Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*

Junping Wang, Jiamin Zhang, Hong Jiang, Chuanfeng Liu, Fuming Yi and Yuanyang Hu

Laboratory of Insect Virology, College of Life Sciences, Wuhan University, Wuhan 430072, China

The nucleotide sequence of a novel icosahedral DNA virus infecting *Dendrolimus punctatus* has been determined. The genome is 5039 nt long and includes inverted terminal repeats of 200 nt containing 131 nt long J-shaped terminal hairpins. The ‘plus’ strand of the genome contains three large open reading frames (ORFs), the left and the mid-ORFs (within the left ORF) in the left-half encoding the non-structural proteins and the right ORF in the right-half encoding viral capsid proteins. NS1 protein contains conserved replication initiation and DNA-dependent ATPase/helicase domains. VP1 protein contains a conserved PGY and phospholipase A2 motifs and shows high identities with VPs of *Casphalia extranea* densovirus and *Bombbyx mori* densovirus-1 belonging to the genus *Iteravirus*. Phylogenetic analysis also revealed that this virus is most closely related to *Casphalia extranea* densovirus and *Bombbyx mori* densovirus-1. Consequently, this virus was considered as a new third member of the genus *Iteravirus* of the subfamily *Densovirinae*, and designated *Dendrolimus punctatus* densovirus.

Densonucleosis viruses or densoviruses (DNVs) are invertebrate viruses belonging to the subfamily *Densovirinae* within the family *Parvoviridae* (Bergoin & Tijssen, 1998). DNVs have linear, single-stranded DNA genomes that are packaged in unenveloped icosahedral virions with a diameter of about 18–26 nm. Based on the structure and organization of their genomes, members of the subfamily *Densovirinae* were grouped into three genera, namely *Densovirus*, *Brevidensovirus* and *Iteravirus* (Borns et al., 2000). Currently, a new genus (*Pefudensovirus*) is created in the subfamily *Densovirinae* with *Periplaneta fuliginosa* densovirus as type species because of distinct genomic organizations (Mayo, 2004).

The members of the genus *Densovirus* have an approximately 6 kb genome with a similar ambisense genomic organization and long inverted terminal repeats (ITRs; >500 bp). They contain open reading frames (ORFs) encoding non-structural and structural proteins located on the 5’-halves of the genome strands, such as *Galleria mellonella* densovirus (GmDNV; Tijssen et al., 2003) and *Junonia coenia* densovirus (JcDNV; Dumas et al., 1992). The members of the genus *Pefudensovirus* possess an approximately 5–5 kb genome with ITRs of about 200 nt, having coding sequences on 5’-halves of ambisense genome. The ‘minus’ strand contains no less than two ORFs encoding structural proteins (Guo et al., 2000). The densovirus in the genus *Brevidensovirus* have genomes of about 4 kb with Y-like terminal hairpin structures, but lacking ITRs, and ORFs of non-structural and structural proteins are located on the same strand. Examples are *Aedes albopictus* parvovirus (AalDNV; Boublík et al., 1994) and *Aedes* denso-nucleosis virus (AaeDNV; Afanasiev et al., 1991). So far, the genus *Iteravirus* only consists of two members: *Bombbyx mori* densovirus-1 (BmDNV-1; Li et al., 2001) and *Casphalia extranea* densovirus (CeDNV; Fédière et al., 2002). The genomic DNAs of BmDNV-1 and CeDNV are 5076 and 5002 nt long, respectively. The CeDNV genome organization shows high similarity to that of BmDNV-1. The ORFs encoding non-structural proteins are on the 5’-half of the ‘plus’ strand whereas the ORF encoding structural proteins is on the 3’-half of the same strand. Unlike other densoviruses and vertebrate parvoviruses, the terminal sequences of their ITRs can be folded into an imperfect J-shaped palindrome. Like other parvoviruses, the conserved PGY motif is located within the unique part of VP1 of CeDNV and BmDNV-1, which contains the conserved motifs of phospholipase A2 (PLA2) domain (Zádori et al., 2001). Their PLA2 activity had been confirmed using the mixed micells assay (Li et al., 2001; Fédière et al., 2002). The tissue tropism of CeDNV and BmDNV-1 is different to that of members of the other three genera. CeDNV and BmDNV-1 can replicate almost exclusively in the columnar cells of...
midgut epithelium, whereas the other densoviruses replicate in most larval tissues except for the midgut (Tijssen & Bergoin, 1995).

