**Aphid small RNAs and viruses**


by


**Diveena Vijayendran**


A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY


Major: Genetics

Program of Study Committee:
Bryony C. Bonning, Major Professor
Wyatt Allen Miller
Lyric Bartholomay
Russell Jurenka
Thomas Baum


Iowa State University

Ames, Iowa

2014

# TABLE OF CONTENTS

iv

**ABSTRACT**

Aphids are important agricultural pests with about 250 species known to cause damage by feeding on plant phloem and transmitting plant viruses. Current management relies primarily on the application of chemical insecticides, which can be deleterious to the environment, and the use of aphid resistant cultivars hampered by the presence of resistant aphid biotypes. We aim to develop novel biotechnology based approaches to better manage pestiferous aphids. To this end, we (i) identified and characterized aphid viruses, and (ii) analyzed a class of small RNA, microRNAs specific to aphids. We used transcriptome and small RNA sequencing datasets to identify viruses that naturally infect populations of various aphid species. We identified and characterized a new isolate of *Aphid lethal paralysis virus*, named ALPV-Ap (Dicistroviridae) from the pea aphid, *Acyrthosiphon pisum* and a novel virus from the soybean aphid, *Aphis glycines*, named *Aphis glycines virus* (AGV; Unclassified). ALPV-Ap, which has a longer genome than other ALPV isolates, is phylogenetically closely related to ALPV isolates from honeybees rather than from aphids. ALPV-Ap localizes to the aphid midgut and is not vertically transmitted from adult to nymph. AGV has a capsid protein similar to those of plant viruses while the RNA-dependent RNA polymerase is more closely related to those of insect viruses. Remarkably, AGV is 100% vertically transmitted. Analysis of small RNA datasets from various aphid species showed the presence of over 100 microRNAs. In addition to detection of many evolutionarily conserved miRNAs, a subset of 12 aphid-specific miRNA was identified. The wealth of information obtained from sequencing datasets will allow for investigation of virus- and miRNAs-based management of aphid populations for reduced damage and increased crop yield worldwide.

## CHAPTER 1

## INTRODUCTION

This chapter is written to provide an overview of research related to certain aspects of aphids, small RNA and viruses. In the first part of the chapter, the aphid as an important agricultural pest and the current management strategies used for aphid control are addressed. In the second part of the chapter, the biological process of RNA interference (RNAi) with a focus on insects is outlined. In the third part, the use of Next generation sequencing technology for virus discovery is discussed. The fourth section contains descriptions and discussion of viruses known to infect aphids. In the final part of this introduction, antiviral immunity in aphids and the potential for application of viruses in aphid management are described.

**Aphids as pests of agricultural crops**

Aphids are soft-bodied arthropods in the order Hemiptera. Approximately 4,400 aphid species have been identified and 250 species of aphids are classified as serious pests of economic importance (1). Aphids use modified mouthparts known as stylets to pierce through the plant epidermis toward the sieve tubes to feed on plant phloem (2). Aphids are known to affect approximately 25% of plant species in temperate regions. Crops affected by these pests include, maize, potatoes, barley and wheat (3). Aphids cause damage by (i) feeding on plant phloem and depriving the host of essential nutrients (4), (ii) aphid honeydew secretions that promote the growth of sooty mold fungus on plant surfaces (5) and (iii) vectoring and transmitting economically damaging plant viruses (6).

Aphids are among the most important insect vectors of plant viruses. Aphids are known to transmit 275 out of the 600 insect-borne viruses (7). Plant virus transmission by aphids is described as "nonpersistent", "semi-persistent" or "persistent". In non-persistent transmission, the plant virus briefly attaches to the aphid stylet during which time, the aphid will move to a new host plant to inoculate the plant with the virus (8, 9). In semi-persistent transmission, the plant virus is acquired and retained in the foregut of the aphid before inoculation into a new host plant. The acquisition time is longer for semi-persistent transmission compared to non-persistent transmission of plant viruses (10) . Persistently transmitted plant viruses that are ingested by the aphids move through the gut epithelium, enter the aphid hemocoel and finally enter the salivary glands from which the virus is secreted into a new plant upon aphid feeding (11). Persistently transmitted viruses are further characterized into propagative and non-propagative based on whether the virus replicates in the insect vector (11). Among the most important plant viruses transmitted by aphids are *Barley yellow dwarf virus* (BYDV), *Beet mild yellowing virus* (BMYV), *Cucumber mosaic virus* (CMV), and *Turnip mosaic virus* (TuMV) (12).

**Current aphid management strategies**

Several methods have been employed to manage populations of aphids. The most commonly used method for aphid management is the application of chemical insecticides. The application of chemical insecticides has increased 130-fold in soybean fields to control populations of the soybean aphid, *Aphis glycines*, an invasive pest in North America (13). Insecticides such as neonicotinoids, imidacloprid and dimethoate while effective against aphids, are damaging to the natural enemies of aphids and other non-target organisms in the field (1). In addition, the intense selection pressure has led to

the development of insecticide resistant aphid populations (14). The use of natural enemies of aphids including pirate bug, *Orius insidious*, Asian lady beetle, *Harmonia axyridis* and brown lacewing, *Micromus tasmaniae* has some effectiveness in aphid control (15).

Another widely adopted management strategy for aphid management is the use of resistant host plants such as *Rag* 1 and *Rag* 2 soybean cultivars for management of *A. glycines* and *Dn 4* and *Dn 7* wheat cultivars for management of Russian wheat aphid, *Diuraphis noxia* (16-18). The initial use of the *Rag* and *Dn* cultivars showed promise in aphid management; however, resistant aphid biotypes have been identified in natural field populations (19-22).

**Potential future aphid management strategies**

Novel biotechnology approaches are at the forefront for pest management approaches. One widely used technology for insect pest suppression is the expression of toxins from the bacterium *Bacillus thuringiensis* (Bt) in transgenic crops. While the Bt toxins are effective against lepidopteran and coleopteran insects such as the western corn rootworm, *Diabrotica virgifera virgifera*, there is little efficacy against hemipteran insects (23). Efforts have been made to adapt existing Bt toxins to be more effective against aphids and this may be an option for aphid management in the near future (24). The ability of plant virus capsid proteins to effectively cross the aphid gut barrier was manipulated to deliver an insecticidal neurotoxin to aphids. The capsid protein of an aphid transmitted luteovirus, *Pea enation mosaic virus* (PEMV) was fused to a spider-derived peptide toxin, Hv1a. The capsid protein of the virus functions to deliver the toxin

across the gut barrier to the aphid hemocoel to be functionally active and kill the aphid. Transgenic *Arabidopsis* plants expressing the fusion protein showed toxicity and induced death by paralysis in the green-peach aphid, *Myzus persicae* (25).

The natural defense response of plants against insects has been explored for enhanced effectiveness against aphids. The plant - produced carbohydrate binding proteins known as lectins have been investigated for potential use in insect pest management. Ingested plant lectins can negatively impact physiological processes in insects. The plant lectin concanavalin A was shown to have toxic effects and to inhibit the growth of pea aphids, *Acyrthosiphon pisum* in feeding bioassays with artificial diet (26). A biotechnology approach for using plant lectins for pest management is to fuse a plant lectin such as the snowdrop lectin, GNA (*Galanthus nivalis agglutinin*) to an insect toxin to be delivered into the insect hemolymph through binding of a glycoprotein receptor on the gut (27, 28) . The plant lectin, GNA was fused to an insecticidal spider venom neurotoxin, SF1 (*Segestria florentina toxin 1*) to induce aphicidal activity in the *M. persicae*. Feeding bioassays showed ~50% aphid mortality after 2 weeks and a significant negative impact on the development and fecundity of the aphid (29).

The gene silencing approach using RNA interference (RNAi) has been explored for aphid management as well as for functional gene studies. Multiple approaches have been used for introduction of exogenous silencing RNA into aphids including microinjection (introduction of silencing RNA into the hemolymph that bypasses the aphid gut barrier), feeding on treated artificial diet and on transgenic plants that expresses silencing RNAs (30-39). Table 1 summarizes research on aphid gene silencing mediated by exogenous short-interfering RNA (siRNA) or double-stranded RNA (dsRNA). A

broad range of responses to gene silencing in aphids was noted with some experiments resulting in significant gene knockdown while others had no observable negative impact on the aphid (35, 39). High variability in the results has also been observed for silencing of the same target gene when carried out by different research groups (35, 36, 39). The potential practical application of the silencing effect for aphid control was primarily tested using *M. persicae* in transgenic plants expressing double-stranded RNA. Aphid mortality does not appear to be severely impacted; however negative effects on fecundity and life span have been observed (34, 37).

**Table1.** Summary of gene silencing experiments conducted with aphids. N/A, not applicable.

| Aphid species | Silencing molecule | Dose/ Concentration | Target genes | Silencing effects | Ref |
|---|---|---|---|---|---|
| **Microinjection** | | | | | |
| Pea aphid (*Acyrthosiphon pisum*) | siRNA | 50 ng | Saliva transcript (Coo2) | 50% mortality at Day 3 | (35, 39) |
| | dsRNA | 276 ng | Calreticulin Cathepsin L (Cat L) | 30 – 40% gene knockdown at Day 5 | (30) |
| | | 80 ng | Ecdysone receptor (EcR) Ultraspiracle (USP) | No effects on survival or developmental | (39) |
| **Feeding on artificial diet** | | | | | |
| Pea aphid (*Acyrthosiphon pisum)* | dsRNA | 1 µg/µL | Putative aquaporin gene (ApAQP1) | Reduction of 2-fold in 24 hours. Transient effect of silencing No significant mortality | (33) |
| | | 0 – 3 µg/µL | Vacuolar ATPase subunit A (vATPase) | 30% gene knockdown | (36, 39) |
| | | 200 ng/µL | Ecdysone receptor (EcR) | No effects on survival or developmental | (39) |
| Grain aphid (*Sitobion avenae*) | dsRNA | 7.5 ng/µL | 5 candidate genes | 3 genes significantly reduced at Day 6 2 genes completely silenced at Day 8 | (38) |
| **Feeding on transgenic plant** | | | | | |
| Green-peach aphid (*Myzus persicae*) | dsRNA | N/A | *Myzus persicae* Coo2 (MpCoo2) Receptor for activated kinase C (Rack-1) | 60% silencing Reduced fecundity No significant mortality | (31) |
| | | | *Myzus persicae* serine protease (MpSP) | No significant mortality Reduced fecundity Reduced parthenogenetic population | (37) |
| | | | *Myzus persicae* hunchback (Mphb) | 31% reduction on Day 7 | (34) |
| Grain aphid (*Sitobion avenae*) | dsRNA | N/A | Carboxylesterase (CbE E4) | 30-60% reduction of expression | (32) |

**RNA interference**

The effective use of RNA interference (RNAi) for insect pest suppression has been demonstrated for the corn rootworm, *Diabrotica virgifera virgifera* and cotton bollworm, *Helicoverpa armigera* (40, 41). Investigation of the use of RNAi for suppression of various other pest species is ongoing, including for use against aphids.

*General mechanism of RNA interference*

This section has been published in a review paper by Vijayendran D, Airs PM, Dolezal K and Bonning BC (42).

RNA interference (RNAi) is an important biological process for silencing nucleic acid in eukaryotic cells. The RNAi pathway is trigged by the presence of endogenous or exogenous double stranded RNA (dsRNA) in the cytoplasm of a cell (43). The dsRNA can be derived from but is not limited to, (i) stem-loop RNA structures such as precursor microRNA (pre-miRNA), (ii) replication intermediates of viruses, (iii) silencing RNAs that is experimentally introduced (soaking, feeding and microinjection) or (iv) transgenes. The long dsRNA is recognized and cleaved into smaller duplexes (21-23nt) of RNA known as short interfering RNA (siRNA) by the protein Dicer, an RNase III endoribonuclease (44). One of the siRNA duplex strands (guide strand) is then 2'-O-methylated at the 3' terminus and loaded onto a multi-protein RNA- induced silencing complex (RISC) with the help of R2D2 protein containing dsRNA binding motifs (45). The RISC complex consists of an Argonaute protein which is responsible for messenger RNA (mRNA) target recognition and cleavage (46). Any mRNA in the cytoplasm that has perfect complementation to the guide siRNA will be cleaved. In the case of imperfect base pairing or base pairing at the "seed" region which is a 2-8 base homology at the 3' terminal of the mRNA,

translational repression of the mRNA occurs. Translational repression is more commonly observed in the miRNAs pathway compared to the siRNA pathway. The translational repression of mRNA is also more common in animal systems compared to plants (47). Some organisms have an additional step of amplifying the initial Dicer produced siRNA signal using a protein called RNA- dependent RNA polymerase (RdRP) (48). The resulting secondary siRNA feed backs into the RNAi pathway which is primarily observed in the model organism, *Caenorhabditis elegans* (49). The effect of RNA silencing has also been shown to spread from cell-to-cell in some invertebrates, a phenomenon known as systemic RNAi (49, 50).

*RNAi in insects*

The RNAi pathway is evolutionarily conserved in insects. Homology analysis of protein sequences has identified two distinct Dicer-like proteins in insects. Dicer-1 is involved in miRNA biogenesis while Dicer-2 is involved in processing dsRNA from other sources and is an important component in the antiviral immune response in invertebrates (51). All core RNAi proteins are encoded by the genome of *A. pisum* (52). A notable finding in the genome of *A. pisum* is the duplication of some genes present in the miRNA pathway. *A. pisum* have distinct Dicer proteins for processing of dsRNA, where two duplicated copies of Dicer-1 are present for processing dsRNA in the miRNA pathway and Dicer-2 is present for processing of other siRNA (53). Another major miRNA pathway component, the Argonaute-1 is also duplicated in *A. pisum*. A related Argonaute-2 is present for the siRNA pathway. Additionally, there is one Drosha, four copies of Pasha, and one Exportin-5 present for the miRNA pathway and one R2D2 and Sid-1 for the siRNA pathway (53).

9

Pioneering studies of RNAi in insects were conducted in the model organisms, vinegar fly, *Drosophila melanogaster,* red flour beetle, *Tribolium castaneum* and silkworm, *Bombyx mori*. The RNAi pathway has been manipulated in multiple insect species for the purposes of understanding the RNAi pathway, studying gene function, understanding antiviral immunity and for pest management approaches (35, 40-42). Increased availability of insect genome sequences revealed the presence and conservation of the core RNAi pathway genes. While all insects appear to have a functional RNAi pathway, there are major differences in the response of the RNAi pathway to exogenously introduced dsRNA. Responses differ according to specific insect species, target genes, length of dsRNA used, region of the gene targeted for silencing and method of delivery. Variation in efficacy makes efforts to use RNAi directly for insect pest management via silencing genes difficult. Some of the major differences in performance of gene silencing observed in model insect species based on the method of delivery are summarized in Figure 1 (54). Although there appears to be major variation in the success rate, specific examples of gene silencing in insects have been highly successful. *T. castaneum* shows a robust RNAi response and the silencing effects are carried through to the next generation, indicating a systemic and possible amplification of the initial siRNA signal (50, 55).
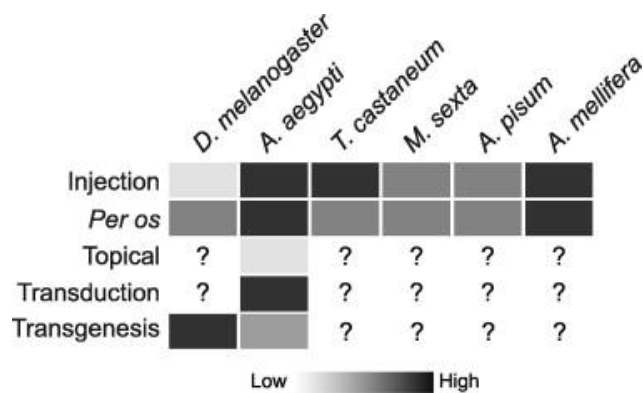


**Figure 1.** Efficacy (low to high) of gene silencing observed in model insect species based on method of delivery, used with permission from reference (54).

*Small RNA in RNA interference*

Small RNAs (sRNA) have emerged as key regulators of important biological processes in eukaryotes. The three major types of small RNA identified in eukaryotes are microRNA (miRNA), small-interfering RNA (siRNA) and PIWI-interacting RNA (piRNA). The biogenesis of eukaryotic sRNA shares a dependence on the Argonaute (Ago) protein family (56). Each small RNA type has distinct functions and characteristics including their (i) biogenesis, (ii) length and modifications, and (iii) targets (57).

*MicroRNAs (miRNAs)*

MicroRNAs (miRNAs) are a group of sRNA of 18-25 nt found in most eukaryotic cells. This class of sRNA regulates gene expression by modulating the availability of mRNA for translation into protein. A primary miRNA is transcribed in the nucleus by cellular RNA polymerase II. The primary miRNA is further processed into a precursor miRNAs by Drosha and Pasha. The precursor miRNA is transported into the cytoplasm to be cleaved into dsRNA duplexes of 18-25 nt by the Dicer protein (44). One strand of the duplex (guide strand) is loaded onto the RISC Argonaute-protein complex and targets mRNA in the cell. Near perfect complementarity of the miRNAs to the target mRNA results in the cleavage and degradation of the target transcript, a process most commonly observed in plants (58). In animals, the miRNAs seed region of 6-8 nt binds the target mRNA and prevents translation (59). Functional studies of miRNAs in arthropods have shown regulation of important biological processes such as wing development, tissue differentiation and cell proliferation (60, 61). The current version (Version 20) of miRBase, a miRNA database contains >1000 insect miRNAs including 103 miRNAs identified from *A. pisum* (62).

*Short-interfering RNA (siRNA)*

Short-interfering RNAs (siRNAs) can be divided into two categories, endogenous and exogenous siRNAs. Endogenous siRNAs (endo-siRNA) are primarily derived from retrotransposons and genomic RNA that forms dsRNA from overlapping transcripts or local secondary structures (63, 64). The production of endo-siRNA is dependent on the host RNAi pathway (65). Endo-siRNAs functions to silence mobile genetic elements and mRNA transcripts in the host (66). Exogenous siRNAs (exo-siRNA) are primarily derived from invading nucleic acids such as those of viruses or experimentally introduced dsRNA. Exo-siRNAs derived from replicating viruses are termed virus-derived small interfering RNAs (vsRNAs). Many parts of the virus genome can serve as a trigger for the host RNAi-based antiviral response. The triggers can include the complementary dsRNA intermediate during replication, local secondary structures in the virus genome, virus encoded siRNA and overlapping virus transcripts (42). Virus sequences can be elucidated from the vsRNA products of the RNAi pathway.

*PIWI-associated RNAs (piRNA)*

PIWI-associated RNAs (piRNA) are 25-30 nt small RNAs produced in a Dicer-independent mechanism (67, 68). Argonaute-3 (Ago3) cleaves piRNA from a precursor sequence, after which the piRNA is incorporated into a RISC, along with PIWI and Aubergine (Aub), to guide degradation of complementary RNA sequences. A "ping-pong" mechanism for amplification of piRNAs is now also supported, in which a sense, primary piRNA as part of a RISC, binds to a complementary, antisense RNA sequence. The target sequence is cleaved with the help of Aub and PIWI such that the sense, primary piRNA and the cleaved, antisense target contain a 10 nt overlap. A conserved feature of this overlap is an A at the first 5' antisense

position and a complementary U at the 10th 5' sense position. That antisense target is then further cleaved to become a 25-30 nt secondary piRNA. This secondary piRNA can then bind to the 10 overlapping nt in other sense sequences, from which the original piRNA was derived, and with the help of Ago3, cleaves another primary piRNA. The production of primary piRNAs thus drives the production of secondary piRNAs, and vice versa (69). The presence and role of piRNAs was initially thought to be solely transposon repression in the germline (70). However, more recent studies have shown piRNA expression in somatic tissues. First, the somatic tissue surrounding the *Drosophila* ovary was found to express PIWI proteins and piRNAs, albeit via an Aub- and Ago-3-independent version of piRNA biogenesis, called the primary pathway (71). While it is still unclear if piRNAs are expressed more broadly in fly tissue, the whole repertoire of PIWI-family proteins and piRNAs are now known to be expressed in the somatic head and thorax tissue of the yellow fever mosquito, *Aedes aegypti* and the tiger mosquito, *Aedes albopictus* (72). These data set the stage to explore piRNA roles beyond transposon regulation, including antiviral activity.

**The use of Next Generation Sequencing technology for insect virus discovery**

This section has been published in a review paper by Liu S, Vijayendran D, and Bonning BC (73).

Next generation sequencing (NGS) is being widely used to understand the role of sRNA in host-virus interactions in addition to the identification and discovery of novel insect viruses. Traditionally, insect viruses were discovered following purification of virus from diseased insects. Conventional approaches to virus discovery include virus purification by ultracentrifugation and sucrose or cesium chloride gradient steps. The sample may then be used

for visualization of virus particles by electron microscopy, infection of cultured insect cells and observation for cytopathic effects, and infection of healthy insects by spraying, injection or oral inoculation (74-76). Viruses would then be identified and further characterized by use of serological methods and nucleic acid hybridization (where specific antisera or probes are available), molecular cloning and genomic sequencing (77). Use of a cell line is beneficial in that it allows for culture and amplification of viruses that cause acute or covert infections in the host. However, the lack of appropriate insect cell lines for such virus screens is a common limiting factor and it cannot be assumed that all viruses present in an insect would replicate in a given cell line. In addition, viruses that accumulate to relatively low titers in the host organism or become latent would not be readily detected using the traditional approaches described.

NGS, which is now being widely adopted for virus discovery, is a non-Sanger-based and high-throughput methodology which allows for generation of millions of sequences at once. Multiple high-throughput sequencing technologies have been developed (78-81). The most common NGS platforms are Roche 454 pyrosequencing (454 Life Science), Illumina (Solexa) sequencing, and SOLiD sequencing (ABI Biosystems). Viral sequences can be extracted from either total DNA (for DNA viruses only) or RNA isolated from insects. Alternatively, prior to viral DNA or RNA extraction, virus purification can be conducted to eliminate host nucleic acid contamination, followed by extraction of viral DNA or RNA. The sequencing reads obtained are then assembled and analyzed using various bioinformatics tools (82, 83). Following detection of viral sequences by NGS technologies, the presence of viral sequences in the sample must be confirmed by PCR (DNA viruses) or RT-PCR (RNA viruses). Small RNA sequencing can also be used to identify viruses in insects. Virus-derived siRNAs can be used to reveal the sequences

of any RNA or DNA virus present in the insect that is susceptible to the RNAi-based antiviral immune response.

The first application of NGS technology that demonstrated the potential for virus discovery was a metagenomic analysis of the honeybee, *Apis mellifera*, conducted to elucidate the causes of colony collapse disorder (CCD) (84). For the metagenomic analysis, total RNA was extracted from bees taken from CCD and non-CCD colonies collected from the US, and Australia and also from royal jelly from China. The RNA libraries were subjected to 454 pyrosequencing, and raw reads were trimmed and assembled into contigs. Contigs were used for BLAST analysis (BLASTn and BLASTx) against the NCBI nr database (83). Seven viruses were identified in bees derived from CCD colonies, compared to five from non-CCD colonies. A wide range of other pathogens was also detected. The presence of the viruses was confirmed by RT-PCR and Sanger sequencing, and the presence of *Israeli acute paralysis virus* (IAPV) was found to be a significant indicator of CCD (84).

The first report of the use of sRNA sequencing for virus identification was for analysis of the sweet potato. In this case, the authors inoculated the plants with the known RNA viruses, *Sweet potato feathery mottle potyvirus* (SPFMV) and *Sweet potato chlorotic stunt closterovirus* (SPCSV). Small RNA was isolated from the inoculated plants and sequenced by using Illumina GAII. The sRNA reads were assembled with three different assembly programs for sequence assembly, and contigs were reassembled to generate longer contigs using the program Contigexpress (Vector NTI, Invitrogen). The contigs were queried by searching the GenBank nr database for viral sequences. SPFMV and SPCSV sequences were successfully recovered. In addition, contigs similar to mastreviruses (ssDNA) and badnaviruses (dsDNA reverse transcribing) were identified from the sRNA sequences (85).

A similar strategy was used to identify viruses present in a *Drosophila* cell line, and from published sRNA datasets for mosquitoes and nematodes (86). Four viruses (two positive single-strand, ssRNA and two dsRNA viruses) were identified from the *Drosophila* S2-GMR cell line. In addition, two viruses, including one new virus, were identified from the mosquito. However, full length genomes of the viruses could not be assembled from the sRNA datasets. RT-PCR, RACE-PCR and sequencing were used to fill gaps in the viral sequences. While many known viruses have been identified using NGS, the true strength of this technology lies in the ability of researchers to identify novel viruses *de novo.* Table 2 shows novel insect viruses that have been identified using NGS.

**Table 2.** Novel insect and nematode viruses identified using NGS data

| Virus | Origin | Ref |
|---|---|---|
| **Birnaviridae (dsRNA)** | | |
| *Drosophila birnavirus* (DBV) | *D. melanogaster* cell line (S2-GMR) | (86) |
| *Culicine-associated Z virus* (CAZV) | *O. caspius, O. detritus* | (87) |
| *Espirito Santo virus* (ESV) | *Ae. albopictus* cell line (C6/36) | (88) |
| **Totiviridae (dsRNA)** | | |
| *Drosophila totivirus* (DTV) | *D. melanogaster* cell line (S2-GMR) | (86) |
| **Dicistroviridae (+ssRNA)** | | |
| *Big Sioux river virus* (BSRV) | *Ap. mellifera* | (89) |
| **Iflaviridae (+ssRNA)** | | |
| *Spodoptera exigua iflavirus-like 1/2* (SeIV-1/2) | *S. exigua* | (90, 91) |
| **Nodaviridae (bipartite +ssRNA)** | | |
| *American nodavirus* (ANV) | *D. melanogaster* cell line (S2-GMR) | (86) |
| *Mosquito nodavirus* (MNV) | *Ae. aegypti*-Liverpool strain | (86) |
| *Santeuil virus Nodavirus* | *C. briggsae* | (92) |
| *Orsay virus Nodavirus* | *C. elegans* | (92) |
| *Le Blanc virus* | *C. briggsae* | (93) |
| **Negeviruses (proposed new taxon) (+ssRNA)** | | |
| *Negev virus* (NEGV) | *C. coronator, C. quinquefasciatus* | (94) |
| *Piura virus* (PIUV) | *Culex sp.* | (94) |
| *Loreto virus* (LORV) | *An. albimanus* | (94) |
| *Dezidougou virus* (DEZV) | *Ae. aegypti* | (94) |
| *Santana virus* (SANV) | *Culex sp.* | (94) |
| *Ngewontan virus* (NWTV) | *C. vishnui* | (94) |

**Table 2.** Novel insect and nematode viruses identified using NGS data (continued)

| | | |
|---|---|---|
| **Togaviridae (+ssRNA)** | | |
| *Eilat virus* (EILV) | *An. coustani* | (95) |
| **Bunyaviridae (-ssRNA)** | | |
| *SCN phlebovirus* (ScPV) | *H. glycines* | (96) |
| **Rhaboviridae (-ssRNA)** | | |
| *SCN rhabdovirus* (ScRV) | *H. glycines* | (96) |
| **Unclassified RNA viruses** | | |
| *Noravirus* (+ssRNA) | *D. melanogaster* ovary cell line | (86) |
| *Lake Sinai Virus 1/2* (LSV1/2;+ssRNA) | *A. mellifera* | (89) |
| *Anopheline-associated C virus* (+ssRNA) | *An. maculipennis* | (87) |
| *SCN nyavirus* (ScNV;-ssRNA) | *H. glycines* | (96) |
| *SCN tenuivirus* (ScTV;-ssRNA) | *H. glycines* | (96) |
| **Nudivirus (dsDNA)** | | |
| *Drosophila innubila Nudivirus* (DiNV) | *D. innubila* | (97) |
| **Parvoviridae (ssDNA)** | | |
| *Culex tritaeniorhynchus densovirus* | *C. pipiens molestus* | (98) |
| **Circoviridae (ssDNA)** | | |
| *The Florida woods cockroach-associated cyclovirus GS140* (FWCasCyV-GS140) | *E. floridana* | (99) |

Although NGS has fundamentally changed the methodology for discovery of viruses from insects, there are some limitations. One limitation is that it is not possible to identify novel viruses that lack homology to known viruses. An exception to this is when the DNA or RNA sequenced is extracted from purified virus, and hence the viral origin of the sequence has already been established. A second limitation to the use of NGS is that full length genome sequences are unlikely to be acquired unless the virus is present in the host insect at high titers. Further sequencing of the genome by other methods will likely be required. In some cases, although most of the sequence is acquired, the 5' and 3' end sequences were difficult to find (89). Hence it is important, where possible, to retain frozen tissues for virus isolation and / or maintain a colony of the insect for virus extraction. A third challenge for NGS methods is the use of non-standardized methods for data analysis. There are no clearly established guidelines as to what is

acceptable read quality, parameters for short read data assembly, and significance of BLAST hits, for example. With the increasing use of NGS, there is a real need to develop tools and software to handle bioinformatics analysis for any organism, rather than just model organisms. In addition to viruses that are known in aphids, NGS is being used for discovery of novel viruses in aphids.

**Viruses in aphids**

The viruses that are known to infect aphids are summarized in Table 3 (100-105) .

*Dicistroviridae*

The virus family Dicistroviridae consists of single-stranded RNA viruses (ssRNA) that infect arthropods. *Aphid-lethal paralysis virus* (ALPV) and *Cricket- paralysis virus* (CrPV) are known to negatively impact agricultural insect pests. Infection of honeybees with *Israeli acute paralysis virus* of bees (IAPV) is being more frequently reported and has been associated with colony collapse disorder (CCD) (84, 89). Figure 2 shows the genome organization of dicistroviruses. The genome has two major open reading frames (ORF). The 5' end of the genome codes for the non-structural polyprotein and the 3' end of the genome encodes for the structural polyprotein. The genome also has two internal ribosomal entry sites (IRES), one at the 5' end of the genome and the other between ORF1 and ORF2, the IGR IRES (106). Dicistroviruses also encode suppressors of RNAi (107-109). The suppressor proteins inhibit the RNAi pathway which targets the replicating virus. The *Drosophila C virus* (DCV) encodes the DCV-1A protein that binds to and inhibits the processing of dsRNA by Dicer (107). CrPV codes for the CrPV-1A protein that binds to Argonaute in RISC to inhibit the cleavage of the target RNA (108).

**Table 3.** Summary of viruses that infect aphids

| Virus | Classification | Host aphid | Genome | Virion structure | Transmission | Ref |
|---|---|---|---|---|---|---|
| *Aphid-lethal paralysis virus* | Dicistroviridae | Bird-cherry oat aphid (*Rhopalosiphum padi*), Pea aphid (*Acyrthosiphon pisum*) | + ssRNA | Icosahedral 27nm | Vertical# | (100) |
| *Rhopalosiphum padi virus* | Dicistroviridae | Bird cherry-oat aphid (*Rhopalosiphum padi*) | + ssRNA | Icosahedral | Horizontal and Vertical | (101) |
| *Brevicoryne brassicae virus* | Iflaviridae | Cabbage aphid (*Brevicoryne brassicae*) | + ssRNA | Icosahedral | No horizontal via plant | (102) |
| *Acyrthosiphon pisum virus* | Unclassified | Pea aphid (*Acyrthosiphon pisum*) | + ssRNA | Icosahedral 31nm | Not determined | (103) |
| *Rosy apple aphid virus* | Unclassified | Rosy apple aphid (*Dysaphis plantaginea*) | + ssRNA | Icosahedral 32nm | Horizontal | (104) |
| *Aphis glycines virus* | Unclassified | Soybean aphid (*Aphis glycines*) | + ssRNA | Icosahedral 30nm | Vertical | Chapter 3 |
| *Myzus persicae densovirus* | Parvoviridae | Green peach aphid (*Myzus persicae*) | ssDNA | Icosahedral 20nm | Not determined | (105) |
| *Dysaphis plantaginea densovirus* | Parvoviridae | Rosy apple aphid (*Dysaphis plantaginea*) | ssDNA | Icosahedral 22nm | Horizontal | (104) |
| *Acyrthosiphon pisum densovirus* | Parvoviridae | Pea aphid (*Acyrthosiphon pisum*) | ssDNA | Icosahedral | Unknown | (104) |

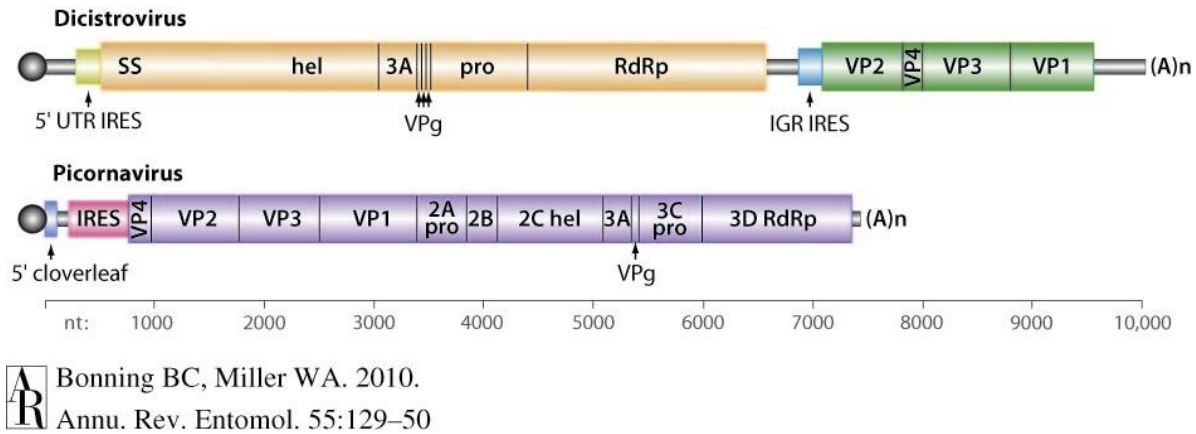# Horizontal transmission not determined

**Figure 2.** Schematic diagram of genome organization of dicistrovirus and picornavirus. Vertical lines indicated proteolytic cleavage sites. SS, silencing suppressor domain, hel, helicase from superfamily 3, Vpg, genome-linked protein, pro, cysteine protease (chymotrypsin-like), RdRp, RNA-dependent RNA polymerase, VP, virion protein, IGR, intergenic region; IRES, internal ribosomal entry site, UTR, untranslated region.

The dicistroviruses known to infect aphids are *Aphid-lethal paralysis virus* (ALPV) and *Rhopalosiphum padi virus* (RhPV). The use of NGS for virus discovery has added to our knowledge of the prevalence of these viruses. ALPV-like viruses have now been identified in organisms other than aphids, including honeybees and from bat fecal samples (Table 4) (89, 100, 110-113) . ALPV was first isolated from the bird cherry-oat aphid, *Rhopalosiphum padi* after observation of infected aphids moving away from the food source and death induced by paralysis (100). ALPV is transmitted vertically in *R. padi* at a 30% rate (n=44) and at a 17% rate in the wheat aphid, *Sitobion avenae*, (n=30) (114). Localization studies of ALPV in *R. padi* tissues using nucleic acid in situ hybridization appear to show localization of viral RNA in the gut, brain and embryonic tissues (115, 116).

RhPV was isolated from *R. padi* after observation of reduced longevity, fecundity and fitness of the aphid population (101). RhPV circulates in the plant phloem and is transmitted

horizontally from plants to uninfected aphids (117). RhPV is also vertically transmitted

transovarially from mother to offspring at 15-28% in *R. padi* (101). Temperature has an effect on

the titer of ALPV and RhPV virus infection in *R. padi* and *S. avenae* when tested using a double

antibody sandwich enzyme linked immunosorbent assay (DAS-ELISA). A temperature

difference of 5°C was sufficient to increase the incidence of both viruses. An increase of 19% of

*R. padi* tested positive for ALPV with an increase of rearing temperature from 15- 20°C

compared to 15 - 24°C. An increase of 46% of aphids was infected with RhPV at rearing

temperatures from 10 -15°C compared to 15 - 20°C (114).

*Other aphid viruses*

The *Brevicoryne brassicae virus* (BrBV) was isolated from the cabbage aphid,

*Brevicoryne brassicae*. This is the first iflavirus described in aphids. The virus has a positive

ssRNA genome of 10,161 nt and has a 3' poly (A) tail. The virus is closely related to other insect

iflaviruses. No virus transcripts were detected in plants infested with BrBV suggesting that BrBV

is probably not transmitted via the plant (102).

The *Acyrthosiphon pisum virus* (APV) is a positive ssRNA virus with a genome of

approximately 10,000 nt. The virus has two major ORFs and the virus encodes four capsid

proteins ranging from 23 kDa to 66 kDa. APV was abundant in the epithelium and lumen of the

digestive tract of three day-old nymphs. The virus was also less frequently localized in muscle

cells and bacteriocytes using immunolocalization with an APV antigen.  Infection with the virus

appears to inhibit the growth of aphids and a decline in the population was observed when the

rearing temperature was increased by 6 – 8 °C (102).

**Table 4.** Evidence for ALPV-like viruses in insects and insect predators. It is unknown if the ALPV strains isolated from bats are derived from prey, or whether they replicate in the bat. CP, capsid protein; NSP, non-structural protein.

| Virus strain | Accession no | Host organism | Genome size | Ref |
|---|---|---|---|---|
| ALPV-RhP | AF536531 | Bird cherry-oat aphid (*Rhopalosiphum padi*) | 9,812 nt | (100) |
| ALPV-Ap | Pending | Pea aphid (*Acyrthosiphon pisum*) | 9,940 nt | Chapter 2 |
| ALPV-An | JX480861 | Milkweed aphid (*Aphis nerii*) | 9,835 nt | (110) |
| ALPV-TJ | JQ320375 | Rickett's big-footed bat (*Myotis ricketti*) | 9,819 nt | (111) |
| ALPV- YNH | JN857319 | Great roundleaf bat (*Hipposideros armiger*) | 648 nt (Partial sequence from NSP) | (111) |
| ALPV-Brookings | HQ871932 | Honeybee (*Apis mellifera*) | 4,125 nt (Partial sequence from CP & NSP) | (89) |
| ALPV- Spain honeybee | JX045858 | Honeybee (*Apis mellifera*) | 9,327 nt (near full length) | (112) |
| ALPV- Belgium honeybee | KC880119.1 | Honeybee (*Apis mellifera*) | 465 nt (Partial sequence from NSP) | (113) |

The *Rosy apple aphid virus* (RAAV) is a positive strand RNA virus isolated from the

rosy apple aphid, *Dysaphis plantaginea* that has a genome organization and a high amino acid

sequence identity to APV. RAAV was present in plant leaves previously exposed to RAAV-

infected aphids and therefore is likely to be horizontally transmitted via the plant phloem. The

incidence of winged aphid morphs was reduced when the aphid was co-infected with both

RAAV and *Dysaphis plantaginea densovirus* (DplDNV) (104).

DplDNV is a densovirus isolated from the rosy apple aphid, *Dysaphis plantaginea*.

DplDNV is a single stranded DNA virus with a genome of approximately 5,000 nt. DplDNV is

transmitted horizontally via the leaf and vertically from adults to nymphs. DplDNV infection

induces winged morphs of the host aphid thereby facilitating virus dispersal. Infection with

DplDNV negatively affects aphid fecundity.

An additional densovirus was also identified from expressed sequenced tag data from *A.

pisum*. The virus is a putative *Acyrthosiphon pisum densovirus*. No further characterization of the

virus has been published (104). Another densovirus was identified from *M. persicae*, named

*Myzus persicae densovirus*. The virus has a genome of approximately 5,700 nt with five ORFs.

The virus infects the stomach of infected aphids and transmission occurs horizontally (through

plant and honeydew) and vertically from mother to offspring (105).

**Antiviral immunity in aphids**

Parts of this section have been published in a review paper by Vijayendran D, Airs PM, Dolezal

K and Bonning BC (42).

Arthropods lack the adaptive immune response of B and T-cells that are found in

vertebrates. A robust immune response against viruses in arthropods includes the signaling

pathways; JAK/STAT, Imd and Toll pathways in addition to the antiviral RNAi pathway (118-120). The majority of RNAi based antiviral studies of arthropods have been carried out in *D. melanogaster* and various mosquito species (86, 121-124). The RNAi response produces virus-derived small interfering RNAs (vsRNAs) to reduce the virus load of replicating viruses. The production of vsRNA is an important part of the innate immune response against viruses in arthropods. The RNAi based-immunity pathway in arthropods is highly developed and recognition of non-native RNAs triggers the defense response (125). In addition to vsRNAs, miRNA and piRNA are also involved in antiviral immunity (72, 126).

The dynamics of RNAi in establishment of persistent virus infection in insects were recently elucidated. An emerging concept was experimentally validated using *Flock house virus* (FHV) *in vitro* in *Drosophila* S2 cells and *in vivo* in adult *D. melanogaster*. Genome segments of FHV were reverse-transcribed to cDNA as early as 12 hours after infection by host reverse transcriptases (originating from retrotransposons or endogenous retroviruses). Surprisingly, the cDNA form of FHV was a reorganized, recombinant form of the 2 RNA segments, RNA 1 and RNA 2. The FHV cDNA was then transcribed and processed into vsRNAs by the host RNAi machinery. These vsRNAs are loaded onto RISC and mediate RNAi of FHV RNA. The cDNA form is thus used early in infection to ensure that the virus forms a persistent infection instead of an acute infection that would compromise the host cells. The authors in addition speculated that insects lacking RNA-dependent RNA polymerase, a protein involved in amplification of the primary vsRNA signal, use this virus cDNA method to ensure a continuous immune response in the insect after virus infection (127).

Sequencing of *A. pisum* genome allowed for analysis of genes involved in immunity in a hemipteran insect. While genes involved in the Toll and JAK/STAT immune signaling pathways

are conserved, *A. pisum* lacks key genes in the IMD pathway that functions to protect arthropods against bacterial infection (52, 128). The close association of aphids and their symbiotic bacteria provides a possible evolutionary reason for the loss of those antimicrobial genes. Aphids harbor the obligate primary endosymbiotic bacterium *Buchnera* housed in specialized aphid cells called bacteriocytes (129). *Buchnera* synthesizes essential amino acids for the host aphid and the aphid in return synthesizes non-essential amino acids from the symbiotic bacterium (130). The close symbiotic relationship has resulted in a reduced *Buchnera* genome compared to its free-living wild type ancestor. In addition to the primary obligate symbiont, aphids harbor various secondary symbiotic bacteria. Secondary symbionts may confer some additional benefits to the host aphid. e.g *Hamiltonella defensa* and *Serratia symbiotica* provide resistance against parasitoid wasps when the adult aphid is parasitized (131) and *Regiella insecticola* reduces mortality in *A. pisum* when exposed to the fungal pathogen, *Zoophthora occidentalis* (132).

