

ANALYSIS OF EST'S ENCODING PEA APHID *ACYRTHOSIPHON PISUM* C002 & THE
EFFECT OF ARMET TRANSCRIPT KNOCKDOWN IN *TRIBOLIUM CASTANEUM*

by

MATTHEW C. HEERMAN

B.S., Kansas State University, 2007

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Biochemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
Dr. Gerald Reeck

Copyright

MATTHEW CHARLES HEERMAN

2012

Abstract

Aphids mount a remarkable salivary secretion to overcome plant host defenses. Our group has previously reported a gene unique to aphids enriched in the salivary glands of the pea aphid *A. pisum*, C002, which is required for successful feeding on its host plant *Vicia faba*. Here I present an analysis of genetic variation within the available EST data for C002 in pea aphids. From 596 total ESTs, 332 are full-length, and segregate into 8 validated haplotypes based on the criteria I set in place to assess the quality of EST data. Additionally, Armet, is a putative multi-functional gene implicated as a neurotrophic factor during development, and as a part of the unfolded protein response during stress. I employ RNA interference in the model organism *T. castaneum* to determine the effect of transcript knockdown during development from early instar larval stages, through pupation, and its effect on adult emergence. I report that knockdown of Armet transcript significantly hinders the ability for beetles to emerge from the pupae.

Table of Contents

List of Figures	v
List of Tables	vi
Chapter 1 - An Analysis of Pea Aphid <i>Acyrtosiphon pisum</i> C002 ESTs	1
Introduction.....	1
Aphid Saliva & Effector Molecules.....	1
Materials & Methods	3
C002 EST Blast Search.....	3
EST Processing.....	4
EST Library Origins	4
Multiple Sequence Alignment and EST Analysis.....	5
Results.....	6
Goal.....	6
Analysis of C002 ESTs from <i>A. pisum</i>	6
Chapter 2 - Effect of Armet Transcript Knockdown in <i>Tribolium castaneum</i>	35
Introduction.....	35
Armet Etymology.....	35
Armet in Endoplasmic Reticulum Stress (ER Stress) & Unfolded Protein Response (UPR)	35
Structure of Armet	36
Armet and Arthropod Development	37
Material & Methods.....	38
Synthesis of dsRNA.....	38
Injection	38
RNA Extraction	39
qRT-RT-PCR	39
Results.....	40
Discussion.....	42
Conclusion	43
References.....	60

List of Figures

Figure 1.1 Nucleotide Alignment of Pea Aphid C002 Sequences.....	9
Figure 1.2 Protein Alignment of Pea Aphid C002 Sequences.....	12
Figure 1.3 Nucleotide alignment of 25 unique ESTs of Pea Aphid C002.....	13
Figure 1.4 Protein alignment of 25 unique ESTs of Pea Aphid C002.....	25
Figure 1.5 Nucleotide Alignment of Aphid C002 Sequences.....	29
Figure 1.6 Alignment of Aphid C002 Amino Acid Sequences.....	31
Figure 2.1 Survival curves after injection of dsVER RNA or dsArmet RNA (#1).....	44
Figure 2.2 Number of insects in various developmental stages after dsVER RNA injections (#1)	45
Figure 2.3 Number of insects in various developmental stages after dsArmet RNA injection (#1)	46
Figure 2.4 Survival curves after injection of dsVER RNA or dsArmet RNA (#2).....	47
Figure 2.5 Number of insects in various developmental stages after dsVer RNA injection (#2). 48	48
Figure 2.6 Number of insects in various developmental stages after dsArmet RNA injection (#2)	49
Figure 2.7 Graph representing live tribolium injected with either dsVER RNA or dsArmet RNA (#3).....	50
Figure 2.8 Number of insects in various developmental stages after dsVER RNA injection (#3)51	51
Figure 2.9 Number of insects in various developmental stages after dsArmet RNA injection (#3)	52
Figure 2.10 Larvae injected with dsVER RNA or dsArmet RNA.....	53
Figure 2.11 dsArmet RNA injected larvae.....	54
Figure 2.12 Armet transcript levels at several stages of development.....	55
Figure 2.13 Armet transcript levels after injection with dsVER RNA or dsArmet RNA.....	56
Figure 2.14 Alignment of Armet amino acid sequences from Homo sapiens, Mus musculus, Drosophila melanogaster, Acyrthosiphon pisum, and Tribolium castaneum.....	57
Figure 2.15 Alignment of Armet exons against the genomic sequence of <i>T. castaneum</i>	58

List of Tables

Table 1.1 Pea Aphid EST Analysis.....	32
Table 1.2 Table of all ESTs included within each validated haplotype.....	34

Chapter 1 - An Analysis of Pea Aphid *Acyrtosiphon pisum* C002

ESTs

Introduction

Aphid Saliva & Effector Molecules

Aphids, an Aphididae family of phytophagous hemipterans, are piercing-sucking organisms that attain their nutrients by accessing the epidermis and sieve element of host plants while concurrently forming a rigid salivary stylet sheath. Finally, the insect is able to ingest the nutrient rich plant phloem sap (Miles 1998). The defense response of the host plant is characterized by a calcium dependent activation of coagulating factors, and a group of unknown responses. The coagulating factors released, in the case of calcium dependence, lead to the occlusion of wounds sustained by the plant (Vos et al. 2009). Microarray analysis of *Arabidopsis thaliana* plants un-infested and infested with green peach aphids *Myzus persicae* demonstrated that plants fed on by aphids indeed responded to aphid feeding (Vos et al. 2009). Also, *A. thaliana* plants over-expressing coagulation factor AtPP2-A1 are refractory to feeding by green peach aphid (Zhang et al. 2011). Additionally, electrical penetration graphs (EPGs) and interaction of vetch aphid *Megoura viciae* saliva with plant forisomes *in vitro* demonstrated that aphid saliva is required to overcome this biochemical barrier (Tjallingii 2005; Will et al. 2007). Within the literature these plant defense suppressing molecules have been deemed avirulence factors. Since, the individual components of arthropod saliva that overcome host defenses have

collectively termed as effector molecules (Hogenhout et al. 2009). Small secreted salivary gland proteins (SSSGPs) is another term in the literature used to describe small salivary secreted protein molecules (Chen et al. 2010).

In the case of phytophagous insects salivary secreted protein molecules are thought to mimic plant defense signaling molecules, thereby modulating the ability of the host plant to mount a response (Bos et al. 2010). Genomic and proteomic techniques have been employed to find candidate effector molecules for further characterization. EST libraries derived from the salivary glands of pea aphid *A. pisum* and green peach aphids were probed to find to find a multitude of putatively secreted salivary enriched proteins (Bos et al. 2010; Carolan et al. 2011). An analysis of pea aphid salivary glands using both a transcriptomic and proteomic approach found more than 300 genes encoding secretion signals. Interestingly nearly half of these genes have no sequence similarity to anything outside of aphids, and some have been directly observed in aphid saliva (Carolan et al. 2011). Collecting saliva in an artificial diet, concentrating, and analyzing with mass spectrometry yielded an active glucose oxidase (GOX) protein in green peach aphid (Harmel et al. 2008). The same group conducted in pea aphids observed a M1 and M2 metalloprotease, GMC oxidoreductase, and regucalcin; genes that do not necessarily fit into the category of SSSGPs (Carolan et al. 2009). The study found an additional 5 proteins that showed no sequence similarity to proteins in databases (Carolan et al. 2009).

To date, only one potential effector molecule, C002, has RNAi evidence linking it to being required for an aphid to feed on its host plant. Injections of dsRNA encoding the salivary enriched transcript C002 abrogated pea aphids' ability to feed on its host plant but not on an artificial diet. Additionally, EPGs demonstrated that injected aphids were unable to establish and maintain feeding host plants (Mutti et al. 2006; Mutti et al. 2008). Recently, transiently

expressed dsRNA molecules designed to interfere with *M. persicae* C002 were transiently expressed within the leaves of *A. thaliana* and *Nicotiana benthamiana*, causing a small decrease in individual aphid nymph production (Pitino et al. 2010).

An emerging theme in the field of effector molecules and SSSGPs is one of high coding region variability that is under intense positive selection (Reviewed in Hogenhout et al 2009). In the gall midge there is a peculiar conservation of non-coding regions in SSSGPs, while coding regions contain high variability with a marked positive selection for non-synonymous changes over synonymous (Chen et al 2010). Within aphids, another group demonstrated a high number of positively selected for non-synonymous changes compared to synonymous in C002 across multiple species of aphids. Their evidence suggests that gene duplication and high GC content may attribute to so called fast evolving genes (Ollivier et al 2010).

Here, with C002, I present a pipe-line method for extracting ESTs from the database at NCBI, restricting ESTs to full length coding region, and analyzing ESTs for genetic variance. I report 8 validated haplotypes of C002, and remarkably, no silent mutations within the 8 EST dataset. The results indicate a high level of positive adaptive selection for C002.

Materials & Methods

C002 EST Blast Search

Blast searches were conducted using the National Center for Biotechnology Information BLAST (Altschul 1990) program (blast.ncbi.nlm.gov). Blastn was employed using an EST encoding full length pea aphid c002 (Accession: HSO93729) as query. The search was against

the Expressed Sequence Tag (EST) library. Sequence searches were restricted to *A. pisum* (taxid: 7029). Max Target Sequences was set to 20,000 to return the maximum number of c002 sequences. Standard penalties were used to assess nucleotide identities and for gaps in sequences. The resulting ESTs were downloaded in batch FASTA format for further processing. Discontiguous blast was performed with the same EST against all ESTs excluding all aphid species.

EST Processing

ESTs were processed with the EMBOSS program GetORF (emboss.sourceforge.net) to remove sequences that were not of full length, and to trim the 5' UTR and 3' UTR to produce only full length coding region sequences. GetORF was set to show 0 flanking nucleotides with respect to the coding region. Additionally, GetORF was set to find full length coding regions between 625 and 725 bp, and report sequences in batch FASTA format.

EST Library Origins

ESTs obtained from blast searches originated from three libraries. The Kirk Pappan library, designated P, was created from adult parthenogenetic female pea aphid salivary glands. This EST library was constructed at Kansas State University in the laboratory of Dr. Gerald Reeck using a non-clonal population of insects. Library P was constructed using the SMART cDNA library construction kit (Clontech, Mountain View, CA). This kit employs the M-MLV reverse transcriptase which is reported to create an error once in 30,000 nucleotides

(www.promega.com). The Nancy Moran library, M, comes from the whole bodies of the pea aphids at all developmental stages and both the winged and wingless phenotype. M was constructed at University of Arizona. The Wayne Hunter library, H, comes from whole bodies of pea aphids at all developmental stages but wingless. A fourth library created by Navdeep Mutti, DV, was created at Kansas State University in the laboratory of Gerald Reeck using the P library protocol. However, this library, as deposited at NCBI, did not produce any full-length blast hits for C002; partial hits were observed but did not cover the entire coding region.

Multiple Sequence Alignment and EST Analysis

Full length ESTs were divided into groups of 25 in order of BLAST score, and Multiple Sequence ClustalW (Thompson 1994) was used to perform alignments on the groups of 25 ESTs. Alignments were analyzed individually, and finally collectively, using the program BioEdit (Hall 1999). ESTs represented 2 or more times were selected for the final analysis. The fourth highest blast score group of 25 ESTs contained only unique sequences or singletons. These 25 singletons were analyzed separately following the same protocol used for the validated haplotypes.

Results

Goal

To explore the evolutionary pressure being applied to the gene encoding *A. pisum* C002 secreted salivary protein, an analysis of the available EST sequences was performed.

Analysis of C002 ESTs from A. pisum

A search for C002 ESTs at <http://www.ncbi.nlm.nih.gov/> using Blastn restricted to *A. pisum* sequences returned 596 ESTs. Flanking 5' & 3' UTR sequences were removed from the coding region of C002 ESTs using the EMBOSS program GetORF (<http://mobyli.e-rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=getorf>). Additionally, GetORF was used to remove sequences that were not full length, resulting in 332 full length C002 coding region ESTs. Alignments of the remaining 332 sequences were performed using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), and were visualized using BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). For an EST to be established as error-free I required that each such sequence, represented by at least 2 identical ESTs, is taken to be error-free. I call these genetic variants or, alternatively, validated haplotypes. This resulted in 8 validated haplotypes of the C002 transcript. Alignments of the 8 variant sequences at the nucleotide and protein levels are shown in Figure 1 and Figure 2 respectively. At every position where a change in nucleotide sequence is observed, a corresponding change in protein sequence occurs. Additionally, alignments of *A. pisum*, *A. gossypii*, *M. persicae*, *T. citricida* C002 at the

nucleotide and protein levels are provided in Figure 5 and Figure 6 respectively. The accession number for the most prevalent haplotype was chosen for the alignment (EX644002_1). The position specific changes in sequence are compared to the changes observed within *A. pisum* at both the nucleotide and protein levels.

25 ESTs of unique sequence, taken essentially randomly from the singleton ESTs, are aligned at the nucleotide and protein levels in Figure 3 and Figure 4 respectively. These are all very similar in sequence to the 8 validated haplotypes. At the nucleotide level it was determined that there were 24 mismatches among the 25 unique ESTs, corresponding to 14 non-synonymous changes and 10 synonymous.

Representative EST counts for each of the 8 validated haplotypes, nucleotide changes, residue changes, and EST library sources are summarized in Table 1. The frequency values represent the frequency of each validated haplotype used in this study. The ESTs selected in this study were obtained from three different EST libraries. A fourth library, DV, was included but returned no full-length ESTs, and was therefore excluded from the analysis. All DV sequences were truncated at the N-terminus, possibly due to an error in processing of the library sequences. The salivary gland derived EST library designated as 'P' provided the majority of ESTs and was represented in 7 of the 8 genetic variants. The two whole body EST libraries designated H and M provided fewer ESTs and was represented two and one times respectively. The 9 genetic validated haplotypes are comprised collectively, of 94 of the 332 full-length ESTs obtained from GETORF. Of the 8 genetic variants 2 predominate: haplotype 3 (46%) and haplotype 7 (30%). Haplotype 3 was found in both the M and P libraries, while haplotype 7 was found in the H and P libraries. The remaining 7 haplotypes range between 2% to 7% and are found only in the P library, except haplotype 5 which was found only in the H library. The two predominant genetic

variants differ at 4 amino acid residues (39, 71, 82, and 141), and one variant, haplotype 5 contains residues from both predominant variants. Two variants haplotype 4 and 6 are predicted to be 7 amino acid residues longer at the C-terminus compared with the other variants.