*Dendrolimus punctatus* larvae (pine caterpillar) are the most destructive defoliator of massopine forests. Recently, a non-enveloped icosahedral DNA virus has been isolated from dead larvae of *Dendrolimus punctatus* in the Xinxian Forestry Center, Henan province, China. This virus is likely to be an important pathogen of *Dendrolimus punctatus*.

Virus particles are about 22 nm in diameter and contain a single-stranded DNA genome of approximately 50 kb. In the present study, we report the nucleotide sequence of this new densovirus from *Dendrolimus punctatus*. The results indicate that it is a new virus that should be classified within the genus *Iteravirus* of the subfamily Densovirinae and we have tentatively named it *Dendrolimus punctatus* densovirus (DpDNV).

Virus was isolated from dead larvae by the method described previously (Jousset *et al.*, 2000). In order to obtain viral DNA, virus was dissociated by incubation at 56 °C for 30 min in a buffer containing 10 mM Tris/HCl (pH 7.5), 100 mM NaCl, 15 mM MgCl2, 0-5 % SDS and 50 μg Proteinase K ml⁻¹. Next, a sample was extracted with an equal volume of phenol, followed by several back extractions of the phenol phase plus interphase with a NaCl solution (100 mM). The pooled, aqueous sample was extracted once with phenol/chloroform (1:1), and once with an equal volume of chloroform. The DNA was precipitated from the supernatant with ethanol and resuspended in water. After treatment with RNase A, the viral nucleic acids migrated as a single 5 kb band in 0-7 % agarose gel electrophoresis.

Extraction of viral DNA, in the presence of high salt concentrations, resulted in double-stranded DNA indicating that the plus and minus strands are packaged in separate virus particles. We failed to obtain recombinant plasmids of this double-stranded DNA, even after blunt-ending with Klenow. Therefore, we cloned *Pst*I fragments into *Pst*I-digested pUC18, whereas the terminal fragments were cloned by a method reminiscent of the 5'-RACE method for mRNA cloning. In this method, the single-stranded DNA of DpDNV, obtained after heating for 5 min at 94 °C and chilling for 1 min on ice, was polyadenylated with terminal deoxynucleotidyl transferase and then amplified by PCR with 5'-CCAGTGAGCAGAGTGACGAGGACTCGAGC(T)15-3' primer and the viral sequence-specific primers. The clones were sequenced on an Applied Biosystems automated sequencer, model 377, using universal sequencing and walking primer methods.

We determined the nucleotide sequence of the DpDNV genome (GenBank accession no. AY665654). The genome of DpDNV was 5039 nt in length with J-shaped hairpin structures (Fig. 1) and its organization is similar to that of members in the genus *Iteravirus* (Fig. 2a). Three large ORFs were found in the plus strand, encoding the non-structural proteins within the 5'-half and the structural proteins within the 3'-half.

![Fig. 1. Terminal hairpin structure of DpDNV genome. The terminal 131 nt within ITRs of DpNDV could be folded into J-shaped hairpin structures. In the flop orientation of DpDNV hairpin structure, sequence of nt 45–87 is replaced by a reverse-complement sequence.](image1)

![Fig. 2. (a) ORFs of the DpDNV genome. The plus strand contains three large ORFs whereas the minus strand does not contain any significant ORF. The replication initiator domain is indicated by a white box, the DNA-dependent ATPase/helicase domain is indicated by a black box and the parvovirus PLA2 domain is marked with a hatched box. (b) Conform of A and polyadenylation sites. The transcription of the NS mRNA most probably starts at position 387, encoding both NS1 and NS2. In addition, there is a potential TATA+1-Inr promoter upstream from the left ORF and thus the first in-frame ATG (nt 328) in this ORF may, after translation, produce a protein of 775 aa (NS1'). Two AATAAA sites at position nt 2704 and 4747 are probably functional for mRNAs of NS and viral proteins.](image2)
the 3’-half. DpDNV genome sequence shared about 60 % identity with that of both CeDNV and BmDNV-1. The base composition of the plus strand of the genome was A/T rich (35-56 % A, 18-49 % C, 19-07 % G and 26-87 % T), similar to other densoviruses (for example, CeDNV 36-34 % A, 25-80 % T; GmDNV 32-43 % A, 31-18 % T; IHHNV, Infectious hypodermal and hematopoietic necrosis virus 20-68 % A, 36-28 % T).

