The Acheta domesticus Densovirus, Isolated from the European House Cricket, Has Evolved an Expression Strategy Unique among Paroviruses\textsuperscript{7,†}

Kaiyu Liu,\textsuperscript{1,2}§ Yi Li,\textsuperscript{1,2}§ Françoise-Xavière Jousset,\textsuperscript{3}§ Zoltan Zadori,\textsuperscript{1,4} Jozsef Szélei,\textsuperscript{1} Qian Yu,\textsuperscript{1} Hanh Thi Pham,\textsuperscript{1} François Lépine,\textsuperscript{1} Max Bergoin,\textsuperscript{1,3} and Peter Tijssen\textsuperscript{1*}

INRS-Institut Armand Frappier, Université du Québec, Laval, Quebec, Canada\textsuperscript{1}; Central China Normal University, Wuhan 430079, People’s Republic of China\textsuperscript{2}; and Laboratoire de Pathologie Comparée, Université Montpellier 2, Montpellier 34095, France\textsuperscript{3}

Received 29 March 2011/Accepted 12 July 2011

The Acheta domesticus densovirus (AdDNV), isolated from crickets, has been endemic in Europe for at least 35 years. Severe epizootics have also been observed in American commercial rearings since 2009 and 2010. The AdDNV genome was cloned and sequenced for this study. The transcription map showed that splicing occurred in both the nonstructural (NS) and capsid protein (VP) multicistronic RNAs. The splicing pattern of NS mRNA predicted 3 nonstructural proteins (NS1 [576 codons], NS2 [286 codons], and NS3 [213 codons]). The VP gene cassette contained two VP open reading frames (ORFs), of 597 (ORF-A) and 268 (ORF-B) codons. The VP2 sequence was shown by N-terminal Edman degradation and mass spectrometry to correspond with ORF-A. Mass spectrometry, sequencing, and Western blotting of baculovirus-expressed VPs versus native structural proteins demonstrated that the VP1 structural protein was generated by joining ORF-A and -B via splicing (splice II), eliminating the N terminus of VP2. This splice resulted in a nested set of VP1 (816 codons), VP3 (467 codons), and VP4 (429 codons) structural proteins. In contrast, the two splices within ORF-B (Ia and Ib) removed the donor site of intron II and resulted in VP2, VP3, and VP4 expression. ORF-B may also code for several nonstructural proteins, of 268, 233, and 158 codons. The small ORF-B contains the coding sequence for a phospholipase A2 motif found in VP1, which was shown previously to be critical for cellular uptake of the virus. These splicing features are unique among paroviruses and define a new genus of ambisense densovirus.

Insect paroviruses (densovirus) belong to the Densovirinae subfamily of the Paroviridae and are small, isometric, nonenveloped viruses (diameter, \textasciitilde25 nm) that contain a linear single-stranded DNA of 4 to 6 kb (2, 3, 27). These viruses can be subdivided into two large groups, those with ambisense genomes and those with monosense genomes. Like vertebrate paroviruses, all densovirus have a genomic DNA with hairpins at both ends, often (but not necessarily for all genera) as inverted terminal repeats (ITRs). All densovirus with ambisense genomes package both complementary strands in equimolecular ratios as single strands in separate capsids (27). The nonstructural (NS) gene cassette is found in the 5’ half of one genome strand, and the structural protein (VP) gene cassette is found in the 5’ half of the complementary strand. By convention, the genome is oriented so that the NS cassette is found in the left half. Expression strategies of densovirus often involve (alternative) splicing and leaky scanning translation mechanisms (28). So far, the near-atomic structures of three densovirus, Penaeus stylirostris densovirus (PstDNV), Bombyl mori densovirus 1 (BmDNSV-1), and Galleria mellonella densovirus (GmDNV), have been solved (10, 11, 21). The capsid of densovirus consists of 60 subunits (T=1) of identical proteins that may contain N-terminal extensions not involved in capsid formation but that confer additional functions to the capsid. One of these functions is a phospholipase A2 (PLA2) activity that is required for genome delivery during infection (34). Densovirus are usually highly pathogenic for their natural hosts (5).

