DEVELOPING AND USING EXPRESSED SEQUENCE TAGS TO STUDY THE PREDATORY MITE *Phytoseiulus persimilis* ATHIAS-HENRIOT (PARASITIFORMES, MESOSTIGMATA, PHYTOSEIIDAE)

by

JU LIN WENG HUANG

B.S. Universidad de Costa Rica, 1997
M.S. Universidad de Costa Rica, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Abstract

The predatory mite *Phytoseiulus persimilis* (Acari, Phytoseiidae) is one of the most frequently released natural enemies for biological control of spider mites in greenhouse and outdoors crops. In this research, I utilized Expresses Sequence Tags (ESTs), the most cost effective approach for transcriptome exploration, to study three different aspects of this arachnid species for which there is little genomic information. I combined two EST datasets from different whole body cDNA libraries and analyzed by bioinformatics means. Approximately 54% of 10,256 uniESTs were annotated based on the homology to sequences in the National Center for Biotechnological Information (NCBI) database. A list of these uniESTs, sorted from most to least likelihood based on the expected value from the blast search in public databases, was used to create tools for each of the three studies. First, I described sixty-one genes encoding products known to be important in pesticide metabolism and in endocrinology, including cytochrome P450s, glutathione-S-transferases, acetylcholinesterase homologs, neuropetides and neurohormones. Findings on arachnid specific esterases and neuropetides, and possible benefits to pest management programs, were discussed. Next, I inferred divergence time for Acari and the point of divergence of two lineages within anactinotrichid mites, *Ixodes scapularis and Phytoseiulus persimilis.* I used expresses sequence tags from the predatory mite *P. persimilis* to pull out 74 orthologous amino acid sequences of invertebrates species: nine insect species, *Daphnia pulex*, *Ixodes scapularis*, *and Caenorhabditis elegans*. I estimated a similar origin for Chelicerata (578.1 ± 38.2 - 482.2 ± 7.2 Mya) as in other recent studies. However, divergence dating using amino acid sequences suggested a Devonian origin of anactinotrichid mites (487.6 ± 32.2 - 410.1 ± 6.1 Mya) based on four reference dates (two fossil records and two molecular clocks) and four amino acid substitution methods; this estimate is much earlier that those in the current literature. This discrepancy of divergence times may be due to the use of a global clock. Finally, I developed molecular markers from the EST dataset to examine inheritance in the haplodiploid system in *P. persimilis*. Biparental contribution of chromosomes is required among the predatory mites but the paternal chromosome set seems to be eliminated or loss (Paternal genome loss, PGL) in male offspring. However, genetic studies in other two phytoseiid species were suggested diploid males with PGL only in the germ cells. In the present study, haploid adult males of *P. persimilis* have been observed using five independent EST-derived markers. Single

mites derived from inter-population crosses were genotyped after whole genome amplification. The parahaploid genetic system in *P. persimilis* is supported by this study, in which both sexes arise from fertilized eggs but the paternal chromosome set is subsequently lost in males.

DEVELOPING AND USING EXPRESSED SEQUENCE TAGS TO STUDY THE PREDATORY MITE *Phytoseiulus persimilis* ATHIAS-HENRIOT (PARASITIFORMES, MESOSTIGMATA, PHYTOSEIIDAE)

by

JU LIN WENG HUANG

B.S. Universidad de Costa Rica, 1997
M.S. Universidad de Costa Rica, 2005

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Co-Major Professor
David Margolies

Approved by:

