Molecular Analysis of Three Ljungan Virus Isolates Reveals a New, Close-to-Root Lineage of the Picornviridae with a Cluster of Two Unrelated 2A Proteins†

Susanne Johansson,1 Bo Niklasson,2 Jacob Maizel,3 Alexander E. Gorbalenya,4* and A. Michael Lindberg1*

Department of Chemistry and Biomedical Sciences, University of Kalmar, S-391 82 Kalmar,1 and Nya Apodemus, S-114 53 Stockholm,2 Sweden, and Laboratory of Experimental and Computational Biology, National Cancer Institute,3 and Advanced Biomedical Computing Center, SAIC/National Cancer Institute-Frederick Cancer Research and Development Center,4 Frederick, Maryland 21702

Received 27 December 2001/Accepted 3 June 2002

Ljungan virus (LV) is a suspected human pathogen recently isolated from bank voles (Clethrionomys glareolus). In the present study, it is revealed through comparative sequence analysis that three newly determined Swedish LV genomes are closely related and possess a deviant picornavirus-like organization: 5' untranslated region-VP0-VP3-VP1-2A1-2A2-2B-2C-3A-3B-3C-3D-3' untranslated region. The LV genomes and the polyproteins encoded by them exhibit several exceptional features, such as the absence of a predicted maturation cleavage of VP0, a conserved sequence determinant in VP0 that is typically found in VP1 of other picornviruses, and a cluster of two unrelated 2A proteins. The 2A1 protein is related to the 2A protein of cardio-, erbo-, tescho-, and aphthoviruses, and the 2A2 protein is related to the 2A protein of parechoviruses, kobuviruses, and avian encephalomyelitis virus. The unprecedented association of two structurally different 2A proteins is a feature never previously observed among picornviruses and implies that their functions are not mutually exclusive. Secondary polyprotein processing of the LV polyprotein is mediated by protease 3C (3Cpro) possessing canonical affinity to Glu and Gln at the P1 position and small amino acid residues at the P1' position. In addition, LV 3Cpro appears to have unique substrate specificity to Asn, Gln, and Asp and to bulky hydrophobic residues at the P2 and P4 positions, respectively. Phylogenetic analysis suggests that LVs form a separate division, which, together with the Parechovirus genus, has branched off the picornavirus tree most closely to its root. The presence of two 2A proteins indicates that some contemporary picornviruses with a single 2A may have evolved from the ancestral multi-2A picornavirus.

Picornviruses are small positive-stranded RNA viruses with a genome comprising approximately 7,500 to 8,200 nucleotides (nt) which is packed into a nonenveloped icosahedral capsid (51). The polyadenylated genome of most picornviruses contains a single expressed open reading frame (ORF) flanked by 5' and 3' untranslated regions (UTRs). The 5' UTR, comprising approximately 600 to 1,300 nt, contains replicative and translational regulator signals and is folded in a complex tertiary structure. Located in the 3' half of the 5' UTR is an internal ribosome entry site (IRES) crucial for the cap-independent initiation of translation. Three types of IRESs are recognized on the basis of structure and are conserved in different branches of picornviruses (62). The ORF encodes a large protein precursor (polyprotein) whose domain backbone contains the following organization: NH2-L-VP0-VP3-VP1-2A1-2A2-2B-2C-3A-3B-3C-3D-3'COOH, where VP0, VP3, and VP1 are paralogous proteins forming the capsid, with all other non-structural proteins being primarily involved in the picornnaviral replicative process. Among the nonstructural proteins, the leader (L) protein has been identified in some but not all picornviruses.

The picornavirus polyprotein is autocatalytically processed at the conserved interdomain junctions by a proteolytic activity associated with the 3C moiety, which, depending on the individual picornavirus, may also be assisted by the (proteolytic) activities of L and/or 2A proteins with different specificities (14, 31, 43, 54). Additional cleavages of the polyprotein at a few alternative sites may take place, resulting in new products and some intermediate precursors, some of which are stable and/or functionally active. In most picornviruses, VP0 is autocatalytically cleaved further into VP4 and VP2 proteins during the final stage of virion maturation. Both the L and 2A proteins have been described as having four apparently unrelated structural forms (10, 24, 54, 64, 65), and this diversity sets them apart from all other proteins conserved across the entire picornavirus family. The conserved proteins include the multifunctional 2C ATPase (2CATPase), the main cytochrome C protease (3Cpro), 3D RNA-dependent RNA polymerase (3Dpol), membrane-associated 2B and 3A proteins, and a small 3B protein (3BVPg) (51). 3BVPg serves as a primer for the RNA synthesis mediated by 3Dpol with the involvement of other nonstructural proteins and remains co-
valently linked to the 5’ end of plus- and minus-strand RNAs (46).

