

## Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family *Dicistroviridae*

M. van Munster,<sup>1</sup> A. M. Dullemans,<sup>1</sup> M. Verbeek,<sup>1</sup> J. F. J. M. van den Heuvel,<sup>2</sup> A. Clérvet<sup>3</sup> and F. van der Wilk<sup>1</sup>

<sup>1</sup> Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands

<sup>2</sup> De Ruiter Seeds, PO Box 1050, 2660 BB Bergschenhoek, The Netherlands

<sup>3</sup> Unité de Phytopathologie, IRD, BP 5045 Montpellier, France

The complete nucleotide sequence of the genomic RNA of an aphid-infecting virus, Aphid lethal paralysis virus (ALPV), has been determined. The genome is 9812 nt in length and contains two long open reading frames (ORFs), which are separated by an intergenic region of 163 nt. The first ORF (5' ORF) is preceded by an untranslated leader sequence of 506 nt, while an untranslated region of 571 nt follows the second ORF (3' ORF). The deduced amino acid sequences of the 5' ORF and 3' ORF products respectively showed similarity to the non-structural and structural proteins of members of the newly recognized genus *Cripavirus* (family *Dicistroviridae*). On the basis of the observed sequence similarities and identical genome organization, it is proposed that ALPV belongs to this genus. Phylogenetic analysis showed that ALPV is most closely related to *Rhopalosiphum padi virus*, and groups in a cluster with *Drosophila C virus* and *Cricket paralysis virus*, while the other members of this genus are more distantly related. Infectivity experiments showed that ALPV can not only infect aphid species but is also able to infect the whitefly *Trialeurodes vaporariorum*, extending its host range to another family of the order Hemiptera.

### Introduction

Aphid lethal paralysis virus (ALPV), an aphid-pathogenic virus, was first identified and characterized in South Africa (Williamson *et al.*, 1988). The incidence of ALPV was associated with rapid decline of the population of the major aphid species colonizing small grains under natural conditions (Laubscher & von Wechmar, 1992, 1993). Although ALPV is biologically well characterized, molecular data are not available. Virus particles are isometric, with a diameter of 26 nm, and contain a 9.7 kb single-stranded positive-sense RNA molecule (Williamson *et al.*, 1988). The capsid is constituted by three major polypeptides, of 34.4, 32 and 31.2 kDa, and a minor one of 40.8 kDa (Williamson *et al.*, 1988). On the basis of its biophysical and biochemical properties, ALPV was previously classified as a picorna-like virus (Williamson *et al.*, 1988), as are many other insect RNA viruses.

Since the molecular characterization of *Acyrtosiphon pisum virus* (van der Wilk *et al.*, 1997), the nucleotide sequences of the genomes of several insect RNA viruses have been determined (Johnson & Christian, 1998; Moon *et al.*, 1998). The availability of molecular data has enabled a new classification of insect picorna-like viruses. While Infectious flacherie virus (Isawa *et al.*, 1998) and Sacbrood virus (Ghosh *et al.*, 1999) have a genomic organization that resembles that of mammalian picornaviruses, *Acyrtosiphon pisum virus* could not be classified in any recognized taxon. Genome sequence analysis has shown that most of the insect RNA viruses belong to the genus of Cricket paralysis-like viruses (Christian *et al.*, 2000), recently renamed *Cripavirus* (family *Dicistroviridae*) (Mayo, 2002). Whilst there is sequence relatedness between these viruses and the mammalian picornaviruses, there are also some fundamental differences. The genome of mammalian picornaviruses consists of a positive-strand RNA containing a single large open reading frame (ORF) that encodes the capsid precursor in its 5' end and the non-structural protein precursor in its 3' part (Minor *et al.*, 1995). In contrast, the genome of the dicistroviruses contains two ORFs that are separated by an intergenic region (IGR). The non-structural proteins are encoded in the 5'

**Author for correspondence:** Manuela van Munster.

Fax +31 31 7410113. e-mail m.vanmunster@plant.wag-ur.nl

The nucleotide sequence reported here has been deposited in EMBL/GenBank/DBJ under accession no. AF536531.

ORF and the capsid proteins are encoded in the downstream ORF. Moreover, it has been shown for *Rhopalosiphum padi* virus (RhpV), *Plautia stali intestine virus* (PSIV) and *Cricket paralysis virus* (CrPV) that translation of ORF2 is cap-independent (Domier *et al.*, 2000; Sasaki & Nakashima, 2000; Wilson *et al.*, 2000a). In these three viruses, the non-coding region upstream of ORF2 functions as an internal ribosome entry site (IRES) that contains a pseudoknot that directs translation initiation at a non-AUG codon. Two other viruses, *Drosophila C virus* (DCV; Johnson & Christian, 1998) and *Triatoma virus* (TrV; Czibener *et al.*, 2000), harbour nucleotide sequences highly similar to the IRES of PSIV, RhpV and CrPV, suggesting the presence of an IRES for these viruses as well. ALPV is serologically related to CrPV (Williamson *et al.*, 1988) and is possibly a member of the *Dicistroviridae*, like *Acute bee paralysis virus* (ABPV; Govan *et al.*, 2000), *Himetobi P virus* (HiPV; Nakashima *et al.*, 1999), *Taura syndrome virus* (TSV; Mari *et al.*, 2002) and *Black queen cell virus* (BQCV; Leat *et al.*, 2000).

