Journal of General Virology (2003), **84**, 165–172 DOI 10.1099/vir.0.18650-0

A new virus infecting *Myzus persicae* has a genome organization similar to the species of the genus *Densovirus*

M. van Munster,[1] A. M. Dullemans,[1] M. Verbeek,[1] J. F. J. M. van den Heuvel,[2] C. Reinbold,[3] V. Brault,[3] A. Clérivet[4] and F. van der Wilk[1]

[1]Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands

[2]De Ruiter Seeds, PO Box 1050, 2660 BB Bergschenhoek, The Netherlands

[3]INRA, 28 rue de Herrlisheim, 68021 Colmar Cédex, France

[4]Unité de Phytopathologie, IRD, BP 5045 Montpellier, France

Correspondence
Manuela van Munster
m.vanmunster@plant.wag-ur.nl

Received 11 June 2002
Accepted 11 September 2002

The genomic sequence of a new icosahedral DNA virus infecting *Myzus persicae* has been determined. Analysis of 5499 nt of the viral genome revealed five open reading frames (ORFs) evenly distributed in the 5′ half of both DNA strands. Three ORFs (ORF1–3) share the same strand, while two other ORFs (ORF4 and ORF5) are detected in the complementary sequence. The overall genomic organization is similar to that of species from the genus *Densovirus*. ORFs 1–3 most likely encode the non-structural proteins, since their putative products contain conserved replication motifs, NTP-binding domains and helicase domains similar to those found in the NS-1 protein of parvoviruses. The deduced amino acid sequences from ORFs 4 and 5 show sequence similarities with the structural proteins of the members of the genus *Densovirus*. These data indicate that this virus is a new species of the genus *Densovirus* in the family *Parvoviridae*. The virus was tentatively named Myzus persicae densovirus.

INTRODUCTION

Several virus agents that infect arthropods have been identified. So far, all the viruses known to infect aphid species are small isometric RNA viruses (Parrish & Briggs, 1966; D'Arcy *et al.*, 1981; Williamson *et al.*, 1988). Recently, we have isolated from the aphid *Myzus persicae* a DNA virus that shares the properties of the subfamily *Densovirinae*. The icosahedral virus particles were 20 nm in diameter and contained a single-stranded DNA genome of approximately 5·7 kb (unpublished results).

Among arthropod hosts, representatives of at least five orders from the class Insecta (Lepidoptera, Diptera, Orthoptera, Odonata and Hemiptera) and one order from the class Crustacea (Shike *et al.*, 2000) are known to be infected by densoviruses (DNVs) (Tijssen & Bergoin, 1995; Thao *et al.*, 2001). DNVs (family *Parvoviridae*) are small non-enveloped viruses of 18–26 nm in diameter containing a single-stranded DNA genome of 4–6 kb (van Regenmortel *et al.*, 2000). Viruses that belong to this family have generally two to four capsid proteins: VP1 (80–96 kDa), VP2 (64–85 kDa), VP3 (60–75 kDa) and VP4 (49–52 kDa) (van Regenmortel *et al.*, 2000). Common symptoms resulting from a densovirus infection include hypertrophy of infected nuclei, progressive paralysis and death of the insect host (Chao *et al.*, 1985; Kawase, 1985).

The overall genome organization of DNVs reveals striking differences both among themselves and with the genome organization of vertebrate parvoviruses (Bando *et al.*, 1987; Dumas *et al.*, 1992; Afanasiev *et al.*, 1991). Based on the characteristics of the genome structure, DNVs are sub-divided into three genera: *Densovirus*, *Iteravirus* and *Brevidensovirus*. In vertebrate parvoviruses and in DNVs of the genera *Iteravirus* and *Brevidensovirus*, the coding sequences of all viral proteins are located on one strand of the viral genome, which is by convention the viral (minus) strand (Berns, 1990). Two sets of open reading frames (ORFs) can be recognized, one encoding the non-structural proteins and the other encoding the capsid proteins. These coding regions are flanked by non-coding sequences of variable length forming terminal palindromic structures (Rhode & Iversen, 1990) that can form either a Y-shaped terminal structure (Astell, 1990), a T-shaped structure (Boublik *et al.*, 1994) or a simple hairpin structure (Bando *et al.*, 1990). These structures are implicated in the initiation of transcription (Astell, 1990) and also contain packaging signals (McLaughlin *et al.*, 1988) that mediate concomitant

The nucleotide sequence reported in this study appears in the EMBL, GenBank and DDBJ nucleotide sequence databases under accession number AY148187.

generation of single-stranded progeny DNA and encapsidation of viral DNA (Muller & Siegl, 1983).

DNVs belonging to the genus *Densovirus* have a unique genome organization characterized by its ambisense structure (Berns *et al.*, 1995). Most of these viruses contain three ORFs encoding the non-structural proteins (NS-1, NS-2 and NS-3) on one strand, while a single ORF encodes the structural proteins on the complementary strand (Dumas *et al.*, 1992; Gross & Tal, 2000). Two mRNA promoters are located in the 5′ inverted terminal repeat (ITR) of each strand and are responsible for the transcription of coding sequences (Berns *et al.*, 1995). Periplaneta fuliginosa densovirus (PfDNV) (Hu *et al.*, 1994), a member of the genus *Densovirus*, has a different coding strategy. PfDNV capsid proteins are encoded by three ORFs (Yamagishi *et al.*, 1999), which are connected in frame by alternative splicing strategies, in a manner similar to vertebrate parvoviruses (Berns, 1990). Moreover, an additional internal promoter, which could lead to the transcription of NS-1 and NS-2, has been detected (Yamagishi *et al.*, 1999).