**Viruses in insect management**

Insect viruses have been effectively used for management of insect pests for many years around the world. Field populations of *H. armigera* have been effectively controlled using a *Helicoverpa armigera stunt virus* (*Tetraviridae*), a sRNA virus (133). The velvetbean caterpillar, *Anticarsia gemmatalis*, is a damaging foliage feeder of soybean plants that has been controlled in Brazil using a nucleopolyhedrovirus (*Baculoviridae*). The virus has been used for over 25 years in one of the largest virus-based insecticide programs in the world (134). In addition to lepidopteran insects, virus-based control has also been applied to management of a coleopteran pest. The rhinoceros beetle, *Oryctes rhinoceros* is a damaging pest in tropical countries that has been controlled for many years using the *Oryctes virus* (Genus Nudivirus; Unclassified) in oil palm plantations in Malaysia (135).

Novel aphid viruses have potential for application for aphid management. Viruses could potentially be produced in large quantities for field applications as previously described for nucleopolyhedrovirus (134, 136). In addition, novel aphid viruses in theory could be used as vectors for delivering silencing RNAs to aphids when they feed on the plant phloem. An example of this approach is the use of a *Sindbis virus* vector to deliver dsRNA targeting the Broad-complex transcription factor in the silkworm, *Bombyx mori* (137). Targeted silencing of aphid genes could further provide information about gene function. Studies on virus discovery and the biological interactions between aphids and viruses could also facilitate understanding of antiviral immune responses in aphids.

**Dissertation organization**

The two chapters that follow this introduction focus on viruses that were identified from aphid sequencing data. Chapters 2 and 3 are on the identification and characterization of *Aphid lethal paralysis virus- Ap* (ALPV-Ap), a virus identified from the sRNAs of *A. pisum*. Chapter 4 is on the identification and characterization of a novel aphid virus, *Aphis glycines virus* (AGV) from *A. glycines*. AGV was identified from transcriptome and sRNA sequencing datasets. Chapter 5 focuses on aphid miRNAs, identification of a subset of 12 aphid-specific miRNAs and the expression of three aphid-specific miRNAs in various life stages in *M. persicae*. In the final chapter the key findings from the research chapters are summarized, with general conclusions and discussion of implications for the development of novel biotechnology tools for aphid management.

# References

1.      Will T & Vilcinskas A (2013) Aphid-proof plants: biotechnology-based approaches for aphid control. *Adv Biochem Eng Biotechnol* (136):179-203.

2.      Will T & van Bel AJ (2006) Physical and chemical interactions between aphids and plants. *J Exp Bot* 57(4):729-737.

3.      Hill DS (1987) *Agricultural insect pests of temperate regions and their control* (Cambridge University Press, Cambridge) p 659.

4.      Tjallingii WF (2006) Salivary secretions by aphids interacting with proteins of phloem wound responses. *J Exp Bot* 57(4):739-745.

5.      Wood BW, Tedders WL, & Reilly CC (1988) Sooty mold fungus on pecan foliage suppresses light penetration and net photosynthesis. *Hort science* 23(5):851-853.

6.      Stafford CA, Walker GP, & Ullman DE (2012) Hitching a ride: Vector feeding and virus transmission. *Commun Integr Biol* 5(1):43-49.

7.      Nault LR (1997) Arthropod transmission of plant viruses: A new synthesis. *Ann Entomol Soc Am* 90(5):521-541.

8.      Martín B, Collar JL, Tjallingii WF, & Fereres A (1997) Intracellular ingestion and salivation by aphids may cause the acquisition and inoculation of non-persistently transmitted plant viruses. *J Gen Virol* 78 ( Pt 10):2701-2705.

9.      Thomas PP & Harris KF (1977) Nonpersistent transmission of plant viruses by aphids. *Annu Rev Phytopathol* 15(1):55-73.

10.     Ng JC & Falk BW (2006) Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu Rev Phytopathol* 44:183-212.

11.     Hogenhout SA, Ammar e-D, Whitfield AE, & Redinbaugh MG (2008) Insect vector interactions with persistently transmitted viruses. *Annu Rev Phytopathol* 46:327-359.

12.     Dedryver CA, Le Ralec A, & Fabre F (2010) The conflicting relationships between aphids and men: a review of aphid damage and control strategies. *C R Biol* 333(6-7):539-553.

13.     Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, & Desneux N (2011) Ecology and management of the soybean aphid in North America. *Annu Rev Entomol* 56:375-399.

14.     Devonshire A (1989) Resistance of aphids to insecticides. *Aphids, their biology, natural enemies and control* 100:123-139.

15. Joshi S, Rabindra RJ, & Rajendran TP (2010) Biological control of aphids. *J Biol Control* 24(3):185-202.

16. Kim KS*, et al.* (2010) Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538. *Theor Appl Genet* 121(3):599-610.

17. Kim KS*, et al.* (2010) Fine mapping the soybean aphid resistance gene Rag1 in soybean. *Theor Appl Genet* 120(5):1063-1071.

18. Anderson GR, Papa D, Peng J, Tahir M, & Lapitan NL (2003) Genetic mapping of Dn7, a rye gene conferring resistance to the Russian wheat aphid in wheat. *Theor Appl Genet* 107(7):1297-1303.

19. Michel AP, Mittapalli O, & Mian RMA (2011) *Evolution of Soybean Aphid Biotypes: Understanding and Managing Virulence to Host-Plant Resistance* (InTech) p 514.

20. Randolph TL*, et al.* (2005) Yield response and categories of resistance to Russian wheat aphid in four Dn4 hard red winter wheat cultivars. *J Econ Entomol* 98(2):588-594.

21. Shufran KA, Mornhinweg DW, Baker CA, & Porter DR (2007) Variation to cause host injury between Russian wheat aphid (Homoptera: Aphididae) clones virulent to Dn4 wheat. *J Econ Entomol* 100(5):1685-1691.

22. Smith CM*, et al.* (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) populations virulent to the Dn4 resistance gene. *J Econ Entomol* 97(3):1112-1117.

23. Zhang Y, Ma Y, Wan PJ, Mu LL, & Li GQ (2013) *Bacillus thuringiensis* insecticidal crystal proteins affect lifespan and reproductive performance of *Helicoverpa armigera* and *Spodoptera exigua* adults. *J Econ Entomol* 106(2):614-621.

24. Chougule NP*, et al.* (2013) Retargeting of the *Bacillus thuringiensis* toxin Cyt2Aa against hemipteran insect pests. *Proc Natl Acad Sci U S A* 110(21):8465-8470.

25. Bonning BC*, et al.* (2014) Toxin delivery by the capsid protein of an aphid-vectored plant virus provides plant resistance to aphids. *Nat Biotechnol* 32(1):102-105.

26. Rahbe Y & Febvay G (1993) Protein toxicity to aphids: an *in vitro* test on *Acyrthosiphon pisum. Entomol Exp Appl* 67:149-160.

27. Bonning BC & Chougule NP (2014) Delivery of intrahemocoelic peptides for insect pest management. *Trends Biotechnol* 32(2):91-98.

29. Down RE*, et al.* (2006) Insecticidal spider venom toxin fused to snowdrop lectin is toxic to the peach-potato aphid, *Myzus persicae* (Hemiptera: Aphididae) and the rice brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae). *Pest Manag Sci* 62(1):77-85.

30.  Jaubert-Possamai S*, et al.* (2007) Gene knockdown by RNAi in the pea aphid *Acyrthosiphon pisum*. *BMC Biotechnol* 7:63.

31.  Pitino M, Coleman AD, Maffei ME, Ridout CJ, & Hogenhout SA (2011) Silencing of Aphid Genes by dsRNA Feeding from Plants. *PLoS One* 6(10):e25709.

32.  Xu L*, et al.* (2013) Silencing of an aphid carboxylesterase gene by use of plant-mediated RNAi impairs *Sitobion avenae* tolerance of Phoxim insecticides. *Transgenic Res* 2:389-396.

33.  Shakesby AJ*, et al.* (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem Mol Biol* 39(1):1-10.

34.  Mao J & Zeng F (2013) Plant-mediated RNAi of a gap gene-enhanced tobacco tolerance against the Myzus persicae. *Transgenic Res* 23:145-152.

35.  Mutti N, Park Y, Reese J, & Reeck G (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. *J Insect Sci* 6:1-7.

36.  Whyard S, Singh AD, & Wong S (2009) Ingested double-stranded RNAs can act as species-specific insecticides. *Insect Biochem Mol Biol* 39(11):824-832.

37.  Bhatia V, Bhattacharya R, Uniyal PL, Singh R, & Niranjan RS (2012) Host generated siRNAs attenuate expression of serine protease gene in *Myzus persicae*. *PLoS One* 7(10):e46343.

38.  Zhang M*, et al.* (2013) Identifying potential RNAi targets in grain aphid (*Sitobion avenae F.*) based on transcriptome profiling of its alimentary canal after feeding on wheat plants. *BMC Genomics* 14(1):560.

39.  Christiaens O, Swevers L, & Smagghe G (2014) DsRNA degradation in the pea aphid (*Acyrthosiphon pisum*) associated with lack of response in RNAi feeding and injection assay. *Peptides*.In Press.

40.  Baum JA*, et al.* (2007) Control of coleopteran insect pests through RNA interference. *Nat Biotechnol* 25(11):1322-1326.

41.  Mao YB*, et al.* (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* 25(11):1307-1313.

42.  Vijayendran D, Airs PM, Dolezal K, & Bonning BC (2013) Arthropod viruses and small RNAs. *J Invertebr Pathol* 114(2):186-195.

43.  Fire A*, et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806-811.

44. MacRae IJ, *et al.* (2006) Structural basis for double-stranded RNA processing by dicer. *Science* 311(5758):195-198.

45. Kawamata T & Tomari Y (2010) Making RISC. *Trends in Biochem Sci* 35(7):368-376.

46. Hammond SM, Bernstein E, Beach D, & Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404(6775):293-296.

47. Ghildiyal M & Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10(2):94-108.

48. Zhang C & Ruvkun G (2012) New insights into siRNA amplification and RNAi. *RNA Biol* 9(8):1045-1049.

49. Hunter CP, *et al.* (2006) Systemic RNAi in *Caenorhabditis elegans*. *Cold Spring Harb Symp Quant Biol* 71:95-100.

50. Tomoyasu Y, *et al.* (2008) Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium. Genome Biol* 9(1):R10.

51. Kim K, Lee YS, Harris D, Nakahara K, & Carthew RW (2006) The RNAi pathway initiated by Dicer-2 in *Drosophila*. *Cold Spring Harb Symp Quant Biol* 71:39-44.

52. Consortium TIAG (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum. PLoS Biol* 8(2):e1000313.

53. Jaubert-Possamai S, *et al.* (2010) Expansion of the miRNAs pathway in the hemipteran insect Acyrthosiphon pisum. *Mol Biol Evol* 27 (5):979-987.

54. Scott JG, *et al.* (2013) Towards the elements of successful insect RNAi. *J Insect Physiol* 59(12):1212-1221.

55. Minakuchi C, Namiki T, Yoshiyama M, & Shinoda T (2008) RNAi-mediated knockdown of juvenile hormone acid O-methyltransferase gene causes precocious metamorphosis in the red flour beetle *Tribolium castaneum*. *FEBS J* 275(11):2919-2931.

56. Hock J & Meister G (2008) The Argonaute protein family. *Genome Biol* 9(2):210.

57. Farazi TA, Juranek SA, & Tuschl T (2008) The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135(7):1201-1214.

58. Brodersen P, *et al.* (2008) Widespread translational inhibition by plant miRNAss and siRNAs. *Science* 320(5880):1185-1190.

59. Lim LP, *et al.* (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433(7027):769-773.

60. Jones CI & Newbury SF (2010) Functions of microRNAs in *Drosophila* development. *Biochem Soc Trans* 38(4):1137-1143.

61. Schnall-Levin M, Zhao Y, Perrimon N, & Berger B (2010) Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3'UTRs. *Proc Natl Acad Sci U S A* 107(36):15751-15756.

62. Kozomara A & Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 39:D152-157.

63. Ghildiyal M, *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320(5879):1077-1081.

64. Okamura K, Balla S, Martin R, Liu N, & Lai EC (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* 15(6):581-590.

65. Kawamura Y, *et al.* (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453(7196):793-797.

66. Kritikou E (2008) Small RNAs: Endo-siRNAs truly endogenous. *Nat Rev Mol Cell Biol* 9:426-427.

67. Houwing S, *et al.* (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129(1):69-82.

68. Vagin VV, *et al.* (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313(5785):320-324.

69. Siomi MC, Sato K, Pezic D, & Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 12(4):246-258.

70. Juliano C, Wang J, & Lin H (2011) Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu Rev Genet* 45:447-469.

71. Malone CD & Hannon GJ (2009) Molecular evolution of piRNA and transposon control pathways in *Drosophila*. *Cold Spring Harb Symp Quant Biol* 74:225-234.

72. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, & Myles KM (2012) Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog* 8(1):e1002470.

73.    Liu S, Vijayendran D, & Bonning BC (2011) Next generation sequencing technologies for insect virus discovery. *Viruses* 3(10):1849-1869.

74.    Hunter WB, *et al.* (2001) Discovering new insect viruses: whitefly iridovirus (Homoptera: Aleyrodidae: Bemisia tabaci). *J Invertebr Pathol* 78(4):220-225.

75.    Funk CJ, Hunter WB, & Achor DS (2001) Replication of insect iridescent virus 6 in a whitefly cell line. *J Invertebr Pathol* 77(2):144-146.

76.    Zirkel F, *et al.* (2011) An insect nidovirus emerging from a primary tropical rainforest. *MBio* 2(3):e00077-00011.

77.    Bai H, *et al.* (2011) Isolation and characterization of a novel alphanodavirus. *Virol J* 8:311.

78.    Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nat Methods* 5(1):16-18.

79.    Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11(1):31-46.

80.    Shendure J & Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26(10):1135-1145.

81.    Shendure JA, Porreca GJ, & Church GM (2008) Overview of DNA sequencing strategies. *Curr Protoc Mol Biol* Chapter 7:Unit 7.1.

82.    Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821-829.

83.    Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.

84.    Cox-Foster DL, *et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318(5848):283-287.

85.    Kreuze JF, *et al.* (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388(1):1-7.

86.    Wu Q, *et al.* (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci U S A* 107(4):1606-1611.

87.    Cook S, *et al.* (2013) Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* 8(11):e80720.

88. Vancini R, *et al.* (2012) Espirito Santo virus: a new birnavirus that replicates in insect cells. *J Virol* 86(5):2390-2399.

89. Runckel C, *et al.* (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS One* 6(6):e20656.

90. Millan-Leiva A, Jakubowska AK, Ferre J, & Herrero S (2012) Genome sequence of SeIV-1, a novel virus from the Iflaviridae family infective to *Spodoptera exigua*. *J Invertebr Pathol* 109(1):127-133.

91. Pascual L, *et al.* (2012) The transcriptome of *Spodoptera exigua* larvae exposed to different types of microbes. *Insect Biochem Mol Biol* 42(8):557-570.

92. Felix MA, *et al.* (2011) Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biol* 9(1):e1000586.

93. Franz CJ, Zhao G, Felix MA, & Wang D (2012) Complete genome sequence of Le Blanc virus, a third Caenorhabditis nematode-infecting virus. *J Virol* 86(21):11940.

94. Vasilakis N, *et al.* (2013) Negevirus: a proposed new taxon of insect-specific viruses with wide geographic distribution. *J Virol* 87(5):2475-2488.

95. Nasar F, *et al.* (2012) Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. *Proc Natl Acad Sci U S A* 109(36):14622-14627.

96. Bekal S, Domier LL, Niblack TL, & Lambert KN (2011) Discovery and initial analysis of novel viral genomes in the soybean cyst nematode. *J Gen Virol* 92(Pt 8):1870-1879.

97. Unckless RL (2011) A DNA virus of *Drosophila*. *PLoS One* 6(10):e26564.

98. Ma M, *et al.* (2011) Discovery of DNA viruses in wild-caught mosquitoes using small RNA high throughput sequencing. *PLoS One* 6(9):e24758.

99. Padilla-Rodriguez M, Rosario K, & Breitbart M (2013) Novel cyclovirus discovered in the Florida woods cockroach *Eurycotis floridana* (Walker). *Arch Virol* 158(6):1389-1392.

100. Williamson C, Rybicki EP, Kasdorf GF, & Von Wechmar MB (1988) Characterization of a new picorna-like virus isolated from aphids. *J Gen Virol* 69:787-795.

101. D'Arcy CJ, Burnett PA, Hewings AD, & Goodman RM (1981) Purification and characterization of a virus from the aphid *Rhopalosiphum padi*. *Virology* 112(1):346-349.

102. Ryabov EV (2007) A novel virus isolated from the aphid *Brevicoryne brassicae* with similarity to Hymenoptera picorna-like viruses. *J Gen Virol* 88(Pt 9):2590-2595.

103.    van den Heuvel JF, Hummelen H, Verbeek M, Dullemans AM, & van der Wilk F (1997) Characteristics of acyrthosiphon pisum virus, a newly identified virus infecting the pea aphid. *J Invertebr Pathol* 70(3):169-176.

104.    Ryabov EV, Keane G, Naish N, Evered C, & Winstanley D (2009) Densovirus induces winged morphs in asexual clones of the rosy apple aphid, *Dysaphis plantaginea*. *Proc Natl Acad Sci U S A* 106(21):8465-8470.

105.    van Munster M, *et al.* (2003) Characterization of a new densovirus infecting the green peach aphid *Myzus persicae*. *J Invertebr Pathol* 84(1):6-14.

106.    Bonning BC & Miller WA (2010) Dicistroviruses. *Annu Rev Entomol* 55:129-150.

107.    Chao JA, *et al.* (2005) Dual modes of RNA-silencing suppression by Flock House virus protein B2. *Nat Struct Mol Biol* 12(11):952-957.

108.    van Rij RP, *et al.* (2006) The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev* 20(21):2985-2995.

109.    Nayak A, *et al.* (2010) Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in *Drosophila*. *Nat Struct Mol Biol* 17(5):547-554.

110.    Dombrovsky A & Luria N (2013) The Nerium oleander aphid *Aphis nerii* is tolerant to a local isolate of Aphid lethal paralysis virus (ALPV). *Virus Genes* 46(2):354-361.

111.    Ge X, *et al.* (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86(8):4620-4630.

112.    Granberg F, *et al.* (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid lethal paralysis, Israel acute paralysis and Lake sinai viruses. *PLoS One* 8(2):e57459.

113.    Ravoet J, *et al.* (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS One* 8(8):e72443.

114.    Laubscher JM & von Wechmar MB (1992) Influence of Aphid lethal paralysis virus and *Rhopalosiphum padi* virus on aphid biology at different temperatures. *J Invertebr Pathol* 60:134-140.

115.    Laubscher JM, Jaffer MA, & von Wechmar MB (1992) Detection by immunogold cytochemical labeling of aphid lethal paralysis virus in aphid *Rhopalosiphum padi* (Hemiptera:Aphididae). *J Invertebr Pathol* 60:40-46.

116. Hatfill SJ, Williamson C, Kirby R, & von wechmar MB (1990) Identification and localization of Aphid lethal paralysis virus in thin tissue sections of the *Rhopalosiphum padi* aphid by in situ nucleic acid hybridization. *J Invertebr Pathol* 55(2):265-271.

117. Gildow FE & D'Arcy CJ (1990) Cytopathology and experimental host range of *Rhopalosiphum padi* virus, a small isometric RNA virus infecting cereal grain aphids. *J Invertebr Pathol* 55(2):245-257.

118. Fragkoudis R, Attarzadeh-Yazdi G, Nash AA, Fazakerley JK, & Kohl A (2009) Advances in dissecting mosquito innate immune responses to arbovirus infection. *J Gen Virol* 90(Pt 9):2061-2072.

119. Kingsolver MB & Hardy RW (2012) Making connections in insect innate immunity. *Proc Natl Acad Sci U S A* 109(46):18639-18640.

120. Merkling SH & van Rij RP (2013) Beyond RNAi: antiviral defense strategies in *Drosophila* and mosquito. *J Insect Physiol* 59(2):159-170.

121. Aliyari R*, et al.* (2008) Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in *Drosophila*. *Cell host & microbe* 4(4):387-397.

122. Flynt A, Liu N, Martin R, & Lai EC (2009) Dicing of viral replication intermediates during silencing of latent *Drosophila* viruses. *Proc Natl Acad Sci U S A* 106(13):5270-5275.

123. Siu RW*, et al.* (2011) Antiviral RNA interference responses induced by Semliki Forest virus infection of mosquito cells: characterization, origin, and frequency-dependent functions of virus-derived small interfering RNAs. *J. Virol.* 85(6):2907-2917.

124. Sabin LR*, et al.* (2013) Dicer-2 processes diverse viral RNA species. *PloS one* 8(2):e55458.

125. Keene KM*, et al.* (2004) RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 101(49):17240-17245.

126. Singh CP, Singh J, & Nagaraju J (2012) A baculovirus-encoded MicroRNA (miRNAs) suppresses its host miRNAs biogenesis by regulating the exportin-5 cofactor Ran. *J Virol* 86(15):7867-7879.

127. Goic B*, et al.* (2013) RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat. Immunol.* 14(4):396-403.

128. Laughton AM, Garcia JR, Altincicek B, Strand MR, & Gerardo NM (2011) Characterisation of immune responses in the pea aphid, *Acyrthosiphon pisum*. *J Insect Physiol* 57(6):830-839.

129. Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol* 43:17-37.

130. Hansen AK & Moran NA (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A* 108(7):2849-2854.

131. Oliver KM, Russell JA, Moran NA, & Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A* 100(4):1803-1807.

132. Parker BJ, Spragg CJ, Altincicek B, & Gerardo NM (2013) Symbiont-mediated protection against fungal pathogens in pea aphids: a role for pathogen specificity? *Appl Environ Microbiol* 79(7):2455-2458.

133. Christian PD*, et al.* (2005) Effective control of a field population of *Helicoverpa armigera* by using the small RNA virus Helicoverpa armigera stunt virus (Tetraviridae: Omegatetravirus). *J Econ Entomol* 98(6):1839-1847.

134. Moscardi F (2007) A Nucleopolyhedrovirus for Control of the Velvetbean Caterpillar in Brazilian Soybeans. *Biological Control: A Global Perspective: Case Studies from Around the World*, eds Vincent C, Goettel SM, & Lazarovits G (CABI), pp 344-352.

135. Huger AM (2005) The Oryctes virus: its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle, *Oryctes rhinoceros* (Coleoptera: Scarabaeidae). *J Invertebr Pathol* 89(1):78-84.

136. Micheloud GA, Gioria VV, Eberhardt I, Visnovsky G, & Claus JD (2011) Production of the Anticarsia gemmatalis multiple nucleopolyhedrovirus in serum-free suspension cultures of the saUFL-AG-286 cell line in stirred reactor and airlift reactor. *J Virol Methods* 178(1-2):106-116.

137. Uhlirova M*, et al.* (2003) Use of Sindbis virus-mediated RNA interference to demonstrate a conserved role of Broad-Complex in insect metamorphosis. *Proc Natl Acad Sci U S A* 100(26):15607-15612.

**CHAPTER 2**

**AN APHID LETHAL PALAYSIS VIRUS ISOLATE FROM THE PEA APHID IS
CLOSELY RELATED TO ALPV DERIVED FROM HONEYBEES**

Sijun Liu[1], Diveena Vijayendran[1], Jimena Carrillo-Tripp[2] W. Allen Miller[2]
and Bryony C. Bonning[1]

[1]Department of Entomology and [2]Plant Pathology and Microbiology,

Iowa State University, Ames, IA, 50011

**Abstract**

*Aphid lethal paralysis virus* (ALPV) is a dicistrovirus that was first isolated from
the bird-cherry oat aphid, *Rhopalosiphum padi*. ALPV-like virus sequences have now
been reported from many insects and insect predators. We identified a new strain of
ALPV from the pea aphid, *Acyrthosiphon pisum*, designated; ALPV-Ap. ALPV-Ap has a
single-stranded RNA genome of 9,940 nucleotides. Phylogenetic analysis of the genome
sequence shows that ALPV-Ap is closely related to ALPV-AM, an ALPV strain isolated
from honeybees in Spain and Brookings, an ALPV strain isolated from honeybees in
South Dakota, USA. The similarity of ALPV-Ap and ALPV-AM is up to 88% at the
RNA level, compared to 78-79% between ALPV-Ap and other ALPV isolates. Similarly,
the sequence identity of proteins between ALPV-Ap and ALPV-AM is 98-99% for both
ORF 1 and ORF2, while only 85-87% for ORF1 and 91-92% for ORF2 between ALPV-
Ap and other ALPV isolates. Sequencing of RACE products and cDNA clones of the
virus genome revealed sequence variation in the 5' untranslated regions (5'-UTR) and in

ORF1, indicative of a mixed population of ALPV-Ap, which could have important biological implications for ALPV host range and infectivity.

**Introduction**

The Dicistroviridae is an emerging family of single-stranded RNA viruses known to infect pest and beneficial arthropods such as aphids, the glassy winged sharpshooter, honeybees and shrimp. Members of this virus family can be highly pathogenic in arthropods (1). For instance, *Israeli acute paralysis virus* (IAPV) of bees and other dicistroviruses have been implicated in colony collapse disorder (CCD) in honeybees (2). Two dicistroviruses are known to infect aphids, *Aphid lethal paralysis virus* (ALPV) and *Rhopalosiphum padi virus* (RhPV) (3).

The original isolate of *Aphid lethal paralysis virus* (ALPV), (NCBI accession No. NC_004365.1) was first isolated from the bird cherry-oat aphid, *Rhopalosiphum padi*) in South Africa. We designated this isolate as ALPV-RhP based on the species name of the original host used in identification of the virus. ALPV-RhP is a single-stranded RNA (ssRNA) of 9,812 nucleotides. *R. padi* infected with ALPV displayed uncoordinated movement and death by paralysis. ALPV infection also dramatically reduced aphid populations in nature (3, 4). Recently, ALPV- like virus isolates has been identified from various species of insects and insect predators (5-8).

An isolate of ALPV (NCBI accession No. JX480861.1) was identified in a wild population of the milkweed aphid, *Aphis nerii* in Northern Israel (ALPV-AN). The ALPV-AN isolate did not show any obvious pathogenic effects in *A. nerii* but was highly pathogenic to the green peach aphid, *Myzus persicae* in laboratory experiments (5). Three

reports describe ALPV-like viruses in honeybee, *Apis mellifera* populations that were being investigated in relation to CCD. A partial sequence of an ALPV isolate was reported in *A. mellifera* in Brookings, South Dakota, designated ALPV-Am-Brookings (NCBI accession No. HQ871932.1) (6). Nearly full length and partial sequences of ALPV isolates were identified in *A. mellifera* from Spain, designated ALPV-Am-Spain (NCBI accession No. JX045858.1) and Belgium, designated ALPV-Am-Belgium (7, 8). ALPV-like sequences (NCBI accession No.  JQ320375.1) were also reported in fecal samples of the Rickett's big-footed bat,  *Myotis ricketti* and the great round leaf bat, *Hipposideros armiger*, designated ALPV-Bf (9). 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.

The identification of ALPV isolates has been accelerated by the use of Next Generation Sequencing (NGS) technologies (10). Most of the ALPV and ALPV-like viruses from *A. mellifera* and bat fecal samples were identified using *de-novo* assembled sequence data. Emerging evidence suggests that ALPV-like viruses have a broad host range. In addition to detection in aphids, honeybees and bats, ALPV-like sequences have been identified in an EST library of the western corn rootworm, *Diabrotica virgifera virgifera* (S. Liu et al. unpublished data) and membrane feeding assays showed that ALPV can infect various species of aphids and whiteflies, *Bemisia tabaci* (11). 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, detection of related viruses in *A. mellifera* may be problematic in this regard. It is unclear if the virus is pathogenic in *A. mellifera*.

In this paper, we describe the discovery of an isolate of ALPV from the pea aphid, *Acyrthosiphon pisum*, which we designate as ALPV-Ap. The virus was assembled *de-novo* from small RNA sequencing data of a laboratory colony of *A. pisum*. We show that ALPV-Ap is closely related to the ALPV strain isolated from *A. mellifera* populations. Phylogenetic analysis suggests that ALPV have two major strains with one comprised of ALPV-Ap and ALPV isolates from *A. mellifera* and the second comprised of ALPV isolates from other aphid species and bat fecal samples. Geographic location does not seem to play a role in formation of these two strains. We also present data on the sequence variation observed in the ALPV-Ap. Our results suggest that variation in the UTR and the ORF1 coding region could be a driving force for the evolution of new strains.

## Materials and Methods

### Insects

Pea aphids, *Acyrthosiphon pisum* were purchased from Berkshire Biological (Westhampton, Massachusetts) and were raised on broad bean, *Vicia faba* in a growth chamber at 24$^{\circ}$C with a 12:12 (light:dark) cycle.

### Total RNA extraction

All instars of *A. pisum* were collected and used for RNA isolation. Approximately 50-60 aphids were homogenized in 1mL of 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 isopropanol at -80°C for optimal recovery of small RNAs (sRNA). Total RNA was re-

suspended in 30 µl nuclease-free water. RNA was quantified using a Nanadrop 2000

(Thermo Scientific) and the RNA quality assessed using a Bioanalyzer (Agilent).

**Small RNA sequencing**

*A. pisum* sRNA was isolated from total RNA and cDNA libraries constructed

using the TruSeq Small RNA Sample Prep Kit (Illumina) according to the manufacturer's

instructions. The sRNAs were sequenced for 50 cycles using an Illumina GA II platform.

sRNA isolation, sequencing library preparation and high-throughput sequencing  were

conducted at the Iowa State University DNA Facility.

**Assembly of sRNA reads**

The Illumina sequencing reads were trimmed with FASTX-Toolkit to remove

adaptors. The sRNA reads were then assembled using the Velvet assembler with multiple

k and c parameters for optimal results (12). 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 with Bioedit software. The contigs that hit

ALPV genomes were manually extracted for further manual assembly using Bioedit (13).

**Purification of ALPV-Ap virus particles from** *A. pisum*

*A. pisum* (~3 g) were ground in liquid nitrogen in a pre-cooled mortar and pestle.

The powdered aphids were transferred to a 30 mL centrifuge tube on ice. Sodium

phosphate buffer, 15 mL of 0.01M, pH 7 was added to the tube and was briefly vortexed.

Chloroform, 7.5 mL (1/2 volume of buffer) was added to the tube. The tube was shaken

vigorously and incubated on ice for 5 minutes. The tube was then centrifuged at 8,000

rpm at $4^oC$ for 25 minutes in a SS34 rotor (Beckman). The resulting supernatant was

passed through a syringe filter sterilizer (0.45 μM) into a 30 mL centrifuge tube. Sodium

phosphate buffer, 5 mL of 0.01M was added to the filtered supernatant. The supernatant

was once again centrifuged at 8,000 rpm at $4°C$ for 25 minutes in a SS34 rotor

(Beckman). The resulting supernatant was then transferred to a 30 mL ultracentrifuge

tube and centrifuged at 40,000 rpm at $4^oC$ for 2.5 hours using a 70 Ti rotor (Beckman).

The supernatant was removed and 5 mL of 0.01M sodium phosphate buffer was added to

the pellet. The tube was covered with parafilm and placed on ice at $4^oC$ overnight with

shaking. The re-suspended virus mixture was transferred into 1.5 mL centrifuge tube and

centrifuged at10, 000 rpm at $4°C$ for 5 minutes. The resulting supernatant was transferred

into a new tube and the final volume was adjusted to 5.5 mL using 0.01M 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 43,000 rpm at $4^oC$ for 3

hours using a 70.1 Ti rotor (Beckman). The resulting liquid was removed and the pellet

re-suspended overnight at $4^oC$ as previously described. The stock of purified virus was

stored at $-20°C$.

**Transmission electron microscopy**

Virions were purified as previously described and re-suspended in nuclease-free

water. Purified virions (10 μL) were placed on a carbon film grid. The grid was

negatively stained with 2% aqueous uranyl acetate for 30 seconds. The virus particles

were visualized using a JEOL 2100 scanning/transmission electron microscope using

standard procedures.

**Analysis of virus genome size**

Purified virions were added to 5X SDS-loading buffer (10% w/v sodium dodecyl sulfate, 10 mM dithiothreitol, 20% v/v glycerol, 0.2M Tris-HCl, pH 6.8 and 0.05% w/v bromophenolblue). The sample was boiled and loaded onto a 5% stacking, 12% resolving SDS-polyacrylamide gel along with 8 µL of Precision Plus, All-Blue Protein ladder (Biorad). The gel was stained with Coomasie Blue to visualize the protein bands. Total RNA was extracted from purified ALPV-Ap virions using TRIzol®reagent (Invitrogen). ALPV-Ap RNA (2 µg) was loaded onto a 1% native agarose gel along with 5 µl of 0.5-10,000 base RNA ladder (Invitrogen).The gel was stained with ethidium bromide for visualization of RNA.

**Sequence confirmation by RT-PCR, 5' and 3' RACE-PCR**

To confirm the *in silico* assembled sequences and to fill the gaps of the resulting genomic sequences, primers (Table 1) 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 reaction kit (Qiagen). The resulting PCR products were isolated and purified from agarose gels using the Qiaquick gel extraction kit (Qiagen) and sequenced by Sanger sequencing. To determine the 5'- and 3'- terminal sequences, rapid amplification of cDNA ends (RACE) experiments were carried out by using the SMARTer RACE cDNA amplification kit (Clontech) with 1 µg of viral RNA as template. The PCR reaction was prepared using the Advantage 2 PCR kit (Clontech) according to the manufacturer's recommendations with primers listed in table 1. PCR was carried out

using a BioRad MyCycler Thermal Cycler with 25 cycles of 94°C for 30 seconds, 68°C

for 30 seconds and 72°C for 3 minutes.  The 5' end RACE PCR product was ligated into

pGEM-T easy (Promega) and plasmids were transformed into TOP 10 competent cells.

Plasmid DNA was isolated and sequenced to determine the nucleotide sequence of the 5'

UTR regions. The 3'end RACE PCR product was gel purified and sequenced as

described above. No template negative controls were included for RACE and RT-PCR

experiments.

**Cloning of the virus genome**

The full length viral genome 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 1) and dNTPs followed by a

denaturing step (65 °C for 5 minutes). The mixture was then added with reaction buffer,

DTT, RNAsin (Promega) and SuperScriptIII retrotranscriptase (Life Technologies) to a

final volume of 20 μl. The reaction was incubated in a thermocycler at different

temperatures following a step program: 37 °C for 20 minutes, 42 °C for 20 minutes,

47 °C for 20 minutes, 51 °C for 20 minutes, 55 °C for 20 minutes and 70 °C for 15

minutes. PCR amplification was conducted using Platinum Taq DNA Polymerase High

Fidelity (Life Technologies) following provider recommendations (for primers AatII-T7-

ALPV-Ap + ALsT30-R) with a ramp program:  95 °C for 3 minutes, 35 cycles of [94 °C

for 45 seconds, 47 °C (increasing 0.1 °C/cycle) for 45 seconds, 68 °C for 10 minutes] and

68 °C for 15 minutes.  Phusion High-Fidelity DNA Polymerase (New England Biolabs)

was used with GC buffer and primers AatII-T7-ALPV-RhP + ALsT30-R (or AatII-T7-

ALPV + ALsKpnI-R, Table 1) following the kit protocol using a similar program: 98 °C

for 30 seconds, 35 cycles of [98 °C for 15 seconds, 47 °C (increasing 0.2 °C/cycle) for 30 seconds, 72 °C for 6 minutes] and 72 °C for 10 minutes. Amplicons were cleaned from the agarose gel using Ultra free-DA Centrifugal Filter Units (Millipore) and cloned with the BigEasy Long PCR Cloning Kit (Lucigen) following the manufacturer's protocol. Positive colonies were selected by PCR screening with primers ALs6701F and ALs7199R (Table 1) and by restriction digestion analysis. Clones with inserts >9.5 kb were fully sequenced. Sequence assembly was performed using DNADynamo software (BlueTractorSoftware Ltd, UK). Alignment and pairwise comparison of resulting clones was done using CLC Main Workbench software (CLC Inc, Aarhus, Denmark).

**Bioinformatics analysis**

Sequence alignment, similarity, identity and homology analyses of RNA and proteins were performed by using Bioedit. ClustalW was used for multiple sequence alignments (EMBL-EBI, UK). Phylogenetic trees were constructed by using the test maximum likelihood tree method and with 500 bootstrap replications, which was performed using MEGA (version 6.0) (14). The RNA secondary structure of the ALPV 5'UTR from multi-aligned sequences was predicted by LocARNA using the default setting (15).

**Results**

**Assembly of the viral genome**

Around 21 million reads of *A. pisum* sRNA was generated from Illumina sequencing. After removing adaptor sequences and trimming low quality bases, the

sRNA reads were assembled by the Velvet assembler (16). A variable number of contigs (50-2,500 nt) was assembled from the sRNA reads depending on the parameters (varying hash length k and coverage c) used in assembly. BLAST analysis against the NCBI nr database or against ALPV-RhP genomic RNA with either BLASTx or BLASTn algorithms were conducted to search for contigs homologous to ALPV. Approximately 1-2% of the contigs (minimum 50 nt) hit the ALPV genome with the longest being 2,388 nt. The ALPV- derived contigs were extracted and manually re-assembled using ALPV-RhP as a reference sequence. Three ALPV fragments of 5,052, 3,766 and 1,110 nt, were assembled by manual alignment. The gaps between the fragments were filled by sequencing of the RT-PCR fragments with primers designed to cover the gaps. The assembled and RT-PCR joining ALPV-Ap genomic sequences have 9,940 nt, which is 128 nt longer than that of the ALPV-RhP sequence (9,812 nt). The assembled ALPV-Ap genome sequence was confirmed by sequencing of RT-PCR fragments. Ninety-eight percent 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 PCR and by sequencing of the cDNA clones of ALPV-Ap (see below). The genome organization of ALPV-Ap is shown in Figure 1 and the proteins encoded are listed in Table 2.

**Confirmation of infection with ALPV-Ap**

Assembled sRNA sequences from *A. 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 the virions from infected aphids. Electron micrographs of purified ALPV-Ap virions showed

icosahedral virus particles with a diameter of ~27 nm, similar to the documented size of
ALPV particles (Figure 2a). In addition, the purified viruses were subjected to SDS-
PAGE and stained using Coomasie Blue. The three major capsid proteins of ALPV were
observed. Capsid proteins one and two have similar molecular masses of 27.03 kDa and
27.06 kDa respectively, and capsid protein three is 29.52 kDa (Figure 2b). To
demonstrate that virions encapsulated the full length ALPV-Ap genome, viral RNA was
isolated from purified virions and visualized on a native agarose RNA gel (Figure 2c). An
RNA band of ~ 10,000 nt long was observed indicating that the virions encapsidate full-
length ALPV-Ap RNA.

**Determination of 5' and 3'end sequences by RACE PCR**

RACE PCR was used to determine the 5' and 3'ends of ALPV-Ap. The RACE
PCR products were generated by using purified RNA as described in the methods section.
The 3'end PCR products were isolated and sequenced. A dominant 3' RACE PCR
product was consistently obtained in two different experiments. The sequenced PCR
product ended with poly-A with the same length of sequence as the assembled 3'end
sequences (excluding the poly-A sequences). The 5' RACE PCR product of ALPV-Ap
was a smear on the agarose gel, ranging from ~350-500 nt in size. The gel band of ~400-
500 nt was purified, cloned and sequenced. A total of 11 clones were sequenced for the 5'
RACE clones. All of the clones showed the same 5'end sequences of ALPV-Ap.
Interestingly, the 5'end sequence observed from RACE PCR was 38 nt shorter (9,888 nt)
than that of the sequences assembled from sRNA reads and differed from the published
ALPV-RhP sequence.

**Confirmation of the 5'sequence of ALPV-Ap by sequencing clones of ALPV-Ap**

The ALPV-Ap genome assembled from sRNA was 9,926 nt in length, while RACE resulted in 9,888 nt with 38 nt less at the 5'end. To confirm the 5'end sequences of ALPV-Ap, we used RT-PCR to generate cDNA from the viral RNA that was isolated from purified ALPV-Ap. Full length genomes were amplified by two step RT-PCR using primers either derived from ALPV-Ap (based on the 5' and 3' RACE results) or from ALPV-RhP (Table 1).

Several rounds of amplification and cloning were carried out, resulting in >1,500 colonies. Among those colonies, 116 positive colonies were obtained from PCR analysis of 879 colonies. After restriction digestion analysis, only 14 of the PCR positive clones had inserts > 9.5 kb. Among the 14 clones, 12 were from the cDNA generated by using the 5'primer derived from RACE results. Only 2 putative full-length clones were obtained from cDNA using the ALPV-RhP 5'primer. The 5'UTR of the 14 clones was sequenced.

Alignment of the 5'end sequences of the 14 clones and the 5'end sequences of the Velvet assembled and RACE PCR-derived sequences (Figure 3) suggested ALPV-Ap has the same 5'end sequence as the documented ALPV-RhP isolate. However, potential variation may exist, because the majority of the cDNA clones were generated by using the RACE product-derived 5'primer. Sequencing of the cloned cDNA also confirmed that the 38 nt sequence which was not found by RACE PCR was part of the ALPV-Ap sequence. (Hence the cDNA generated from the ALPV-RhP primer has 53 more bases at the 5'end than that of the 5'end sequencing result from RACE). The total genomic RNA

of ALPV-Ap is 9,940 nt (excluding the poly A tail), 129 nt more than that of ALPV-RhP, which is the longest ALPV- like virus discovered so far.

**Comparison of sequences of ALPV isolates**

The genome organization of ALPV is shown in Figure 1. The ALPV-Ap genome has a G+C content of 38.3%. The non-structural ORF1 of ALPV-Ap has 2,037 amino acids (aa), which is 2 aa more than that of ALPV-RhP, ALPV-Am-Spain and ALPV-Bf, but 1 aa less than  that of ALPV-AN.  The structural protein precursor has 809 aa, which is the same as ALPV-Am-Spain, but 2 aa more than that of ALPV-RhP and 1 aa more than  that of ALPV-AN and ALPV-Bf.  Numbers of aa of each peptide and molecular mass (Mr) of the peptides are summarized in Table 2.