In this paper, the complete nucleotide sequence of the ALPV genomic RNA is reported. Analysis of the sequence revealed the presence of two ORFs. The non-structural proteins were mapped in the 5' ORF, while the 3' ORF encoded the capsid protein subunits. It is concluded from the genomic organization that ALPV is a member of the *Dicistroviridae*. Phylogenetic analysis of the putative non-structural polyprotein and capsid proteins enabled the classification of ALPV among members of the *Dicistroviridae*. In addition, infectivity studies showed that the occurrence of ALPV was not restricted to aphid hosts, and the virus could also infect a member of the family Aleyrodidae.

## Methods

**■ Virus purification and RNA isolation.** Initial infections of nymphs of *Rhopalosiphum padi* (family Aphididae) were established by feeding on an ALPV-containing diet for 2 days. Purified ALPV was a kind gift of M. B. von Wechmar (University of Cape Town, Rondebosch, South Africa). After 2 days, the infected aphids were transferred to *Avena sativa* plants at 22 °C with a photoperiod of 16 h per day. Virus was purified essentially as described for *Acyrtosiphon pisum virus* (van den Heuvel *et al.*, 1997). Genomic RNA was extracted from the purified virus by using the RNeasy Total kit (Qiagen), according to the manufacturer's instructions. The integrity of the viral RNA was verified by denaturing agarose gel electrophoresis (Sambrook *et al.*, 1989).

**■ Northern blot analysis.** Total RNA from 100 mg of healthy and ALPV-infected aphids was isolated by employing the Plant Total RNA kit (Qiagen), according to the manufacturer's instructions. In order to detect specific ALPV sequences, 16 µg of the purified RNA was separated on a 1% agarose gel containing formaldehyde (Sambrook *et al.*, 1989), transferred to a Hybond-N membrane (Amersham) and probed with a 720 nt radiolabelled cDNA fragment (nt 9092–9812 of the ALPV genome).

**■ Synthesis and cloning of cDNA.** Synthesis of cDNA was carried out by priming with oligo(dT) primers using the SuperScript System

(Gibco-BRL) according to the supplier's instructions. The double-stranded cDNA fragments were ligated in the *EcoRI* site of the  $\lambda$ ZAP II vector (Stratagene). The ligation mixture was used to transform *Escherichia coli* host strain XL-1 Blue MRF'. Identification and isolation of recombinant clones was done by following standard procedures (Sambrook *et al.*, 1989).

The cDNA of the 5'-terminal sequence of the ALPV genomic RNA was synthesized by RT-PCR using a 5' RACE kit (Gibco-BRL). The tailing reaction in the RACE procedure was performed with TdT and dCTP. PCR fragments were cloned using the TA cloning kit (Invitrogen), according to the manufacturer's instructions.

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

The sequence of the ALPV genomic RNA was determined by sequencing both strands of three independently obtained, overlapping cDNA clones covering the whole genome, and was confirmed by sequence analysis of another three cDNA clones on one strand. RT-PCR was carried out to amplify a 300 nt region to elucidate the beginning of the first ORF. Fragments were cloned and four of the clones obtained were analysed. The primers used in the RT-PCR procedure were identical to nt 392–412 and complementary to nt 707–727 of the ALPV genome.

**■ Computer analysis of nucleic acid and deduced protein sequences.** All computational sequence analysis was done using the Wisconsin package version 10.1 [Genetics Computer Group (GCG), Madison, WI, USA] and the BLAST suite (Altschul *et al.*, 1990). Multiple alignments were performed with CLUSTAL (Thompson *et al.*, 1994). Phylogenetic trees were constructed using the neighbour-joining method (Saitou & Nei, 1987). For each tree, confidence levels were estimated using the bootstrap resampling procedure (1000 trials). The sequences (with accession numbers) used in the alignments were: DCV (AF014388), PSIV (AB006531), *Cowpea mosaic virus* (CPMV; P03600, X00729), RhpV (AF022937), BQCV (AF183905), HiPV (AB017037), CrPV (AF218039), TrV (AF178440), ABPV (AF150629) and TSV (AF277675).

**■ Infection experiments.** Membrane feeding experiments were carried out for inoculation with ALPV of small-grain aphids (*Rhopalosiphum padi*, *Schizaphis graminum* and *Metopolophium dirhodum*), the green peach aphid *Myzus persicae* and the whitefly *Trialleurodes vaporariorum*. Aphid nymphs and adult whiteflies were allowed to feed on a 15% sucrose diet containing 0.01 and 1 mg/ml, respectively, of purified ALPV. Aphids and whiteflies feeding on a virus-free diet served as negative controls. After an acquisition period of 24 h, nymphs of *R. padi*, *S. graminum* and *M. dirhodum* were transferred to *Avena sativa* plants. *M. persicae* nymphs and adults of *T. vaporariorum* were respectively transferred to plants of *Brassica oleracea* and *Phaseolus vulgaris*. The presence of the virus was assayed 10 days later by double-antibody sandwich (DAS)-ELISA using two aphids or ten whiteflies per sample. Samples of 30 µl of the virus-free diet and of the diet containing 1 mg/ml ALPV were respectively used as negative and positive controls. Specific antibodies directed against ALPV were used to detect the virus in the samples. Moreover, the presence of the viral RNA within whiteflies was checked by RT-PCR. The total RNA of ten mock-inoculated or ten ALPV-inoculated whiteflies was extracted using the RNeasy Total kit (Qiagen) according to the manufacturer's instructions. Ten µl of the total RNA obtained (diluted to 30 µl final volume) was used in an RT-PCR procedure with specific ALPV primers. Primers sequences were as follows: upstream primer ALPV1 (5' ATCTTCGCAGTGTATTGGC 3'); downstream primer ALPV2 (5' GGAATAGCATCTAAGGGTTC 3'). PCR products obtained were analysed on agarose gels.

