Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

VIROLOGY 194, 585–596 (1993)

Complete Genomic Sequence and Phylogenetic Analysis of the Lactate Dehydrogenase-Elevating Virus (LDV)

E. K. GODENY,*,[1] L. CHEN,* S. N. KUMAR,* S. L. METHVEN,* E. V. KOONIN,† AND M. A. BRINTON*,[2]

*Department of Biology, P.O. Box 4010, Georgia State University, Atlanta, Georgia 30302; and †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894

Received September 3, 1992; accepted February 16, 1993

The apparently complete sequence of the RNA genome of the neurovirulent isolate of lactate dehydrogenase-elevating virus (LDV-C) has been determined. The LDV-C genome is at least 14,222 nucleotides in length and contains eight open reading frames (ORFs). ORF 1a, which encodes a protein of 242.8 kDa and is located at the 5′ end of the genome, contains at least two putative papain-like cysteine protease domains, and one putative chymotrypsin-like serine protease domain. This ORF terminates with a UAG stop codon that can be bypassed if a −1 frameshift occurs. The frameshift region consists of a heptanucleotide "slippery" sequence, 5′-UUUAAAC-3′, followed by a putative pseudoknot. ORF 1b encodes a protein of 155.4 kDa containing, in its N-terminal portion, an RNA-dependent RNA polymerase and an RNA helicase domain separated by a Zn finger domain. Another domain of unknown function that is also conserved in coronaviruses and toroviruses is located at the C-terminus of the ORF 1b product. Three cleavage sites in the ORF 1a polyprotein and three in the ORF 1b polyprotein were predicted for the chymotrypsin-like protease and tentatively delimit the mature nonstructural proteins of LDV. Six small, overlapping 3′ ORFs (ORFs 2 through 7) encode proteins with calculated sizes of 25.8, 21.6, 19.8, 23.9, 18.9, and 12.3 kDa. ORF 7 encodes the virion nucleocapsid protein Vp-1, while ORF 6 encodes the nonglycosylated envelope protein Vp2. ORFs 5, 4, 3, and 2 each encode glycoproteins which may be virion envelope proteins. LDV is closely related to equine arteritis virus, Lelystad virus (LV), and simian hemorrhagic fever virus. These four viruses belong to a new group of positive-strand RNA viruses and are related to coronaviruses and toroviruses.

INTRODUCTION

Lactate dehydrogenase-elevating virus (LDV) infects only mice and always causes a persistent infection (Notkins, 1965; Brinton, 1982). Infected mice display permanently elevated (5- to 10-fold) serum levels of lactate dehydrogenase and, to a lesser degree, of six additional serum enzymes. The increase in LDH levels is observed by 4 days after infection and is due primarily to a decrease in the rate of enzyme clearance (Notkins, 1965; Brinton, 1982). It has been postulated that Kupffer cells involved in enzyme turnover are infected and killed as a direct result of LDV replication (Smit *et al.*, 1989). The elevated serum LDH allowed the original discovery of the virus, provided the name of the virus, and is utilized as an endpoint for titration of viral infectivity (Brinton, 1982; Rowson and Mahy, 1985).

Infected mice display a lifelong viremia (Notkins, 1965). Antiviral antibodies are elicited in infected animals, but virus is not cleared from the bloodstream; viral–immune complexes are infectious and presumably continue to attach to viral target cells, a subpopulation of macrophages, via Fc receptors. Since an established cell line has not yet been identified which is permissive for LDV replication, LDV is grown in primary murine cell cultures which contain macrophages (Rowson and Mahy, 1985). Although the majority of the LDV isolates produce inapparent infections in the mice they infect, one isolate, designated LDV-C, induces a sometimes fatal poliomyelitis in immunosuppressed individuals of a few susceptible inbred mouse strains (Martinez *et al.*, 1980).

LDV was initially classified within the Togaviridae family based on the infectivity of its single-stranded RNA and its virion morphology (Fenner, 1977) and is currently included in the genus *Arterivirus* belonging to this family (Francki *et al.*, 1991). The mapping of the LDV capsid protein to the 3′ end of the LDV genome (Godeny *et al.*, 1990) and the demonstration of a 3′ coterminal nested set of LDV subgenomic mRNAs in infected macrophages (Kuo *et al.*, 1991) indicate that LDV is not a togavirus. Although we refer to these viruses as arteriviruses for convenience in this paper, it is quite possible that the name denoting this group of viruses will be changed at the time of their reclassification and removal from the Togaviridae family. In order

[1] Present address: Department of Veterinary Microbiology and Parasitology, Louisiana State University, School of Veterinary Medicine, Baton Rouge, LA 70803.

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

0042-6822/93 $5.00

to further characterize the genome structure and the encoded proteins of LDV, the genome of the neurotropic LDV-C isolate was sequenced.

MATERIALS AND METHODS

Virus and genome RNA

Adult (5- to 6-week-old) C58/J mice were injected intraperitoneally with 10^7 ID_{50} of LDV-C in 0.1 ml of MEM containing 10% FCS. Plasma was obtained from these mice 24 hr after infection. The LDV-C virions were pelleted through a discontinuous glycerol gradient consisting of layers of 5, 10, and 20% glycerol (Brinton *et al.*, 1986). Pelleted virions were incubated for 15 min at 37° with pronase (2 μg) and vanadyl ribonucleoside complex (2 m*M*) and then disrupted by addition of SDS to a final concentration of 1%. Genomic RNA was purified by sedimentation through a 15–35% SDS–sucrose gradient as previously described (Brinton *et al.*, 1986). The RNA in the peak fractions was then ethanol precipitated, aliquoted, reprecipitated, and stored under ethanol at −70° until use.

Cloning of LDV cDNA

Genome RNA was used as template for oligodeoxythymidine (dt)-primed or calf-thymus pentameric DNA (ct)-primed reverse transcription. Second-strand cDNA was made according to the method of Okayama and Berg (1982). The cDNA was then methylated with *Eco*RI methylase, ligated to *Eco*RI linkers, digested with *Eco*RI, and inserted into a pUC13 (Pharmacia, Piscataway, NJ) plasmid vector (Rice *et al.*, 1985). *Escherichia coli,* strain JM 103, was transformed with recombinant pUC13 DNA, and the resulting colonies were screened for inactivation of β-galactosidase gene expression (Close *et al.*, 1983). Miniprep DNA was prepared from selected clones and digested with *Eco*RI. The DNA inserts were sized on 1% agarose gels and tested for virus-specificity by Southern blot analysis (Southern, 1975) using a radiolabeled cDNA probe primed with random ct oligomers and reverse transcribed from the LDV-C genome RNA.

For regions not represented in the cDNA libraries, additional clones were generated using cDNA–PCR products made from the LDV-C genome RNA with specific primers synthesized by the Georgia State University DNA synthesis facility. The PCR products were gel purified (Maniatis *et al.*, 1989) prior to unidirectional cloning into the pCR2000 plasmid (Invitrogen, San Diego, CA).

Sequencing methods

Double-stranded miniprep cDNA was sequenced by a modified dideoxy chain-termination method (Sanger *et al.*, 1977; Winship, 1989) with a Sequenase kit (USB, Cleveland, OH) using M13-universal and M13-reverse primers. Internal virus-specific primers were synthesized to complete the sequencing of long inserts. PCR products were sometimes sequenced directly as described previously (Nainan *et al.*, 1991) or more recently by the cycle sequencing method using a Δ Taq Cycle Sequencing Kit (USB).

A total of 161 cDNA clones containing over 47,000 nucleotides were sequenced and aligned. The cDNA clones sequenced ranged in size from 100 to 1250 nucleotides with the majority being in the size range of 500 to 800 nucleotides. The density of clones covering most genome regions was at least three and in many regions eight or more. Regions for which fewer than three clones were available and six gaps of 170, 150, 90, 340, 260, and 690 nucleotides in the genome consensus sequence were sequenced directly from genome RNA (Nainan *et al.*, 1991) or from cloned cDNA–PCR products generated with specific primers.