Here, we report the molecular characterization of a new virus infecting the green peach aphid *M. persicae* (order Hemiptera). This icosahedral virus possesses a DNA genome of $\pm 5{\cdot}7$ kb, of which 5499 nt have been sequenced. Analysis of the sequence revealed the presence of five ORFs on the viral genome. The non-structural proteins were mapped in the 5′ half of one strand, while the ORFs encoding the structural proteins segregated in the 5′ half of the complementary strand. It is concluded from the genomic organization and from phylogenetic analysis that this virus is related to DNVs of the genus *Densovirus* and is most likely a new member of this genus.

METHODS

Aphid culture and virus purification. Infected *M. persicae* aphids were maintained on *Brassica oleracea* plants at 22 °C with a photoperiod of 16 h per day. Aphids collected for virus purification were stored at −80 °C. The virus was purified essentially as described for Acyrthosiphon pisum virus (van den Heuvel *et al.*, 1997). A major difference with the established protocol consisted of the buffer system used (0·1 M Tris/HCl, pH 7·4, instead of 0·1 M sodium phosphate, pH 7) and the purified virus was collected from a caesium chloride density gradient.

Nature of the nucleic acid. To isolate the genomic nucleic acid of Myzus persicae densovirus (MpDNV), purified virus particles were disrupted in 0·5 % SDS and extracted with phenol at 65 °C. The nucleic acid was extracted three times with phenol/chloroform (1 : 1), submitted to an ethanol precipitation and resuspended in water. To determine the nature of the nucleic acid, the MpDNV genome was subjected to endonuclease treatments (Sambrook *et al.*, 1989). A sample of 2 μg of the nucleic acid was incubated with 0·4 units of RNase A (Boehringer) or with 10 units of DNase I (Boehringer) for 30 min or 1 h, respectively. The samples were subsequently analysed on 1 % agarose gel containing 1 μg ethidium bromide ml^{-1}.

DNA isolation and cloning of the MpDNV genome. Genomic DNA was extracted from the purified virus using the DNeasy kit (Qiagen), according to the manufacturer's instructions. The viral DNA was digested with *Bam*HI and the obtained fragments of 140 bp were subsequently ligated into a *Bam*HI-digested pUC19 vector. The ligation mixture was used to transform competent *Escherichia coli* cells. Identification and isolation of recombinant clones was done following standard procedures (Sambrook *et al.*, 1989). Recombinant clones with an insert of 140 bp were isolated and sequenced. The obtained sequences were used to design specific primers for subsequent sequencing of the MpDNV genome.

PCR fragments corresponding to the MpDNV genome were synthesized using specific primers.

Sequence determination. Nucleotide sequencing was performed with an Applied Biosystems automated sequencer, model 373, employing a sequencing kit with AmpliTaq DNA polymerase (Applied Biosystems) and universal and MpDNV sequence-specific primers.

The nucleotide sequence of the MpDNV genome was determined by sequencing the purified viral DNA directly. To confirm the sequence results, 13 independently obtained and overlapping PCR fragments spanning the MpDNV genome were sequenced on both strands.

Computer analysis on nucleic acid and protein sequences. Computational sequence analysis was done using the computer package WISCONSIN, version 10.1 (GCG, Madison, Wisconsin, USA) and the BLAST suite (Altschul *et al.*, 1990). Transcriptional control signals were assessed by NEURAL NETWORK (PPNN) software (Ohler *et al.*, 1999) and the splice site prediction was done with the GENIE software package (Reese *et al.*, 1997).

Multiple alignments were performed with CLUSTAL X (Thompson *et al.*, 1994). Phylogenetic trees were constructed using the neighbour-joining method (Saitou & Nei, 1987) and the program TREECON, version 1.3b (Yves van de Peer, University of Antwerpen, Belgium). For each tree, confidence levels were estimated using the bootstrap resampling procedures (1000 trials). The sequences used in the alignments are: Diatraea saccharalis densovirus (DsDNV), *Junonia coenia densovirus* (JcDNV), PfDNV, *Galleria mellonella densovirus* (GmDNV), *Aedes aegypti densovirus* (AaDNV), *Aedes albopictus densovirus* (AlDNV), *Bombyx mori densovirus* (BmDNV-1), Infectious hypodermal and haematopoietic necrosis virus (IHHNV), Planococcus citri densovirus (PcDNV), Casphalia extranea densovirus (CeDNV) and *Mice minute virus* (MMV).