## Results

### Infection experiments

ALPV was originally isolated from the small-grain aphid *R. padi* and has been shown to infect a wide range of grain aphid species (Laubscher & von Wechmar, 1993). In this study, we have investigated the host range of ALPV. To this end, the small-grain aphids *R. padi*, *S. graminum* and *M. dirhodum* and the green peach aphid *M. persicae* were allowed to feed on ALPV in feeding experiments using an artificial diet mixed with virus. Moreover, to determine whether the host range of ALPV extends beyond the family Aphididae, whiteflies (*T. vaporariorum*) were inoculated by feeding experiments. Whiteflies belong to the Aleyrodidae, another family of the order Hemiptera.

At a concentration of 0.01 mg/ml purified virus, 50–60% of *R. padi* aphids inoculated became ALPV-infected when tested by ELISA after 10 days. The other aphid species tested, *S. graminum*, *M. dirhodum* and *M. persicae*, also became infected after membrane-feeding experiments (Table 1). Surprisingly, adult whiteflies could also acquire the virus from a diet containing 1 mg/ml purified ALPV. ALPV was still present in the whiteflies 10 days later when assayed by ELISA (Fig. 1). It is most likely that the whiteflies became infected with ALPV, since the viral RNA was still present in the insect 10 days after the end of the feeding period, as shown by RT-PCR (Fig. 1).

### Nucleotide sequence

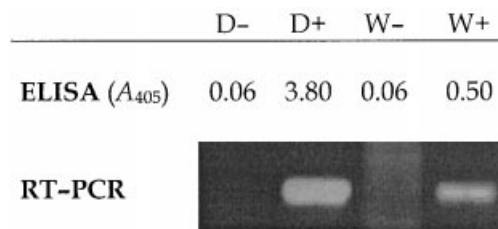
The nucleotide sequence of the ALPV genome was obtained from six overlapping cDNA clones. Each nucleotide was determined from at least two independent clones. The size of the genome was 9812 nt, excluding the poly(A) tail, which is consistent with the predicted size (9.7 kb) of the viral RNA determined from agarose gel electrophoresis (Williamson *et al.*, 1988). The 5' end of the genomic RNA was determined by a 5' RACE procedure and subsequent sequence analysis of five independently obtained clones.

The base composition of the entire genome is A (31.2%), U (30.2%), C (19.4%), G (19.2%). ALPV therefore resembles dicistroviruses and other insect picorna-like viruses.

**Table 1. Results of infection of aphid species with ALPV**

Aphids were fed on a diet containing 0.01 mg/ml purified ALPV. Infection was determined by DAS-ELISA.

Species	Infection by ALPV
<i>Rhopalosiphum padi</i>	+
<i>Metopolophium dirhodum</i>	+
<i>Schizaphis graminum</i>	+
<i>Myzus persicae</i>	+



**Fig. 1.** ELISA and RT-PCR on whiteflies. Sample groups: D-, 15% sucrose virus-free diet; D+, 15% sucrose diet containing 1 mg/ml purified ALPV; W-, whiteflies feeding on a 15% sucrose virus-free diet; W+, whiteflies feeding on a 15% sucrose diet containing 1 mg/ml purified ALPV.

Some 87% of the genome encodes two large ORFs, which are located on the same positive RNA strand but not in the same reading frame (Fig. 2). The other 13% of the genome consists of untranslated regions (UTRs). These UTRs span positions 1–506 of the 5' end (5'-UTR), the last 571 nt of the 3' end (3'-UTR) and the IGR (162 nt). In both the 5'-UTR and the IGR, three stem-loops were identified using the program MFOLD (Jacobson & Zuker, 1993), respectively between nt 300–493 and nt 6660–6802. Sequence comparison revealed that the IGRs of the different members of the *Dicistroviridae* and ALPV were well conserved (Fig. 3), whereas no sequence similarity could be detected in the 5'-UTRs of the viruses. The IGR of ALPV shared 66% sequence identity with the equivalent region of DCV and 64% with the IGR of RhPV. Moreover, a 5 nt inverted repeat sequence previously noted in other dicistroviruses was also present in the ALPV IGR. The first part of this inverted repeat sequence (located 20 nt upstream of the capsid coding region) is part of a loop in a stem-loop structure, while the second part of the inverted repeat is probably part of the first codon of the second ORF (CCU), as demonstrated for CrPV (Wilson *et al.*, 2000a), PSIV (Sasaki & Nakashima, 1999) and RhPV (Domier *et al.*, 2000).

### Non-structural polyprotein

The first ORF (5' ORF) is located between positions 507 and 6647. The first AUG codon, present at position 540, leads to the translation of a polypeptide with a molecular mass of 230.7 kDa. However, in one of three cDNA clones analysed, a U was present at position 510 instead of a C, resulting in an AUG start codon. The presence of this start codon in the same reading frame as the 5' ORF leads to a shift of the start of this ORF to position 510. To elucidate the beginning of the coding region, the nucleotide sequence of this region was determined from four independently obtained RT-PCR fragments. Three of the four PCR fragments contained a C at position 510, and only one clone contained a U residue. Therefore, it was concluded that the start of the 5' ORF is probably at position 540. However, it can not be entirely excluded that the start of the 5' ORF is at position 510.

