Analysis of new aphid lethal paralysis virus (ALPV) isolates suggests evolution of two ALPV species

Sijun Liu,¹ Diveena Vijayendran,¹ Jimena Carrillo-Tripp,² W. Allen Miller² and Bryony C. Bonning¹

¹Department of Entomology, Iowa State University, Ames, IA 50011, USA
²Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA

Aphid lethal paralysis virus (ALPV; family Dicistroviridae) was first isolated from the bird cherry-oat aphid, Rhopalosiphum padi. ALPV-like virus sequences have been reported from many insects and insect predators. We identified a new isolate of ALPV (ALPV-AP) from the pea aphid, Acyrthosiphon pisum, and a new isolate (ALPV-DvV) from western corn rootworm, Diabrotica virgifera virgifera. ALPV-AP has an ssRNA genome of 9940 nt. Based on phylogenetic analysis, ALPV-AP was closely related to ALPV-AM, an ALPV isolate from honeybees, Apis mellifera, in Spain and Brookings, SD, USA. The distinct evolutionary branches suggested the existence of two lineages of the ALPV virus. One consisted of ALPV-AP and ALPV-AM, whilst all other isolates of ALPV grouped into the other lineage. The similarity of ALPV-AP and ALPV-AM was up to 88% at the RNA level, compared with 78–79% between ALPV-AP and other ALPV isolates. The sequence identity of proteins between ALPV-AP and ALPV-AM was 98–99% for both ORF1 and ORF2, whilst only 85–87% for ORF1 and 91–92% for ORF2 between ALPV-AP and other ALPV isolates. Sequencing of RACE (rapid amplification of cDNA ends) products and cDNA clones of the virus genome revealed sequence variation in the 5’ UTRs and in ORF1, indicating that ALPV may be under strong selection pressure, which could have important biological implications for ALPV host range and infectivity. Our results indicated that ALPV-like viruses infect insects in the order Coleoptera, in addition to the orders Hemiptera and Hymenoptera, and we propose that ALPV isolates be classified as two separate viral species.

INTRODUCTION

The Dicistroviridae is an emerging family of ssRNA viruses known to infect both pest and beneficial arthropods, including aphids, the glassy winged sharpshooter, honey bees and shrimp. Members of this virus family can cause severe pathogenic effects in arthropods (Bonning & Miller, 2010). For instance, Israeli acute paralysis virus and other dicistroviruses have been implicated in colony collapse disorder of honey bees, Apis mellifera (Cox-Foster et al., 2007). Two dicistroviruses are known to infect aphids: aphid lethal paralysis virus (ALPV) and Rhopalosiphum padi virus (RhPV) (Williamson et al., 1988).

ALPV-RhP (GenBank accession number NC_004365.1) was first isolated from the bird cherry-oat aphid, R. padi, in South Africa. The virus genome is an ssRNA of 9812 nt. R. padi infected with ALPV displayed uncoordinated movement and death by paralysis. In addition, ALPV infection also dramatically reduced aphid populations in nature (D’Arcy et al., 1981; Williamson et al., 1988). ALPV-like virus isolates have now been identified from various species of insects and insect predators (Dombrovsky & Luria, 2013; Granberg et al., 2013; Ravoet et al., 2013; Runckel et al., 2011).

An isolate of ALPV, ALPV-AN (GenBank accession number JX480861.1), was identified in a wild population of the milkweed aphid, Aphis nerii, in northern Israel. The ALPV-AN isolate did not show any obvious pathogenic effects to Aphis nerii, but was highly pathogenic to the green peach aphid, Myzus persicae, in laboratory experiments (Dombrovsky & Luria, 2013). Three reports describe ALPV-like viruses in Apis mellifera populations that were being investigated in relation to colony collapse disorder. A partial sequence of an ALPV isolate was reported in Apis mellifera in Brookings, SD, USA (ALPV-Brookings; GenBank accession number HQ871932.1) (Runckel et al., 2011). Near-full-length and partial sequences of ALPV isolates were identified in Apis mellifera from Spain (ALPV-AM; GenBank accession number JX045858.1) and Belgium (ALPV-Belgium) (Granberg et al., 2013; Ravoet et al., 2013). ALPV-like sequences (ALPV-BF; GenBank accession number JQ320375.1) were also reported in faecal samples of two different species

The GenBank/EMBL/DDBJ accession number for aphid lethal paralysis virus isolate AP is KJ817182.

One supplementary table and two supplementary figures are available with the online Supplementary Material.

References

Bonning & Miller, 2010. For instance, Israeli acute paralysis virus and other dicistroviruses have been implicated in colony collapse disorder of honey bees, Apis mellifera (Cox-Foster et al., 2007). Two dicistroviruses are known to infect aphids: aphid lethal paralysis virus (ALPV) and Rhopalosiphum padi virus (RhPV) (Williamson et al., 1988).