The monosense densovirus have been classified into 3 uniform genera, i.e., Iteravirus, with a 5.0-kb genome, 0.25-kb ITRs, and a PLA2 motif in VP; Brevidensovirus, with a 4.0-kb genome, no ITRs but terminal hairpins, and no PLA2 motif; and Hepanvirus, with a single member, hepatopancreatic parovirus, with a 6.3-kb genome also lacking a PLA2 motif and ITRs but with 0.2-kb terminal hairpins (23, 27). In contrast, the ambisense densovirus have been classified into one uniform genus, Densovirus, with a 6-kb genome and 0.55-kb ITRs, and a second genus, Pefudensovirus, with only Periplaneta fuliginosa densovirus (PfdDV) as a member, with a 5.5-kb genome, 0.2-kb ITRs, and a split VP gene cassette (2, 26). Ribosome frameshifts have been proposed to connect its VP open reading frames (ORFs) (33). So far, all ambisense densovirus have an N-terminal PLA2 motif in their largest VP. Some sequenced ambisense densovirus, e.g., Myzus persicae densovirus (MpDNV) (32), Blattella germanica densovirus (BgDNV) (18), and Planococcus citri densovirus (PcDNV) (25), are as yet unclassified. The ambisense virus Culex pipiens densovirus (CpDNV) has a different genome organization for both the NS
and VP proteins and will have to be classified in a different genus (1).

_Acheta domestica _densovirus (AdDNV) was isolated from diseased _Acheta domestica _L, house crickets from a Swiss commercial mass rearing facility (16). The virus spread rapidly and was responsible for high mortality rates, such that the rearing could not be saved. This was the first observation of a densovirus in an orthopteran species. Infected tissues included adipose tissue, the midgut, the hypodermis, and particularly the Malpighi tubules, but the most obvious pathological change was the completely empty digestive cecae (24). The cecae, which flank the proventriculus, are the sites where most enzymes are released and most absorption of nutrients occurs. Feulgen-positive masses were observed in the nuclei of infected cells (16). Commercial production facilities for the pet industry or for research mass rearings of house crickets in Europe are frequently affected by this pathogen. This virus was previously not known to circulate in North America, except for a small epidemic in the Southern United States in the 1980s (22). Beginning in 2009, sudden, severe outbreaks were observed in commercial facilities in Canada and the United States, leading to losses of hundreds of millions of dollars and to an acute crisis in the pet food industry (24). In this study on AdDNV, we observed that over the last 34 years the annual replacement rate was about 2.45 × 10⁻⁴ substitution/site/nt and that the VP gene cassette consists of two ORFs, a characteristic of the _Pheidole _densovirus genus (24).

In the present study, the complete genome and the expression strategy of AdDNV were investigated and showed features not yet described for other densoviruses or vertebrate paroviruses. The most striking observation was the intricate splicing pattern of its VP ORFs, resulting, in contrast to the case for all paroviruses studied so far, in unrelated N-terminal extensions of its two largest structural proteins and in the probable production of several supplementary NS proteins from the VP cassette.

**MATERIALS AND METHODS**

**Rearing of crickets.** _A. domestica _L, house crickets were obtained from a commercial supplier and were reared under conditions of 65% relative humidity, 25°C, and a 16-h-8-h light-dark cycle. Diet conditions and drinking water supply, as well as conditions for perch, hiding, and oviposition, were as described previously (24).

**Infection techniques.** The visceral cavity of nymphs of about 1.5 to 2 cm was injected with an inoculum consisting of a viral suspension obtained by grinding infected cricket in 1× phosphate-buffered saline (PBS) plus 2% ascorbic acid, clarifying the mixture by centrifugation for 10 min at 8,000 × g, and filtering it through 450-nm membranes. Mortality was usually 100% within 2 weeks. Alternatively, infection was achieved by feeding with a virus-contaminated diet as previously described (24).

**Virus and DNA preparation.** Virus was purified as previously described (29). Lysis buffer [500 μl of 6 M guanidine-HCl, 10 mM urea, 10 mM Tris-HCl, 20% Triton X-100, pH 4.4, containing 80 μM poly(A) carrier RNA] and 200 μl sample were mixed and incubated for 10 min at 70°C. The sample was vortexed after adding 125 μl isopropanol, and the DNA was then purified on High-Pure plasmid spin columns (Roche Molecular Biochemicals) according to the supplier's instructions.