Picornaviruses infect mammals, including humans, and birds (28). Picornavirus-like viruses that infect invertebrates have also been identified (7). Depending on the nature of the individual picornavirus, the infection may cause severe ailments of the gastrointestinal tract and the respiratory, neural, hepato-cellular, and cardiomuscular systems (23, 42). Likewise, the host range, progeny yield, and reproductive cycle mechanisms differ dramatically among picornaviruses. This phenotypic diversity of picornaviruses is ultimately linked to the plasticity of the picornavirus genome.

The family Picornaviridae was originally classified into four genera based on the antigenic and biophysical properties of the virions (35). Subsequent molecular analysis of the viral genomes supported this classification for the majority of picornaviruses. Such characterizations also led to the creation of two additional genera, Hepatovirus and Parechovirus, comprising two former enteroviruses which have diverged substantially from other picornaviruses (28). In order to accommodate the molecular diversity of newly recognized picornaviruses, three other genera have recently been introduced (49). The constant increase in the number of picornavirus genera, the current bias towards picornaviruses that infect humans and a few other species, and the huge disparity in the numbers of viruses assigned to different genera (28) indicate that the diversity of Picornaviridae is far from being fully described.

During a search for an infectious agent linked to myocarditis in humans, a new virus, Ljungan virus (LV), was recently isolated from bank voles (Clethrionomys glareolus) trapped near the Ljungan River in Medelpad County, Sweden (38, 39). The partial sequences of the LV prototype strain, 87-012, and two other serologically related LV isolates, 174F and 145SL, suggested that LV is most closely related to human parechoviruses (HPEVs) (39). In the present study, bioinformatic techniques were employed to analyze the newly determined, nearly complete viral genomes of three LV isolates. Evidence was obtained showing that LV is a prototype of a new picornavirus lineage clustering with, but distinct from, parechoviruses. The organization of the LV genome, a tentative polyprotein map of processing signals recognized by 3Cpro with a unique specificity, the unusual sequence characteristics of capsid proteins, and the unprecedented association of two different 2A proteins are also described. Due to their possession of these unique features, it is proposed that LVs be classified into a separate genus of the Picornaviridae.

MATERIALS AND METHODS

Viruses and cells. LV was isolated from bank voles in BHK-21 cells as previously described (39). The 145SL and 174F isolates were passaged four times in A549 cells and then once in Vero cells, whereas strain 87-012 was passaged six times in Vero cells, causing a mild and delayed cytopathic effect (CPE).

Purification of LV RNA and cDNA synthesis. Viral RNA was extracted from Vero cell culture-propagated LV stocks (clarified by slow-speed centrifugation) with Ultraspec II (Biotec Laboratories, Inc., Houston, Tex.). The RNA was reverse transcribed into cDNA by using the RNase H-deficient reverse transcriptase Superscript II (Life Technologies, Taunton, Sweden) and the primer Noti25 (5′-ATAGAACATGGCGGCTGCGC-3′) at 45°C for 2 h at 5°C as previously described (33). After cDNA synthesis was completed, the RNA template was hydrolyzed by the addition of NaOH to a final concentration of 0.2 M and subsequent incubation at 37°C for 15 min. The samples were then neutralized with sodium acetate at a final concentration of 0.4 M, and single-stranded cDNA was precipitated with ethanol, dissolved in Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA [pH 7.5]) and stored at −20°C before amplification. For reverse transcription-PCR, several different types of primers were initially used in this study. The oligo(dT)-rich Noti25 primer was used for both oligo(dT)-primed reverse transcription and for amplification of the 3′ end of the genome by use of the 3′ RACE (rapid amplification of cDNA ends) technique (3). For amplification of the 5′ end of the genome, two HPEV-based primers covering the 17 or 25 5′-proximate nt (LV5nc17; 5′-GATCTTAATATTG-3′) and LV5nc25 (5′-GCTTTAATATTG-3′) were designed to previously described nucleotide sequences being derived from the parechovirus nucleotide sequence) were used. A number of primers were derived from previously published partial sequences of LV, which include a 2,118-nucleotide segment that covers part of the 5′ UTR, VP0, and VP3 of strain 145SL and a 264-nucleotide segment that covers part of the 5′ UTR of strains 87-012 and 174F (39). Primer design was also based on comparisons with HPEV sequences listed in the GenBank database.