Comparisons of ALPV-Ap RNA and protein sequences with known ALPV sequences are summarized in Table 3 and Table 4. The overall RNA homology between ALPV-Ap and the other aphid ALPV strains and bat fecal-derived ALPV is 78-79%. Interestingly the RNA homology between ALPV-Ap and ALPV-Am-Spain (lacking parts of the 5' and 3'UTR sequences), is 88%, about 10% higher than that of the other ALPV isolates.  The intergenic region – IRES (IGR-IRES) sequences among all ALPV isolates are highly conserved (97-99% homology). The RNA sequences of the coding regions were relatively conserved with more than 81% homology between ALPV-Ap and other isolates. Again, sequence homology between ALPV-Ap and *A. mellifera* isolates of ALPV was much higher (92-96%) in ORF1 and ORF2 relative to other isolates.

Comparison of ORF1 and ORF2 protein sequences also shows that ALPV-Ap has higher sequence identity with ALPV-AM (Table 4, Fig 4). Protein identities between

these two isolates are 98.4% for ORF1 and 98.5% for ORF2, while identities of ALPV-Ap and three other ALPV isolates are 87% (ORF1) and 91-92% (ORF2) (Table 4). Although only a partial sequence is available for the ALPV-Am-Brookings strain, sequence comparison of the Brookings strain with the corresponding sequences of other ALPV strains showed that the Brookings strain is highly homologous to ALPV-Ap and the ALPV-Am-Spain isolate (data not shown). Indeed, phylogenetic analysis shows that the ALPV isolates form two groups with ALPV-Ap and the ALPV *A. mellifera* isolates grouping together, separate from other ALPV isolates (Figure 5).

**The 5' and 3'UTRs of ALPV-Ap**

The 5'UTR of ALPV-Ap was aligned to the known ALPV-RhP 5'UTR sequence (Figure 6a) and a consensus secondary structure of ALPV (Figure 6b) was constructed by using LocARNA program. Comparison of the 5'UTR and 3'UTR between the two ALPV-Ap and ALPV-RhP isolates showed that multiple insertions occurred in several regions of ALPV-Ap UTRs, which explains why the genomic sequence of ALPV-Ap is longer compared to the published ALPV-RhP sequence. Indeed, the 5'UTR and 3'UTR regions showed much less similarity between ALPV-Ap and other ALPV isolates with only 63% sequence similarity between ALPV-Ap and other ALPVs at the 5'UTR region, while the other isolates share 90-95% homology at the 5'UTR. Lower similarity was also observed at the 3'UTR between ALPV-Ap and other ALPV sequences (Table 3). These results suggest that the 5' and 3'UTR regions of ALPV are less conserved than other regions of the virus genome. Although significant sequence variation was observed in the 5' and 3' UTR regions, 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 the LocARNA folding program (Figure 6b). This suggests that most of the inserted sequence in the 5'end of ALPV-Ap is at the structural loop region and therefore may not affect the common 5'UTR structure of ALPV.

**Sequence analysis of ALPV-Ap full length clones**

For determination of the 5'sequence of ALPV-Ap, we sequenced the 5'UTR regions of 14 near full-length cDNA clones of ALPV-Ap. In addition to the variations observed in the 5'UTR, we also noticed sequence variation in the coding sequences. To assess variations in the ALPV-Ap sequence in the coding region, we sequenced full genomic sequences of 14 full length clones. The sequencing results showed that a single mutation in three of the 14 cDNA clones introduced a stop codon or fragment deletion that truncated the protein product of ORF1.This is expected to be lethal to the virus clones. Two clones had short 3'UTR, which might also affect transcription of the viral RNA. Interestingly, none of the 14 clones had mutations that might disrupt translation of structural protein from ORF 2 (data not shown). As each genome was PCR-amplified in one piece, these results represent a snap shot of sequence diversity within the viral population.

**Discussion**

In this paper, we identified and characterized the genome of ALPV-Ap, a novel isolate of ALPV with a longer genome sequence (9,940 nt) than other isolates reported so far. ALPV-Ap is the first ALPV isolate infecting aphids that is phylogenetically more closely related to *A. mellifera* isolates of ALPV than to other aphid isolates. Previous

studies using both virus isolation followed by cloning and metagenomic approaches have identified seven different ALPV isolates from aphids, honeybees and bat feces. In addition to ALPV-Ap, we have identified ALPV sequences from soybean aphid, *Aphis glycines* and western corn rootworm, *Diabrotica virgifera virgifera* (Liu *et al.* unpublished data), demonstrating that ALPV-like viruses have a broad host range.

To obtain the complete genome sequence of ALPV-Ap, we employed Illumina sRNA sequencing and assembly, RT-PCR and RACE PCR methods. RACE PCR was limiting and uncovered only part of the 5' UTR region but was used to determine the 3' UTR region presumably facilitated by the poly (A) tail of the virus genome. The assembled sequence and the RT-PCR method used to amplify the 5' UTR sequence resulted in a longer genome sequence than previously reported for this virus. These results highlight the limitations of the commonly used RACE PCR for identification of the terminal regions of the genome of newly discovered viruses. The presence of stable secondary structures in the 5' UTR may block the RT reaction and prevent acquisition of sequence using this approach. Virus secondary structures have been known to interfere with RACE PCR (17). No additional stem-loop secondary structures were predicted in the 5' UTR of the ALPV-Ap genome. It remains unclear whether the sequence of the ALPV-Ap represents a full-length sequence of the virus, considering that the 5'end primer used for generating the viral sequence was based on the ALPV-RhP sequence.

The current ALPV classification is determined by the similarity of the capsid proteins (CP). If the homology between the CP sequences is 90% or more, viruses are considered different isolates of the same species (18). Indeed, the similarity of all ALPV CP sequences identified so far is over 90%, irrespective of the insect host of the ALPV

isolate (Table 4). This suggests that the sequence of ORF2 is conserved among ALPV-like viruses. However, sequence comparison between ALPV-Ap, ALPV isolates from *A. mellifera* and ALPV isolated from other aphids demonstrates that sequence diversity is much greater at the RNA level. For instance, the ALPV-Ap genome is only 78.3% similar to that of ALPV-RhP, isolated from *R. padi*. In addition, although ALPV-Ap is also isolated from aphid, phylogenetic analysis shows that ALPV-Ap has a greater sequence similarity to ALPV isolates from *A. mellifera*. The fact that there are two distinguishable genomic branches suggests that the current ALPV classification criteria may not be sufficient for ALPV-like viruses.

In begomoviruses, two viruses with CP homology about 90% are classified as two different viruses. For example, *African cassava mosaic virus* (ACMV) and *East African cassava virus* (EACMV) are two different viruses, but their CP protein identity is up to 93% (19). Hence, it is possible that ALPV-Ap is a different species, which has two isolates, e.g. ALPV-Ap and ALPV honeybee isolate. The first identified ALPV, ALPV-RhP, was shown to cause disease symptoms in *R. padi* in the field (3). However, no acute pathogenicity was observed from ALPV-Ap infection of *A.pisum* under ideal rearing conditions.

Sequence comparison of the different ALPV-Ap and other ALPV isolates show major nucleotide sequence variations (>60%) in the 5' and 3' UTR (Table 3). Changes in the UTR region may enable evasion from the host RNA interference-based antiviral immune response. While some isolates may trigger the RNAi response, others could evade and successfully replicate in the host. This strategy is used by the flaviviruses, *West Nile virus* (WNV) and *dengue virus* (DENV) (20). Sequence variation in the UTR

may also correlate with pathogenicity of the virus isolate. Structural differences in the 5'
UTR of DENV correlates with disease severity in humans (21).  Experiments designed to
take a closer look at the UTR of these different isolates would help in understanding the
variations observed.

None of the 14 clones of ALPV were infectious and acquisition of an infectious
clone of ALPV has proven challenging. A common problem encountered for infectious
clones is the presence of natural mutations in all regions of many genomes within a virus
population. The concept of viral quasispecies with a high rate of mutation resulting in
many natural variants may be at play in this scenario. 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 noninfectious
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. These new data
will help in our understanding of virus evolution, and toward understanding the
determinants of pathogenicity and host specificity. Previous regions of the virus genome
that have not been heavily investigated such as the UTRs could emerge as key players in
host-virus molecular interactions.

**Author Contributions**

SL performed the bioinformatics analysis. DV conducted all laboratory experiments except for the cloning of ALPV-Ap which was carried out by JCT. BCB and WAM was involved in experimental design and discussion of data analysis. All authors contributed to data analysis and writing of the manuscript.

**Acknowledgements**

**References**

1. Bonning BC & Miller WA (2010) Dicistroviruses. *Annu Rev Entomol* 55:129-150.

2. Cox-Foster DL*, et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318(5848):283-287.

3. Williamson C, Rybicki EP, Kasdorf GF, & Von Wechmar MB (1988) Characterization of a new picorna-like virus isolated from aphids. *J Gen Virol* 69:787-795.

4. D'Arcy CJ, Burnett PA, Hewings AD, & Goodman RM (1981) Purification and characterization of a virus from the aphid *Rhopalosiphum padi*. *Virology* 112(1):346-349.

5. Dombrovsky A & Luria N (2013) The Nerium oleander aphid *Aphis nerii* is tolerant to a local isolate of Aphid lethal paralysis virus (ALPV). *Virus Genes* 46(2):354-361.

6. Runckel C*, et al.* (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS One* 6(6):e20656.

7. Ravoet J*, et al.* (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS One* 8(8):e72443.

8. Granberg F*, et al.* (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid lethal paralysis, Israel acute paralysis and Lake sinai viruses. *PLoS One* 8(2):e57459.

9. Ge X*, et al.* (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86(8):4620-4630.

10. Liu S, Vijayendran D, & Bonning BC (2011) Next generation sequencing technologies for insect virus discovery. *Viruses* 3(10):1849-1869.

11. Van Munster M*, et al.* (2002) Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family Dicistroviridae. *J Gen Virol* 83(Pt 12):3131-3138.

12. Liu S, Chougule NP, Vijayendran D, & Bonning BC (2012) Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. *PLoS One* 7(9):e45161.

13.     Hall (2005) Bioedit 7.0.5.  (Department of Microbiology, North Carolina State University).

14.     Tamura K, Stecher G, Peterson D, Filipski A, & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729.

15.     Will S, Joshi T, Hofacker IL, Stadler PF, & Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* 18(5):900-914.

16.     Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821-829.

17.     Murakami R, Suetsugu Y, Kobayashi T, & Nakashima N (2013) The genome sequence and transmission of an iflavirus from the brown planthopper, *Nilaparvata lugens*. *Virus Res* 176(1-2):179-187.

18.     Chen YP, Nakashima, N., Christian, P. D., Bakonyi, T., Bonning, B.C., Valles, S. M. and Lightner, D. (2012) Family *Dicistroviridae*. *Virus taxonomy, ninth report of the international committee on taxonomy of viruses*, ed King AMQ, Adams, Michael J., Carstens, Eric B. and Lefkowitz, Elliot J. (Elsevier Academic, London), pp 840-845.

19.     Oteng-Frimpong R*, et al.* (2012) Complete genome sequencing of two causative viruses of cassava mosaic disease in Ghana. *Acta Virol* 56(4):305-314.

20.     Schnettler E*, et al.* (2012) Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells. *J Virol* 86(24):13486-13500.

21.     Leitmeyer KC*, et al.* (1999) Dengue virus structural differences that correlate with pathogenesis. *J Virol* 73(6):4738-4747.

**Table 1.** Primers for RT (R), PCR (P), sequencing (S) and RACE PCR (RACE) reactions. Letters at the end of primer name, F denotes forward primer and R denotes reverse primer.

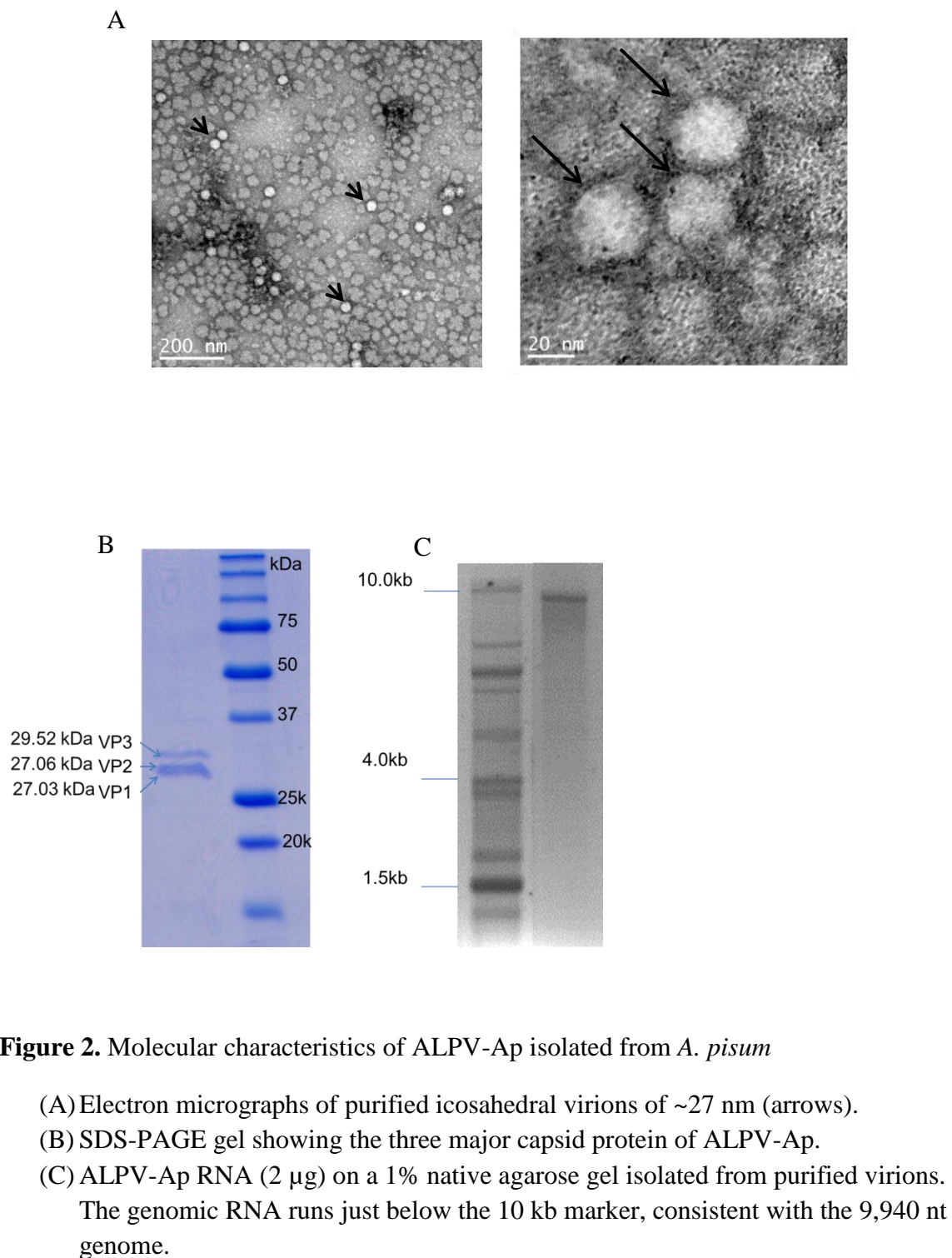| Primer name | Sequence (5'-3') | Genome position |
|---|---|---|
| **(P)** AatII-T7-APV-AP | ACTGACGTCTAATACGACTCACTATAGGG AATAACCTGAAACCTTAGACC | 52-75 |
| **(P)** AatII-T7-ALPV-RhP | ACTGACGTCTAATACGACTCACTATAGGG TTAAATAAGAAACTATATAT | 1-19 |
| **(P)** ALsT30-R | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG AAAAATAACATTTACATACAATACGTACTA AACC | 9,914-9,940 |
| **(P)** AlsKpnI-R | ATCGGTACCAGAAAAATAACATTTACATA CAATACGTACTAAACC | 9,914-9,940 |
| **(P)** Als6701F | GTAATTTAAGGCTTAGTTATTTAAC | 6,749-6,774 |
| **(P)** Als7199R | ACTAATAATAAACGCCCAGCTTG | 7,225-7,248 |
| **(P)** 1-1 F | CTTAGACCCGTAAGATCCTAAG | 67-89 |
| **(R/P)** 1-1R | AGTGTCAAATGCGTAAATCA | 546-566 |
| **(R/P)** 1 R | CCATGATGGTTGTTCCCATA | 3,208-3,228 |
| **(P)** 1-2 F | GGGAATATCTATTATATCACACACG | 3,226-3,251 |
| **(P)** 3-1 F | GTCCAGTTCAAGAAGACAATAC | 8,981-9,003 |
| **(R/P)** 3 R | CCTAAGTTACACTTCTTGATCCTC | 9,882-9,906 |
| **(R/P)** AP1-769F | CTTCAACTAACGGTATCTCACG | 820-842 |
| **(R/P)** AP1-1539F | GTGGCTTTGATTACGACAATC | 1,590-1,611 |
| **(R/P)** AP1-2300F | CGTATTTGGAGATATTCCAGCT | 2,351-2,373 |
| **(R/P)** AP1-3081F | GCTCTATTAAGGAGAACTAAACG | 3,132-3,155 |
| **(R/P)** Als3750F | AATTGCGTGTGTCTGTTGACTCG | 3,795-3,818 |
| **(R/P)** AP1-4523 | ACAAGCTTACAAACACATTATCG | 4,574-4,597 |
| **(R/P)** AP1-5295F | GCTATTATTGGAATAAACGGAGAC | 5,346-5,370 |
| **(R/P)** API-6059F | CGTAAATCATTGTTGTCTTGCTG | 6,110-6,133 |
| **(R/P)** ALPVp5 | GTAATTTAAGGCTTAGTTATTTAAC | 6,749-6,774 |
| **(R/P)** Als7096F | ATGTATAAAGCGAAATGCGAACG | 7,144-7,167 |
| **(R/P)** AP1-7775F | CCAATGCAGATGGAGTTGATATG | 7,826-7,849 |
| **(R/P)** AP1-9142R | CTGCTTCTGTGAGTATATTGACC | 9,171-9,194 |
| **(S)** SL1 | CAGTCCAGTTACGCTGGAGTC | vector |
| **(S)** NZ RevC | AAA TGG TCA GTT AAT CAG TTC T | vector |
| **(RACE)** 5' end | ACGAGCTCCGACTAAGGACGTACGCAGT ATCCT | 375-409 |
| **(RACE)** 3' end | TGCGAGACTCAGAAGAGAGCGTGAGCGT CAGGT | 9,571-9,605 |
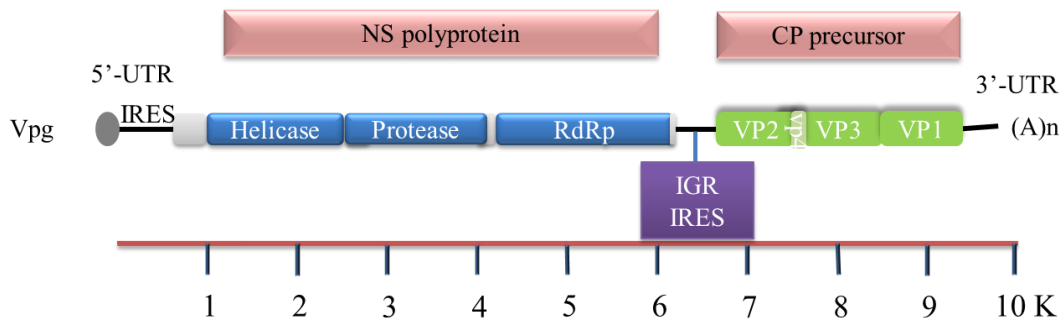
**Figure 1.** Genome organization of ALPV-Ap. The virus encodes two polyproteins, the non-structural (NS polyprotein) and the capsid protein (CP precursor polyprotein). Two internal ribosomal entry sites (IRES) are present at the 5' end and the intergenic region (IGR) between ORF1 and 2. The virus genome is capped at the 5' terminus with a viral protein genome-linked (Vpg) while the 3' end has a poly (A) tail.

**Table 2.** Summary of predicted ALPV-Ap encoded proteins

| Proteins | Size (aa) | Molecular mass (kDa) |
|---|---|---|
| ORF1 (NS polyprotein) | 2037 | 231.46 |
| ORF2 (CP precursor ) | 809 | 89.58 |
| VP1 | 242 | 27.03 |
| VP2 | 241 | 27.06 |
| VP3 | 268 | 29.52 |
| VP4 | 58 | 6.01 |

**Figure 2.** Molecular characteristics of ALPV-Ap isolated from *A. pisum*

(A) Electron micrographs of purified icosahedral virions of ~27 nm (arrows).
(B) SDS-PAGE gel showing the three major capsid protein of ALPV-Ap.
(C) 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 9,940 nt genome.
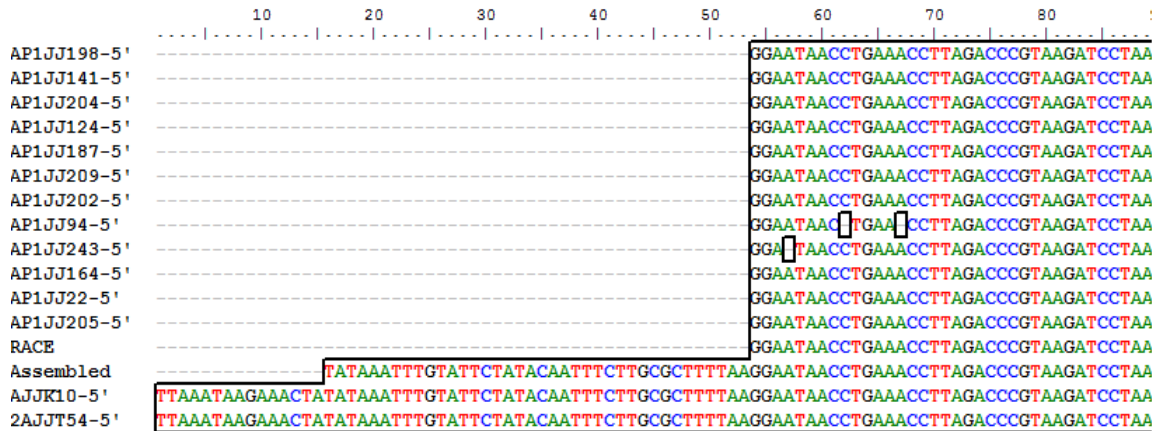
**Figure 3.** 5'end alignment of ALPV-Ap cDNA clones, assembled ALPV-Ap and RACE results. 5'end sequences of 14 ALPV-Ap cDNA clones, the 5'end sequences from sequencing of clones of RACE products (RACE) and the 5'end of in silico assembled sequences (Assembled) were aligned. While the majority of ALPV-Ap clones aligned with the RACE results some clones had the same sequence as the bioinformatics assembled ALPV sequences. Start codon begins at nucleotide 583.

**Table 3.** Percentage nucleotide sequence identity between different ALPV isolates. Ap: ALPV isolate from *A. pisum*, AN: ALPV isolate from *A. nerii*, RhP: ALPV isolate from *R. padi*, Bf: ALPV isolate from bat fecal sample, and Am-Spain: ALPV isolate from *A. mellifera* in Spain.

| | Ap | AN | RhP | Bf | Am-Spain |
|---|---|---|---|---|---|
| **Full length** | | | | | |
| Ap | 100 | | | | |
| AN | 78.9 | 100 | | | |
| RhP | 78.3 | 89.3 | 100 | | |
| Bf | 78.7 | 89.3 | 88.7 | 100 | |
| Am-Spain | 87.7 | 78.4 | 77.7 | 78.5 | 100 |
| **5'- UTR /3'-UTR/IGR-IRES** | | | | | |
| Ap | 100 | | | | |
| AN | 63.4/60.1/97.3 | 100 | | | |
| RhP | 63.4/59.3/97.8 | 89.6/91.9/98.4 | 100 | | |
| Bf | 62.2/60.1/96.8 | 92.7/95.2/97.3 | 95.5/94.1/98.9 | 100 | |
| Am-Spain | ND/ND/99.4 | ND/ND/96.8 | ND/ND/98.4 | ND/ND/97.3 | 100 |
| **ORF1/ORF2** | | | | | |
| Ap | 100 | | | | |
| AN | 81.1/83.3 | 100 | | | |
| RhP | 80.8/81.4 | 90.7/84.5 | 100 | | |
| Bf | 81.0/83.1 | 87.7/90.2 | 88.8/85.4 | 100 | |
| Am-Spain | 95.9/92.2 | 81.1/83.7 | 80.9/81.5 | 81.4/83.7 | 100 |

**Table 4.** Protein sequence identity between ALPV isolates. Ap: ALPV isolate from *A. pisum*, AN: ALPV isolate from *A. nerii*, RhP: ALPV isolate from *R. padi*, Bf: ALPV isolate from bat fecal sample, and Am-Spain: ALPV isolate from *A. mellifera* in Spain.

| | Ap | AN | RhP | Bf | Am-Spain |
|---|---|---|---|---|---|
| **ORF 1/ORF2** | | | | | |
| Ap | 100 | | | | |
| AN | 87.5/92.0 | 100 | | | |
| RhP | 87.1/90.2 | 95.6/93.6 | 100 | | |
| Bf | 87.5/91.3 | 94.0/97.4 | 95.5/94.3 | 100 | |
| Am-Spain | 98.4/98.5 | 87.7/92.9 | 87.2/90.7 | 87.8/91.9 | 100 |
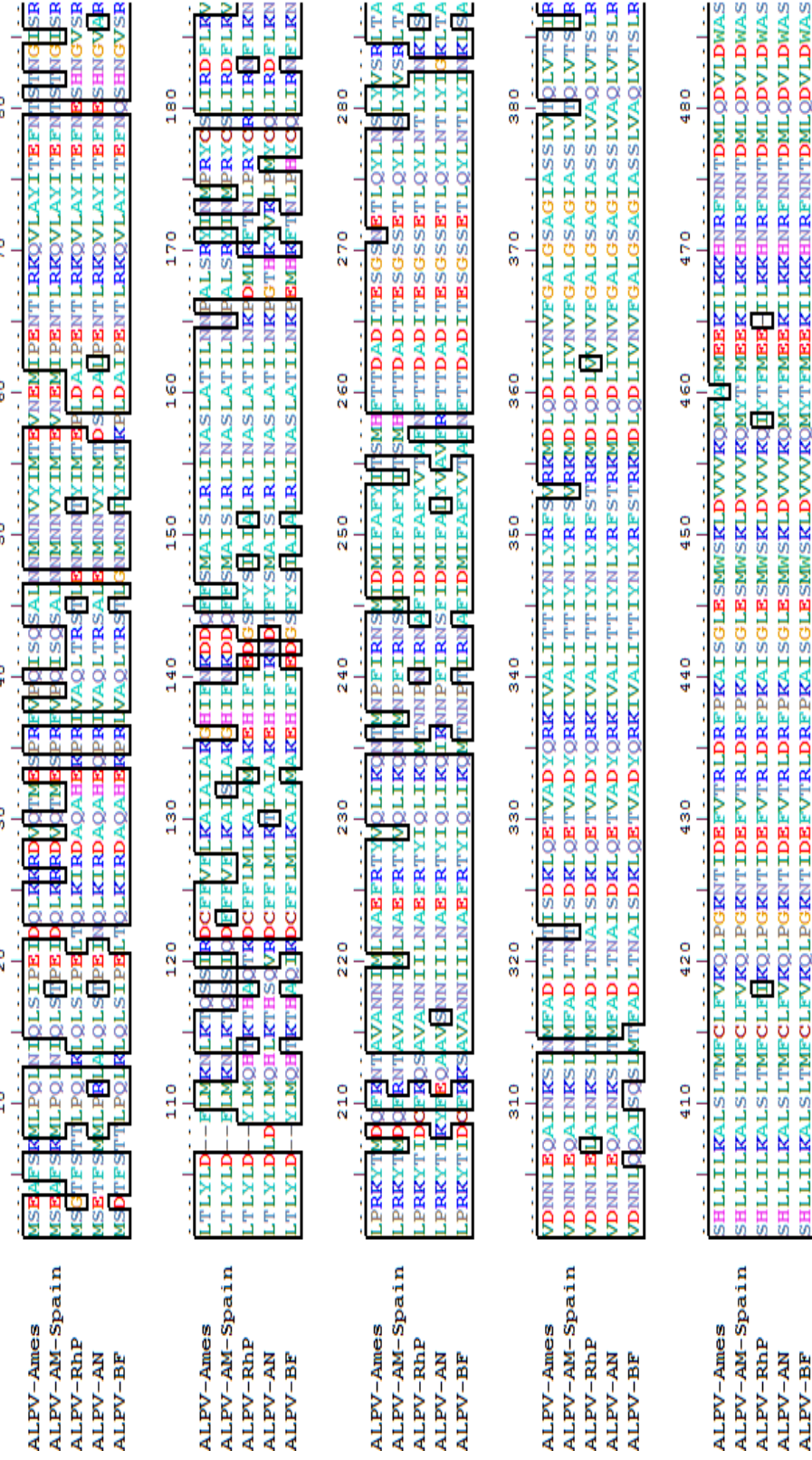
**Figure 4.** Protein sequence alignment between different ALPV isolates. A) Partial sequence alignment for non-structural proteins. B) Sequence alignment for structural proteins. ClustalW multiple alignment program was used for conduct alignment. The alignments were performed using BioEdit. Figure shows sequence similarities of ALPV-Ap to ALPV-AM from *A. mellifera*

B
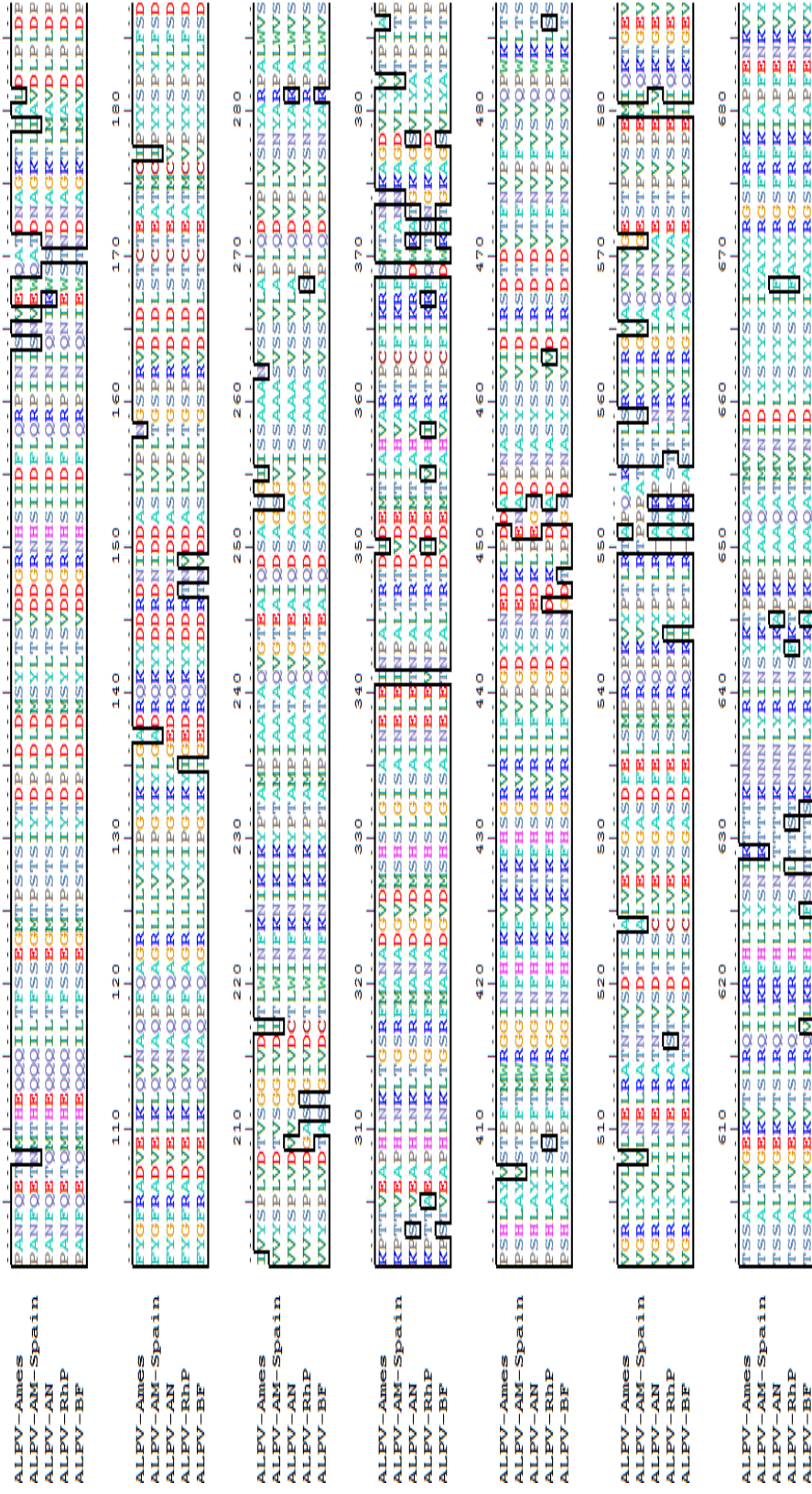


**Figure 4.** Protein sequence alignment between different ALPV isolates. A) Partial sequence alignment for non-structural proteins. B) Sequence alignment for structural proteins. ClustalW multiple alignment program was used for conduct alignment. The alignments were performed using BioEdit. Figure shows sequence similarities of ALPV-Ap to ALPV-AM from *A. mellifera*
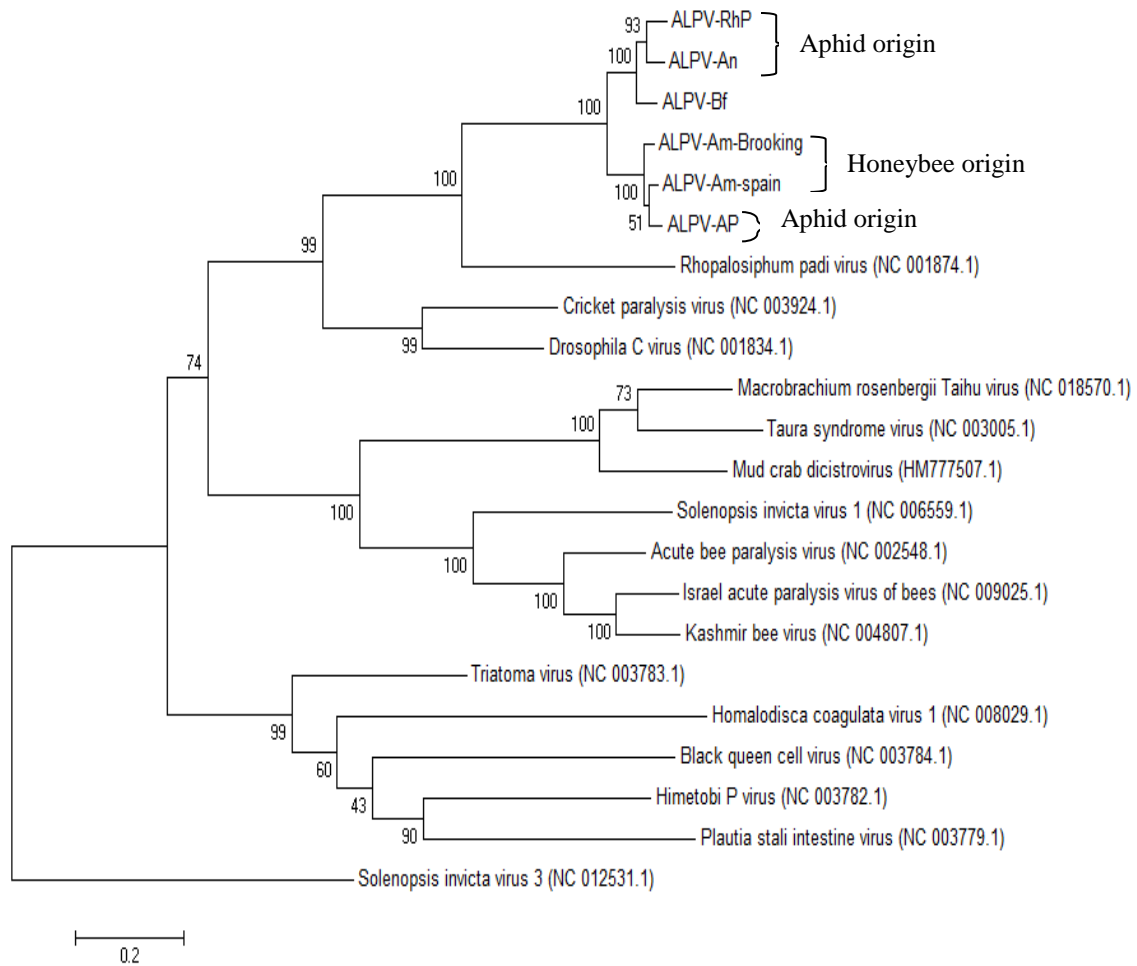
**Figure 5.** Phylogenetic tree of dicistroviruses infecting insects. ALPV genomic sequences were downloaded from NCBI (see accession number in main text). Sequence of ALPV-Am-Spain is only near full-length. ALPV-Am- Brookings is only partial sequence. The Sequences were aligned using ClustalW. Phylogenetic tree was constructed with Test Maximum Likelihood method with 500 bootstrap replications by MEGA6.0.

A.
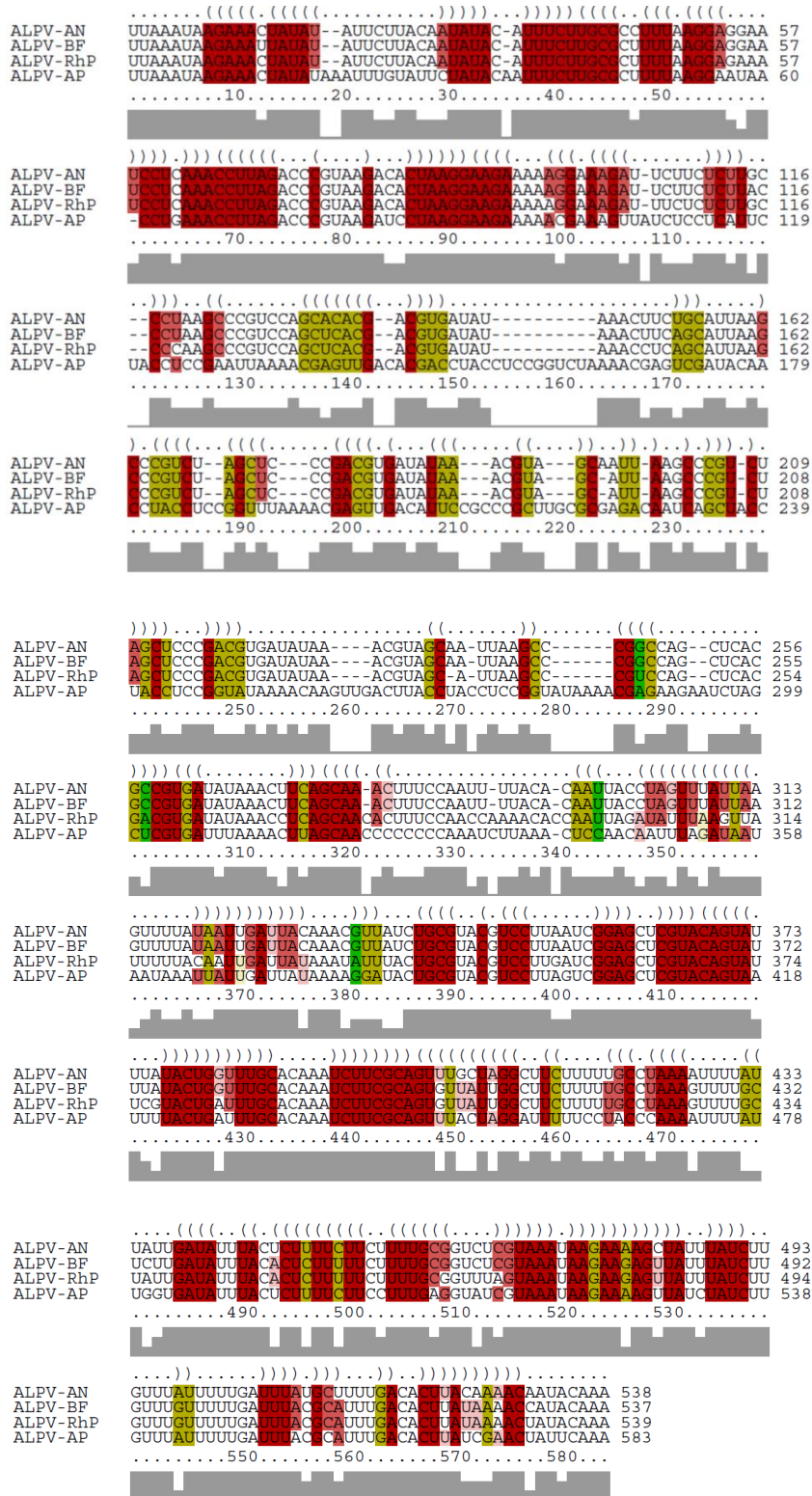
B.
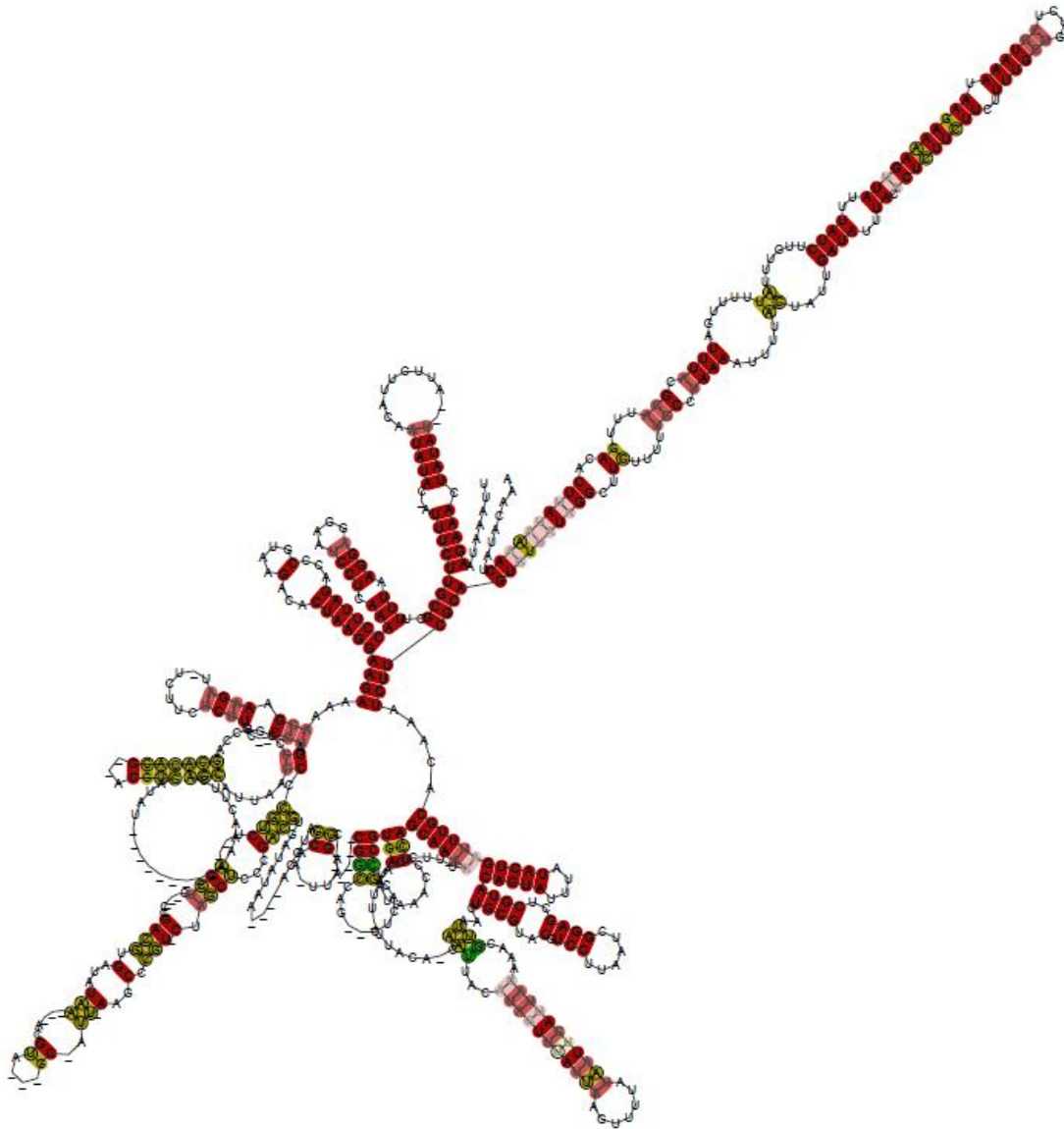


**Figure 6.** Analysis of ALPV-Ap 5' UTR. A) Alignment of ALPV-Ap 5'UTR. Top lines show putative stem-loop formation pattern. Top line showing structural constraints; '(' and ')' indicate base-pairs forming stems, '.' indicates nucleotides forming loop structure. The hue shows sequence conservation of the base pair. The saturation decreases with the number of incompatible base pairs. Thus, pattern of the hues indicates the structural conservation of the base pair. Gray box at the bottom shows homologous sequences. B) Graphic representation of the predicted consensus 5' UTR structure of ALPV. See A. for corresponding sequences and detail of the sequence alignment. Structure prediction was generated by locARNA online multiple RNA structure prediction program with standard setting.