RESULTS

Virus purification and genome analysis

The initial virus purification was adapted from a protocol described previously by van den Heuvel *et al.* (1997). Using this protocol we were able to obtain approximately 8–10 μg of virus g^{-1} of aphids. Viral nucleic acid was extracted from the purified virus isolate and submitted to endonuclease treatments. This revealed that the viral genome was sensitive to DNase but not to RNase (data not shown). Moreover, the nucleic acid migrated as a single band on a 1 % agarose gel in electrophoresis (Fig. 1). Therefore, it was concluded that MpDNV possesses a DNA genome of approximately 5·7 kb in size.

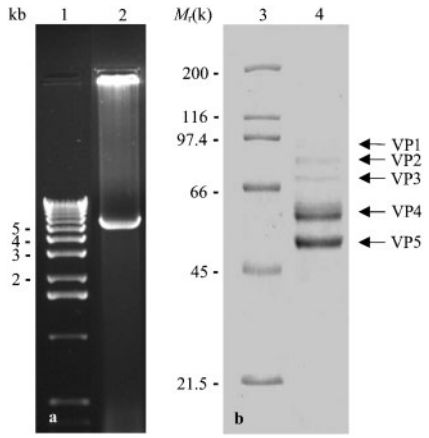

Fig. 1. (a) Agarose gel electrophoresis (1 %) of extracted MpDNV DNA (lane 2) showing a single band. Lane 1, molecular mass markers. (b) MpDNV structural proteins on a 10 % SDS–polyacrylamide gel stained with Coomassie brilliant blue (lane 4) and protein markers (lane 3, relative molecular mass M_r).

Nucleotide sequence

The nucleotide sequence of the MpDNV genome was obtained from direct sequencing by primer walking on the purified viral DNA and confirmed by sequencing 13 overlapping PCR fragments. No sequence heterogeneity was detected in the virus isolate. The viral sequence obtained was 5499 nt. The non-coding region of the 3′ extremity of the viral (minus) strand sequenced so far did not exceed 130 nt. In contrast, the obtained sequence of the non-coding region at the 5′ extremity of this strand was 201 nt long. Considering the size of the viral DNA estimated from agarose gel electrophoresis (Fig. 1), these extremities may be slightly longer.

ORFs

As shown in Fig. 2, computer-assisted analysis of the MpDNV genome revealed the presence of five potential coding domains. Both strands of the MpDNV genome have the capacity to encode peptides, like all species of the genus *Densovirus* characterized so far. Three ORFs (ORFs 1–3) are located on the 5′ part of one strand, while two additional ORFs (ORFs 4 and 5) span the 5′ half of the complementary strand (Fig. 2).

By convention, the ambisense genome of MpDNV is presented in such a way that the ORFs encoding the non-structural proteins are located to the left, analogous to the representation of the genomic organization of vertebrate parvoviruses (van Regenmortel *et al.*, 2000).

ORFs 1–3

ORFs 1–3 are located on the 5′ half on the same strand of the MpDNV genome. A computer search for putative transcription regulatory sequences detected one putative promoter (P1) at position 2–65 of the obtained sequence, in the 5′ ITR. P1 contains a putative TATA box (nt 57) and an appropriately positioned upstream GC-rich region (nt 2), which may act as an activator sequence (Bensimhon *et al.*, 1983) (Table 1). However, no conserved CATG boxes, known as an initiator and characteristic of invertebrate transcription (Cherbas & Cherbas, 1993), have been identified. As no likely promoter was detected on this strand using the PPNN program, it is suggested that P1 is responsible for the transcription of ORFs 1–3.

The small ORF1 starts at nt 208 and stops at nt 528. The first ATG codon is located at position 238 and translation of this ORF leads to a peptide of 97 aa, which corresponds to a molecular mass of 10·3 kDa. ORFs 2 and 3 overlap on different reading frames to ORF1. ORF2 starts at nt 395 with an ATG codon and terminates with a TAA codon at nt 1142. Two additional ATG codons at positions 443 and 467 may possibly act as start codons. The third in-frame ATG codon (nt 467) matches 6 of the 7 nt of the consensus sequence (ACCAUGG) for initiation of eukaryotic protein translation (Kozak, 1986), possibly initiating translation of ORF2. Assuming that this ATG codon is functional, the translation of ORF2 leads to a protein composed of 225 aa, with a molecular mass of 28 kDa. ORF3 overlaps ORF2, starts at position 525 and stops with a TAA codon at nt 2667. The first in-frame ATG codon is present at position 573 and is the putative start codon of this ORF. Translation of ORF3 leads to a protein of 698 aa, corresponding to a molecular mass of 81 kDa.

Several polyadenylation sites can be identified on this strand but only the AATAAA site at position 2642 is surrounded by a CAYTG sequence (CATTG, nt 2715) and a downstream GT-rich sequence (TTGATTTTT, nt 2682) necessary for transcription termination (Birnstiel *et al.*, 1985). Consequently, this region appears to be the most likely polyadenylation site for the RNA transcription of ORFs 1–3.