DNA sequencing. The nucleotide sequences of the LV genomes were determined with purified PCR products (QiAquick gel extraction kit; Qiagen GmbH) by using AmpliTaq DNA polymerase (Perkin-Elmer-Cetus) and cycle sequencing ready reaction kit according to the manufacturer’s instructions (PE Biosystems, Stockholm, Sweden). Each nucleotide was determined at least twice in each direction, except for the 3′ ends of the genomes, which were sequenced multiple times in the same direction. To avoid the possible inclusion of mutations generated by the amplification process, the final genome sequences were derived from sequences determined for at least two batches of viral RNAs extracted and amplified independently. Sequence data were recorded with an ABI Prism 310 genetic analyzer (PE Biosystems) and assembled and edited by using Sequencer 3.0 software (Gene Codes, Ann Arbor, Mich.).

Computer-aided comparative sequence analysis. Sequence alignments were generated by using the Clustal W program, a windows-based interface of the Clustal W program (59, 60). The alignments were manually edited and all positions containing a gap(s) were removed by using the Data Analysis in Molecular Biology and Evolution (DAMBE) package (version 4.0.30; X. Xia, Department of Ecology and Biodiversity, University of Hong Kong). To investigate whether a data set contained phylogenetic signals corresponding to tree-, star-, or net-like evolution, an likelihood mapping analysis of the aligned sequences was performed (57). Phylogenetic tree reconstructions were conducted by using the neighbor-joining method using quartet puzzling as implemented in Tree-Puzzle version 5.0 software (58) and the Proml program of an alpha release of the PHYLIP version 3.6a2 package (J. Felsenstein, Department of Genetics, University of Washington, Seattle). The significance of the inferred phylogeny was evaluated by bootstrap analysis with 100 pseudoreplicates as implemented in the PHYLIP SeqBoot program (16). The resulting trees were visualized by using the Treeview program (41). The RNA secondary structures were predicted by using the RNAGene consensus program (5) and the MFOLD 3.1 program (67).

Nucleotide and protein sequences. The genome sequences, and the polyprotein derived from them, that were used for comparisons in this study were from the following picornaviruses (with their GenBank accession numbers): Aphthovirus genus, foot-and-mouth disease virus (FMDV) (M110975) and equine rhinitis A virus (ERAV) (L43052); Cardioivirus genus, encephalomyocarditis virus (EMCV) (M22457) and Thiel’s murine encephalomyelitis virus (TMEV) (M23030); Enterovirus genus, poliovirus type 1 strain Sabin (PV1S) (V01150) and A-2 plaque virus (A2pV) (AAF585765); Erbovirus genus, equine rhinovirus B virus (ERBH) (X96871); Hepatovirus genus, hepatitis A virus (HAV) (M14707 and M59810) and avian encephalomyelitis virus (AEV) (AJ225173); Kobuvirus genus, Aichi virus (AIV) (AB010145); Parechovirus genus, HPEV1 strain Harris (HPEV1H) (S45504), HPEV2 strain Williamson (HPEV2W) (AJ005695), and HPEV4 strain CT6676-67 (HPEV4) (A055846); Picornavirus genus, human rhinovirus 2 (HRV2) (X02316); and Tospovirus genus, porcine teschovirus 1 (PTV1) (AJ011380). The protein sequences of two insect viruses, infectious flacherie virus (InFV) (AB000906) and saclbrood virus (SBV) (AF092924), that are distantly related to picornaviruses were also used as out-groups in the phylogenetic analysis.

Downloaded from http://jvi.asm.org/ on October 30, 2017 by guest
FIG. 1. Predicted stem-loop structures of the LV 5’ UTR. Stem-loop structures in the 5’ UTR were predicted for three LV isolates as described in Materials and Methods. Illustrated is the predicted organization of the 5’ UTR of strain LV 87-012 drawn with the XRNA program (B. Wasei and H. F. Noller, unpublished data). The secondary-structure elements recognized in the 5’ UTRs harboring group II IRESs are labeled. Also highlighted are various conserved sequence elements discussed in the text. Stem-loop structures A through H were derived from the RNAGA consensus program (5) which compared LV strains and three parechoviruses. The folding of the I and J-K stem-loop regions was supported by the consensus analysis and was predicted using the MFOLD program (67). Replacements in strains 145SL and 174F are indicated in boldface uppercase and lowercase letters, respectively, and insertions and deletions are marked with asterisks and arrows, respectively. Base pairings predicted for LV 87-012 are indicated with continuous lines, and base pairings that are possible in either of the two other isolates are indicated with broken lines. The initiator codon is indicated with a shaded box.
Nucleotide sequence accession numbers. The genome sequences of LV strains 87-012, 174F, and 145SL described in this study have been submitted to GenBank and have been assigned accession no. AF327920, AF327921, and AF327922, respectively.