Haplotype 4 shows greater similarity to haplotype 3, while haplotype 6 shows greater similarity to haplotype 7, and are only found in the P library. Remarkably, there are no silent mutations observed in this analysis. Each predicted nucleic acid change results in an amino acid residue change.

Figure 1.1 Nucleotide Alignment of Pea Aphid C002 Sequences.

Alignment of 8 C002 validated haplotypes at the nucleotide level. All EST used are accounted for by accession number in Table 2. Differences in nucleotide sequence are presented in bold face font. Sequences differences are compared across different aphid species nucleotide changes shown in Figure 5.

```
6      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
8      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
7      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
2      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
5      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
      *****

6      GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTTTCGAA 120
8      GTTAGATGCGATTGGTCTGCCACTGAACCGTACGATGAGCAGGAAGAAGCGTCTTTCGAA 120
7      GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTTTCGAA 120
1      GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTCGAA 120
2      GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTCGAA 120
5      GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTCGAA 120
      *****

6      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
8      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
7      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
1      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
2      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
5      TTACCGATGGAGCACCGTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
      *****

6      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
8      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
7      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
1      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
2      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
5      AGCAACCAGGAGGCTATGCAGCTGATGGAAATAACGTTTAATACAGGTAAGGAATTAGGC 240
      *****
```

6 TCC**A**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
8 TCC**A**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
7 TCC**A**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
1 TCC**C**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
2 TCC**C**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
5 TCC**A**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTAACTTCGTCGATGTGATGGCC 300
*** *****

6 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
8 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
7 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
1 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
2 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
5 ACCAACCAGAACGCCATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360

6 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
8 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
7 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
2 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
5 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420

6 **A**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
8 **A**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
7 **A**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
1 **G**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
2 **G**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
5 **G**CGAAAAC**T**CCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480

6 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
8 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
7 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
2 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
5 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540

6 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
8 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
7 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
2 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
5 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600

6 CCCGCCAAGATGGCTATGGGAGCCATAAATAAGTTT-GGAAGTTTCCTTCG**A**CGTTTTTA 659
8 CCCGCCAAGATGGCTATGGGAGCCATAAATAAGTTT-GGAAGTTTCCTTCG**A**CGTTTT-- 657
7 CCCGCCAAGATGGCTATGGGAGCCATAAATAAGTTT-GGAAGTTTCCTTCG**A**CGTTTT-- 657
1 CCCGCCAAGATGGCTATGGGAGCCATAAATAAGTTT-GGAAGTTTCCTTCG**G**CGTTTT-- 657
2 CCCGCCAAGATGGCTATGGGAGCCATAAATAAGTTT**T**GGAAGTTTCCTTCG**A**CGTTTTTA 660
5 CCC**A**CCAAGATGGCTATGGGAGCCATAAATAAGTTT-GGAAGTTTCCTTCG**A**CGTTTT-- 657
*** *****

6 TAAGCGCGTCCATACAGAC 678
8 -----
7 -----
1 -----
2 -----
5 -----

Figure 1.2 Protein Alignment of Pea Aphid C002 Sequences.

Alignment of 8 validated haplotypes variants at the protein level named after single EST representatives (See Table 2). Amino acid residue changes in protein sequences are represented in bold face font. Changes in amino acid sequences were compared to species specific changes provided in Figure 6.

```

6      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASFELPMEHRQCDEYKSKIWDKAF 60
8      MGSYKLYVAVMAIAIAVAVQEVRCDSTEEPYDEQEEASFELPMEHRQCDEYKSKIWDKAF 60
7      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASFELPMEHRQCDEYKSKIWDKAF 60
5      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
2      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
      *****:*****.*****.*****

6      SNQEAMQLMEITFNTGKELGSNEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
8      SNQEAMQLMEITFNTGKELGSNEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
7      SNQEAMQLMEITFNTGKELGSNEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
5      SNQEAMQLMELTFNTGKELGSNEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
1      SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
2      SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
      *****:*****:*****

6      REVDTTSNKFKETKEVFERITKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
8      REVDTTSNKFKETKEVFERITKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
7      REVDTTSNKFKETKEVFERITKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
5      REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
1      REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
2      REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLEFKVVKAFEG L 180
      *****:*****

6      IKPSENEELVKQRLKRLITNAPAKMAMGAINKFGSFLRRFYKRVHTD 226
8      IKPSENEELVKQRLKRLITNAPAKMAMGAINKFGSFLRRF----- 219
7      IKPSENEELVKQRLKRLITNAPAKMAMGAINKFGSFLRRF----- 219
5      IKPSENEELVKQRLKRLITNAPTKMAMGAINKFGSFLRRF----- 219
1      IKPSENEELVKQRLKRLITNAPAKMAMGAINKFGSFLRRF----- 219
2      IKPSENEELVKQRLKRLITNAPAKMAMGAINKFWKFPSTFL----- 220
      *****:*****.*
  
```

Figure 1.3 Nucleotide alignment of 25 unique ESTs of Pea Aphid C002

25 unique ESTs (by accession number) are aligned at the nucleotide level. Differences are marked in bold face.

```
FP889223_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS094889_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS093815_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092939_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092790_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS093201_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS094575_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092845_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS096096_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS095744_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS093225_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS094600_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092884_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS096119_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS093801_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092279_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092907_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092773_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS095838_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS092820_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS093195_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS094993_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGTTGTAGTACAGGAA 60
HS095235_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS095505_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
HS095664_1      ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGCCATAGCTGTAGTACAGGAA 60
*****
```

FP889223_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS094889_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS093815_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092939_1 GTTAGATGCGATTGGTCTGCC**A**CTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092790_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS093201_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS094575_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092845_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS096096_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGT**A**TGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS095744_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS093225_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS094600_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092884_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS096119_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS093801_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092279_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092907_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092773_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS095838_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS092820_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS093195_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS094993_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS095235_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS095505_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120
 HS095664_1 GTTAGATGCGATTGGTCTGCCGCTGAACCGTACGATGAGCAGGAAGAAGCGTCTGTTCGAA 120

FP889223_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS094889_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS093815_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092939_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092790_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS093201_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS094575_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092845_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS096096_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS095744_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS093225_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS094600_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092884_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS096119_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS093801_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092279_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092907_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092773_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS095838_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS092820_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS093195_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS094993_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS095235_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS095505_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180
HS095664_1 TTACCGATGGAGCACCGTCAGTGCGATGAATACAAATCGAAGATCTGGGACAAAGCATT 180

FP889223_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS094889_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS093815_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092939_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092790_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS093201_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS094575_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092845_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS096096_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS095744_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS093225_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS094600_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092884_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS096119_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS093801_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092279_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092907_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092773_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS095838_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS092820_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS093195_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS094993_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS095235_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS095505_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240
HS095664_1 AGCAACCAGGAGGCTATGCAGCTGATGGAACCTAACGTTTAATACAGGTAAGGAATTAGGC 240

FP889223_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS094889_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS093815_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092939_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092790_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS093201_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS094575_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092845_1 CCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS096096_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS095744_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS093225_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS094600_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092884_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS096119_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS093801_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092279_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092907_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092773_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS095838_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS092820_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS093195_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS094993_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS095235_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS095505_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300
HS095664_1 TCCCACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTCGATGTGATGGCC 300

FP889223_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS094889_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS093815_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092939_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092790_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS093201_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS094575_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092845_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS096096_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS095744_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS093225_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS094600_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092884_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS096119_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS093801_1 ACCAACCAGAACGCCCTATTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092279_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092907_1 ACCAACCAGAACGCCCATTTGCTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092773_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS095838_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS092820_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS093195_1 ACCAACCAGAACGCCCCTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS094993_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS095235_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS095505_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
HS095664_1 ACCAACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATA 360
***** ** *****

FP889223_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS094889_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS093815_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092939_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092790_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGGTTTTCGAACGCATC 420
HS093201_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS094575_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092845_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS096096_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS095744_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS093225_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS094600_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092884_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS096119_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS093801_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092279_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092907_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092773_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS095838_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS092820_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS093195_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS094993_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS095235_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS095505_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420
HS095664_1 AGAGAGGTGGACACGACGTCCAACAAATTCAAAGAGACGAAGGAGGTTTTCGAACGCATC 420

FP889223_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS094889_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS093815_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092939_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092790_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS093201_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS094575_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092845_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS096096_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS095744_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS093225_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS094600_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092884_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS096119_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS093801_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092279_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092907_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092773_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS095838_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS092820_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS093195_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS094993_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS095235_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGTACACGACCGCCCGGACCGTCGACTTG 480
HS095505_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480
HS095664_1 GCGAAAAC TCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGGACCGTCGACTTG 480

FP889223_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS094889_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS093815_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092939_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092790_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS093201_1 CTCAAAGAGCCCG**C**GATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS094575_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092845_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS096096_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS095744_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGG**C**GAAAGCCTTCGAGGGTCTG 540
 HS093225_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS094600_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092884_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS096119_1 CTCA**C**AGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS093801_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092279_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092907_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092773_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS095838_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS092820_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS093195_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS094993_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS095235_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS095505_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540
 HS095664_1 CTCAAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAAGCCTTCGAGGGTCTG 540

FP889223_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS094889_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS093815_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092939_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092790_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS093201_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS094575_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092845_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS096096_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS095744_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS093225_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS094600_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092884_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS096119_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS093801_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092279_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092907_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092773_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS095838_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS092820_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS093195_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS094993_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS095235_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS095505_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600
HS095664_1 ATAAAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCT 600

FP889223_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS094889_1 CCCGCCAAGATGGCTATGGGAGCCATAA-TAAGTTTGG-AAGTTTCC-TTCGACGTTTT**T** 657
 HS093815_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS092939_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS092790_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS093201_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS094575_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG**GA**AAGTTTCC-TTCGACGTTTT**T** 659
 HS092845_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS096096_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS095744_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS093225_1 CCCGCCAAGAT**GA**CTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS094600_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC**C**TTCGACGTTTT**T** 659
 HS092884_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS096119_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS093801_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTT-**GGA**AAGTTTCC-TTCGACGTTTT- 657
 HS092279_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTT**TTGGA**AAGTTTCC-TTCGACGTTTT**T** 659
 HS092907_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTT-**GGA**AAGTTTCC-TTCGACGTTTT- 657
 HS092773_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTT-**GGA**AAGTTTCC-TTCGACGTTTT- 657
 HS095838_1 CCCGCCAAGATGGCTGTGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS092820_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS093195_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT- 657
 HS094993_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT**T** 658
 HS095235_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT**T** 658
 HS095505_1 CCCGCCAAGATGGCTATGGGAG**TC**ATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT**T** 658
 HS095664_1 CCCGCCAAGATGGCTATGGGAGCCATAAAATAAGTTTGG-AAGTTTCC-TTCGACGTTTT**T** 658
 ***** ** ***** ***** ***** * *** ***** *****

FP889223_1	AAT-----	660
HS094889_1	AATAAGCGCGTCCATACAGAC	678
HS093815_1	-----	
HS092939_1	-----	
HS092790_1	-----	
HS093201_1	-----	
HS094575_1	A-----	660
HS092845_1	-----	
HS096096_1	-----	
HS095744_1	-----	
HS093225_1	-----	
HS094600_1	A-----	660
HS092884_1	-----	
HS096119_1	-----	
HS093801_1	-----	
HS092279_1	A-----	660
HS092907_1	-----	
HS092773_1	-----	
HS095838_1	-----	
HS092820_1	-----	
HS093195_1	-----	
HS094993_1	ATAAGCGCGTCCATACAGAC-	678
HS095235_1	ATAAGCGCGTCCATACAGAC-	678
HS095505_1	ATAAGCGCGTCCATACAGAC-	678
HS095664_1	ATAAGCGCGTCCATACAGAC-	678

Figure 1.4 Protein alignment of 25 unique ESTs of Pea Aphid C002

25 unique ESTs were translated to protein sequence and aligned. Differences in amino acid residues are indicated in bold face font.

```
HS092279_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS094575_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS094600_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS093801_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS093195_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS095505_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS095235_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS094993_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS095664_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS093201_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092884_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092790_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS096096_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS093815_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
FP889223_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS093225_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092907_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092820_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS096119_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS095744_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS095838_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092773_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092845_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS092939_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAATEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
HS094889_1      MGSYKLYVAVMAIAIAVAVQEVRCDSAAEPYDEQEEASVELPMEHRQCDEYKSKIWDKAF 60
*****.*****:*****
```

HS092279_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS094575_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS094600_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS093801_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNA~~Y~~YSLGMMNKMLAFI I 120
 HS093195_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNA~~P~~PYSLGMMNKMLAFI I 120
 HS095505_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS095235_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS094993_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS095664_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS093201_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092884_1 SNQEAMQLM~~K~~ELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092790_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS096096_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS093815_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 FP889223_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS093225_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092907_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAH~~C~~SLGMMNKMLAFI I 120
 HS092820_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS096119_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS095744_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS095838_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092773_1 SNQEAM~~R~~LMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092845_1 SNQEAMQLMELTFNTGKELG~~P~~HEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS092939_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 HS094889_1 SNQEAMQLMELTFNTGKELGSHEVCSDTTRAI FNFVDVMATNQNAHYSLGMMNKMLAFI I 120
 *****: ** : ***** . ***** ***** ***** *****

HS092279_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS094575_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS094600_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS093801_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS093195_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS095505_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS095235_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKYTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS094993_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS095664_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS093201_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPAIRGRLFKVVKAFEG 180
 HS092884_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092790_1 REVDTTSNKFKETKEAFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS096096_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS093815_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 FP889223_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS093225_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092907_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092820_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS096119_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLTEPVIRGRLFKVVKAFEG 180
 HS095744_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVAKAFEG 180
 HS095838_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092773_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092845_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS092939_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 HS094889_1 REVDTTSNKFKETKEVFERIAKTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEG 180
 *****.*****:*****.***.*****.*****

HS092279_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFWKFPSTFL	220
HS094575_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGKFPSTFL	220
HS094600_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFPSTFL	220
HS093801_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS093195_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS095505_1	IKPSENEELVKQRLKRLITNAPAKMAMGVIN----KFGSFLRRFYKRVHTD	226
HS095235_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRFYKRVHTD	226
HS094993_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRFYKRVHTD	226
HS095664_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRFYKRVHTD	226
HS093201_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS092884_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS092790_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS096096_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS093815_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
FP889223_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRFN	220
HS093225_1	IKPSENEELVKQRLKRLITNAPAKMTMGAIN----KFGSFLRRF	219
HS092907_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS092820_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS096119_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS095744_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS095838_1	IKPSENEELVKQRLKRLITNAPAKMAVGAIN----KFGSFLRRF	219
HS092773_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS092845_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS092939_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIN----KFGSFLRRF	219
HS094889_1	IKPSENEELVKQRLKRLITNAPAKMAMGAIISLEVSDVFNKRVHTD	226

*****:*****:*. * . * .

Figure 1.5 Nucleotide Alignment of Aphid C002 Sequences.