**CHAPTER 3**

**CHARACTERIZATION OF AN APHID LETHAL PARALYSIS VIRUS ISOLATE**

**FROM A MODEL APHID SPECIES, THE PEA APHID (*Acyrthosiphon pisum*)**

Diveena Vijayendran, Sijun Liu, and Bryony C. Bonning

Department of Entomology, Iowa State University, Ames, IA, 50011

**Abstract**

Aphids are pests of major crops and ornamental plants worldwide. Aphid management has relied mostly on the use of chemical insecticides and resistant host plant cultivars. Both methods show promise for suppression of aphid populations but are increasingly challenged by the presence of resistant aphid biotypes. The use of viruses for the suppression of insect pests is receiving increased attention. To this end, we identified a novel isolate of *Aphid lethal paralysis virus* (ALPV-Ap) from the pea aphid, *Acyrthosiphon pisum* (Harris). Transmission of ALPV-Ap is horizontal with no vertical transmission detected. ALPV-Ap was detected in aphid gut tissue. Replication of ALPV-Ap in a cell line derived from the spotted cucumber beetle, *Diabrotica undecimpunctata*, DU182A was demonstrated. The characterization increases our knowledge on ALPV isolates and would aid in efforts for use in insect pest management.

**Introduction**

Aphids are important plant pests that cause major annual losses in many crops including wheat, potatoes, sugar beet, citrus plants, soybean and pea (1). Aphids are also among the most common vectors of plant viruses and transmit economically damaging plant viruses such as the *Barley yellow dwarf virus* (*Luteoviridae*) (2). While there are many plant viruses known to be vectored by aphids, only a few viruses are known to infect and replicate in the aphid. A total of eight aphid viruses have been identified: *Aphid lethal paralysis virus* (*Dicistroviridae*) (3), *Rhapalosiphum padi virus* (*Dicistroviridae*) (4), *Brevicoryne brassicae virus* (*Iflaviridae*) (5), *Acyrthosiphon pisum virus* (Unclassified) (6), *Rosy apple aphid virus* (Unclassified) (5, 7), *Myzus persicae densovirus* (*Parvoviridae*) (8), *Acythosiphon pisum densovirus* (*Parvoviridae*) and *Dysaphis plantaginea densovirus* (*Parvovoridae*) (7). These viruses were identified following observation of negative impacts on field or laboratory aphid populations. The number of aphid viruses identified remains relatively small compared to the more than 4,000 species of aphids that have been identified.

The traditional approach to virus discovery is being complemented by next generation sequencing and bioinformatics-based identification of viral sequences (9). Recently, known and novel virus sequences have been discovered from insects and from insect cell lines from next generation sequencing data (10, 11). Small RNA sequencing data also provide for identification of viruses that replicate in the host. Replicating viruses in insects are targeted by the RNA interference (RNAi) based immune response which generates virus-interfering small RNAs (vsRNA). For RNA viruses, the double stranded RNA (dsRNA) intermediate of a replicating virus is targeted by the dicer protein

for cleavage into 21-22nt vsRNA duplexes. The vsRNAs are then loaded into the multi protein RISC complex to target and silence the virus genome (12). Small RNA sequencing data can encompass such vsRNAs that can be used to assemble the viral genome.

We used sRNA sequencing data to identify viruses infecting a laboratory colony of the pea aphid, *Acyrthosiphon pisum* and identified a new strain of ALPV (*Dicistroviridae*), named ALPV-Ap (3). This virus belongs to the family *Dicistroviridae* along with several other insect infecting viruses such as the *Cricket paralysis virus* and *Israeli acute paralysis virus* of bees (13). The virions of ALPV-Ap encapsidate a single-stranded RNA genome of 9,940 nucleotides that was confirmed using molecular methods (Chapter 2). This virus is closely related to ALPV isolates found in honeybees, *Apis mellifera* identified in Europe and USA (10, 14, 15). In this paper, we present information on the biological interaction of ALPV-Ap with its *A. pisum* host. ALPV-Ap does not infect all adults in an infected population. The virus is present in the midgut tissue as shown by fluorescence in-situ hybridization (FISH). An increase in virus titer and cytopathic effects were observed on transfection of the spotted cucumber beetle cell line, *Diabrotica undecimpunctata,* DU182A with purified ALPV-Ap RNA. Our study using FISH corroborates a previously published study on the localization of ALPV in the gut tissue of infected aphids. This study provides new information on virus prevalence in an aphid population and identification of a cell line that could be used for in vitro production of ALPV toward development of virus- based aphid management strategies.

**Materials and Methods**

**Aphid rearing**

Pea aphid, *A. pisum* was maintained on broad bean, *Vicia faba* in a growth chamber at 24°C with a 12 hour light and 12 hour dark cycle. Soybean aphid, *Aphis glycines* was maintained on soybean, *Glycine max* in a growth chamber at 24°C with a 12 hour light and 12 hour dark cycle. Green-peach aphid, *Myzus persicae* was maintained on Chinese cabbage, *Brassica rapa* at room temperature (RT).

**Total RNA extraction for detection of ALPV-Ap in multiple aphid species**

Aphids were collected and placed in tubes containing TRIzol® Reagent (Invitrogen). The aphids were crushed directly in the TRIzol® Reagent using a pestle and RNA was extracted as per the manufacturer's instructions with overnight precipitation of RNA in isopropanol at -80°C. Total RNA was re-suspended in nuclease-free water. RNA quality was quantified using NanoDrop 2000 (Thermo Scientific) and the quality was assessed using a Bioanalyzer (Agilent).

**Determination of prevalence of ALPV-Ap by RT-PCR**

To test for the prevalence of ALPV-Ap in *A. pisum* population, individual adults were placed in TRIzol® reagent and total RNA was extracted as previously described. One-Step RT-PCR (Qiagen) was carried out using ALPV-Ap, 3-1F forward primer, 5' GTCCAGTTCAAGAAGACAATAC 3' and ALPV-Ap, 3R reverse primer 5' CCTAAGTTACACTTCTTGATCCTC 3'. The RT-PCR conditions were as follows: 50°C for 30 minutes, 94°C for 5 minutes, 30 cycles of 94°C for 30 seconds, 46°C for 30

seconds, 72°C for 2 minutes, followed by one cycle of 72°C for 10 minutes. The RT-PCR

products were run on a 1% agarose gel, and stained with ethidium bromide. We also

assessed the prevalence of ALPV-Ap across aphid species. Laboratory colonies of

soybean aphid, *A. glycines*, green peach aphid, *M. persicae* and two *A. pisum* colonies,

one colony harboring the ALPV-Ap virus (positive control) and the other with the virus

removed (negative control), were tested. Total RNA extraction and RT-PCR using

ALPV-Ap 3-1F and 3R primers were conducted as described above.

**Vertical transmission of ALPV-Ap in *A. pisum***

For vertical transmission studies, individual adult *A. pisum* were isolated in petri

dishes containing a moist Kimwipe® and a leaf from the pea plant, *Pisum sativum*. The

aphids were then monitored constantly for production of newborn nymphs. Newborn

nymphs were isolated from the adult *A. pisum* before contact with the leaf. The newborn

nymphs were isolated in petri dishes and maintained to the adult stage. The aphids were

then placed in TRIzol® reagent (Invitrogen) and total RNA was extracted. ALPV-Ap

detection was carried out using RT-PCR as described above.

**Establishing an ALPV-Ap free *A. pisum* colony**

Individual *A. pisum* were isolated in petri dishes containing a moist Kimwipe® on

a pea leaf. The aphids were then monitored constantly for production of nymphs and

newborn nymphs isolated in petri dishes prior to contact with the leaf, as described

above. The resulting adult aphids were transferred to individual *P. sativum* plants and F2

generation aphids tested for the presence of virus by RT-PCR. Virus-negative families (n

= 3) were then pooled to create the ALPV-Ap-free *A. pisum* colony.

**Purification of ALPV-Ap virus particles from *A. pisum***

*A. pisum* (~3 g) were ground in liquid nitrogen in a pre-cooled mortar and pestle. The powdered aphids were transferred to a 30 mL centrifuge tube on ice. Sodium phosphate buffer, 15 mL of 0.01M, pH 7 was added to the tube and the tube briefly vortexed. Chloroform, 7.5 mL (1/2 volume of buffer) was added and the tube shaken vigorously before incubation on ice for 5 minutes. The tube was then centrifuged at 8,000 rpm at 4°C for 25 minutes in a SS34 rotor (Beckman). The resulting supernatant was passed through a syringe filter sterilizer (0.45 μM) into a 30 mL centrifuge tube. Sodium phosphate buffer, 5 mL of 0.01M was added to the filtered supernatant, which was then centrifuged at 8,000 rpm at 4°C for 25 minutes in a SS34 rotor (Beckman). The resulting supernatant was then transferred to a 30 mL ultracentrifuge tube and centrifuged at 40,000 rpm at 4°C for 2.5 hours using a 70 Ti rotor (Beckman). The supernatant was removed and 5 mL of 0.01M sodium phosphate buffer was added to the pellet. The tube was covered with parafilm and place on ice at 4°C overnight with shaking. The re-suspended virus mixture was transferred into 1.5 mL centrifuge tube and centrifuged at10, 000 rpm at 4°C for 5 minutes. The resulting supernatant was transferred into a new tube and the final volume was adjusted to 5.5 mL using 0.01M 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 43,000 rpm at 4°C for 3 hours using a 70.1 Ti rotor (Beckman). The resulting liquid was removed and the pellet re-suspended overnight at 4°C as previously described. The purified virus was stored in -20°C.

**Infection of *A. pisum* with ALPV-Ap via feeding and microinjection**

Virus free *A. pisum* were infected using two different methods, membrane feeding and microinjection. For membrane feeding bioassays, second to third instar *A. pisum* (~10-15 aphids) were fed overnight on complete artificial diet (16) containing BSA at varying concentrations and purified ALPV-Ap virions. For injection bioassays, adult *A. pisum* were injected with ~0.2 µL purified ALPV-Ap virions. The fed and injected *A. pisum* were transferred onto *P. sativum* plants and the aphid population was allowed to grow for 7 days. Aphids were then tested for the presence of ALPV-Ap by RT-PCR with actin as an internal control gene (NCBI accession no. NM_001126200.2). The internal control actin primers used were, Forward 5' ATTGAACCCCAAGGCTAATC 3' and Reverse 5' GATCGAGACGAAGGATAGCA 3'.

**Localization of virus in aphid tissues using fluorescence in situ hybridization**

Guts from adult *A. pisum* were dissected in 1X phosphate buffer saline (PBS). The guts were then placed on a glass cover slip capsided with 0.01% poly – L – Lysine (BD Biocapsid ™) and fixed in 4% paraformaldehyde in PBS for 20 minutes. The guts were then washed 3 times in 1X PBS for 5 minutes each time, and equilibrated for 10 minutes in hybridization buffer which contains, 5X SSC (750 mM NaCl, 75 mM Na-Citrate, pH 7), 50% formamide , 200 mg/mL dextran sulfate , 0.1M DTT , 0.5X Denhardt's solution (1% Ficoll type 400, 1% polyvinylpyrrolidone, 1% bovine serum albumin, BSA). The gut tissue was incubated with 10 ng of 5' end labeled probes in hybridization buffer at 60°C overnight in a humid chamber. Guts were washed once for 5 minutes in 1X wash buffer (150 mM NaCl, 15 mM Na-Citrate, pH7, 10 mM DTT), twice

for 15 minutes in wash buffer at 55°C and twice for 15 minutes in 0.5X wash buffer. Tissues were then rinsed three times for 5 minutes in 1X PBS and mounted in Fluoro-Gel anti-fade solution (Electron Microscopy Sciences) with 1 ng/µL DAPI (4',6-diamidino-2-phenylindole) on a microscope slide. The slides were stored at 4°C for 2 hours. The tissues were observed using a Carl Zeiss Axio Vert.A1 florescence microscope and imaged using the Zeiss Zen 2012 software. The probe sequences used for ALPV-Ap: 5'-/56 -FAM/CCGTGGATTTATCATGCATAG - 3'

5'/ALEXA/CACTCTGGTGACATTGGATTAATACCTACAGCAATAGAATTTTGCA CGCGATTACGCATAGAGTGAGCAA- 3'.

**Anti-ALPV antibody design**

*Cricket Paralysis virus* (CrPV: Dicistroviridae) VP1 protein which is a conserved domain model for picornaviruses (Pfam: cd00205) was analyzed in the Φ-Ψ Map viewer interphase of the VIPER data base (17). The exposed amino acids (on the surface) were identified to use as reference in a secondary structure alignment with the VP1 proteins from ALPV. The secondary structure alignment was performed using the CN3D tool available at National Center for Biotechnology Information (NCBI) website (18). The selected epitope PVQEDNTVIDETRT was analyzed to verify a high antigenicity index (above or close to 1) using the Antigenicity Plot tool from the JaMBW website [http://www.bioinformatics.org/JaMBW/]. The peptide and the polyclonal antibody were synthesized by GenScript (Piscataway, NJ, USA).

**Localization of virus in aphid microtome sections using immunogold labelling**

Adult *A. pisum* were fixed and sectioned as previously described (19). The treatments included no probe controls, primary antibody only control and primary and secondary antibody treatments. Microtome sections on microscope slides were blocked with 200 µL blocking buffer (1X PBS, 3% BSA, 1% normal goat serum and 0.1% fish gelatin). Blocking was carried out in a humidified chamber for one hour at room temperature (RT). The blocking buffer was replaced with 200 µL anti-ALPV primary antibody diluted 1:50 in blocking buffer. The primary antibody treatment was carried out in a humid chamber for two hours at 37°C. The sections were then washed in blocking buffer three times for 10 minutes at RT. The goat anti-rabbit secondary antibody conjugated to 10 nm gold was diluted in1:40 in blocking buffer and 200 µL applied to the microtome sections. The secondary antibody treatment was carried out for two hours at 37°C. The slides were washed four times in blocking buffer for 5 minutes at RT. The gold particles on the secondary antibody were enhanced with silver. Silver enhancer and silver developer solutions were mixed in equal volumes to create the working silver enhancement solution. Silver enhancement was carried out with 150 µL of solution for 20 minutes at RT in the dark. The slides were subsequently washed in sterile water four times for 5 minutes at RT and then allowed to air dry. Sections were visualized using a BX40 Olympus light microscope.

**Screening of insect cell lines for ALPV-Ap replication**

Various insect cell lines were screened for their ability to support the replication of ALPV-Ap RNA. The media used to culture the various cell lines are listed in the Table

1 below. The cell lines were seeded in a 6 well plate in medium supplemented with fetal

bovine serum (FBS). The cells were allowed to grow to 60-70% confluence at 28°C. The

cell line medium was removed and 2 mL medium without FBS was added to the cells.

After 20 minutes, the cells were transfected with various treatments without FBS. The

cell lines were transfected with three different treatments: A control treatment of 1 mL

cell culture medium only, a transfection control of 1 mL cell culture medium with 20 μL

Cellfectin II (Invitrogen) reagent and a virus treatment of 1 mL cell culture medium with

20 μL Cellfectin II reagent (Invitrogen) and 3 μg of ALPV-Ap RNA purified from

ALPV-Ap virions. The transfection procedure was carried out for 4 hours at room

temperature with 2-4 shaking motions per minute. The transfection reagents were

removed and replaced with 2 mL of the appropriate cell line medium with FBS. The cell

lines were incubated at 28°C and monitored daily for cytopathic effects (CPE). Following

collection at various time points post transfection, the cells were centrifuged at 5,000 rpm

for 5 minutes, medium removed and the cell pellets were stored at -80°C.

**Table 1.** Composition of media used for cell lines screened for infection by ALPV-Ap.

| Cell line | Composition of Growth medium |
|---|---|
| Glassy winged sharp shooter – GWSS Z15 (*Homalodisca vitripennis*) | Excell 420® (SAFC Biosciences) + 2.5% FBS (Atlanta Biologics) |
| Leafhopper – Ac – 20 (*Agallia constricta*) | Excell 420® (SAFC Biosciences) + 2.5% FBS (Atlanta Biologics) |
| Fall armyworm – Sf9 (*Spodoptera frugiperda*) | Sf-900™ SFM (Gibco) |
| Cabbage looper – Hi5 (*Trichoplusia ni*) | EXPRESS FIVE® SFM (Gibco) |
| Mosquito – C6/36 (*Aedes albopictus*) | Liebovitz's L-15 (Life technologies) + 1% L-glutamine (SIGMA) + 10% FBS (Atlas Biologics) |
| Fruit fly – S2 (*Drosophila melanogaster*) | Schneider's Drosophila medium (Gibco) + 2.5% FBS (Gibco) |

**Table 1.** Composition of media used for cell lines screened for infection by ALPV-Ap (Continued)

| | |
|---|---|
| Wasp – Tex-2 (*Trichogramma exiguum*) | TNM-FH (Sigma-Aldrich) + 10% FBS (Gibco) |
| Spotted cucumber beetle – DU182A (*Diabrotica undecimpunctata*) | Excell 420® (SAFC Biosciences) + 3% FBS (Gibco) |

**Relative quantification of ALPV-Ap titer in DU182A cell line**

Transfected cells were collected at various time points post transfection. Total RNA was extracted from the cells using TRIzol® reagent. Total RNA (100 ng) was used for qRT-PCR analysis. Two technical replicates were tested for each sample. The fold-change of ALPV-Ap RNA in the cells was determined using the ΔΔCt method (20). The Ct values were normalized to the internal control gene, 28S ribosomal RNA (NCBI accession no. AY171445.1) and the negative control calibrator sample (cells in medium only).

**Results**

**ALPV-Ap was detected in *A. pisum,* but not in *A. glycines* or *M. persicae*.**

The presence of ALPV-Ap was tested by RT-PCR with primers designed to amplify the 3' end of ALPV-Ap genome from nucleotide 8,971to 9,906 (Figure 1). A PCR product was detected in *A. pisum* but not in *A. glycines* and *M. persicae.*

**ALPV-Ap infectivity and transmission**

Individual adult *A. pisum* were tested by RT-PCR to assess the prevalence of ALPV-Ap in the population. In the 24 adult *A. pisum* tested, 17 adults (70.8%) showed

the presence of ALPV-Ap (Figure 2). No vertical transmission of ALPV-Ap from adult

*A. pisum* to newborn nymphs was detected using single family studies (n=70).

Establishing infection of ALPV-Ap in virus-free *A. pisum* was successful when purified

virus was either microinjected directly into the aphid hemolymph or delivered orally in

artificial diet containing BSA. *A. pisum* free of the virus was not infected when fed on

artificial diet containing purified ALPV-Ap in the absence of BSA (Figure 3). Virus was

also detected in the honeydew of aphids fed virus and BSA.

**Localization of ALPV-Ap RNA in *A. pisum* gut tissues**

ALPV-Ap was localized in *A. pisum* using fluorophore labelled probes for viral

RNA, and anti-ALPV primary antibody generated against the capsid protein region for

localization of virions using immunogold experiments. ALPV-Ap RNA was localized to

the gut of *A. pisum* with relatively high abundance in the midgut cell using FISH (Figure

4). No fluorescence was observed in the negative control gut tissue. In the immunogold

silver-enhancement experiments, the no probe controls showed no binding to the

microtome section of the gut tissue and the primary antibody only control showed very

low binding. Strong silver enhancement was observed for the treatment with both

primary and secondary antibody used for ALPV-Ap virion detection (Figure 5).

**Screening of cell lines for ALPV-Ap replication**

Various insect cell lines were screened for their ability to support replication of

ALPV-Ap (Table 2). The cell lines were transfected with 3 µg of RNA from purified

virions. Two cell lines; GWSS Z15, *Homalodisca vitripennis* and Tex-2, *Trichogramma*

*exiguum* showed cytopathic effects (CPE) but no increase in virus positive strand RNA

based on qRT-PCR analysis. The spotted cucumber beetle, *Diabrotica undecimpunctata* cell line, DU182A showed cytopathic effects (Figure 6) and an increase in virus positive strand RNA. Cytopathic effects observed included lower cell density and rounding of the cells three days post transfection. A 60-fold increase of ALPV-Ap positive sense RNA was observed relative to control treatments one day post transfection compared to the negative controls, but levels of positive sense RNA dropped to basal levels by day three. By five days post transfection the cells had recovered and displayed similar cell morphology and density to the negative control treatments (Figure 6).

**Discussion**

Current knowledge of the biology of ALPV is limited to only one isolate of ALPV, ALPV-RhP from *R. padi* (21). Localization of ALPV-RhP was detected in the gut, brain and embryonic tissue (22, 23), with a vertical transmission rate of 30% in *R. padi* and a 17% rate in the wheat aphid, *Sitobion avenae* (24). High temperatures (increase of 5°C) increased the incidence of virus in the aphid colony. Here, we studied the biology of ALPV-Ap, a new isolate of ALPV from *A. pisum* that is related to ALPV-RhP but more closely related to the honeybee, *A. mellifera* isolates (ALPV-AM) (see Chapter 2). We showed that ALPV-Ap does not infect all *A. pisum* adults in the population, is not vertically transmitted in the pea aphid, and is detected in the aphid midgut cells. In addition, our study identified an insect cell line (DU182A) that could support the initial replication of the virus.

No vertical transmission was detected using single family assays for ALPV-Ap. This result is in contrast to data for an ALPV isolate from *R. padi* , ALPV-RhP which

had a vertical transmission rate of ~ 30% in *R. padi* and a 17% vertical transmission rate in *S. avenae* (24). There is a possibility that ALPV-Ap may still be vertically transmitted but at very low levels. Instead, ALPV-Ap may rely primarily or exclusively on horizontal transmission either via the plant phloem as shown for *Rhopalosiphum padi virus* (RhPV) and *Dysaphis plantaginea densovirus* (DplDNV) or via honeydew contamination of the plant surface (7, 25). Several other dicistroviruses, *Kashmir bee virus* (KBV), *Black queen cell virus* (BQCV), *Acute bee paralysis virus* (ABPV) and *Sacbrood virus* (SBV) have been detected in pollen fed upon by *A. mellifera* (26). The source of the virus inoculum in the pollen is unknown but could have been introduced by another infected insect that fed on the plant or inoculation via *A. mellifera* saliva.

Given that ALPV isolates have been detected in diverse species of insects, including honeybees, aphids, corn rootworms, and whiteflies, this virus may rely on horizontal transmission to multiple host species, rather than being highly adapted for infection of a single or a few closely related host insect species. The potential 100% horizontal transmission rate of ALPV-Ap makes the virus an ideal candidate for field application by spraying. Although sRNA and transcriptome sequencing data for *A. glycines* indicated that an ALPV-like virus is present in the laboratory colony (data not shown), the 3' end of the genome between the two different ALPV isolates may be dissimilar as suggested by the lack of amplification of the specific RT-PCR product from *A. glycines* samples.

ALPV-Ap RNA was detected in the gut tissue of *A. pisum*, similar to ALPV-RhP which also localized to the gut tissue of *R. padi* (22, 23). We established that infection of *A. pisum* with ALPV-Ap was enabled by the presence of BSA, with no infection of

aphids fed on purified virions in diet in the absence of BSA. A recent publication (27) and unpublished work (Linz LB, unpublished) demonstrated the increase in plant virus uptake by aphids when co-fed with BSA or with several other non-phloem proteins such as casein and cytochrome C (27, 28). Possible explanations for this observation include: (i) BSA attaches to the virus and facilitates entry of the virion into the gut cells via an additional receptor on the gut surface or (ii) BSA provides additional substrate for gut proteases thereby protecting virions from degradation by gut proteases. In nature, horizontally transmitted ALPV-Ap could receive similar benefits from plant phloem proteins if horizontally transmitted through plant phloem.

On the basis that research on ALPV-Ap would be facilitated by a cell line that supports replication of this virus, we screened cell lines for their ability to support virus replication. In the absence of aphid cell lines, the following cell lines derived from other hemipteran insects were screened: lines were derived from the glassy – winged sharp shooter (GWSS), *H. vitripennis*, and the leafhopper, *A. constricta*, Ac-20. While the Ac-20 cell line showed no CPE following transfection of ALPV-Ap RNA, CPE were observed in the GWSS cell line. No increase in virus titer as indicated by the presence of increased amounts of positive strand RNA was detected however. A similar observation was made with the Tex-2 cell line, a hymenopteran cell line tested on the basis that an ALPV-like virus, ALPV- Am-Brookings, was detected in *A. mellifera* (10). On further analysis of infection of the GWSS Z15 cell line, we determined that icosahedral virus-like particles ranging from 20 to 45nm in diameter were present in non-transfected GWSS cells by TEM indicating that this cell line is covertly infected with other viruses (Supplementary Figure 1).

The DU182A cell line showed an initial 60-fold increase in viral RNA but the cells recovered from the infection by day 3 post transfection. This result indicates that this cell line can support ALPV-Ap replication but only in the short-term. Recovery of the transfected cells may result from an RNAi-based antiviral response. ALPV-Ap capsid proteins were not detected by western blot in the DU182A cell line (data not shown). A similar observation (increased viral RNA in the absence of detectable capsid protein) was made for the replication of *Homalodisca coagulata virus-1* (HoCV-1: Dicistroviridae) in the GWSS Z15 cell line (J.A. Kroemer et al. unpublished). Further experiments are needed to elucidate why the capsid proteins are not produced via the IGR IRES as expected on entry of the RNA into these cell lines (29). The IGR IRES have not been tested for functionality in a hempiteran cell line.

Since the first described ALPV isolate (ALPV-RhP), a second isolate was identified in *A. mellifera* colonies 23 years later (10). Since then, ALPV isolates have been identified in multiple colonies of *A. mellifera*, *A. nerii* and in bat fecal samples (14, 15, 30, 31). Laboratory studies have further established that ALPV can infect multiple other aphid species and whiteflies (30, 32). These publications show the wide host range of ALPV isolates in nature and the ability of this virus to infect multiple insects. It appears that ALPV may be a virus that is less adapted to one specific host like some other dicistroviruses such as RhPV and *Drosophila C virus* that would have a higher rate of vertical transmission (33-35). The specific ecological strategy of ALPV to infect a host insect is similar to that of another dicistrovirus *Cricket paralysis virus* (CrPV) that has a wide host-range and the ability to infect multiple insects (35). CrPV has been showed to

85

transmit horizontally in populations of olive fruit fly, *Dacus oleae* by fecal contamination (36).

The study provides characterization of ALPV-Ap and its interactions with the model aphid species, *A. pisum*. The availability of the genome sequence for *A. pisum* will make future studies for aphid – virus interactions a possibility especially in aspects of antiviral immunity. Analysis of antiviral immunity responses other than RNAi using *A. pisum* and ALPV isolate from *R. padi* showed no significant differences in the immune response (37). There appears to be natural variation in the pathogenicity of the various ALPV isolates with some isolates being highly virulent to the host aphid (21) and others having no apparent negative impact on the host aphid (30). Surprisingly, the ALPV-AN isolate from *A. nerii* was not virulent to the natural host aphid but showed high virulence towards another species of aphid, *M. persicae* in laboratory assays. Additional ALPV– host studies will further facilitate our understanding of ALPV and potentially pave the way toward development of virus – based insecticides that could be used in insect pest management.

**Author Contributions**

DV conducted all experiments and wrote the manuscript. SL was involved in experimental design and data analysis. BCB was involved in experimental design, discussion of data analysis and writing of the manuscript.

**Acknowledgements**

## References

1.   Dedryver CA, Le Ralec A, & Fabre F (2010) The conflicting relationships between aphids and men: a review of aphid damage and control strategies. *C R Biol* 333(6-7):539-553.

2.   Ng C. K. J & Perry L. K (2004) Transmission of plant viruses by aphid vectors. *Mol. Plant. Pathol.* 5(5):505-511.

3.   Williamson C & Rybicki EP (1989) A comparative study on the cell-free translation of the genomic RNAs of two aphid picorna-like viruses. *Arch Virol* 109(1-2):59-70.

4.   D'Arcy CJ, Burnett PA, Hewings AD, & Goodman RM (1981) Purification and characterization of a virus from the aphid *Rhopalosiphum padi*. *Virology* 112(1):346-349.

5.   Ryabov EV (2007) A novel virus isolated from the aphid *Brevicoryne brassicae* with similarity to Hymenoptera picorna-like viruses. *J Gen Virol* 88(Pt 9):2590-2595.

6.   Richards S*, et al.* (2010) Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *Plos Biology* 8(2).

7.   Ryabov EV, Keane G, Naish N, Evered C, & Winstanley D (2009) Densovirus induces winged morphs in asexual clones of the rosy apple aphid, *Dysaphis plantaginea*. *Proc Natl Acad Sci U S A* 106(21):8465-8470.

8.   van Munster M*, et al.* (2003) Characterization of a new densovirus infecting the green peach aphid *Myzus persicae*. *J Invertebr Pathol* 84(1):6-14.

9.   Liu S, Vijayendran D, & Bonning BC (2011) Next generation sequencing technologies for insect virus discovery. *Viruses* 3(10):1849-1869.

10.   Runckel C*, et al.* (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *nosema*, and *crithidia*. *PLoS ONE* 6(6):e20656.

11.   Wu Q*, et al.* (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci U S A* 107(4):1606-1611.

12.   Csorba T, Pantaleo V, & Burgyan J (2009) RNA Silencing: An Antiviral Mechanism. *Advances in Virus Research, Vol 75,* Advances in Virus Research, (Elsevier Academic Press Inc, San Diego), Vol 75, pp 35-71.

13. Bonning BC & Miller WA (2010) Dicistroviruses. *Annu Rev Entomol* 55:129-150.

14. Granberg F*, et al.* (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid lethal paralysis, Israel acute paralysis and Lake sinai viruses. *PLoS One* 8(2):e57459.

15. Ravoet J*, et al.* (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS One* 8(8):e72443.

16. Febvay G, Delobel B, & Rahbe Y (1987) Influence of the amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (Homoptera: Aphididae). *Can.J.Zool* 66:2449-2453.

17. Carrillo-Tripp M*, et al.* (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res* 37(Database issue):D436-442.

18. Wang Y, Bryant S, Tatusov R, & Tatusova T (2000) Links from genome proteins to known 3-D structures. *Genome Res* 10(10):1643-1647.

19. Boyapalle S, Pal N, Miller WA, & Bonning BC (2007) A glassy-winged sharpshooter cell line supports replication of Rhopalosiphum padi virus (Dicistroviridae). *J Invertebr Pathol* 94(2):130-139.

20. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408.

21. Williamson C, Rybicki EP, Kasdorf GF, & Von Wechmar MB (1988) Characterization of a new picorna-like virus isolated from aphids. *J Gen Virol* 69:787-795.

22. Laubscher JM, Jaffer MA, & von Wechmar MB (1992) Detection by immunogold cytochemical labeling of Aphid lethal paralysis virus in aphid *Rhopalosiphum padi* (Hemiptera:Aphididae). *J Invertebr Pathol* 60:40-46.

23. Hatfill SJ, Williamson C, Kirby R, & von wechmar MB (1990) Identification and localization of Aphid lethal paralysis virus in thin tissue sections of the *Rhopalosiphum padi* aphid by in situ nucleic acid hybridization. *J Invertebr Pathol* 55(2):265-271.

24. Laubscher JM & von Wechmar MB (1992) Influence of Aphid lethal paralysis virus and *Rhopalosiphum padi* virus on aphid biology at different temperatures. *J Invertebr Pathol* 60:134-140.

25.    Ban L, Didon A, Jonsson LM, Glinwood R, & Delp G (2007) An improved detection method for the Rhopalosiphum padi virus (RhPV) allows monitoring of its presence in aphids and movement within plants. *J Virol Methods* 142(1-2):136-142.

26.    Chen Y, Evans J, & Feldlaufer M (2006) Horizontal and vertical transmission of viruses in the honey bee, Apis mellifera. *J Invertebr Pathol* 92(3):152-159.

27.    Bencharki B, *et al.* (2010) Phloem protein partners of Cucurbit aphid borne yellows virus: possible involvement of phloem proteins in virus transmission by aphids. *Mol Plant Microbe Interact* 23(6):799-810.

28.    Bonning BC & Chougule NP (2014) Delivery of intrahemocoelic peptides for insect pest management. *Trends Biotechnol* 32(2):91-98.

29.    Brödel AK, *et al.* (2013) IRES-Mediated Translation of Membrane Proteins and Glycoproteins in Eukaryotic Cell-Free Systems. *PLoS One* 8(12):e82234.

30.    Dombrovsky A & Luria N (2013) The Nerium oleander aphid *Aphis nerii* is tolerant to a local isolate of Aphid lethal paralysis virus (ALPV). *Virus Genes* 46(2):354-361.

31.    Ge X, *et al.* (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86(8):4620-4630.

32.    Van Munster M, *et al.* (2002) Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family Dicistroviridae. *J Gen Virol* 83(Pt 12):3131-3138.

33.    Gildow FE & D'Arcy CJ (1990) Cytopathology and experimental host range of *Rhopalosiphum padi* virus, a small isometric RNA virus infecting cereal grain aphids. *J Invertebr Pathol* 55(2):245-257.

34.    Kapun M, Nolte V, Flatt T, & Schlötterer C (2010) Host range and specificity of the Drosophila C virus. *PLoS One* 5(8):e12421.

35.    Plus N, Croizier G, Reinganum C, & Scott PD (1978) Cricket paralysis virus and Drosophila C virus: serological analysis and comparison of capsid polypeptides and host range. *J Invertebr Pathol* 31(3):296-302.

36.    Manousis T & Moore NF (1987) Cricket Paralysis Virus, a Potential Control Agent for the Olive Fruit Fly, Dacus oleae Gmel. *Appl Environ Microbiol* 53(1):142-148.

37.     Gerardo NM, *et al.* (2010) Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome Biol* 11(2):R21.
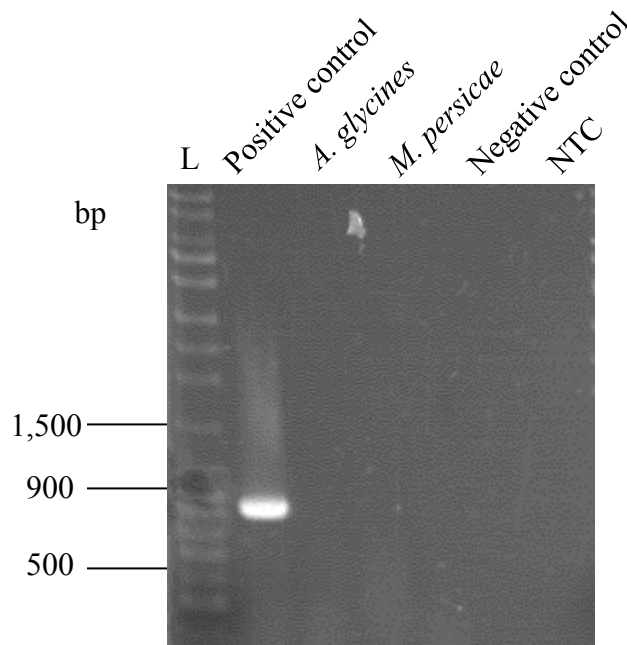
**Figure 1.** ALPV-Ap was not detected in other species of aphids. *A. glycines* and *M. persicae* were tested for the presence of ALPV-Ap by RT-PCR. Positive control, ALPV-Ap infected *A. pisum*; negative control, ALPV-Ap free *A .pisum*. NTC, no template control. Primers designed to amplify 3' end of virus genome from 8,971nt to 9,906nt were used for detection. L, DNA ladder.
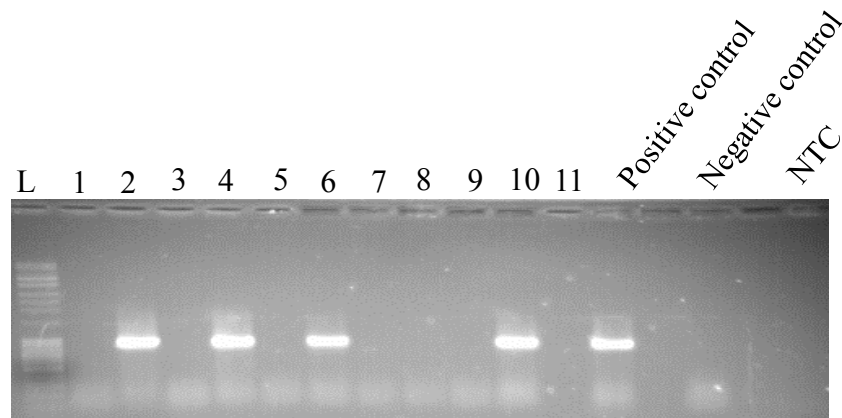
**Figure 2.** Prevalence of ALPV-Ap in *A. pisum* population. A total of 70 adult *A. pisum* were tested for the presence of the virus by RT-PCR and a representative gel is shown. Positive control, random sample of multiple *A. pisum* from the infected colony; negative control, *A. pisum* from the ALPV-Ap – free colony. NTC, no template control. L, DNA ladder.

93
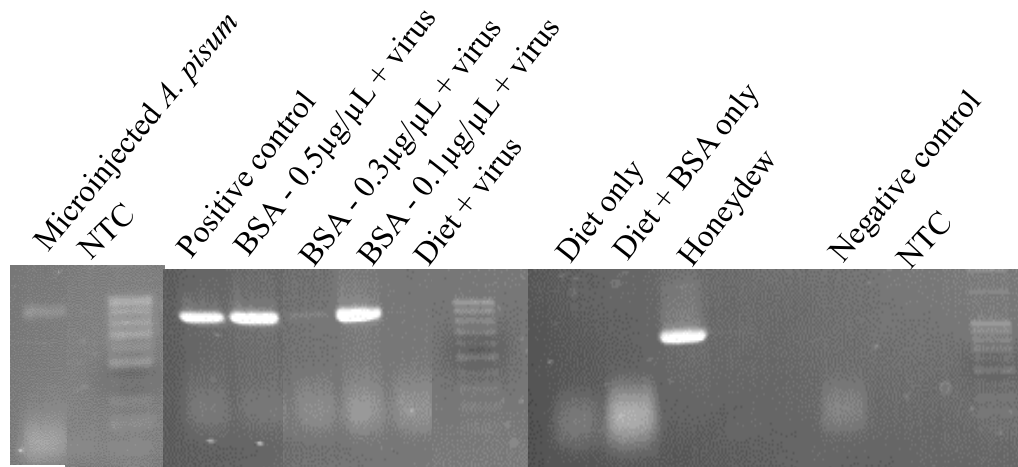


**Figure 3.** BSA is required for oral infection of virus free *A. pisum* with ALPV-Ap. Microinjection and feeding bioassays with varying concentrations of bovine serum albumin (BSA) or diet only were used. Droplets of honeydew in the feeding dish containing the virus and BSA mixed with the diet were pooled for the honeydew sample. Positive control, ALPV-Ap infected *A. pisum*; negative control, ALPV-Ap-free *A. pisum*; NTC, no template control for the RT-PCR experiment. Aphids were microinjected with or fed on virus and tested after 7 days.
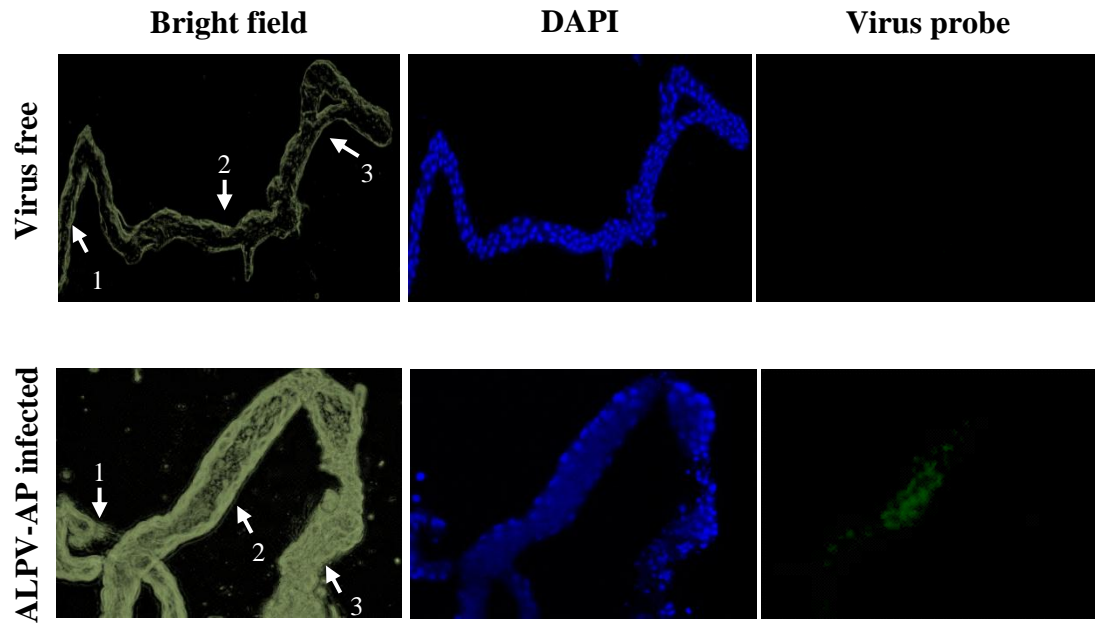
**Figure 4.** Presence of ALPV-Ap in *A. pisum* gut. Virus was detected using fluorescence in situ hybridization. The negative control guts were from ALPV-Ap-free *A.pisum*. DAPI staining of DNA was used for visualization of nuclei. ALEXA 488 conjugated probe antisense to the virus genome was used in the probe treatment. Images are representative of two independent experiments. 1, foregut; 2, midgut; 3, hindgut.