ALPV-RhP (GenBank accession number NC_004365.1) was first isolated from the bird cherry-oat aphid, R. padi, in South Africa. The virus genome is an ssRNA of 9812 nt. R. padi infected with ALPV displayed uncoordinated movement and death by paralysis. In addition, ALPV infection also dramatically reduced aphid populations in nature (D’Arcy et al., 1981; Williamson et al., 1988). ALPV-like virus isolates have now been identified from various species of insects and insect predators (Dombrovsky & Luria, 2013; Granberg et al., 2013; Ravoet et al., 2013; Runckel et al., 2011).

An isolate of ALPV, ALPV-AN (GenBank accession number JX480861.1), was identified in a wild population of the milkweed aphid, Aphis nerii, in northern Israel. The ALPV-AN isolate did not show any obvious pathogenic effects to Aphis nerii, but was highly pathogenic to the green peach aphid, Myzus persicae, in laboratory experiments (Dombrovsky & Luria, 2013). Three reports describe ALPV-like viruses in Apis mellifera populations that were being investigated in relation to colony collapse disorder. A partial sequence of an ALPV isolate was reported in Apis mellifera in Brookings, SD, USA (ALPV-Brookings; GenBank accession number HQ871932.1) (Runckel et al., 2011). Near-full-length and partial sequences of ALPV isolates were identified in Apis mellifera from Spain (ALPV-AM; GenBank accession number JX045858.1) and Belgium (ALPV-Belgium) (Granberg et al., 2013; Ravoet et al., 2013). ALPV-like sequences (ALPV-BF; GenBank accession number JQ320375.1) were also reported in faecal samples of two different species
of bat: the Rickett’s big-footed bat, *Myotis ricketti*, and the great roundleaf bat, *Hipposideros armiger* (Ge et al., 2012). Whether these insect predators were infected with the ALPV-like viruses or whether the viral sequences were derived from insects ingested by these bats is unknown, although dicistrovirus infection of bats is unlikely.

The identification of ALPV isolates has been accelerated by the use of next-generation sequencing technologies (Liu et al., 2011). Most of the ALPV and ALPV-like viruses from *Apis mellifera* and bat faecal samples were identified using de novo assembled sequence data. Membrane feeding assays showed that ALPV can infect various species of aphids and whiteflies, *Bemisia tabaci* (Van Munster et al., 2002). The pathology of ALPV in aphid populations and the ability of this virus to infect multiple species of insect pest make it an attractive avenue to explore for insect pest management. However, the presence of related viruses in *Apis mellifera* may be problematic in this regard. It is unclear if the virus causes pathogenic effects in *Apis mellifera*.

Here, we describe the discovery of an isolate of ALPV from the pea aphid, *Acyrthosiphon pisum*, which we designated as ALPV-AP. The genome sequence was assembled de novo from small RNA (sRNA) sequencing data of a laboratory colony of *Acyrthosiphon pisum*. We showed that ALPV-AP was closely related to the ALPV isolate from *Apis mellifera* populations (ALPV-AM). In addition, we discovered another ALPV isolate designated ALPV-DvV from the western corn rootworm, *Diabrotica virgifera virgifera*, a devastating pest of maize. Phylogenetic analysis indicated that ALPV consisted of two major lineages: one composed of ALPV-AP and ALPV-AM, and the second composed of ALPV isolates from other aphid species, bat faeces and *D. v. virgifera*. Geographical location did not seem to play a role in formation of these two lineages. Our results suggested that variation in the UTRs and the ORF1 coding region could drive the evolution of new viruses. Hence, we propose that ALPV isolates belong to two different viral species.

### RESULTS

**Assembly of the ALPV-AP genome**

Around 21 × 10^6 reads of *Acyrthosiphon pisum* sRNA were generated from Illumina sequencing. A variable number of contigs (50–2500 nt) was obtained on assembly depending on the parameters used (hash length k and coverage c). Approximately 1–2% of the contigs (minimum 50 nt) hit the ALPV genome, with the longest being 2388 nt. The ALPV contigs were extracted and reasssembled manually, resulting in three fragments of 5052, 3766 and 1110 nt. The gaps between the fragments were filled by sequencing of the reverse transcription (RT)-PCR fragments with primers designed to cover the gap regions. The assembled and RT-PCR joined ALPV-AP genomic sequences had 9940 nt, which was 128 nt longer than that of the ALPV-RhP sequence (9812 nt). The assembled ALPV-AP genome sequence was confirmed by dideoxy sequencing of RT-PCR fragments. Ninety-eight per cent of the assembled Illumina-derived sequences matched sequences obtained from the Sanger sequencing. The 5’- and 3’-end sequences of ALPV-AP were obtained by RACE (rapid amplification of cDNA ends) and by sequencing full-genome cDNA clones of ALPV-AP (see below). The genome organization of ALPV-AP is shown in Fig. 1(a) and the predicted proteins encoded are listed in Table 1.