RESULTS AND DISCUSSION

Sequencing genomes of three LV isolates. To determine the genomic sequence of LV, three field isolates were propagated through several different cell cultures (see Materials and Methods). LV replication in Vero cells induced a delayed and less pronounced CPE than that normally facilitated by many enteroviruses (our unpublished data). Despite serial passages in several cell lines, no evidence of adaptation of LV was observed within 2 weeks. The genomic LV RNA was isolated from infected cells and used to determine the nucleotide sequence from overlapping PCR-generated amplicons. Amplifications of the extreme 5'/H11032 UTR by use of different protocols of the RACE technique (15) were not successful, possibly due to a stable secondary structure of the 5'/H11032 terminal region, as has previously been suggested for parechoviruses (17) (Fig. 1). The most 5'-terminal LV sequences, which have been deposited in GenBank (17 nt for the prototype strain 87-012 and 25 nt for strains 174F and 145SL), remain of parechovirus origin and should be verified in the near future. Based on the results of the primer extension and Southern blot analyses (data not shown), the length of the amplified LV genomes reported here must be close or identical to the length of wild-type LV genomes. The LV 145SL sequence reported here and a fragment of it published previously (39) differ in four positions (A157G, T733C, T1894C, and G2267C, numbered according to our nucleotide sequence, AF327921), the last three of which are nonsynonymous. The newly described genomic sequences were determined from PCR-generated amplicons to ensure that they correspond to the dominant genotypes of the semiadapted LV present during propagation in Vero cells. In contrast, previously reported LV sequences were determined from PCR-generated individual clones (39) and may deviate from the consensus genotype. Indeed, at position 733, a nucleotide determined in the previous study (39) destroyed the initiator codon predicted here, which in the present analysis resides in an optimal Kozak context (29).

The GC content of the LV genomes is 42%, which is similar to those of rhino- (39% for HRV2), hepato- (38% for HAV), and enteroviruses (38% for TMEV, 38% for PV1, 38% for ERBV, 38% for ERBOV, 38% for HEV, 38% for AEV, 38% for AIV, 38% for HEV, 38% for HRV2, 38% for PTV1).

### TABLE 1. Features of proteins of LV strains 145SL and 174F and the prototype strain 87-012a

<table>
<thead>
<tr>
<th>Protein</th>
<th>No. of amino acids in protein of strain:</th>
<th>Amino acid positions for the prototype protein</th>
<th>Predicted activity or function of viral protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP0</td>
<td>259</td>
<td>1Met–Gln259</td>
<td>Capsid protein</td>
</tr>
<tr>
<td>VP3</td>
<td>244</td>
<td>260Gly–Glu503</td>
<td>Capsid protein</td>
</tr>
<tr>
<td>VP1</td>
<td>297</td>
<td>504Gly–Glu800</td>
<td>Capsid protein</td>
</tr>
<tr>
<td>2A1</td>
<td>20</td>
<td>801Met–Gly820</td>
<td>Primary polyprotein processing</td>
</tr>
<tr>
<td>2A2</td>
<td>135</td>
<td>821Pro–Glu955</td>
<td>Inhibition of cell growth</td>
</tr>
<tr>
<td>2B</td>
<td>140</td>
<td>956Ser–Glu1093</td>
<td>Membrane associated</td>
</tr>
<tr>
<td>2C</td>
<td>334</td>
<td>1094Gly–Asn1426</td>
<td>ATPase</td>
</tr>
<tr>
<td>3A</td>
<td>130</td>
<td>1427Glu–Glu1556</td>
<td>Membrane associated</td>
</tr>
<tr>
<td>3B</td>
<td>29</td>
<td>1557Arg–Glu1585</td>
<td>Initiation of RNA synthesisb</td>
</tr>
<tr>
<td>3C</td>
<td>198</td>
<td>1586Ala–Gln1783</td>
<td>Protease</td>
</tr>
<tr>
<td>3D</td>
<td>470</td>
<td>1784Gly–Asp2253</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
</tbody>
</table>

a The polyproteins of strains 145SL, 174F, and 87-012 contain 2,252, 2,253, and 2,253 amino acids in all, respectively.

b Protein 3B1 (VPg).