Computer analysis of sequence data

The consensus nucleotide sequence for the LDV-C genome was assembled and analyzed using the fragment assembly software provided in version 7.0 of the University of Wisconsin Genetics Computer Group (GCG) (Devereux *et al.*, 1984). The amino-acid sequences of the ORF 1a and 1b products were aligned with the homologous sequences from Lelystad virus (LV) and equine arteritis virus (EAV) using the MACAW software (Schuler *et al.*, 1991). Portions corresponding to highly conserved domains were cut out and fit into previously published multiple alignments of viral proteins. Determination of the nucleotide identity and amino acid similarity between two viral sequences was performed using the GCG GAP program. A gap weight of 5.0 and a gap length weight of 0.3 were used for calculating nucleotide identity, and a gap weight of 3.0 and gap length weight of 0.1 were used for amino acid comparisons. Evolutionary relationships between the conserved domains of different viruses were analyzed using version 3.4 of the Phylogeny Inference Package (Phylip; Felsenstein, 1989).

RESULTS

Cloning and sequencing of the LDV-C genome

The mapping of seven of the cDNA library clones to the 3′ terminus of the LDV-C genome was described previously (Godeny *et al.*, 1990). Five of these clones generated a consensus sequence of 1064 nucleotides that contained a terminal poly(A), the 3′ noncoding region and two complete, overlapping ORFs (Godeny *et al.*, 1989). The longest cloned 3′ poly(A) tract was 52 nucleotides in length. A poly(A) tract of about 50 nucleotides had previously been demonstrated by di-

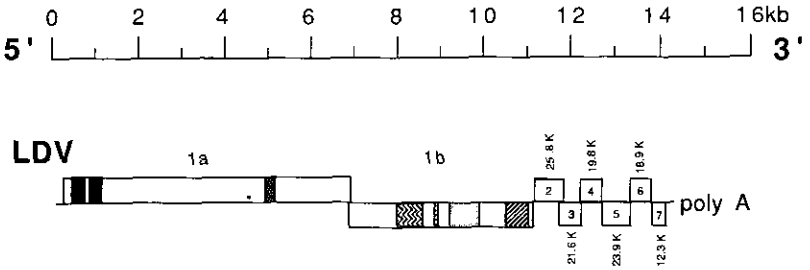

FIG. 1. Organization of the LDV-C genome. The 5′ portion contains two long ORFs (1a and 1b) with a frameshift region located at the junction between them. Six conserved domains in ORF 1a/ORF 1b are indicated by shaded boxes. From left to right these are: a papain-like cysteine protease (repeated), a trypsin-like serine protease, an RNA polymerase, a zinc finger, a helicase, and a domain identified in coronaviruses and toroviruses (Snijder *et al.*, 1990). The 3′ portion of the genome contains six overlapping ORFs. Adjacent ORFs in the 3′ region are in different frames. The estimated sizes of the proteins encoded by the 3′ ORFs are indicated.

rectly sequencing end-labeled LDV-C genome RNA (Brinton *et al.*, 1986).

Clones from the cDNA library, genome RNA, and virus-specific PCR products were sequenced to obtain a consensus sequence for the LDV-C genome RNA. This sequence, consisting of 14,222 nucleotides, has been submitted to the Genbank database (accession number L13298) and will not be duplicated here. The length of the consensus sequence agrees well with the size (14 kb) of the LDV genome previously calculated from gradient sedimentation experiments (Brinton-Darnell and Plagemann, 1975). The organization of the LDV-C genome is shown in Fig. 1. The LDV genome encodes two large 5′ ORFs, ORF 1a and ORF 1b, separated from each other by a frameshift junction region and six small, overlapping 3′ ORFs. Each 3′ ORF is in a different reading frame from its neighboring ORFs.

Even though the LDV-C genome RNA was not obtained from plaque-purified virus, it was possible to easily determine the majority nucleotide for almost all of the positions in the LDV-C sequence. However, for 12 nucleotide positions, 470, 681, 5160, 5192, 5198, 5253, 5870, 6888, 6955, 7015, 8983, and 9763, a single majority nucleotide could not be determined. These ambiguous nucleotides are indicated in the consensus sequence by the appropriate symbol specified by the GCG program. At all but two of these ambiguous positions, the nucleotide substitution did not result in an amino acid change. At position 5253, the amino acid encoded was either a serine or a proline, depending on whether a T or a C, respectively, was present in the cDNA sequence. At position 6888, the T to C shift changed the encoded amino acid from a serine to a phenylalanine.

The 5′ noncoding region sequence obtained thus far for LDV-C is 158 nucleotides in length. The consensus sequence generated from the cDNA library was extended in the 5′ direction by primer extension using the viral RNA as template. No sequence could be read beyond a strong stop located 19 nucleotides after the end of the consensus sequence. Since the 5′ noncoding regions obtained for EAV and LV are longer than that of LDV by 49 and 27 nucleotides, respectively, efforts are continuing to determine whether the strong stop observed at the 5′ terminus of the LDV-C sequence is due to the presence of the terminal cap structure or to a hairpin structure.

The LDV-C 3′ noncoding region is 80 nucleotides in length. For comparison, the 3′ noncoding region of EAV (den Boon *et al.*, 1991) is 59 nucleotides, that of LV (Meulenberg *et al.*, 1993) is 114 nucleotides, and that of simian hemorrhagic fever virus (SHFV) is 76 nucleotides in length (Godeny, unpublished data). The sequence identity for the entire 3′ noncoding region of LDV, EAV, LV, and SHFV ranges from 33 to 47%. The first nine 3′ nucleotides of the 3′ noncoding region showed the highest degree of conservation (Fig. 2). Of the four genomes compared, EAV was the most divergent. EAV differs in the first two nucleotides of the 3′ end and contains an extra nucleotide at the seventh position (Fig. 2). Conservation of the 3′ terminal nucleotides among these four viruses suggests that this region may function as a signal for viral RNA replication (Strauss and Strauss, 1983).

The LDV-C sequence was used to search the current sequence databases for related sequences with programs FASTA (Pearson and Lipman, 1988) and BLASTP (Altschul *et al.*, 1990). No significant similarities, beyond those with LV and EAV, that could indicate recombination between LDV-C and any currently available virus or cell gene sequence were detected.

Analysis of the ORF 1a/ORF 1b junction region

Within the LDV-C ORF 1a/ORF 1b junction region, a potential slippery sequence (5′-UUUAAAC-3′) is located just upstream from the ORF 1a stop codon and a

```
          5'                              3'
LDV       ATTTGGCTGGGCC.GGAATT-poly A
LV        GAACCATGTGACC.GAAATT-poly A
SHFV      ACTGGTATATACC.ATAATT-poly A
EAV       TTATTAGCCACCCAGGAACC-poly A
```

FIG. 2. Alignment of the first 20 nucleotides of the 3′ noncoding regions of the genomes of LDV, LV, SHFV, and EAV. Conserved nucleotides are underlined.