**Confirmation of the presence of ALPV-AP in the pea aphid**

Assembly of sRNA sequences of *Acyrthosiphon pisum* revealed a new ALPV-like viral genome. However, no obvious disease symptoms associated with the virus were observed in the aphid population. To confirm infection with ALPV-AP, we purified virions from infected aphids. Electron micrographs showed icosahedral virus particles with a diameter of ~27 nm, similar to the documented size of ALPV particles (Fig. 1b). In addition, the purified virions were subjected to SDS-PAGE and proteins were stained using Coomassie blue. Three major capsid proteins (VP1, VP2 and VP3) were observed, which migrated in good agreement with their predicted molecular masses of 27.03, 27.06 and 29.52 kDa (Fig. 1c). To demonstrate that virions encapsulated the full-length ALPV-AP genome, viral RNA was isolated from purified virions and visualized by electrophoresis on a native agarose gel (Fig. 1d). An RNA band of ~10,000 nt was observed, indicating that the virions encapsidated full-length ALPV-AP RNA.

**Determination of 5’- and 3’-end sequences by RACE**

RACE was performed to determine the 5’- and 3’-end sequences of ALPV-AP RNA using purified RNA as template. A dominant 3’-RACE product was obtained consistently in two different experiments. This product had the same sequence as the assembled 3’-end sequence [excluding the poly(A) tail]. A total of 11 clones of the 5’-RACE product of ALPV-AP showed the same 5’-end sequence. Interestingly, the 5’-RACE sequence was 38 nt shorter (9888 nt) than the sequence assembled from sRNA reads.

**Confirmation of the 5’-end sequence by sequencing full-genome clones of ALPV-AP**

The ALPV-AP genome assembled from sRNA was 9926 nt in length, whilst RACE resulted in 9888 nt with 38 nt fewer at the 5’ end. To confirm the 5’-end sequences of ALPV-AP, we generated cDNA from viral RNA isolated from purified ALPV-AP virions. Full-length genomes were amplified by two-step RT-PCR using primers derived either from the ends of the ALPV-AP genome (based on the 5’- and 3’-RACE results) or from the ALPV-RhP genome (Table S1, available in the online Supplementary Material). Fourteen PCR-positive clones with inserts
9.5 kb were obtained. Among the 14 clones, 12 were from the cDNA generated using the primer derived from 5' -RACE results. Only two putative full-length clones were obtained using the ALPV-RhP 5' primer. The 5' UTRs of the 14 clones were sequenced. Alignment of the 5' -end sequences of the 14 clones and the 5' -end sequences of the Velvet-assembled and RACE-derived sequences suggested that ALPV-AP had the same 5' -end sequence as the documented ALPV-RhP isolate (Fig. 2). However, our results indicated that variation may exist. Sequencing of the cloned cDNAs also confirmed that the 38 nt sequence that was not found by RACE was indeed part of the ALPV-AP genome. The total genomic RNA of ALPV-AP was 9940 nt [excluding the poly(A) tail], 128 nt more than that reported for ALPV-RhP (the longest ALPV-like virus discovered so far).

### Comparison of sequences of ALPV isolates

The genome organization of ALPV-AP is shown in Fig. 1(a). The non-structural polyprotein (ORF1) had 2037 aa, which was 2 aa more than that of ALPV-RhP, ALPV-AM and ALPV-BF, but 1 aa less than that of ALPV-AN. The structural protein precursor had 809 aa, which was the same as that of ALPV-AM, but 2 aa more than that of ALPV-RhP and 1 aa more than that of ALPV-AN and ALPV-BF. Numbers of amino acids of each peptide and molecular masses of the peptides are summarized in Table 1.

Comparisons of ALPV-AP RNA and protein sequences with known ALPV sequences are summarized in Tables 2 and 3. The overall RNA identity between ALPV-AP and the other aphid ALPV isolates and bat faecal-derived ALPV was 78–79 %. Interestingly, the RNA identity between

### Table 1. Summary of predicted ALPV-AP-encoded proteins

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Size (aa)</th>
<th>Molecular mass (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1 (NS polyprotein)</td>
<td>2037</td>
<td>231.46</td>
</tr>
<tr>
<td>ORF2 (VP precursor)</td>
<td>809</td>
<td>89.58</td>
</tr>
<tr>
<td>VP1</td>
<td>242</td>
<td>27.03</td>
</tr>
<tr>
<td>VP2</td>
<td>241</td>
<td>27.06</td>
</tr>
<tr>
<td>VP3</td>
<td>268</td>
<td>29.52</td>
</tr>
<tr>
<td>VP4</td>
<td>58</td>
<td>6.01</td>
</tr>
</tbody>
</table>