**Cloning, mutation analysis, and sequencing of viral DNA.** The 1977 isolate of AdDNV was cloned into the pCR-XL-TOPO vector (Invitrogen Life Sciences), using supercompetent Sure 2 Escherichia coli cells (Stratagene) at 30°C. Point mutations in the AdDNV genome were generated with a QuikChange site-directed mutagenesis kit (Stratagene), whereas deletion mutants were obtained via the type IIb restriction endonuclease strategy (7). Independent clones were sequenced in both directions by primer walking. The terminal hairpins yielded compressions that did not affect sequence; however, inclusion of 1 M betaine (Sigma) and 3% dimethyl sulfoxide (DMSO) or restriction in the hairpin by DraI yielded clean sequence reads. DNAs from subsequent isolates were amplified by PCR and sequenced between the ITRs.

**Isolation and characterization of viral RNA.** Total RNAs were isolated from 30 mg adipose tissue from infected cricket larvae (2 to 5 days postinfection [p.i.]) and from recombinant baculovirus-infected cells at 48 h.p.i. by use of an RNaseasy minikit from Qiagen. The DNase I treatment was extended from 15 to 30 min or repeated twice. A PCR test was included to verify the absence of DNA. Total extracted RNA was subjected to mRNA purification using an mRNA isolation kit (Roche).

**Northern blots.** About 20 to 30 μg total RNA in a 6-μl volume was added to 18 μl buffer (1× MOPS [20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, 0.5 mM EDTA adjusted to pH 8 with NaOH], 18.5% formamide, 50% formamide), 5 μl loading buffer was added, and the mixture was incubated for 5 to 10 min at 65 to 70°C and separated by electrophoresis on a 1% formaldehyde-agarose gel. Parallel lanes contained RNA size markers (Promega). After migration and washing, RNAs were transferred to positively charged nylon membranes (Roche) by capillary blotting overnight. The blotted membranes were prehybridized with 10 mg/ml herring sperm DNA in 50% formamide before hybridization with 3²P-labeled probes. The probes corresponded to a 1.5-kb BamHI-DraII restriction fragment specific for the VP coding sequence and a 6.8-kb EcoRI-HindIII DraI restriction fragment specific for NS. Hybridized probes were visualized with a Storm 840 phosphorimager.

**Mapping of 5’ ends, 3’ ends, and introns of viral transcripts.** The most probable locations of the transcripts were predicted from the ORFs obtained by sequence analysis. A 3’ rapid amplification of cDNA ends (3’ RACE) system was used to characterize the 3’ ends of the polyadenylated viral transcripts using the RNAtag and ADAP primers (Table 1) and PCR (28), whereas the 5’ ends were determined with a FirstChoice RLM RACE kit (Ambion) according to the instructions of the supplier. The locations of introns were determined after reverse transcription of the transcripts by use of avian myeloblastosis virus (AMV) reverse transcriptase (Promega) in a final volume of 20 μl for 1 h at 42°C, PCR using internal DNV-specific primers for overlapping regions, andideoxy sequencing of the amplicons according to standard methods (28).

**Promoter activity in AdDNV genome.** Promoter regions were amplified by PCR and cloned upstream of the luciferase gene in the pG3-basic system. The ProNSF and ProNSR primers were used for the NS promoter, the ProVP1F and ProVP1R primers were used for the _Myrmicina loreyi _densovirus (MDNV) NS promoter (control) (Table 1). Sequencing was performed to confirm the promoter direction. For the assay, Lb652 cells were seeded into wells of 24-well cell culture plates. Each well contained about 0.5 ml of cells at 5 × 10⁵ cells/ml. The cells were cultured overnight. Transfection was performed with 2.5 μl DOTAP reagent and 0.6 μg DNA in 15 μl HEPES, and the mixture was added to 245 μl medium (without antibiotic or fetal bovine serum [FBS]) per well. Cells were harvested at 48 or 60 h posttransfection, washed twice with PBS, and resuspended in 100 μl of Bright-Glo lysis buffer (Promega). Cell lysates were quickly centrifuged to remove cell debris, and 25-μl aliquots of the liquid were injected into the luciferase activity assay according to the instructions for a luciferase assay system (Promega).