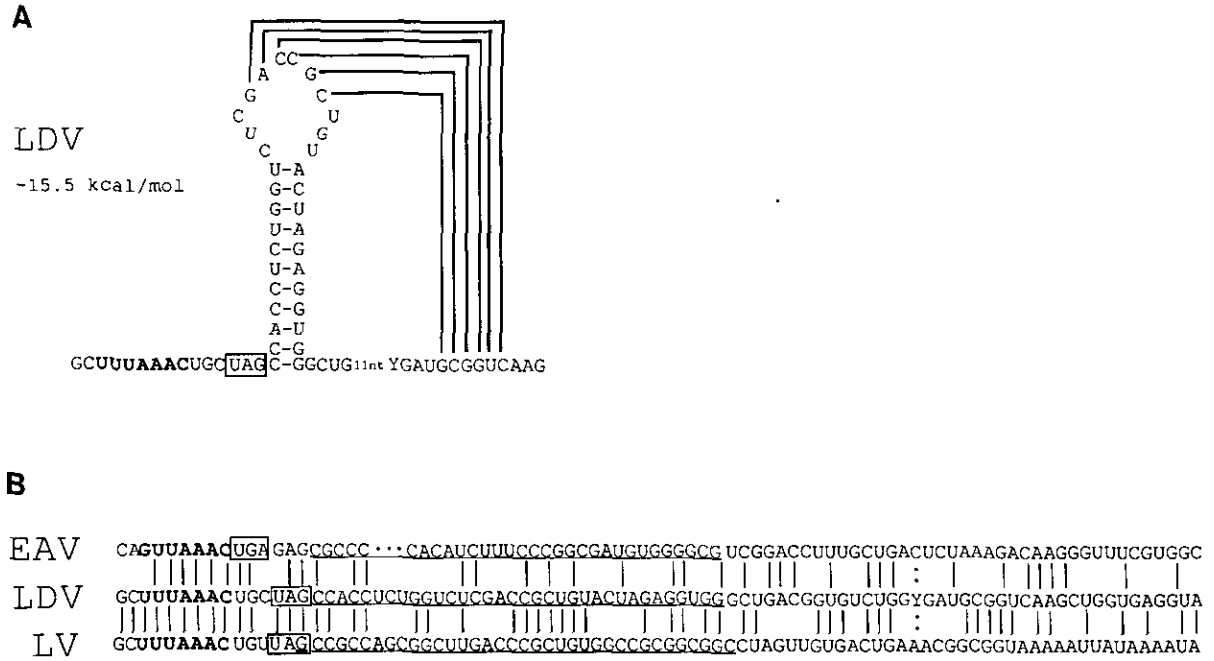

FIG. 3. The LDV-C frameshift region. (A) The slippery sequence is shown in bold letters. The ORF 1a stop codon is boxed. A putative pseudoknot can be folded from the sequence which follows the slippery sequence. (B) Alignment of the ORF 1a/1b junction (frameshift) regions of LDV-C, EAV, and LV. The sequence comprising the first stem loop in each sequence is underlined.

putative RNA pseudoknot is located immediately downstream of this stop codon (Fig. 3).

Comparison of the putative LDV slippery sequence and pseudoknot to those found in the coronaviruses (Brierley *et al.*, 1987, 1989; Lee *et al.*, 1991), toroviruses (Snijder *et al.*, 1990), EAV (den Boon *et al.*, 1991), and LV (Meulenberg *et al.*, 1993) genomes indicate that these structural elements have been highly conserved in each of these viruses. The coronavirus, torovirus, LDV, and LV "slippery" sequences (5′-UUUAAAC-3′) are identical and conform with the definition that "slippery" sequences consist of AAA, UUU, or GGG followed by UUUA, UUUU, or AAAC (Jacks *et al.*, 1988; ten Dam *et al.*, 1990). The EAV genome is the only one analyzed to date which contains a nucleotide substitution which causes it to differ from the consensus "slippery" sequences; however, the EAV junction region was demonstrated to be functional (den Boon *et al.*, 1991). In the genome of the coronavirus infectious bronchitis virus (IBV), both the "slippery" sequence and the RNA pseudoknot structure have been shown to be essential for efficient frameshifting at the ORF 1a/ORF 1b junction (Jacks *et al.*, 1988).

The particular stop codon used to terminate ORF 1a is not conserved among the viruses. The distance between the slippery sequence and the ORF 1a stop codon varies between the different viruses, with the distance being very short (0 to 4 nucleotides) in the genomes of LDV, LV, EAV, and a torovirus, Berne virus (BEV; Snijder *et al.*, 1990), and much longer in two coronaviruses, MHV and IBV (18 to 27 nucleotides).

Comparison of the pseudoknot structures of the various genomes indicates a general similarity in the size and stability of the first and second stems as well as in the length of the first loop (L1) of the pseudoknot structures. In contrast, the second loop (L2) varies considerably, with LDV, LV, and BEV having the shortest L2 loops (19, 14, and 11 nucleotides, respectively), EAV having a very long L2 loop (69 nucleotides), and MHV and IBV having intermediate-sized L2 loops (33 and 32 nucleotides, respectively). Alignment of the sequences of the frameshift regions of the different viruses indicated that LDV and EAV or LDV and LV show a lower degree of nucleotide identity with each other in this region than do the two coronaviruses MHV and IBV; the LDV and EAV sequences show a 45% nucleotide identity, the LDV and LV sequences show a 59% identity (Fig. 3), and the MHV and IBV show a 68% identity (data not shown).

Analysis of the LDV ORF 1a sequence

The LDV-C ORF 1a encodes a polyprotein of 242.8 kDa (Table 1). Alignment of the amino acid sequences of the ORF 1a products of LDV, EAV, and LV indicated *that conserved blocks of sequence are interspersed* with regions of little or no similarity (Fig. 4). The LDV ORF 1a is 499 amino acid residues longer than the EAV ORF 1a, and the LV sequence is 170 residues longer than that of LDV. The majority of these differences are accounted for by two large inserts in the LDV and LV

TABLE 1

PHYSICAL PROPERTIES OF THE PEPTIDES ENCODED BY THE LDV GENOME

Open reading frame (ORF)	Nucleotide location	Number of amino acids	Calculated size (kDa)	*pI*	Number of potential N-linked glycosylation sites	Comments
ORF 1a	162–6842	2226	242.8	7.8	7	replication proteins
ORF 1b	6839–11074	1411	155.4	7.4	2	replication proteins
ORF 2	11114–11797	227	25.8	10.5	4	envelope protein?
ORF 3	11665–12240	191	21.6	5.2	6	envelope protein?
ORF 4	12132–12659	175	19.3	8.1	5	envelope protein?
ORF 5	12611–13255	214	23.9	8.8	2	envelope protein?
	12656–13255	199	22.4	8.8	2	
ORF 6	13243–13757	171	18.9	10.2	0	envelope protein (Vp2)
ORF 7	13745–14092	115	12.3	11.4	3	capsid protein (Vp1)

sequences; however, EAV also has two sizable unique inserts (Fig. 4).

Papain-like cysteine protease domains have been identified in the N-terminal portion of the ORF 1a of two coronaviruses (Lee *et al.*, 1991; Gorbalenya *et al.*, 1991), EAV (den Boon *et al.*, 1991), and LV (Meulenberg *et al.*, 1993). The cysteine at position 164 and the histidine at position 230 have recently been identified as the probable catalytic residues of the EAV papain-like protease (Snijder *et al.*, 1992). Inspection of the ORF 1a alignment showed that the EAV cysteine protease domain has a counterpart in LDV and LV, with the cysteine at position 76 and the histidine at position 158 in the LDV sequence aligning with the predicted catalytic residues of EAV (Fig. 4). The unique insert downstream of the first papain-like protease domain in LDV and LV contains a second domain of this type. This observation makes a striking parallel with the situation among coronaviruses; MHV encodes two whereas IBV apparently encodes only one papain-like protease (Lee *et al.*, 1991; Gorbalenya *et al.*, 1991). Comparison of the amino acid sequences of eight coronavirus and arterivirus putative papain-like proteases revealed little overall conservation beyond the short stretches around the catalytic residues (Fig. 5A), which is in accord with a previous report on the variability in this type of protease among positive-strand RNA viruses (Gorbalenya *et al.*, 1991). Interestingly, LDV, EAV, and LV appear to lack the conserved domain of unknown function ("X domain") that flanks the papain-like protease domain in the polyproteins of all other animal positive-stand RNA viruses encoding this enzyme (Gorbalenya *et al.*, 1991; Koonin *et al.*, 1992).