The DpDNV genome had ITRs of 200 nt, which was 30 nt shorter than those of CeDNV or BmDNV-1 and their terminal 131 nt could be folded into J-shaped hairpin structures (Fig. 1). These structures are characteristic of the CeDNV and BmDNV-1 genome ITRs, but do not resemble the typical Y- or T-shaped hairpin structures found in other paroviruses. The secondary structure of DpDNV consisted of a stem with basepairing nt 1–44 with nt 88–131 and an arm of J-shaped structure from nt 45 to 87 that was present in two alternative sequences (flip or flop) (Fig. 1). The sequences of CeDNV and BmDNV-1 ITRs are highly conserved, especially the flip/flop sequences (98 %) (Fédière et al., 2002). However, the flip/flop sequence of DpDNV displayed low identity with that of CeDNV (39 %) and BmDNV-1 (37 %). It is interesting that the flip/flop sequence of DpDNV has the same length, 43 nt, as that of CeDNV and BmDNV-1.

A direct repeat of 45 nt, containing a putative TATA-box, has been observed between the NS1 and VP ORFs of the BmDNV-1 but not for CeDNV (Li et al., 2001). The DpDNV sequence also revealed a direct repeat between the NS1 and VP ORFs, but smaller (23 nt between 2656 and 2691), did not contain the putative TATA-box. The corresponding sequence of DpDNV and BmDNV-1 contains a conserved TCTAAATC.

Three potential promoters were detected for the putative NS transcript of DpDNV according to the characteristics of invertebrate transcription initiation (Cherbas & Cherbas, 1993; Purnell et al., 1994) (Fig. 2b). The third putative NS promoter at nt position 356–419, which contained the typical downstream promoter element (DPE) (Kutach & Kadonaga, 2000) and the motif ten element (MTE) (Ohler et al., 2002; Lim et al., 2004) of invertebrate promoters, was most likely to be the functional promoter, named by ‘P7’ (at map unit 7). The first AUG (nt 494) in this transcript had been suggested to act as the initiation codon for NS2 protein translation. Since this initiation site conformed poorly to the Kozak consensus sequence (A/GCCaugG) (Kozak, 1987, 1999) (Fig. 2b), a leaking scanning may occur as with the other densoviruses. When an AUG codon is flanked by A–3 or by G–2 and G+4, the rest of the consensus sequence contributes only marginally (Kozak, 1999). Consequently, some 40S ribosomal subunits may bypass the AUG (nt 494) and then initiate at the next downstream in-frame AUG (the left frame) at position 799, producing NS1 protein. Moreover, a potential promoter (P54) upstream from the right ORF contained a TATA-box (TATAAT) at nt position 2700, and an initiator (Inr; TCAGT) at nt 2730 might be responsible for the transcription of mRNA encoding viral capsid proteins. Nine AATAAA sites, specific sequences for polyadenylation signal, were found downstream of the left and right ORFs. However, only two AATAAA sites, at nt positions 2704 and 4747, were most likely to be polyadenylation sites for DpDNV pre-mRNAs since those were followed by a CAYTG sequence and G/T-rich sequence, which are typical for eukaryotic transcription terminators (Birnstiel et al., 1985). The AATAAA site (nt 2704) at the end of the left ORF was followed by CATTC (nt 2727) and TTGCGGG (nt 2754), while another AATAAA site (nt 4747) at the end of the right ORF was flanked by CAATA (nt 4804) and GTGTGTG (nt 4951). Interestingly, the polyadenylation signal at position nt 4747 overlapped the stop codon of the right ORF.

The left ORF (nt 799–2655) encoded a putative NS1 of 618 aa, which had high identities with NS1 of two members of Iteravirus (CeDNV 44 %; BmDNV-1 44 %). The mid ORF (nt 494–1855) was located entirely within the left (NS1) ORF but in a different reading frame and encoded a putative NS2 of 453 aa with unknown function. It had 47 % homology to putative NS2s of CeDNV and BmDNV-1, members of Iteravirus, and shared no similarity with that of other paroviruses. The right ORF spanned from nt 2745 and terminated at nt 4751. Translation from the first in-frame ATG would produce a predicted 668v, 74 kDa protein. The amino acid sequence of the protein showed low identities (< 15 %) with that of members of the Densovirus, Pseudoendovirus and Brevidensovirus genera but 76 and 72 % with VP1s of CeDNV and BmDNV-1, respectively. Additional structural proteins (VP2–4) could be synthesized by a leaking scanning mechanism by initiation of translation from downstream AUGs.