Co-Major Professor
Yoonseong Park

Copyright

Abstract

The predatory mite *Phytoseiulus persimilis* (Acari, Phytoseiidae) is one of the most frequently released natural enemies for biological control of spider mites in greenhouse and outdoors crops. In this research, I utilized Expresses Sequence Tags (ESTs), the most cost effective approach for transcriptome exploration, to study three different aspects of this arachnid species for which there is little genomic information. I combined two EST datasets from different whole body cDNA libraries and analyzed by bioinformatics means. Approximately 54% of 10,256 uniESTs were annotated based on the homology to sequences in the National Center for Biotechnological Information (NCBI) database. A list of these uniESTs, sorted from most to least likelihood based on the expected value from the blast search in public databases, was used to create tools for each of the three studies. First, I described sixty-one genes encoding products known to be important in pesticide metabolism and in endocrinology, including cytochrome P450s, glutathione-S-transferases, acetylcholinesterase homologs, neuropetides and neurohormones. Findings on arachnid specific esterases and neuropetides, and possible benefits to pest management programs, were discussed. Next, I inferred divergence time for Acari and the point of divergence of two lineages within anactinotrichid mites, *Ixodes scapularis and Phytoseiulus persimilis.* I used expresses sequence tags from the predatory mite *P. persimilis* to pull out 74 orthologous amino acid sequences of invertebrates species: nine insect species, *Daphnia pulex*, *Ixodes scapularis*, *and Caenorhabditis elegans*. I estimated a similar origin for Chelicerata (578.1 ± 38.2 - 482.2 ± 7.2 Mya) as in other recent studies. However, divergence dating using amino acid sequences suggested a Devonian origin of anactinotrichid mites (487.6 ± 32.2 - 410.1 ± 6.1 Mya) based on four reference dates (two fossil records and two molecular clocks) and four amino acid substitution methods; this estimate is much earlier that those in the current literature. This discrepancy of divergence times may be due to the use of a global clock. Finally, I developed molecular markers from the EST dataset to examine inheritance in the haplodiploid system in *P. persimilis*. Biparental contribution of chromosomes is required among the predatory mites but the paternal chromosome set seems to be eliminated or loss (Paternal genome loss, PGL) in male offspring. However, genetic studies in other two phytoseiid species were suggested diploid males with PGL only in the germ cells. In the present study, haploid adult males of *P. persimilis* have been observed using five independent EST-derived markers. Single

mites derived from inter-population crosses were genotyped after whole genome amplification. The parahaploid genetic system in *P. persimilis* is supported by this study, in which both sexes arise from fertilized eggs but the paternal chromosome set is subsequently lost in males.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I want to thank first my co-major advisors Dr. Margolies and Dr. Park, who I deeply admire and grateful for their guidance and extraordinary patience. I have been blessed for the opportunity to experience from two laboratories with very different directions, and have had the chance to learn from the diversity at each site. I also want to thank my committee members Dr. Jeremy Marshal and Dr. Mark Ungerer for the improvement of my thesis through their insightful comments and recommendations. I am very grateful to the Bioinformatics Center at Kansas State University, to Dr. Susan J. Brown, Dr. Doina Caragea, and Dr. Sanjay S. Chellapilla for their help and collaboration in various steps of my research. Also, I thank Dr. Brian Ellis in the Michael Smith Laboratories at the University of British Columbia for sharing the expresses sequence tag data set. Also, my acknowledgements to Xiaoli Wu for her teaching in mites rearing and maintenance techniques as well as collaboration in my experiments.

In my research journey at K-State I have met wonderful people that contributed to my degree completion through their understanding, support, and help. I want to thank Dr. Jim Nechols for his logistic support in my transition to the mite's project. Thanks to my friends/family at K-State: Mukta Pahwa, Xiaoli Wu, Bin Li, Ladislav Simo, Khurshida Begum, Punya Nachappa, and my lab mates from both laboratories. I am grateful to Alberto and Bobby Broce for their caring and support. To my friends at KSU, "los ticos" and those who kindly shared your joy and spirit with me, thank you. Finally, to my parents, siblings and Gilbert, thanks for your support and understanding through all these years, for carrying me in your heart and your prayers, and the best long-distance logistic support ever.