Fig. 1. Genome organization and molecular characteristics of ALPV-AP. (a) Genome organization of ALPV-AP. The virus encodes two polyproteins: the non-structural and the capsid protein (VP). Two internal ribosomal entry sites (IRESs) are present at the 5' end and the intergenic region (IGR) between ORF1 and ORF2. The virus genome is capped at the 5' end with VPg (‘viral protein genome-linked’), whilst the 3' end has a poly(A) tail. RdRp, RNA-dependent RNA polymerase. (b) Electron micrographs of purified icosahedral virions of ~27 nm. (c) SDS-PAGE gel showing the three major virus capsid proteins of ALPV-AP. (d) ALPV-AP RNA (2 µg) on a 1 % native agarose gel isolated from purified virions. The genomic RNA runs just below the 10 kb marker, consistent with the 9940 nt genome.
ALPV-AP and ALPV-AM-Spain (lacking parts of the 5′- and 3′-UTR sequences) was 88%, ~10% higher than that of the other ALPV isolates. The intergenic region internal ribosomal entry site (IGR IRES) sequences among all ALPV isolates were highly conserved (97–99% identity). The RNA sequences of the coding regions were relatively conserved with ~81% homology between ALPV-AP and other isolates. Again, sequence identity between ALPV-AP

![Alignment of the 5′ ends of ALPV-AP cDNA clones, assembled ALPV-AP and RACE results.](image)

Table 2. Nucleotide sequence identity (%) between different ALPV isolates

<table>
<thead>
<tr>
<th></th>
<th>AP</th>
<th>AN</th>
<th>RhP</th>
<th>BF</th>
<th>AM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-length</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>78.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RhP</td>
<td>78.3</td>
<td></td>
<td>89.3</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>BF</td>
<td>78.7</td>
<td></td>
<td>89.3</td>
<td>88.7</td>
<td>100</td>
</tr>
<tr>
<td>AM</td>
<td>87.7</td>
<td>78.4</td>
<td>77.7</td>
<td>78.5</td>
<td>100</td>
</tr>
<tr>
<td><strong>5′ UTR/3′ UTR/IGR IRES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>63.4/60.1/97.3</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RhP</td>
<td>63.4/59.3/97.8</td>
<td>89.6/91.9/98.4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>62.2/60.1/96.8</td>
<td>92.7/95.2/97.3</td>
<td>95.5/94.1/98.9</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>ND/ND/99.4</td>
<td>ND/ND/96.8</td>
<td>ND/ND/98.4</td>
<td>ND/ND/97.3</td>
<td>100</td>
</tr>
<tr>
<td><strong>ORF1/ORF2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>81.1/83.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RhP</td>
<td>80.8/81.4</td>
<td>90.7/84.5</td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>BF</td>
<td>81.0/83.1</td>
<td>87.7/90.2</td>
<td>88.8/85.4</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>AM</td>
<td>95.9/92.2</td>
<td>81.1/83.7</td>
<td>80.9/81.5</td>
<td>81.4/83.7</td>
<td>100</td>
</tr>
</tbody>
</table>

ND, Not determined.
and ALPV-AM-Spain was much higher (92–96%) in the ORF1 and ORF2 relative to other isolates.

Comparison of polyprotein sequences also showed that ALPV-AP had higher sequence identity with ALPV-AM (Table 3). Protein identities between these two isolates were 98.4% for ORF1 and 98.5% for ORF2, whilst identities of ALPV-AP and three other ALPV isolates were 87% (ORF1) and 91–92% (ORF2) (Table 3). Although only a partial sequence was available for the ALPV-AM-Brookings, sequence comparison of the Brookings isolate with the corresponding sequences of other ALPV isolates showed that the Brookings isolate was highly similar to ALPV-AP and the ALPV-AM-Spain isolate (data not shown). Indeed, phylogenetic analysis showed that the ALPV isolates formed two groups, with ALPV-AP and the ALPV-AM isolates grouping together, separate from other ALPV isolates (Fig. 3).

### Table 3. Protein sequence identity (%) between ALPV isolates for ORF1 (non-structural polyprotein)/ORF2 (structural polyprotein)

<table>
<thead>
<tr>
<th></th>
<th>AP</th>
<th>AN</th>
<th>RhP</th>
<th>BF</th>
<th>AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>AN</td>
<td>87.5/92.0</td>
<td>100</td>
<td>95.6/93.6</td>
<td>100</td>
<td>95.5/94.3</td>
</tr>
<tr>
<td>RhP</td>
<td>87.5/91.3</td>
<td>94.0/97.4</td>
<td>95.5/94.3</td>
<td>87.2/90.7</td>
<td>87.8/91.9</td>
</tr>
<tr>
<td>BF</td>
<td>98.4/98.5</td>
<td>87.7/92.9</td>
<td>87.2/90.7</td>
<td>87.8/91.9</td>
<td>100</td>
</tr>
</tbody>
</table>