Alignment of C002 ESTs from *A. pisum* against: *A. gossypii*, *M. persicae*, and *T. citricida*.

Changes in protein sequence are highlighted in Figure 6 with bold face font.

```

A. gossypii      ATGGGACGTTACCAATTATATGTGGCCGTCATGGCAATATCTAGCTTAGCCGTAATACAG 60
T. citricida     ATGGGACGTTACCAATTATATGTGGCCGTCGTTGGCAATATCTAGCTTAGCCGTAATACAG 60
A. pisum         ATGGGAAGTTACAAATTATACGTAGCCGTCATGGCAATAGC---CATAGCTGTAGTACAG 57
M. persicae     ATGGCAAGTTACAAATTGTACCTGGCCGTCATGGCAATAGCTGTCATAGCTGCAGTTCAG 60
                ***** ** * ***** ***** * * ***** * * * **

A. gossypii      AAAGCTAGTTGTGCTGGTGGGCCTAACGCCTACAACACT----- 99
T. citricida     AAAGCTAGTTGTGCTGGTGGGTCTAACGCCTACCCAGT----- 99
A. pisum         GAAGTTAGATGCGA---TTGGTCTGCCGC----- 83
M. persicae     GAAATTAGTTGCAAGTTTCAGACTTCCGAACAGGACGATGATCAGGAAGGATATTACGAT 120
                ** *** ** * * ** **

A. gossypii      -----ACCGAACAAATA-----CATTGAGTCAAAAAGA----- 125
T. citricida     -----ACTGAACAATA-----CGATGAATCAAAAAGA----- 125
A. pisum         -----TGAACCGTA-----CGATGAGCAGGAAGAAGCGT-- 112
M. persicae     GATGAGGGAGGAGTGAACGATAATCAGGGAGAAGAGAACGATAATCAGGGAGAAGAGAAC 180
                **** ** * * * ***

A. gossypii      -----CGGATTA 132
T. citricida     -----CAAATTA 132
A. pisum         -----CTTTCGAATTA 123
M. persicae     GATAATCAGGGAGAAGAGAACGATAATCAGGGAGAAGAGAAGGAAGCTTCCGAACCA 240
                * * *

A. gossypii      GAAATGGAGCATCATCAGTGTGATGAATACAAATCGAAAATCTGGAATAAGGCATTTAGC 192
T. citricida     GAAATGGAGCACCATCAGTGTGATGAGTACAAATCGAAGATCTGGAATAAGGCATTTAGC 192
A. pisum         CCGATGGAGCACCCTCAGTGCATGAATACAAATCGAAGATCTGGGACAAAGCATTTAGC 183
M. persicae     GAGATGGAGCACCATCAGTGCGAAGAATACAAATCGAAGATCTGGAACGATGCATTTAGC 300
                ***** * ***** ** * ***** ***** * * *****

A. gossypii      AACCCAGCGGCTATGCAACTGGTGGACGTAGTGCTTAAACAGCTAAGGAAATGGGAACC 252
T. citricida     TACCCAGCGGCTATGCAACTGATGGACGTAATTTTTGAAACAGCTAAGGAAATGGGAACC 252
A. pisum         AACCCAGGAGGCTATGCAGCTGATGGAACTAACTTTAATACAGGTAAGGAATTAGGCTCC 243
M. persicae     AACCCGAAGGCTATGAACCTGATGAAACTGACGTTTAAATACAGCTAAGGAATTGGGCTCC 360
                *** ***** * ** * * * * * * * ***** * * * **

```

A. gossypii **G**AATAACGTGTGCTCAGACACGATTCCGGTCTTGTCTAATTTTCATCGATGTGATGGCCACC 312
T. citricida **A**ATGACGTGTGCTCAGACACAATTCGGGTCTTGTCTAACTTCATCGATGTGATGGCCACC 312
A. pisum **C**ACGAAGTGTGCTCGGACACGACGCGGGCCATTTTTAACTTCGTGATGTGATGGCCACC 303
M. persicae **A**ACGAAGTGTGCTCGGACACGACCCGGGCCTTATTTAACTTCGTGATGTGATGGCCACC 420
* * * * *

A. gossypii AATCAGAACTCCCACTACTCGGTGGGAATGTTGGCAAAGATGCTGGCATTTCATTGCGAGA 372
T. citricida AATCAGAACTCCCACTACTCGGTGGGAATGTTGGGAAAGATGCTGGCATTTCATTGCGAGA 372
A. pisum AACCAGAACGCCCATTTACTCGCTGGGTATGATGAACAAGATGTTGGCGTTCATCATAAGA 363
M. persicae AGCCCGTACGCCCACTTCTCGCTAGGTATGTTTAAACAAGATGGTGGCGTTTATTTTGAGG 480
* * * * *

A. gossypii GAAGCGGATATGACGTCGGACAAATTCAGAGATACAAAGGAGGTGTTTCGATCGCAT**C**GTA 432
T. citricida GAAGTGGATACGACGTCGGACAAATTCAGAGAGACAACGGAGGTGTTTCGAACGCAT**C**GCG 432
A. pisum GAGGTGGACACGACGTCCTCAACAAATTCAAAGAGACGAAGGAGTTTTTCGAACGCAT**C**GCG 423
M. persicae GAGGTGGACACGACATCGGACAAATTTAAAGAGACGAAGCAGGTGGTTCGACCGTAT**C**TCG 540
* * * * *

A. gossypii CAAAATGCTGATATCCGTGACTATATCAGAAACACGGCCTCCCGGTGGTTCGACTTGCTC 492
T. citricida AAAAAATGCTGATATCCGTGACTATATTAGGCAC----- 465
A. pisum AAAACTCCAGAGATCCGAGACTATATCAAGCACACGACCGCCCGACCGTTCGACTTGCTC 483
M. persicae AAAACTCCAGAGATCCGTGACTATATCAGAACTCGGCCGCAAGACCGTTCGACTTGCTC 600
* * * * *

A. gossypii AAGTTACCCGTGATGAGAAACCGATTAGCCAGAGTGTTTAAAGCCTTTGAGACTTTGTAT 552
T. citricida -----
A. pisum AAAGAGCCCGTGATTAGAGGCCGACTGTTCAAAGTGGTGAAGCCTTCGAGGGTCTGATA 543
M. persicae AAGGAACCCAAGATTAGAGCACGACTGTTTCAGAGTGTGAAAGCCTTCGAGAGTCTGATA 660

A. gossypii AATCCATCCAAAAACAGCAAATGATCAAGCAGAGGATTATGGGACTTACCAACACTCCT 612
T. citricida -----
A. pisum AAACCGTCCGAAAACGAGGAATTGGTCAAGCAGAGGCTTAAGAGGATAACCAATGCTCCC 603
M. persicae AAACCAAACGAAAACGAAGCATTAAATCAAACAGAAGATTAAGGGTTAACCAATGCTCCC 720

A. gossypii **T**CCAAAATAGCTATGGGTACCATGAATAAGGT**T**GGAAATATGTTTCA**A**AATGTC 666
T. citricida -----
A. pisum **A**CCAAGATGGCTATGGGAGCCATAAATAAGTT**T**GGAAGTTTCCTTC**G**ACGTTTT 657
M. persicae **G**TCAAGTTAGCCAAGGGTGCCATGAAAACGGT**T**GACGTTTCTTTAG**A**CATTTT 774

Figure 1.6 Alignment of Aphid C002 Amino Acid Sequences.

Alignment of translated validated C002 haplotypes for *A. pisum*, *A. gossypii*, *M. persicae*, and *T. citricida*. Variable positions in amino acid sequence observed in Figure 2 are compared to species-specific changes and are represented by bold face font. Of the 8 changes observed in Figure 2, 7 are represented in the main sequence, and 2 are represented by the immediate C-Terminus.

```

A. gossypii      MGRYQLYVAVMAISSLAVIQKASCAGGP-NAYNTTEQYI----- 38
T. citricida     MGRYQLYVAVVAISSLAVIQKASCAGGS-NAYPSTEQYD----- 38
A. pisum         MGSYKLYVAVMAIA-IAVVQEVRCDSAAEEPYDEQE----- 35
M. persicae      MASYKLYLAVMAIAVIAAVQEISCKVQTSEQDDDDQEGYYDDEGGVNDNQGEENDNQGEEN 60
                *. *:**:**:**: :*.:*: * . : *

A. gossypii      -----ESKDGLEMEHHQCDEYKSKIWNKAFSNPAAMQLVDVVLKTAKEMGT 84
T. citricida     -----ESKDKLEMEHHQCDEYKSKIWNKAFSYPAAMQLMDVIFETAKEMGT 84
A. pisum         -----EASVELPMEHRQCDEYKSKIWDKAFSNQEAMQLMELTFNTGKELGS 81
M. persicae      DNQGEENDNQGEEEEEASEPEMEHHQCEEYKSKIWNDAFSNPKAMNLMLTFNTAKELGS 120
                *      ***:**:*****:.*** **:*:..: :*.**:**:

A. gossypii      DNVCSDTIRVLSNFIDVMATNQNSHYSVGMLAKMLAFIAREADMTSDKFRDTKEVFDRIV 144
T. citricida     NDVCSDTIRVLSNFIDVMATNQNSHVVGMLGKMLAFIAREVDTTSDKFRETTEVFERIAA 144
A. pisum         NEVCSDTTRAIFNFVDVMATNQNAHSLGMMNKMLAFIIREVDTTSNKFKETKEVFERIAA 141
M. persicae      NEVCSDTTRALFNFVDVMATSPYAHFSLGMFNKMVAFILREVDTTSDKFKETKQVVDRIS 180
                .:***** *.: **:*****. :*: **: **:* **.* **:*:*.:.:.*

A. gossypii      QNADIRDYIRNTASRVVDLLKLPVMNRRLARVFKAFETLYNPSKNQQMIKQRIMGLTNTP 204
T. citricida     KNADIRDYIRH----- 155
A. pisum         KTPEIRDYIKHTTARTVDLLKEPVIRGRLFKVVKAFEGLIKPSENEELVKQRLKRITNAP 201
M. persicae      KTPEIRDYIRNSAAKTVDLLKEPKIRARLFRVMKAFESLIKPNENEALIKQKIKGLTNAP 240
                :...:*****:

A. gossypii      SKIAMGTMNKVGNMFQNV 222
T. citricida     -----
A. pisum         AKMAMGAINKFGSFLRRF 219
M. persicae      VKLAKGAMKTVGRFFRHF 258

```

Table 1.1 Pea Aphid EST Analysis

Summary of the amino acid and nucleotide changes for the 8 validated haplotypes. An EST count and frequency of the EST found out of the 94 ESTs from the dataset is provided in columns 2 & 3 of the nucleotide & amino acid sequences respectively. Discreet nucleotide & amino acid sequence changes are provided in columns 4-12 in the table respectively. The C-terminal amino acid sequence & library source are provided in columns 13 & 14 respectively.

Haplotype	EST Count	EST Frequency ¹	Nuc. 11	Nuc. 82	Nuc. 115	Nuc. 211
1	2	2%	A	G	G	C
2	2	2%	A	G	G	C
3	45	48%	A	G	G	C
4	7	7%	A	G	G	C
5	2	2%	A	G	G	C
6	4	4%	A	G	T	A
7	29	31%	A	G	T	A
8	2	2%	A	A	T	A
9	3	3%	G	G	T	A

Haplotype	Nuc. 244	Nuc. 421	Nuc. 604	Nuc. 637	Nuc. 652
1	C	G	G	Gap	G
2	C	G	G	T	A
3	C	G	G	Gap	A
4	C	G	G	Gap	A
5	A	G	A	Gap	A
6	A	A	G	Gap	A
7	A	A	G	Gap	A
8	A	A	G	Gap	A
9	A	A	G	Gap	A

Haplotype	EST Count	EST Frequency ¹	Residue 4	Residue 28	Residue 39	Residue 71	Residue 82
1	2	2%	Tyr	Ala	Val	Leu	His
2	2	3%	Tyr	Ala	Val	Leu	His
3	45	48%	Tyr	Ala	Val	Leu	His
4	7	7%	Tyr	Ala	Val	Leu	His
5	2	2%	Tyr	Ala	Val	Leu	Asn
6	4	4%	Tyr	Ala	Phe	Ile	Asn
7	29	31%	Tyr	Ala	Phe	Ile	Asn
8	2	2%	Tyr	Thr	Phe	Ile	Asn

Haplotype	Residue 141	Residue 202	Residue 213	Residue 214	C-Terminus	EST Source
1	Ala	Ala	Gly	Ser	LRRF	P(2)*
2	Ala	Ala	Trp	Pro	STFL	P(2)*
3	Ala	Ala	Gly	Ser	LRRF	P(40)*, M(5)**
4	Ala	Ala	Gly	Ser	LRRFYKRVHTD	P(7)*
5	Thr	Thr	Gly	Ser	LRRF	H(2)***
6	Thr	Ala	Gly	Ser	LRRFYKRVHTD	P(4)*
7	Thr	Ala	Gly	Ser	LRRF	P(28)*, H(1)***
8	Thr	Ala	Gly	Ser	LRRF	P(2)*

*Kirk Pappan EST Library of Pea Aphid salivary glands from adult parthenogenetic females

http://www.ncbi.nlm.nih.gov/biosample/LIBEST_026773

**Nancy Moran EST library of Pea Aphid whole body from nymph, adult, winged, and wingless parthenogenetic females

http://www.ncbi.nlm.nih.gov/biosample/?term=LIBEST_021987

***Wayne Hunter EST library of Pea Aphid whole body from nymph and adult parthenogenetic females

http://www.ncbi.nlm.nih.gov/biosample/?term=LIBEST_015734

The number parenthesis represents the number of ESTs obtained from each library

¹Frequency is the percentage of ESTs representing each validated haplotype with respect to the total number of ESTs in the validated haplotypes. Frequency does not to describe the frequency of haplotypes in nature or within the total number of C002 ESTs found at NCBI.

Table 1.2 Table of all ESTs included within each validated haplotype

94 ESTs used to create the 8 validated haplotypes are provided. EST accession numbers presented in Figures 1-4 are shown in bold face font.