**Expression of structural proteins and analysis of VP ORFs by use of a baculovirus system.** The potential VP coding sequences (see below) were cloned into the _Autographica california _nuclear polyhedrosis virus (AcNPV) downstream of the polyhedrin promoter by use of the Bac-To-Bac baculovirus expression system (14) (Invitrogen) via the pFastBac1 and pFastBacHT vectors according to the supplier’s instructions. In constructs involving expression of VP1, the initiation codon had to be moved closer to the start of the transcript. For this purpose, an EcoRI site was introduced 100 bp upstream of the multiple cloning site (MCS), using the pUC7 and pUC7VIR mutation primers (Table 1), followed by removal of the small EcoRI fragment between the new and MCS EcoRI sites. Inserts were generated by PCR (28) with the primers given in Table 1, using the wild-type (wt) template or a template in which intron II (see below) splicing sites had been mutated. The forward primer with an EcoRI site was either AdATG1B, which coincided with the initiation codon of VP1, or an equivalent in which the initiation codon ATG was mutated to ACC (AdATG1B), and the reverse primer AdIHar, containing an Xbal site, was used (Table 1). All pFastBac recombinant constructs were verified by sequencing.

**Protein analysis by SDS-PAGE.** Western blotting and _N-acetylated _amino acid sequencing. Capsid proteins were analyzed by SDS-PAGE (13), using the structural proteins of _Janninia coenia _densovirus (JcDVN) or broad-range standards (Bio-Rad) as size markers. Expressed proteins were analyzed by Western blotting (28, 30), using polyvinylidene difluoride (PVDF) membranes and Roche blocking reagent. For amino acid sequencing, structural proteins from AdDNV
were separated by SDS-PAGE on 10% polyacrylamide gels and were electroblotted onto nitrocellulose membranes (Westran, Schleicher & Schuell, Keene, NH) and sequenced according to the method of Matsudaira (15).

MS. Expressed proteins from baculovirus constructs and native proteins from the virus were analyzed by mass spectrometry (MS) after separation by SDS-PAGE. The proteins, dissociated with 2% SDS at 95°C for 5 min, were run in a 10% acrylamide gel (13). The protein bands were cut from the gel and destained by PAGE. The proteins, dissociated with 2% SDS at 95°C for 5 min, were run in a 10% acrylamide gel (13). The protein bands were cut from the gel and destained. The proteins were then digested with trypsin and the resulting peptides were separated by reverse-phase liquid chromatography (SB-C18 Zorbax 300 reversed-phase column (150 mm x 4.6 mm), Agilent Nanopump instrument using a C18 Zorbax trap and an acetonitrile containing 0.1% trifluoroacetic acid). The tryptic peptides were then reduced with dithiothreitol (DTT) and alkylated with iodoacetamide prior to in-gel digestion.

Nucleotide sequence accession number. The AdDNV sequence is available in the GenBank database under accession number HQ827781.

RESULTS

AdDNV infection of A. domesticus. AdDNV is a frequent cause of epizootics in commercial or research mass rearing facilities for house crickets in Europe. The highest mortality is observed in the last larval stage and in young adults. These crickets die slowly over a period of several days; although they appear healthy, they lie on their backs and do not move. The guts of infected A. domesticus crickets that are still alive and no longer move are almost completely empty. Beginning in September 2009, mass epizootics have also occurred in rearing facilities throughout North America.

DNA sequence and organization of AdDNV isolates. Three full-length genomic clones in the pCR-XL-TOPO vector, namely, pAd22, pAd25, and pAd35, were obtained from the 1977 AdDNV isolate. Both strands of the viral genomes were sequenced (for the full annotated sequence, see Fig. S1 in the supplemental material). Nucleotide substitutions in more recent isolates have been reported elsewhere (24). The total length of the genome was 5,425 nt and contained ITRs of 144 nt, of which the distal 114 nt could fold into a perfect I-type palindromic hairpin (Fig. 1A, B, and D). The side arms in the typical Y-shaped terminal palindromes of many paroviruses were missing in the case of AdDNV.