The deduced amino acid sequences of ORFs 1–3 were compared with entries in protein sequence databases. In the putative ORF2 product, a highly conserved consensus sequence of a metal-binding domain of DNA-binding proteins [$HXH(X)_2H$] was detected at residues 163–168 (Fig. 3a). The putative product of ORF3 revealed similarities with the non-structural protein NS-1 of PfDNV, DsDNV, JcDNV and PcDNV (37, 38, 38 and 33 % identity in a cluster of 200 aa, respectively). Indeed, the C-terminal region of the putative ORF3 product (aa 334–432) contains motifs characteristic of the NTP-binding and helicase domains of the NS-1 polypeptide shared by all parvoviruses (Afanasiev *et al.*, 1991; Iversen & Rhode, 1990) (Fig. 3b). However, no significant similarity was detected between the ORF1 product and the proteins of other parvoviruses reported so far.

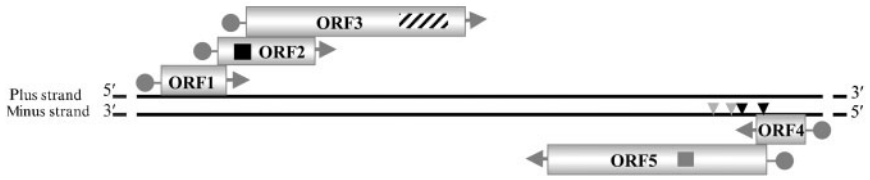

Fig. 2. Schematic presentation of the genomic organization of MpDNV. Boxes indicate putative ORFs. The position of the replication initiation motifs is indicated by a black box and the NTP-binding and helicase domains by a hatched box. In the structural proteins the dark grey box represents the glycine-rich motif of VP4. The donor and acceptor sites are indicated by black and grey arrows, respectively.

ORFs 4 and 5

ORFs 4 and 5 cluster in the 2·3 kb of the 5′ half of the complementary strand. Two potential promoters P2 and P3 were detected using the PPNN program. The putative P2 promoter, located in the 3′ ITR (nt 5410–5350) harbours a TATA box at position 5360 and an activator region (nt 5408) (Table 1). P2 is likely to be responsible for the transcription of ORF4. Another promoter P3 (nt 4895–4835) with a TATA box at nt 4887 and an activator region at nt 4881, located upstream the ORF5, may initiate the transcription of this ORF (Table 1).

ORF4 starts at position 5352 and terminates with a TAA stop codon at nt 4758. The ATG codon at nt 5337 is likely to be the translation initiation point and leads to a protein of 20·8 kDa. ORF5 is located in another reading frame and overlaps with ORF4. ORF5 starts at nt 4883 and spans to the TGA stop codon at nt 2999. Assuming that the first in-frame ATG codon (nt 4856) of ORF5 is functional, this ORF has a coding capacity of 68·3 kDa.

Three potential polyadenylation signals were identified on this strand at positions 5250, 3455 and 2694. However, only the polyadenylation signal at nt 2694 fulfils the criteria proposed by Birnstiel *et al.* (1985) for eukaryotic transcription terminators, as it is flanked by a CAYTG sequence (CAATG, nt 2719) and a downstream GT-rich sequence (GTTTTTT, nt 2598).

The deduced amino acid sequences of ORFs 4 and 5 were compared with entries in protein sequence databases. The putative MpDNV ORF5 product showed sequence similarities with the VP4 capsid protein of PcDNV (28 %), PfDNV (27 %), JcDNV (23 %), GmDNV (25 %) and DsDNV (25 %).

A stretch of 48 aa of the MpDNV ORF4 product (aa 30–78) exhibits 35 % identity with the JcDNV VP1 structural polypeptide but not with the other DNVs. This region is part of the 'PGY' conserved region of the VP1 capsid protein of most vertebrate parvoviruses. The results from these BLAST searches indicate that ORFs 4 and 5 of the MpDNV genome encode the five capsid proteins of 92, 85, 68, 64 and 57 kDa of MpDNV detected on polyacrylamide gels (Fig. 1). The 92 kDa protein could only be detected occasionally, suggesting that this protein is present in low amounts.

Potential splicing sites

Using the GENIE software package (Reese *et al.*, 1997), several potential splicing sites were detected in the ORFs encoding the structural proteins but none in the ORFs encoding the non-structural proteins of MpDNV. Two putative 5′ donor sites, D1 (score 0·93) at nt 5207 and D2 (score 0·99) at nt 4863, and two putative acceptor sites, A1 (score 0·77) at nt 4835 and A2 (score 0·93) at nt 4718, were detected on the strand shared by ORFs 4 and 5. In the event of splicing, the predicted sizes of the generated capsid proteins are shown in Table 2.

Relationship with other members of the subfamily *Densovirinae*

Phylogenetic analysis was carried out to compare the non-structural (Fig. 4a) and the structural proteins (Fig. 4b) of

Table 1. Potential promoter sequences on the MpDNV genome

Nucleotide positions are shown as the position of the first nucleotide in the genome and the position of the last nucleotide in the genome.