The EAV cysteine protease has been shown to cleave the polyprotein of ORF 1a between residues 260 and 261 (YG↓G) generating a 30-kDa N-terminal peptide (Snijder *et al.*, 1992). This protease is also capable of cleaving at YG↓A as shown by site-directed mutagenesis of the EAV cleavage site (Snijder *et al.*, 1992). LV contains the sequence YGA (data not shown), which probably serves as an efficient cleavage site, in place of the YGG site of EAV. In contrast, in the LDV sequence the second Gly residue appears to be deleted, resulting in a YGY sequence (Fig. 4). The substitution of the bulky Tyr residue for Gly may have a significant effect on the accessibility of this cleavage site. Replacement of the second glycine by valine severely, but not completely, impaired cleavage by the EAV protease (Snijder *et al.*, 1992). It remains to be elucidated whether the cysteine protease of LDV cleaves at this site. If it does, a 43-kDa N-terminal polypeptide would be produced. *In vitro* translation of gradient-purified LDV-C genome RNA yielded, among others, a prominent virus-specific band of 41 kDa (data not shown); however, this protein has not yet been mapped on the genome. The processing of the LDV polyprotein may be more complicated than that of EAV because of the presence to two putative papain-like protease domains.

A chymotrypsin-like serine protease domain is located near the C-terminus of the LDV-C ORF 1a. This domain is also highly conserved in LV and in EAV (Fig. 4). The coronavirus ORF 1a contains a putative cysteine protease domain, which is similar to the 3C-proteases of picornaviruses, in the same relative position (Gorbalenya *et al.*, 1989c; Lee *et al.*, 1991). The picornavirus and coronavirus proteases belong to a specific family within the chymotrypsin-like class of proteases (Bazan and Fletterick, 1988; Gorbalenya *et al.*, 1989b). The sequence segments in the LDV, EAV, and LV protease domains surrounding the three (putative) catalytic residues as well as two short sequences implicated in substrate binding could be confidently aligned with those in the coronavirus proteases despite the replacement of the principle catalytic amino acid (Fig. 5B; den Boon *et al.*, 1991). However, the arterivirus proteases are more closely related to the serine 3C-like protease of sobemoviruses and luteoviruses (Gorbalenya *et al.*, 1988a; Bazan and Fletterick, 1990; A. E. Gorbalenya and E. V. Koonin, unpublished observa-

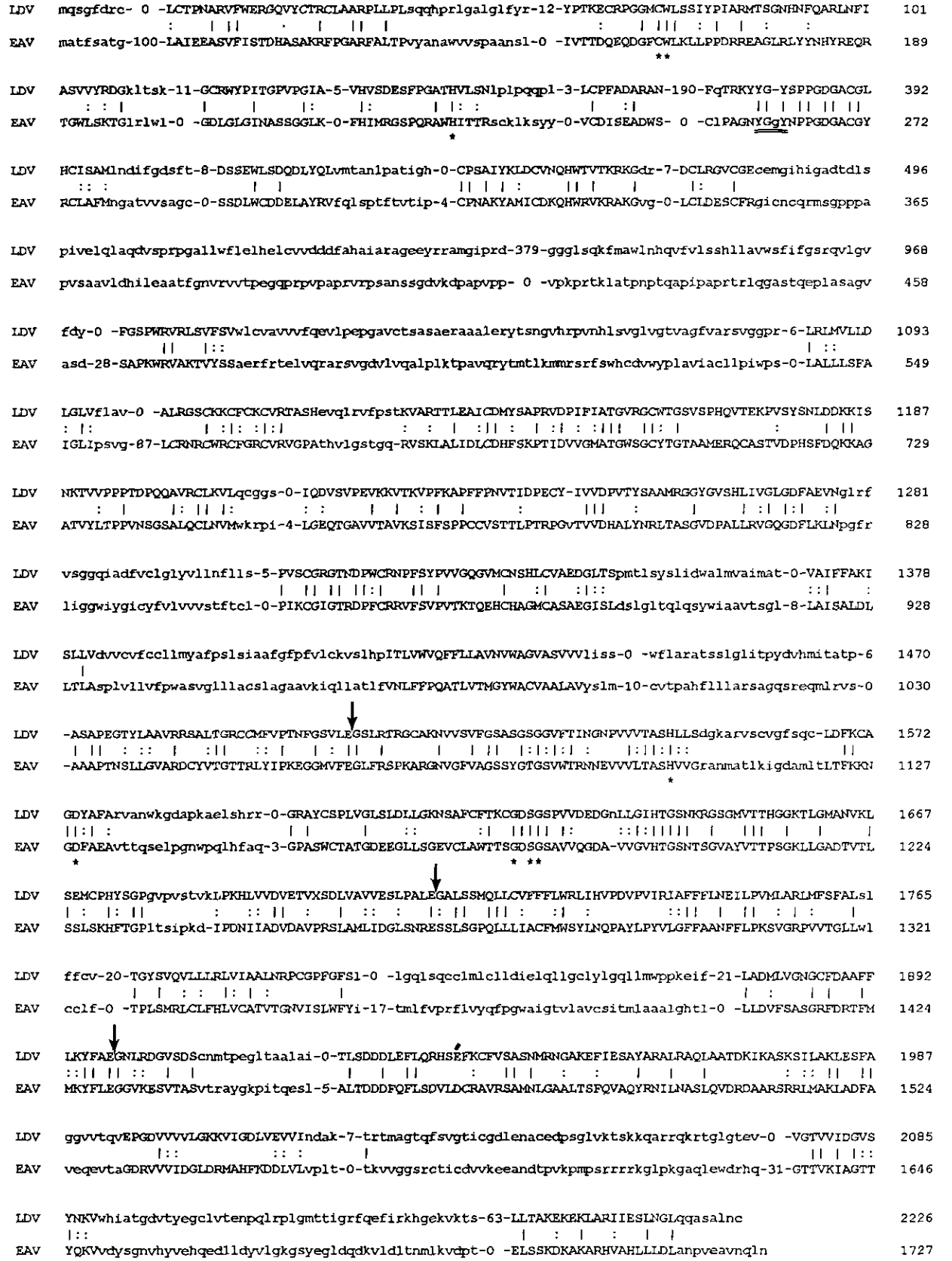

FIG. 4. Alignment of the ORF 1a polyprotein amino acid sequences of LDV and EAV. The alignment was generated using the MACAW software (Schuler *et al.*, 1991). Regions of significant similarity are shown in upper-case letters, and regions with little or no similarity are shown in lower-case. Inserts in either sequence are indicated by the number of amino-acid residues they contain. The EAV sequence is from den Boon *et al.* (1991) and the LV sequence is from Meulenberg *et al.* (1993). The LV sequence is not shown. Vertical lines indicate identical residues and colons indicate similar residues in all three sequences. Predicted catalytic and conserved residues of two putative protease domains are indicated by asterisks. The N-terminal domain is a papain-like cysteine protease, while the C-terminal domain is a chymotrypsin-like serine protease. The putative cleavage site of the cysteine protease is indicated by a double underline. Putative cleavage sites of the serine protease are indicated by arrows.