Both complementary strands contained large ORFs in their 5’ halves; one strand had 3 large ORFs (ORFs 1 to 3), 2 of which were overlapping, and its complementary strand had 2 large ORFs (ORF-A and -B) (Fig. 1A). ORFs 1 to 3 potentially code for proteins consisting of 576, 286, and 213 amino acids (aa), respectively, whereas ORFs A and B potentially code for proteins of 597 and 268 aa, respectively. nBLAST analysis (http://www.ncbi.nlm.nih.gov) of the 5 ORFs revealed that the ORF-1 product is a homologue of densovirus NS1 proteins, the ORF-2 product is a homologue of densovirus NS2 proteins, the ORF-A product is a homologue of densovirus VP proteins, the ORF-B product is a homologue of paroviral phospholipase A2 proteins (N-terminal sequence of VP1), and the ORF-3 product does not have homologous proteins. NS1 ORF-1 can also be recognized by the presence of rolling circle replication and Walker A and B motifs, and the VP ORF can be recognized by the presence of a PLA2 motif (Fig. 1A; see Fig. S1 in the supplemental material). Since the convention for all paroviruses is to have the genes coding for the nonstructural proteins in the left half of the genome, it was decided to define the strain of the ambisense AdDNV genome con-

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence*</th>
<th>Position (nt) in AdDNV</th>
<th>Target or use</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdVPRI</td>
<td>TTTTGGAGTTCTCTGAGATGAC</td>
<td>2610–2633</td>
<td>Near 3’ end of VP mRNA</td>
</tr>
<tr>
<td>NADRI</td>
<td>ACATCTTGAAAGGCTGTTTACACCT</td>
<td>3892–3915</td>
<td>Just upstream of VP4</td>
</tr>
<tr>
<td>Adsp</td>
<td>CAGCTTCTTGTGGATGAGG</td>
<td>4362–4380</td>
<td>19-37 nt into VP2</td>
</tr>
<tr>
<td>Adsv</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Ads</td>
<td>GCTCTAGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Adl</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Ad</td>
<td>GCTCTAGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Adm</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>AdmF</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>AdmR</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Adr</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Adl</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>Adm</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
<tr>
<td>AdmR</td>
<td>GCCAGCTGATAGAGGCTGTTTACGTA</td>
<td>4453–4474</td>
<td>End of ORF-B (TGA)</td>
</tr>
</tbody>
</table>

* The AdDNV sequence is shown in capital letters. Underlined nucleotides indicate stop codons (e.g., TGA) or restriction sites (e.g., ggtacc [KpnI]).
containing the ORFs for the NS genes as the plus strand so the genes would be located similarly.

SDS-PAGE revealed that the capsid is composed of 4 structural proteins with estimated molecular masses ranging from 43 to 110 kDa (Fig. 2A), although a fifth protein may arise during purification, probably due to proteolysis (not shown). Attempts to obtain N-terminal sequences failed for VP1, VP3, and VP4, but the sequence TPPLKPHP(I)(E) was obtained for VP2, which indicated that its translation started at the AUG start codon of ORF-A, at nt 4398 to 4396 (Fig. 1D; see Fig. S1 in the supplemental material), and predicted a molecular mass of 65.3 kDa for VP2. ORF-B encoded the PLA2 motif recently identified in the structural proteins of most parvoviruses (4, 6, 28, 34) but was too small to code for a VP1 of 110 kDa as