No antibody control    No primary antibody    Primary and
secondary antibody
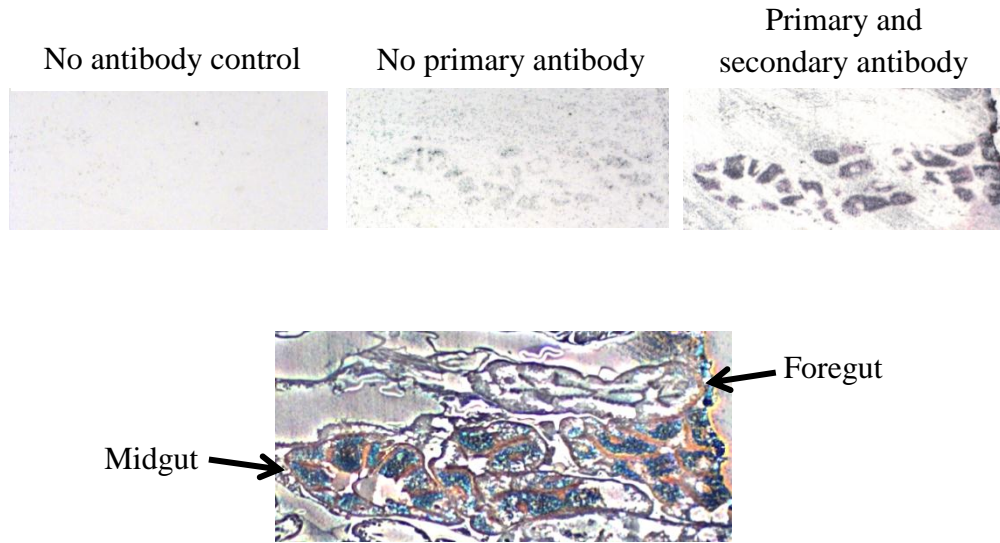


**Figure 5.** Localization of ALPV-Ap to *A. pisum* gut. Immunogold labelling with silver enhancement was used for microtome sections of *A. pisum* gut for localization of ALPV-Ap. No antibody and no primary antibody controls serve as negative controls. Images are representative of four independent experiments. Lower image shows the light image of the aphid gut tissue microtome section shown above.

**Table 2.** Insect cell lines screened for their ability to support replication of ALPV-Ap. Cell lines were monitored for cytopathic effects (CPE) and qRT-PCR conducted to monitor for changes in virus titer based on the abundance of positive sense RNA.

| Insect order | Insect cell line screened | Cytopathic effects | Increase in viral RNA |
|---|---|---|---|
| Hemiptera | Glassy winged sharp shooter – GWSS Z15 (*Homalodisca vitripennis*) | Yes* | No |
| Hemiptera | Leafhopper – Ac-20 (*Agallia constricta*) | No | No |
| Lepidoptera | Fall armyworm – Sf9 (*Spodoptera frugiperda*) | No | No |
| Lepidoptera | Cabbage looper – Hi5 (*Trichoplusia ni*) | No | No |
| Diptera | Mosquito – C6/36 (*Aedes albopictus*) | No | No |
| Diptera | Fruit fly – S2 (*Drosophila melanogaster*) | No | No |
| Hymenoptera | Wasp – Tex-2 (*Trichogramma exiguum*) | Yes | No |
| Coleoptera | Spotted cucumber beetle – DU182A (*Diabrotica undecimpunctata*) | Yes | Yes |

* GWSS Z15 contains viruses that naturally infect the cell line (Supplementary Figure 1).
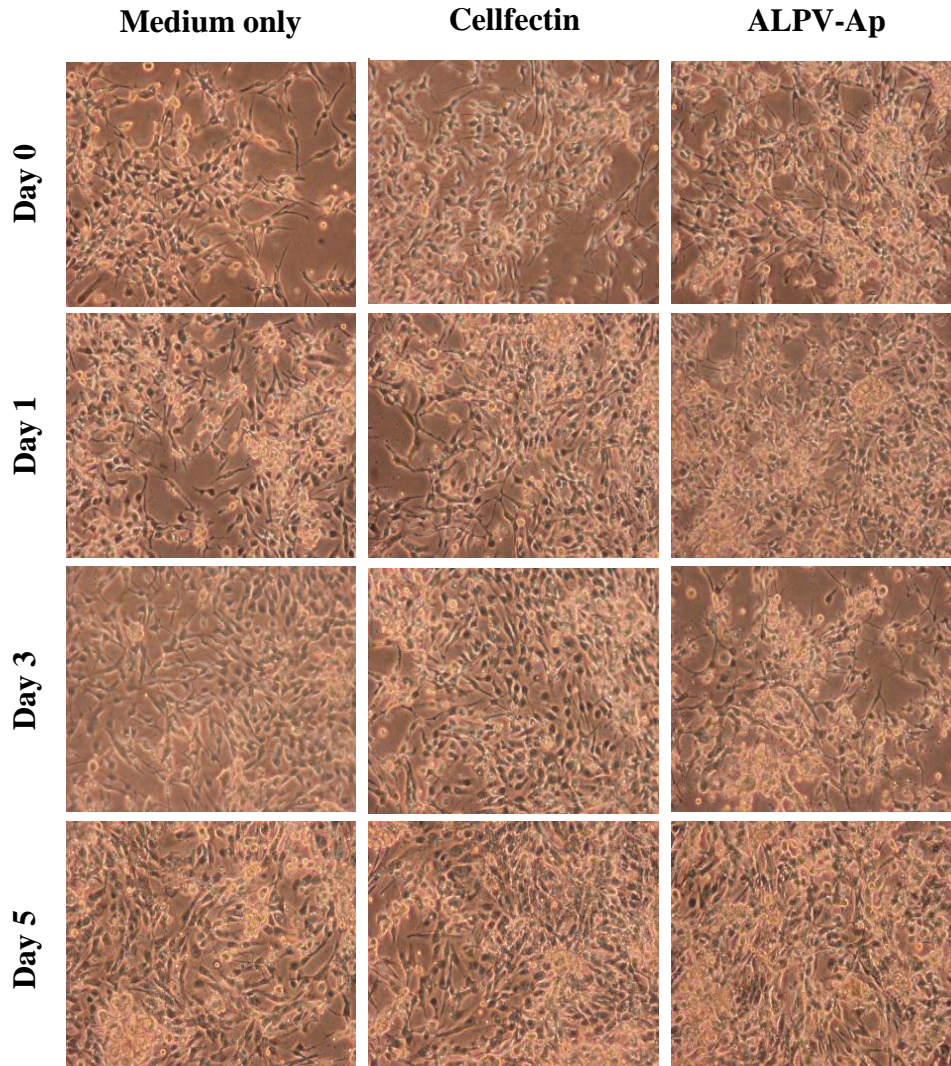
**Figure 6.** Cytopathic effects in DU182A cells following transfection with ALPV-Ap RNA. Images show cells at different days post treatment. Cells were transfected with 3 µg of virus RNA. Cytopathic effects were evident at one and three days post transfection, but cells had recovered by day 5. Medium only and Cellfectin show cells in negative control treatments.

**Supplementary information**

**Supplementary Figure 1. The GWSS Z15 cell line is naturally infected with viruses**

Glassy – winged sharp shooter (GWSS) Z15 cells were grown in large quantities (~30 mL of 100% confluent cells) and were kept frozen at -80°C. The tube containing the cells was thawed and vortexed to break open the cells and release the virus particle. The tube was centrifuged at 3,200 x g for 15 minutes at 4°C. The supernatant was collected after centrifugation and used to precipitate virus particles using 7.5 mL Polyethylene glycol (Biovision). Virus particles were precipitated overnight at 4°C. The tube was centrifuged at 3,200 x g for 30 minutes at 4°C to pellet the precipitated virus particles. The pellet was re-suspended in 0.5 mL1X PBS. Virus particles were negatively stained and viewed using a transmission electron microscope as previously described (Liu *et al.* 2014).
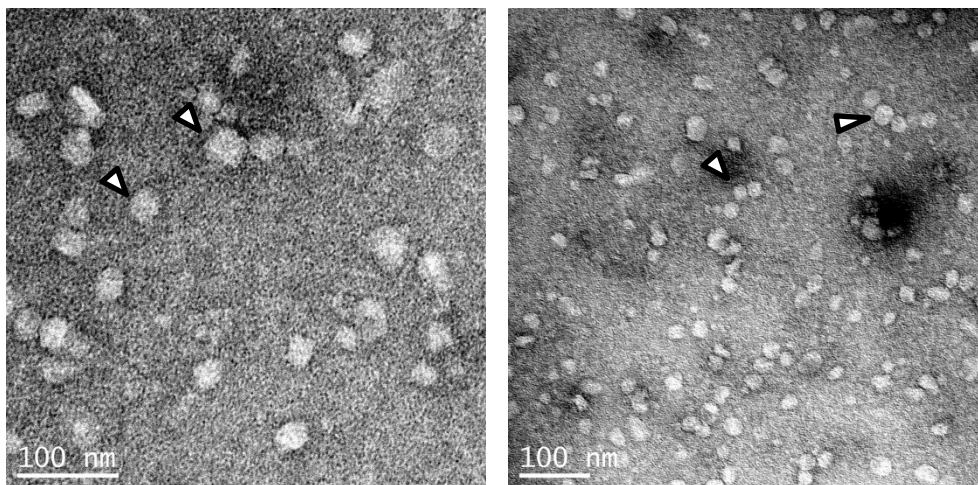


Transmission electron micrograph of purified virions from GWSS Z15 cells, negative stained with uranyl acetate. In addition to some cellular debris, icosahedral structures that are consistent with virus particles were observed (arrow heads). The diameter of the virus-like particles ranged from 20 nm – 45 nm.

CHAPTER 4

# A NOVEL VIRUS OF SOYBEAN APHID WITH PLANT VIRUS CHARACTERISTICS

Diveena Vijayendran, Sijun Liu and Bryony C. Bonning

Department of Entomology, Iowa State University, Ames, IA, 50011

## Abstract

The soybean aphid, *Aphis glycines* is a major pest on soybean worldwide resulting in losses in the billions of dollars per year. We identified a novel *A. glycines* virus named *Aphis glycines virus* (AGV) from transcriptome sequences. This virus has a single-stranded RNA genome of ~ 4,771 nt and potentially encodes four proteins. The RNA-dependent RNA polymerase (RdRP) showed similarity to insect tetraviruses while the capsid protein show similarity to plant sobemoviruses. AGV is wide-spread in *A. glycines* populations and persistently infects the aphid with a 100% vertical transmission rate, suggesting that AGV is highly adapted to this species. AGV genome sequences was not detected in the *A. glycines* genome. AGV was suseptible to the RNAi pathway, but only in the presence of a second *A. glycines* virus, indicating that AGV may have evolved to avoid the host immune response and persistently infect soybean aphids.

## Introduction

Aphids are major pest of agricultural crops worldwide because of their ability to feed on plant phloem and transmit a significant number of economically important plant viruses (1, 2). The soybean aphid, *Aphis glycines* Matsumura was introduced into North

America in 2000 and has since become one of the primary pests in soybean fields (3).

The estimated losses resulting from low yields combined with the cost of management of

this pest in North America alone were $1.6 billion over a 10 year period (4). Current

management strategies to control *A. glycines* populations include the use of chemical

insecticides, the use of predatory arthropods and development of aphid resistant soybean

cultivars (1, 5, 6). Although soybean cultivars such as *Rag 1* and *Rag 2* showed promise

in aphid management, resistant biotypes of *A. glycines* have been identified (7). One

novel avenue that has yet to be explored for aphid management is the identification and

utilization of insect viruses.

Insect viruses were traditionally identified when symptoms were observed in the

infected organism. While symptoms usually result from an acute infection of the host

insect, viruses with low pathogenicity or that are asymptomatic were not commonly

identified. Next generation sequencing (NGS) has been successfully used for

identification of viruses from multiple insect species that lacked disease symptoms. The

first report of insect viruses identified from sequencing data was from a metagenomic

analysis of the honey bee, *Apis mellifera* conducted to elucidate the causative agents of

colony collapse disorder (CCD) (8). Since this report, there have been more than 20

reports of novel viruses identified from insects and from insect cell lines (9-12). It is also

being realized that some insect viruses such as the *Aphid lethal paralysis virus* (ALPV)

have a much broader host range than previously thought with detection of this virus in

aphids, honeybees and corn rootworm (11, 13-15). Sequencing data therefore provide a

useful resource for the identification of novel viruses that can enhance our understanding

of insect virus evolution and insect-virus interactions.

We used our previously published *A. glycines* transcriptome sequencing data for identification and assembly of virus sequences. We report on the identification of a novel *A. glycines* virus, *Aphis glycines virus* (AGV). Remarkably, the 5' end of the genome that encodes nonstructural proteins resembles an insect virus, while the 3' end of the genome resembles a plant virus. Although all *A. glycines* tested harbor this virus, symptoms were not apparent. Small RNA sequencing data showed that AGV is susceptible to the host RNA interference (RNAi) - based antiviral immune response. We detected 22 nt virus-derived small interfering RNAs (vsRNAs) which are characteristic of dicer processing of viral replication intermediates (16). This study describes the first *A. glycines* virus in addition to providing evidence for a functional RNAi-based antiviral response in aphids.

**Materials and Methods**

**Aphid rearing**

Soybean aphid, *Aphis glycines* Matsumura was reared in the laboratory from Iowa and Ohio field collected samples. The aphids were reared on soybean plants, *Glycine max* at room temperature (20°C to 25°C) with a continuous light cycle.

**AGV genome discovery**

The methods for assembly of the *A. glycines* transcriptome sequencing data generated using Illumina sequencing have been described (17). Briefly, the sequencing reads were trimmed to remove adaptor sequences, and the resulting reads were assembled by using the Velvet assembler (18). The resulting contigs were used for BLAST against the NCBI nr database by using the Blast2GO platform (19). Contigs with viral genes as top hits were extracted. The contigs with viral hits were then used as subject sequences

and the total contigs as query sequences for local BLAST to search for additional

homologous sequences derived from viruses. The final AGV genomic sequences were

assembled by reassembling the contigs with viral hits. The transcriptome reads were

subsequently reassembled by using the Trinity assembler using the default settings (20).

**Sequence confirmation and 5' and 3' end determination**

The assembled genome sequence of *Aphis glycines virus* (AGV) was confirmed

using RT-PCR with primers based on the assembled AGV genome sequence (Table 1).

Purified AGV virions were used for total RNA isolation using TRIzol® reagent

(Invitrogen). Total RNA was extracted according to the manufacturer's recommendations

with overnight precipitation of RNA in isopropanol at -80°C. RT-PCR was carried out

using a One-Step RT-PCR reaction kit (Qiagen).  The resulting PCR products were

isolated and purified from agarose gels using the Qiaquick gel extraction kit (Qiagen) and

sequenced by Sanger sequencing.

The ends of the virus genome were determined using rapid amplification of

cDNA ends (RACE) with the SMARTer RACE cDNA amplification kit (Clontech). Viral

RNA (2 µg) was used as template to generate RACE ready cDNA. The Advantage 2 PCR

kit (Clontech) was used for the PCR reaction according to the manufacturer's protocol.

The primers used for the RACE-PCR are listed in Table 1. A touchdown PCR reaction

was carried out in a thermocycler (BioRad MyCycle Thermal Cycler) for 5 cycles of

94°C for 30 seconds and 72°C for 2 minutes, 5 cycles of 94°C for 30 seconds, 68°C for

30 seconds and 72°C for 2 minutes followed by 94°C for 30 seconds, 66°C for 30

seconds and 72°C for 2 minutes. The RACE-PCR products were separated on 1%

agarose gels and 5' RACE products > 515 bp and 3' RACE products >340 bp were ligated into pGEM-T easy (Promega). Plasmids were transformed into TOP 10 competent cells, isolated and sequenced to determine the end sequences.

**Sequence alignment and phylogenetic analysis**

Local BLAST and pairwise sequence alignments of RNA and protein sequences were performed using the BioEdit Sequence Alignment Editor (version 7.2.5) (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Phylogenetic analyses were performed using the Phylogeny.fr website (www.phylogeny.fr) for the RNA-dependent RNA polymerase (RdRP) and the capsid protein (CP) amino acid sequences of AGV (21). The top hits from protein BLAST analyses were extracted and used in the phylogenetic analyses (22). Multiple sequence alignment was carried out using MUSCLE software (23). The phylogeny was constructed using PhyML, providing an estimate of maximum likelihood method for amino acids with posterior probability values (24). The phylogenetic tree was visualized using the Phylodendron Phylogenetic tree printer (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html).

**Homology modeling of RdRP and CP structures**

To predict the putative protein structure of RdRP and CP, homology modeling of proteins was performed by LOMET (Local Meta-Threading-Server) (25).

**Isolation of dicistrovirus-free, AGV infected *A. glycines***

As *A. glycines* were also found to harbor dicistroviruses, we isolated single families to establish an AGV-positive, dicistrovirus-negative *A. glycines* colony.

Individual *A. glycines* were isolated in petri dishes containing a moist Kimwipe® and a soybean leaf. The aphids were monitored constantly for production of newborn nymphs. The newborn nymph was isolated from the adult *A. glycines* before contact with the leaf. The newborn nymphs were isolated in petri dishes and maintained to adults. The aphids were transferred to individual soybean plants and F2 generation aphids tested for the presence of the dicistroviruses, *Rhopalosiphum padi virus* (RhPV) and ALPV-like virus using RT-PCR with primers designed based on the assembled virus contigs (Table 1). *A. glycines* colonies that tested negative for the viruses were allowed to grow for 2 weeks and re-tested for the presence of viruses using the same RT-PCR method. The colonies with no dicistrovirus sequence amplification were then pooled as the AGV-only infected *A. glycines* colony.

**AGV virion purification**

AGV was purified from the laboratory colony of *A. glycines* infected only with AGV. *A. glycines* (~40 mL) were homogenized in liquid nitrogen and transferred to 0.01M sodium phosphate buffer, pH 7.0. The homogenate was centrifuged at 7,000 rpm (5,856 x g) for 10 minutes to pellet aphid debris. The supernatant was filtered through four layers of cheesecloth to remove contaminating lipids. The supernatant was added to the top of 30 mL of a 10% sucrose cushion in ultra-centrifuge tubes. The tubes were centrifuged at 27,000 rpm at 4°C for 3.5 hours in a SW28 swinging bucket rotor (Beckman). Sodium phosphate buffer (400 μL of 0.01M) was added to each pellet and the tubes were covered with Parafilm and placed on ice at 4°C with gentle shaking overnight.

The pellets were re-suspended in 400 µL of 0.01M sodium phosphate buffer with additional buffer added to completely dissolve the pellet. The tubes were centrifuged at 14,000 rpm (23,425.7 x g) for 10 minutes at 4°C to pellet any undissolved proteins. The supernatant was passed through a sucrose cushion as previously described. The resulting pellet was re-suspended in 500 µL of 0.01M sodium phosphate buffer on ice at 4°C. The virus was stored at -20°C in 100 µL aliquots.

**Sample preparation for electron microscopy**

Purified AGV virions (10 µL) were pipetted onto a carbon film grid. The grid was negatively stained with 2% aqueous uranyl- acetate for 30 seconds. The virus particles were visualized using a JEOL 2100 scanning/transmission electron microscope.

**Vertical transmission of AGV in *A. glycines***

For vertical transmission studies, individual adult *A. glycines* were isolated in petri dishes containing a moist Kimwipe® and a soybean leaf. The aphids were monitored constantly for production of nymphs. Each newborn nymph was isolated from the adult *A. glycines* before contact with the leaf. The newborn nymphs were isolated in petri dishes. The adult aphids were then placed in TRIzol® reagent (Invitrogen) and total RNA extracted using the manufacturer's protocol. A proportion of newborn nymphs were directly placed in TRIzol® reagent.

AGV detected using RT-PCR with primers specific for the RNA-dependent RNA polymerase region with Tet-F and Tet-R primers (Table 1). The RT-PCR reactions were carried out using the One-Step RT-PCR kit (Qiagen). The reactions were incubated in the

BioRad MyCycle Thermal Cycler with the following cycle, 1 cycle of 50°C for 30 minutes, 95°C for 15 minutes, 35 cycles of 94°C for 30 seconds, 53°C for 30 seconds and 72°C for 2 minutes with a final extension step of 72°C for 10 minutes. A no template negative control and a random sample of AGV infected aphids (positive control) were included for RT-PCR.

**Horizontal transmission of AGV in soybean plants**

Leaves from three individual soybean plants infested with AGV- infected *A. glycines* were collected in a petri dish. The aphids were gently removed and the leaves were washed in 10% bleach solution and deionized water for 1 minute to surface sterilize the leaf material and to remove honeydew and shed cuticles. Approximately 2-3 leaves were grouped together for each sample tested. The leaves were homogenized in liquid nitrogen. Total RNA extraction and testing for the presence of AGV by RT-PCR were conducted as described above.

**Prevalence of AGV in field collected *A. glycines* and in other aphid species**

*A. glycines* were collected from soybean fields in Iowa, Ohio and Michigan and laboratory colonies established. Laboratory colonies of three aphid species, the green peach aphid, *Myzus persicae*, the bird cherry-oat aphid, *Rhopalosiphum padi* and the pea aphid, *Acyrthosiphon pisum* were also tested for the presence of AGV. Aphids (~0.1 mL) of all stages were ground in TRIzol® reagent. RNA extraction and AGV testing using Tet-F and Tet-R primers were conducted as described above.

**Testing for the presence of AGV sequences in the *A. glycines* genome**

To assess whether AGV sequences were present in *A. glycines* genome, *A. glycines* (~0.1 mL) of all stages were ground in 800 µL of DNAzol® (Invitrogen). Total *A. glycines* DNA were used for PCR with primer combinations to span the entire length of the AGV genome (Table 1). The PCR was carried out using Choice Taq DNA Polymerase with 1 cycle of 94°C for 3 minutes, 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 3 minutes and 1 cycle of 72°C for 10 minutes. The PCR was repeated with a reduced annealing temperature of 50°C to amplify products with low melting temperature primers. The amplified PCR products were separated on 1% agarose gels, excised from the gels and purified using the Qiaquick gel extraction kit (Qiagen). The PCR products were sequenced to determine the identity of amplified products.

**Mapping small RNA reads to the AGV genome**

Total RNA extraction and sRNA sequencing from *A. glycines* were as previously described (Chapter 2). The sRNA sequencing reads were trimmed to remove adaptors and lower quality sequences using FastQC programs (Babraham Bioinformatics, Babraham Institute, Cambridgeshire, UK). A script written in Perl was used to map the sRNA reads to AGV sequences. The script was designed to only map reads with perfectly matched bases to the AGV genome sequence. Positions and orientation in the AGV genome were recorded for each mapped read.

**Analysis of expression levels of AGV, Dicer-2 and Argonaute-2 in *A. glycines***

Total RNA was extracted from *A. glycines* infected only with AGV and from *A. glycines* co-infected with AGV and an ALPV-like virus using TRIzol® reagent as

previously described. qRT-PCR was carried out using the iTaq™ Universal SYBR®

Green One-Step Kit (Biorad) in 10 µL reactions with 50 ng of total RNA. Primers used

for amplification of the internal control gene actin and each target template are listed in

Table 1. The qRT-PCR reaction was incubated in the Bio-Rad My iQ5 thermocycler

using the following cycle: 50°C for 30 minutes and 95°C for 5 minutes followed by 35

cycles of 95°C for 10 sec and 53.5°C for 30 seconds. Melt curve analysis was carried out

to ensure that there was no non-specific amplification. No template controls were

included for each primer pair. The experiments were carried out with three biological

replicates and two technical replicates. The data was normalized to the internal control

gene actin and fold change calculated relative to transcript levels in the AGV only

infected *A. glycines*.

**Results**

**Assembly of the AGV genome**

Analysis of our *A. glycines* transcriptome dataset indicated the presence of two

known aphid viruses, ALPV and  RhPV (17).  In addition to these two aphid

dicistroviruses, four contigs of 1,004, 1,279, 3,147 and 3,280 nt had nucleotide BLAST

hits to the putative RNA-dependent RNA polymerase (RdRP) of either *Drosophila A

virus* (DAV, Unclassified), *Euprosterna elaeasa virus* (EEV; Tetraviridae) or *Thosea

asigna virus* (TAV; Tetraviridae). These contigs were then used as subject sequences for

BLAST to search for additional putative genomic sequence derived from AGV.  Three

additional viral contigs of 450-2,090 nt were identified. The contigs of putative viral

origin were assembled into a 4,811nt contig, which was predicted to be the near full-

length viral genome. We then re-assembled the transcriptome reads using the Trinity

assembler and obtained a viral contig of 4,795 nt. Confirmation of the genome sequence

using RT-PCR and RACE PCR resulted in a genome size of 4,771 nt.

The newly identified virus was named *Aphis glycines virus* (AGV) after the host

aphid. To facilitate further studies of this virus, an *A. glycines* colony infected with only

AGV was established using single family isolation. This *A. glyciens* colony was tested

using RT-PCR specific to ALPV and RhPV to ensure complete removal of these

dicistroviruses from the colony before further analysis of AGV.

AGV has a single-stranded, positive sense RNA (ssRNA) genome. The 4,771nt

genome is predicted to encode three major ORFs (P145, P28 and P35) and one additional

small ORF (P10) (Figure 1). The capsid protein ORF (P28) overlaps with the RdRP ORF

(P145). The ORF encoding P35 is a putative read through domain (RTD) which is

predicted to be translated by read through of the capsid protein (CP) stop codon (UAG) to

generate a 63 kDa CP-RTD protein (Figure 1 and Table 2). AGV may encode a

subgenomic RNA for the expression of CP-RTD protein.

**5' and 3' UTR of the AGV genome**

The complete AGV genome sequence was confirmed using RT-PCR with primers

designed to span the entire assembled genome of AGV except for the extreme 5' and 3'

untranslated regions (UTR). The terminal ends of the genome were determined using

RACE PCR. The genome sequence of assembled AGV matched the RACE results well at

the 5' UTR except for an additional 24 nt in the extreme 5' end from the genome

assembled using Trinity software. The assembled 3' UTR extreme end was confirmed

using RT-PCR but was not obtained using RACE PCR. The RACE PCR results did not

extend beyond the already assembled sequence.

**The RdRPs of AGV and insect tetraviruses are closely related**

The 145 kDa RdRP of AGV consists of 1,121 amino acids (aa). BLAST analysis

of the sequences suggested similarities between AGV RdRP and those of the insect

tetraviruses, *Euprosterna elaeasa virus* (EEV) and *Thosea asigna virus* (TAV) in addition

to *Drosophila A virus* (DAV). The tetraviruses, EEV and TAV were identified from

Lepidoptera while DAV is currently unclassified (26-28). Amino acid sequence

comparison of AGV RdRP with the RdRPs of the tetraviruses and DAV showed 33-34%

sequence identities. To further examine the similarities between the RdRP of AGV and

tetraviruses, the RdRP tertiary structures of AGV, DAV, and EEV RdRP were predicted

by homology modeling using the LOMETS server. The structure prediction results

clearly demonstrated that the RdRP of AGV has a similar structure to that of the

tetraviruses and DAV (Figure 2), although they only share 33-34% amino acid sequence

identities.

Some tetraviruses encode a permuted RdRP core motif (29). A permuted RdRP is

a non-conical arrangement of the core motifs present in the palm subdomain of the

protein encoded by RNA viruses. The typical arrangement of A-B-C subdomains is

inverted to a C-A-B arrangement (30). AGV also encodes an RdRP with a permuted core

motif. Phylogenetic analysis was carried out for the RdRP of AGV with the RdRP amino

acid sequences from tetraviruses (EEV and TAV) and members of the Birnaviridae family

(Figure 3). The results show that AGV is closely related to DAV and more closely related

to the insect tetraviruses with the permuted RdRP compared to the birnaviruses. This result suggests that the permuted RdRP of the insect viruses have the same evolutionary origin. EEV, TAV and DAV also contain viral protein genome-linked (VPg) sequences in the N-termini of their RdRP sequences. Sequence alignments of the RdRP N-termini of AGV, EEV, TAV and DAV showed putative VPg sequences present in the AGV genome (Figure 4), suggesting that the AGV genome is capped with a VPg.

**The CP of AGV is structurally similar to the CP of plant viruses**

The 28 kDa AGV capsid protein contains 217 aa. Interestingly, BLAST analysis showed similarity to the capsid proteins of plant viruses. The top hit was from a newly identified plant virus isolated from bat fecal material, *Bat sobemovirus* (BSV; NCBI Accession No: AGN73380.1 (31) ) with 31% amino acid identity and 79% sequence coverage. For comparison, protein BLAST analysis of AGV CP to DAV CP showed a 25% amino acid identity and 46% sequence coverage. Homology modeling to determine the tertiary structure of AGV CP suggested that it could form a similar structure to the CP of *Tobacco necrosis virus* (TNV), a tombusvirus (Figure 5). These results indicate that AGV is structurally closer to plant viruses with T=3 virion structures than to insect viruses. Phylogenetic analysis of the AGV CP amino acid sequence shows a close relationship to the CP of BSV and other sobemoviruses (Figure 6). In addition to the major CP, AGV also encodes a putative 35 kDa read through domain (RTD) of 267 aa, which may be translated into a 63 kDa and 485 aa CP-RTD protein. The N-terminus of RTD contains a poly-proline track, a common motif found in the RTD of luteoviruses, a group of plant viruses encoding CP-RTD proteins (32).  BLAST analysis of the RTD

sequence did not result in significant hits of any viral genes, but had low similarity (29% identity) to the methyltransferase genes of bacteria and insects.

**Electron micrograph of purified AGV virions**

AGV virions were icosahedral with an estimated diameter of 30 nm (Figure 7). The virion diameter is consistent with most small RNA viruses. The AGV encoded capsid protein is similar to the CP of sobemoviruses that have a diameter in the range of 25 – 33 nm.

**AGV is widely distributed in field samples of *A. glycines* and may infect multiple species of aphid**

To examine the distribution of AGV, total RNA isolated from field collected *A. glycines* was tested using RT-PCR for the presence of AGV. AGV was detected in *A. glycines* samples collected from Iowa, Ohio and Michigan (Figure 8a). An RT-PCR product of 929 bp corresponding to the RdRP region of AGV was detected for all samples. The same testing method was used to check for AGV in laboratory colonies of other species of aphids. In addition to the laboratory colony of the *A. glycines*, RT-PCR products of the correct size were detected from the *M. persicae* and *R. padi* but not *A. pisum* (Figure 8b).

**AGV has a 100% vertical transmission rate**

To assess the rate of vertical transmission of AGV, newborn *A. glycines* nymphs were isolated before contact with the leaf. Some newborn aphids were directly used for RT-PCR while other nymphs were maintained to the adult stage before testing, to ensure

that detection was possible in the event that the virus titer was too low in newborn nymphs. RT-PCR detection of RdRP viral sequences showed that AGV was present in all tested nymphs and single families. The virus was 100% vertically transmitted from adult females to the newborn nymphs (n=63). This highly efficient vertical transmission rate thwarted efforts to isolate a virus-free *A. glycines* colony. The possibility for horizontal transmission through the soybean plant was also investigated by testing 2-3 leaves from three soybean plants infested with AGV infected *A. glycines*. No AGV sequences were identified from any of the leaves tested.

**AGV is not detected in the genome of the *A. glycines***

Insertion of sequences of persistent insect RNA viruses into the host genome has been reported for *Flock house virus* (FHV, Nodaviridae) and *Israeli acute paralysis virus* of bees (IAPV, Dicistroviridae) (33, 34). The detection of AGV in every *A. glycines* colony tested and the 100% vertical transmission rate prompted testing for the possibility of AGV presence in the genome of *A. glycines*. A combination of primer pairs spanning the entire genome of the virus was used in PCR reactions with *A. glycines* total DNA. The PCR reactions amplified only non-specific unknown products or *A. glycines* genes. None of the sequenced PCR products matched any part of the AGV genome sequence (Supplementary Figure 1).

**AGV is targeted by the RNAi pathway of *A. glycines***

To determine whether silencing of AGV is regulated by the RNAi pathway, we sequenced sRNAs of *A. glycines* infected by AGV. Illumina sequencing sRNA reads were trimmed and mapped to the AGV genome. Around 0.4% of reads (20.6 million

reads in total) matched AGV genomic RNA. A typical vsRNA distribution pattern with a peak at 22 nt derived from both strands of the replicating virus was observed (Figure 9), suggesting AGV was degraded by the RNAi pathway. Over 20,000 vsRNA reads were derived from AGV with the most abundant peak at 22 nt (13,534 reads). Fewer vsRNA reads were detected at other sizes (19 nt-21 nt, 23 nt). The vsRNA reads were derived from all regions of the virus genome with some regions of the genome targeted more frequently for degradation. The region targeted most frequently for dicing corresponds to the RdRP region (Supplementary Figure 2 A-D).

**RNAi pathway is affected by an ALPV-like virus in *A. glycines***

In order to confirm that we established a dicistrovirus free *A. glycines* colony, we sequenced sRNA from the AGV- only infected *A. glycines* colony. The sRNAs were assembled by Velvet assembler, followed by aligning the obtained contigs to NCBI nr database with BLASTx algorithms, or mapped to *Aphid lethal paralysis virus* (ALPV) genomes, a virus previously identified in our *A. glycines* colony. No ALPV and other known virus sequences were identified by BLAST analysis or by ALPV genomic mapping. When the sRNA reads were mapped to the AGV genome, we found that vsRNAs (16-30 nt) derived from the AGV genome were reduced 10-fold when compared to AGV-derived vsRNAs in the presence of the ALPV-like virus (Figure 10). The total reads of sRNA isolated from *A. glycines* infected with both viruses were 20.6 million and reads from the AGV only *A. glycines* were 18.4 million. Hence, the reduction of AGV specific vsRNAs was not the result of the number of reads generated by Illumina sequencing. *A. glycines* collected from soybean fields in Ohio were tested for the presence of ALPV-like virus and AGV using RT-PCR. The Ohio colony harbored AGV

but not the ALPV-like virus. Mapping of sRNA reads from the Ohio field colony to the

AGV genome further confirmed that few reads were derived from AGV (Figure 10).

Taken together, these results suggest that RNAi-mediated degradation of AGV was

promoted by infection with the ALPV-like virus in *A. glycines*.

To test whether the presence of the ALPV-like virus could stimulate either

increased replication of AGV or an up-regulation of the aphid RNAi machinery resulting

in degradation of AGV, we investigated potential differences in transcript levels for

AGV, Dicer-2 and Argonaute-2 using quantitative RT-PCR. The presence of the ALPV-

like virus did not result in any significant differences in transcript levels for AGV, Dicer-

2 or Argonaute-2 (Student's t-test, $p > 0.05$) (Figure 11).

**Discussion**

In this study, we described the identification and characterization of a novel *A.

glycines* virus, *Aphis glycines virus*. The virus has a unique genome structure that appears

to result from a fusion between the polymerase sequence of an insect virus and the capsid

protein sequence of a plant virus. AGV is the first insect virus identified with a capsid

protein similar to the CP-RTD of plant viruses. The CP-RTD of AGV suggests that the

capsid protein of AGV originated from a plant virus. The RdRP of AGV has the motif

arrangements of a permuted RdRP that was first described in tetraviruses and

birnaviruses (30), none of which have been previously described in aphids. It has been

proposed that viruses with a permuted RdRP share a common evolutionary event that

resulted in switching of the sequence motifs. While the RdRP of tetraviruses and

birnaviruses may be related, discrepancies exist when phylogenetic analysis is carried out

using capsid protein sequence even among birnaviruses with the same permuted RdRP
(35). Phylogenetic analysis and virus classification may have to be re-evaluated to
accommodate current outlier viruses such as AGV and DAV that have the same permuted
RdRP and a capsid protein similar to those of plant viruses (26). The role of the RTD
encoded by the AGV genome is unclear in the life cycle of this *A. glycines* virus. In aphid
transmitted luteoviruses, the RTD is implicated in facilitating virion transcytosis across
the aphid accessory salivary gland, basal lamina and cell plasma membrane and may
affect transmission efficiency (36).

AGV is widespread in field populations of *A. glycines* within the United States.
The presence of the faint RT-PCR products in multiple aphid species suggests that AGV-
like viruses infect other aphid species. AGV persistently infects *A. glycines* with a 100%
vertical transmission rate, remarkable among insect viruses. Polydnaviruses are also
100% vertically transmitted in insects but these viruses function more as symbionts with
genes integrated into the parasitic wasp genome. Polydnavirus gene products function to
suppress the host immune response to allow for wasp larval development in the
parasitized host (38).

Based on our experiments, we conclude that genome sequences of AGV have not
inserted into the genome of *A. glycines*. AGV sequences were also not identified using
bioinformatics analysis with an unpublished *A. glycines* genome draft. AGV therefore
may represent a new group of RNA viruses with a unique transmission mode to persist in
aphid populations. Whether AGV has established a true symbiotic relationship with *A.*
*glycines* remains to be determined.

All major components of the RNAi pathway are present in the *A. pisum* genome (40). The role of RNAi in antiviral defense in aphids has not been demonstrated until now. Our data show that RNAi functions in silencing the replication intermediate and local secondary structures of AGV. It further demonstrates that AGV is an actively replicating virus in *A. glycines* with a genome that is susceptible to the host antiviral immune response. The susceptibility of AGV to the immune response is however more pronounced on co-infection of *A. glycines* with a dicistrovirus (ALPV). We hypothesized that AGV remains elusive to the host RNAi response in the absence of infection by a second virus. However, no significant differences were observed in Ago-2 and Dicer-2 transcript levels in the presence of AGV versus AGV and the ALPV-like virus.

In our study we were unable to detect viral RNA using gel electrophoresis. A smear was always observed for RNA extracted from purified AGV virions. A similar observation was reported for DAV. The RNA of DAV was shown to be degraded rapidly during extraction and electrophoresis (26). It has been demonstrated that RNA extraction from tetraviruses results in high levels of degradation presumably because the RNA is not stable outside the virion (37). We were also unable to obtain the protein profile for this virus. A protein with a molecular mass of ~58 kDa was consistently observed. N-terminal sequencing revealed the identity of the protein to be similar to glycoside hydrolase, a common protein found in all organisms.

In summary, we describe a virus of *A. glycines* that represents a new type of insect infecting RNA virus along with DAV. AGV and DAV may represent an ancestral group of viruses from which plant and insect viruses emerged or represent a more recent recombination between insect and plant viruses. The ability of hemipteran insects such as

aphids to acquire and transmit plant viruses provides opportunity for recombination between plant viruses and insect viruses.  Future studies on the origin of AGV and DAV may increase our understanding of plant and insect virus evolution.

## Author Contributions

SL performed bioinformatics analysis. DV conducted all laboratory experiments and construction of phylogenetic trees. BCB was involved in experimental design and discussion of data analysis. All authors were involved in data analysis and writing of the manuscript.

## Acknowledgements

# References

1.	Will T & Vilcinskas A (2013) Aphid-proof plants: Biotechnology-Based Approaches for Aphid Control. *Adv Biochem Eng Biotechnol*.

2.	Nault LR (1997) Arthropod transmission of plant viruses: A new synthesis. *Ann Entomol Soc Am* 90(5):521-541.

3.	Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, & Desneux N (2011) Ecology and management of the soybean aphid in North America. *Annu Rev Entomol* 56:375-399.

4.	Kim CS, Schaible G, Lynn G, Lubowski R, & Lee D (2008) Economic Impacts of the U.S. Soybean Aphid Infestation: A Multi-Regional competitive Dynamic Analysis. *Agricultural and Resource Economics Review* 37(2):227-242.

5.	Kim KS*, et al.* (2010) Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538. *Theor Appl Genet* 121(3):599-610.

6.	Kim KS*, et al.* (2010) Fine mapping the soybean aphid resistance gene Rag1 in soybean. *Theor Appl Genet* 120(5):1063-1071.

7.	Michel AP, Mittapalli O, & Mian RMA (2011) *Evolution of Soybean Aphid Biotypes: Understanding and Managing Virulence to Host-Plant Resistance* (InTech) p 514.

8.	Cox-Foster DL*, et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318(5848):283-287.

9.	Wu Q*, et al.* (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci U S A* 107(4):1606-1611.

10.	Padilla-Rodriguez M, Rosario K, & Breitbart M (2013) Novel cyclovirus discovered in the Florida woods cockroach *Eurycotis floridana* (Walker). *Arch Virol* 158(6):1389-1392.

11.	Runckel C*, et al.* (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS One* 6(6):e20656.

12.	Vijayendran D, Airs PM, Dolezal K, & Bonning BC (2013) Arthropod viruses and small RNAs. *J Invertebr Pathol* 114(2):186-195.

13.	Dombrovsky A & Luria N (2013) The Nerium oleander aphid *Aphis nerii* is tolerant to a local isolate of Aphid lethal paralysis virus (ALPV). *Virus Genes* 46(2):354-361.

14.    Granberg F, *et al.* (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid lethal paralysis, Israel acute paralysis and Lake sinai viruses. *PLoS One* 8(2):e57459.