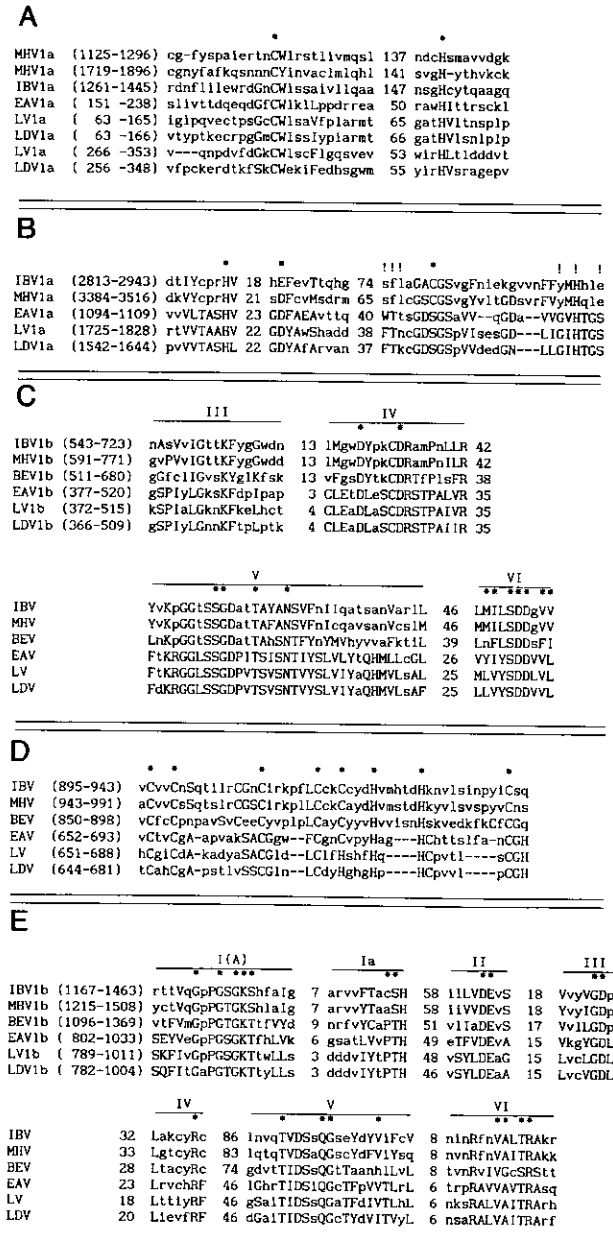

FIG. 5. Conserved domains in the nonstructural polyproteins of LDV, EAV, LV, coronaviruses and toroviruses. (A) Putative papain-like cysteine protease domains. Asterisks designate the putative catalytic residues. MHV, LV, and LDV each have two predicted proteases of this type, whereas IBV and EAV appear to have only one such domain. (B) Putative 3C-like serine protease domain. Asterisks designate the putative catalytic residues and exclamation marks designate the residues implicated in substrate binding (Gorbalenya *et al.*, 1989b). (C) RNA-dependent RNA polymerase domain. The conserved motifs are designated after Koonin (1991). Asterisks indicate the residues conserved in the majority of positive-strand RNA virus polymerases. (D) Putative Zn finger domain. Asterisks designate conserved Cys and His residues. (E) RNA helicase domain. The conserved motifs are designated after Gorbalenya *et al.* (1989a). Asterisks designate the residues conserved in the majority of (putative) helicases of superfamily I. (F) A conserved domain of unknown function specific for arteriviruses, coronaviruses, and toroviruses ("domain 4" in den Boon *et al.*, 1991). Asterisks designate residues conserved in all aligned sequences. The positions of the aligned regions in the respective polyproteins are given in parentheses and the distances between the aligned conserved segments are indicated. Residues that are conserved (identical or similar) in the three arterivirus sequences, and identical or similar to residues in the other virus sequences compared are indicated by upper-case letters. IBV, infectious bronchitis virus; MHV, mouse hepatitis virus; BEV, Berne virus.

tions). A more detailed analysis of the evolutionary relationships between the proteases of coronaviruses and arteriviruses will be published elsewhere.

Conserved domains in the ORF 1b

The LDV-C ORF 1b encodes a polyprotein of 155.4 kDa (Table 1). Four conserved domains were identified in this region (Figs. 1 and 5C–5F). Starting from the N-terminal side of the ORF, the first domain is the putative RNA-dependent RNA polymerase (Kamer and Argos, 1984; Koonin, 1991). A unique feature of the putative polymerases of coronaviruses, toroviruses, and arteriviruses is the substitution of serine for glycine in the GDD "signature sequence," which is invariant among other positive-strand RNA viruses (Gorbalenya *et al.*, 1989c; Koonin, 1991; Fig. 5C).

The second conserved domain is the putative Zn finger. Eight cysteine and histidine residues are conserved in this domain among coronaviruses, toroviruses, and arteriviruses (Fig. 5D; den Boon *et al.*, 1991). However, the pattern formed by these residues appears to allow the formation of only one finger instead of the two or even three postulated for coronaviruses (Gorbalenya *et al.*, 1989c; Lee *et al.*, 1991). Nevertheless, it is likely that this single conserved Zn finger may be involved in an interaction with RNA which is important for viral genome replication and/or transcription.

The third conserved domain is the putative RNA helicase (Figs. 1 and 5E). The coronavirus, torovirus, and arterivirus helicases belong to the so-called superfamily I that also includes the putative helicases of alpha-like viruses (Gorbalenya *et al.*, 1988b, 1989a; Hodgman, 1988; Snijder *et al.*, 1990; den Boon *et al.*, 1991). The location of the helicase on the C-terminal side of the polymerase (Fig. 1) represents a genome organization which, among positive-strand RNA viruses, is unique to coronaviruses, toroviruses, and arteriviruses

```
EAV1a  1060   GmVFE/GlFRS   N-terminal region/3C-like protease
LV1a   1689   GsLLE/GaFRT
LDV1a  1506   GsVLE/GsLRT

EAV1a  1264   LsNRE/SSLSG   3C-like protease/?
LV1a   1892   VpVVE/GGLST
LDV1a  1708   LpALE/GALSS

EAV1a  1426   KYFlE/GqVKe   ?/?
LV1a   2062   RYFaE/GnLRk
LDV1a  1894   KYFaE/GnLRd

EAV1b   198   eaVTD/GtnvI   ?/polymerase
LV1b    192   apVSD/GkstL
LDV1b   186   vpVSD/SetlV

EAV1b   626   aARtD/GVeFP   polymerase/helicase
LV1b    619   cARqD/GYsFP
LDV1b   613   cARkE/GFrFP

EAV1b  1075   lwSnE/Gleyy   helicase/"domain 4"
LV1b   1083   saSlE/Gscmp
LDV1b  1063   lmGlE/Gtasp
```

FIG. 6. Predicted cleavage sites for the 3C-like serine protease. The cleavage site position is indicated by a slash. The position of the N-terminal residue of each of the aligned sequences in its respective polyprotein is indicated. Residues conserved (identical or similar) in all three sequences are indicated by upper-case letters. The two mature proteins resulting from cleavage are indicated for each of the putative sites. Question marks designate putative proteins of unknown function.

(Spaan *et al.*, 1988; Gorbalenya *et al.*, 1989c; Snijder *et al.*, 1990; den Boon *et al.*, 1991).

Finally, the domain located nearest the C-terminus of the ORF 1b polyprotein is conserved only in coronaviruses, toroviruses and arteriviruses, but has not been found in other positive-strand RNA viruses (Spaan *et al.*, 1988; Snijder *et al.*, 1990; den Boon *et al.*, 1991; Fig. 5F). LDV and LV lack the counterpart to the C-terminal portion of this domain which is conserved in the coronaviruses and in BEV, and apparently also in EAV (den Boon *et al.*, 1991). No function for this domain has yet been determined.

Prediction of cleavage sites for the serine protease

The 3C-like proteases cleave predominantly at Q,E/G,S dipeptides (reviewed by Bazan and Fletterick, 1990). Although no experimental data for coronaviruses are yet available, cleavage sites have been predicted in coronavirus polyproteins on the basis of conservation of the respective dipeptides and the surrounding sequences (Lee *et al.*, 1991). Inspection of the alignments of the ORF 1a (Fig. 4) and ORF 1b (not shown) products of LDV, EAV, and LV allowed the prediction of three putative cleavage sites for the 3C-like serine protease in each of these polyproteins (Fig. 6). The sites in the ORF 1a polyprotein, particularly the two which flank the serine protease could be predicted with confidence since they are highly conserved in the three viruses and have the consensus structure E/G(S)*x*[VLF] (*x*, any residue; bulky hydrophobic residues which can occupy the +3 position are bracketed). The sites in the ORF 1b polyprotein are considered speculative since they are less conserved and some of them apparently contain the relatively unusual substitution of Asp for Glu in the −1 position (Fig. 6).