FIG. 1. Genome organization of AdDNV. (A) Overview of genome organization, positions, and sizes of ITRs, ORFs, and introns. The Intron in the NS mRNA, between nt 223 and nt 855, occurs in about half of the NS transcripts. Ia (nt 4403 to 4758), Ib (nt 4403 to 4533), and II (nt 4260 to 4434) are introns that occur in alternative VP transcripts. vPLA2 indicates the position of the viral phospholipase A2 motif. (B) Left ITR and regulation of production of NS transcripts. NS transcripts start at 192-A and yield NS3 from 225-A. However, a fraction of these transcripts are spliced just upstream of this start codon (intron In), leading to translation of NS1 from 856-A (AUG with a poor initiation environment) and, through leaky scanning, of NS2 from 875-A. Inr and DPE are promoter elements. (C) Like the case for all members of the Densovirus genus, the 3'-ends of AdDNV NS and VP transcripts overlap in the middle of the genome. The stop codons and AATAAA motifs are underlined. (D) Right ITR, VP transcription sites, and splicing in ORF-B on the complementary strand. Transcription starts at nt 5235, and VP1 initiation is at nt 5230. The short 5'-UTR predicts an inefficient initiation (leaky scanning) and could be responsible for the production of a nested set of N-terminally extended viral proteins. However, removal of either of the two alternative introns in ORF-B (Ia or Ib) did not connect the exons in ORF-B and ORF-A in frame, so only nonstructural proteins could be produced from nt 5230 and VP2 could be produced directly from the first AUG in ORF-A when this splicing occurred. (E) An alternative intron II, which is mutually exclusive with introns Ia and Ib because the ORF-B splice acceptor is removed, connects ORF-B and ORF-A (both shaded) in frame so that VP1 can be produced from nt 5230. The VP1 sequence around the splicing site is underlined and shown above the nt sequence.
estimated by SDS-PAGE (Fig. 2A). Therefore, splicing of ORF-A and ORF-B seemed necessary to code for the largest AdDNV structural protein. Since the N-terminal coding region of ORF-A before its first ATG overlapped with the C-terminal coding region of ORF-B (Fig. 1E; see Fig. S1 in the supplemental material), an unspliced transcript could also code for VP1 by a ribosome frameshift in the ORF-A–ORF-B overlapping region, as suggested for PIDNV (33). These hypotheses were investigated further by transcript mapping and expression studies.

**Northern blotting and mapping of viral transcripts.** Northern blotting of RNAs obtained from infected Acheta larvae revealed two bands of NS transcripts (about 2.5 and 1.8 kb) and one band of VP transcripts (about 2.5 kb) (Fig. 2B). The transcript maps for RNAs isolated from both diseased crickets and recombinant baculovirus-infected cells were established by 5'- and 3'-RACE and are shown in Fig. 1B to D. The 3' termini of the NS and VP transcripts had a 34-nt overlap (Fig. 1C), similar to the situation observed with members of the Densovirus genus (27). NS transcription and splicing followed the same strategy as that previously described for GmDNV (28). A large unspliced transcript (nt 192 to 2596) was found to code for NS3 (first AUG in ORF-3), starting at nt 225. The NS3 coding sequence was removed in the spliced form in roughly half of these transcripts, with an intron from 221-AG/GT to 5235-AG/GT (Fig. 1D and 3A and B). Both introns failed to yield an in-frame coding sequence with ORF-A. The stop codon in the spliced ORF-B overlapped the start codon for VP2 (ugUAAUGa) (Fig. 1D). In some systems, e.g., influenza B virus (19, 20) and some non-long-terminal-repeat (non-LTR) retrotransposons (12), reinitiation occurs at such stop-start sequences. Expression of the intronless sequence from nt 4546 (before the small intron) to nt 3892 in the baculovirus system via pFastbacHTb, using primers Adsp and NAdR (Table 1) and cloning using EheI and EcoRI sites, did not yield products larger than the 29 aa expected from the baculovirus/ORF-B construct, arguing against reinitiation. These results gave credence to the previously suggested ribosome frameshift for PIDNV to generate a nested set of N-terminally extended structural proteins, as observed for all parvoviruses studied thus far (33). To test this hypothesis, we made several recombinant baculovirus constructs such that their expression products would be amenable to N-terminal sequencing in the potential frameshift region and they would be of reasonable size for mass spectrometry (Fig. 1E).

**Expression of VP1.** Figure 1E shows the 43-nt overlap of ORF-B with the N-terminal extension of ORF-A and the positions of primers (after the ORF-B and before the ORF-A stop codon) used to study the potential translational frameshift. The pFastbacHTb vector was used to yield products that could be purified via their N-terminal His tails and cleaved with the tobacco etch virus (TEV) protease, leaving only one codon (Gly) upstream of the insertion at the blunt-end EheI restriction site (Fig. 4A). The PCR products obtained using the forward primers Ad3s, Ad6s, Ads, Adm, and Adl (Table 1) (the distance from the ORF-B stop codon is indicated, in codons, in Fig. 4A) and the reverse primer NAdR, chosen at the beginning of the extension of VP4 to ensure stable prod-
ucts, were cloned into the EheI site, and the clones were selected for orientation by XbaI analysis (the XbaI sites in the MCS and the NAdR primer should be close to each other) and then verified by sequencing.