Chapter 1 - Introduction

The predatory mite, *Phytoseiulus persimilis* Athias-Henriot, is a specialist on herbivorous tetranychid mites (McMurtry and Croft, 1997) and one of the most widely-used and effective phytoseiid species for biological control of twospotted spider mites, *Tetranychus urticae* Koch on food and ornamental crops (Gilstrap and Friese, 1985; Sabelis, 1985; Sabelis and van der Meer, 1986 cited in Grostal and Dicke, 1999; Van Lenteren and Woets, 1988). Because of its importance in pest management, a large amount of biological information about *P. persimilis* has been developed over the last 40 years (Sabelis, 1985), making this species one of the best known biological control agents. *P. persimilis* originated in Chile and was introduced to Germany in 1958, from where it was soon disseminated to and became established in other parts of the world (Dosse, 1958). The life history patterns of *P. persimilis* depend on the presence of their prey, which they require for survival, development and reproduction (McMurtry and Croft, 1997). Perhaps because of this dependence, *P. persimilis* has evolved ecological and behavioral traits that make them to be one of the most effective natural enemies of the spider mites, including having a short development time, high rate of prey consumption, and search behavior attuned to cues of the prey and the prey host plants (Jarošík, 1990; Devonshire and Fielde, 1991; Zhang and Sanderson, 1995). The predatory mites are most commonly used in an inundated manner for biological control (McMurtry and Croft, 1997), which requires periodic releases in large numbers. Due to its tropical origin, *P. persimilis* does not have a diapause stage, unlike other predatory mites used in biological control, such as *Galendromus* (=*Metaseiulus*) *occidentalis*. Therefore, it is active year-round even in template zones when used against spider mites in enclosed habitats such as interior plant landscapes and in greenhouses (Hajek, 2004). Some of these traits have been demonstrated to be heritable in phytoseiids, i.e. diapause, rate of prey consumption, short-range dispersal, efficiency of conversion from prey item to reproduction (Nachappa et al., 2009); hence, selection to enhance predator efficiency may be possible possible (Hoy, 1986). In addition to predator efficiency, pesticide resistance is another possible target for improving biological control agents. Heritable pesticide resistance has been observed in populations of *P. persimilis*, allowing for selection for resistant strains (i.e. methidathion, Fournier et al. 1987). One to few major genes seem to determine pesticide resistance in *G. occidentalis* (Hoy, 1987) and a stable genetic transformation is already possible using the

candidate gene approach (Presnail et al., 1997; Hoy, 2000). Hence, genetic enhancement of the predatory mite *P. persimilis* is being considered to improve pest management.

Currently, genomic approaches are being applied to the study and enhancement of some biological control agents to complement up-dated pest management programs (Roderick and Navajas, 2003; Al-Tawaha et al., 2008). *P. persimilis* is relatively easy to maintain in laboratory conditions and has a short life cycle (~5 days from egg to adult, Abad-Moyano et al., 2009), which helps to fit into a genetic enhancement program. However, little genetic, let alone genomic information is available for *P. persimilis*. Research on mite genetic enhancement and transformation is still in the pre-genomic era; whole genome sequence data from predatory mites is not yet available. However, nucleotide and amino acid sequences from various species of predatory mites are being uploaded sporadically to public databases. Although we may not be able to obtain whole genomes in the near future, the use of Expresses Sequence Tags (ESTs) may be a suitable alternative method, as was first proposed for sequencing the complete human genome (Adam et al., 1991). Many researchers considered this as a cost-effective approach to increase the amount of genetic information for organisms lacking genomic information, such as the predatory mite *P. persimilis*. Therefore, the rationale for my study is to pursue the development of molecular tools and candidate genes of interest for future research on this valuable predatory mite. To this end, I have taken ESTs derived from whole body *P. persimilis* and applied them to: 1) identification of candidate genes for pesticide resistance studies, 2) dating Acari divergence, and 3) development of molecular markers and observations for the genetic system in *P. persimilis*. In the rest of this chapter I will briefly introduce the components to the approaches I have taken using the ESTs.

Expressed sequence tags

Expresses sequence tags (ESTs) are partial sequences of a transcript randomly selected from a complementary DNA (cDNA) library. These DNA sequences, usually about 300-500 base pairs long, are typically produced in large batches that can be used to assemble fragments of the gene. ESTs are useful in new gene discovery and genome mapping and identification of coding regions, (see Hatey et al., 1989 for amore complete review of the uses of ESTs).

Currently, ESTs are the most cost-effective approach for transcriptome exploration (Nagaraj et al., 2007). As high-throughput sequencing technology and its analysis methods through bioinformatics become widely available, the production of ESTs is becoming accessible for almost all type of organisms (Church, 2006; Nagaraj et al., 2007). Sequence tags are informative not only for the species from which mRNA was used to develop the cDNA library, but comparative analysis between tissues, life stages, and even between species can be performed, and sharing of these biological sequences is possible through public databases. Most EST projects publish their sequences in databases such as dbEST (database of Expresses Sequence Tags) and GeneBank, turning ESTs into one of the largest resources for biological sequence data for identification of gene products, for gene annotation and mapping, and discovery of potential variations (polymorphisms) to develop molecular markers (Boguski et al; 1993).