### TABLE 2. Comparison of the LV prototype strain 87-012 with selected picornavirusesa

<table>
<thead>
<tr>
<th>Region</th>
<th>Aphthovirus</th>
<th>Cardiovirus</th>
<th>Enterovirus</th>
<th>Erbovirus</th>
<th>Hepatovirus</th>
<th>Kobuvirus</th>
<th>Parechovirus</th>
<th>Rhinovirus</th>
<th>Teschovirus</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMDV</td>
<td>46</td>
<td>35</td>
<td>48</td>
<td>46</td>
<td>36</td>
<td>48</td>
<td>39</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>ERV</td>
<td>23</td>
<td>24</td>
<td>29</td>
<td>23</td>
<td>17</td>
<td>26</td>
<td>24</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>EMCV</td>
<td>22</td>
<td>25</td>
<td>25</td>
<td>27</td>
<td>19</td>
<td>22</td>
<td>26</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>TMEV</td>
<td>17</td>
<td>22</td>
<td>17</td>
<td>24</td>
<td>19</td>
<td>22</td>
<td>31</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>PV1</td>
<td>2A1</td>
<td>44</td>
<td>50</td>
<td>41</td>
<td>50</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>VP0</td>
<td>2A2</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>VP3</td>
<td>3B</td>
<td>28</td>
<td>21</td>
<td>26</td>
<td>27</td>
<td>15</td>
<td>27</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>VP1</td>
<td>3A</td>
<td>28</td>
<td>27</td>
<td>19</td>
<td>24</td>
<td>9</td>
<td>24</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>2A1</td>
<td>3B</td>
<td>50</td>
<td>42</td>
<td>40</td>
<td>17</td>
<td>23</td>
<td>25</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td>2A2</td>
<td>3C</td>
<td>25</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>23</td>
<td>21</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>ND</td>
<td>3D</td>
<td>26</td>
<td>26</td>
<td>29</td>
<td>31</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>ND</td>
<td>3' UTR</td>
<td>28</td>
<td>46</td>
<td>29</td>
<td>35</td>
<td>47</td>
<td>41</td>
<td>52</td>
<td>44</td>
</tr>
</tbody>
</table>

a Values are percent amino acid identities and percent nucleotide identities (for the 5' and 3' UTRs) between 87-012 and the selected picornaviruses. Values were calculated by using the GCG GAP program. ND, not determined.

b It should be noted that the true length of the extreme 5' end has been estimated by Southern blots only.

c FMDV protein 3B1 (VPg1) is used for comparison.
and parechoviruses (39% for HPEV1). The major features of the polyproteins encoded by the genomes of the three LVs are summarized in Table 1. Excluding the poly(A) tract, the LV 87-012, 174F, and 145SL genomes comprise 7,606, 7,608, and 7,609 nt, respectively. The 5'/H11032 UTRs of the LV 87-012, 174F, and 145SL genomes consist of 733, 735, and 731 nt, respectively, and are followed by single ORFs encoding polyproteins containing 2,253, 2,253, and 2,256 amino acids (aa), respectively, while the 3'/H11032 UTRs possess 111, 111, and 107 nt, respectively, and poly(A) tails. The genomes and polyproteins of the LV isolates have 80 to 93% identical nt and 89 to 99% identical amino acid residues, respectively. Previously, based on their antigenic properties, LV 87-012 and 174F had been separated from 145SL (39); the genomic sequence analysis reported here confirms this grouping. A detailed domain comparison of LV with representatives of all genera of the Picornaviridae confirmed the clustering of LV with parechoviruses (Table 2).

FIG. 2. Comparison of the terminal regions and the predicted cleavage sites of the polyproteins of LV strains 87-012, 174F, and 145SL with those of parechoviruses (HPEV1H, HPEV2W, and HPEV2C). *, identical residues; : and ·, strong and weak conserved groups, respectively, of amino acid residues as defined by the Clustal X program (59, 60). The RGD motif in the C-terminal end of VP1 in HPEV is highlighted in the smaller gray box and is located in the last conserved region between HPEV and LV in VP1. The conserved DvExNPG/P motif putatively responsible for the LV 2A1 C-terminal polyprotein processing is highlighted in the larger gray box. Two of the previously predicted cleavage sites for parechoviruses, 2A/H20841 and 3B/H20841, are revised here based on the now available LV amino acid sequences. The numbers of amino acids separating the amino acids shown in detail around the proposed scissile bonds are indicated in the alignments.
the apical 14-nucleotide loop of the J domain. The 5' half of the 5' UTR of LVs may also contain stem-loop structures found in other group II picornaviruses, although no equivalent to the element C was identified by our analysis. However, at the sequence level, conservation is limited, with a pentapeptide AAUAA present in a loop of the stem-loop D being the longest region conserved between LVs and parechoviruses. Stem-loop structures A, F, I, and J are supported by compensatory mutations found in the three isolates of LV, while the element D may vary in LVs (Fig. 1).