The amino acid sequence of NS1 was further found to share the functional domains of the replication initiator and of the DNA-dependent ATPase/helicase with other paroviruses, which may be involved in the initiation of DNA replication (Ilyina & Koonin, 1992; Koonin, 1993). Motif I (H&H&&&) for metal binding site and motif II (Y&.K/R) for cleavage-ligation reaction (Nüesch et al., 1995), characteristic for the replication initiator domain, were located at aa 207–212 and 254–257, respectively (Fig. 3a). The C-terminal sequence (aa 461–580) contained the typical sequences of superfamily III-type ATPase/helicase (Koonin, 1993) (Fig. 3b). The sequence of the DNA-dependent ATPase/helicase domain of DpDNV shared 58 % identity with that of CeDNV and BmDNV-1. Motif A and B diverged slightly from the conserved sequences (GKN; &;&&ED/ED/E) of the tripartite superfamily III-type ATPase/helicase motifs. So far, the biological functions of NS1 protein of Densovirus have not been investigated in detail as vertebrate paroviruses. However, it was reported that NS1 of AaeDNV could stimulate expression of the viral protein gene (Afanasiev et al., 1994; Ward et al., 2001) and JcDNV NS1 possesses activities common to the superfamily of
rolling-circle replication initiator proteins especially parvovirus replication proteins (Ding et al., 2002).

A phylogenetic tree was generated with CLUSTALX based on the conserved sequence of DNA-dependent ATPase/helicase domain of NS1 proteins from parvoviruses (Fig. 3d). DpDNV was observed to be rather closely related to BmDNV-1 and CeDNV from the genus I terivirus, but less closely related to DsDNV, JcDNV, GmDNV and MIDNV from the genus D ensovirus and PidNV from the genus Pefudensovirus. However, DpDNV seemed to be distantly related to AaeDNV, AalDNV and IHHNV, members of the genus Brevidensovirus.

Recently, a conserved PLA2 domain, resembling the conserved motifs of secreted PLA2 was identified by sequence alignment in the VP1 unique region of parvoviruses (Zádori et al., 2001). The conserved PGY motif was located between aa 2 and 55 in VP1 of DpDNV. This motif contained YIGPG (aa 9–13) and HDLAY(x)2D (aa 32–49) sequences corresponding to the consensus sequences of the conserved $\text{Ca}^{2+}$ binding loop and catalytic site of secreted PLA2 domain, respectively (Fig. 3c) (Zádori et al., 2001; Fédère et al., 2004). Moreover, they demonstrated parvovirus PLA2 (pvPLA2) activity of Porcine parvovirus (PPV) and Erythro- virus B19 (B19), both in expressed viral proteins as well as in the infectious clones. Knock-out mutants displayed, by in situ hybridization, a defect in the transfer of the viral genome to the nucleus and cumulated in a perinuclear accumulation of virions (Zádori et al., 2001; Girod et al., 2002). Furthermore, pvPLA2 domain in VP1 of BmDNV-1, CeDNV and GmDNV has been demonstrated to be PLA2 activity (Li et al., 2001; Fédère et al., 2002; Tijssen et al., 2003). Therefore, the conserved motifs of pvPLA2 domain of DpDNV may play a similar role in viral infectivity as other parvoviruses.

Fig. 3. (a) Alignment of conserved replication initiator domain from DpDNV and other parvoviruses. Residues identical in at least 60% of the viruses are highlighted. An ampersand represents bulky hydrophobic residues and a dot represents any residue. (b) Alignment of the putative DNA-dependent ATPase/helicase of DpDNV and of other parvoviruses. Motifs A, B and C are present at the C terminus of NS1. Consensus amino acid residues are conserved either in all aligned sequences (uppercase) or in only some aligned sequences (lowercase). (c) Alignment of PLA2 domain in VP1 of DpDNV with that of other paroviruses. No typical conserved PLA2 domain was found in the VP1 of the densoviruses from the genus Brevidensovirus. (d) Phylogenetic analysis of DpDNV and other parvoviruses. The neighbour-joining algorithm was used to compare the putative DNA-dependent ATPase/helicase sequences shown in Fig. 3(b). The analysis was performed with the help of TREEVIEW. GenBank accession numbers of the viruses from the family Parvoviridae are shown on the tree. AAV2, Adeno-associated virus 2; and MVM, Minute virus of mice.
In conclusion, similarities in the genome organization, structural characteristics of the genome and sequence identities all suggest that DpDNV is a new third member of the genus Iteravirus of the subfamily Densovirinae. These results further support the classification of the Iteravirus as a separate genus within the subfamily Densovirinae.

Acknowledgements

The authors thank Professor Peter Tijssen of Université du Québec and Dr Haitao Guo of Fox Chase Cancer Center for suggestions of revising the manuscript.

References