The deduced amino acid sequence of the ALPV 5' ORF was compared with entries in protein sequence databases. This

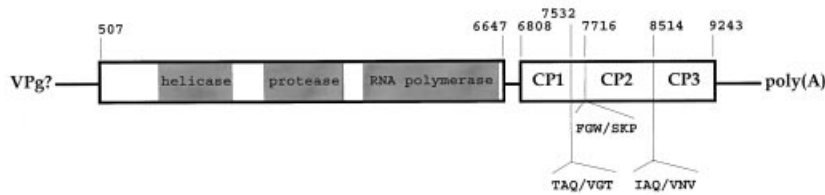


Fig. 2. Schematic presentation of the genomic organization of ALPV. Non-coding regions (single thick line) and ORFs (open boxes) are represented. The nucleotide positions at the start and the end of each ORF and of the putative cleavage sites are indicated. The conserved regions of the helicase, protease and RNA polymerase domains are shaded.

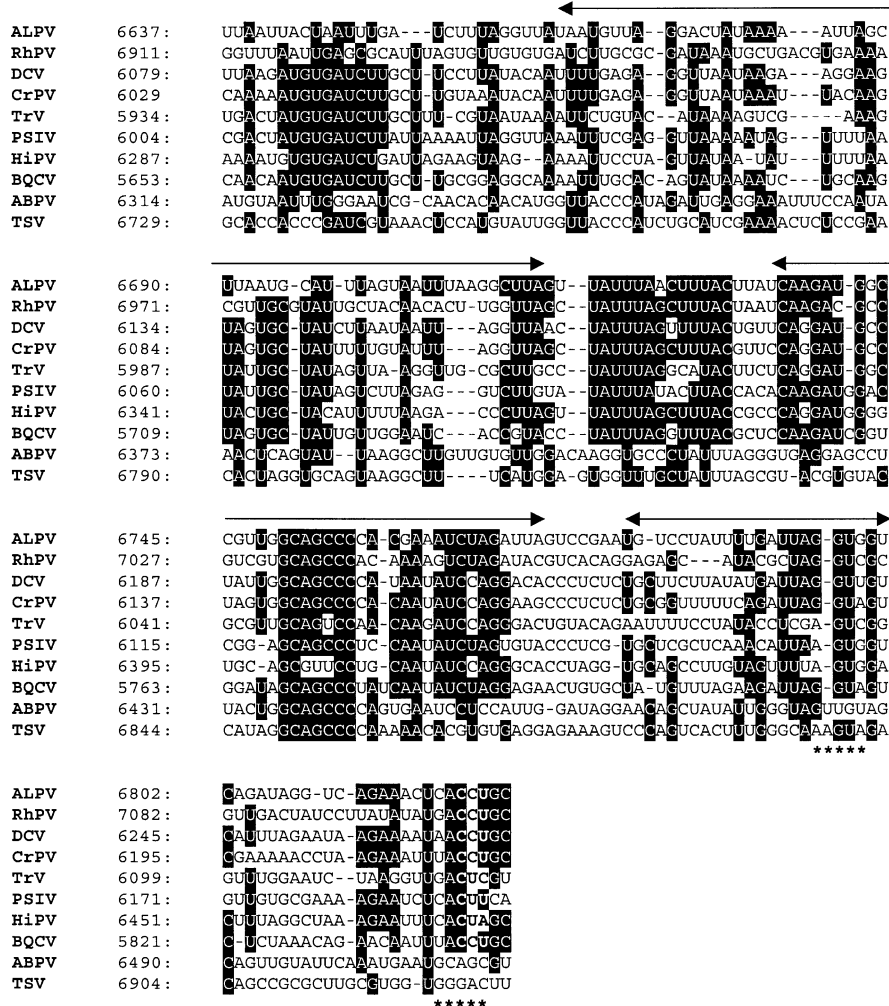


Fig. 3. Alignment of the nucleotide sequences of the IGRs of dicistroviruses. Numbers on the left indicate the starting positions of the aligned sequences. Residues conserved in at least six of the viruses are shown in white on black. Double-headed arrows above the alignment correspond to stem-loop structures predicted for the ALPV IRES. The inverted 5 nt repeat is indicated by asterisks and the ORF2 putative initiation codon is in bold.

revealed the highest similarity to the non-structural proteins of RhPV, DCV and CrPV (respectively 38, 33 and 29% identity). Sequence similarity was also found to the non-structural proteins of the *Picornaviridae*, *Sequiviridae* and *Caliciviridae*. Motifs characteristic of helicases, 3C-like proteases and RNA-dependent RNA polymerases, conserved among the *Picornaviridae*, *Comoviridae*, *Sequiviridae* and *Caliciviridae* (Koonin &

Dolja, 1993), were identified in the 5' ORF product of ALPV (Fig. 2).

The N-terminal region of the 5' ORF product (aa 400–800) contained conserved motifs characteristic of helicases and showed 67% identity to the putative helicase domain of RhPV. Two (A and B) of the three motifs conserved in helicase domains were recognized (Koonin & Dolja, 1993). Motif A,

(a)

CP1/CP4

ALPV	236:	<b>IAATAQ/VGTEAI*</b>
RhPV	231:	<b>PTSLAQ/VGEEGS</b>
DCV	273:	<b>DKIFAQ/VASEIT*</b>
CrPV	288:	<b>ARIYAQ/AAKELK*</b>
BQCV	231:	<b>SGMLAQ/AGLKVQ</b>
PSIV	254:	<b>TDLILQ/SGETSG</b>
HiPV	256:	<b>AVAREQ/VNLNSE</b>
TrV	250:	<b>SAIVAQ/AGKEQL</b>
ABPV	318:	<b>KNVTMQ/INSKKN</b>
TSV	267:	<b>VYPSTQ/SGYDDD</b>