Haplotype	EST Accession Number
1	HS092421_1 HS094460_1
2	HS096025_1 HS095987_1
3	HS092499_1 HS092366_1 HS092178_1 HS095667_1 HS095418_1 HS095361_1 HS095309_1 HS095230_1 HS095137_1 HS095096_1 HS094949_1 HS094937_1 HS094728_1 HS096917_1 HS094651_1 HS094399_1 HS094363_1 HS094259_1 HS094252_1 HS094193_1 HS094149_1 HS094074_1 HS092340_1 HS092532_1 EX644002_1 FF294143_1 FP892242_1 FP889576_1 FP912637_1 HS095776_1 HS095912_1 HS096124_1 HS096240_1 HS096382_1 HS096492_1 HS096533_1 HS096621_1 HS092805_1 HS092911_1 HS093002_1 HS093029_1 HS093219_1 HS093295_1 HS093579_1 HS093936_1
4	HS095737_1 HS096280_1 HS096452_1 HS094169_1 HS094199_1 HS095663_1 HS092544_1
5	CN759948_1 CN759518_1
6	HS093775_1 HS093629_1 HS096126_1 HS092406_1
7	HS096239_1 HS096118_1 HS095920_1 HS095836_1 HS095716_1 CN763229_1 HS094752_1 HS096888_1 HS094569_1 HS094542_1 HS094316_1 HS094310_1 HS094289_1 HS094285_1 HS094280_1 HS094270_1 HS094237_1 HS093696_1 HS093646_1 HS093573_1 HS093192_1 HS093156_1 HS092994_1 HS092992_1 HS092969_1 HS092818_1 HS092808_1 HS096667_1 HS096304_1
8	HS095928_1 HS092906_1

Chapter 2 - Effect of Armet Transcript Knockdown in *Tribolium castaneum*

Introduction

Armet Etymology

The Armet sequence was first discovered from a human cDNA, and designated arginine rich protein (ARP) due to a 55 tri-nucleotide repeat at the 5' end encoding arginine (Shridhar et al., 1996). Polymorphisms of Armet were later found in a variety of early stage tumors, and Armet was renamed arginine rich mutated in early state tumors (Armet) (Shridhar et al., 1996, 1997). Later, Armet was cloned from human ventral mesencephalic cell line 1 (VMCL1). This translated product was found to not contain the 55 tri-nucleotide repeat, but was instead a secretory protein containing a signal peptide. Both the arginine rich and tumor based mutations of Armet turned out to be incorrect. The designation of mesencephalic astrocyte-derived neurotrophic factor (MANF) was assigned to Armet (Petrova et al., 2003). Armet is still a prevalent title for the gene in the literature and will be in this thesis.

Armet in Endoplasmic Reticulum Stress (ER Stress) & Unfolded Protein Response (UPR)

Armet was first implicated in ER stress response when the addition of chemical ER stress inducers (tunicamycin, thapsigargin, and DDT) to mouse pancreatic β cells induced an increase

in both transcript and protein levels. The increased expression correlated to increases of two other known ER stress genes, Grp78 and Hsp70 (Mizobuchi et al., 2007). Additionally, microarray analysis of U2OS and HEK293 challenged by the same chemical inducers found Armet to be up-regulated. Armet was found to be localized in the ER, Golgi, and secreted. Knockdown of Armet showed a higher propensity towards apoptosis, while over-expression conferred a degree of protection to cells (Apostolou et al., 2008).

The protective effects of Armet have been studied mainly in neural and cardiac tissues. It has been demonstrated that expression of recombinant human Armet and addition to dissected mouse mesencephalic tissue provided increased protection to dopaminergic neurons (Petrova et al., 2003). Immunohistochemical studies showed that Armet is expressed during neural development and in adult mice. Induction of epilepsy caused a transient increase of Armet in the thalamus and hippocampus. Also, forebrain ischemia induced a large increase in Armet expression in the hippocampus and a small, restricted increase in the cerebral cortex (Lindholm et al., 2008). Armet has been shown to decrease apoptosis and necrosis in mice that have undergone cerebral ischemia. Recovery of mouse behavior is also associated with the action of Armet (Airavaara et al., 2009; Yu et al., 2010). Armet is also secreted by ventricular myocytes, and during simulated cardiac ischemia (Tadimalla et al., 2008; Glembotski 2011).

Structure of Armet

The structure of human Armet was determined using x-ray crystallography (Parkash et al., 2009). Additionally mouse and human Armet structures were determined using solution state

nuclear magnetic resonance (Hoseki et al., 2010; Hellman et al., 2010; Hellman et al., 2010). All structural studies showed a protein containing two domains separated by a cysteine-bridge linker region. The N-terminal domain has a saposin like structure while the C-terminal end is predicted to be intrinsically unstructured.

Armet and Arthropod Development

D. melanogaster lines containing P-element insertions to abrogate Armet expression showed a distinct phenotype. Larvae homozygous for defective Armet died as first instar larvae. Lack of Armet led to loss of dopaminergic cells and the degradation of the developing larval axonal bundles. It was also noted that the cuticle of the larvae was malformed by the observations that antibodies could pass through without permeabilization (Palgi et al., 2009), likely due to a lack of dopamine to act a protein cross linking agent.

Here I probe the effects on development of the red flour beetle, *Tribolium castaneum*, by knocking down Armet in late (determined by mass) larvae. *Tribolium* is the model RNAi system for insects, owing to the systemic effect of injection of dsRNAs. We demonstrate that depletion of Armet leads to an arrest in development during the eclosion from pupa to adult. It is possible that a loss of dopaminergic neurons is causing an inability for these insect to properly cross-link their adult cuticle.

Material & Methods

Synthesis of dsRNA

The T-7 promoter PCR based method was used for preparation of dsRNAs (Tomoyasu et al 2005). For TcArmet dsRNAs the primers with the following sequence were used: 5'- TAA TAC GAC TCA CTA TAG GGC TTC ACT GTT CTA TTC GCAA G -3' and 5'- TAA TAC GAC TCA CTA TAG GGA GTT CGG TAT GTT TAT GCT TCA -3' (forward and reverse respectively). The product created was 486 bp in length and began at position 18 of the open reading frame and terminated at position 504. For TcVer dsRNAs the T7 template for vermilion was provided by Dr. Muthukrishnan.

PCR products were purified using the Qiaquick PCR Purification Kit (Qiagen, Valencia, CA). dsRNAs were synthesized using the Ampliscribe™ T7-Flash™ Kit (Epicentre biotechnologies, Madison, WI). dsRNAs were purified by adding phenyl chloroform, extracting with chloroform, and precipitating with 3M sodium acetate and isopropyl alcohol. dsRNA pellets were washed with 75% ethanol in DEPC treated deionized distilled water, and resuspended in DEPC treated deionized distilled water. dsRNA concentrations were confirmed using a Nanodrop 1000 (Thermoscientific, Wilmington, DE).

Injection

Larvae weighing 2 mg (late instar larvae) were selected for injection. Larvae were injected with 200 ng of dsRNA using a pulled glass capillary needle. Capillaries (outer diameter of 1.0 mm,

inner diameter of 0.5 mm, Sutter Instruments, Navato, CA) were pulled using a Model P-87 capillary puller and filament (Sutter Instruments, Navato, CA).

RNA Extraction

RNA from injected late instar larvae was extracted 48 hrs after injection using the TriZol reagent (Invitrogen, Grand Island, NY), followed by extraction with chloroform, and precipitation with 3M sodium acetate and isopropyl alcohol. RNA was converted to cDNA using Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen, Grand Island, NY).

qRT-RT-PCR

Quantitative real time reverse transcription polymerase chain reaction was employed to create an age dependent expression profile of Armet with respect to housekeeping gene RPS6. Insects from early larvae (<2 mg), late larvae (>2 mg), prepupating, pupa, newly emerged adults, and adults aged 1 week were selected for expression analysis. Also, qRT-RT-PCR was used to quantify the knockdown levels for dsArmet and dsVermilion. 2 mg larvae were injected with dsArmet or dsVermilion and RNA was extracted 48 hours post injection. For RPS6 the forward and reverse primers 5'-AGA TAT ATG GAA GCA TCA TGA AGC-3' and 5'-CGT CGT CTT CTT CTC TGC TCA AAT TG-3' respectively were used. For Armet the forward and reverse primer used were 5' AAG ACT GCC AAG GGT GCA T-3' and 5'-CAT GCA CAT TCA CTG GGA AC-3' respectively. qRT-RT-PCR was conducted using a Bio-Rad iCycler iQ Real-Time PCR machine (Hercules, CA). Bio-Rad SYBER Green SsoFast™ Evagreen SuperMix was used

to fluorescently label PCR products. Melting temperature of 94⁰ Celsius, annealing temperature of 55⁰ Celsius, and extension temperature of 68⁰ Celsius were used for 40 cycles of PCR. Results were analyzed using ExPCRt program (<http://www.expertqpcr.com>) based on the $\Delta\Delta C_t$ method.

Results

To determine the level of Armet transcript in different developmental stages, qRT-RT-PCR was used. 3 biological replicates of 5 pooled insects were measured for their Armet transcript level relative to ribosomal protein S6 (RPS6) transcript. Early instar larvae, late instar larvae, prepupa, pupa, newly emerged adults, and adults aged 1 week were probed for their Armet transcript level in Figure 12. Armet transcript level reaches its peak levels during the late instar larvae, newly emerged adults, and 1 week old adult stages.

To understand the role of Armet in development of *T. castaneum* dsRNA mediated depletion of Armet's transcript was employed. Late instar larvae were injected with 200ng of either double stranded RNA encoding the gene Armet (dsArmet) or Vermilion (dsVER). Three sets of injections were made and development of larvae to adult stage was monitored. The numbers of living insects with respect to days after injection are reported in Figures 1, 4, and 7. Insects injected with dsVER had a higher number of surviving insects in all three replicates. The developmental stages of all three replicates were also monitored. The number of insects in each developmental stage is provided in Figures 2, 3, 5, 6, 8, and 9. In all three cases *Tribolium* injected with dsVER were able to progress through pupation and eclosion into adulthood, and showed the typical phenotype of white or pigmentless eyes. The insects injected with dsVER

were able to progress through pupation, but development was arrested at the pharate adult stage and they were unable to undergo eclosion to adulthood in the majority of injected insects.

Pictures of *Tribolium* injected with either dsVER or dsArmet are provided in Figure 10 and 11. In Figure 10 insects injected with either dsVER or dsArmet are presented. The dsVER insect shows pigmentless eyes and was able to undergo eclosion to adulthood. The insect injected with dsArmet was able to progress to late pupa stage but arrested before eclosion to adulthood. In Figure 11 three insects injected with dsArmet are shown at various levels of pupa development. The insect on the left and in the middle are living injected insects that had tanned normally and are just prior to the lethal pupa phenotype. The insect on the right depicts a dead dsArmet injected insect that was unable to eclose to adult.

To assess the efficiency of the dsRNA mediated knockdown of Armet. qRT-RT-PCR measurements were taken for Armet in 2 biological replicates of 5 pooled insects each. Injected larvae were harvested for mRNA content as larvae 48 hours post injection. Relative expression levels of Armet transcript are displayed in Figure 13. Insects injected with dsArmet but no dsVER showed lower relative Armet transcript level.

Armet exists in insects and all vertebrates. An alignment of *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Acyrtosiphon pisum*, and *Tribolium castaneum* is provided in Figure 14. The structure of the *T. castaneum* Armet gene is presented in Figure 15. TcArmet contains 3 exons and 2 introns. The final mRNA is 763 bp in length.

Discussion

Knocking out Armet in *D. melanogaster* showed lethal effects during the first instar larval stages (Palgi et al 2009). I have shown by depleting Armet transcript level in late instar larvae that Tribolium their development during the progression of pupa to adult. Although these two results are not comparable at developmental levels, the Armet transcript is crucial for Tribolium development. By measuring the insects for dopamine content and probing at the morphology of the insect brain by dissection and microscopy, a connection to the results seen in *Drosophila* could be made. By injecting adults and allowing them to reproduce and transfer the knock down to their offspring by way of parental RNAi we would be able to compare early embryogenesis and larva development in *T. castaneum* to that of *D. melanogaster* (Bucher et al 2002).

I present a phenotype that is easily interpreted visually, the insects are able to progress from last instar to pupae, but die when attempting to emerge. However, there is no evidence here to elucidate the molecular mechanism of dsArmet killing. It is possible that death is occurring, as in the case with *D. melanogaster*, from the degradation of dopaminergic neurons in the brain. Protein secretion & cross-linking by the cuticle epithelium is an event of importance during eclosion that may require the presence of Armet. Although I injected much earlier, Armet may not be crucial until this step in development. Secretion by the midgut and epithelium is a tightly regulated process involving the secretion of numerous factors, possibly causing ER stress, to proceed correctly (Chaudhari et al 2011). It is important that a future developmental expression profile be created for TcArmet in individual tissues. Also, injected beetles be dissected and/or sectioned at the late pupae stage and first be assayed with microscopy for morphological

differences. Additionally, using immunohistochemistry to determine the state of known secreted factors during development will allow us to identify the molecular basis for dsArmet killing.

It might also be possible to tease out a weak phenotype by injecting lower concentrations of dsRNA as was done in the case of *Laccase-2* (Arakane et al 2005). It was shown that high concentrations of injected dsRNA for *Laccase-2* resulted in a lethal phenotype and failure to undergo cuticle tanning or sclerotization. At lower levels of injected dsRNAs the insects were able to survive and show a similar but weaker phenotype. By injecting smaller concentrations of dsArmet, insects may be able to proceed beyond the pupa stage and eclose to adult. In this situation a greater number of phenotypes could be visually assessed with living specimens.

Conclusion

Armet is differentially expressed at the various stages of *T. castaneum* development. Depletion of TcArmet transcript in late instar larvae leads to a late pupa lethal phenotype. The likely cause of death is due to a lack of dopaminergic neurons surviving the development of the adult brain, and an inability for the insects to properly form their adult cuticle.

Figure 2.1 Survival curves after injection of dsVER RNA or dsArmet RNA (#1)
9 Insects were injected with 200 ng of either dsVER RNA (blue) or dsArmet RNA (Red).
#1 denotes the first replication of this experiment.

