ end (Table 2). The identity of the heptanucleotide junctions between the 5′ leader and the bodies of the subgenomic mRNAs suggests that the joining is a highly site-specific process (*13*). It is of interest that the segments on either side of each junction preceding ORFs 2–7 are also highly conserved in LDV-P and LDV-C (data not shown).

Most of the apparent nucleotide differences between the genomes of the two LDV strains are located in the ORFs. ORFs 1a and 1b differ in length and exhibit only 76 and 77% nt identity, respectively (Table 2). The segments containing ORFs 2–7 of LDV-P and LDV-C have the same lengths and ORFs 2–7 of LDV-P and LDV-C overlap identically with their neighboring ORFs by between 1 and 129 nt (*11;* see Fig. 1), but nucleotide identity is only 84%. Many of these nucleotide differences result in amino acid differences of the predicted gene products (see later).

However, some of the apparent nucleotide differences in the ORFs and differences in the lengths of the genomes of LDV-P and LDV-C could not be confirmed. For example, a number of nucleotide differences between ORF 1a of LDV-P and that reported for LDV-C were unusual in that they involved compensating single nucleotide deletions/insertions that would result in transient frame shifts and thus in segments with quite different amino acids in the predicted proteins of the two strains. One such example is shown in Fig. 2 (see *). Mutations in viral RNA sequences resulting in such alterations seem very uncommon, if they exist at all (*19*). We have therefore resequenced segments of ORF 1a of LDV-C and LDV-P that contained such apparent frame shift mutations via generation of RT/PCR products (a total of about 1.4 kb; LDV-C was provided by M. Brinton). In general we found good agreement between the LDV-C-PCR sequence and that reported for LDV-C (Figs. 2 and 3). However, in the segment depicted in Fig. 2, we could not confirm any of the six apparent nucleotide deletions in the reported LDV-C sequence that resulted in frame shifts as well as the loss of two amino acids in the predicted LDV-C ORF 1a protein. Throughout the depicted segment the LDV-C-PCR sequence was colinear with that of LDV-P and had an identical number of nucleotides. In addition there were 13 nt in the LDV-C-PCR sequence that differed from those reported for LDV-C but instead were identical to those in LDV-P. These corrections significantly reduced the number of amino acid differences in this segment of the predicted ORF 1a proteins of LDV-C and LDV-P (from about 47 to 22%).

A similar situation pertained to at least three other segments of ORF 1a. For instance, the ORF 1a segment depicted in Fig. 3A contained four apparent compensating nucleotide deletions/insertions in the LDV-P and LDV-C genomes that resulted in transient frame shifts. None was confirmed by the sequence analyses of LDV-C-PCR products and an additional 6 nt in the LDV-C-PCR sequence were identical to those in LDV-P rather than those reported for LDV-C. Furthermore, in the 358-nt-long segment directly upstream of that depicted in Fig. 3A (i.e., LDV-C nt 1822–2254) a total of 21 nt (or 4.9%) of the LDV-C-PCR products differed from those reported for LDV-C and instead were identical to those in LDV-P (data not shown).

In a further downstream segment in ORF 1a, the reported LDV-C sequence contained a 59-nt-long direct repeat (Fig. 3B). The first repeat plus one additional nucleotide were lacking in two independent clones of LDV-P as well as in two independent clones of LDV-C-PCR. Thus, it seems likely that this segment of ORF 1a of LDV-C is 60 nt shorter than previously reported. Another segment of ORF 1a of LDV-C (nt 5685–5760) contained six nucleotides more than the corresponding ORF 1a segment of LDV-P. The presence of only three additional nucleotides was indicated by our analysis of LDV-C-PCR (data not shown). Thus, overall, our results suggest that the ORF 1a of LDV-C may have only three more nucleotides than that of LDV-P resulting in one additional amino acid in the ORF 1a protein, i.e., 2207 amino acids (see Fig. 1) rather than the reported 2226 amino acids (*7*).

There are also three pairs of compensating nucleotide deletions/insertions in ORF 1b of LDV-C and LDV-P, 38, 18, and 30 nt apart, respectively (data not shown). They fall within a 340-nt segment (nt 7432–7770 for LDV-P) and result in transient frame shifts. These and additional nucleotide differences account for some amino acid divergence of the predicted protein products. One such divergent protein sequence of 34 amino acids is illustrated in Fig. 3C (amino acids 323–356 in ORF 1b of LDV-C). This segment of the LDV-C ORF 1b protein exhibits only 23% amino acid identity with that of LDV-P and even less with the corresponding PRRSV-LV and EAV ORF 1b segments. In contrast, the LDV-P segment exhibits 83 and 65% amino acid identity with the corresponding segments of PRRSV-LV and EAV, respectively. Another upstream stretch of 14 amino acids of the ORF 1b protein of LDV-P has no amino acid identity with the reported LDV-C segment, whereas exhibiting 64% amino acid identity with the PRRSV-LV segment. Also, 3 nt of the ORF 1b of LDV-C are missing in the corresponding LDV-P segment, resulting in the deletion of amino acid 183 (proline).