ESTs as resource for gene discovery and candidate genes

Expressed sequence tags are currently being used in the study of several species of economic important mites, such as the dust mite (Ljunggren et al., 2003), to isolate gene products that could be used for control strategies (Angus et al. 2004; Holt et al., 2004). I have followed this same approach in my research to find important genes for life processes and candidate genes that might be of benefit for pest control. Once the partial sequences of the gene(s) are found, their function might be surmised by comparison to similar sequences in the genomic databases using bioinformatics. For example, ESTs have been a good resource for finding genes that are potentially involved in pesticide resistance (Khajuria et al., 2009). Several genes involved in pesticide resistance have been observed among predatory mites. Other genes related to neurohormones and cuticle formation have also been found. My goal was to identify those candidate genes from a *P. pesimilis* EST dataset.

Pesticide resistance in arthropods may occur through a variety of means. The most common mechanism conferring pesticide resistance to insects and mites is metabolic alteration of xenobiotics. It results in detoxification either by increasing enzyme titer, producing more efficient enzymes, or both. The most common enzymes involved in this form of resistance are cytochrome P450 (CYP) mono oxygenases, hydrolytic esterases, and glutathione S-transferases

(GST). However, different mechanisms may be used for insecticides with same mode of action. For example, in two cases in phytoseiid mites in which resistance has been attributed to acetylcholinesterase inhibitors, resistance to methidathion (organophosphate) in *P. pesimilis* is due to GST activity through gene amplification (Fournier, et al, 1988; Devonshire and Fielde, 1991), while in *G. occidentalis* enhanced CYP in response to carbaryl (carbamates) is responsible for resistance (Knowles, 1997). If it were possible to identify the genes involved in the resistance mechanisms in *P. persimilis*, it might then be possible to track the overall physiological response and determine the pathways involved in resistance. This information can support genetic enhancement of beneficial mites that could be used in an integrated fashion in agricultural situations.

ESTs in development of molecular markers to observe pseudoarrhenotoky

Another potential use for ESTs is for development of molecular markers, as ESTs provide distinctive landmarks produced by polymorphism in the nucleotide sequence, which could be used to distinguish between individual genomes. Most polymorphisms are caused by single nucleotide changes, insertion/deletion, or simple sequence repeats, and these can be identified when ESTs are overlapped and arranged into clusters. These elements are useful for identification of and distinction between individuals or between populations; ESTs have already proven useful for marker development in agriculture and evolutionary studies (Pashley, 2006; Ayeh, 2008). I used this approach in my research to develop a tool to assess haplodiploidy in male *P. persimilis*.

P. persimilis is a bisexual species in which males are haploid (N) and females are diploid (2N) (Wysoki & Swirski, 1968; Schulten et al., 1978). Four short cephalobrachial chromosomes are observed in the male karyotype, while females possess eight chromosomes. This cytologenetic condition initially led to the assumption of arrhenotoky in phytoseiids, in which unfertilized eggs develop into haploid males (Olivier Jr., 1977). However, subsequent work has shown genetic evidence of biparental contribution in both sexes in two predatory mites, *Amblyseius bibens* and *P. persimilis* (Helle et al., 1978). Furthermore, mating is required for oviposition in phytoseiids (Schulten, 1985; Norton et al., 1993; Sabelis and Nagelkerke, 1993).

This genetic system is called pseudoarrhenotoky. Under it, the haploid condition in males seems to be achieved by heterochromatinization of the paternal set of chromosomes during early embryo development (Nelson-Rees et al., 1980; Schulten, 1985). However, it is not certain whether paternal genome is eliminated from all tissues (Sabelis and Nagelkerke, 1993); some karyotyped cells from central nervous tissues in males appear with diploid chromosomes, and there is evidence of paternal genome in male somatic cells from genetic marker studies (Perrot-Minnot et al. 2000). The remaining diploid cells could have important physiological effects in the males but the trait may not be inherited by the next generation. Implications of the mite's genetic system in the build up of pesticide resistance and their maintenance have been discussed by Sabelis et al. (2002). An understanding of pseudoarrhenotoky has been useful in explaning uniparental inheritance of miticides resistance in *G. occidentalis* (Hoy, 1977). A better understanding of the reproduction in phytoseiid mites would assist the design and use of effective transgenic predators.