15.    Ravoet J, *et al.* (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS One* 8(8):e72443.

16.    Aliyari R & Ding SW (2009) RNA-based viral immunity initiated by the Dicer family of host immune receptors. *Immunol Rev* 227(1):176-188.

17.    Liu S, Chougule NP, Vijayendran D, & Bonning BC (2012) Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. *PLoS One* 7(9):e45161.

18.    Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821-829.

19.    Conesa A, *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676.

20.    Haas BJ, *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494-1512.

21.    Dereeper A, *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server issue):W465-469.

22.    Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.

23.    Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.

24.    Guindon S, Lethiec F, Duroux P, & Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33(Web Server issue):W557-559.

25.    Wu S & Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 35(10):3375-3382.

26.    Ambrose RL, *et al.* (2009) Drosophila A virus is an unusual RNA virus with a T=3 icosahedral core and permuted RNA-dependent RNA polymerase. *J Gen Virol* 90(Pt 9):2191-2200.

27.     Zeddam JL*, et al.* (2010) Euprosterna elaeasa virus genome sequence and evolution of the Tetraviridae family: emergence of bipartite genomes and conservation of the VPg signal with the dsRNA Birnaviridae family. *Virology* 397(1):145-154.

28.     Pringle FM, Kalmakoff J, & Ward VK (2001) Analysis of the capsid processing strategy of Thosea asigna virus using baculovirus expression of virus-like particles. *J Gen Virol* 82(Pt 1):259-266.

29.     Gordon KH, Williams MR, Hendry DA, & Hanzlik TN (1999) Sequence of the genomic RNA of nudaurelia beta virus (Tetraviridae) defines a novel virus genome organization. *Virology* 258(1):42-53.

30.     Gorbalenya AE*, et al.* (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J Mol Biol* 324(1):47-62.

31.     Li L*, et al.* (2010) Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* 84(14):6955-6965.

32.     Gray S & Gildow FE (2003) Luteovirus-aphid interactions. *Annu Rev Phytopathol* 41:539-566.

33.     Goic B*, et al.* (2013) RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. *Nat Immunol* 14(4):396-403.

34.     Maori E, Tanne E, & Sela I (2007) Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* 362(2):342-349.

35.     Gibrat JF, Mariadassou M, Boudinot P, & Delmas B (2013) Analyses of the radiation of birnaviruses from diverse host phyla and of their evolutionary affinities with other double-stranded RNA and positive strand RNA viruses using robust structure-based multiple sequence alignments and advanced phylogenetic methods. *BMC Evol Biol* 13:154.

36.     Brault V*, et al.* (1995) Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. *EMBO J* 14(4):650-659.

37.     Gordon KH & Hanzlik TN (1998) Tetraviruses. *The insect viruses*, (Springer), pp 269-299.

38.     Strand MR & Burke GR (2013) Polydnavirus-wasp associations: evolution, genome organization, and function. *Curr Opin Virol* 3(5):587-594.

39.     Longdon B & Jiggins FM (2012) Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proc Biol Sci* 279(1744):3889-3898.

40.     Richards S, *et al.* (2010) Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *Plos Biology* 8(2).

**Table 1.** Primers used in RT-PCR and RACE PCR for AGV, ALPV and RhPV in *A. glycines*.  F denotes forward primer while R denotes reverse primer. N/A, not applicable.

| Primer name | Sequence (5' – 3') | AGV genome position |
|---|---|---|
| Tet – F | AGTGGCTGCGCATGCTCGTT | 1,733-1,753 |
| Tet – R | ACGCGCCTCTCCGTTGAACT | 2,642-2,662 |
| AGV 3'S – R | CCTGCTGAACCGCTCTT | |
| AGV 3' – F | CAGTACAGCAATACGGCTCATT | 4,322-4,344 |
| AGV 5' – F | GCAGGACCTTGCCTCGCTCAAA | 198-220 |
| AGV R/C – F | CACGCGCGGAATCTTTGCAG | 3,014-3,034 |
| AGV R/C – R | TCGGTCTTGGCGGCGTCATA | 3,863-3,883 |
| AGV C/T – F | TGTGACTCCGACACCGTCGAA | 3,923-3,944 |
| AGV C/T – R | GCACCGGGAGAAATCCCAGAGT | 4,462-4,484 |
| AGV 3' RACE | TCTCCCGGTGCCTCGTCTCACCACAGG | 4,473-4,500 |
| AGV 5' RACE | AAGTGCCGTAGCGCTGCCTCGAGCAC | 514-540 |
| ALPV – F | TGAACTTCGTGCAACGAACACTGTT | N/A |
| ALPV – R | TCCGCCTGCGTTAGGAAGAAGA | N/A |
| RhPV – F | AATCTGGCGTTGACGCGCTC | N/A |
| RhPV – R | TCCCCCATCATCAACATAGATGCGT | N/A |
| qRT-PCR AGV – F | TCCCCGCCACGTGAAGTGAA | 2,662-2,687 |
| qRT-PCR AGV – R | GCTACTGCGTGCGTGGTGAA | 2,860-2,880 |
| qRT-PCR Ago2 – F | CGTGTGTCAATGGTGTCATTTT | N/A |
| qRT-PCR Ago2 – R | ACTCGAGAGAGCTAAAGAGTCAGG | N/A |
| qRT-PCR Dicer2 – F | CGTTCGTCAAGTGTCAAGTGT | N/A |
| qRT-PCR Dicer2 –R | ACCAGGAGATCCATTGTTTGGTT | N/A |
| qRT-PCR Actin – F | ATTGAACCCCAAGGCTAATC | N/A |
| qRT-PCR Actin – R | GATCGAGACGAAGGATAGCA | N/A |
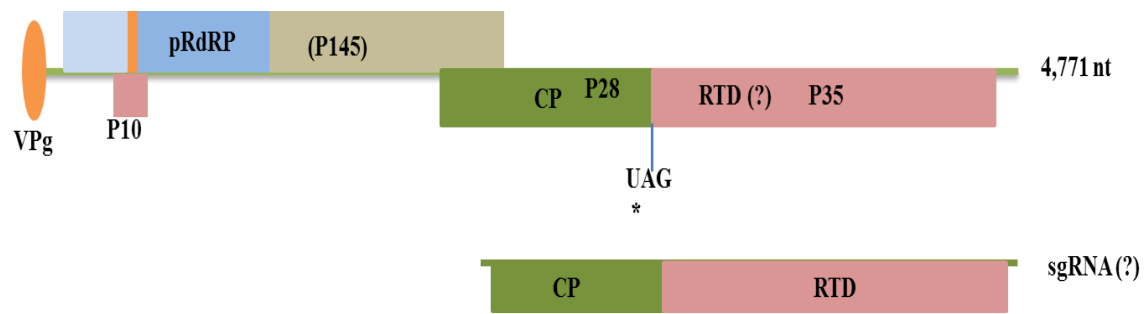
**Figure 1.** Schematic representation of the predicted AGV genome organization. The virus genome is capped at the 5' terminus with the putative RNA-dependent RNA polymerase (RdRP) encoded by the 5' ORF. The capsid protein (CP) is encoded towards the 3' terminus along with a putative read-through domain (RTD). AGV may encode a subgenomic RNA for the expression of CP and CP-RTD. VPg is encoded at nucleotides 580 to 619.

**Table 2.** Predicted AGV ORFs with the corresponding nucleotide position in the genome sequence, amino acid size, molecular mass and identity of encoded proteins.

| ORF | Positions on genome (nt) | Size (aa) | Mr (kD) | Proteins |
|------|--------------------------|-----------|---------|---------------------|
| P10  | 744 – 977                | 78        | 10.41   | Unknown             |
| P145 | 119 –3481                | 1121      | 145.51  | RdRP                |
| P28  | 3,351 - 4,001            | 217       | 28.18   | Capsid protein      |
| P35  | 4,005-4,805              | 267       | 35.03   | Read through domain |
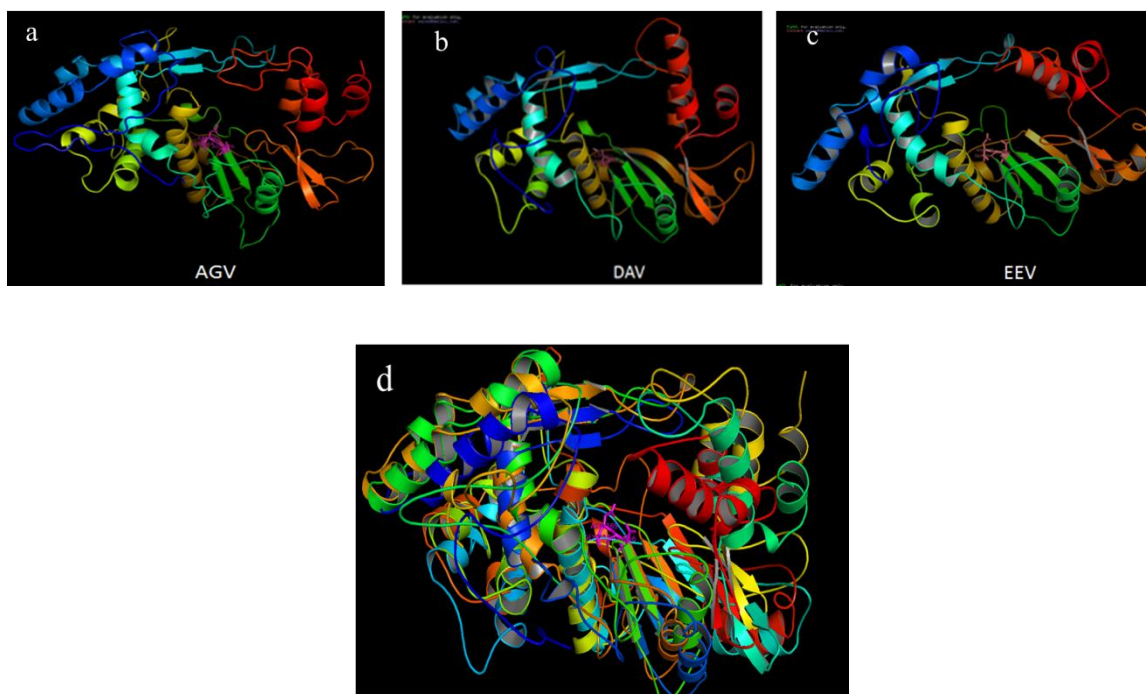
**Figure 2.** The predicted RdRP structures for AGV, DAV and EEV are similar. Homology modeling of tertiary protein structures of RNA-dependent RNA polymerase from (a) *Aphis glycines virus* (AGV), (b) *Drosophila A virus* (DAV) and (c) *Euprosterna elaeasa virus* (EEV). Image (d) shows the superimposed images of all three predicted RdRP structures. The images were generated using the LOMETS server.
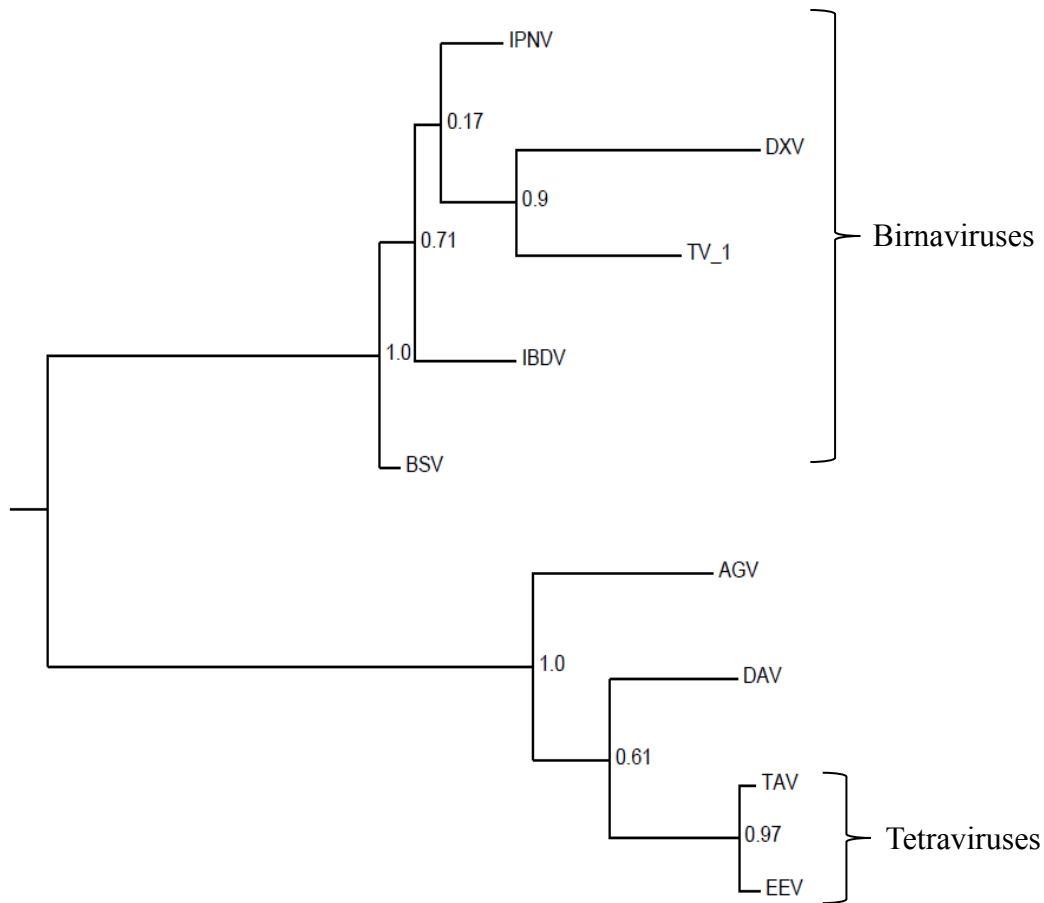
**Figure 3.** Phylogenetic tree of AGV RNA-dependent RNA polymerase with closely related viruses based on BLAST results of amino acid identity. Phylogenetic trees were constructed using the Phylogeny.fr tool with the maximum likelihood method. The numerical numbers on the branches represent the posterior probability values (a fraction closer to 1 indicates higher confidence levels). *Thosea asigna virus* (TAV; Accession no. AAQ14329.1) and *Euprosterna elaeasa virus* (EEV; NP573541.1) are closely related to but phylogenetically distinct from *Drosophila A virus* (DAV; YP003038595.1) and AGV. The birnaviruses form a distinct group from the tetraviruses. Birnaviruses used in the analysis, *Infectious pancreatic necrosis virus* (IPNV; AAV48847.1), *Drosophila X virus* (DXV; NP690806), *Tellina virus 1* (TV_1; CAI74982.1), *Infectious bursal disease virus* (IBDV; ACS44343.1), *Blotched snakehead virus* (BSV; YP052864).

```
EEV   MEASNPVIAPTRLSLEAMLEERASVVRQDLQSLRVRLEGAPRTLAPQAPEKQGRDSAQAA   60
TAV   MEASNPVIAPTRLSLEAMLAERAMVARQDLAGLKRKLAGADRVLAPQSPEQCGRESAQAQ   60
AGV   MDCSNPVNVNDRITVGDARIAAAGYIRQDLASLKNAAAKVPRNYLPTSTPRP-PDVVRAD   59
DAV   MDASNPIISSDRSTVNQVYAKMVRRKRQDLAYLQELVRTAHREILPQGELPHE--SEAKQ   58

EEV   AKSIVTELRDAVKEAQGLEHK---SLDFLEQLGEYPVCGIIHGDHPIHPSGTNNNKGKVS   117
TAV   ARSVTSELKSAVKEAQGLEHQ---TLDFLEQLGEYPVCGILHGDHPVHPSGTHNNNGKVS   117
AGV   LLRVKTSRRTVPNDPEGLARLNDLSEFLTLEAESIPICGYVGHQGNVHPSGVKMKQGRSL   119
DAV   YVAEYKARLPVVDTKKVG--ELPTEVFQVALPSIPVRGIMDAVGVPLHPPSINNRGGHTH   116

EEV   VQRQFTTSMNSVDALQVALRYEDVPLVRDIALKTTYTDGSLAGFVERLKLQTTRPCVQEK   177
TAV   VKRQFAAGVNTSDALTCAFRFEDSDLVRETALKTTYTDGTWAGFVQRLKMQTTRKCVQEK   177
AGV   ATRTFGGQTRDDIVLEAALRHFPYMDVAAAYSSAVYTSGTPSGFLERMRLNMTRPAVSSL   176
DAV   VTRQMGGVPHPPLGLIPIAQKYG---VEILDQKHVYTGGTWDGFLSRLAQQMGRTTVPLA   173
```

VPg

**Figure 4.** The AGV genome is predicted to be capped by a VPg. Clustal W sequence alignment of N-terminal sequences of *Euprosterna elaeasa virus* (EEV), *Thosea asigna virus* (TAV), *Aphis glycines virus* (AGV), and *Drosophila A virus* (DAV). The sequence alignment shows conservation of amino acid sequences (highlighted sequences) for the viral protein genome-linked (VPg) that caps the virus genome sequence at the 5' terminal.
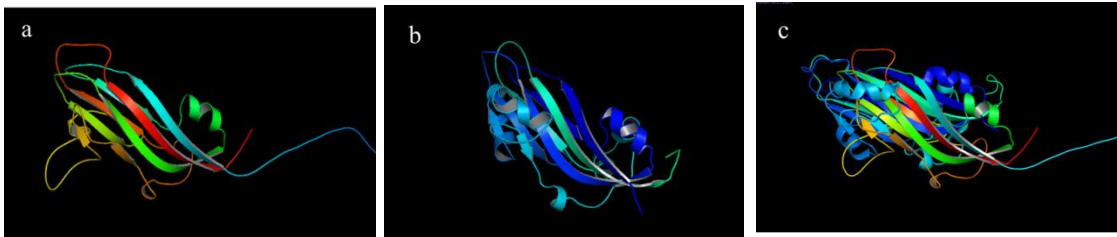
**Figure 5.** The predicted structures of AGV and TNV capsid proteins are similar. Homology modeling of tertiary protein structures of viral capsid protein for (a) *Aphis glycines virus* and (b) *Tobacco necrosis virus*, a plant virus. Image (c) shows superimposed images of the two predicted CP structures. The images were generated using the LOMETS server.
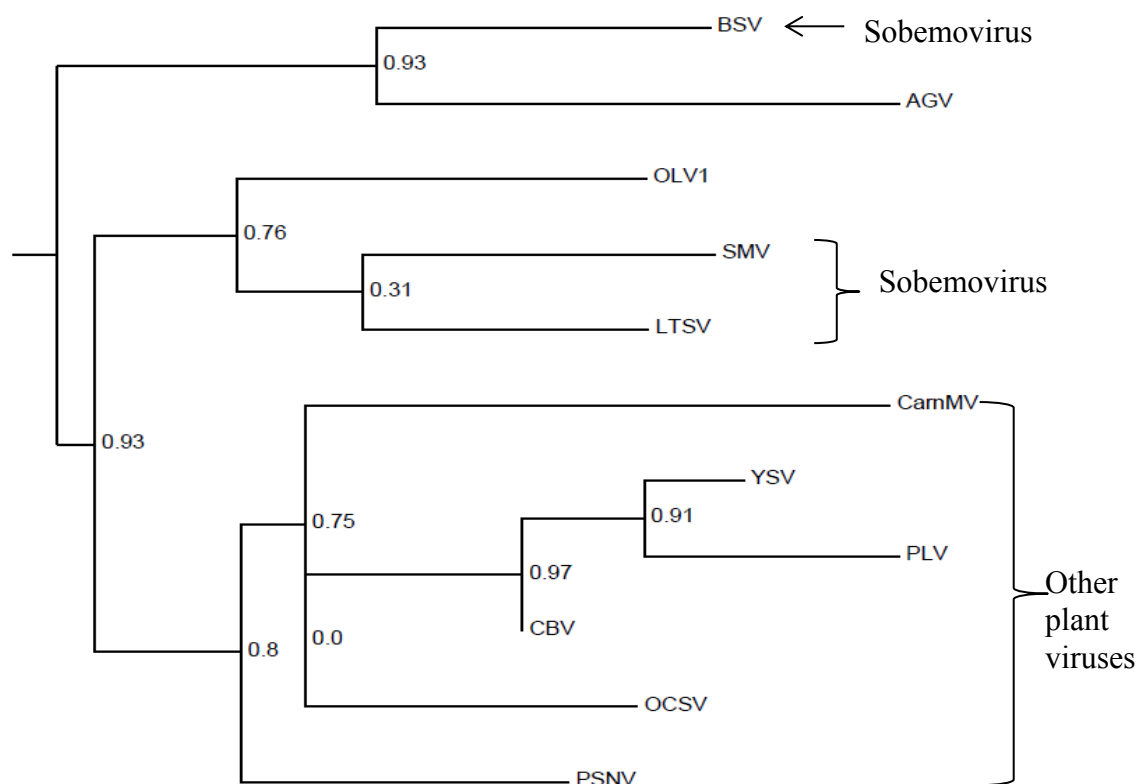
**Figure 6.** Phylogenetic tree of AGV capsid protein with closely related viruses based on BLAST results of amino acid identity. Phylogenetic trees were constructed using the Phylogeny.fr tool with a maximum likelihood method. The numerical number on the branches represents the posterior probability values (a fraction closer to 1 indicates higher confidence levels). AGV CP groups closely with the CPs of sobemoviruses, *Bat sobemovirus* (BSV; AGN73380.1), *Sowbane mosaic virus* (SMV; YP002158815.1) and *Lucerne transient streak virus* (LTSV; AAA79990.1). Other viruses used in this analysis, *Olive latent virus 1* (OLV1; AHE40766.1), *Carnation mottle virus* (CarnMV; CAH59636.1), *Yam spherical virus* (YSV; YP008828158.1), *Pothos latent virus* (PLV; CAA60597.1), *Cucumber Bulgarian virus* (CBV; NP835255.1), *Oat chlorotic stunt virus* (OCSV) and *Pea stem necrosis virus* (PSNV; NP862839.1).
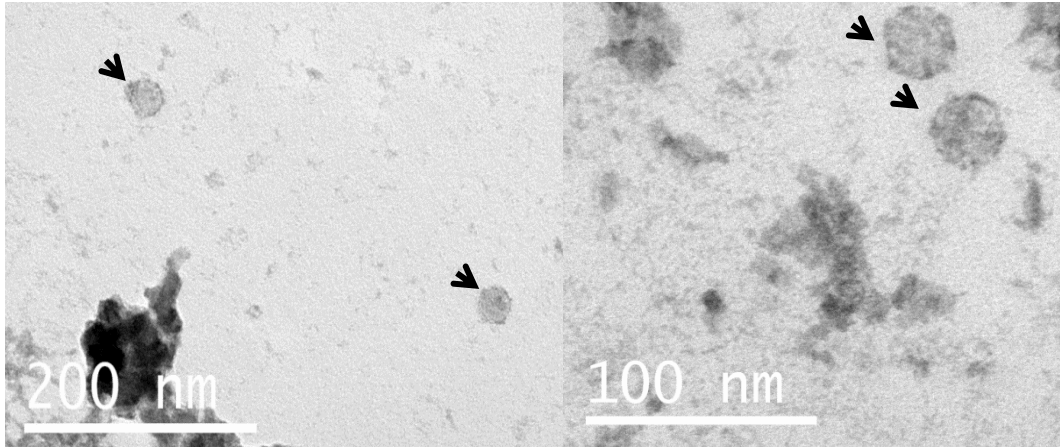
**Figure 7.** Virions of *Aphis gycines virus* viewed using a transmission electron microscope. Arrows indicate AGV virions. The virus has an icosahedral shape and a diameter of ~30 nm.
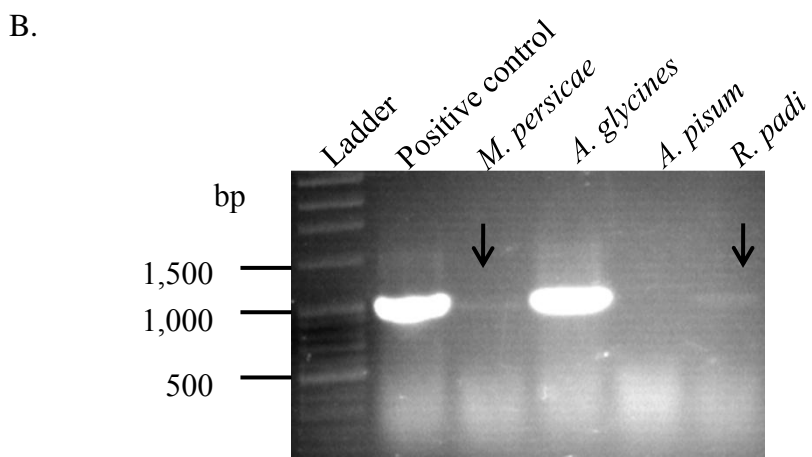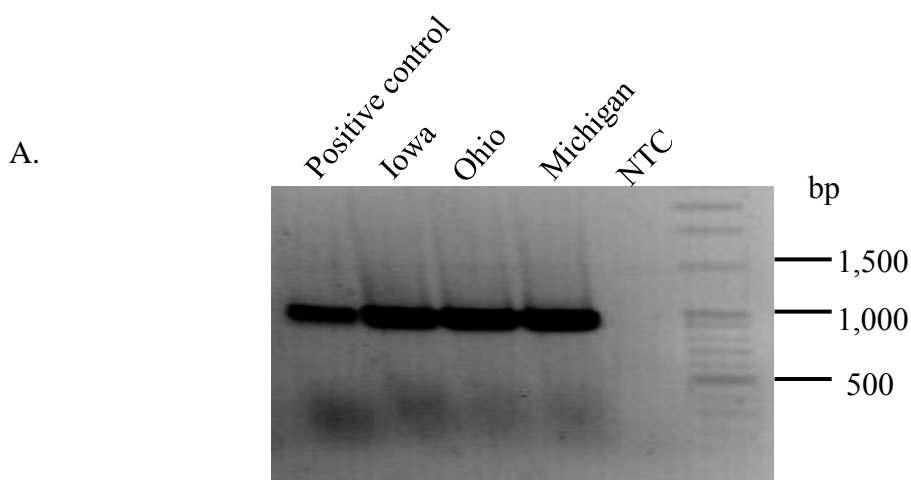
A.



B.



**Figure 8.** Detection of AGV RdRP sequence from (A) field collected *A. glycines* and (B) laboratory colonies of different aphid species. The 929 bp RT- PCR product was detected in field collected *A. glycines* from Iowa, Ohio and Michigan. The virus was also detected at low levels in the green peach aphid, *Myzus persicae* and bird cherry-oat aphid, *Rhopalosiphum padi* (arrows). The positive control and soybean aphid, *Aphis glycines* samples amplified the expected PCR products. No amplification was detected in the no template control (NTC) sample.

**Total number of reads**
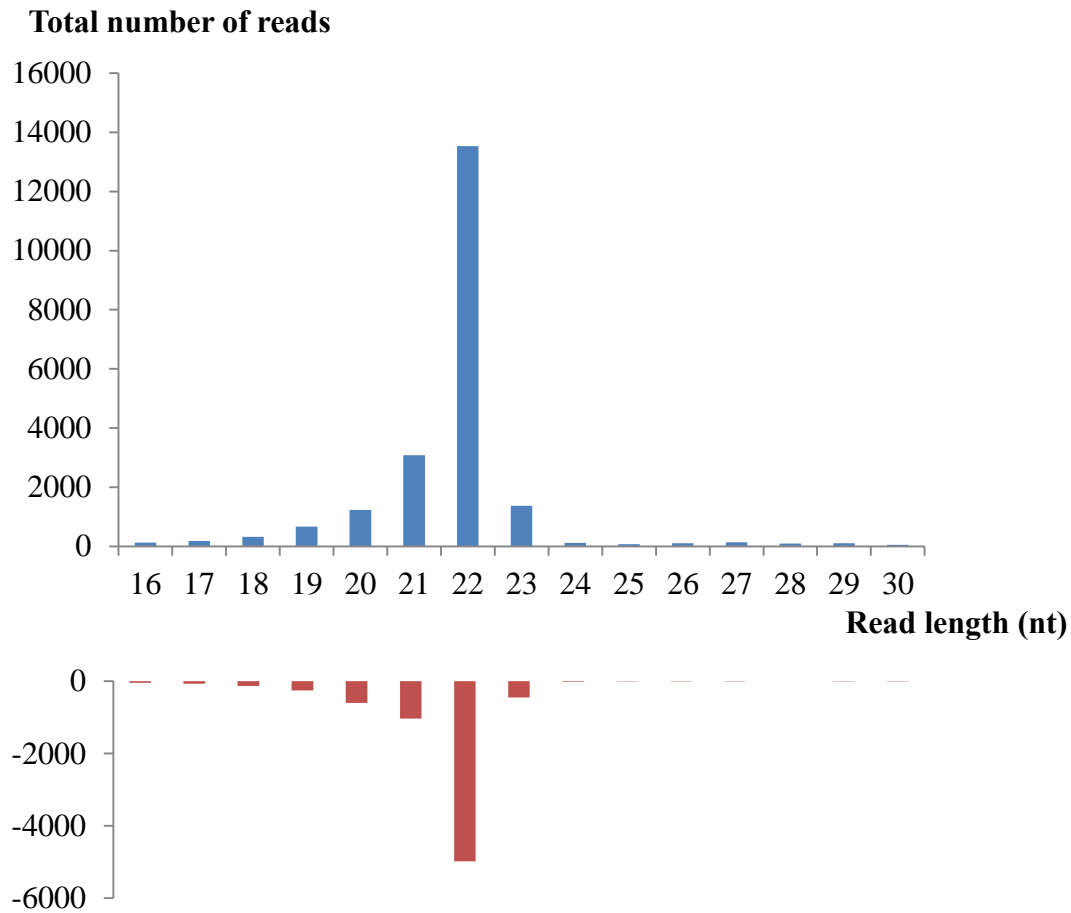


**Figure 9.** Virus-derived small interfering RNA (vsRNA) sequences derived from the sense (blue) and anti- sense (red) genome of AGV. Reads were extracted from a small RNA sequencing dataset of *A. glycines* infected with AGV and an ALPV-like virus. The majority of the AGV vsRNAs are 22 nt long, characteristic of double stranded RNA processing by the dicer protein.
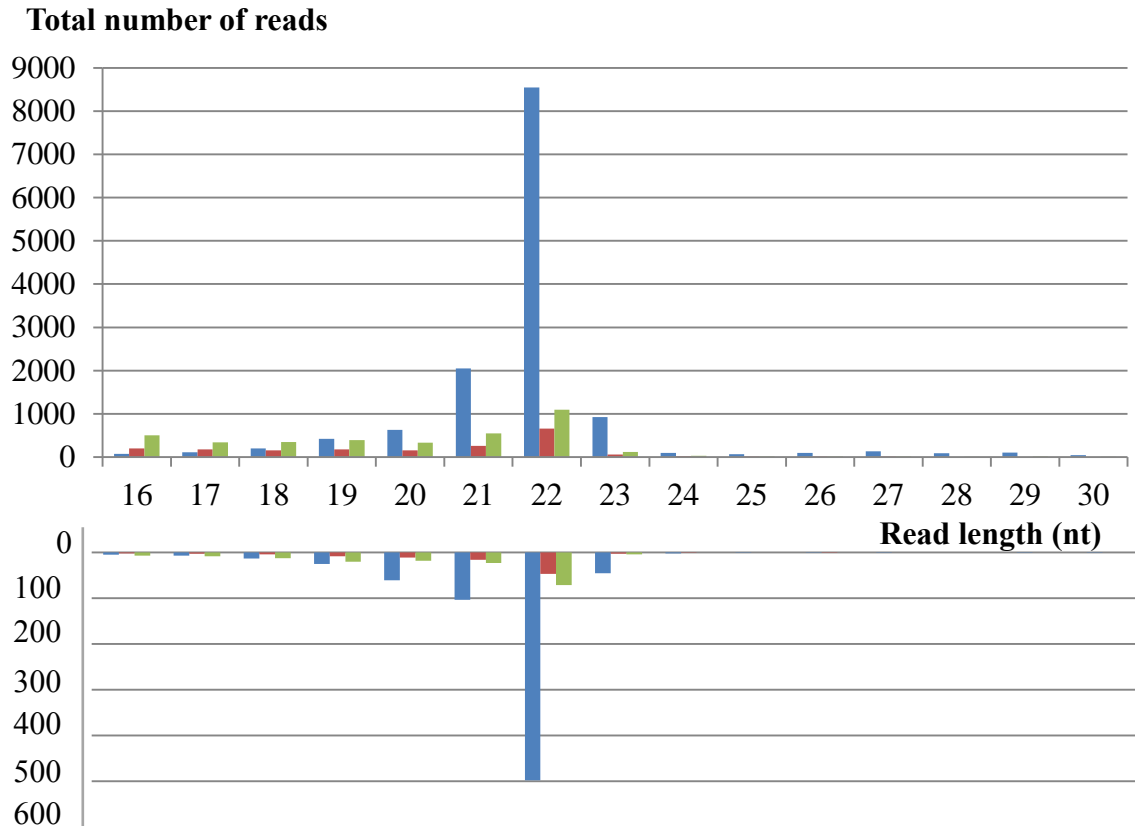
**Total number of reads**



**Figure 10.** AGV vsRNA are more abundant in the presence of an ALPV-like virus. The vsRNAs were isolated from *A. glycines* populations infected with both AGV and an ALPV-like virus (blue bars), AGV alone from Iowa (red) and AGV alone from Ohio (green). VsRNA that mapped to the positive strand (above) and to the negative strand (below) of the AGV genome are shown.
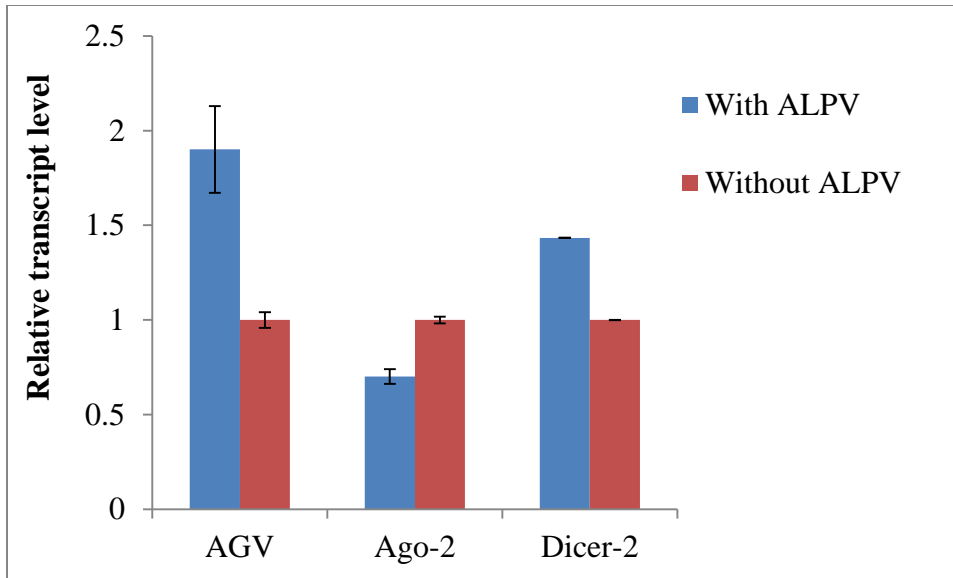
**Figure 11**. Relative transcript levels of AGV, Argonaute-2 and Dicer-2 in *A. glycines* infected with AGV in the presence or absence of an ALPV-like virus. Three independent biological replicates were carried out for each target transcript with data normalization to the internal control, actin. Black lines show standard error bars. There were no significant differences in transcript levels between the two treatments (Student's t-test, p<0.05).
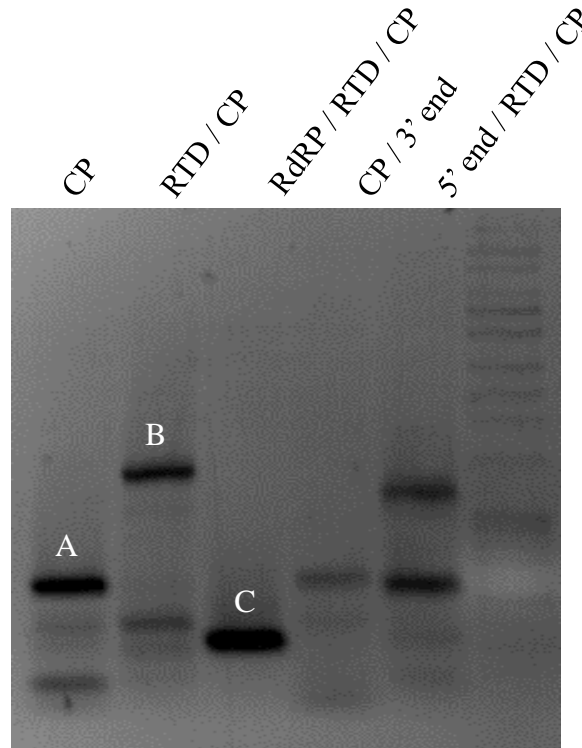
**Supplementary information**



**Figure 1.** AGV sequences are not detected in the *A. glycines* genome. Primers spanning the entire AGV genome were used to test for the presence of AGV sequence in the *A. glycines* genome. A representative agarose gel with amplified PCR products is shown. Labels indicate region of AGV genome spanned by the PCR primers. CP, capsid protein; RTD, read through domain and RdRP, RNA-dependent RNA polymerase. The sequenced PCR product from A and B did not have any significant matches to AGV or the NCBI database. PCR product C hit a pea aphid prestin-like transcript variant 5.
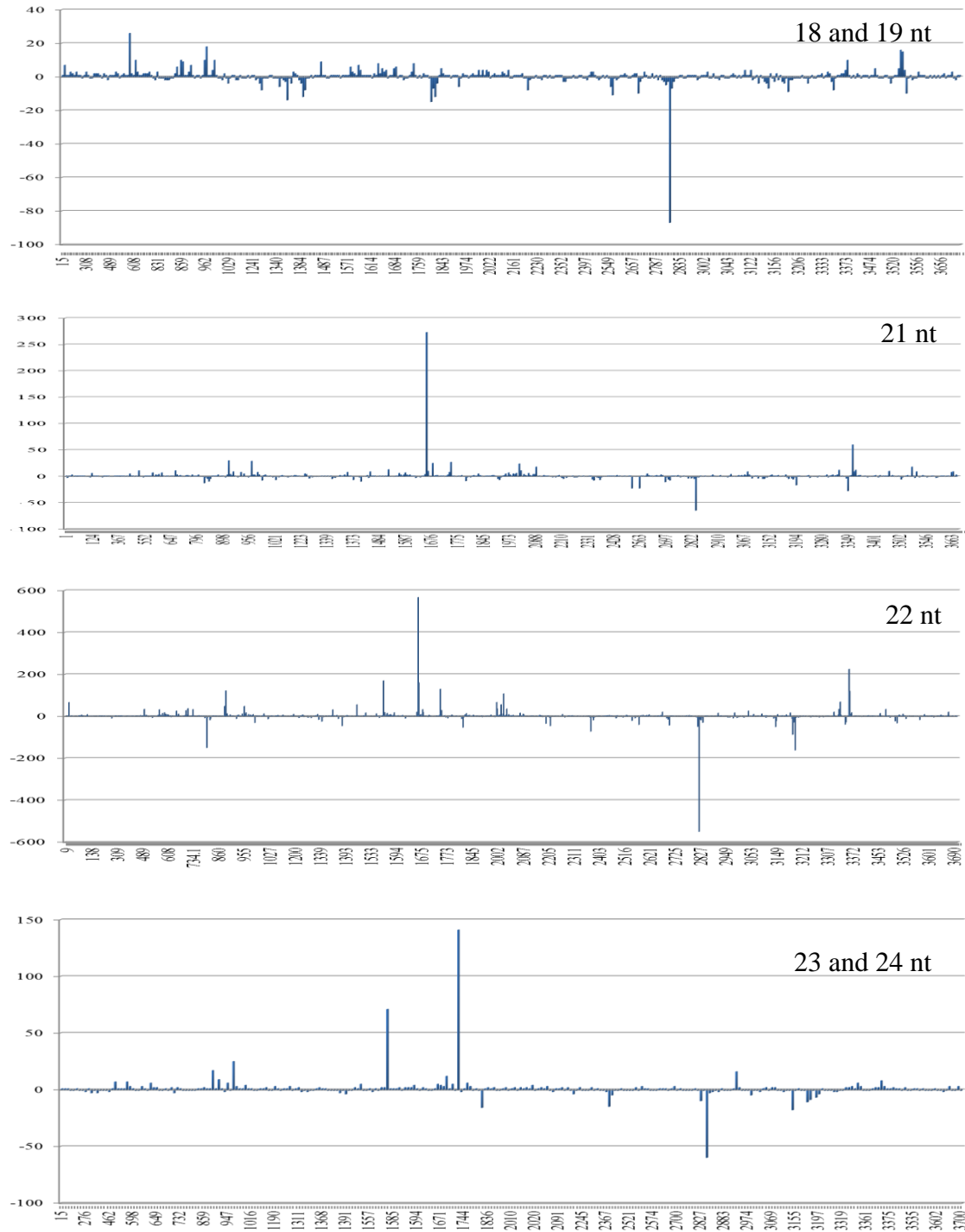
**Figure 2.** Distribution of vsRNAs on the sense (upper) and antisense genome (lower) of AGV based on small RNA read size. The X-axis shows the position on the virus genome while the Y-axis shows number of reads.

**CHAPTER 5**

**NEXT GENERATION SEQUENCING FOR IDENTIFICATION OF APHID-SPECIFIC MICRORNAS IN FOUR SPECIES OF APHIDS**

Diveena Vijayendran and Bryony C. Bonning

Department of Entomology, Iowa State University, Ames, IA, 50011

**Abstract**

Aphids are serious agricultural pests of economic importance. Current management options rely heavily on the use of chemical insecticides and on the use of aphid resistant cultivars, but these options are being challenged by the development of resistant aphid biotypes. In our efforts to identify novel avenues to manage these pests, we studied the microRNA (miRNA) expression profile from the small RNA (sRNA) datasets of four pest species of aphid, the pea aphid, *Acyrthosiphon pisum*, the soybean aphid, *Aphis glycines*, the green-peach aphid, *Myzus persicae* and the bird cherry-oat aphid, *Rhopalasiphum padi*. We describe the identification and abundance of conserved miRNAs in all four species and further identified a subset of 12 aphid-specific miRNAs. Relative quantification of three candidate aphid-specific miRNAs showed miRNA expression in multiple life stages of *M. persicae*. The identification of such pest-specific miRNAs could lead to development of miRNA-based management options unique to the targeted pest, in this case aphids.