There is also a difference in the length of the 5′ leader. The 5′ end of the LDV-P genome has been accurately

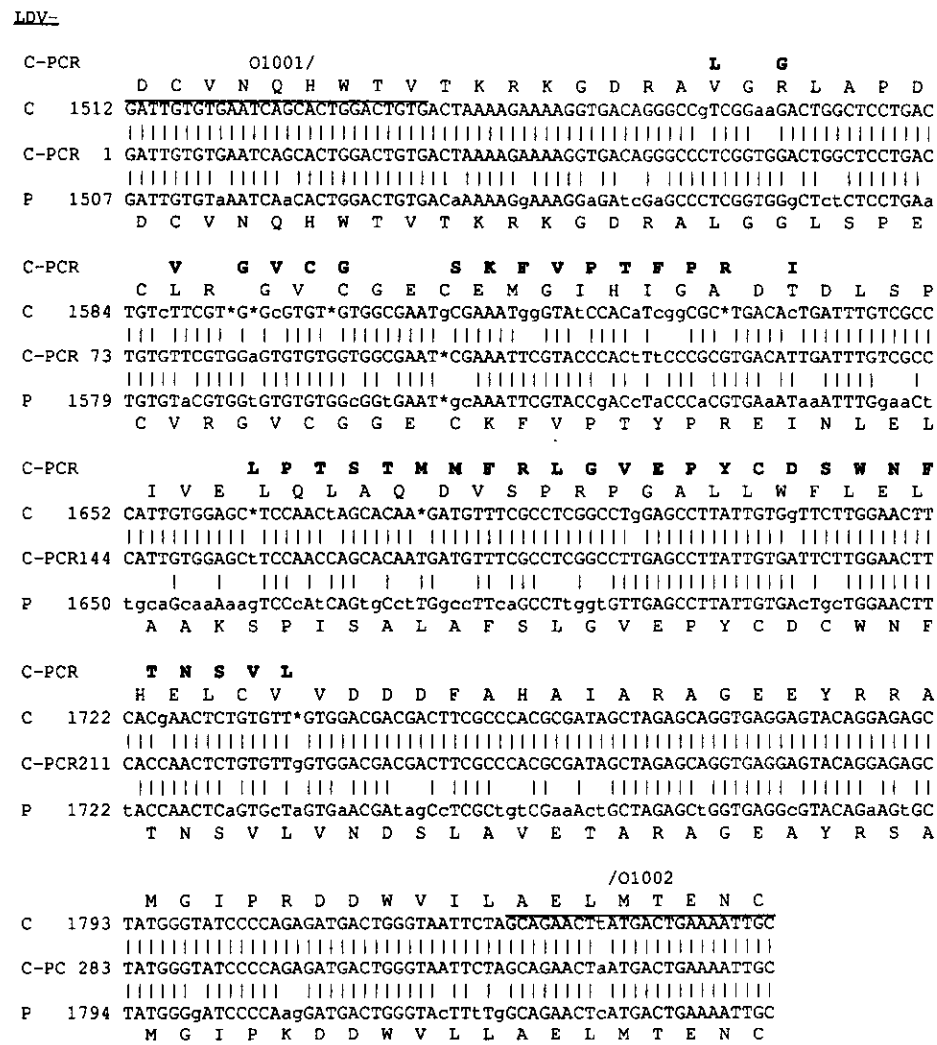

FIG. 2. Comparison of an ORF 1a segment (nt 1507–1851) of the LDV-P genome and of the encoded amino acid sequence with those of the LDV-C genome (*7*) and a RT/PCR product of this segment of LDV-C (LDV-C-PCR). The LDV-C-PCR sequence was generated by RT of LDV-C genomic RNA using random hexanucleotides as primer. The first strand products were amplified by PCR using, as indicated, antisense oligonucleotide /01002 and sense oligonucleotide 01001/ as primers. *denotes apparent gaps in the nucleic acid sequences. The amino acids encoded by the LDV-C-PCR segment that differ from those reported for LDV-C are indicated by boldface letters.

determined by two independent methods (*12*). The leader is 156 nt long inclusive of the 3′ end heptanucleotide junction segment. That of the LDV-C genome has been reported to be at least 5 nt longer, contributing to the apparent differences in the lengths of the genomes (Table 2).