CP2/CP3

ALPV	559:	<b>IRGIAQ/VNVAES*</b>
RhPV	553:	<b>ITSLAQ/VGTDTG</b>
DCV	629:	<b>ARIVAQ/VMGEDL</b>
CrPV	644:	<b>SRIVAQ/VMGEDQ*</b>
BQCV	573:	<b>KGMVAQ/SNSGTE</b>
PSIV	578:	<b>SQLTLQ/SGDTFG</b>
HiPV	611:	<b>STAQEQ/ANFAST</b>
TrV	591:	<b>GVPFAQ/VGFASA</b>
ABPV	701:	<b>IDASMQ/INLANK</b>
TSV	817:	<b>LYPSTH/AGLDYS</b>

(b)

CP4/CP2

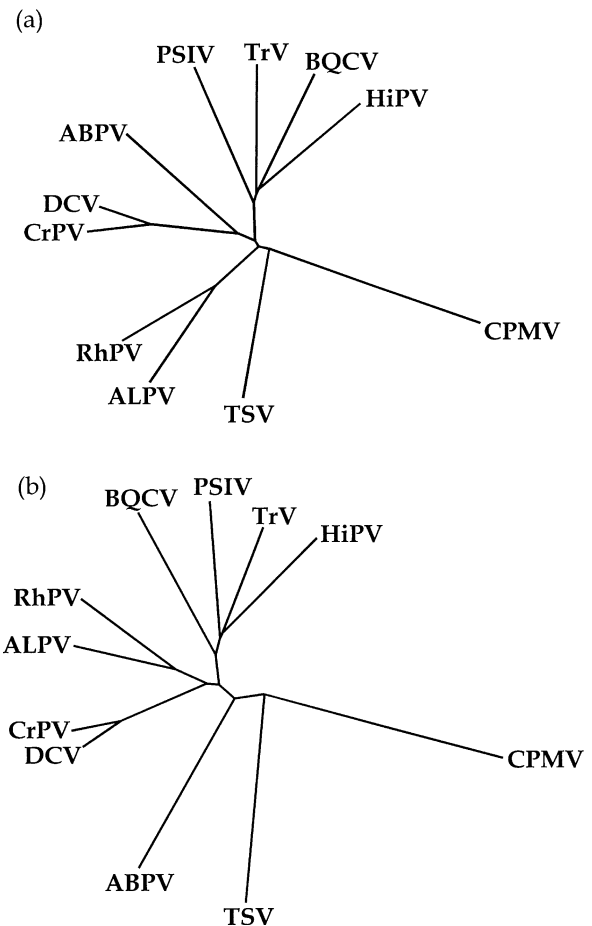
ALPV	294:	<b>AKHFGW/SKPTTA*</b>
RhPV	293:	<b>ASTHGW/SKPLNP</b>
DCV	338:	<b>FKMLGF/SKPTVQ</b>
CrPV	346:	<b>AKLFGF/SKPTVQ*</b>
BQCV	306:	<b>LSLFGF/SKPLLP</b>
PSIV	310:	<b>IAAFGF/SKPQKL</b>
HiPV	324:	<b>SMIPGF/KKPDKT</b>
TrV	307:	<b>LSALGF/SKPLTT</b>
ABPV	401:	<b>AAIFGW/SKPRNL</b>
TSV	324:	<b>ASMPGF/SKDRDM</b>

**Fig. 4.** Comparison of the amino acid sequences of dicistroviruses at the boundaries of the putative cleavage sites of the capsid polyprotein. (a) Alignment of peptide sequences at the proteolytic cleavage sites. (b) Alignment of peptide sequences at the putative catalytic cleavage site. Numbers on the left refer to amino acid positions in the polyprotein. The scissile bond of the cleavage site is represented by a forward slash (/). Highly conserved amino acid residues are in bold. Putative cleavage sites identified by sequence comparison are marked with an asterisk.

with the consensus sequence  $GX_4GK$  (aa 569–576), which is proposed to be responsible for nucleotide binding (Gorbalenya *et al.*, 1989), and motif B, WDGY (aa 612–615), were highly conserved in the putative 5' ORF product of ALPV.

A region similar to the conserved domain of 3C-like proteases was detected in the 5' ORF product (aa 1200–1500). The core motif, GXCG, was identified at residue 1411 of the putative ALPV 5' ORF product. However, the RNA-binding region of picornavirus 3C proteases (KFRDI; Ryan & Flint, 1997) was not detected.

In the C-terminal region of the putative 5' ORF product (aa 1550–2000), the conserved motifs of the putative RNA-dependent RNA polymerase (Koonin, 1991) were present. The highest sequence identity was observed to the putative RNA-dependent RNA polymerase of RhPV, which showed 53% identity to this region of the ALPV 5' ORF product.



**Fig. 5.** Phylogenetic analysis of ALPV-encoded proteins. The predicted amino acid sequences of the complete non-structural polyproteins (a) and capsid polyproteins (b) of the dicistroviruses were aligned and used to produce unrooted trees. Branch lengths are proportional to distances between sequences. CPMV was used as an outgroup.

### Capsid proteins

The second ORF (3' ORF), located in a different reading frame from the 5' ORF, starts at position 6808 and stops at position 9243 of the ALPV genome. The first AUG start codon in the 3' ORF is present 36 nt downstream of the start of the 3' ORF. However, on basis of sequence similarities of ALPV to the dicistroviruses and the presence of a CCU codon at position 6820, which is part of the aforementioned 5 nt inverted repeat, it is assumed that the beginning of the coding region is at position 6820, not at residue 6844. The 3' ORF encodes a polypeptide of 807 aa, corresponding to a molecular mass of 89.2 kDa. Amino acid sequence comparisons revealed that the ALPV 3' ORF product showed sequence similarity to the capsid proteins of the other members of the *Dicistroviridae*, *Picornaviridae*, *Sequiviridae* and *Caliciviridae*. The ALPV 3' ORF product respectively shared 38, 29 and 28% sequence identity with the capsid protein sequences of RhPV, DCV and CrPV. This indicates that the three major capsid proteins of 34, 32, 31 kDa and the minor one of 40.8 kDa detected on