All constructs (His6-TEV recognition site-EheI site-insert-XbaI site) yielded proteins that could be purified via their His tails and that had molecular masses indicating that the two frames were connected. The Ad3s and Ads products were treated with TEV protease and were subjected to N-terminal sequencing after the N-terminal His tail/TEV site fragments and the protease were removed by affinity chromatography on Ni-nitrilotriacetic acid (Ni-NTA) columns and analyzed by SDS-PAGE (Fig. 4B). The sequences obtained (Ad3s sequence, GQP??RAEE; and Ads sequence, GAPQQP??QPPaAE), containing an N-terminal G and GA that were introduced via the primers and remained after TEV cleavage, indicated that nt 4435 in ORF-B was connected in frame with nt 4259 in ORF-A. Mass spectrometry analysis of the purified Ads product yielded a mass of 13,652 Da, with masses of 13,732 and 13,812 Da for phosphorylated species, for a predicted mass of 13,649.64 Da for the sequence GAPQQP...GGKRSR (Fig. 4C). These results confirmed the occurrence of a splice between nt 4435 and nt 4259. The predicted mass of VP1 is thus 88 kDa less than that estimated by SDS-PAGE (Fig. 2A), which may be explained by the phosphorylation observed by mass spectrometry.

Splicing was further investigated by RT-PCR of mRNAs extracted from infected crickets, using 2 sets of primers, namely, AdlgF/NAdR and Ads/AdVPR (Table 1), covering the whole coding sequence of VP1 except for the common VP4 sequence, with estimated products of 1,357 and 1,847 bp without splicing and 1,182 and 1,672 bp after intron II splicing, respectively. The intron II splice was also confirmed by sequencing of the VP1 cDNA (Fig. 3D). As illustrated in Fig. 1D and E, the Ia or Ib intron and the intron II splice were mutually exclusive, i.e., the intron II splice removed the acceptor site for the two intron I splices, whereas the intron I splices removed the donor site for intron II splicing. This expression strategy was further confirmed using recombinant baculovirus constructs. The VP1 sequence from which intron II was removed, and which was thus rendered resistant to ORF-B splicing (Ia and Ib), did not yield VP2 (Fig. 5A). In contrast, constructs with a mutated VP1 initiation codon and a normal VP template yielded a strong VP2 band but also some VP3 and VP4 (Fig. 5A), because type II splicing could remove the VP2 ini-
tiation codon and thus allow downstream initiation. When the template without intron II was used in combination with the mutated VP1 sequence, it yielded, as expected, VP3 and VP4 only. The leaky scanning of the VP3 initiation codon can probably be explained by its weak environment (uccAUGa), in contrast to the strong environment of the VP4 initiation codon (agaAUGg). Therefore, the VP multicistronic cassette yielded 2 sets of structural proteins (Fig. 5B).

Analysis of promoter activity. The promoter elements as well as the poly(A) signals were predicted by the mapping of transcription starts and polyadenylation sites of both NS and VP transcripts (Fig. 1B to D). To assess and compare their functionality, promoter regions (including the start of transcription) were amplified by PCR and cloned upstream of the luciferase gene in the pGL3-basic system. Their relative activities were determined by luciferase assays in independent duplicates at either 40 or 60 h posttransfection. The activity of the NS promoter of AdDNV was very significant in Ld652 cells from gypsy moths but was lower than that of the NS promoter of MlDNV, a lepidopteran densovirus (Fig. 2C). However, the VP1 promoter activity was significantly lower, suggesting the need for trans-activation, the absence of a critical factor reacting with the non-ITR region of the VP1 promoter, or differences in transcription factors between the cricket and gypsy moth systems.