3' UTR. The 3' UTRs of picornaviruses vary in length and are organized in a higher-order RNA structure thought to be involved in RNA replication (19). The two most closely related LV strains, 87-012 and 174F, both have a 111-nucleotide-long 3' UTR, while strain 145SL has a 4-nucleotide deletion in this region, corresponding to nt 7538 to 7541 in the prototype strain. This region seems to be unique for LVs, although it is predicted to fold into two hairpins (data not shown) structurally similar to those identified in the 3' UTR of poliovirus and EMCV (1, 9, 48).

LV polyprotein processing. The polyprotein interdomain junctions processed by 3Cpro in all picornaviruses are confined to small regions that contain primary- and tertiary-structure elements recognized by 3Cpro (14, 31, 43, 54). An alignment of the polyproteins of LVs and their closest relatives, parechoviruses, was generated and analyzed for the presence of conserved 3Cpro sequence substrate determinants at interdomain junctions to predict the 3Cpro-mediated processing map of LV polyproteins (Fig. 2). It was noted that two previously determined sites, VP0/VP3 and VP3/VP1, and four predicted sites, 2B/2C, 2D/3B, and 3C/3D, of parechoviruses (17, 25, 40) are also conserved in LVs (Fig. 2). In contrast, the previously predicted 3Cpro sites at the A2/A and B3B junctions in parechoviruses were not conserved in LVs. However, two other dipeptides in these regions, which are conserved in LVs and parechoviruses and resemble 3C cleavage sites elsewhere, could actually be processed in these viruses (Fig. 2).

Like 3Cpro's of other picornaviruses, the LV main protease is predicted to prefer Gln and Glu at the P1 position and a small amino acid residue at the P2 position, although the latter requirement seems not to be stringent (Fig. 3A). The 3Cpro's of parechoviruses also have a similar albeit relaxed specificity with respect to the P1 position (Fig. 3B). These proteases may also have affinity to structurally similar Asn, Gln, and Asp at the P2 position. This specificity is unique among 3Cpro's of picornaviruses, although 3C-like proteases of nido- and potyviruses and 2A proteases of entero- and rhinoviruses possess other restricted specificities at the same position (14, 54, 66). When the putative 3Cpro cleavage sites of LVs and parechoviruses were analyzed together, the affinities of the 3Cpro's to the bulky hydrophobic amino acid residues at the P4 position became evident (Fig. 3C). This P4 substrate position is also recognized by other 3Cpro's, e.g., poliovirus 3Cpro is known for its preference for Ala and Thr residues at this position (4). Comparison of LV- and parechovirus-specific logos (Fig. 3A and B, respectively) suggested that the 3Cpro of parechoviruses, but not of LVs, may also recognize the P2' substrate position.

Molecular modeling of polyprotein processing at the LV VP1/2A1 border presented a special challenge. In HPEVs, this
junction is likely to be processed by 3Cpro (Fig. 2). The 3C cleavage site separating VP1 and 2A in HPEVs resides in an ~20-aa sequence corresponding to a larger region in LVs (70 aa) lacking reliable similarity (Fig. 2 and data not shown). A separate analysis of this region of LVs revealed that it, surprisingly, contains a homologue of a 17-aa peptide conserved at the 2A/2B junction in cardio-, erbo-, tescho-, and aphthoviruses. The C terminus of this peptide homologue conforms to the signature DvExNPG, with the uppercase letters denoting absolutely conserved residues (Fig. 4A) (2, 8, 10, 32, 47, 52, 63, 65). This peptide is sufficient to promote separation of the 2A and 2B moieties at the NPGP junction in mengovirus and FMDV (11, 12, 21, 45) by a mechanism that is currently the subject of intensive debate (13, 20). In aphthoviruses and probably in erbo- and teschoviruses, this conserved peptide, excluding the most C-terminal Pro residue, is released as a mature 17- to 20-aa-long product through cleavage by 3C at the N terminus and autoprocessing at the C terminus (11). A similar 20-aa-long peptide could be produced in LVs assuming that 3Cpro cleaves at the N-terminal HSDE site (amino acid position 800) that conforms to the pattern conserved in other 3Cpro sites of LVs (Fig. 2 and 3A). This model implies that, unlike with all other picornaviruses, including parechoviruses, the LV polyprotein is processed at an extra site in a region between VP1 and 2B which generates two 2A proteins (2A1 and 2A2) rather than one. If LVs actually deviate further from picornaviruses, their 3Cpro might cleave another site (e.g., the FLNQ/C site, 11 aa further downstream [Fig. 2]) that would make the regulation of the expression of this region even more complex.