polyacrylamide gels (Williamson *et al.*, 1988) result from processing of the ALPV 3' ORF product. The structural protein of dicistroviruses, like picornaviruses (Palmenberg, 1990), possesses specific cleavage sites that are recognized by the 3C-like protease. Published peptide sequences of cleavage sites determined by N-terminal sequencing of the capsid proteins of several dicistroviruses were aligned with the ALPV 3' ORF product. The putative cleavage sites at CP1/CP4 (analogous to VP4 of picornaviruses) (TAQ/VGT), CP4/CP2 (FGW/SKP) and CP2/CP3 (IAQ/VNV) were identified in the ALPV 3' ORF product at positions 241, 299 and 564, respectively (Fig. 4), leading to proteins of 25, 32 and 28 kDa.

### Relationships with other dicistroviruses

Phylogenetic analysis indicated that the deduced non-structural (Fig. 5a) and structural (Fig. 5b) polyprotein sequences of ALPV were most closely related to those of RhPV, CrPV and DCV. These four viruses seem to belong to one cluster, while another cluster, grouping HiPV (Nakashima *et al.*, 1999), TrV, PSIV and BQCV (Leat *et al.*, 2000), appears to be more distantly related. TSV is quite distinct from the other members and forms a separate cluster.

### Discussion

The molecular data and analysis of the ALPV genome show that this virus belongs to the recently recognized family *Dicistroviridae* (Mayo, 2002). The ALPV genome is 9812 nt in length, excluding the poly(A) tail, and contains two ORFs. Because of the use of the RACE method, it can not be excluded that one or more additional G residues were present at the 5' terminus. The 5' ORF encodes the non-structural proteins and the 3' ORF encodes the structural polyprotein. It is noteworthy that the 3'-UTRs of ALPV and RhPV (591 and 495 nt, respectively) are significantly longer than the 3'-UTRs of the other dicistroviruses sequenced so far.

The IGR of the dicistroviruses forms stem-loop structures that act as an IRES (Sasaki & Nakashima, 1999) in which the 3'-terminal stem-loop is part of a pseudoknot structure. This structure leads to the start of translation of the 3' ORF at a non-AUG codon (Sasaki & Nakashima, 2000). Recently, it has been shown that ribosomes bind the IGR-IRES of CrPV directly at the pseudoknot, and do not require the complete set of initiation factors, suggesting a novel mechanism of initiation of translation (Wilson *et al.*, 2000b). The IGR of ALPV shows structural similarities to the IGRs of various dicistroviruses and harbours a 5 nt inverted repeat at the 3' part of the IGR that is conserved in these viruses. Moreover, amino acid sequence similarities between the putative 3' ORF product of ALPV and the 3' ORF product of the dicistroviruses were detected upstream of the first methionine present in the second ORF. Furthermore, ALPV, like other dicistroviruses, does not produce subgenomic RNA, as demonstrated when carrying out a Northern blot procedure using infected *R. padi* aphids and purified virus (data not shown). Therefore, it is concluded

that the start of translation of the ALPV 3' ORF probably occurs at a non-AUG codon via a pseudoknot structure. Analysis of the 5'-UTR of ALPV revealed the presence of three stem-loop structures, which may also act as IRESs (data not shown). The 5'-UTRs of CrPV and RhPV have recently been shown to contain IRES elements that function in different *in vitro* translation systems (Wilson *et al.*, 2000b; Woolaway *et al.*, 2001) in a manner similar to those of picornaviruses (Belsham & Brangwyn, 1990).

Based on amino acid sequence similarities to known cleavage sites of dicistroviruses and by analogy with those of picornaviruses, three putative cleavage sites could be identified in the ALPV 3' ORF product, at residues 241 (TAQ/VGT), 299 (FGW/SKP) and 564 (IAQ/VNV). It is known that most cleavage reactions catalysed by the picornavirus 3C protease occur within a small subset of dipeptides comprising Q-G, -S, -T, -V, -A and -M (Hellen *et al.*, 1989). The putative cleavage sites of CP1/CP4 and CP2/CP3 of the ALPV 3' ORF product have the same dipeptide, Q-V, and the sequences flanking these sites are similar to each other. These residues surrounding the Q-V dipeptide might influence cleavage efficiency and are therefore conserved (Pallai *et al.*, 1989). While most of the cleavage sites of the picornaviruses harbour a proline residue in position 2 upstream of the cleavage junction (P2), an alanine residue seems to be prevalent at this position in dicistroviruses. Moreover, the prevalence of an alanine or other aliphatic residue at the P4 position has been shown in the cleavage site of picornaviruses (Nicklin *et al.*, 1986). The same observation could be made when the peptide sequences of cleavage sites of dicistroviruses were aligned. The putative cleavage site at CP4/CP2 occurs at a W-S dipeptide. By analogy with picornaviruses, it is suggested that the CP4/CP2 cleavage in dicistroviruses occurs after capsid formation (Sasaki *et al.*, 1998). This cleavage site is located inside the mature particle, as shown for CrPV (Tate *et al.*, 1999), and is therefore inaccessible to the viral proteases. In picornaviruses, the cleavage of VP0 into CP4 and CP2 probably results from the suitable juxtaposition of catalytic residues in the assembled particle (Basavappa *et al.*, 1994), and the same mechanism may occur in dicistroviruses. The cleavage site at CP4/CP2 (F-S, W-S or F-K) of ALPV and other dicistroviruses does not have the dipeptide preferred by 3C-like proteases, and the cleavage probably occurs catalytically.