Location	Activator sequence	TATA box	Putative start codon
Plus-strand ORFs 1–3			
P1	CCCCCCGG (2–15)	TAATAA (57–62)	TATAATGA (234–241)
Minus-strand ORFs 4 and 5			
P2	GGTGGGGG (5408–5401)	ATATTT (5360–5355)	TAACATGT (5341–5334)
P3	GCGGCAGC (4881–4873)	TATTTA (4887–4882)	AAGTATGC (4860–4853)

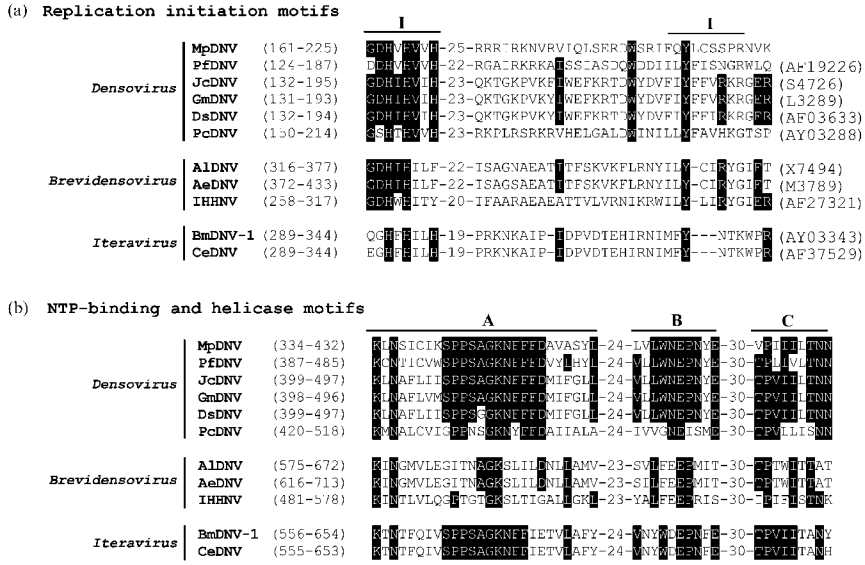

Fig. 3. Alignment of the amino acid sequences of (a) replication initiation motifs I and II and (b) NTP-binding and helicase domains A, B and C, corresponding to the highly conserved 119 aa of NS-1 of MpDNV and other densoviruses. Full names of the viruses are given in Methods. Numbers at the left indicate the starting amino acid positions of the aligned sequences. Residues conserved in at least six of the viruses are shown in reverse. GenBank accession numbers are shown in parentheses.

the different members of the subfamily *Densovirinae* with the deduced polyproteins sequences of MpDNV. The results revealed that MpDNV clusters with the species of the genus *Densovirus*, while members of the genera *Brevidensovirus* and *Iteravirus* are more distantly related.

DISCUSSION

The DNA genome of a new virus isolated recently from *M. persicae* was sequenced. A total of 5499 nt of the viral genome, estimated to be 5·7 kb, was obtained. The 5′ and 3′ extremities contained ITRs, with the presence of short palindromic sequences capable of forming hairpin-like structures by base pairing (data not shown), a feature shared by all parvoviruses (Astell, 1990). However, computer analysis failed to generate structures typical for the extremities of densoviruses (Bergoin & Tijssen, 1998). This indicates that the terminal sequences we have obtained so far are not complete. The overall genome organization, the size of the coding sequences and similarities in amino acid sequences with members of the genus *Densovirus* support the classification of MpDNV in this genus, despite some differences in the putative expression strategy. Two ORFs code for the structural proteins of MpDNV instead of one ORF, as is classically reported for DNVs.

Phylogenetic analysis comparing the non-structural proteins of the different members of the subfamily *Densovirinae* with the deduced non-structural proteins of MpDNV (Fig. 4a) showed that MpDNV was related most closely to PfDNV, a DNV isolated from the smoky-brown cockroach *Periplaneta fuliginosa*. However, comparison of the putative structural proteins (Fig. 4b) showed that MpDNV capsid proteins were related more to PcDNV. Interestingly, both PfDNV and PcDNV have an unique genome organization with more than one ORF encoding the structural proteins (Guo *et al.*, 2000; Thao *et al.*, 2001). It has also been demonstrated for PfDNV that its capsid proteins were generated by alternative splicing and using differential transcription termination in a manner similar to many vertebrate parvoviruses (Cassinotti *et al.*, 1988; Cotmore & Tattersall, 1990). As the coding capacities of ORFs 4 and 5 of MpDNV do not correspond to the final products detected on gels, it is suggested that the MpDNV genome uses the same strategy as PfDNV for translation of its capsid proteins. The presence of several putative splicing sites in the region encoding the capsid proteins of MpDNV appears to support this hypothesis. Noteworthy is that all the identified splicing events (Table 2) allow the connection of ORF4 with ORF5, resulting in the generation of larger ORFs. The size of the putative structural proteins generated by D1–A2 splicing (67·5 kDa) and D2–A1 splicing (84·5 kDa) are in agreement with the molecular mass of the proteins detected on gels (85 and 68 kDa). However, the size of VP1 (92 kDa) estimated by PAGE analysis cannot be explained by this alternative splicing strategy only. One possible explanation for the observed difference could be *N*-glycosylation of the polypeptide, since appropriate potential sites (N : X : S or N : X : T) (Marshall, 1972) are present on the putative amino acid sequences of the structural proteins. Moreover,

Table 2. Location of the 5′ donor and 3′ acceptor splicing sites and putative polypeptides generated by spliced mRNAs transcribed from potential P1 and P2 promoter sites on the MpDNV genome

5′ Donor sites (site's position in nt)	3′ Acceptor sites (site's position in nt)	Size of putative proteins (kDa)
GGTCCAG/GTAATTT (5214–5201)	TATTTAG/GTCCTCC (4841–4828)	72
	GTTTTAG/GTGGCGT (4724–4711)	67·5
GTAGCCG/GTAAGTA (4869–4856)	TATTTAG/GTCCTCC (4841–4828)	84·5
	GTTTTAG/GTGGCGT (4724–4711)	80

by analogy with PfDNV, which, in addition to alternative splicing, uses alternative transcription termination for the generation of the structural proteins (Yamagishi *et al.*, 1999), it cannot be excluded that the two other polyadenylation sites detected on the strand encoding the structural proteins of MpDNV are functional.