ALPV-AP RNA 5' and 3' UTRs

The 5' UTR of ALPV-AP was aligned to the known ALPV-RhP 5'-UTR sequence (Fig. 4a) and a consensus secondary structure of ALPV (Fig. 4b) was constructed by using the LocARNa folding program. Comparison of the 5’ and 3’ UTRs between ALPV-AP and ALPV-RhP isolates showed that multiple insertions occurred in several regions of ALPV-AP UTRs, which explained why the genomic sequence of ALPV-AP was longer compared with the published ALPV-RhP sequence. Indeed, the 5’ and 3’ UTRs showed much less similarity between ALPV-AP and other ALPV isolates. There was 63% sequence identity between ALPV-AP and other ALPV, and ALPVs at the 5’ UTR, whereas the other isolates shared 90–95% identity at the 5’ UTR. Lower similarity was also observed at the 3’ UTR between ALPV-AP and other ALPV sequences (Table 2). These results indicated that the 5’ and 3’ UTRs of ALPV were less conserved than other regions of the virus genome. Although significant sequence variation was observed in the 5’ and 3’ UTRs, multiple RNA sequence folding results suggested that conserved RNA sequences could be identified among the ALPV 5’ UTR. A consensus RNA secondary structure was predicted by LocARNa (Fig. 4a). The resulting structure showed that most of the inserted sequence in the 5’ end of ALPV-AP was at the loop region and therefore may not affect the common ALPV 5’-UTR structure (Fig. 4b).

Fig. 3. Phylogenetic tree of dicistroviruses. ALPV genomic sequences were downloaded from GenBank (see accession numbers in text). The sequence of ALPV-AM-Spain is only near full length. ALPV-AM-Brookings is only a partial sequence (4175 nt). The sequences were aligned using CLUSTAL W. The phylogenetic tree was reconstructed with the maximum-likelihood method with 500 bootstrap replications in MEGA6.0. (ALPV-DvV was not included due to limited sequence availability.)
Sequence analysis of ALPV-AP full-length clones

To assess variations in the ALPV-AP sequence in the coding region, we sequenced 14 full-genomic clones. Three of the 14 cDNA clones had a point mutation that truncated ORF1 upstream of the GDD active site. The 3’ UTR of two of the clones was shorter (by 10 and 212 nt) than that of ALPV-AP, which might also have affected transcription of the viral RNA. In total, five clones (36%) contain lethal mutations in their genomic RNA, three had a single nucleotide insertion,
which disrupted ORF1 via a premature stop codon, whilst two clones had short fragment deletions that truncated ORF1. Interestingly, none of the 14 clones had mutations that might have disrupted translation of structural proteins from ORF2. As each genome was PCR-amplified in one piece, these results represented a snapshot of sequence diversity within the viral population.

**Discovery of ALPV in western corn rootworm (ALPV-DvV)**

We identified an RNA fragment (Fragment 1) of 412 nt derived from an ALPV-like virus from an expressed sequence tag (EST) library of *D. v. virgifera* (GenBank accession number EW772779.1). In addition, we isolated and sequenced mRNA from *D. v. virgifera* larvae collected...
from Kondoros, Hungary. Two contigs of 329 and 754 nt (Fragments 2 and 3, respectively) were derived from ALPV-like viruses. Sequence comparison showed that all fragments derived from ALPV-DvV had high sequence identities to ALPV-AN and ALPV-RhP, particularly for Fragment 3, which was partially derived from the 3′ UTR (Table 4). Only 66% sequence identity was found between ALPV-DvV and ALPV-AP isolates in this region. In comparison, the ALPV-DvV fragments were 91–93% identical to ALPV-AN, ALPV-RhP and ALPV-BF, indicating that ALPV-DvV was closely related to these isolates, and differed from the ALPV-AP and ALPV-AM isolates (Table 4).

**DISCUSSION**

We described the identification and genome characterization of ALPV-AP, a novel isolate of ALPV with a longer genome (9940 nt) than other isolates reported so far. We also identified partial sequences of ALPV-DvV had high sequence identities to ALPV-AN and ALPV-RhP, particularly for Fragment 3, which was partially derived from the 3′ UTR (Table 4). Only 66% sequence identity was found between ALPV-DvV and ALPV-AP isolates in this region. In comparison, the ALPV-DvV fragments were 91–93% identical to ALPV-AN, ALPV-RhP and ALPV-BF, indicating that ALPV-DvV was closely related to these isolates, and differed from the ALPV-AP and ALPV-AM isolates (Table 4).