The molecular masses of CP1, CP2 and CP3 calculated from the putative 3' ORF product (respectively 25, 32 and 28 kDa) are fairly consistent with the masses determined on SDS-PAGE (34, 32 and 31 kDa). However, the discrepancy between the values for the mass of CP1 (25 versus 31 kDa) could be attributed to post-translational modifications, as suggested for BQCV (Leat *et al.*, 2000).

Sequence comparison showed that the putative non-structural proteins and the putative capsid proteins of ALPV shared a high level of sequence similarity with RhPV, a member of the *Dicistroviridae*. RhPV is also a virus that infects

small-grain aphids and was first purified from *R. padi* (D'Arcy *et al.*, 1981). Phylogenetic analysis suggested the presence of two major clusters among the dicistroviruses. One cluster contains ALPV, RhPV, DCV, CrPV and possibly ABPV, while the second cluster harbours PSIV, BQCV, TrV and HiPV. TSV seems to be more distantly related and forms a separate group.

Studies to date have shown that ALPV infects mainly aphids that infest small grains, such as *S. graminum*, *M. dirhodum*, *Rhopalosiphum maidis* and *Diuraphis noxia* (Williamson *et al.*, 1988). In our study, we have shown that ALPV can also infect a member of the family Aleyrodidae. The host range of dicistroviruses varies considerably depending on the virus. While CrPV has a broad host range and can infect species of different insect orders (Scotti *et al.*, 1981), RhPV infection is restricted to aphid species (D'Arcy *et al.*, 1981). To date, there are only a few reports of viruses that infect whiteflies. Costa *et al.* (1996) described the presence of virus-like particles approximately 30 nm in diameter in the mycetocytes of *Bemisia tabaci*. Recently, a DNA virus belonging to the family Iridoviridae was reported to infect *B. tabaci* (Hunter *et al.*, 2001). However, the pathogenicity of these viruses was not determined.

Whiteflies and aphids cause economic losses worldwide because of their feeding behaviour and their role as vectors of plant viruses. Therefore, control of these insect pests is needed. Current measures of control rely largely upon frequent applications of hazardous insecticides during the growing season. The extensive use of insecticides has led to widespread development of resistance in many aphid and whitefly species and this represents one of the major threats to the future success of chemical pest control. The development of non-chemical biological agents like insect viruses may lead to a significant reduction in the use of and dependence on insecticides. In this context, the potential of ALPV to infect species of both the families Aphididae and Aleyrodidae might open a window of opportunity for biological control of these insect pests.

The authors acknowledge Professor M. B. von Wechmar for generously providing purified ALPV and antiserum to ALPV. This research has been supported by a Marie Curie Research Fellowship of the European Community under contract number FAIR-BM-98-3051.

## References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- Basavappa, R., Syed, R., Flore, O., Icenogle, J. P., Filman, D. J. & Hogle, J. M. (1994). Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: structure of the empty capsid assembly intermediate at 2.9 Å resolution. *Protein Science* **3**, 1651–1669.
- Belsham, G. J. & Brangwyn, J. K. (1990). A region of the 5' noncoding region of foot-and-mouth disease virus RNA directs efficient internal initiation of protein synthesis within cells: involvement with the role of L protease in translational control. *Journal of Virology* **64**, 5389–5395.
- Christian, P., Carstens, E., Domier, L., Johnson, K., Nakashima, N., Scotti, P. & van der Wilk, F. (2000). Genus "Cricket paralysis-like viruses". In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 678–683. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.
- Costa, H. S., Westcot, D. M., Ullman, D. E., Rosell, R. C., Brown, J. K. & Johnson, M. W. (1996). Virus-like particles in the mycetocytes of the sweetpotato whitefly, *Bemisia tabaci* (Homoptera, Aleyrodidae). *Journal of Invertebrate Pathology* **67**, 183–186.
- Czibener, C., La Torre, J. L., Muscio, O. A., Ugalde, R. A. & Scodeller, E. A. (2000). Nucleotide sequence analysis of Triatoma virus shows that it is a member of a novel group of insect RNA viruses. *Journal of General Virology* **81**, 1149–1154.
- 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.
- Domier, L. L., McCoppin, N. K. & D'Arcy, C. J. (2000). Sequence requirements for translation initiation of *Rhopalosiphum padi* virus ORF2. *Virology* **268**, 264–271.
- Ghosh, R. C., Ball, B. V., Willcocks, M. M. & Carter, M. J. (1999). The nucleotide sequence of sacbrood virus of the honey bee: an insect picorna-like virus. *Journal of General Virology* **80**, 1541–1549.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Research* **17**, 4713–4730.
- Govan, V. A., Leat, N., Allsopp, M. & Davison, S. (2000). Analysis of the complete genome sequence of acute bee paralysis virus shows that it belongs to the novel group of insect-infecting RNA viruses. *Virology* **277**, 457–463.
- Hellen, C. U. T., Krausslich, H.-G. & Wimmer, E. (1989). Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**, 9881–9890.
- Hunter, W. B., Patte, C. P., Sinisterra, X. H., Achor, D. S., Funk, C. J. & Polston, J. E. (2001). Discovering new insect viruses: whitefly iridovirus (Homoptera: Aleyrodidae: *Bemisia tabaci*). *Journal of Invertebrate Pathology* **78**, 220–225.
- Isawa, H., Asano, S., Sahara, K., Iizuka, T. & Bando, H. (1998). Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses. *Archives of Virology* **143**, 127–143.
- Jacobson, A. B. & Zuker, M. (1993). Structural analysis by energy dot plot of a large mRNA. *Journal of Molecular Biology* **233**, 261–269.
- Johnson, K. N. & Christian, P. D. (1998). The novel genome organization of the insect picorna-like virus *Drosophila* C virus suggests this virus belongs to a previously undescribed virus family. *Journal of General Virology* **79**, 191–203.
- Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197–2206.
- Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 375–430.
- Laubscher, J. M. & von Wechmar, M. B. (1992). Influence of aphid lethal paralysis virus and *Rhopalosiphum padi* virus on aphid biology at different temperatures. *Journal of Invertebrate Pathology* **60**, 134–140.