Cleavage of the ORF 1a/1b polyprotein at the predicted sites (Fig. 6) would result in seven mature proteins including one of 202 residues containing the LDV 3C-like protease, one of 427 residues containing the RNA-dependent RNA polymerase, and one of 450 residues containing the helicase. The sizes of these proteins are within the range of replicative enzymes produced from the polyproteins by other positive-strand RNA viruses. However, it is possible that the cleavage site upstream of the conserved polymerase domain may not be cleaved (Fig. 6). This would result in the production of a larger protein resembling the predicted coronavirus polymerase (Lee *et al.*, 1991). The cleavage pattern predicts that the Zn finger domain would comprise the N-terminal portion of the helicase protein rather than the C-terminal portion of the polymerase, which is similar to the situation in coronaviruses and toroviruses (Lee *et al.*, 1991; and E. V. Koonin, unpublished observations).

In vitro translation of the LDV-C genome RNA yielded major products of 41, 23, 22, and 19.6 kDa (data not shown). As mentioned in the previous section on ORF 1a, the largest *in vitro* translation product might correspond to the N-terminal leader peptide presumably cleaved by the papain-like protease; alternatively, it might represent the product generated from ORF 1b polyprotein cleavage which contains domain 4 (352 residues). Two of the smaller translation products might correspond to the 3C-like protease (202 residues) and the protein flanking its C-terminus (186 residues). The actual relationship between the *in vitro* translation products and the putative cleavage products remains to be elucidated.

The six 3′ ORFs

The LDV genome encodes six small 3′ ORFs (Fig. 1). ORFs 2, 3, 4, and 5 encode proteins with calculated sizes of 25.8, 21.6, 19.8, and 23.9 kDa (Table 1). The isoelectric points (p*I*s) of these four proteins vary considerably and each protein contains 2 or more potential N-linked glycosylation sites (Table 1). The proteins encoded by ORFs 2 through 5 each contain a putative N-terminal signal peptide sequence (data not shown). Hydropathic analyses of the LDV 3′ ORFs were done according to the method of Kyte and Doolittle (1982). ORFs 4 and 2 both contain hydrophobic C-terminal regions, while ORFs 3, 4, and 5 have one or more internal hydrophobic regions (data not shown). The general characteristics of these four proteins are consistent with those of virion envelope proteins. A recent report has identified the EAV ORF 5 product as the main envelope glycoprotein of EAV (de Vries *et al.*, 1992). Further work is necessary to determine whether the pro-

teins encoded by ORFs 2 through 4 represent additional virion envelope proteins.

Each of the LDV 3′ ORF proteins is translated from a separate subgenomic mRNA (Kuo *et al.*, 1991, 1992). The consensus intergenic sequence for LDV and LV is 5′-(U/C/A)-AACC-3′ (Plagemann and Moennig, 1992; Meulenberg *et al.*, 1993). In contrast, the EAV intergenic consensus sequence is 5′-UCAAC-3′ (den Boon *et al.*, 1991). The published LDV intergenic sequence was determined using the LDV-P (LDV-1) isolate (Plagemann and Moennig, 1992). The LDV-C genome contains a sequence similar to the LDV-P intergenic consensus sequence upstream of each of the 3′ ORFs and ORF 1a. The only possible intergenic region for ORF 2 is deleted at the second position of the consensus, 5′-U · ACC-3′. The genomic ORF 5 encodes a protein of 214 amino acids (Table 1). However, since the consensus intergenic sequence is located 3′ of the initiating AUG of this ORF, it seems likely that translation of the ORF 5 subgenomic mRNA initiates at the second AUG of the genomic ORF yielding a 199 amino acid protein (Table 1).

The product of ORF 6 has a predicted size of 18.9 kDa and contains no N-linked glycosylation sites (Fig. 1; Table 1). A short amino-acid sequence (5 amino acids) obtained by N-terminal sequencing of purified LDV-C Vp2, a nonglycosylated, virion envelope-associated protein (Brinton-Darnell and Plagemann, 1975), mapped to this ORF. This protein contains an N-terminal signal peptide sequence (data not shown) and internal hydrophobic regions. The EAV ORF 6 product is similar to that of LDV in having no N-linked glycosylation sites (den Boon *et al.*, 1991). In contrast, the ORF 6 product of LV contains two potential N-linked glycosylation sites (Meulenberg *et al.*, 1993).

The mapping of the capsid protein to ORF 7 was reported previously (Godeny *et al.*, 1990). Interestingly, three potential glycosylation sites were found at the C-terminus of the LDV ORF 7 sequence. During previous attempts to radioactively label the sugars of virion proteins (Brinton-Darnell and Plagemann, 1975), no evidence was obtained to indicate that the mature LDV capsid protein found in virions is glycosylated. These same three potential N-linked glycosylation sites are also present in the capsid protein sequence reported by Kuo *et al.* (1991) for the LDV-P isolate. One potential N-linked glycosylation site is present in the N-terminal portion of the LV capsid protein, but no glycosylation sites were observed in the EAV capsid protein sequence (Meulenberg *et al.*, 1993; den Boon *et al.*, 1991).

Comparison of the amino acid sequences of each of the LDV-C 3′ ORFs as well as each of the conserved ORF 1a/1b domains with those of EAV and LV indicated that these three viruses are closely related to each other and that among them LDV and LV show the highest degree of sequence similarity (data not shown). Although preliminary analysis of the SHFV 3′ ORFs and subgenomic mRNAs (Godeny, unpublished data) strongly suggests that SHFV is closely related to LDV, LV, and EAV, sufficient sequence data has not yet been obtained for SHFV to include it in the comparison shown in Fig. 7.

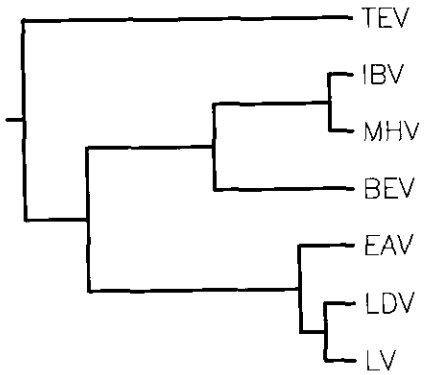

FIG. 7. Phylogenetic analysis of the RNA-dependent RNA polymerase domains of coronaviruses, toroviruses, and arteriviruses. The analysis was performed using an alignment spanning the eight conserved polymerase motifs delineated previously (Koonin, 1991). The tree was generated using the Fitch–Margoliash distance matrix algorithm (Fitch and Margoliash, 1967) implemented in the Fitch program of the PHYLIP package (Felsenstein, 1989). The root position was inferred by using the tobacco etch potyvirus (TEV) polymerase sequence as an outgroup. The branch lengths are proportional to the calculated evolutionary distances between the compared sequences.

Possible phylogenetic relationships between arteriviruses, coronaviruses, and toroviruses

It has been shown previously that the RNA-dependent RNA polymerases of coronaviruses and toroviruses constitute a distinct lineage within the tentative phylogenetic tree of positive-strand RNA virus polymerases (Koonin, 1991). Phylogenetic analysis of polymerase domains of two coronaviruses, of the single available torovirus sequence, and of LDV, EAV, and LV were performed. The potyvirus polymerase sequence, which shows the highest degree of similarity to the coronavirus sequences among those of other positive strand RNA viruses (Gorbalenya *et al.*, 1989c) was used as an outgroup to infer the root position. The tentative phylogenetic tree produced by the Fitch-Margoliash distance matrix algorithm obviously split into two distinct lineages, one leading to the arteriviruses, and the other to the coronaviruses and toroviruses (Fig. 7). Identical tree topology was produced by other algorithms used, namely the UPGMA clustering method, neighbor-joining, and protein parsimony. Also, very similar trees were generated for the helicase domain using the putative helicase of hepatitis E virus, which is most closely related to the arterivirus heli-

cases, as an outgroup (figure not shown). These observations indicate that at least the genes for two principal replication enzymes coevolved in this lineage of positive-strand RNA viruses.