Two conserved regions in the capsid proteins of mammalian and insect parvoviruses were detected in the ORFs encoding the capsid proteins of MpDNV. The first conserved region corresponds to the beginning of VP4, harbouring a glycine-rich motif located in channels along the virion fivefold axes in vertebrate parvoviruses (Tsao *et al.*, 1991). These channels serve to externalize the N-terminal end of the capsid proteins in the virus particles (Tsao *et al.*, 1991). By analogy, as this glycine-rich motif is conserved in insect parvoviruses, it has been suggested to have a similar function for DNVs (Simpson *et al.*, 1998). The function of the second region, the so-called PGY motif, highly conserved (70–98 % identity within a stretch of 39 aa in VP1) among vertebrate parvoviruses and DNVs (Dumas *et al.*, 1992), was elucidated recently. Zadori *et al.* (2001) showed that this region was part of the catalytic site of secretory phospholipase A_2 (PLA_2). The activity of PLA_2 is critical for the efficient transfer of the viral genome from late endosomes to the nucleus to initiate replication and thus for parvovirus infectivity (Zadori *et al.*, 2001). This enzyme activity of the capsid was also demonstrated for BmDNV-1 (Li *et al.*, 2001) and CeDNV (Fédière *et al.*, 2002). Interestingly, this motif appears to be less conserved in the MpDNV structural proteins (30–36 % identity within a stretch of 39 aa with the other DNVs).

The MpDNV ORF5 product showed the highest sequence similarity with the putative structural proteins of PcDNV (28 %). PcDNV was isolated recently from a mealybug belonging to the order Hemiptera (Thao *et al.*, 2001). Moreover, the first 100 aa from ORF5 of MpDNV showed sequence similarity with PcDNV but with none of the other DNVs. Notably, the first 33 aa are highly conserved between the structural proteins of these two viruses (70 % of sequence similarity). The presence of this conserved region in the capsid proteins of two parvoviruses infecting two members of the order Hemiptera possibly suggests a role in host range recognition.

Two main domains of the ORFs encoding the putative non-structural proteins showed sequence similarities with the

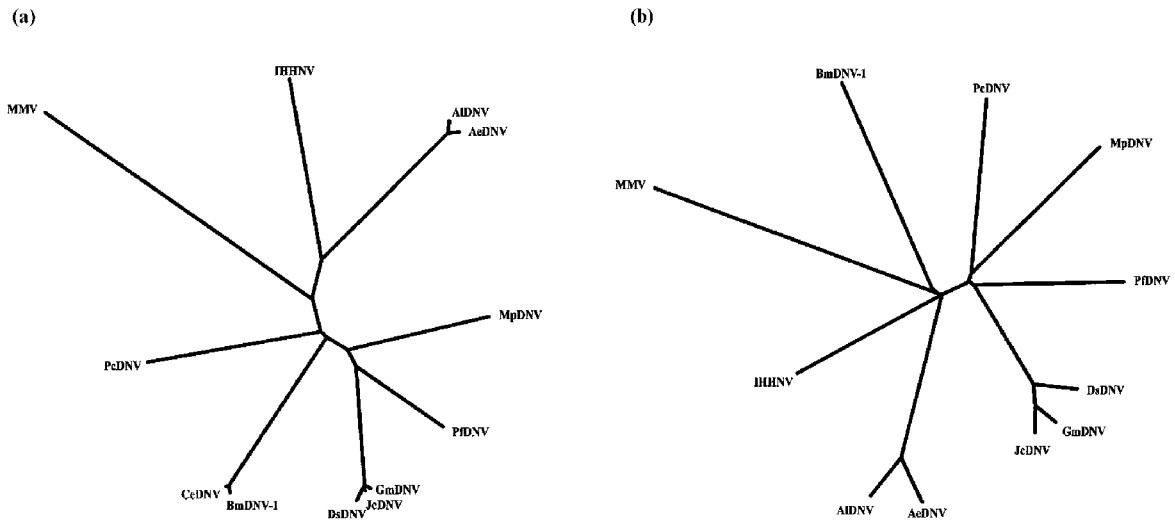

Fig. 4. Phylogenetic analysis of MpDNV-encoded proteins. The predicted amino acid sequences of the (a) 119 aa region of NS-1 shown in Fig. 3(b) and (b) VP4 of the insect parvoviruses were aligned and used to produce an unrooted tree. Branch lengths are proportional to distances between sequences. MMV was used as an outgroup (accession no. NC001510).