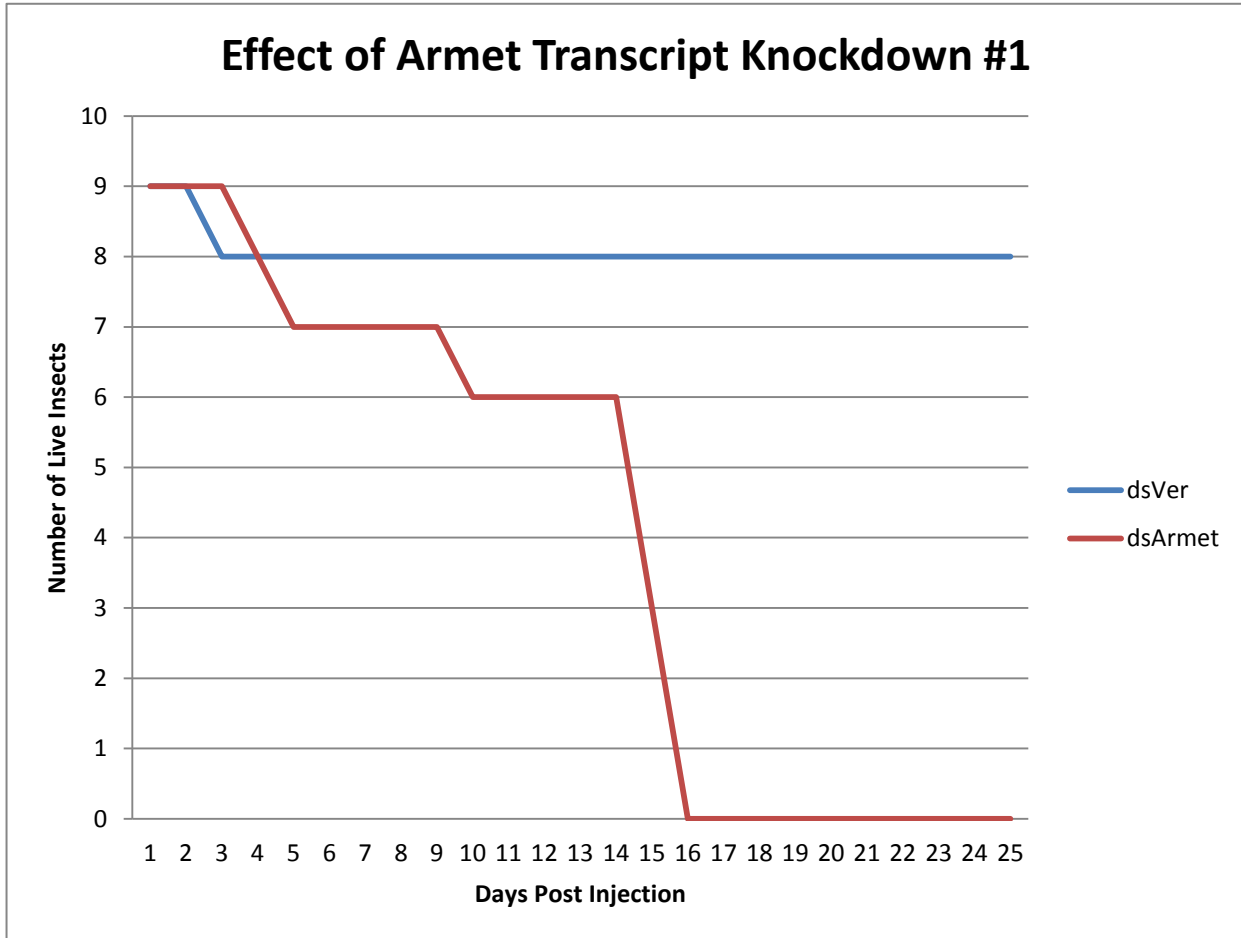


Figure 2.2 Number of insects in various developmental stages after dsVER RNA injections (#1)

9 Insects were injected with 200 ng of dsVER RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple). #1 denotes the first replication of this experiment

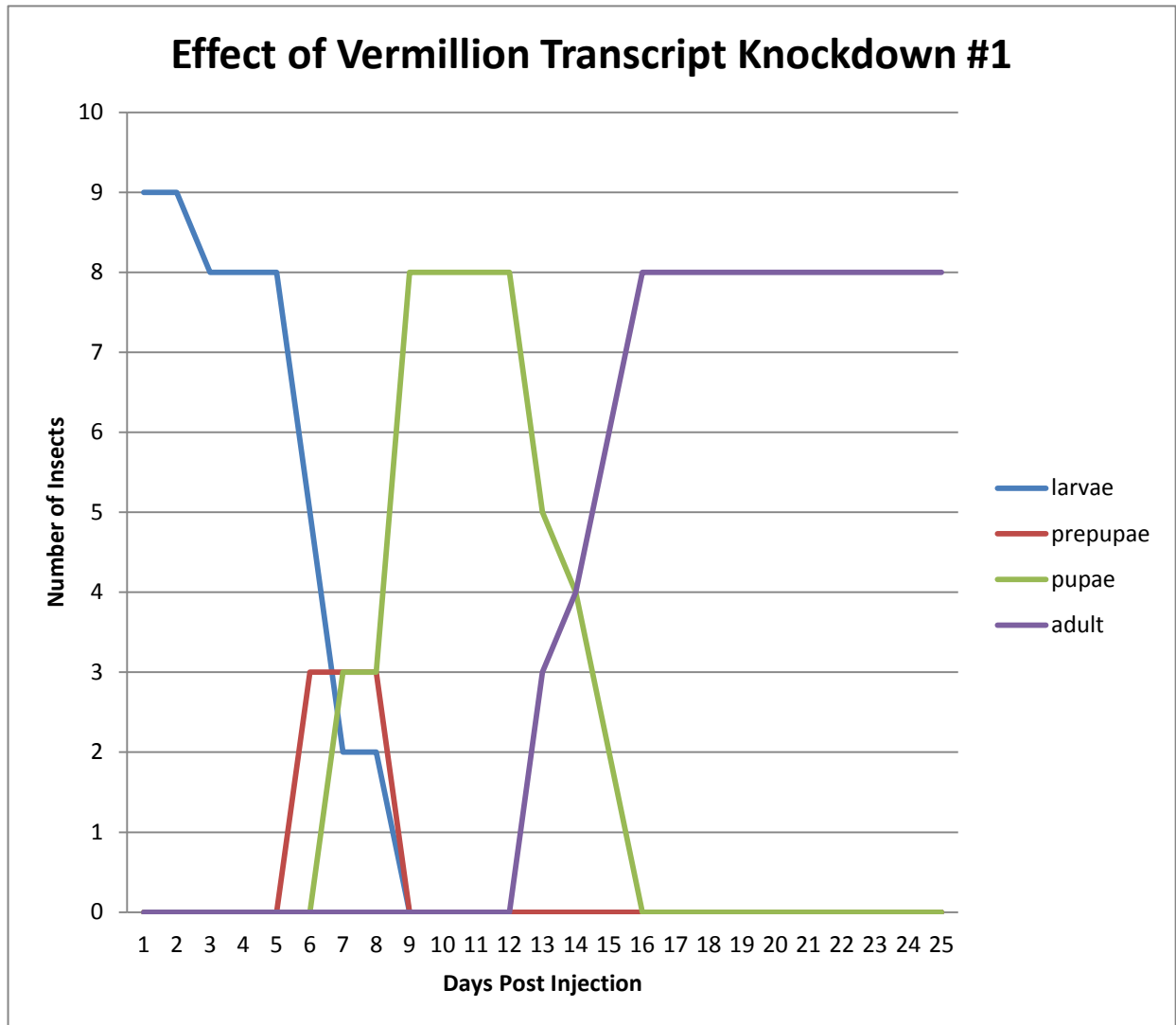


Figure 2.3 Number of insects in various developmental stages after dsArmet RNA injection (#1)

9 Insects were injected with 200 ng of dsArmet RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple).

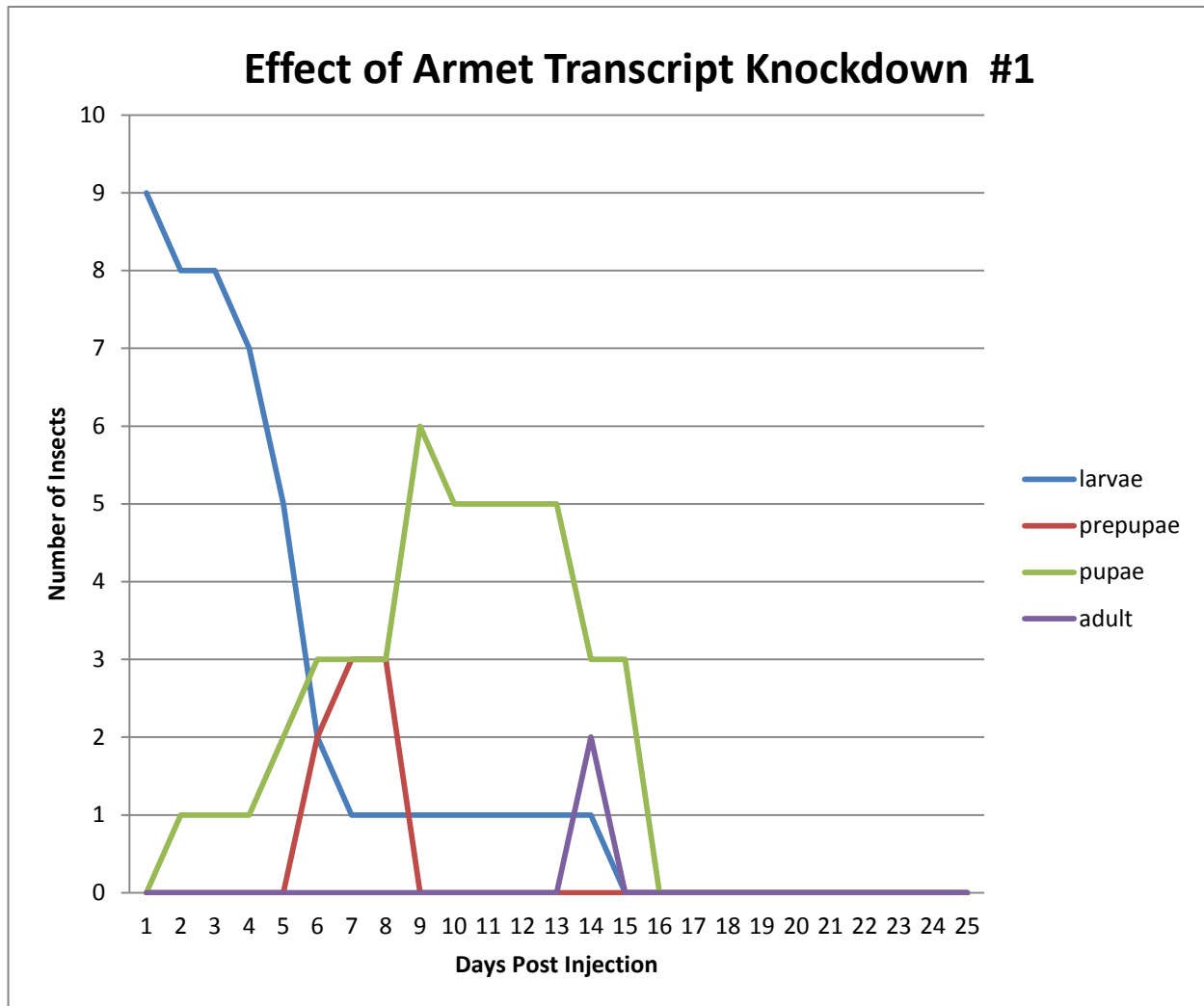


Figure 2.4 Survival curves after injection of dsVER RNA or dsArmet RNA (#2)
10 Insects were injected with 200 ng of either dsVER RNA (blue) or dsArmet RNA (Red).
One dsArmet injected insect was not counted due to injection related death .#2 denotes the
first replication of this experiment.

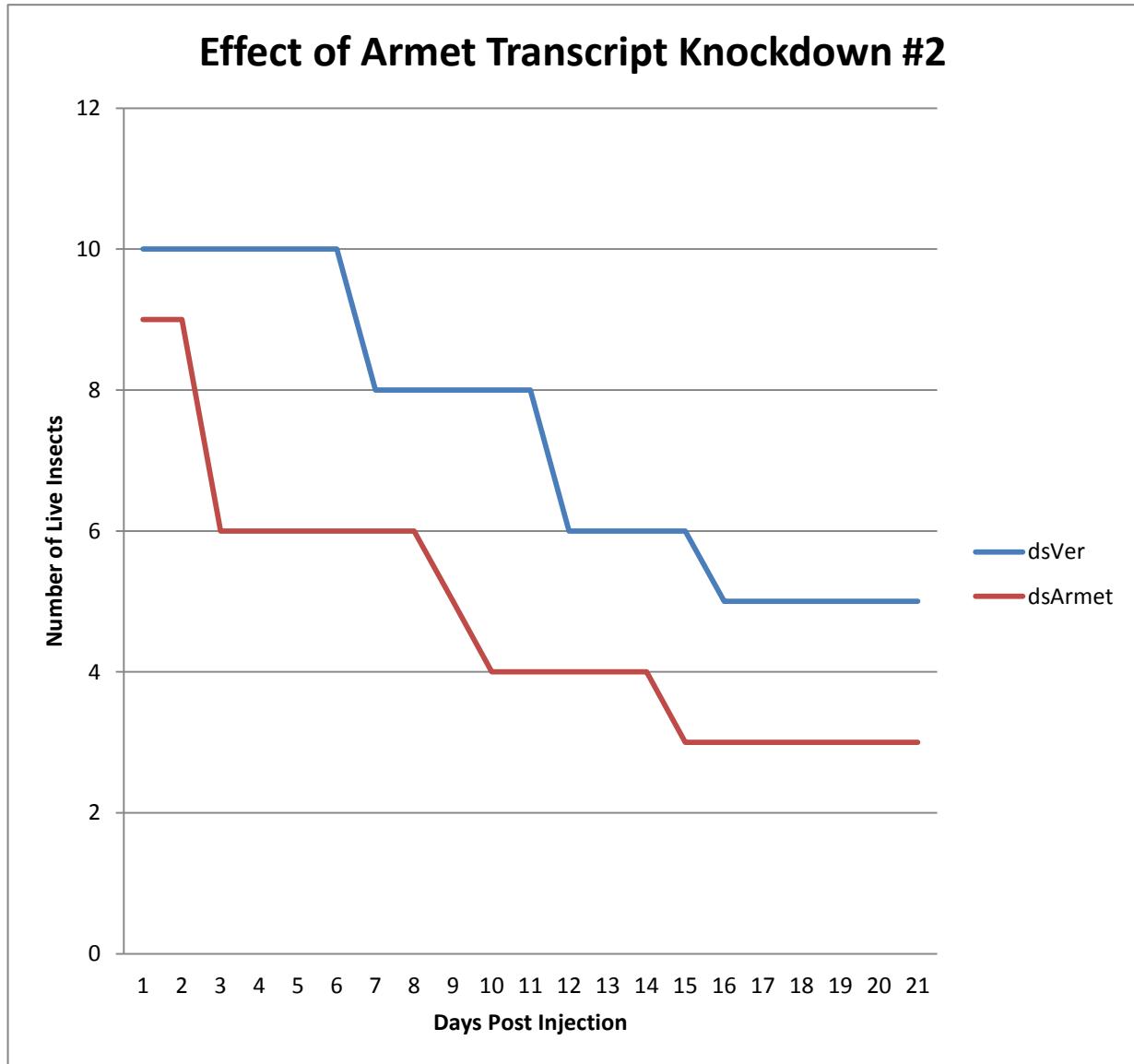


Figure 2.5 Number of insects in various developmental stages after dsVer RNA injection (#2)

10 Insects were injected with 200 ng of dsVER RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple). #2 denotes the second replication of this experiment.