DISCUSSION

Comparison of the genomes of LDV, EAV, and LV indicated that these viruses are closely related, but that they have diverged from each other by point mutation as well as through recombination. The lengths and properties of the individual 3′ ORF proteins differ to some degree among the three viruses. The ORF 1a regions of LDV, EAV, and LV also differ in length, with EAV being the shortest (1727 amino acids), LV the longest (2396 amino acids), and LDV intermediate in length (2226 amino acids). LDV and LV have two putative papain-like cysteine protease domains, while EAV has only one. EAV is also divergent from LDV and LV in its intergenic consensus sequence, its "slippery sequence" and the terminal nucleotides of the 3′ noncoding region. The distinction between LDV and LV, on the one hand, and EAV, on the other hand, was confirmed by the phylogenetic trees generated which suggest that EAV diverged from the common ancestor prior to LDV and LV (Fig. 7).

LDV, EAV, LV, and SHFV differ from the coronaviruses and toroviruses in several respects. Using LDV as an example for this group of viruses, the diameter of the LDV virion and the length of the LDV genome are about half those of the coronavirus and torovirus virions and genomes. The morphology of the LDV virion differs markedly from that of the coronaviruses. Coronaviruses gained their name from the large spikes which protrude from the virion surface. No spikes are visible on LDV virions (Brinton-Darnell and Plagemann, 1975). Instead, cup-like subunits 10 nm in diameter have been observed on the LDV surface. The nucleocapsid of LDV appears to be icosahedral (Brinton-Darnell and Plagemann, 1975), while those of the coronaviruses and toroviruses are helical. LDV encodes a serine 3C-like protease near the C-terminus of ORF 1a which is only distantly related to the cysteine 3C-like protease of the coronaviruses. The number and sizes of the LDV 3′ ORFs differ from those of the coronaviruses and show little sequence similarity with them.

Both phylogenetic analysis of individual conserved domains and comparison of genome organization strongly suggest that coronaviruses, toroviruses, and arteriviruses comprise a distinct evolutionary lineage among positive-strand RNA viruses. Unique features of this virus lineage are the location of the helicase domain on the C-terminal side of the polymerase and the presence of the conserved domain 4 in the ORF 1b polyprotein. A less significant but characteristic "birthmark" is the presence of SDD in place of the universal GDD signature in the RNA-dependent RNA polymerase domain.

It has been postulated previously that the coronaviruses, toroviruses, and EAV represent three separate lineages of divergence from a common ancestor (Spaan *et al.*, 1990). Phylogenetic analysis using the polymerase domains of these viruses (Fig. 7) indicated that two lineages diverged from a common ancestor, one generating a BEV-like virus and the other an EAV-like virus. Viruses in these two lineages have continued to diverge from each other by both mutation and recombination. LDV and LV diverged from the EAV-like progenitor, while the coronaviruses diverged from the BEV-like progenitor. Interestingly, in both branches the genomes grew longer as evolution continued. The variation in genome length is much greater in the BEV lineage than in the EAV lineage. The increase in genome length indicates that recombination has played a significant role in the evolution of these viruses.

Coronaviruses have been shown to undergo recombination by a copy-choice mechanism during which the polymerase jumps from one template to another resulting in continued extension of the nascent strand from the new template. Recombination events have been documented in the laboratory between two different coronaviruses co-infecting the same cell (Keck *et al.*, 1987; Makino *et al.*, 1986, 1987) and have also been shown to occur in animals (Keck *et al.*, 1988; Kusters *et al.*, 1989). Recombination events can result in the acquisition of new genes as well as in the loss of genes. Evidence for the insertion of genes from other viruses, such as influenza virus, into the coronavirus genome has been reported (Luytjes *et al.*, 1988). Recombination events are random and the majority of the "hybrid" viruses created by copy-choice recombination are not viable. Thus, virus viability provides a strong selective pressure for the retention of all genomic regions that are functionally critical for virus reproduction.

It is not known whether the common ancestor of the BEV and EAV lineages had a helical or an icosahedral nucleocapsid. The replacement of one or more of the structural proteins in the ancestral virus genome via a recombination event with another virus would have had a major impact on the further evolution of the recombinant viruses. Such a recombination event may well have been the initial event leading to the divergence of the coronavirus/torovirus branch from the ancestral virus. If the original prototype virus had an icosahedral capsid that was replaced through recombination by a helical nucleocapsid, the viral genome in the resulting recombinant virus would have been freed from previous packaging constraints so that it could continue to grow significantly in length during subsequent random recombination events. The coronavirus genome is the largest of the known positive-strand

RNA virus genomes and the genomic material located between the conserved domains in the ORF 1a/ORF 1b region shows little sequence similarity among the genomes of the various coronaviruses (Lee *et al.*, 1991). On the other hand, there is no known mechanism by which a longer ancestral genome in a helical nucleocapsid could have been deleted rapidly and at multiple positions to meet the packaging constraints of a newly acquired icosahedral capsid while still preserving all six of its conserved nonstructural protein domains. Using this line of reasoning, we hypothesize that the common ancestral virus of both the BEV-like and EAV-like lineages had an icosahedral nucleocapsid rather than a helical one and, therefore, was morphologically more similar to the EAV-like virus branch in genome size and virion morphology. The progression from shorter to longer genomes in both lineages is consistent with our hypothesis.

Viruses such as LDV, EAV, LV, and SHFV are very successful at surviving because of the persistent and nonlethal characteristics of the infections they cause in their natural hosts. It is likely that additional members of this virus group will be isolated from other host species in the future.

ACKNOWLEDGMENTS

This work was supported by Public Health Service Grant, NS19013 from NINCDS. We are grateful to Drs. J. J. M. Meulenberg and R. J. M. Moorman for providing us with a copy of the LV genome sequence prior to publication. E.V.K. acknowledges numerous helpful discussions with Alexander E. Gorbalenya. We thank Jerry Blackwell for technical assistance and Amy Black for typing the manuscript.

REFERENCES

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., and LIPMAN, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

BAZAN, J. F., and FLETTERICK, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: Structural and functional implications. *Proc. Natl. Acad. Sci. USA* **85,** 7872–7876.

BAZAN, J. F., and FLETTERICK, R. J. (1990). Structural and catalytic models of trypsin-like viral proteases. *Semin. Virol.* 1, 311–322.

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

BRIERLEY, I., DIGGARD, P., and INGLIS, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57,** 537–547.

BRINTON, M. A. (1982). Lactate dehydrogenase-elevating virus. *In* "The Mouse in Biomedical Research" (H. L. Foster, J. D. Small, and J. G. Fox, Eds.), Vol. 2, pp. 194–208. Academic Press, NY.

BRINTON, M. A., GAVIN, E. I., and FERNANDEZ, A. V. (1986). Genetic variation among strains of lactate dehydrogenase-elevating virus (LDV). *J. Gen. Virol.* **67,** 2673–2684.

BRINTON-DARNELL, M., and PLAGEMANN, P. G. W. (1975). Structure and chemical-physical characteristics of lactate dehydrogenase-elevating virus and its RNA. *J. Virol.* **16,** 420–433.

CLOSE, T. J., CHRISTMANN, J. L., and RODRIGUEZ, R. L. (1983). M13 bacteriophage and pUC plasmids containing DNA inserts but still capable of β-galactosidase α-complementation. *Gene* **23,** 131–136.

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

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

DE VRIES, A. A. F., CHIRNSIDE, E. D., HORZINEK, M. C., and ROTTIER, P. J. M. (1992). Structural proteins of equine arteritis virus. *J. Virol.* **66,** 6294–6303.