The amino acid identities of the predicted gene products of LDV-P and LDV-C are as follows. The predicted ORF-1a and -1b proteins exhibit only 82.2 and 83.9% amino acid identity, respectively. These values are slightly underestimated since the nucleotide deletions/insertions in the reported LDV-C genome that cause transient usage of different reading frames in the two genomes seem in error. All putative functional motifs that have been identified (*7, 11*), i.e., two papain-like cysteine protease (PCP) motifs and a serine protease motif (SP) in ORF 1a (see Fig. 4) and the RNA replicase, helicase, zinc-binding motifs in ORF 1b are identical for LDV-P and LDV-C. These motifs are shared with PRRSV, EAV, and coronaviruses (*7*). A hydrophobic moment analysis of the ORF 1a protein of LDV-P revealed the presence of 11 potential transmembrane segments which flank the SP motif (Fig. 4). Hydrophobic moment analyses of the ORF 1a proteins of EAV and PRRSV-LV revealed the presence of similar 11 or 12 potential transmembrane segments flanking the SP motif (data not shown), even though the ORF 1a proteins of these three viruses differ considerably in size (2306, 2206, and 1727 amino acids for the ORF 1a proteins of PRRSV-LV, LDV-P, and EAV, respectively) and exhibit only 30–40% amino acid identities. The EAV ORF 1a protein possesses only one PCP which autocatalytically cleaves off the 29-kDa N-terminal end of the ORF 1a protein (*21*). The two PCPs in the ORF 1a protein of LDV (Fig. 4) autocatalytically release proteins of 22 and