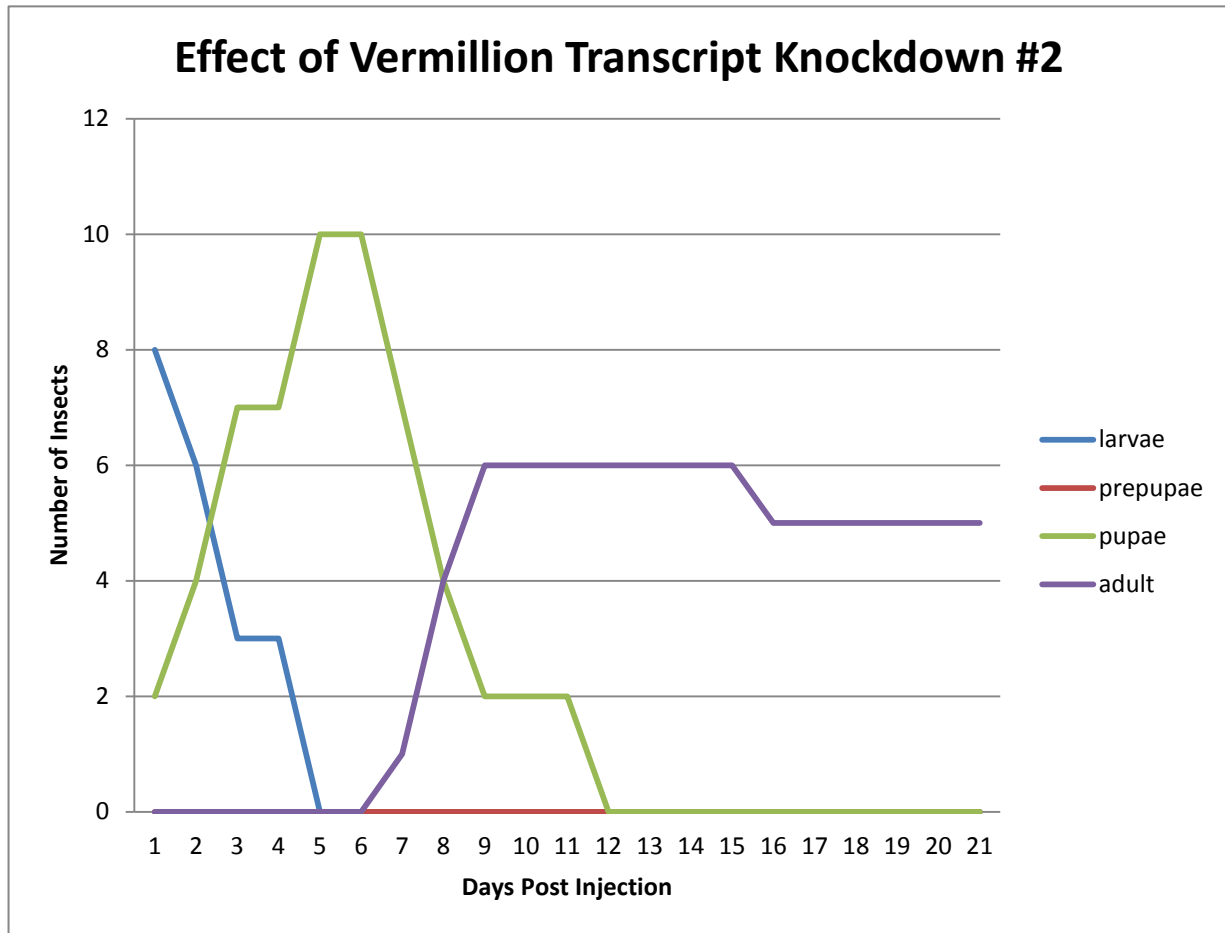


Figure 2.6 Number of insects in various developmental stages after dsArmet RNA injection (#2)

9 Insects were injected with 200 ng of dsArmet RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple). #2 denotes the second replication of this experiment.

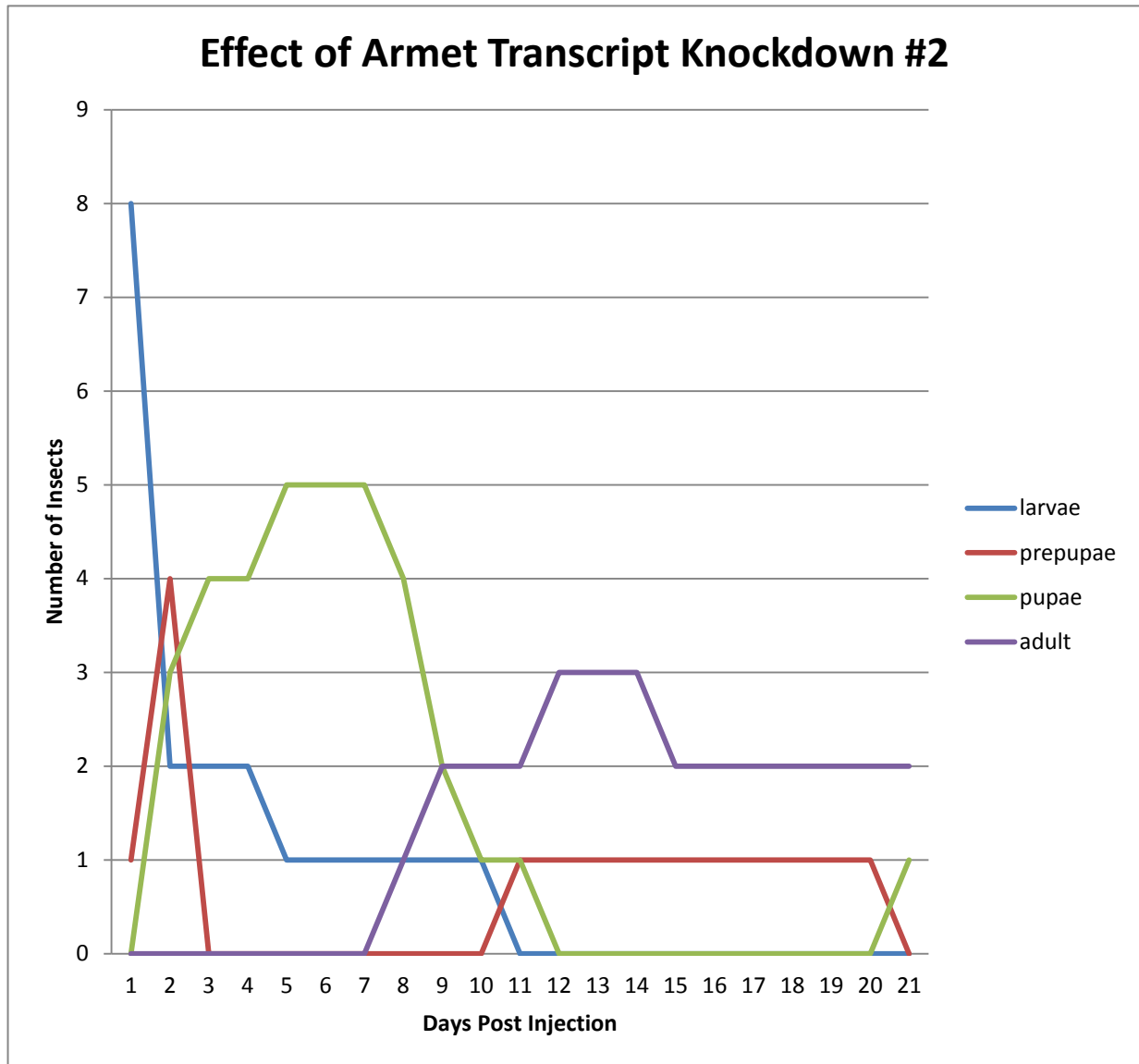


Figure 2.7 Graph representing live tribolium injected with either dsVER RNA or dsArmet RNA (#3)

9 Insects were injected with 200 ng dsVER RNA (blue) and 15 with dsArmet RNA (Red).

#3 denotes the third replication of this experiment

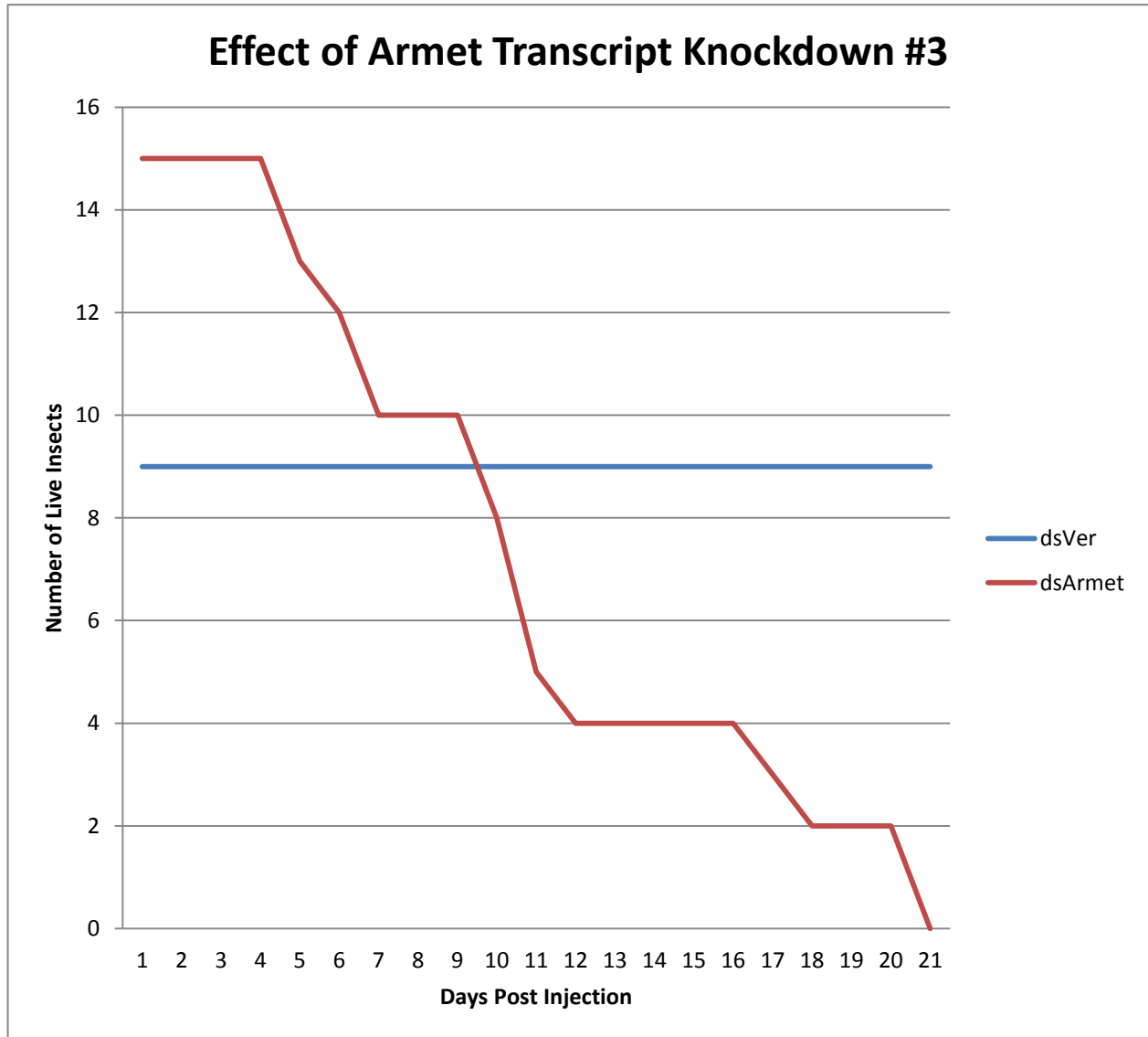


Figure 2.8 Number of insects in various developmental stages after dsVER RNA injection (#3)