FELSENSTEIN, J. (1989). PHYLIP-Phylogeny Interference Package (version 3.2). *Cladistics* **5,** 164–166.

FENNER, F. (1977). Classification and nomenclature of viruses. *Intervirology* **7,** 44–47.

FITCH, W. M., and MARGOLIASH, E. (1967). Construction of phylogenetic trees. *Science* **155,** 279–284.

FRANCKI, R. I. B., FAUQUET, C. M., KNUDSON, D. L., and BROWN, F. (1991). Classification and nomenclature of viruses. Fifth report of the International committee on Taxonomy of Viruses. *Arch. Virol.* (Suppl. 2), 1–450.

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., WERNER, M. R., and BRINTON, M. A. (1989). The 3′ terminus of lactate dehydrogenase-elevating virus genome RNA does not contain togavirus or flavivirus conserved sequences. *Virology* **172,** 647–650.

GORBALENYA, A. E., BLINOV, V. M., DONCHENKO, A. P., and KOONIN, E. V. (1989a). An NTP-binding motif is the most conserved sequence in a highly diverged group of proteins involved in positive strand RNA viral replication. *J. Mol. Evol.* **28,** 256–268.

GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M., and KOONIN, E. V. (1989b). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common structural fold. *FEBS Lett.* **243,** 103–114.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1988a). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* **236,** 287–290.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1988b). A novel superfamily of nucleoside triphosphate-binding motif-containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* **239,** 16–24.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M. (1989c). Coronavirus genome: Prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **17,** 4847–4861.

GORBALENYA, A. E., KOONIN, E. V., and LAI, M. M.-C. (1991). Putative papain-related thiol proteases of positive strand RNA viruses. *FEBS Lett.* **288,** 201–205.

HODGEMAN, T. C. (1988). A new superfamily of replicative proteins. *Nature (London)* **333,** 22–23.

JACKS, T., MADHANI, D. M., MASIARZ, F. R., and VARMUS, H. E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag–pol region. *Cell* **55,** 449–458.

KAMER, G., and ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12,** 7269–7282.

KECK, J. G., MATSUSHIMA, G. K., MAKINO, S., FLEMING, J. O., VANNIER, D. M., STOHLMAN, S. A., and LAI, M. M. C. (1988). *In vivo* RNA–

RNA recombination of coronavirus in mouse brain. *J. Virol.* **62,** 1810–1813.

KECK, J. G., STOHLMAN, S. A., SOE, L. H., MAKINO, S., and LAI, M. M. C. (1987). Multiple recombination sites at the 5'-end of murine coronavirus RNA. *Virology* **156,** 331–341.

KOONIN, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive strand RNA viruses. *J. Gen. Virol.* **72,** 2197–2207.

KOONIN, E. V., GORBALENYA, A. E., PURDY, M. A., ROZANOV, M. N., REYES, G. R., and BRADLEY, D. W. (1992). Computer-assisted assignment of functional domains in the non-structural polyprotein of hepatitis E virus: Delineation of an additional group of positive strand RNA plant and animal viruses. *Proc. Natl. Acad. Sci. USA* **89,** 8259–8263.

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

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

KUSTERS, J. G., NIESTERS, H. G. M., LENSTRA, J. A., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1989). Phylogeny of antigenic variants of avian coronavirus IBV. *Virology* **169,** 217–221.

KYTE, J., and DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

LEE, H. J., SHIEH, C. K., GORBALENYA, A. E., KOONIN, E. V., LA MONICA, N., TULER, J., BAGDZHADZHYAN, A., and LAI, M. M. C. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative protease and RNA polymerase. *Virology* **180,** 567–582.

LUYTJES, W., BREDENBEEK, P. J., NOTEN, A. F., HORZINEK, M. C., and SPAAN, W. J. M. (1988). Sequence of mouse hepatitis virus A59 mRNA2: Indications for RNA recombination between coronavirus and influenza C virus. *Virology* **166,** 415–422.

MAKINO, S., FLEMING, J. O., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C. (1987). RNA recombination of coronaviruses: Localization of neutralizing epitopes and neuropathogenic determinants on the carboxyl terminus of peplomers. *Proc. Natl. Acad. Sci. USA* **84,** 8567–8571.

MAKINO, S., KECK, J. G., STOHLMANN, S. A., and LAI, M. M. (1986). High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57,** 729–737.

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MARTINEZ, D., BRINTON, M. A., TACHOVSKY, T. G., and PHELPS, A. H. (1980). Identification of lactate dehydrogenase-elevating virus as the etiological agent of genetically restricted, age-dependent polioencephalomyelitis of mice. *Infect. Immun.* **27,** 979–987.

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. (1993). Lelystad virus, the causative agent of poscine epidemic abortion and respiratory syndrome (PEARS) is related to LDV and EAV. *Virology* **192,** 62–72.

NAINAN, O. V., MARGOLIS, H. S., ROBERTSON, B. H., BALAYAN, M., and BRINTON, M. A. (1991). Sequence analysis of a new hepatitis A virus infecting cynomolgus macaques (*Macaca fascicularis*). *J. Gen. Virol.* **72,** 1685–1689.

NOTKINS, A. L. (1965). Lactic dehydrogenase virus. *Bacteriol. Rev.* **29,** 143–160.

OKAYAMA, H., and BERG, P. (1982). High efficiency cloning of full length cDNA. *Mol. Cell Biol.* **2,** 161–170.

PEARSON, W. R., and LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.

PLAGEMANN, P. G. W., and MOENNIG, 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.

RICE, C. M., LENCHES, E. M., EDDY, S. R., SHIN, S. J., SHEETS, R. L., and STRAUSS, J. M. (1985). Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* **229,** 726–735.

ROWSON, K. E. K., and MAHY, B. W. J. (1985). Lactate dehydrogenase-elevating virus. *J. Gen. Virol.* **66,** 2297–2312.

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

SCHULER, D., ALTSCHUL, F., and LIPMAN, J. (1991). A workbench for multiple alignment construction and analysis. *Proteins: Struct. Funct. Genet.* **9,** 180–190.

SMIT, M. J., DUUERSMA, A. M., KOUDSTAAL, J., HARDONK, M. J., and BOUMA, J. M. W. (1989). Infection of mice with lactate dehydrogenase-elevating virus destroys the subpopulation of Kupffer cells involved in receptor-mediated endocytosis of lactate dehydrogenase and other enzymes. *Hepatology* **12,** 1192–1199.

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

SNIJDER, E. J., WASSENAAR, A. L. M., and SPAAN, W. J. M. (1992). The 5' end of the equine arteritis virus replicase gene encodes a papain-like cysteine protease. *J. Virol.* **66,** 7040–7048.

SOUTHERN, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98,** 503–517.

SPAAN, W. J. M., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69,** 2939–2952.

SPAAN, W. J. M., DEN BOON, J. A., BREDENBEEK, P. J., CHIRNSIDE, E. D., NOTEN, A. F. H., SNIJDER, E. J., DE VRIES, A. A. F., and HORZINEK, M. C. (1990). Comparative and evolutionary aspects of coronaviral arteriviral, and toroviral genome structure and expression. *In* "New Aspects of Positive Strand RNA Viruses" (M. A. Brinton and F. X. Heinz, Eds.), pp. 12–19. ASM Press, Washington DC.

STRAUSS, E. G., and STRAUSS, J. H. (1983). Replication strategies of the single-stranded RNA viruses of eukaryotes. *Curr. Top. Microbiol. Immunol.* **105,** 1–98.

TEN DAM, E. B., PLEIJ, C. W. A., and BOSCH, L. (1990). RNA pseudoknots: Translational frameshifting and read through on viral RNAs. *Virus Genes* **4,** 121–136.

WINSHIP, P. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulfoxide. *Nucleic Acids Res.* **17,** 1266.