Sequence comparison of ALPV-AP with other ALPV isolates shows major nucleotide sequence variations (>60%) in the 5′ and 3′ UTRs (Table 2). Furthermore, variations were also observed within various clones of ALPV-AP, although some changes may have resulted from PCR errors. However, RNA structures in the UTRs remain conserved (Fig. 4), suggesting that the UTR structure plays important functions. Secondary structures in the UTRs of RNA viruses can play major roles in the regulation of virus replication and translation. In addition, the UTRs may contribute to mediation of the virus-host immune response via the RNA interference (RNAi) pathway. The UTR of West Nile virus and dengue virus (DENV) enable evasion from the host RNAi-based antiviral immune response. Whilst some isolates may trigger the RNAi response, others could evade and replicate successfully in the host (Hyde et al., 2014; Schnettler et al., 2012). Sequence variation in the UTRs may also correlate with pathogenicity of the virus isolates. For instance, structural differences in the 5′ UTR of DENV correlate with disease severity in humans (Leitmeyer et al., 1999). The driving forces that caused variation in different ALPV isolates are currently unknown.

The concept of viral quasispecies with a high rate of mutation resulting in many natural variants may be at
play in ALPV populations (Domingo et al., 1998). In an environment where the virus is constantly being targeted for suppression, natural variants would enable a small proportion of the virus genomes to replicate and to persist in the host with non-infectious genomes serving as decoys for the host immune system, and providing for greater adaptability under changing conditions. With the increase in metagenomic sequencing and associated increase in virus discovery, more novel isolates and related viruses are being identified (Li et al., 2011). These new data will help our understanding of virus evolution, and our understanding of the determinants of pathogenicity and host specificity. Previous regions of the virus genome that have not been investigated in depth, such as the UTRs, could emerge as key players in host–virus molecular interactions.

Current classification of ALPV is based solely on the similarity of the VP capsid proteins. VP sequence identity of <90% is the criterion used for designating a new dicistrovirus (Chen et al., 2012). The similarity of all ALPV VP sequences identified so far is >90%, irrespective of the insect host of the ALPV isolates (Table 3). Hence, all of the ALPV sequences identified represent different isolates of ALPV and not new virus species. However, sequence comparison between ALPV isolates demonstrates that sequence diversity is much greater at the RNA level. For instance, the ALPV-AP genome is only 78.3% identical to that of ALPV-RhP, isolated from R. padi. Although ALPV-AP is also isolated from aphids, phylogenetic analysis using the viral RNA sequence shows clearly that ALPV-AP has a greater sequence similarity to ALPV isolates from Apis mellifera. In addition to sequence variation at the RNA level, a difference of pathogenicity is also observed for the different isolates. ALPV-RhP was shown to cause disease symptoms in R. padi in the field (Williamson et al., 1988). ALPV-AN also caused pathogenicity to M. persicae (Dombrovsky & Luria, 2013). However, no acute pathogenicity was observed from ALPV-AP infection of Acyrthosiphon pisum under ideal rearing conditions. These observations indicate that the two ALPV ‘isolates’ may have variable biological properties, e.g. ability to cause acute versus chronic infection of the host. The fact that there are two phylogenetically distinguishable genomic branches based on the entire genome sequence of ALPV (Fig. 3) and protein sequences (Fig. S1a, b) suggests that the current ALPV classification criterion may not be sufficient for ALPV-like viruses.

As discussed above, the UTRs may be involved in regulation of RNA replication, translation, disease severity and host immune responses. Hence, these regions may be under much stronger selection pressure to meet changing environments. Lower RNA sequence identity in UTRs and ORF1 regions may reflect evolution of ALPV to meet environmental challenges. Therefore, we suggest that these regions be taken into account for species determination similar to classification of other virus groups. For example, African cassava mosaic virus and East African cassava mosaic virus are two different begomoviruses, although their VP protein identity is up to 93% (Oteng-Frimpong et al., 2012). Similarly, reliance on polyhedrin, the primary structural protein of baculoviruses, was not reliable for predicting viral relatedness (Jehe et al., 2006). Considering greater variation in the ALPV UTR and ORF1 regions, the criteria for ALPV taxonomy should be revisited. We propose that ALPV isolates be classified into two species with ALPV-AP as a new species, which has two isolates: ALPV-AP and ALPV-AM.

**METHODS**

**Insects.** Pea aphids, *Acyrthosiphon pisum*, were purchased from Berkshire Biological and were raised on broad bean, *Vicia faba*, in a growth chamber at 24°C with a 12 h–12 h (light–dark) cycle. Western corn rootworm, *D. v. virgifera*, larvae were collected from corn fields near Kondoros, Hungary, and stored in 90% ethanol before storage at −80°C and shipping to Iowa State University for RNA isolation.