9 Insects were injected with 200 ng of dsVER RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple).

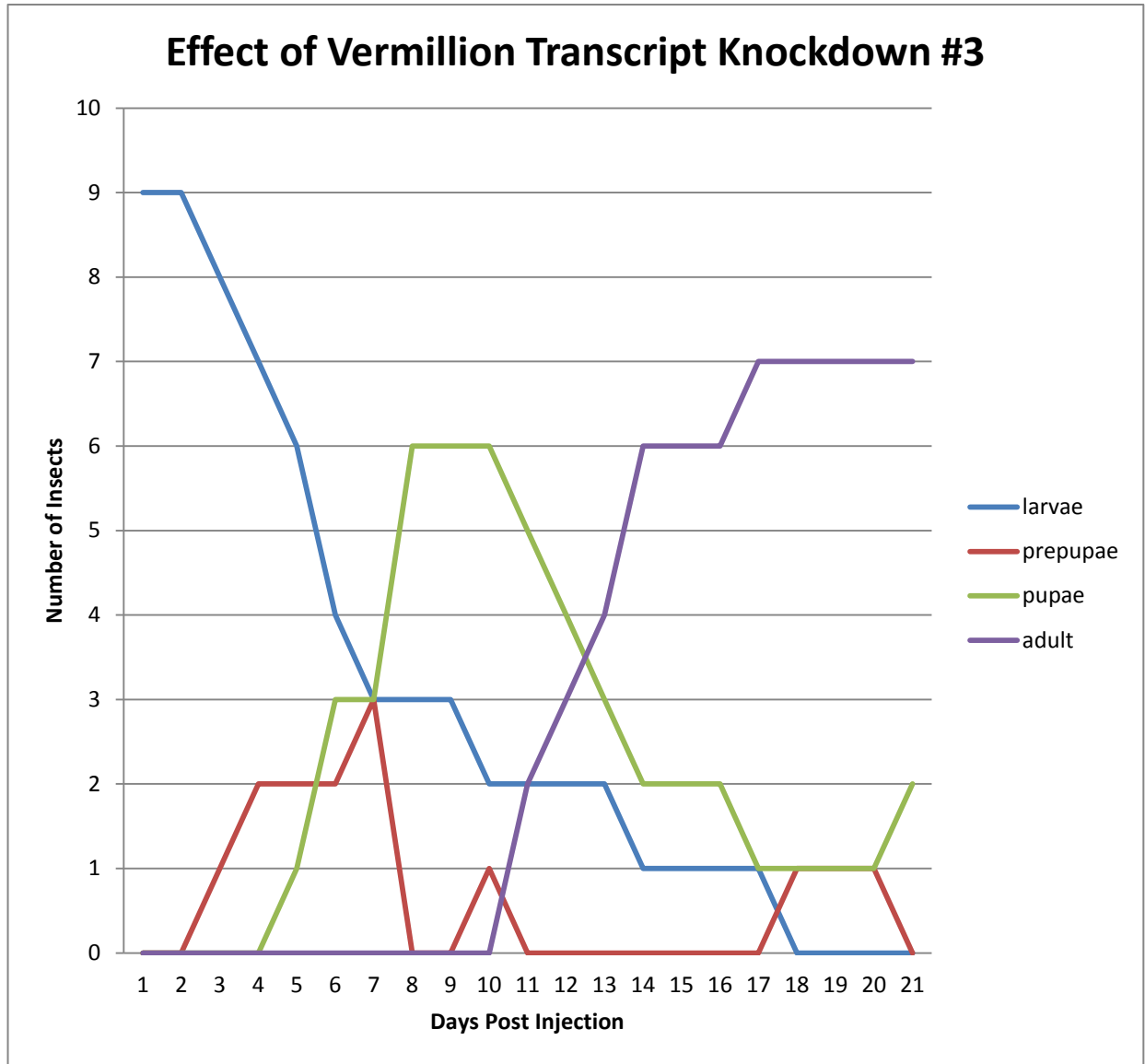


Figure 2.9 Number of insects in various developmental stages after dsArmet RNA injection (#3)

Insects were injected with 200 ng of dsArmet RNA. The number of insects in developmental stages: larvae (blue), Pre-pupa (red), pupa (green), and adult (purple). #3 denotes the third replication of this experiment.

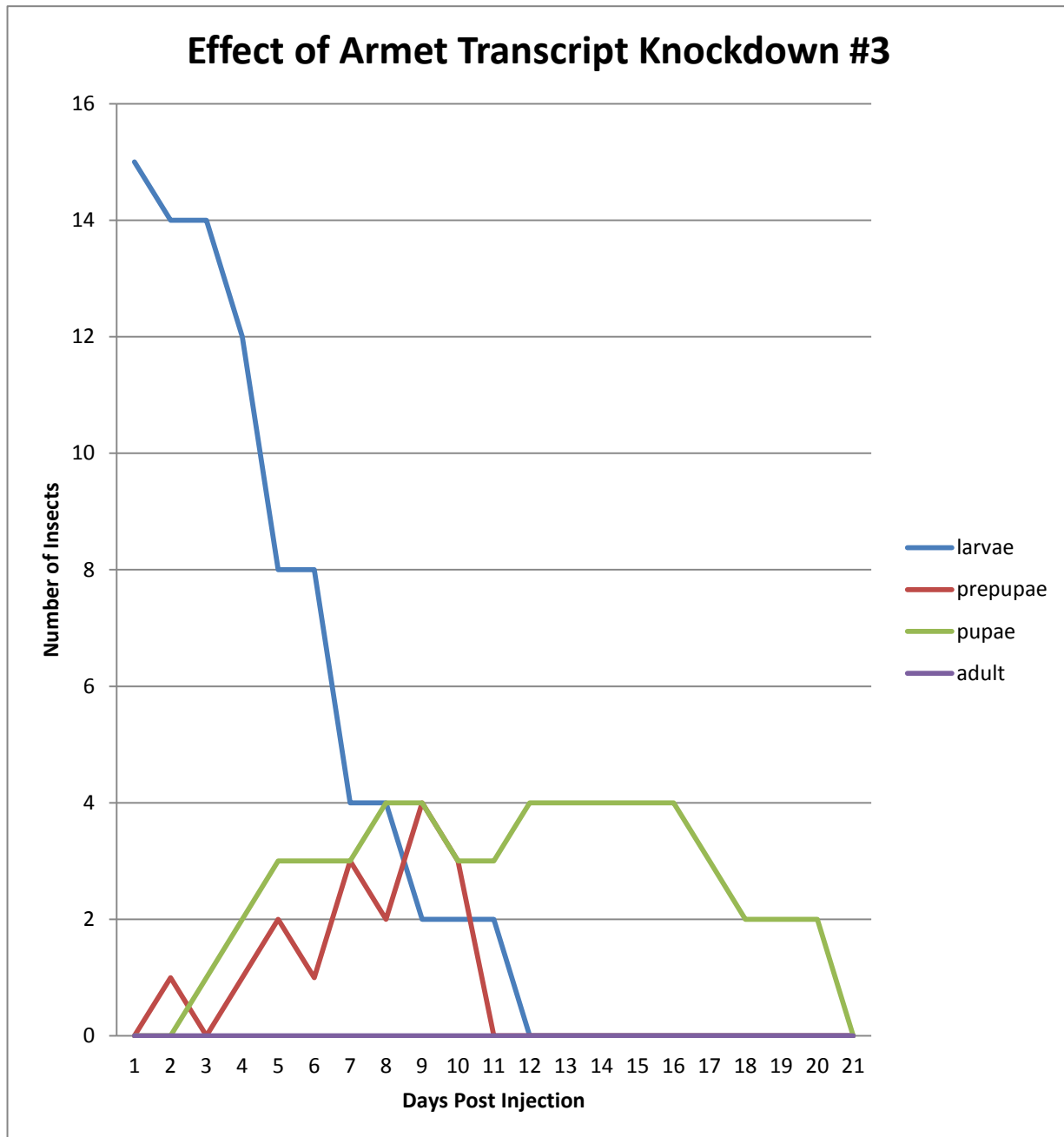


Figure 2.10 Larvae injected with dsVER RNA or dsArmet RNA

Insects were injected with 200 ng of dsVER or dsArmet and allowed to develop normally. Insects injected with dsVER lack of pigment in the eyes of pupa. Insects injected with dsArmet showed normal eye development but their development arrested at late pupa stage.

Pupa Lethal Phenotype

dsVer RNA

dsArmet RNA



Figure 2.11 dsArmet RNA injected larvae

Left: Living dsArmet injected insect showing normal tanning prior to lethal phenotype.

Middle: Living dsArmet injected insect showing normal tanning just prior to lethal phenotype.

Right: Dead dsArmet injected insect after lethal phenotype.

Pupa Lethal Phenotype



Figure 2.12 Armet transcript levels at several stages of development

RT-PCR expression profile of Armet including Early instar larvae (YL), late instar larvae (LL), prepupal (PP), pupal (P), young adult (YA), and 1 week aged adult beetles (LA) relative to RPS6. At each stage 3 biological replicates (n=3) pooled from 5 insects were used. Larvae weighing less than 2 mg were selected for being early instar larvae (YL). Larvae greater than 2 mg were designated as late instar larvae (LL). Insects identified as prepupal (PP) due to the signature elongated banana prepupal shape. Pupa (P). Adult's newly emerged from pupa were pooled for young adult (YA). Adults aged 1 week were designated late adults (LA). Armet transcript level measurements were done with respect to RPS6 transcript level. Error bars represent the standard error of each sample.

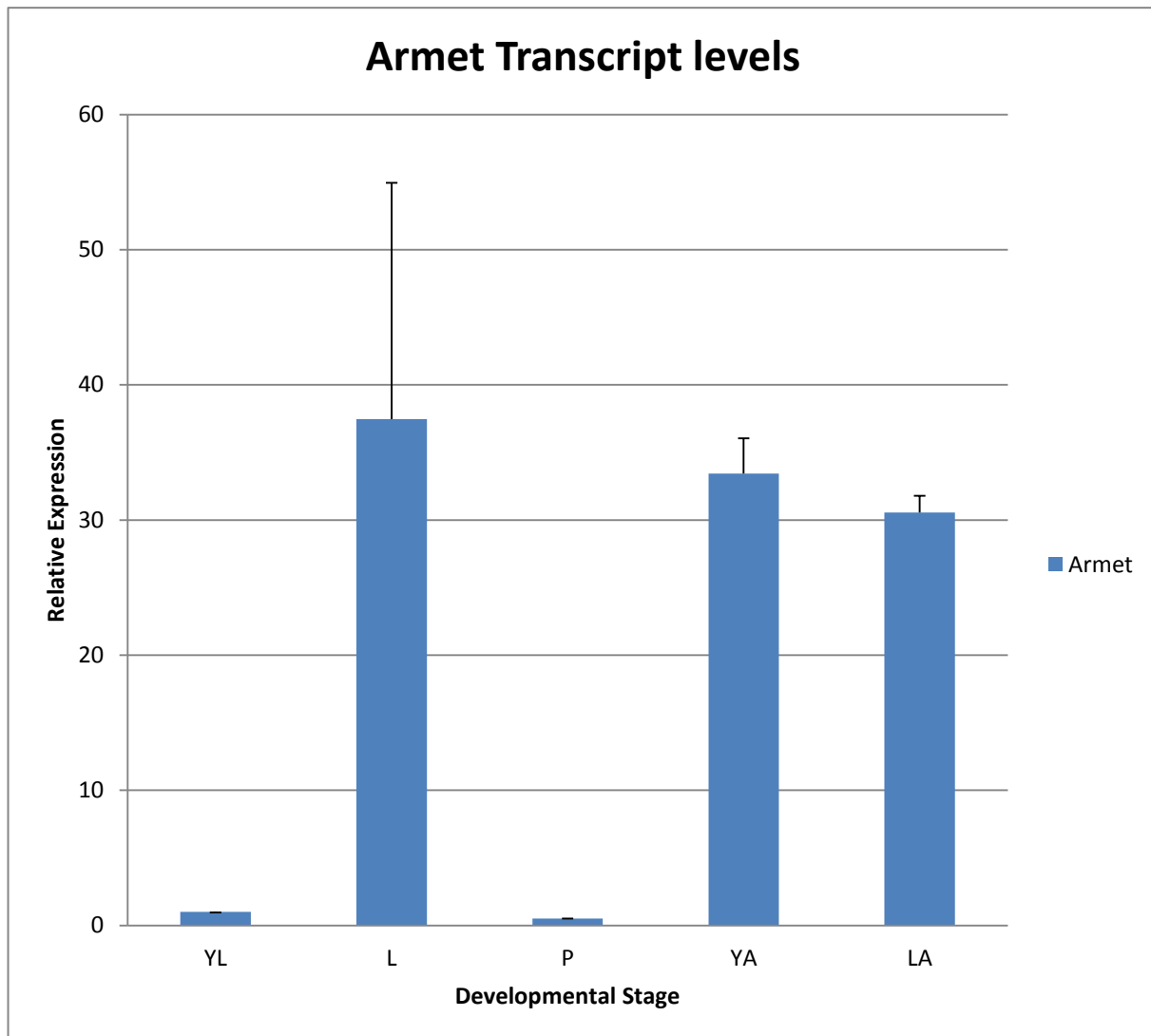


Figure 2.13 Armet transcript levels after injection with dsVER RNA or dsArmet RNA. Graph of qRT-PCR data for dsArmet and dsVER knock-down beetles relative to RPS6. Duplicate measurements (n=2) were conducted on pools of 5 larvae. Larvae injected with 200 ng of dsArmet or dsVER were harvested as larvae 48 hrs post injection. Armet transcript level was measured relative to RPS6.



Figure 2.14 Alignment of Armet amino acid sequences from Homo sapiens, Mus musculus, Drosophila melanogaster, Acyrthosiphon pisum, and Tribolium castaneum

ClustalW multiple sequence alignment for Armet sequences. Consensus sequence: “*” identical amino acid, “:” conserved amino acid, “.” semi-conserved amino acid.