**Introduction**

MicroRNA (miRNA) is a class of small RNA (sRNA) identified in eukaryotes for the regulation of genes. A primary miRNA (pri-miRNA) is transcribed by cellular RNA polymerase II in the nucleus; the pri-miRNA is processed by Drosha with cofactor Pasha into pre-cursor miRNA (pre-miRNA) of ~70 nt. The characteristic hairpin loop structure of the precursor miRNA is recognized by the cellular protein Dicer which cleaves it into smaller duplexes of 18-25 nt long (1). One strand of the duplex is then loaded onto an Argonaute-protein complex (RISC complex) guiding it to target messenger RNA (mRNA). Micro RNA is known to bind most frequently at the 3' UTR of the target mRNA although other regions of the mRNA can also be targeted for gene regulation (2-4). Increasing research in animal models has shown the ability of a single miRNA to regulate many mRNA targets in a cell (5). Although much less common, incidents of multiple miRNAs simultaneously regulating a particular target mRNA have also been observed (6).

The current version of miRBase, a miRNA database includes over 1000 insect miRNAs including 103 miRNAs identified from *A. pisum* (7). Targets have been identified and validated for a small subset of these insect miRNAs primarily in *Drosophila* (8-10). The miRNA functional studies carried out in insects have shown regulation of mRNAs involved in important biological functions such as wing-development, tissue differentiation and cell proliferation (10, 11). The miRNAs of the cotton-melon aphid, *Aphis gossypii* were differentially expressed when the aphid fed on host plant cultivars that were either susceptible or resistant to the aphids (12).

The importance of miRNA function in insects makes it an attractive biological phenomenon that could be manipulated for the development of management strategies for pests such as aphids. Aphids are of major economic importance worldwide (13). *A. glycines* alone is estimated to have resulted in $1.6 billion in losses over a 10 year period in the United States, resulting both from yield loss and from application of chemical insecticides for management (14). The propensity for aphid populations to increase exponentially when conditions are favorable combined with their ready evolution of resistance to chemical insecticides (15, 16) make them particularly problematic. In addition, aphids are not particularly susceptible to toxins derived from the bacterium *Bacillus thuringiensis* which have increasingly been adopted for development of insect pest resistant transgenic plants (17). Although two promising strategies for development of aphid resistant transgenic plants have recently been published (18, 19), additional strategies for aphid management would allow for more durable management to be achieved when multiple physiological target sites in the pest are used (20).

In this study, we looked at the abundance of conserved and unique miRNAs in four species of economically important aphids, the pea aphid, *Acyrthosiphon pisum*, the soybean aphid, *Aphis glycines*, the green-peach aphid, *Myzus persicae* and the bird cherry-oat aphid, *Rhopalasiphum padi.* . We show that many conserved insect miRNAs are expressed in all four aphid species and identify a subset of aphid-specific miRNAs. We also present data on the expression profile of three of the aphid-specific miRNAs which are present in multiple life stages in *M. persicae*.

## Methods and Materials

### Insects

Pea aphids, *Acyrthosiphon pisum* were purchased from Berkshire Biological (Westhampton, Massachusetts) and were raised on broad bean, *Vicia faba* in a growth chamber at $24^{o}$C with a 12:12 (light:dark) cycle. Field collected soybean aphids, *Aphis glycines* Matsumura were reared on soybean, *Glycine max* at room temperature ($20°$C to $25°$C) with a continuous light cycle. Green peach aphids, *Myzus persicae* were reared on Chinese cabbage, *Brassica rapa* at room temperature ($20°$C to $25°$C) with a continuous light cycle and the bird cherry-oat aphids, *Rhopalasiphum padi* were reared on corn, *Zea mays* in a growth chamber at $25\,^{o}$C with a 12:12 (light:dark) cycle.

### Total RNA extraction

All instars of each aphid species were collected and used for RNA isolation. Approximately 50-60 aphids were homogenized in 1mL of TRIzol® reagent (Life Technologies) with a pestle in a 1.5 ml micro centrifuge tube. Procedures for RNA isolation followed the manufacturer's instructions. RNA was precipitated overnight in isopropanol at -80°C for optimal recovery of sRNA. Total RNA was re-suspended in 30 µl nuclease-free water. RNA was quantified using a Nanadrop 2000 (Thermo Scientific) and the RNA quality assessed by using a Bioanalyzer (Agilent).

### Small RNA sequencing

sRNAs were isolated from total RNA and cDNA libraries constructed using the TruSeq Small RNA Sample Prep Kit (Illumina) according to the manufacturer's instructions. The sRNAs were sequenced for 50 cycles using an Illumina GA II platform.

sRNA isolation, sequencing library preparation and high-throughput sequencing were conducted at the Iowa State University DNA Facility.

**MicroRNA analysis**

sRNA reads were trimmed with custom Perl code. Reads (> 2 copies) were grouped together for analysis. The sRNA reads from each sample was analyzed for presence of known miRNAs using miRanalyzer (http://bioinfo5.ugr.es/miRanalyzer/miRanalyzer.php) with no mismatches allowed for mature miRNA reads matched to mirBase.

**Staging developmental stages of *M. persicae***

*M. persicae* were isolated into individual petri dishes on a leaf with a wet paper towel. The adults were monitored for new born nymphs. Five newborns were pooled into 1.5 mL tube containing 500 µL TRIzol® reagent. The remaining newborns were further isolated into petri dishes as previously described. The individual aphids were monitored for molting and isolated based in the number of molts which correspond to the life stages. Aphids were collected at 24 hour time points with an average molting time of ~30-35 hours/instar stage. A total of five aphids were grouped together and three biological replicates were collected for each life stage. Life stages collected were, new born (NB), first and second instar (1-2 instar), third and fourth instar (3-4 instar) and adults (A).

**Relative quantification of miRNA levels**

Expression of the most abundant aphid-specific miRNAs (miR-3024, miR-3027 and miR-3050) was determined for staged *M. persicae* using stem-loop RT-PCR. Total RNA was extracted as previously described. Primers used are listed in Table 1. The RT-step was carried out in a 10 µL reaction with 500 ng of total RNA using Superscript III®RT (Life technologies) according to the manufacturer's recommendations with gene specific primers. The qRT-PCR was carried out using the iQ SYBR® Green Supermix (Biorad) in 10 µL reactions with 100 ng of cDNA template. The protocols for RT and PCR were as previously described (21). The experiments were carried out with three biological replicates. The relative expression level was calculated by data normalization to ubiquitin as the internal control gene and the adult life stage used as the calibrator sample. Standard errors were calculated and significance was determined using student's t-test with p-value <0.05.

**Results**

**Many conserved insect miRNAs are expressed in aphids**

Over 100 mature miRNAs were identified from all four species of aphids (Table 2). The highest number of miRNAs was identified from *A. pisum* with a total of 94 miRNAs. This represents 91.3% of all miRNAs that have been identified from *A. pisum* by previous researches. Of the 94 identified miRNAs, 42 have sequence conservation with miRNAs identified from other organisms. In comparison, only ~68% of miRNAs identified from the other aphid species were conserved relative to *A. pisum* miRNAs. A subset of 16 miRNAs is *A. pisum-* specific and these miRNAs are not conserved with

those of other insects or other species of aphid. The miRNA fraction of the total sRNA ranged from 27.8% in *A. glycines* to 7.4% in *M. persicae*. Although the percentage of miRNA reads was low in *M. persicae*, the number of miRNAs identified was not low compared to the other species of aphids. Figure 1 shows miRNA conservation in the four species of aphids.

**Abundance of miRNAs in multiple aphid species**

Several miRNAs were highly abundant in all species of aphids including miR-8, miR-1, miR-184 and mir-276. The least abundant miRNAs in aphids included miR-iab-4, miR-92 and miR-100 which are universal miRNAs found in other species of insects. The majority of the miRNAs identified were present at very low levels with fewer than 10 copies. The number of miRNAs identified was reduced as the copy number increased with relatively few (six or less) miRNAs expressed at >100,000 reads, and 34 to 45 miRNAs with 10 or fewer reads (Figure2).

**A group of aphid-specific miRNAs**

Analysis of miRNA profiles for each species of aphid revealed conservation of 12 miRNAs that were expressed in all aphid species analyzed. These miRNAs have thus far only been described in aphids with no sequence conservation with any other organism (Table 3). The aphid-specific miRNAs range in size from 21 nt to 24 nt. These miRNAs were generally present at low abundance compared to other conserved miRNAs. Some of the aphid-specific miRNAs are encoded in *A. pisum* genome as clusters; miR-3016, 3024 and 3032 (superscript a) are encoded together while miR-3031, 3047 and 3037

(superscript b) are encoded together. The other miRNAs are expressed individually or with other miRNAs that are not unique to aphids.

**Three aphid-specific miRNAs are expressed in all life stages of *M. persicae***

Expression of the three aphid-specific miRNAs with the highest abundance, miR-3024, miR-3027 and miR-3050 was determined in *M. persicae*. Figure 3 shows the relative expression of these miRNAs in the various aphid life stages compared to the adults. All three miRNAs were detected in the various life stages. Peak miR-3024 expression was detected in the third and fourth instars with the lowest expression levels in newborns. The expression of miR-3024 increased with the development of the aphid with levels in the newborns and the first and second instar stages being significantly lower than the levels in the adult life stage (Student's t-test, $p<0.05$). miR-3027 was detected at similar levels throughout the nymphal stages with similar levels in the adults and newborn. Levels of miR-3027 were significantly higher in the first and second instars (Student's t-test, $p<0.05$). miR-3050 was present at all life stages with significantly higher expression in the newborns compared to adults (Student's t-test, $p<0.05$). The standard error obtained between biological replicates for miR-3027 and miR-3050 was notably larger than for miR-3024.

**Discussion**

In this paper, we present data on the identification and abundance of miRNAs from four species of aphids. We further describe a subset of 12 miRNAs that is unique to aphids. Some of these aphid-specific miRNAs are encoded in clusters in the genome of *A. pisum* . MicroRNA biogenesis with the transcription of pri-miRNAs of these clusters

suggests that these miRNAs are expressed at the same level and may target related genes in a pathway for modulation of gene expression. The abundance of miRNAs in aphids varies from low copy number (<10) to more than 100,000 copies. The highly abundantmiRNAs are conserved with other organisms and appear to be universal miRNAs such as miR-8 that is involved in determination of body size in *Drosophila* (22). The organism-specific miRNAs appear to be present at much lower levels in comparison. Low abundance however should not be taken to indicate that the miRNAs are any less important to the organism as small changes in expression of a gene can have major effects on an organism. The role of low abundance miRNAs as an important regulator of colon cancer was recently demonstrated (23).

Aphid miRNAs have been previously examined in *A. pisum* and in *A. gossypii* (12, 24). Key bioinformatics analysis of the *A. pisum* genome revealed duplications of genes involved in the miRNA pathways, specifically detection of two copies of Dicer-1 and Argonaute-1 each and four copies of Pasha (25). It was further determined that one copy of Dicer-1 and Argonaute-1 is evolving slowly while the other copy is evolving rapidly in the aphid (26). This duplication and differential rate of evolution of genes involved in the miRNA pathway may provide aphids with the ability to generate novel miRNAs from new transcripts with potential development of aphid-specific miRNAs (25). MicroRNA studies in the cotton-melon aphid in relation to susceptible and resistant melon plants showed the presence of a subset of 9 novel *A. gossypii* miRNAs that were differently regulated between the treatments (12). Using sequence comparison, we detected 8 of the 9 novel *A. gossypii* miRNAs in our sRNA datasets from other aphid species (data not shown). This result suggests that there may be little conservation of

miRNAs at the species level and perhaps more conservation at the family level. Additional sequencing data from other aphid species across different subfamilies would provide greater insight into miRNA conservation. The complete genome sequence would however be needed to identify novel species-specific miRNAs. When discovered, species-specific miRNAs could provide insights into novel regulation of pathways as revealed in other organisms (27, 28).

Our analysis of the abundance of three aphid-specific miRNAs in *M. persicae* shows that these miRNAs are expressed at all life stages of the aphid with variation in levels at different developmental stages. Statistical analysis showed large variations in miR-3027 and miR-3050 levels between three biological replicates but almost no variation was observed for miR-3024. This could be attributed to the function of the miRNAs in various biological processes and the targets of these specific miRNAs could be rapidly modulated resulting in the large variation observed. More precise developmental staging of the aphids, increasing the representative number of aphids used in each biological replicate and target predictions could further clarify the results obtained.

From an aphid management perspective, it is important for a target miRNA to be expressed in all life stages for effective targeting of the pest. Molecular studies to disrupt miRNA function could be used to further characterize these miRNAs and assess their importance for aphid survival. To establish which miRNAs would be good candidates for further studies, bioassays using synthetic miRNA mimics and inhibitors can be utilized. The effectiveness of this method has been demonstrated for other pest insects. Disruption of miR-275 function using antagomiR (modified miRNA inhibitors) effectively blocked

blood digestion and egg development in *Aedes aegypti* mosquitoes (29). Over-expressing and inhibiting miR-2002b by feeding larvae on synthetic inhibitors and mimics resulted in negative effects on the cotton bollworm, *Helicoverpa amigera* (30).

The ultimate goal for aphid management is the ability to target the miRNAs using a plant-mediated approach. It has been demonstrated that stable transgenic plants can be used to express double stranded RNA for targeted gene silencing in the green peach aphid (31, 32). MicroRNAs inhibition has been shown to be an effective method not only for gene function assays but also for management strategies. Over-expression or inhibition of miRNAs will result in mis-regulation of protein synthesis which could be devastating to the pest. Our identification of these aphid-specific miRNAs presents a unique opportunity for development of an aphid-specific insecticide for generation of aphid resistant transgenic plants.

**Author Contributions**

DV conducted bioinformatics and laboratory experiments. BCB was involved in experimental design and discussion of data analysis. Both authors wrote the paper.

**Acknowledgements**

**Reference**

1.      MacRae IJ*, et al.* (2006) Structural basis for double-stranded RNA processing by dicer. *Science* 311(5758):195-198.

2.      Behura SK (2007) Insect microRNAs: Structure, function and evolution. *Insect Biochem Mol Biol* 37(1):3-9.

3.      Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136(2):215-233.

4.      Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4):642-655.

5.      Pillai RS (2005) MicroRNA function: multiple mechanisms for a tiny RNA? *RNA* 11(12):1753-1761.

6.      Wu S*, et al.* (2010) Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* 29(15):2302-2308.

7.      Kozomara A & Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 39(Database issue):D152-157.

8.      Brennecke J, Hipfner DR, Stark A, Russell RB, & Cohen SM (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113(1):25-36.

9.      Enright AJ*, et al.* (2003) MicroRNA targets in Drosophila. *Genome Biol* 5(1):R1.

10.     Schnall-Levin M, Zhao Y, Perrimon N, & Berger B (2010) Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3'UTRs. *Proc Natl Acad Sci U S A*, United States 107(36):15751-15756.

11.     Jones CI & Newbury SF (2010) Functions of microRNAs in Drosophila development. *Biochem Soc Trans* 38(4):1137-1143.

12.     Sattar S*, et al.* (2012) Expression of small RNA in *Aphis gossypii* and its potential role in the resistance interaction with melon. *PLoS One* 7(11):e48579.

13.     Blackman RL & Eastop VF (2000) *Aphids on the world's crops: An identification and information guide* (John Wiley and Sons Ltd, Chichester) 2nd Ed.

14.     Kim CS, Schaible G, Lynn G, Lubowski R, & Lee D (2008) Economic Impacts of the U.S. Soybean Aphid Infestation: A Multi-Regional competitive Dynamic Analysis. *Agricultural and Resource Economics Review* 37(2):227-242.

15.    Devonshire A (1989) Resistance of aphids to insecticides. *Aphids, their biology, natural enemies and control* 100:123-139.

16.    Minks AK & Harrewijn P (1987) *Aphids. Their biology, natural enemies and control. Volume 2A* (Elsevier).

17.    Chougule NP & Bonning BC (2012) Toxins for transgenic resistance to hemipteran pests. *Toxins (Basel)* 4(6):405-429.

18.    Bonning BC*, et al.* (2014) Toxin delivery by the coat protein of an aphid-vectored plant virus provides plant resistance to aphids. *Nat Biotechnol* 32(1):102-105.

19.    Chougule NP*, et al.* (2013) Retargeting of the Bacillus thuringiensis toxin Cyt2Aa against hemipteran insect pests. *Proc Natl Acad Sci U S A* 110(21):8465-8470.

20.    Gould F (1998) Sustainability of transgenic insecticidal cultivars: integrating pest genetics and ecology. *Annu Rev Entomol* 43:701-726.

21.    Varkonyi-Gasic E, Wu R, Wood M, Walton EF, & Hellens RP (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 3:12.

22.    Jin H, Kim VN, & Hyun S (2012) Conserved microRNA miR-8 controls body size in response to steroid signaling in Drosophila. *Genes Dev* 26(13):1427-1432.

23.    Parasramka MA*, et al.* (2012) A role for low-abundance miRNAs in colon cancer: the miR-206/Krüppel-like factor 4 (KLF4) axis. *Clin Epigenetics* 4(1):16.

24.    Legeai F*, et al.* (2010) Bioinformatic prediction, deep sequencing of microRNAs and expression analysis during phenotypic plasticity in the pea aphid, *Acyrthosiphon pisum*. *BMC Genomics* 11:281.

25.    Jaubert-Possamai S*, et al.* (2010) Expansion of the miRNA pathway in the hemipteran insect *Acyrthosiphon pisum*. *Mol Biol Evol* 27(5):979-987.

26.    Ortiz-Rivas B*, et al.* (2012) Evolutionary study of duplications of the miRNA machinery in aphids associated with striking rate acceleration and changes in expression profiles. *BMC Evol Biol* 12:216.

27.    Bao H*, et al.* (2013) Expansion of ruminant-specific microRNAs shapes target gene expression divergence between ruminant and non-ruminant species. *BMC Genomics* 14:609.

28.    Mor E*, et al.* (2011) Species-specific microRNA roles elucidated following astrocyte activation. *Nucleic Acids Res* 39(9):3710-3723.

29.     Bryant B, Macdonald W, & Raikhel AS (2010) microRNA miR-275 is indispensable for blood digestion and egg development in the mosquito *Aedes aegypti*. *Proc Natl Acad Sci U S A* 107(52):22391-22398.

30.     Jayachandran B, Hussain M, & Asgari S (2013) An insect trypsin-like serine protease as a target of microRNA: utilization of microRNA mimics and inhibitors by oral feeding. *Insect Biochem Mol Biol* 43(4):398-406.

31.     Pitino M, Coleman AD, Maffei ME, Ridout CJ, & Hogenhout SA (2011) Silencing of aphid genes by dsRNA feeding from plants. *PLoS One* 6(10):e25709.

32.     Bhatia V, Bhattacharya R, Uniyal PL, Singh R, & Niranjan RS (2012) Host generated siRNAs attenuate expression of serine protease gene in *Myzus persicae*. *PLoS One* 7(10):e46343.

**Table 1.** Primers used for stem-loop RT-PCR analysis of miRNA expression in *M. persicae*.

| Primer name | Sequence (5' to 3') |
|---|---|
| api-miR-3024-F | CGCCGGTCTTTGGGATTTAATAG |
| api-miR-3024-RT | GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATAC GACACCGGC |
| api-miR-3027-F | CGACGCCAGTCTTGCATTTATTC |
| api-miR-3027-RT | GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATAC GACAGTGGA |
| api-miR-3050-F | CGAGCTGTGAGATCTTGATAAACT |
| api-miR-3050-RT | GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATAC GACAGGCGA |
| Universal-RT | GTGCAGGGTCCGAGGT |
| Ubiquitin-F | GCTACTGGAAGCCGTCACTG |
| Ubiquitin-R | GGCGAGCCTTCTTGTTCTTGG |

**Table 2.** Mature miRNAs identified from small RNA datasets of multiple aphid species

| Aphid species | Total number of miRNAs | Conserved with pea aphid | Conserved with other organism | Percentage of known miRNAs from pea aphid | Percentage of sRNA reads corresponding to miRNAs |
|---|---|---|---|---|---|
| Pea aphid, *Acyrthosiphon pisum* | 136 | 94 | 42 | 91.3% | 16.5% |
| Soybean aphid, *Aphis glycines* | 125 | 71 | 54 | 68.9% | 27.8% |
| Green-peach aphid, *Myzus persicae* | 147 | 70 | 77 | 68% | 7.4% |
| Bird cherry-oat aphid, *Rhopalosiphum padi* | 141 | 72 | 69 | 69.9% | 23.7% |

**A. pisum**



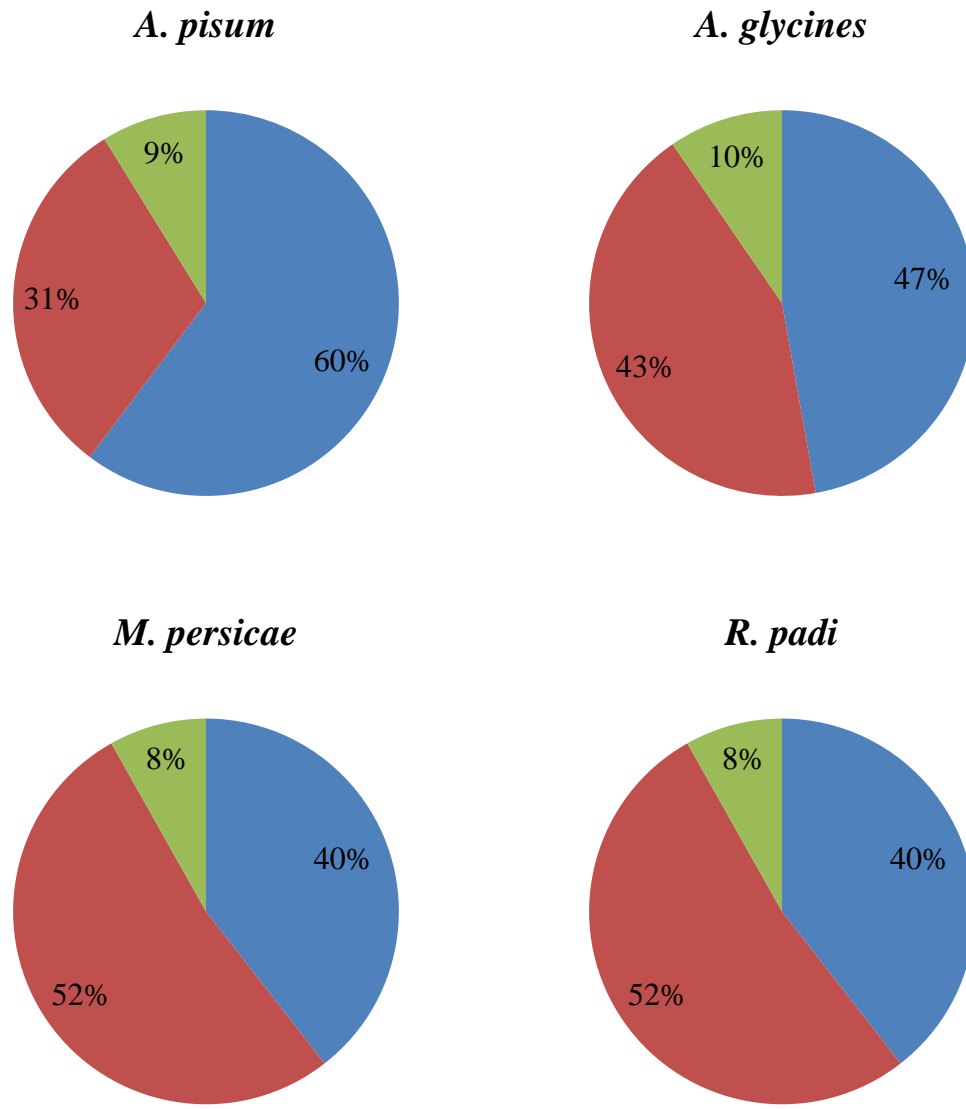**A. glycines**



**M. persicae**



**R. padi**



**Figure 1.** Conservation of miRNAs identified in four species of aphid. Blue represents percentage of miRNAs conserved with previously identified *A. pisum* miRNAs. Red represents percentage of miRNAs conserved with other organisms and green represents percentage of miRNAs that are aphid-specific.
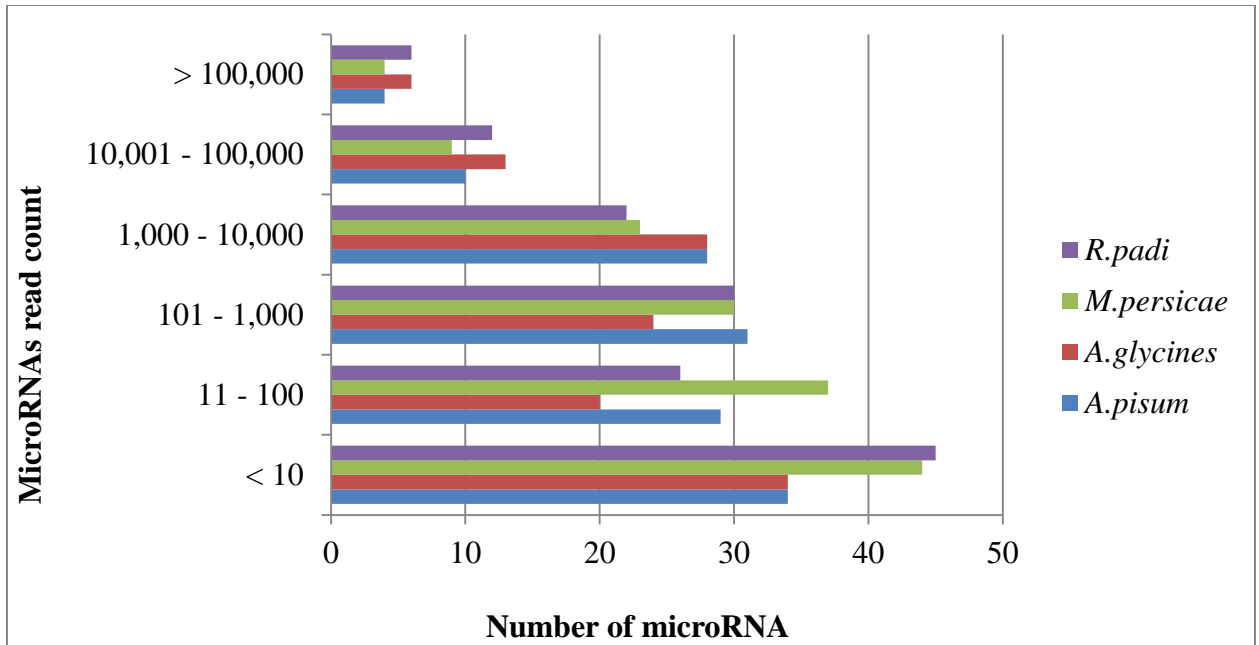
**Figure 2.** Abundance of miRNAs identified from multiple species of aphids. miRNA reads from aphids were identified with reference to mature miRNA sequences in miRBase. Identical reads of >2 with no sequence mismatch were used for the analysis. Among the most highly expressed miRNAs (> 100,000) are miR-1, miR-184, miR-276 and miR-263 which are highly conserved in insects.

**Table 3.** Aphid-specific miRNAs identified from four species of aphids. The different superscript letters indicate miRNAs that are encoded in clusters. api-miR refers to miRNAs identified from *A. pisum*.

| api-miR | Sequence | Length |
|---|---|---|
| miR-3016[a] | AUUGGUAACACAUACGUCUUUAG | 23nt |
| miR-3024[a] | UCUUUGGGAUUUAAUAGAGCCGGU | 24nt |
| miR-3032[a] | UGUUAGUAUAACUCUUAGUAACA | 23nt |
| miR-3031[b] | UUGCUUUUUAACAAGUUUCACUA | 23nt |
| miR-3047[b] | CAAAACAUUCAAAACUCCCUAC | 22nt |
| miR-3037[b] | UUACAAAACAUUCAGAAUUUUG | 22nt |
| miR-3027 | CCAGUCUUGCAUUUAUUCCACU | 22nt |
| miR-3039 | AAAAACGUCAAAACACGGUGG | 21nt |
| miR-3040 | CAGCCGGUGGUGACUGUUUCCACA | 24nt |
| miR-3041 | UUAAAGCUUUGAUGACGGGAUA | 22nt |
| miR-3050 | UGAGAUCUUGAUAAACUCGCCU | 22nt |
| miR-3051 | AAGGAACGUUAAAAACCAUUGU | 22nt |
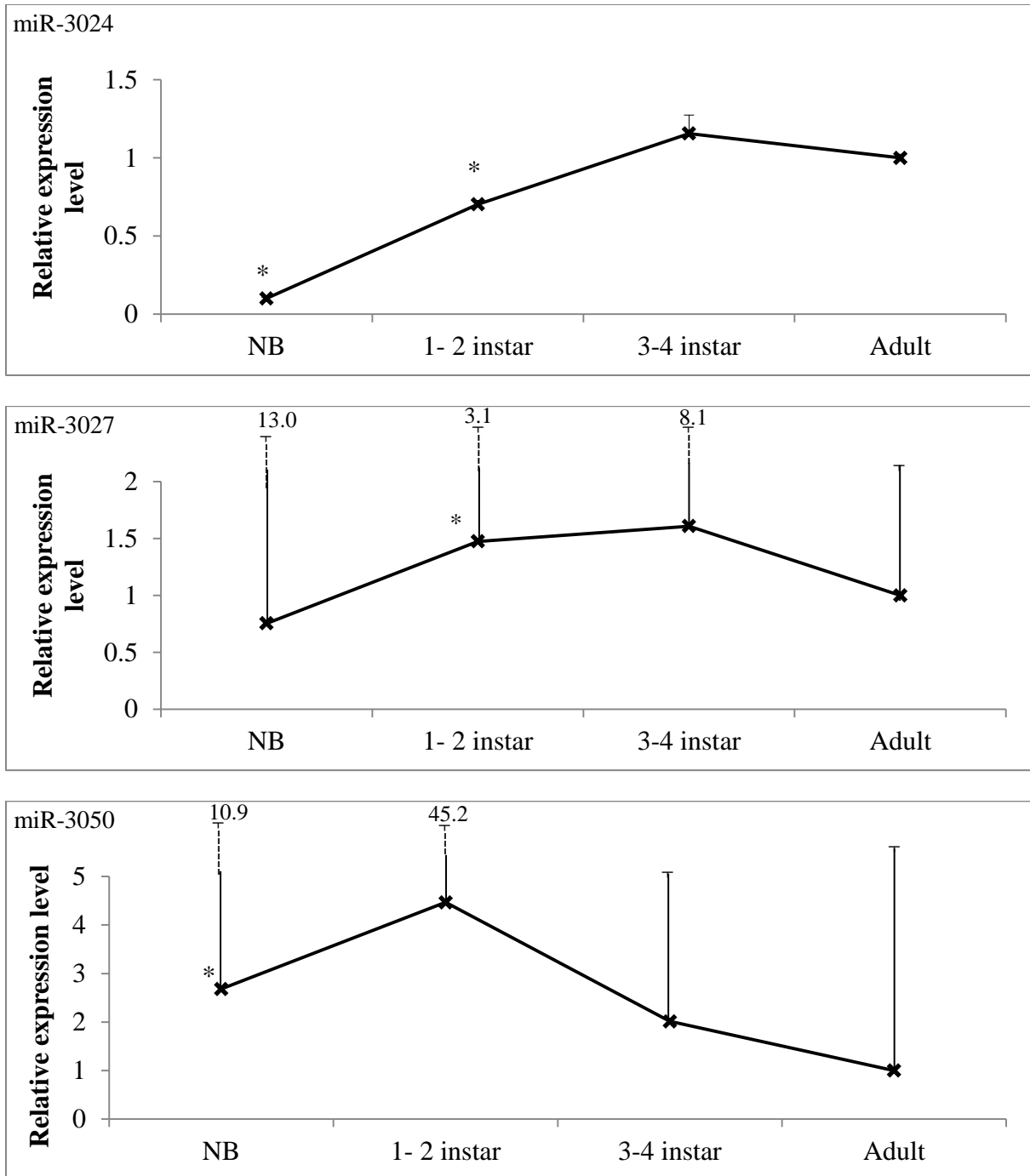
**Figure 3.** Abundance of three aphid-specific miRNAs, miR-3024, miR-3027 and miR-3050 during various developmental stages of *M. persicae*. The experiments were carried out with 3 biological replicates with 5 aphids in each replicate. The relative expression was normalized to ubiquitin and relative to the expression levels in the adults, which was set at 1.0. NB, newborn. Lines represent standard error and (*) indicate expression levels that are significantly different from the adult (Student's t-test, $p < 0.05$).

## CHAPTER 6

## SUMMARY AND CONCLUSIONS

Aphids negatively impact plants of economic importance by feeding and by transmission of plant viruses, and present a challenge to our efforts to increase crop yields. Efforts to identify novel avenues for management of aphids and aphid transmitted viruses include identification of aphid gut binding peptides that interfere with plant virus transmission (1), modification of *Bacillus thuringiensis* toxins for increased effectiveness against aphids (2) and the use of a plant virus coat proteins to deliver neurotoxin into the aphid hemocoel (3). In this dissertation, additional avenues for development of aphid management tools were explored. Next-generation sequencing of aphid transcriptomes and small RNA (sRNA) was used to identify novel viruses infecting aphids and microRNAs (miRNAs) common to multiple species of aphids.

In Chapter 2 of this dissertation, the genome of a novel isolate of Aphid lethal paralysis virus (ALPV) from *A. pisum* was described. The genome of ALPV-Ap was 9,940 nt, which is longer than the genomes of other ALPV isolates. The genome sequence was determined using bioinformatics analysis of sequence data, RACE PCR, molecular cloning and sequencing of clones. The use of any one method would have resulted in acquisition of a shorter genome sequence. Our experience with genome assembly emphasizes the need for multiple approaches for assembly and validation of the genome sequence of novel viruses. For example, the published genome sequence of *Infectious myonecrosis virus* (IMNV) that infects shrimp was shorter at the untranslated regions (UTR) of the virus genome compared to the recently assembled sequence (4). The additional sequence was confirmed to be part of the virus genome (Loy *et al*,

unpublished). Phylogenetic analysis showed that ALPV-Ap is related to the ALPV

isolates derived from *A. mellifera* (5-7), although other isolates of this virus have been

identified from two different species of aphid, *R. padi* and *A. nerii* (8, 9).

Variation between ALPV isolates was more common in the 5' UTR of the virus

genome with high conservation in the protein coding regions. An interesting observation

with the various ALPV isolates is the report of disease symptoms due to ALPV infection.

The first identified strain of ALPV was ALPV-RhP from *R. padi* which was shown to

cause a reduction of aphid populations due to death by paralysis (10). The ALPV-An

strain isolated from *A. nerii* was not pathogenic to *A. nerii* but was highly pathogenic to a

different species of aphid, *M. persicae* (8). The ALPV isolates from *A. mellifera* were

also not associated with obvious disease symptoms.

Several factors may contribute to the presence or absence of symptoms in infected

hosts. The regions with variation (5' UTR) may be involved with the pathogenicity of the

virus. To address this, an ALPV infectious clone could be constructed with different UTR

to test for effects on virus properties in a cell line. The ability of ALPV-Ap to replicate to

a limited extent in the DU182A cell line derived from the spotted cucumber beetle,

*Homalodisca vitripennis* (Chapter 3) provides an additional tool that could potentially be

used for investigation of the different isolates of ALPV. The titer of virus in the host

insect is expected to be a major determinant of pathogenicity. *R. padi* was shown to

exhibit behavioral changes when infected with ALPV-RhP. The aphids moved away from

the food source, perhaps to reduce the virus load acquired by the aphid (10). Differences

in tissue tropism between ALPV isolates may also contribute to differences in

pathogenicity. The ALPV isolate known to cause paralysis in *R. padi* localized to the

160

nervous system (11). Therefore, pathogenicity (paralysis in particular) could be

associated with the ability of the virus to escape the gut barrier and infect the nervous

system of the aphid.

In Chapter 3, the characterization of ALPV-Ap and *A. pisum* host-virus

interaction was described. In an effort to more effectively work with this virus, a screen

to identify an insect cell line that will support replication was carried out. The DU182A

cell line supported limited replication of ALPV-Ap. The primary justification for testing

this particular cell line for ALPV-Ap replication was the discovery of ALPV-like

sequences in corn rootworm transcriptome datasets (Liu *et al.* unpublished). The

DU182A cell line showed initial symptoms of infection with ALPV-Ap but recovered

three days post infection. This could be attributed to a robust immune response by the

cell. Suppressors of RNA silencing have been identified from many dicistroviruses

(Chapter 1) that help the virus to replicate without being silenced by the host RNA

interference-based immune response. Future work with the cell line could involve co-

transfection of a suppressor of RNAi to limit the effectiveness of the RNAi immune

response to further promote replication of ALPV-Ap.

Sequence analysis of ALPV did not identify any region of the virus genome with

known motifs of a silencing suppressor. A potent silencing suppressor is encoded by

CrPV that is associated with production of acute disease. The first isolate of ALPV,

ALPV-RhP caused acute infection with paralysis but no obvious symptoms were

observed for the other isolates. The silencing suppressor of the non-paralytic isolates of

ALPV may differ from that of ALPV-RhP, which could contribute to lower

pathogenicity. The UTR regions could also be involved in silencing. Non-conical

suppressors of silencing such as decoy small RNA have been discovered for other viruses (12, 13). Examination of the silencing suppressors of different ALPV isolates would be of interest to assess if the efficacy of the suppressor correlates directly with pathogenicity, as suggested by previous research (14). Replacement of the DCV suppressor with the *Cricket paralysis virus* (CrPV) suppressor resulted in acute infection and paralysis not normally seen in DCV infections of *Drosophila melanogaster* (14)

ALPV is emerging as a virus that can infect many different host insects similar to another discistrovirus, CrPV(15). This reflects a dichotomy seen in several insect virus families where viruses are either adapted for efficient replication and vertical transmission in one or a few closely related host insects (e.g. RhPV), or adapted to infect multiple species in a given order with low or no vertical transmission in any given species. These two strategies exploit two very different ecological niches contributing to the overall success of the insect viruses.

ALPV-Ap is not vertically transmitted suggesting that the transmission route is entirely horizontal. This complete reliance on horizontal transmission is ideal for an insect virus that might be applied by spray for suppression of aphid populations on the basis that horizontal transmission is likely to be efficient. However, the potential for this virus to infect beneficial species of insects (specifically bees) would need to be examined before ALPV-Ap could be deployed as an insecticide.

Our efforts to identify a cell line that will support replication of ALPV-Ap virus lead to the discovery of a natural virus infection in a hemipteran cell line, GWSS Z15 derived from *H. vitripennis*. This result highlights the need for researchers to be cautious

with cell lines and laboratory organisms used for scientific experiments, particularly for virus-related research.

During the course of the experiments, the ALPV-Ap virus was lost from the *A. pisum* colony which prevented further work on this virus isolate. The most plausible explanation for loss of the virus is a combination of 1) increased temperature in a new incubator that most recently housed the colony triggered ALPV-Ap replication resulting in death of infected aphids, and 2) loss of viability in virus that was stored long term at -20°C. An increase in ALPV-RhP in *R. padi* when the rearing temperature was increased by only 5°C has previously been observed (16). Another possibility is that the aphids developed immunity to the virus and effectively cleared virus infection. However, no virus infection was obtained on microinjection of virus directly into the hemocoel, or by feeding using the same *A. pisum* colony or using newly acquired *A. pisum* (Berkshire Biologics, MA). These negative results suggest that the virus stock used for infection had lost viability, rather than the lack of infection resulting from the aphid immune response.

Following the detection of ALPV negative strand RNA in the honeybee population in Belgium, the *A. mellifera* isolate of ALPV was shown to infect and replicate in *A. mellifera* (6). It remains to be determined if the virus could cause severe pathogenic effects in the bees, but given recent challenges associated with CCD and the fact that bees are already known to be infected by more than 18 viruses (17), caution is required regarding use of this particular virus for aphid suppression. ALPV isolates have also been identified from bat fecal samples (18). These viruses are most likely to have been acquired by the bat when feeding on insects infected with ALPV. In contrast to detection of *A. mellifera* virus negative strand RNA in *A. mellifera*, virus negative strand

has not been detected in bats. The extensive use of next-generation sequencing is likely to uncover more isolates of ALPV and sequence comparisons will facilitate our understanding of virus host range and evolution.

In Chapter 4 of the dissertation, we described the identification and characterization of *Aphis glycines virus* (AGV), a unique insect virus with both insect and plant virus-like properties. The genome sequence of AGV was assembled from the *A. glycines* transcriptome. Assembly of the near full length virus genome was not possible from the small RNA sequencing data. It is unclear if this virus represents the ancestor of plant and insect viruses or is an evolutionarily more recent virus that developed from a recombination event between an insect virus and a plant virus. AGV is 100% vertically transmitted and is highly adapted to the host aphid. Vertical transmission could occur from infection of the developing embryo or from a dose of virus acquired in the birth canal. Localization studies of AGV in *A. glycines* would be needed to determine the exact mechanism of vertical transmission.

The susceptibility of AGV to *A. glycines* RNAi pathway was demonstrated with observation of a 22 nt peak of virus-derived small RNAs (vsRNA). This result demonstrated a key aspect in aphid antiviral immunity showing a functional response from the RNAi pathway. This response was however only observed on co-infection of the host with an ALPV-like virus. Mapping of AGV vsRNAs against the genome of AGV indicated hotspot regions i.e., regions of the virus genome that are more often targeted by the dicer protein for silencing.

In Chapter 5 of this dissertation, the identification of microRNAs (miRNAs) from multiple aphid species was described. A total of 12 miRNAs were further classified as being aphid-specific. These 12 miRNAs were identified in all aphid species analyzed but not previously described in any other organisms. Duplication of miRNA pathway genes has been suggested as a possible means for aphids to generate novel miRNAs that could aid in the evolution and adaptability of these insects in nature (19). The miRNAs detected so far represent a snap-shot of the current miRNA profile of aphids. Three of the aphid-specific miRNAs in *M. persicae* were detected at all developmental stages. The large variation in levels of some miRNAs between biological replicates may represent a relatively recent functionality of the aphid-specific miRNAs with possible roles in regulating multiple mRNA at multiple developmental stages. Improvements in target prediction tools and technology will aid in the identification of targets for aphid-specific miRNAs for a better understanding of their role in the aphid.