**Total RNA extraction.** All instars of *Acyrthosiphon pisum* were collected and used for RNA isolation. Approximately 50–60 aphids and 100 *D. v. virgifera* larvae were homogenized in 1 ml TRIZOL Reagent (Life Technologies) with a pestle in a 1.5 ml microcentrifuge tube. Procedures for RNA isolation followed the manufacturer’s instructions. RNA was precipitated overnight in 2-propanol at −80°C for optimal recovery of sRNAs. Total RNA was resuspended in 30 μl nuclease-free water. RNA was quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and the RNA quality assessed using a Bioanalyzer (Agilent).

**sRNA and mRNA sequencing.** sRNA from *Acyrthosiphon pisum* and mRNA from *D. v. virgifera* were isolated from total RNA. cDNA libraries of sRNA and mRNA libraries were constructed using the TruSeq Small RNA Sample Prep and TruSeq RNA Sample Prep v2 kits (Illumina), respectively, according to the manufacturer’s instructions. The sRNAs and mRNAs were sequenced for 50 and 100 cycles, respectively, using an Illumina GA II platform. sRNA and mRNA isolation, sequencing library preparation, and high-throughput sequencing were conducted at the Iowa State University DNA Facility.

**Assembly of RNA and sRNA reads.** The Illumina sequencing reads were trimmed with FASTX-Toolkit to remove adaptors. The sRNA reads were then assembled using the Velvet assembler (Zerbino & Birney, 2008) with multiple *k* and *c* parameters for optimal results (Liu et al., 2012), and RNA reads were assembled using the Trinity assembler (Haas et al., 2013). The assembled contigs were then used for BLAST searches to identify virus-derived contigs. BLAST analysis was also conducted with assembled contigs against the ALPV genome to identify potential ALPV-derived fragments. Local BLAST was performed using BioEdit software. The contigs that hit ALPV genomes were extracted manually for further manual assembly using BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The library for *D. v. virgifera* was prepared from a normalized and subtracted adult female head cDNA library of clone ST020028B10H10 (Bonaldo et al., 1996).

**Purification of ALPV-AP virus particles from *Acyrthosiphon pisum*.** *Acyrthosiphon pisum* (−3 g) were ground in liquid nitrogen using a pre-cooled mortar and pestle. The powdered aphids were transferred to a 30 ml centrifuge tube on ice. An aliquot of 15 ml 0.01 M sodium phosphate buffer, pH 7, was added to the tube and vortexed briefly. Then, 7.5 ml chloroform (0.5 vol. buffer) was added to the tube. The tube was shaken vigorously and incubated on ice for 5 min. The tube was then centrifuged at 7649 g at 4°C for 25 min in
an SS34 rotor (Sorvall). The resulting supernatant was passed through a syringe filter sterilizer (0.45 μm) into a 30 ml centrifuge tube. An aliquot of 5 ml 0.01 M sodium phosphate buffer was added to the filtered supernatant. The supernatant was once again centrifuged at 7649 g at 4 °C for 25 min in an SS34 rotor (Sorvall). The resulting supernatant was then transferred to a 30 ml ultracentrifuge tube and centrifuged at 164,243 g at 4 °C for 2.5 h using a 70 Ti rotor. The supernatant was removed and 5 ml 0.01 M sodium phosphate buffer was added to the pellet. The tube was covered with Parafilm and placed on ice at 4 °C overnight with shaking. The resuspended virus mixture was transferred into a 1.5 ml centrifuge tube and centrifuged at 9279 g at 4 °C for 5 min. The resulting supernatant was transferred into a new tube and the final volume was adjusted to 5.5 ml using 0.01 M sodium phosphate buffer. The supernatant was added to the top layer of a 2.5 ml 30% sucrose cushion in an ultracentrifuge tube. The tube was centrifuged at 169,357 g at 4 °C for 3 h using a 70.1 Ti rotor (Beckman). The resulting liquid was removed and the pellet resuspended overnight at 4 °C as described previously. The stock of purified virus was stored at −20 °C.

Transmission electron microscopy. Purified virions (10 μl) were placed on a carbon film grid. The grid was negatively stained with 2% aqueous uranyl acetate for 30 s. The virus particles were visualized using a JEOL 2100 scanning/transmission electron microscope using standard procedures.

Virus genome characteristics. Purified ALPV-AP virions were added to 5 × SDS loading buffer [10% (v/v) SDS, 10 mM DTT, 20% v/v glycerol, 0.2 M Tris/HCl, pH 6.8, and 0.05% (w/v) bromophenol blue]. The sample was boiled and loaded onto a 5% stacking, 12% resolving SDS-polyacrylamide gel along with 8× Protean Plus Protein All Blue ladder (Bio-Rad). The gel was stained with Coomassie blue to visualize the protein bands. Total RNA was extracted from purified ALPV-AP virions using TRizol reagent (Life Technologies). ALPV-AP RNA (2 μg) was loaded onto a 1% native agarose gel along with 5 μl of a 0.5–10 kb RNA ladder (Invitrogen). The gel was stained with ethidium bromide for visualization of RNA.