The comparisons of polyprotein sequences of LVs and other picornaviruses do not provide positive evidence for the presence of an L peptide at the N terminus or cleavage of VP0 into VP4 and VP2. Overall, the LV polyprotein appears to be processed into 11 mature products by two different proteolytic activities (Fig. 2).

**Nonstructural proteins.** LV polyproteins possess a full complement of nonstructural proteins identified previously in parechoviruses (17, 25, 40). With the exception of 2A moieties, all nonstructural proteins are conserved across picornaviruses; however, they are larger in LVs than in parechoviruses, mostly because of extra sequences at one or both termini (Fig. 2). Conserved sequence motifs characteristic of picornavirus 2CATPase, 3BVPg, 3Cpro, and 3DPol proteins (18) have been identified in the LV homologues without substitutions known to compromise their respective functions (data not shown). Hence, the core enzymatic mechanisms of replication and expression employed by picornaviruses appear to be conserved in LVs. Likewise, the stringent sequence conservation that is suggestive of functional competence is evident for the 2A moieties. The N-terminal 2A1 is related to the NPGP family of 2As

![Table A](https://example.com/tableA.png)

![Table B](https://example.com/tableB.png)

**FIG. 4.** Multiple sequence alignments of 2A1 and 2A2 of LVs and selected picornaviruses. Black and gray backgrounds highlight alignment columns with 100 and 60% conserved residues, respectively, as defined in the GeneDoc default similarity groups (37). (A) Alignment of the LV 2A1 and the C-terminal regions of the picornavirus NPGP 2A protein family. (B) Alignment of the LV 2A2 and the picornavirus H-NC 2A protein family. The C-terminal border of parechovirus 2As is depicted as it was predicted (Fig. 2). The actual boundaries of 2A proteins from AIV and AEV were not defined. The conserved H-box and NC-box and a putative transmembrane domain made up of an approximately 20-aa-long hydrophobic region are labeled. *., each noneven count of 10.
encoded by cardio-, erbo-, tescho-, and aphthoviruses (Fig. 4A) (2, 8, 10, 32, 47, 52, 63, 65), and the C-terminal 2A2 belongs to the H-NC family of 2As encoded by parechoviruses, kobuviruses, and AEVs (Fig. 4B) (25, 34, 64). It is predicted (see above) that 2A1 mediates the separation between 2A1 and 2A2. The function of the picornavirus H-NC 2A2 is yet to be established, but they, as well as the 2A2 of LVs, may be involved in an aspect of virus-host interaction related to cell growth control (24). An unprecedented association of two structurally different 2As may also result in a unique regulation of the life cycle of LVs. Future comparative studies of parechoviruses and LVs should be especially useful in helping us to understand the unique benefits to a picornavirus associated with a cluster of two 2A proteins.

Capsid proteins. Alignment of the capsid proteins of LVs and other picornaviruses revealed the conservation of the major elements forming an eight-stranded antiparallel β-barrel structure (jelly roll) in VP0, VP3, and VP1 (53). The most pronounced divergence between LVs and parechoviruses has occurred outside the β-strand elements at the terminal regions that have accepted large insertions and deletions. These include the N termini of the capsid proteins and the C terminus of VP1 (Fig. 2). Compared to parechoviruses, LVs have a 30-aa-shorter N terminus of VP0 that does not contain the myristylation signal GXXX(S/T) (6), which is typical of the majority of picornaviruses. The latter feature puts them in one group with Parechovirus and Hepatovirus, whose virion morphogenesis may not rely upon the myristylation of VP0.

Unlike other picornaviruses, LVs and parechoviruses contain 18-aa-long and 24- to 30-aa-long extensions, respectively, that are enriched with basic residues at the N terminus of VP3 (Fig. 2). This region is immunogenic in parechoviruses (27) but is yet to be characterized in LVs. Furthermore, a functionally uncharacterized 10- to 11-aa insertion was found in the vicinity of the N terminus of VP1 in the LV isolates, while there was a major difference between parechoviruses and LVs at the C terminus of VP1 (Fig. 2). In this region, the parechoviruses contain an Arg-Gly-Asp (RGD) motif important for cell surface interactions (50, 56, 61), while LVs contain a unique 43-aa extension (Fig. 2).

Picornaviruses have evolved three paralogous copies of the jelly roll domain to take advantage of the hetero-oligomer multisubunit organization for virion-associated activities. This evolution was accompanied by a substantial divergence of VP0, VP3, and VP1 as each protein assumed specific functions (53). Surprisingly, LVs and parechoviruses possess atypically high sequence similarity between the paralogous C-terminal regions of VP0 and VP1 encompassing β-strands I (Fig. 5). This conserved region is flanked from the C terminus by a tetrapeptide CysProArgPro in VP0 of LVs. Remarkably, the same peptide is also conserved in the equivalent position of paralogous VP1 of a majority of picornaviruses (Fig. 5) and is considered the most conserved sequence of this protein (44; A. E. Gorbale-nya, unpublished observations). These observations indicate that some conserved function(s) may be performed by different capsid proteins in LVs and other picornaviruses.