<i>D. melanogaster</i>	MKTWYMVVVIGFLAT--LAQTSALKEEDCEVCVKTVRRFADSLDD--TKKDYKQI	53
<i>A. pisum</i>	MDKHILLVCVFFIVFHVFAQSRTTFTEEDCPVCVLTIDKFSKTLEG---ELNPKNI	53
<i>T. castaneum</i>	MEFQLIFTVLFASIV-----AVNSLKQGECEVCIKVLDKFAASLSDD--VKKDPKLI	50
<i>M. musculus</i>	MWATRGLAVALALSVP----DSRALRPGDCEVCISYLGREFYQDLKDRDVTFSPTI	53
<i>H. sapiens</i>	MRRMWATQGLAVALALSVP----GSRALRPGDCEVCISYLGREFYQDLKDRDVTFSPTI	56
	: . :	: : * ** : : * *.. . *
<i>D. melanogaster</i>	ETAFKKFCKAQK-NKEHRFCYYLGGLEESATGILNELSKPLSWSMPAEKICEKLNKDAQ	112
<i>A. pisum</i>	EEQFKKYCLSTKIDKEKRLCYLGGLEDSATGILSEMSKPLSWSMPALKICERLKKMDAQ	113
<i>T. castaneum</i>	ESKFRDYCKNTR-NKENRFCYYLGGLEESATGILGEMSKPLSWSMPDCKICEKLNKDAQ	109
<i>M. musculus</i>	EEELIKFCREAR-GKENRLCYIIGATDDAATKIINEVSKPLAHHIPVEKICEKLNKDSQ	112
<i>H. sapiens</i>	ENELIKFCREAR-GKENRLCYIIGATDDAATKIINEVSKPLAHHIPVEKICEKLNKDSQ	115
	* : .:* : .**:*:*:*:*.*. :*:** *:.*:*****: :* *****:** *:*	
<i>D. melanogaster</i>	ICDLRYEKQIDLNSVDLKKLKVRLDKKILNDWDESCDGCLEKGFDFIKRIEELKPKYSRSE	172
<i>A. pisum</i>	VCDIKYDKEIDWKTVNLKMKVKDLKKILDNWGEICDGCLEKTDYIKRVEELKPSYVKEE	173
<i>T. castaneum</i>	ICELRYDVEIDLKTVDLKKLKVRLDKKIINDWGEDCQGCIEKSEFIQRIEELKHKHT--E	167
<i>M. musculus</i>	ICELKYDKQIDLSTVDLKKLRVKELKKILDDWGEMCKGCAEKSDYIRKINELMPKYAPKA	172
<i>H. sapiens</i>	ICELKYDKQIDLSTVDLKKLRVKELKKILDDWGETCKGCAEKSDYIRKINELMPKYAPKA	175
	:*::*: :** .:*:*:*:*:*:*:*:*:*:*:* *.* ** ** :*:::*:** .:	
<i>D. melanogaster</i>	L	173
<i>A. pisum</i>	L	174
<i>T. castaneum</i>	L	168
<i>M. musculus</i>	ASARTDL	179
<i>H. sapiens</i>	ASARTDL	182

Figure 2.15 Alignment of Armet exons against the genomic sequence of *T. castaneum*
Tribolium Armet consists of 3 exons and 2 introns.

```

Chromosome      TACTTAAATAAAGTCATTAAGACAGTCAACTTCATAGAGCCCACCTTCCACACTAAAGC 60
TcArmet         -----

Chromosome      CTTGAGAGAGACGTGGACCAATCACGGTGCGCCAACGATGTACCCACAAAATCCAAAATC 120
TcArmet         -----

Chromosome      ACGTTCTGTCAAATCAATTTTTAATAACTGTCAACGTGTAAAACAAGCCAATAGTTAAT 180
TcArmet         -----ATTTTTAATAACTGTCAACGTGTAAAACAAGCCAATAGTTAAT 43
                  *****

Chromosome      TATTTTTGATAAAAAGTGTAACAATCAGTGATTTTTTACGAAATAAAAGCTCAAGTGAGT 240
TcArmet         TATTTTTGATAAAAAGTGTAACAATCAGTGATTTTTTACGAAATAAAAGCTCAA----- 97
                  *****

Chromosome      AGTTCTTTTCATTGTTTCATTCACAAACCCACACTTTTACACAAACCGGTCAAGACAATT 300
TcArmet         -----

Chromosome      GACATAACCTCACTTTGCAGCCCCAGAAATGGAATTTCAATTAATCTTCACTGTTCTATT 360
TcArmet         -----CCCAGAAATGGAATTTCAATTAATCTTCACTGTTCTATT 137
                  *****

Chromosome      CGCAAGCATCGTCGCGGTGAACAGTTTAAAGCAAGGAGAATGCGAAGTCTGTATCAAAGT 420
TcArmet         CGCAAGCATCGTCGCGGTGAACAGTTTAAAGCAAGGAGAATGCGAAGTCTGTATCAAAGT 197
                  *****

Chromosome      CTTGGACAAATTCGCGGCCAGTTTATCAGACGACGTGAAAAAAGACCCGAAATTGATAGA 480
TcArmet         CTTGGACAAATTCGCGGCCAGTTTATCAGACGACGTGAAAAAAGACCCGAAATTGATAGA 257
                  *****

Chromosome      ATCCAAATTTCCGGGACTATTGTAAAAACACGAGAAACAAGGAGAATCGATTTGTAAGCAG 540
TcArmet         ATCCAAATTTCCGGGACTATTGTAAAAACACGAGAAACAAGGAGAATCGATTT----- 309
                  *****

```


Chromosome	CAATTGTTGATTTCGTCAAGTTTCACAAAAATAAAAATGCAATAAAAATTTTCAGTGTTACTA	600
TcArmet	-----TGTTACTA	317

Chromosome	CTTGGGGGGCCTTGAGGAAAGTGCCACTGGTATTTTGGGCGAAATGTCCAAACCGCTTTC	660
TcArmet	CTTGGGGGGCCTTGAGGAAAGTGCCACTGGTATTTTGGGCGAAATGTCCAAACCGCTTTC	377

Chromosome	ATGGTCAATGCCATCGGACAAAATCTGCGAAAAATTGAAGAAAAAGACGCCCAAATCTG	720
TcArmet	ATGGTCAATGCCATCGGACAAAATCTGCGAAAAATTGAAGAAAAAGACGCCCAAATCTG	437

Chromosome	CGAATTGCGCTACGACGTCGAAATCGATTAAAGACAGTTGATTTGAAGAACTCAAAGT	780
TcArmet	CGAATTGCGCTACGACGTCGAAATCGATTAAAGACAGTTGATTTGAAGAACTCAAAGT	497

Chromosome	GAGGGATTTGAAGAAGATTATCAACGACTGGGGCGAAGACTGCCAAGGGTGCATCGAAAA	840
TcArmet	GAGGGATTTGAAGAAGATTATCAACGACTGGGGCGAAGACTGCCAAGGGTGCATCGAAAA	557

Chromosome	AAGCGAGTTCATCCAAAGAATAGAAGAATTGAAGCATAAACATACCGAACTTTAAGTGTG	900
TcArmet	AAGCGAGTTCATCCAAAGAATAGAAGAATTGAAGCATAAACATACCGAACTTTAAGTGTG	617

Chromosome	TTGTTACTAGTGAAATACTGGACTGATTATTTATTACTGGAGCTCCAAGCGCACCGTTCC	960
TcArmet	TTGTTACTAGTGAAATACTGGACTGATTATTTATTACTGGAGCTCCAAGCGCACCGTTCC	677

Chromosome	CAGTGAATGTGCATGAAATTAACCTTTATTTTCATAGTTCGGTTTATTAGGATTGTATTTAT	1020
TcArmet	CAGTGAATGTGCATGAAATTAACCTTTATTTTCATAGTTCGGTTTATTAGGATTGTATTTAT	737

Chromosome	ATTTGTAATAAACATGAATTGAAAAACAGTTTATTTATTAATAAATGATACATTTTTATTA	1080
TcArmet	ATTTGTAATAAACATGAATTGAAAAA-----	763

Chromosome	CATTACTATCTACATCAGGGGGGGGGCAATAACCTGAAAAAACAATTTTATCAGTAG	1140
TcArmet	-----	
Chromosome	GTTTTATTTGAACTTCCCGCGGGCTCAGCTGAGTACAGTG	1181
TcArmet	-----	

References

- Airavaara M, Shen H, Kuo CC, Peränen J, Saarma M, Hoffer B, Wang Y. Mesencephalic astrocyte-derived neurotrophic factor reduces ischemic brain injury and promotes behavioral recovery in rats. *The Journal of Comparative Neurobiology* 2009;515(1):116-124.
- Apostolou A, Shen Y, Liang Y, Luo J, Fang S. Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced cell death. 2008;314(13):2454-2467.
- Bos J, Prince D, Pitino M, Massimo M, Win J, Hogenhout S. A functional genomics approach identifies candidate effectors from the aphid species *myzus persicae* (green peach aphid). *PLoS Genetics* 2010;6(11).
- Candeias M, Malbert-Colas L, Powell D, Daskalogianni C, Maslon M, Naski N, Bourougaa K, Calvo F, Fåhræus R. P53 mRNA controls p53 activity by managing Mdm2 functions. *Nat Cell Bio* 2008;10(9).
- Carolan J, Caragea D, Reardon K, Muttie N, Dittmer N, Pappan K, Cui F, Cataneto M, Poulain J, Dossat C, et al. Predicted effector molecules in the salivary secretome of the pea aphid (*acyrthosiphon pisum*): A dual Transcriptomic/Proteomic approach. *J Proteome Res* 2011;10(4):1505-1518.
- Carolan J, Fitzroy C, Ashton P, Douglas A, Wilkinson T. The secreted salivary proteome of the pea aphid *acyrthosiphon pisum* characterised by mass spectrometry. *Proteomics* 2009;9:2457, 2467.
- Chaudhari S, Arakane Y, Specht C, Moussian B, Boyle D, Park Y, Kramer K, Beeman R, Muthukrishnan S, Knickkopf protein protects and organizes chitin in the newly synthesized insect exoskeleton. *PNAS* 2011;108:17028-17033.

- Chen M, Liu X, Yang Z, Zhao H, Shukle R, Stuart J, Hulbert S. Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. *BMC Evolutionary Biology* 2010;10(296).
- Glembotski CC. Functions for the cardiomyokine, MANF, in cardioprotection, hypertrophy and heart failure. *Journal of Molecular and Cellular Cardiology* 2011;51(4):512-517.
- Hall T. BioEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999;41:95-98.
- Harmel N, Letocart E, Cherqui A, Giordanengo P, Mazzuccheli G, Guillonneau F, De Pauw E, Haubruge N, Francis F. Identification of aphid salivary proteins: A proteomic investigation of *myzus persicae*. *Insect Molecular Biology* 2008;17(2):165-174.
- Hellman M, Arumäe U, Yu L, Lindholm P, Peränen J, Saarma M, Permi P. Mesencephalic astrocyte-derived neurotrophic factor (MANF) has a unique mechanism to rescue apoptotic neurons. *The Journal of Biological Chemistry* 2011;286:2675-2680.
- Hellman M, Peränen J, Saarma M, Permi P. ^1H , ^{13}C and ^{15}N resonance assignments of the human mesencephalic astrocyte-derived neurotrophic factor. *Biomolecular NMR Assignments* 2010;4(2):215-217.
- Hogenhout S, Hoorn R, Terauchi R, Kamoun S. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *The American Phytopathological Society* 2009;22(2):115-112.
- Hoseki J, Sasakawa H, Yamaguchi Y, Maeda M, Kubota H, Kato K, Nagata K. Solution structure and dynamics of mouse ARMET. *FEBS Letters* 2010;584(8):1536-1542.

- Lindholm P, Peränen J, Andressoo J, Kalkkinen N, Kokaia Z, Lindvall O, Timmusk T, Saarma M. MANF is widely expressed in mammalian tissues and differently regulated after ischemic and epileptic insults in rodent brain. 2008;39(30):356-371.
- Miles P. Aphid saliva. Biol Rev 1999;74:41-85.
- Mizobuchi N, Hoseki J, Kubota H, Toyokuni S, Nozaki J, Naitoh M, Koizumi A, Nagata K. ARMET is a soluble ER protein induced by the unfolded protein response *via* ERSE-II element. 2007;32(1):41-50.
- Mutti N, Louis J, Pappan L, Pappan K, Begum K, Chen M, Park Y, Dittmer N, Marshall J, Reese J, et al. A protein from the salivary glands of the pea aphid, *acyrthosiphon pisum* is essential in feeding on a host plant. PNAS 2008;105(29):9965-9969.
- Mutti N, Park Y, Reese J, Reeck G. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *acyrthosiphon pisum*. JIS 2006;6(38).
- Ollivier M, Legeai F, Rispe C. Comparative analysis of the acyrthosiphon pisum genome and expressed sequence tag-based gene sets from other aphid species. Insect Molecular Biology 2010;19:33-45.
- Palgi M, Lindström R, Peränen J, Piepponen P, Saarma M, Heino T. Evidence that DmMANF is an invertebrate neurotrophic factor supporting dopaminergic neurons. PNAS 2009;106(7):2429-2434.
- Parkash V, Lindholm P, Peränen J, Kalkkinen N, Oksanen E, Saarma M, Leppänen VM, Goldman A. The structure of the conserved neurotrophic factors MANF and CDFN explains why they are bifunctional. PEDS 2009;22(4):233-241, PEDS Volume 22, Issue 4 Pp. 233-241.

- Petrova P, Raibekas A, Pevsner J, Vigo N, Anafi M, Moore M, Peaire A, Shridhar V, Smith D, Kelly J. MANF A new mesencephalic, astrocyte-derived neurotrophic factor with selectivity for dopaminergic neurons. *Journal of Molecular Neuroscience* 2003;20(2):173-187.
- Pitino M, Coleman A, Maffei M, Ridout C, Hogenhout S. Silencing of Aphid Genes by dsRNA Feeding from Plants. *PLoS One* 2011;6(10).
- Shridhar V, Rivard S, Shridhar R, Mullins C, Bostick L, Sakr W, Grignon D, Miller O, Smith D. A gene from human chromosomal band 3p21.1 encodes a highly conserved arginine-rich protein and is mutated in renal cell carcinomas. 1996;20(2):1931-1939.
- Shridhar V, Rivard S, Wang X, Shridhar R, Paisley C, Mullins C, Beirnat L, Dugan M, Sarkar F, Miller O, et al. Mutations in the arginine-rich protein gene (ARP) in pancreatic cancer. *Oncogene* 1997;14(18):2213-2216.
- Tadimalla A, Belmont PJ, Thuerauf DJ, Glassy MS, Martindale JJ, Gude N, Sussman MS, Glembotski CC. Mesencephalic astrocyte-derived neurotrophic factor is an ischemia-inducible secreted endoplasmic reticulum stress response protein in the heart. 2008;103:1249-1258.
- Tang J, Vosman B, Voorrips R, Linden C, Leunissen J. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploidy species. *BMC Bioinformatics* 2006;7:426.
- Thompson J, Higgins D, Gibson T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res* 1994;22(22):4673-4680.

- Ting-Cai C, Qing-You X, Ji-Feng Q, Chun L, Ying L, Xing-Fu Z, Zhong-Huai X. Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain *Dazao*. *Insect Biochemistry and Molecular Biology* 2004;24:523-530.
- Tjanlingii W. Salivary secretions by aphids interacting with proteins of phloem wound responses. *Journal of Experimental Botany* 2006;57(4):739-745.
- Vos M, Jander G. *Myzus persicae* (green peach aphid) salivary components induce defense responses in *arabidopsis thaliana*. *Plant, Cell and Environment* 2009;32(11):1548-1560.
- Will T, Tjallingii F, Thonnessen A, van Bel A. Molecular sabotage of plant defense by aphid saliva. *PNAS* 2007;104(25):10536-10541.
- Yu YQ, Liu LC, Wang FC, Liang Y, Cha DQ, Zhang JJ, Shen YJ, Wang HP, Fang S, Shen YX. Induction profile of MANF/ARMET by cerebral ischemia and its implication for neuron protection. *Journal of Cerebral Blood Flow & Metabolism* 2010;30:79-91.
- Zhang C, Shi H, Chen L, Wang X, Lü B, Zhang S, Liang Y, Liu R, Qian J, Sun W, et al. Harpin-induced expression and transgenic overexpression of the phloem protein gene AtPP2-A1 in arabidopsis repress phloem feeding of the green peach aphid *myzus persicae*. *BMC Plant Pathology* 2011;11(11).