NS-1 protein of parvoviruses. The metal-binding motif detected at nt 161–225 is part of the replication initiator motifs common to all parvoviruses (Nuesch *et al.*, 1995) and is involved in initiation and termination of rolling-circle replication (Ilyina & Koonin, 1992). Sequence similarities were also found with the domains of the helicase superfamily III present in NS-1 of the vertebrate parvoviruses. This protein is a multifunctional nuclear protein required for parvovirus replication (Nuesch *et al.*, 1992). Ding *et al.* (2002) have shown recently that the NS-1 protein of JcDNV possesses activities common to the superfamily of rolling-circle replication initiator proteins similar to those of the vertebrate parvoviruses.

ACKNOWLEDGEMENTS

This research has been supported by a Marie Curie Research Fellowship of the European Community under contract number FAIR-BM-98-3051.

REFERENCES

Afanasiev, B. N., Galyov, E. E., Buchatsky, L. P. & Kozlov, Y. V. (1991). Nucleotide sequence and genomic organization of *Aedes densonucleosis virus*. *Virology* **185**, 323–336.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Astell, C. R. (1990). Terminal hairpins of parvovirus genomes and their role in DNA replication. In *CRC Handbook of Parvoviruses*, vol. 1, pp. 59–79. Edited by P. Tijssen. Boca Raton, FL: CRC Press.

Bando, H., Kusuda, J., Gojobori, T., Maruyama, T. & Kawase, S. (1987). Organization and nucleotide sequence of a densovirus genome imply a host-dependant evolution of the parvoviruses. *J Virol* **61**, 553–560.

Bando, H., Choi, H., Ito, Y. & Kawase, S. (1990). Terminal structure of a densovirus implies hairpin transfer replication which is similar to the model for AAV. *Virology* **179**, 57–63.

Bensimhon, M. J., Gabarro-Arpa, J., Ehrlich, R. & Reiss, C. (1983). Physical characteristics in eucaryotic promoters. *Nucleic Acids Res* **11**, 4521–4540.

Bergoin, M. & Tijssen, P. (1998). Biological and molecular properties of densoviruses and their use in protein expression and biological control. In *The Insect Viruses*, pp. 141–169. Edited by L. K. Miller & L. A. Ball. New York: Plenum.

Berns, K. I. (1990). Parvovirus replication. *Microbiol Rev* **54**, 316–329.

Berns, K. I., Bergoin, M., Bloom, M., Lederman, M., Muzyczka, N., Siegl, G., Tal., J. & Tattersall, P. (1995). The family *Parvoviridae*. In *Virus Taxonomy: Classification and Nomenclature of Viruses*. pp. 169–178. Edited by P. Murphy. New York & Berlin: Springer-Verlag.

Birnstiel, M. L., Busslinger, M. & Strub, K. (1985). Transcription termination and 3′ processing: the end is in site! *Cell* **41**, 349–359.

Boublik, Y., Jousset, F.-X. & Bergoin, M. (1994). Complete nucleotide sequence and genomic organization of the *Aedes albopictus* parvovirus (AaPV) pathogenic for *Aedes aegypti* larvae. *Virology* **200**, 752–763.

Cassinotti, P., Weitz, M. & Tratschin, J. D. (1988). Organization of the adeno-associated virus (AAV) capsid gene: mapping of a minor spliced mRNA coding for virus capsid protein 1. *Virology* **167**, 176–184.

Chao, Y. C., Young, S. Y., III, Kim, K. S. & Scott, H. A. (1985). A newly isolated densonucleosis virus from *Pseudoplusia includens* (Lepidoptera: Noctuidae). *J Invertebr Pathol* **46**, 70–82.

Cherbas, L. & Cherbas, P. (1993). The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* **23**, 81–90.

Cotmore, S. F. & Tattersall, P. (1990). Alternate splicing in a parvoviral nonstructural gene links a common amino-terminal sequence to downstream domains which confer radically different localization and turnover characteristics. *Virology* **177**, 477–487.

D'Arcy, C. J., Burnett, P. A., Hewings, A. D. & Goodman, R. M. (1981). Purification and characterization of a virus from the aphid *Rhopalosiphum padi*. *Virology* **112**, 346–349.

Ding, C., Urabe, M., Bergoin, M. & Kotin, R. M. (2002). Biochemical characterization of *Junonia coenia* densovirus nonstructural protein NS-1. *J Virol* **76**, 338–345.

Dumas, B., Jourdan, M., Pascaud, A. M. & Bergoin, M. (1992). Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densovirus reveals an organization unique among parvoviruses. *Virology* **191**, 202–222.

Fédière, G., Li, Y., Zadori, Z., Szelei, J. & Tijssen, P. (2002). Genome organization of *Casphalia extranea* densovirus, a new *Iteravirus*. *Virology* **292**, 299–308.

Gross, O., & Tal, J. (2000). Expression of the insect parvovirus *Gm*DNV *in vivo*: the structural and non-structural proteins are encoded by opposite DNA strands. *J Invertebr Pathol* **75**, 126–132.