Evolution and taxonomy of LVs. LVs and parechoviruses cluster together over the entire polyprotein (Table 2). To define the position of LVs among picornaviruses more precisely, a phylogenetic analysis of the 3Dpol proteins of LV and representative sequences in the vicinity of the β-strands I of VP0 and VP1 split into four virus-protein groups separated by substantial gaps. The top group comprises VP0 of all picornavirus genera except those of LVs and parechoviruses, which form the second group; the bottom group comprises VP1 of all picornavirus genera except those of LVs and parechoviruses, which form the third group. The amino acid conservation is highlighted separately for the top, bottom, and two middle groups. Black, dark gray (with white letters) and light gray (with black letters) backgrounds highlight alignment columns with 100, 80, and 60% conserved residues, respectively, as defined in the GeneDoc default similarity groups (37). The CysProArgPro (CPRP) tetrapeptide is indicated by italic type. Arrows indicate the position of the scissile bond at the VP0-VP3 junction.

![FIG. 5. Conservation between the C-terminal regions of VP0 and VP1 of picornaviruses. Multiple alignments of VP0 and VP1 of a set of picornaviruses representing all genera and LVs were generated as described in Materials and Methods. Presented is an alignment of VP0 and VP1 of a set of picornaviruses representing all genera and LVs were generated as described in Materials and Methods.](http://jvi.asm.org)
It is informative to compare the tree in Fig. 6 with a distribution of different molecular forms of 2A among picornaviruses. In this context, it becomes apparent that contemporary picornaviruses employing either the NPGP or H-NC types of 2A may have originated from the ancestral picornavirus having the LV-like organization of the 2A locus through selective and repeated loss of one of the 2As. Alternatively, the immediate ancestor of LV could have acquired the 2A1 gene from an outside source, although no specific sequence affinity was found between the 2A1 of LVs and any homologue of picornavirus or other origin (11) (data not shown). The generation of the entire molecular diversity of 2A(-like) proteins in picornaviruses and related viruses must have also involved other mechanisms in addition to that proposed above.

Regardless of whether LVs have inherited a cluster of two 2As from either a common picornavirus or an immediate LV ancestor, it is evident for the first time that, functionally, the two structural forms of 2A are not mutually exclusive. This compatibility of different 2As contributes to the evolutionary potential of picornaviruses (19), whose correct assessment may be vital for efforts to control picornavirus infections and could be used in the generation of new virus vectors.

Should a new LV-prototyped genus of the *Picornaviridae* be created? We are inclined to suggest so given the unique features of the genetic organization of LVs that correlate well with the substantial divergence of LVs from other picornaviruses. If the LV-based genus is established, it would join the majority of picornavirus genera (excluding hepatoviruses), which are characterized by a uniform combination of molecular forms of the L and 2A proteins (Fig. 6).

Overall, the analysis of the newly determined LV genomes presented here demonstrates that the basic picornavirus genetic plan has sufficient plasticity to evolve a cluster of two structurally different 2A proteins. The expression of the LV genomes is predicted to involve 3Cpro with a restricted specificity, which uniquely includes the P2 substrate position. Future studies should determine how these characteristics are translated into the biology of LV. The first step in this direction is to develop a cell culture system to support efficient multicycle production of LV progeny.

**FIG. 6.** Phylogenetic relationships among picornaviruses. A phylogenetic tree of picornavirus 3D pol protein sequences (left) representing all genera and LVs was generated by using the Fossil program in the PHYML 3.6a2 package as described in Materials and Methods. Prior to the tree inference, a likelihood mapping analysis of the 3D pol alignment showed that the phylogenetic signal could be inferred by using bifurcating trees (data not shown). The tree was rooted by using two 3D-like sequences encoded by InFV and SBV as the out-group. All bifurcations supported in more than 70 of 100 bootstrap trials are labeled. Bar, 0.1 substitutions per site. The LV division is highlighted in the gray oval. The picornavirus genera are listed to the right of the tree. To illustrate the diversity of picornaviruses outside 3D pol, structural variants of 2A and L proteins were arbitrarily assigned unique numbers, and these are listed along with their respective virus genera (right). The number 0 indicates the lack of L protein. Note that AEV and HAV of hepatoviruses encode different combinations of L and 2A proteins.
REFERENCES