Mass spectrometry of AdDNV capsid proteins. AdDNV was purified and the proteins separated by SDS-PAGE and subjected to mass spectrometry analysis in order to confirm the results obtained by analyzing the baculovirus constructs of the viral proteins. The proteins purified from the gel were digested with trypsin, and sequences of the peptides were determined. Analysis of VP1 and VP2 confirmed the results obtained with the baculovirus expression experiments with respect to the in-frame linking of ORF-B and ORF-A. The peptide sequences obtained covered 33% of VP1, 26% of VP2, 50% of VP3, and 42% of VP4 (Fig. 3E). One outlier peptide identified for VP3 was found in VP2, but with an ion score of 4, this was considered background.

Splicing of the Ia and Ib donor sites with the intron II acceptor site would theoretically also be possible and would yield additional products of 783 and 708 amino acids, close to the 816 amino acids observed for VP1. However, only 4 structural proteins were observed, and mass spectrometry demonstrated the presence of two VP1-specific peptides located in the introns of these potential supplementary products, of 708
DISCUSSION

The 1977 isolate of AdDNV was cloned and its expression strategy analyzed. Additional AdDNV isolates from Europe, isolated in 2004, 2006, 2007, and 2009, and from North America, isolated in 1988 (Tennessee) and 2009 (Quebec, Alberta, British Columbia, and Washington State), were amplified by PCR targeting the region between the ITRs and then sequenced (reported elsewhere [24]). All 2009 North American isolates had identical sequences, suggesting a common source, and differed from the 1977, 2004, and 2006 isolates by 49, 18,
and 16 nt substitutions, respectively. The genome organizations of these isolates were identical.

The sequences of the AdDNV ORFs were compared, using nBLAST, with those of other densoviruses, in particular with those of viruses such as PIDNV, PcDNV, BgDNV, MpDNV, and Dysaphis plantaginea densovirus (DplDNV), which also have split VP ORFs (see Fig. S3 in the supplemental material). Surprisingly, the highest identities by far for the AdDNV NS1 and ORF-A proteins were found for proteins of PcDNV from Planococcus citri (a citrus mealybug belonging to the Pseudococcidae family of the Hemiptera insect order, whereas Acheta domestica belongs to the Gryllidae family of the insect order Orthoptera).

The VP transcript was found to start 23 nt upstream of the 3'-ITR, at nt 5467, and the starts of both NS transcripts were at nt 573, 23 nt downstream of the 5'-ITR. This suggested that many promoter elements would be located reasonably well with the consensus sequence for Inr boxes (TCAGTG); however, the promoter activity in lepidopteran cells differed considerably (Fig. 2C). The region from the 5'-untranslated region (5'-UTR) in the VP mRNA to the putative VP1 AUG was only 5 nt long, whereas for the two NS transcripts, the 5'-UTRs were 32 (1.8-kb transcript) and 30 (2.5-kb transcript) nt long.

The expression strategy of the NS cassette is identical to that for the other members of the Densovirus genus. In the unspliced transcript (Fig. 1A), NS3 is translated, whereas in the spliced form the ORF for NS3 is removed and translation starts at the weak initiation codon of NS1 or, due to leaky scanning, at the coding sequence for NS2, 19 nt downstream.

In contrast to this conserved strategy, VP expression has unique features that so far have not been observed for vertebrate paroviruses and densoviruses, which all have a perfect nested set of N-terminally extended structural proteins. AdDNV displays split VP ORFs, and its two largest structural proteins have different extensions to which no roles have yet been assigned. PIDNV (33), PcDNV (25), BgDNV (9, 17), and MpDNV (31, 32), which all have a split VP ORF, with the smaller ORF encoding the PLA2 motif, probably have similar expression strategies (27). For PfDNV, a large number of donor and acceptor splicing sites have also been identified in the VP ORFs. Splicing of the NS cassette structure of PfDNV (33), PcDNV (25), BgDNV (9, 17), and MpDNV (31, 32), which all have a split VP ORF, with the smaller ORF encoding the PLA2 motif, probably have similar expression strategies (27).

ACKNOWLEDGMENTS

P.T. acknowledges support from the Natural Sciences and Engineering Research Council of Canada.

We thank the professionals in the cricket-rearing industry for their work.

REFERENCES