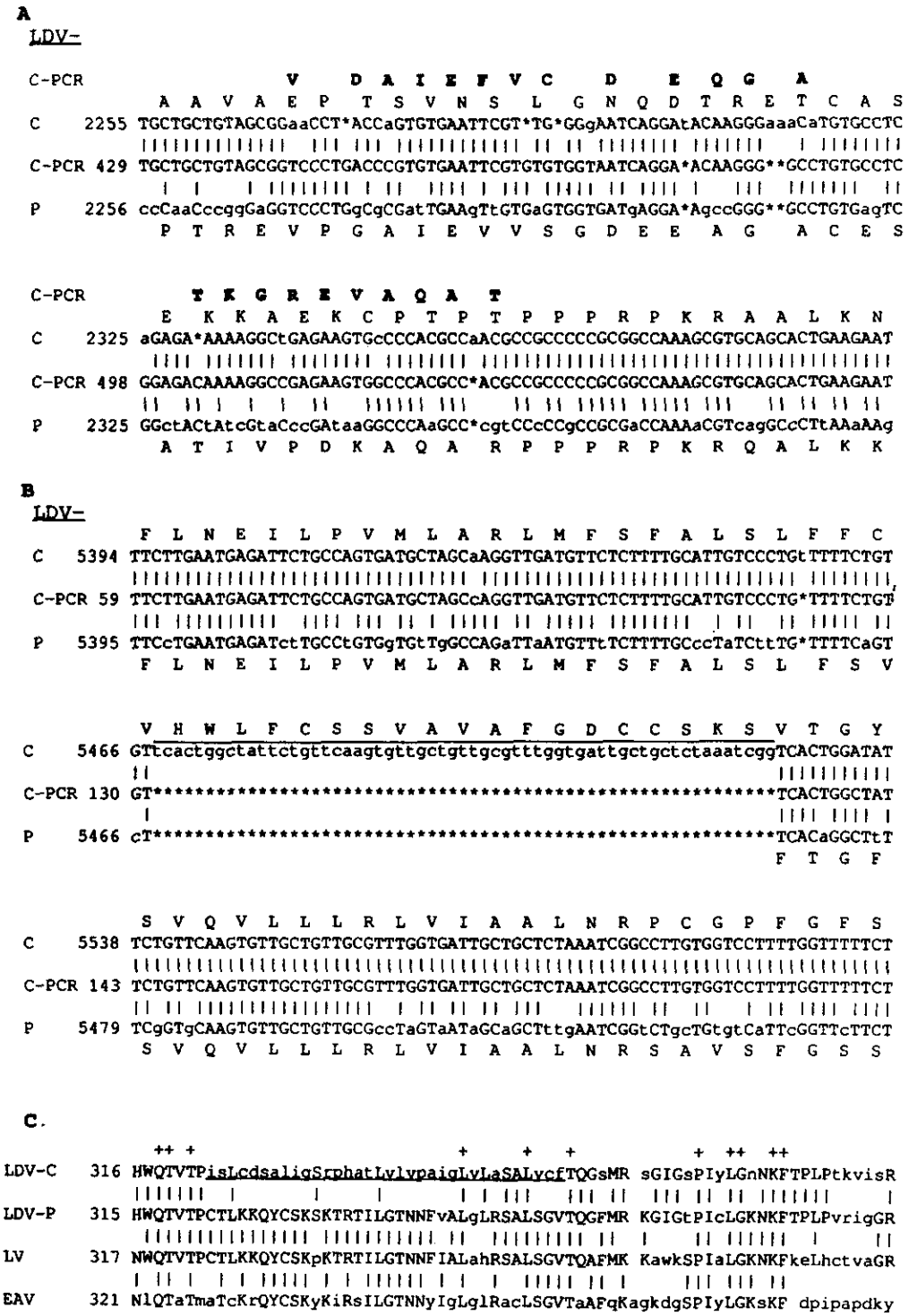

FIG. 3. Comparison of nt 2356–2297 (A) and 5394–5550 (B) of ORF 1a of LDV-P and of the encoded amino acid sequence with those of the corresponding segments of the LDV-C genome and the LDV-C RT/PCR products (LDV-C-PCR) and (C) comparison of the amino acid sequences of a portion of the predicted ORF 1b proteins of LDV-C, LDV-P, PRRSV-LV, and EAV. The RT/PCR products were generated by RT of genomic LDV-C RNA using random hexanucleotides as primers. The first strand products were amplified by PCR using LDV-C-specific sense oligonucleotides 01003/ (nt 1827–1860) and antisense oligonucleotide /G1047 (nt 2403–2428) in A and sense oligonucleotide G1044/ (nt 5336–5363) and antisense oligonucleotide G1040 (nt 5826–5850) in B as primers. The predicted products were 601 and 449 nt long, respectively. Only pertinent portions of these products are shown. *denotes apparent gaps in the nucleic acid sequences. In A, the amino acids predicted for the LDV-C-PCR sequence that differed from the one predicted by the reported LDV-C sequence are indicated by boldface letters. In B the 59-nt-long nucleotide direct repeat segment of LDV-C is overlined. In C, a segment in the ORF 1b protein of LDV-C that greatly diverges from those of the other three viruses is underlined. Amino acids that are identical in all four viruses are indicated by a + sign.

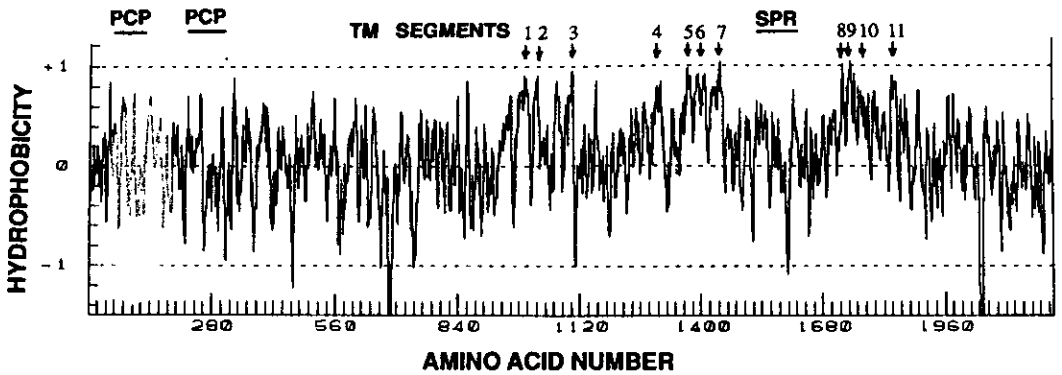

FIG. 4. Hydrophobic moment analysis of the predicted ORF 1a protein of LDV-P according to Eisenberg *et al.* (*20*). Potential transmembrane (TM) segments 1–11 were identified on the basis of mean hydrophobicities (H) as follows: Amino acids 980–1000, 1011–1031, 1084–1103, 1290–1310, 1359–1379, 1388–1408, 1427–1448, 1718–1738, 1739–1759, 1762–1782, and 1832–1852 with H values of 73.3, 52.7, 59.0, 62.4, 71.0, 77.0, 76.9, 58.1, 66.0, 58.7, and 56.5%, respectively. Also indicated are the location of the papain-like cysteine protease (PCP) and the serine protease (SP) motifs.

26 kDa, respectively (J. den Boon *et al.*, submitted for publication).