- Laubscher, J. M. & von Wechmar, M. B. (1993).** Assessment of aphid lethal paralysis virus as an apparent population growth-limiting factor in grain aphids in the presence of other natural enemies. *Biocontrol Science and Technology* **3**, 455–466.
- Leat, N., Ball, B., Govan, V. & Davison, S. (2000).** Analysis of the complete genome sequence of black queen-cell virus, a picorna-like virus of honey bees. *Journal of General Virology* **81**, 2111–2119.
- Mari, J., Poulos, B. T., Lightner, D. V. & Bonami, J.-R. (2002).** Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus *Cricket paralysis-like viruses*. *Journal of General Virology* **83**, 915–926.
- Mayo, M. A. (2002).** Virology Division news: virus taxonomy – Houston 2002. *Archives of Virology* **147**, 1071–1076.
- Minor, P. D., Brown, F., Domingo, E., Hoey, E., King, A., Knowles, N., Lemon, S., Palmenberg, A., Rueckert, R. R., Stanway, G., Wimmer, E. & Yin-Murphy, M. (1995).** Family *Picornaviridae*. In *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*, pp. 329–336. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo & M. D. Summers. Vienna & New York: Springer-Verlag.
- Moon, J. S., Domier, L. L., McCoppin, N. K., D'Arcy, C. J. & Jin, H. (1998).** Nucleotide sequence analysis shows that *Rhopalosiphum padi* virus is a member of a novel group of insect-infecting RNA viruses. *Virology* **243**, 54–65.
- Nakashima, N., Sasaki, J. & Toriyama, S. (1999).** Determining the nucleotide sequence and capsid-coding region of himetobi P virus: a member of a novel group of RNA viruses that infect insects. *Archives of Virology* **144**, 2051–2058.
- Nicklin, M. J. H., Toyoda, H., Murray, M. G. & Wimmer, E. (1986).** Proteolytic processing in the replication of polio and related viruses. *Bio/Technology* **4**, 33–42.
- Pallai, P. V., Burkhardt, F., Skoog, M., Schreiner, K., Bax, P., Cohen, K. A., Hansen, G., Palladino, D. E. H., Harris, K. S., Nicklin, M. J. & Wimmer, E. (1989).** Cleavage of synthetic peptides by purified poliovirus 3C proteinase. *Journal of Biological Chemistry* **264**, 9738–9741.
- Palmenberg, A. C. (1990).** Proteolytic processing of picornaviral polyprotein. *Annual Review of Microbiology* **44**, 603–623.
- Ryan, M. D. & Flint, M. (1997).** Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.
- Saitou, N. & Nei, M. (1987).** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **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.
- Sasaki, J. & Nakashima, N. (1999).** Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus *in vitro*. *Journal of Virology* **73**, 1219–1226.
- Sasaki, J. & Nakashima, N. (2000).** Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proceedings of the National Academy of Sciences, USA* **97**, 1512–1515.
- Sasaki, J., Nakashima, N., Saito, H. & Noda, H. (1998).** An insect picorna-like virus, *Plautia stali* intestine virus, has genes of capsid proteins in the 3' part of the genome. *Virology* **244**, 50–58.
- Scotti, P. D., Longworth, J. F., Plus, N., Croizier, G. & Reinganum, C. (1981).** The biology and ecology of strains of an insect small RNA virus complex. *Advances in Virus Research* **26**, 117–143.
- Tate, J., Liljas, L., Scotti, P., Christian, P., Lin, T. & Johnson, J. E. (1999).** The crystal structure of cricket paralysis virus: the first view of a new virus family. *Nature Structural Biology* **6**, 765–774.
- 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 Research* **22**, 4673–4680.
- van den Heuvel, J. F. J. M., Hummelen, H., Verbeek, M., Dullemans, A. M. & van der Wilk, F. (1997).** Characteristics of *Acyrtosiphon pisum* virus, a newly identified virus infecting the pea aphid. *Journal of Invertebrate Pathology* **70**, 169–176.
- van der Wilk, F., Dullemans, A. M., Verbeek, M. & van den Heuvel, J. F. J. M. (1997).** Nucleotide sequence and genomic organization of *Acyrtosiphon pisum* virus. *Virology* **238**, 353–362.
- Williamson, C., Rybicki, E. P., Kasdorf, G. G. F. & Von Wechmar, M. B. (1988).** Characterization of a new picorna-like virus isolated from aphids. *Journal of General Virology* **69**, 787–795.
- Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. (2000a).** Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular and Cellular Biology* **20**, 4990–4999.
- Wilson, J. E., Pestova, T. V., Hellen, C. U. T. & Sarnow, P. (2000b).** Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511–520.
- Woolaway, K. E., Lazaridis, K., Belsham, G. J., Carter, M. J. & Roberts, L. O. (2001).** The 5' untranslated region of *Rhopalosiphum padi* virus contains an internal ribosome entry site which functions efficiently in mammalian, plant, and insect translation systems. *Journal of Virology* **75**, 10244–10249.

---

Received 16 July 2002; Accepted 12 August 2002