Sequence confirmation by RT-PCR, and 5’ and 3’ RACE. To confirm the in silico assembled ALPV-AP sequences and to fill the gaps of the resulting genomic sequences, primers (Table S1) were designed to amplify different regions of the ALPV-AP genome based on the assembled viral sequences. Total RNA extracted from purified ALPV-AP virions was used as template for RT-PCR carried out using a One-Step RT-PCR kit (Qiagen). The resulting PCR products were isolated and purified from agarose gels using a QIAquick gel extraction kit (Qiagen) and sequenced by Sanger sequencing. To determine the 5’- and 3’-terminal sequences, RACE experiments were carried out using a SMARTer RACE cDNA amplification kit (Clontech) with 1 μg viral RNA as template. The PCR was prepared using an Advantage 2 PCR kit (Clontech) according to the manufacturer’s recommendations with primers listed in Table S1. PCR was carried out using a Bio-Rad MyCycler Thermal Cycler with 25 cycles of 94 °C for 30 s, 68 °C for 30 s and 72 °C for 3 min. The 5’-end RACE products were ligated into pGEM-T Easy (Promega) and plasmids were transformed into TOP10 competent cells. Plasmid DNA was isolated and sequenced. The 3’-end RACE product was gel purified and sequenced as described above. Negative controls (no template) were included for RACE and RT-PCR experiments.

Cloning of ALPV-AP. The full-length viral genome of ALPV-AP was amplified from viral RNA extracted from particles using TRizol reagent as described previously. To synthesize the cDNA, 40 ng RNA was mixed with T30-ALPV-AP reverse primer (Table S1) and dNTPs followed by a denaturing step (65 °C for 5 min). Reaction buffer, DTT, RNAsin (Promega) and SuperScript III retrotranscriptase (Life Technologies) were then added to the mixture to a final volume of 20 μl. The reaction was incubated in a thermocycler following a step programme: 37 °C for 20 min, 42 °C for 20 min, 47 °C for 20 min, 51 °C for 20 min, 55 °C for 20 min and 70 °C for 15 min. PCR amplification was conducted using Platinum Taq DNA Polymerase High Fidelity (Life Technologies) following the manufacturer’s recommendations (for primers AatII-T7-ALPV-AP designed from the 5’-RACE results, and ALS30-R) with a ramp programme: 95 °C for 3 min, 35 cycles of [94 °C for 45 s, 47 °C (increasing 0.1 °C per cycle) for 45 s, 68 °C for 10 min] and 68 °C for 15 min. Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used with GC buffer and primers designed based on ALPV-Rhp (AatII-T7-ALPV-Rhp and ALS30-R or AatII-T7-ALPV and AlSKpnl-R, Table S1) following the kit protocol using a similar programme: 98 °C for 30 s, 35 cycles of [98 °C for 15 s, 47 °C (increasing 0.2 °C per cycle) for 30 s, 72 °C for 6 min] and 72 °C for 10 min. Amplicons were cleaned from the agarose gel using Ultrafree-DA Centrifugal Filter Units (Millipore) and cloned using a BigEasy Long PCR Cloning kit (Lucigen) following the manufacturer’s protocol. Positive colonies were selected by PCR screening with primers ALS6701F and ALS7199R (Table S1), and by restriction digestion analysis. Clones with inserts >9.5 kb were fully sequenced. Sequence assembly was performed using DNADynamo software (BlueTractorSoftware). Alignment and pairwise comparison of resulting clones was done using CLC Main Workbench software (CLC bio).

Bioinformatics analysis. Sequence alignment, similarity, identity and homology analyses of RNA and proteins were performed using BioEdit software. CLUSTALW was used for multiple sequence alignments (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Phylogenetic trees were reconstructed using the test maximum-likelihood tree method with 500 bootstrap replications in MEGA software (version 6.0) (Tamura et al., 2013). The RNA secondary structure of the ALPV 5’ UTR from multi-aligned sequences was predicted by LocARNa using the default setting (Will et al., 2012).

ACKNOWLEDGEMENTS

The authors would like to thank Dr John VanDyk for IT assistance, and Dr Stefan Toepfer (CABI, c/o Plant Protection and Soil Conservation Directorate, Hodmezovasarhely, Hungary) and Dr Thomas W. Sappington (USDA-ARS, CICGRU, Genetics Laboratory, ISU, Ames, IA, USA) for D. v. virgifera sample collection. This work was funded by the North Central Branch Soybean Research Program, the Iowa State University Plant Sciences Institute Virus–Insect Interactions Initiative and the Monsanto Company Corn Rootworm Knowledge Program.

REFERENCES