The nucleocapsid protein (N/VP-1) and the nonglycosylated envelope protein (M/VP-2) encoded by ORF 7 and 6, respectively, are highly conserved in LDV-P and LDV-C (98.7 and 95.4% amino acid identity, respectively; *7, 10*), probably because their interaction with genomic RNA in the assembly of the nucleocapsid and their function in the subsequent budding of the nucleocapsid into single membrane, cytoplasmic cisternae, respectively (*1, 2*), allow for little sequence flexibility. On the other hand, the glycoproteins predicted for ORFs 2–5 of LDV-P and LDV-C differ markedly (75.3, 88.0, 88.6, and 86.9 amino acid identity, respectively). The amino acid identities of the products predicted for each ORF of LDV-P and LDV-C are higher than the nucleotide identities of the ORFs. Thus, many of the nucleotide differences between the ORFs of the LDV-P and LDV-C genomes are silent.

ACKNOWLEDGMENTS

We thank Zongming Chen and John Torseth for isolating the G1 and G2 cDNAs, respectively (Fig. 1), Gao Feng Lin for participating in some of the sequencing work, Ernie Retzel for advice in some of the computer analyses, and Melodie Bahan for excellent secretarial assistance. This work was partly supported by USPHS Research Grant AI15267 and Training Grant CA09138 (G.A.P. and K.S.F.).

REFERENCES

1. Plagemann, P. G. W., *In* "Virology" (B. N. Fields, D. M. Knipe, and D. M. Howley, Eds.), 3rd Ed., Raven Press, New York, in press.
2. Plagemann, P. G. W., and Moennig, V., *Adv. Virus Res.* **41,** 99–192 (1992).
3. Holmes, K. V., *In* "Virology" (B. N. Fields, Ed.), 2nd Ed., pp. 841–855. Raven Press, New York, 1990.
4. Lai, M. M. C., *Annu. Rev. Microbiol.* **44,** 303–333 (1990).
5. den Boon, J., Snijder, E. J., Chirnside, E. D., de Vries, A. A. F., Horzinek, M. E., and Spaan, W. J. M., *J. Virol.* **65,** 2910–2920 (1993).
6. Meulenberg, J. J. M., Hulst, M. M., de Meijer, E. J., Moonen, P. J. L. M., den Besten, A., de Kluyver, E. P., Wensvoort, G., and Moormann, R. J. M., *Virology* **192,** 62–72 (1993).
7. Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V., and Brinton, M. A., *Virology* **194,** 585–596 (1993).
8. de Vries, A. A. F., Chirnside, E. D., Horzinek, M. C., and Rottier, P. J. M., *J. Virol.* **66,** 6294–6303 (1992).
9. Kuo, L., Harty, J. T., Erickson, L., Palmer, G. A., and Plagemann, P. G. W., *J. Virol.* **65,** 5118–5123 (1991).
10. Kuo, L., Chen, Z., Rowland, R. R. R., Faaberg, K. S., and Plagemann, P. G. W., *Virus Res.* **23,** 55–72 (1992).
11. Chen, Z., Kuo, L., Rowland, R. R. R., Even, C., Faaberg, K. S., and Plagemann, P. G. W., *J. Gen. Virol.* **74,** 643–660 (1993).
12. Chen, Z., Faaberg, K. S., and Plagemann, P. G. W., *J. Gen. Virol.* **75,** 925–930 (1994).
13. Chen, Z., Faaberg, K. S., and Plagemann, P. G. W., *Virus Res.* **34,** 167–177 (1994).
14. Karlin, S., Doerfler, W., and Cardon, L. R., *J. Virol.* **68,** 2889–2897 (1994).
15. Brinton-Darnell, M., and Plagemann, P. G. W., *J. Virol.* **16,** 420–433 (1975).
16. Martinez, D., Brinton, M., Tachovsky, T. G., and Phelps, A. H., *Infect. Immun.* **27,** 979–987 (1980).
17. Rowson, K. E. K., and Mahy, B. W. J., "Virology Monograph," Vol. 13. Springer-Verlag, Wien, 1975.
18. Needleman, S. B., and Wunsch, C. D., *J. Mol. Biol.* **48,** 443–453 (1970).
19. Bonilla, P. J., Gorbalenya, A. E., and Weiss, S. R., *Virology* **198,** 736–740 (1993).
20. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R., *J. Mol. Biol.* **179,** 125–142 (1984).
21. Snijder, E. J., Wassenaar, A. L. M., and Spaan, W. J. M., *J. Virol.* **66,** 7040–7048 (1992).