Guo, H., Zhang, J. & Hu, Y. (2000). Complete sequence and organization of *Periplaneta fuliginosa densovirus* genome. *Acta Virol* **44**, 315–322.

Hu, Y., Zheng, J., Iizuka, T. & Bando, H. (1994). A densovirus newly isolated from the smoky-brown cockroach *Periplaneta fuliginosa*. *Arch Virol* **138**, 365–372.

Ilyina, T. V. & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* **20**, 3279–3285.

Iversen, P. & Rhode, S. L., III. (1990). Parvovirus proteins. In *CRC Handbook of Parvoviruses*, vol. 1, pp. 81–90. Edited by P. Tijssen. Boca Raton, FL: CRC Press.

Kawase, M. (1985). Pathology associated with densovirus. In *Viral Insecticides for Biological Control*, pp. 197–231. Edited by K. Maramosh & E. Sherman. New York: Academic Press.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Li, Y., Zadori, Z., Bando, H., Dubuc, R., Fédière, G., Szelei, J. & Tijssen, P. (2001). Genome organization of the densovirus from *Bombyx mori* (*Bm*DNV-1) and enzyme activity of its capsid. *J Gen Virol* **82**, 2821–2825.

McLaughlin, S. K., Collis, P., Hermonat, P. L. & Muzyczka, N. (1988). Adeno-associated virus general transduction vectors: analysis of proviral structures. *J Virol* **62**, 1963–1973.

Marshall, R. D. (1972). Glycoproteins. *Annu Rev Biochem* **41**, 673–702.

Muller, D.-E. & Siegl, G. (1983). Maturation of parvovirus LuIII in a subcellular system. I. Optimal conditions for *in vitro* synthesis and encapsidation of viral DNA. *J Gen Virol* **64**, 1043–1054.

Nuesch, J. P. F., Cotmore, S. F. & Tattersall, P. (1992). Expression of functional parvoviral NS1 from recombinant vaccinia virus:

effects of mutations in the nucleotide-binding motif. *Virology* **191**, 406–416.

Nuesch, J. P. F., Cotmore, S. F. & Tattersall, P. (1995). Sequence motifs in the replicator protein of parvovirus MVM essential for nicking and covalent attachment to the viral origin: identification of the linking tyrosine. *Virology* **209**, 122–135.

Ohler, U., Harbeck, S., Niemann, H., Noth, E. & Reese, M. G. (1999). Interpolated Markov chains for eukaryotic promoter recognition. *Bioinformatics* **15**, 362–369.

Parrish, W. B. & Briggs, J. D. (1966). Morphological identification of virus-like particles in the corn leaf aphid, *Rhopalosiphum maidis* (Fitch). *J Invertebr Pathol* **8**, 122–123.

Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. (1997). Improved splice site detection in Genie. *J Comput Biol* **4**, 311–323.

Rhode, S. L. III & Iversen, P. (1990). Parvovirus genomes: DNA sequences. In *Handbook of Parvoviruses*, vol. 1, pp. 31–57. Edited by P. Tijssen. Boca Raton, FL: CRC Press.

Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Shike, H., Dhar, A. K., Burns, J. C., Shimizu, C., Jousset, F.-X., Klimpel, K. R. & Bergoin, M. (2000). Infectious hypodermal and hematopoietic necrosis virus of shrimp is related to mosquito brevidensoviruses. *Virology* **277**, 167–177.

Simpson, A. A., Chipman, P. R., Baker, T. S., Tijssen, P. & Rossmann, M. G. (1998). The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution. *Structure* **6**, 1355–1367.

Thao, M. L., Wineriter, S., Buckingham, G. & Baumann, P. (2001). Genetic characterization of a putative densovirus from the mealybug *Planococcus citri*. *Curr Microbiol* **43**, 457–458.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tijssen, P. & Bergoin, M. (1995). Densonucleosis viruses constitute an increasingly diversified subfamily among the parvoviruses. *Semin Virol* **6**, 347–355.

Tsao, J., Chapman, M. S. & Agbandje, M. (1991). The three-dimensional structure of canine parvovirus and its functional implications. *Science* **25**, 1456–1464.

van den Heuvel, J. F. J. M., Hummelen, H., Verbeek, M., Dullemans, A. M. & van der Wilk, F. (1997). Characteristics of acyrthosiphon pisumvirus (APV), a newly identified virus infecting the pea aphid. *J Invertebr Pathol* **70**, 169–176.

van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. & Wickner, R. B. (editors) (2000). *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press.

Williamson, C., Rybicki, E. P., Kasdorf, G. G. F. & von Wechmar, M. B. (1988). Characterization of a new picorna-like virus isolated from aphids. *J Gen Virol* **69**, 787–795.

Yamagishi, J., Hu, Y., Zheng, J. & Bando, H. (1999). Genome organization and mRNA structure of *Periplaneta fuliginosa* densovirus imply alternative splicing involvement in viral gene expression. *Arch Virol* **144**, 2111–2124.

Zadori, Z., Szelei, J., Lacoste, M.-C., Li, Y., Gariépy, S., Raymond, P., Allaire, M., Nabi, I. R. & Tijssen, P. (2001). A viral phospholipase A_2 is required for parvovirus infectivity. *Dev Cell* **1**, 291–302.