All of the knowledge gained from this dissertation provides the foundation for further work on both aphid viruses and aphid miRNAs. The newly identified virus, AGV or the full length clones of ALPV-Ap may provide candidates for development of an infectious clone for use as a virus insecticide or as a vector for gene silencing. Both ALPV-Ap and AGV will present unique challenges for development as insecticides. ALPV-Ap may not be aphid-specific and may infect beneficial insects, specifically *A. mellifera*. Knowledge of the bases of pathogenicity of ALPV-Ap would facilitate development of this virus as an insecticide for aphid control. While AGV is highly adapted to *A. glycines*, it is unknown whether AGV also infects other insects. The difficulty in obtaining intact genomic RNA and protein profiles makes AGV a

particularly challenging virus to work with. Studies of AGV in the absence of AGV-free

*A. glycines* would also be challenging. Laboratory experiments using miRNA mimics and

inhibitors will be needed to demonstrate the importance of the aphid-specific miRNAs

prior to engineering transgenic plants for miRNA-mediated aphid resistance.

**Reference**

1.      Liu S, Sivakumar S, Sparks WO, Miller WA, & Bonning BC (2010) A peptide that binds the pea aphid gut impedes entry of Pea enation mosaic virus into the aphid hemocoel. *Virology* 401(1):107-116.

2.      Chougule NP, *et al.* (2013) Retargeting of the *Bacillus thuringiensis* toxin Cyt2Aa against hemipteran insect pests. *Proc Natl Acad Sci U S A* 110(21):8465-8470.

3.      Bonning BC & Chougule NP (2014) Delivery of intrahemocoelic peptides for insect pest management. *Trends Biotechnol* 32(2):91-98.

4.      Poulos BT, Tang KF, Pantoja CR, Bonami JR, & Lightner DV (2006) Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *J Gen Virol* 87(Pt 4):987-996.

5.      Runckel C, *et al.* (2011) Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, nosema, and crithidia. *PLoS ONE* 6(6):e20656.

6.      Ravoet J, *et al.* (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS One* 8(8):e72443.

7.      Granberg F, *et al.* (2013) Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid lethal paralysis, Israel acute paralysis and Lake sinai viruses. *PLoS One* 8(2):e57459.

8.      Dombrovsky A & Luria N (2013) The Nerium oleander aphid *Aphis nerii* is tolerant to a local isolate of Aphid lethal paralysis virus (ALPV). *Virus Genes* 46(2):354-361.

9.      Van Munster M, *et al.* (2002) Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family Dicistroviridae. *J Gen Virol* 83(Pt 12):3131-3138.

10. Williamson C, Rybicki EP, Kasdorf GF, & Von Wechmar MB (1988) Characterization of a new picorna-like virus isolated from aphids. *J Gen Virol* 69:787-795.

11. Hatfill SJ, Williamson C, Kirby R, & von wechmar MB (1990) Identification and localization of Aphid lethal paralysis virus in thin tissue sections of the *Rhopalosiphum padi* aphid by in situ nucleic acid hybridization. *J Invertebr Pathol* 55(2):265-271.

12. Siu RW*, et al.* (2011) Antiviral RNA interference responses induced by Semliki Forest virus infection of mosquito cells: characterization, origin, and frequency-dependent functions of virus-derived small interfering RNAs. *J Virol* 85(6):2907-2917.

13. Sabin LR*, et al.* (2013) Dicer-2 processes diverse viral RNA species. *PLoS One* 8(2):e55458.

14. Nayak A*, et al.* (2010) Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in Drosophila. *Nat Struct Mol Biol* 17(5):547-554.

15. Plus N, Croizier G, Reinganum C, & Scott PD (1978) Cricket paralysis virus and Drosophila C virus: serological analysis and comparison of capsid polypeptides and host range. *J Invertebr Pathol* 31(3):296-302.

16. Laubscher JM & von Wechmar MB (1992) Influence of Aphid lethal paralysis virus and Rhopalosiphum padi virus on aphid biology at different temperatures. *J Invertebr Pathol* 60:134-140.

17. Genersch E, Evans JD, & Fries I (2010) Honey bee disease overview. *J Invertebr Pathol* 103 Suppl 1:S2-4.

18. Ge X*, et al.* (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86(8):4620-4630.

19. Jaubert-Possamai S*, et al.* (2010) Expansion of the miRNA pathway in the hemipteran insect *Acyrthosiphon pisum*. *Mol Biol Evol* 27(5):979-987.

# ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Bryony C. Bonning for her endless patience, mentoring, encouragement and understanding through my journey in graduate school. She offered invaluable support and guidance in all aspects of my life, in good and bad times. She has been an inspiration and a great role model.

I would also like to thank my committee members, Dr. W. Allen Miller, Dr. Lyric Bartholomay, Dr. Russell Jurenka and Dr. Thomas Baum for providing guidance on different aspects of my dissertation work. Thank you to current and past members of the Bonning lab who have been an important part of my graduate school life providing both practical help and moral support. A very special thank you to my co-authors especially Dr. Sijun Liu for involving me in his research program, tirelessly training me in bioinformatics and constantly providing suggestions for my research projects.

Most importantly, I would like to thank my amazing family. My mom, who taught me that anything is possible if you work for it, my uncles who always encouraged me to be more, my loving grandmothers and my sister who always showed me a different perspective. A special thank you to my host parents, Jon and Carol Reese for their endless support in everything I do and for giving me a sense of family away from home.

**APPENDIX**

**DOUBLE-STRANDED RNA-MEDIATED KNOCKDOWN OF CATHEPSIN**

**GENES IN THE PEA APHID GUT**

Diveena Vijayendran, Nanasaheb P. Chougule and Bryony C. Bonning

Department of Entomology, Iowa State University, Ames, IA, 50011

**Introduction**

RNA interference (RNAi) is an important biological phenomenon in eukaryotes that has been manipulated in insects for functional genomics studies and explored for novel pest management strategies. The pioneering study for RNAi in pest management was carried out in a coleopteran pest, the western corn rootworm, *Diabrotica virgifera virgifera*. The *vacuolar ATPase subunit A* (*V-ATPase*) gene was silenced when double stranded RNA (dsRNA) was orally administered in artificial diet. Feeding of the rootworms on transgenic corn plants engineered to express the same dsRNA resulted in larval stunting and mortality. There was also reduced damage to the plant roots (1). This was the first study that showed the potential use of RNAi for pest management. Many studies have since been carried out to demonstrate gene silencing in pest insects such as that of the *cytochrome P450* gene in cotton bollworm, *Helicoverpa armigera* (2) and the *trehalose phosphate synthase* gene in brown planthopper, *Nilaparvata lugens* (3).

Silencing studies have also been carried out in aphids using feeding or microinjection for the delivery of silencing RNA. Microinjection of short-interfering RNA (siRNA) targeting the Coo2 transcript in pea aphid saliva, *Acyrthosiphon pisum*

resulted in a 2-fold reduction of the target transcript after 24 hours (4). Microinjection of dsRNA targeting a calcium binding protein, calreticulin and cathepsin L resulted in ~40% silencing of target genes by Day 5 (5). Gene silencing by feeding on artificial diet mixed with dsRNA was demonstrated for silencing of a putative aquaporin gene (ApAQP1) (6), the V-ATPase gene in *A. pisum* (7) and five candidate target genes in the grain aphid, *Sitobion avenae* (8). Transgenic plants expressing dsRNA have also been used for gene silencing in aphids. Genes targeted include, *Myzus persicae* Coo2 (MpCoo2), receptor for activated kinase C (Rack-1), *Myzus persicae* serine protease (MpSP) and the *Myzus persicae* hunchback gene (Mphb) in the green peach aphid, *Myzus persicae* (9-11). The carboxylesterase gene (CbE E4) was targeted in *S. avenae* using transgenic wheat, *Triticum aesticum* (12). Target gene knockdowns were observed in the range of 30-80% in these assays.

In this study, we attempted silencing of gut-specific cathepsin genes in *A. pisum*. The study follows from published results of silencing of Cathepsin L in *A. pisum* by dsRNA microinjection (5). Cathepsins are a large family of cysteine protease genes. A subset of cathepsin genes, namely Cathepsins L, B84, B16, B16D, B1874 and B2744 were identified to be preferentially and widely expressed in the gut cells of aphids (13, 14). Aphid gut genes were chosen because they would be ideal candidates for silencing effects in gut cells upon acquisition of silencing dsRNA by feeding. Aphid feeding bioassays were hugely variable with high mortality rates. The cathepsin B16 gene was moderately silenced in the gut of *A. pisum* fed on 0.25 µg/µL dsRNA after 3 days of continuous feeding.

**Materials and Methods**

**Aphid guts dissections and total RNA extraction**

Adult pea aphids, *Acyrthosiphon pisum* were dissected in 1X phosphate buffer saline (PBS). A total of 200 dissected aphid guts were directly placed into TRIzol® reagent (Life Technologies). Aphid guts were gently homogenized in TRIzol® reagent using a pellet pestle and total RNA was extracted from the aphid guts using the TRIzol® reagent according to the manufacturer's protocol with overnight precipitation of RNA in isopropanol at -80°C to maximize the amount of RNA precipitated.

**Reverse transcription and PCR amplification of *A. pisum* cathepsin genes**

Total RNA extracted from *A. pisum* guts was quantified using Nanodrop (Thermo Scientific). Approximately 5 μg of total RNA from *A. pisum* guts was reverse transcribed to cDNA using Superscript III Reverse Transcriptase (Life Technologies). *A. pisum* cathepsin genes (B84, B16, B16D, B2744, B1874 and L) were PCR amplified using the gut cDNA and primers listed in Table 1 without the T7 promoter sequence. PCR reactions (25 μL) were carried out for each target cathepsin gene using the PCR conditions: 94°C for 10 minutes, 30 cycles of 94°C for 30 seconds, 57°C for 30 seconds, 72°C for 2 minutes followed by one cycle of 72°C for 5 minutes. The PCR products were run on a 1% agarose gel, stained with ethidium bromide and imaged using a gel imager (Fotodyne). Amplified PCR products were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

The purified PCR products were then used as templates for PCR amplification of *A. pisum* cathepsin genes with T7 overhangs. Primers that were used are listed in Table 1.

PCR conditions were as described above. The negative control gene, Lac Z was PCR amplified directly from 50 ng of pGEM T-easy (Promega) using the T7 primer listed in Table 1. PCR conditions used: 94°C for 2 minutes, 30 cycles of 94°C for 30 seconds, 53°C for 30 seconds, 72°C for 2 minutes followed by one cycle of 72°C for 5 minutes. An additional negative control gene, GFP was PCR amplified directly from a previously cloned plasmid in the lab, refer to (15). PCR conditions used for GFP amplification : 94°C for 2 minutes, 30 cycles of 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 2 minutes followed by one cycle of 72°C for 5 minutes. PCR products were visualized and purified as previously described. All PCR products were sequenced at the ISU DNA Facility to ensure correct amplification of the target gene.

**Double stranded RNA synthesis**

Double stranded RNA for silencing all genes was produced in-vitro using the MEGAscript RNAi Kit (Life Technologies). Approximately 2 µg of purified PCR product with a T7 overhang was used in each dsRNA synthesis reaction. Reactions were carried out according to the manufacturer's directions.

*A. pisum* **feeding bioassays**

*A. pisum* (3rd and 4th instars) were starved at 4°C for approximately 4 hours before feeding bioassays. The feeding bioassays were carried out using a Parafilm ® membrane feeding method in a petri dish. In-vitro synthesized cathepsin, lac Z or GFP dsRNA was added to aphid complete diet (16) at concentrations of either 0.25 µg/µL or 0.5 µg/µL. A complete diet only control was also included in each bioassay. Aphids were monitored daily for feeding behavior, percent mortality and number of progeny produced. Aphid

mortality data was analyzed using Student's t-test, data is significantly different if p-value <0.05.

## Quantitative RT-PCR amplification of cathepsin gene knockdown

Aphid guts were dissected from *A. pisum* from the diet only control, GFP dsRNA fed and cathepsin B16 dsRNA fed treatment at Day 3. Total RNA was extracted from the guts using TRIzol® (Life Technologies) according to the manufacturer's directions. qRT-PCR was carried out using the iScript One-Step RT-PCR with the SYBR Green kit (Bio-Rad). The reactions were carried out with 10 ng of total gut RNA in a 10 µL reaction using gene-specific primers (Table 1). Ribosomal protein (Rp L7) was used as an internal control gene. All reactions were carried out with duplicate technical replicates. Additional no template negative control reactions were included. qRT-PCR was carried out in the iCycler IQ system (Bio-Rad) using the protocol of 1 cycle of 50°C for 30 minutes and 95°C for 5 minutes, 30 cycles of 95°C for 10 seconds and 53°C for 30 seconds. A dissociation melt curve analysis was carried out with 81 cycles of 15 second increments of 0.5°C starting at 55°C. The data were analyzed using the $\Delta\Delta C_T$ method (17) with data normalized to the internal control gene and the diet only treatment.

## Double stranded RNA stability assay

In-vitro synthesized dsRNA that was stored at -20°C was thawed on ice. dsRNA (0.2 µL) was incubated at room temperature for 3 hours. Equal volumes of the thawed -20°C samples and the room temperature incubated samples were loaded on a 1% agarose gel. The gel staining and viewing as previously described.

**Results**

**Amplification of cathepsin genes from *A. pisum* and dsRNA synthesis**

The annotated sequences of *A. pisum* cathepsin genes were obtained from the National Center for Biotechnology Information (NCBI) sequence database. Primers were designed to amplify ~ 600-700 bases of the target cathepsin genes from the guts of *A. pisum*. The list of gene accession numbers and primer sequences is provided in Table 1. All target cathepsin genes amplified except for cathepsin B2744 and B1874 (Figure 1). PCR amplification using 2 additional primer sets did not change the outcome (data not shown). These cathepsin genes were not included in further assays. The T7 gene-specific primers yielded PCR products at the expected sizes (Figure 2) for the target cathepsin genes, Lac Z and GFP controls. Double stranded RNA (dsRNA) was produced for each target cathepsin and for Lac Z. The dsRNA migrated through the agarose gel slower than the DNA counterpart as observed with the Lac Z sample (Figure 3).

**Impact of dsRNA feeding on *A. pisum* mortality and fecundity**

*A. pisum* fed on 0.5 µg/µL target dsRNA had a higher mortality rate compared to diet-only fed aphids. However, *A. pisum* fed on negative control dsRNA Lac Z also had high mortality (Figure 4). The fecundity of *A. pisum* is affected by feeding on cathepsin dsRNA compared to the diet-only control but was not significantly different when compared to the negative control Lac Z dsRNA treatment using Student's t-test ($p > 0.05$) (Figure 5). Reasoning that Lac Z might be having a non-target silencing effect in *A. pisum*, a new negative control gene was used instead. The dsRNA synthesized from the GFP gene was used for the second set of bioassays.

The mortality of *A. pisum* is higher in the cathepsin B16 dsRNA at 0.25 µg/µL fed aphid sample compared to the diet-only and GFP controls at Day 3 (Figure 6a). The mortality is however not significantly different (p>0.05, Student's t-test). Cathepsin L dsRNA feeding for 5 days at 0.25 µg/µL did not result in significant mortality (Figure 6b).

**Silencing of cathepsin B16 in *A. pisum* gut**

The guts from surviving *A. pisum* in the cathepsin B16 knockdown feeding bioassay (10-15 guts) were dissected out at day 3. qRT-PCR was carried out using total RNA extracted from the gut samples from each treatment with data normalization of B16 mRNA expression to the diet-only fed aphid guts. The internal control gene used for the analysis was Rpl7, the same internal control gene used in a previous cathepsin L knockdown experiment by microinjection (5). The cathepsin B16 mRNA was reduced 30% in *A. pisum* that fed on dsRNA targeting cathepsin B16. There was an increase in cathepsin B16 expression level in aphids that fed on dsRNA targeting GFP (Figure 7).

**Stability of in-vitro synthesized dsRNA at room temperature**

The stability of the cathepsin dsRNA used in the experiments was tested with incubation for 3 hours at room temperature and the original dsRNA thawed from the freezer. Significant degradation of the dsRNA is observed after 3 hours at room temperature (~ 21°C) (Figure 8). This could explain some of the lack of silencing observed in feeding bioassays.

**Discussion**

The results obtained from the silencing of cathepsin genes in *A. pisum* guts were highly variable. The *A. pisum* feeding bioassays present one challenging aspect of the experimental design. The use of negative control dsRNA Lac Z induced high levels of mortality in *A. pisum* although this same negative control has been successfully used in previous aphid knockdown bioassays (5). Additional factors that may play a crucial role in determining the success of gene knockdown studies in aphids include the abundance of the target gene in tissues targeted at various life stages, the activation of the RNAi pathway by the exogenous dsRNA, stability of dsRNA fed to aphids and the mode of delivery.

Although all the core RNAi pathway genes are present in *A. pisum* (18), silencing of aphid genes has proved to be challenging. High concentrations of dsRNA are required for even very low levels of gene silencing (4, 5). Continuous feeding on siRNA in plant phloem reduced the target transcripts in *M. persicae* by 60% (11). There appears to be much that we still do not understand about the RNAi pathway, especially in non-model organisms such as *A. pisum*. Several recent publications provide information about unsuccessful RNAi-based gene knockdown experiments in insects that could facilitate improved experimental design (19, 20).

**References**

1.      Baum JA*, et al.* (2007) Control of coleopteran insect pests through RNA interference. *Nat Biotechnol* 25(11):1322-1326.

2.      Mao YB*, et al.* (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* 25(11):1307-1313.

3.      Chen J*, et al.* (2010) Feeding-based RNA interference of a trehalose phosphate synthase gene in the brown planthopper, *Nilaparvata lugens*. *Insect Mol Biol* 19(6):777-786.

4.      Mutti N, Park Y, Reese J, & Reeck G (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum. *J Insect Sci* 6:1-7.

5.      Jaubert-Possamai S, *et al.* (2007) Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum. *BMC Biotechnol* 7:63.

6.      Shakesby AJ*, et al.* (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem Mol Biol* 39(1):1-10.

7.      Whyard S, Singh AD, & Wong S (2009) Ingested double-stranded RNAs can act as species-specific insecticides. *Insect Biochem Mol Biol* 39(11):824-832.

8.      Zhang M*, et al.* (2013) Identifying potential RNAi targets in grain aphid (Sitobion avenae F.) based on transcriptome profiling of its alimentary canal after feeding on wheat plants. *BMC Genomics* 14(1):560.

9.      Bhatia V, Bhattacharya R, Uniyal PL, Singh R, & Niranjan RS (2012) Host generated siRNAs attenuate expression of serine protease gene in *Myzus persicae*. *PLoS One* 7(10):e46343.

10.     Mao J & Zeng F (2013) Plant-mediated RNAi of a gap gene-enhanced tobacco tolerance against the *Myzus persicae*. *Transgenic Res* 23:145-152.

11.     Pitino M, Coleman AD, Maffei ME, Ridout CJ, & Hogenhout SA (2011) Silencing of Aphid Genes by dsRNA Feeding from Plants. *PLoS One*, United States), Vol 6, p e25709.

12.     Xu L*, et al.* (2013) Silencing of an aphid carboxylesterase gene by use of plant-mediated RNAi impairs Sitobion avenae tolerance of Phoxim ins*ecticides*. *Transgenic Res* (Epub ahead of print).

13.    Deraison C*, et al.* (2004) Cloning and characterization of a gut-specific cathepsin L from the aphid *Aphis gossypii*. *Insect Mol Biol* 13(2):165-177.

14.    Rispe C*, et al.* (2008) Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Mol Biol Evol* 25(1):5-17.

15.    Liu S, Sivakumar S, Sparks WO, Miller WA, & Bonning BC (2010) A peptide that binds the pea aphid gut impedes entry of Pea enation mosaic virus into the aphid hemocoel. *Virology* 401(1):107-116.

16.    Febvay G, Delobel B, & Rahbe Y (1987) Influence of the amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (Homoptera: Aphididae). *Can.J.Zool* 66:2449-2453.

17.    Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408.

18.    Consortium TIAG (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol* 8(2):e1000313.

19.    Terenius O*, et al.* (2011) RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. *J Insect Physiol* 57(2):231-245.

20.    Scott JG*, et al.* (2013) Towards the elements of successful insect RNAi. *J Insect Physiol* 59(12):1212-1221.

**Table 1.** The primers used for amplification of cathepsin genes from *A. pisum* guts and Lac Z from pGEM T-easy vector. The T7 promoter sequence was incorporated at the 5' end of the T7 gene-specific primers

T7 promoter sequence: 5' TAATACGACTCACTATAGGGAGA 3'

| Target | Accession No. | Primer sequence (5' to 3') |
|---|---|---|
| Cathepsin B84 | gi 209863087 | Forward: T7- GTGTGGACAGATATTAAACC<br>Reverse: T7- GTGTCTTTCTGCATGGTTG |
| Cathepsin B16 | gi 161343834 | Forward: T7- CGATTATTTTCCAGCAAAACC<br>Reverse: T7- ATTTTTTTCAATTGGCTTACCG |
| Cathepsin B16D | gi 201023314 | Forward: T7- AATAATTGGCGAAAAGTTGC<br>Reverse: T7- TACATCTGTGATTCGATTCCC |
| Cathepsin B2744 | gi 209863078 | Forward: T7- CGAAGACGATTTACATAAAAC<br>Reverse: T7- GTAGGTAGGTACTACTAAAG |
| Cathepsin B1874 | gi 201023320 | Forward: T7- TTGTCAACCATATACAATCCC<br>Reverse: T7- AATTATTTTATGACATCAAGACC |
| Cathepsin L | gi 209693434 | Forward: T7- ATTATAGTCTACGTTATTGTCTG<br>Reverse: T7- GTCATCCTCGGCTTCATAT |
| Lac Z | pGEM T easy vector | Forward: T7- CGTAATCATGGTCATAGCTG<br>Reverse: T7- TTCATTAATGCAGCTGGCAC |
| GFP | GBP3.1-EGFP vector | Forward: T7- TAAACGGCCACAAGTTCAG<br>Reverse: T7- TGTACAGCTCGTCCATGC |
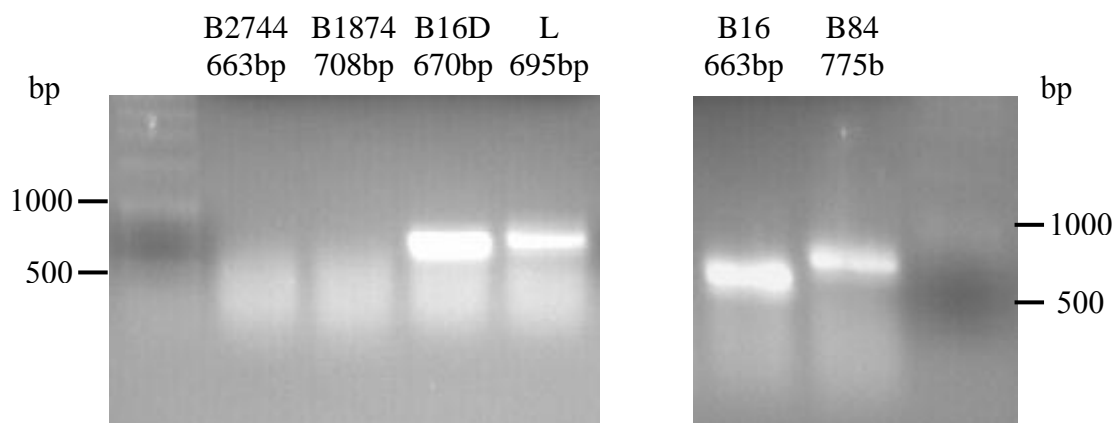
**Figure 1.** RT-PCR amplification of cathepsin genes from *A. pisum* gut total RNA using gene-specific primers. Amplification of cathepsin genes B2744 and B1874 was unsuccessful. All other target cathepsin genes amplified and were visualized at the expected sizes.
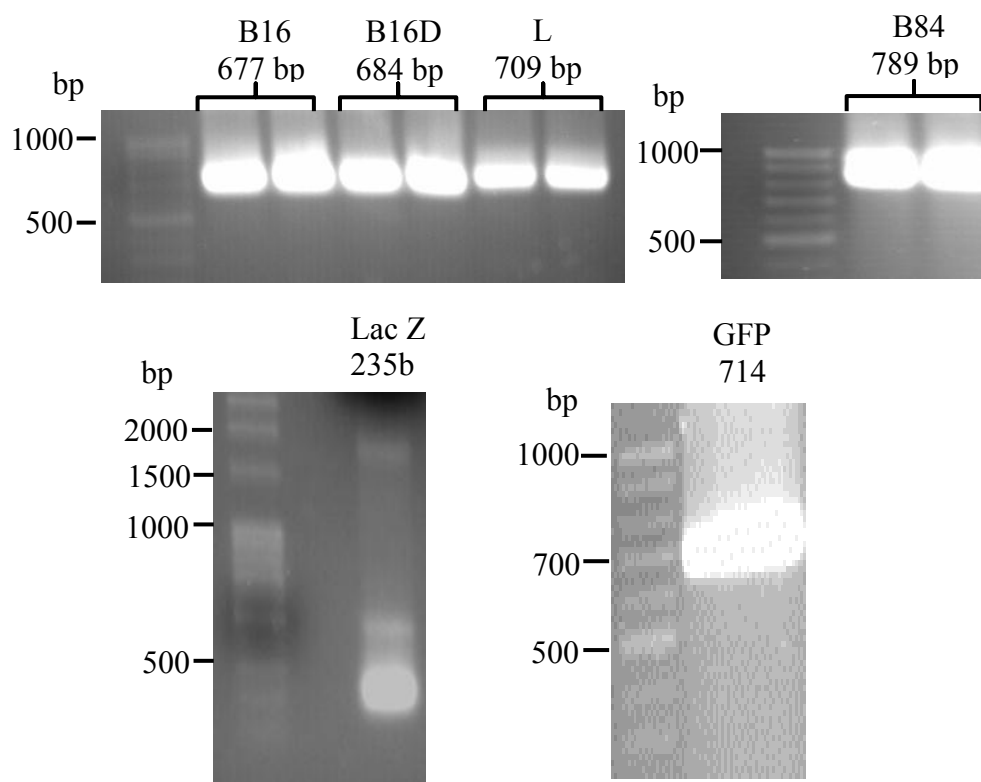
**Figure 2.** PCR amplification of cathepsin genes, Lac Z and GFP with T7 promoter sequence overhangs using T7-gene-specific primers from purified PCR products.
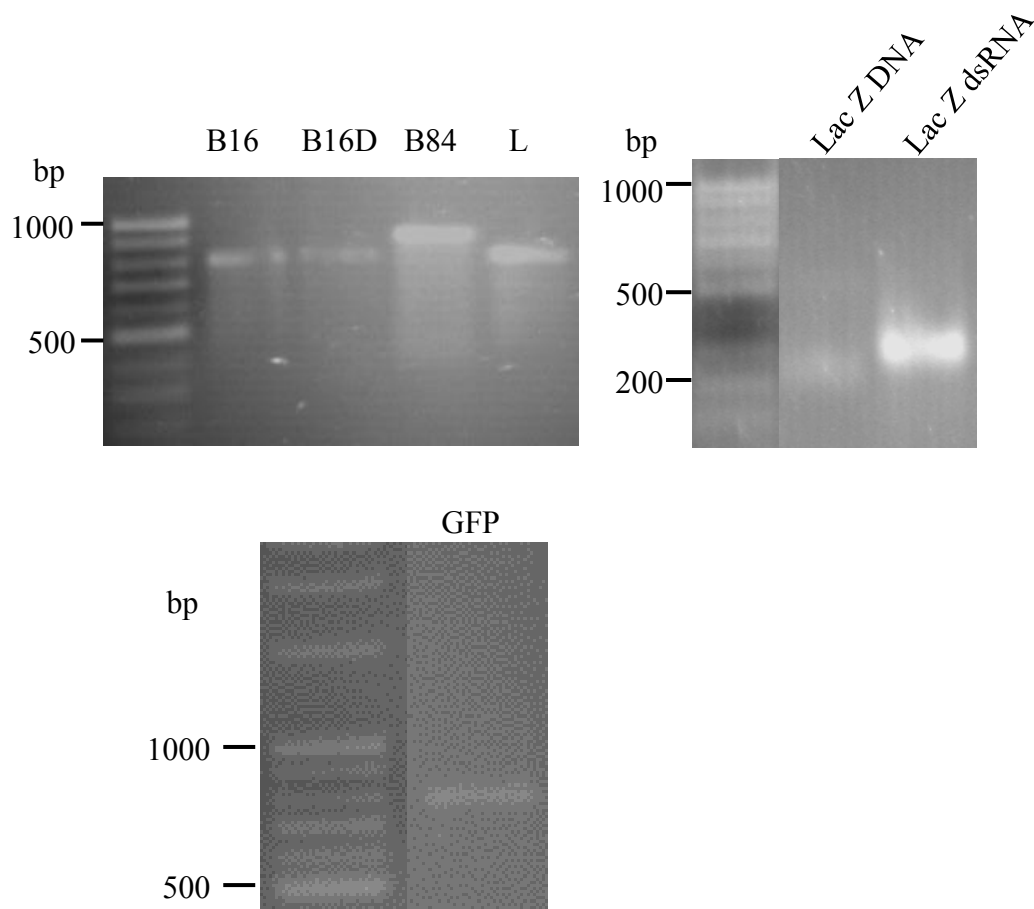
**Figure 3.** In-vitro synthesized dsRNA for *A. pisum* cathepsins B16, B16D, B84, L and the negative controls Lac Z and GFP. The dsRNA appears to run at a higher size than expected. Molecular size comparison with Lac Z PCR product and Lac Z dsRNA shows that dsRNA migrates more slowly in the gel.
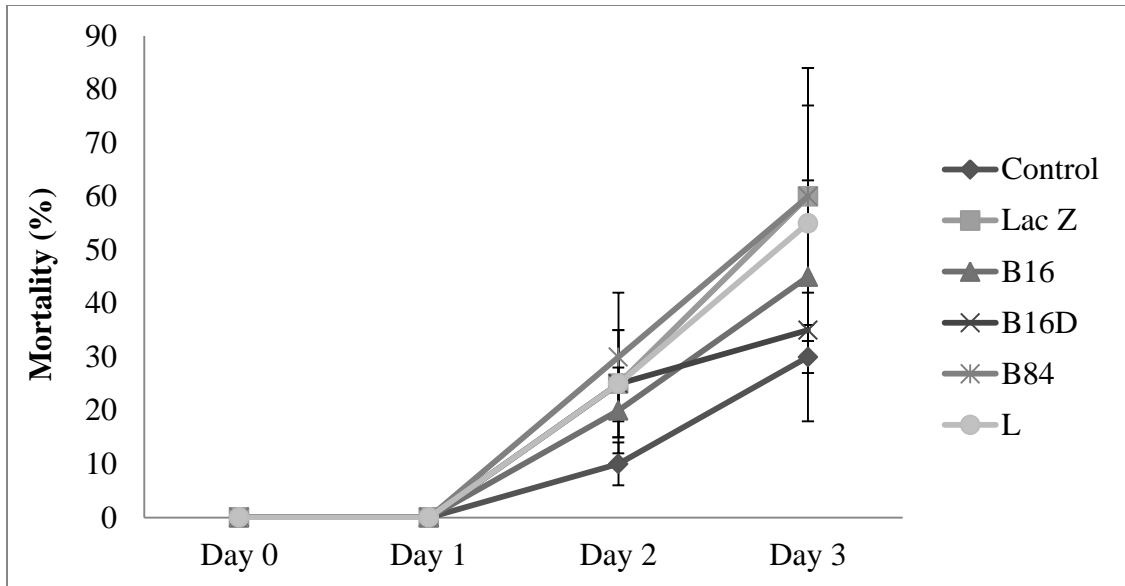
**Figure 4.** Mortality of *A. pisum* fed on artificial diet containing 0.5 µg/µL dsRNA against target genes, cathepsin B16, B16D, B84 and L. Control represents aphid fed on complete diet only. Lac Z dsRNA served as a negative control. 20 adult *A. pisum* were used for each treatment (2 technical repeats of 10 aphids per feeding plate). Mortality data was recorded daily for 3 days. No significant differences was observed for cathepsin dsRNA treatment compared to controls, $p>0.05$ (Student's t-test).
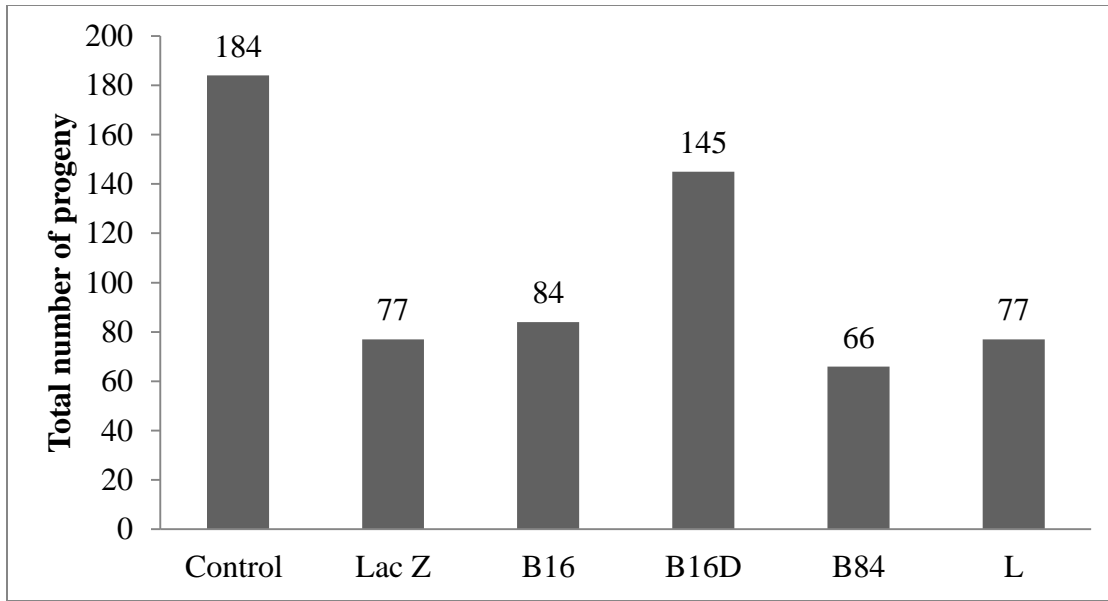
**Figure 5.** Fecundity of *A. pisum* fed on artificial diet containing 0.5 µg/µL dsRNA against target genes. Control represents aphids fed on complete diet only. Lac Z dsRNA served as a negative control. Adult *A. pisum* (20) were used for each treatment (2 technical repeats of 10 aphids per feeding plate). Newborn aphids were counted daily for 3 days then transferred out of the feeding dish.
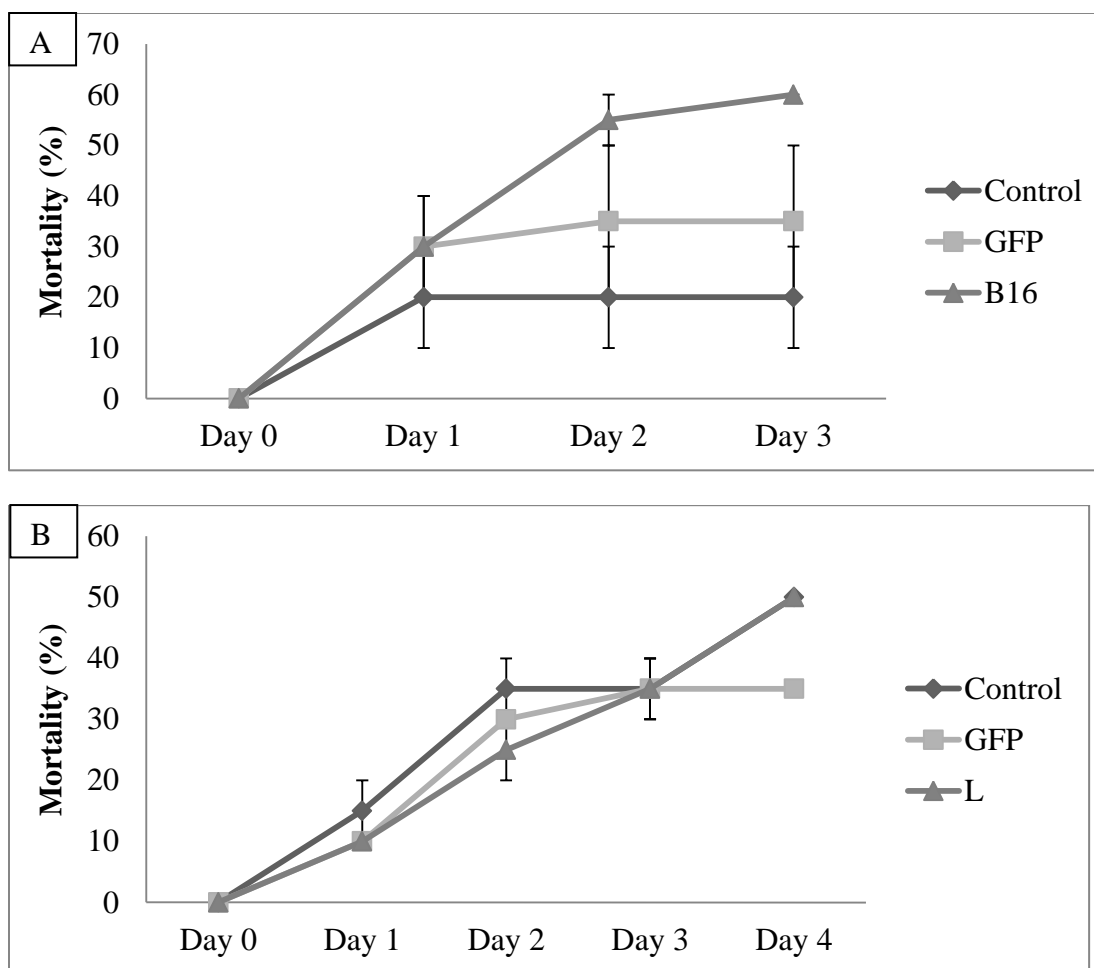
**Figure 6.** Mortality of *A. pisum* fed on artificial diet containing 0.25 µg/µL dsRNA against Cathepsin B16 (A) and Cathepsin L (B). Control represents aphid fed on complete diet only. GFP dsRNA served as a negative control. 30 *A. pisum* between 3$^{rd}$ and 4$^{th}$ instar were used for each treatment (2 technical repeats of 10 aphids per feeding plate). Mortality data was recorded daily for 3 days. No significant differences was observed for both B16 and L dsRNA treatment compared to the controls, p>0.05 (Student's t-test).

185

**Table 2.** Primers used in qRT-PCR analysis of cathepsin B16 gene silencing in *A. pisum* guts at Day 3 of feeding on dsRNA at 0.25 µg/µL.

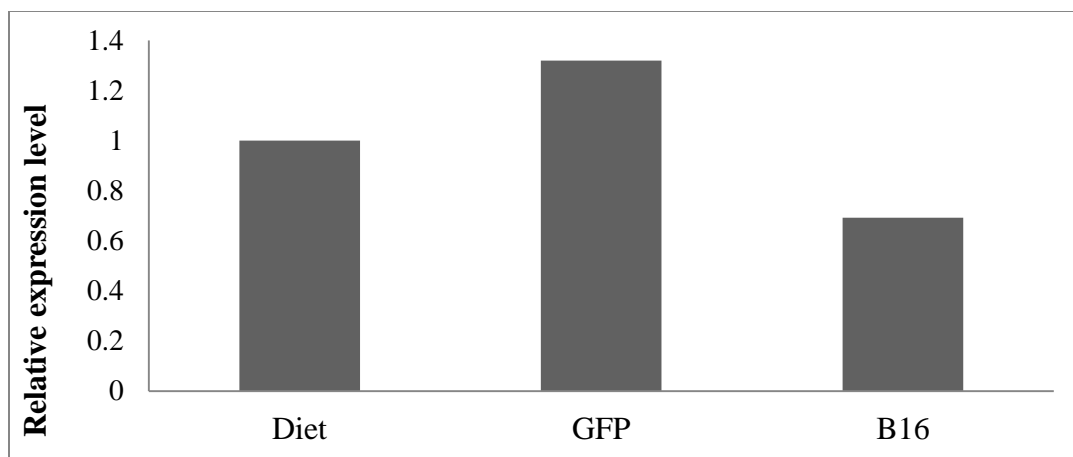| Target | Accession No. | Primer sequence (5' to 3') |
|---|---|---|
| Cathepsin B16 | gi 161343834 | Forward: ACAATAACGGCTATATCCCGAGGACA |
| | | Reverse: GGCTTTGATCGGGTATCCTCCG |
| RpL7 | gi 209571476 | Forward: CGCAAAGCCCGTACAGCATTCA |
| | | Reverse: GGAGCTACTTGGTTCACACCACGA |

**Figure 7.** The relative expression level of cathepsin B16 gene in guts of *A. pisum* fed 3 days on GFP dsRNA and B16 dsRNA at 0.25 µg/µL concentration. The qRT-PCR data were normalized to cathepsin B16 expression in gut samples from diet only fed *A. pisum.*
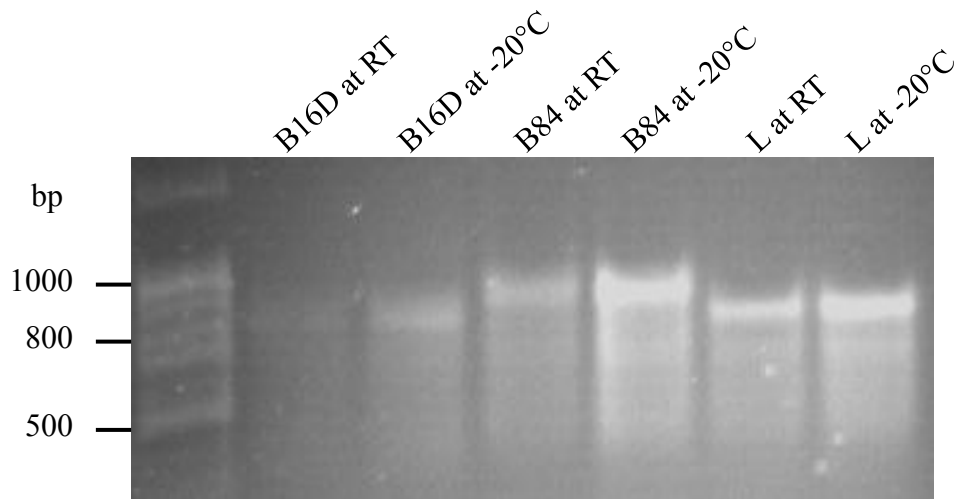
**Figure 8.** Stability of in-vitro synthesized dsRNA directly thawed from stored samples at -20°C versus dsRNA that has been incubated at room temperature for 3 hours. Equal volumes of the same sample aliquot were loaded on a 1% agarose gel for each assayed cathepsin.