Pseudoarrhenotoky in two species of phytoseiid mites was observed using two types of PCR based marker systems: RAPD (random amplification of polymorphic DNA) in *Typhlodromus pyri* males (Perrot-Minnot and Navajas, 1995) and DALP (direct amplification of length polymorphism) in *Neoseiulus californicus* (Perrot-Minnot et al., 2000). In both studies, arbitrary primers were used to generate genomic fingerprints to identify paternal gene loss. Both methods offer relatively fast and inexpensive means to detect polymorphisms, however their use is limited because they can suffer from poor reproducibility in different labs, especially in the RAPD markers (Black, 1993). The utility of ESTs for marker development has been demonstrated in ecology, behavior and genetics (see review in Behura, 2006), and I used them to study the genetic system in *P. persimilis*.

Estimating Acari divergence time using ESTs

Mites are considered amongst the oldest terrestrial animals (Norton et al., 1988). Time of divergence time of the Acari from the ancestral lineage has been determined mainly by fossil records and from molecular sequences. However, the latest Acari paleontological update has shown a conflict between estimates of divergence time using fossil dating and that inferred from

molecular clocks through phylogenetic analysis. In fossil data the origin of the mites was suggested to occur in the Ordovician (ca. 475 Mya; Bernini et al. 2002; see also Dunlop and Selden 2009), while dating using contemporary molecular clock methods based on mitochondrial genes suggested a later departure of ticks and mites from the lineage of sea spiders (Pycnogonida), around 424 ± 21 Mya (Jeyaprakash and Hoy, 2009). This discrepancy is a reversal of the usual cases in which divergence time inferred from molecular clocks are usually much earlier than that suggested by missing fossil records. In general, paleontological data sets the limits for evolutionary events and is considered to be the best reference parameter to calibrate molecular clocks (Donoghue and Benton, 2007). However, their use carries the assumption of an accurate fossil record dating and correct identification (Dunlop and Selden 2009), and is subject to change due to discovery of more ancient evidence. Fossil deposits of soft body organisms are harder to find in the stratification deposits than those with hardened skeletal structures, and environmental factors also affect the fossil formation and conservation in generating gaps in the stratification records (Kidwell and Holland, 2002). Therefore, the stratification data is mostly used to set minimum age constraints (Marshall, 2008). Use of molecular data, on the other hand, is tempered by widespread heterogeneity in molecular evolutionary rates among and within lineages (Britten, 1986) which seriously compromise the assumption of uniform rate across lineages (Zuckerkandl and Pauling 1962, 1965). In this respect, two major concerns in the use of molecular clocks are: 1) choosing the reference for calibration, either from fossil records or from other molecular clocks, and 2) the type (i.e. DNA, amino acid sequences) and size of informative biological sequences being used for the considered evolutionary distance (i.e species, family, phylum level).

In Acari, mitochondrial genes and nuclear ribosomal RNA have been used for phylogenetic analyses from population to phylum level (Navajas and Fenton, 2000; Cruickshank, 2002). However, Cruickshank (2002) suggested that nuclear protein-coding genes may be more suitable than mitochondria genes for inferring intermediate to deeper phylogenetic relationships, noting possible difficulties due to the high AT content in 12S and 16S rDNA, emphasizing the faster evolutionary rate in mitochondrial genes in comparison to nuclear protein-coding genes. EST projects on *P. persimilis* developed by Dr. B. Ellis' lab (University of British Columbia) and in-house (Drs. D.C. Margolies and Y. Park, Kansas State University) have provided a good

source of data to look for orthologous sequences (single genes in different species that are descended from the same ancestral gene) between taxa. In my research, I searched for 1:1 orthologous genes between P. persimilis and insects, crustaceans, and ticks to conduct a phylogenetic analysis of these arthropod clades. In addition, multiple fragments of all 74 amino acid sequences were concatenated after trimming each individual sequence to generate a "super" sequence for alignment. This is considered a better approach than using single genes because, while different evolutionary rates may occur for each gene, concatenation evens the rates and produces a more resolved phylogeny than using single genes (Rokas et al., 2003; Gadagkar et al., 2005). I used both fossil estimates and molecular clocks as reference time points for the analysis.

Objectives

The major goal of this study was to develop genomic tools for future research in molecular aspects of the predatory mite *P. persimilis*. I then demonstrated the usefulness of these tools to address specific questions about *P. persimilis*. Specifically, I took the ESTs derived from a cDNA library made from whole *P. persimilis* to explore for candidate genes for genetic enhancement and molecular marker development, as well as the findings for ortholog genes.

The objectives of this study were to:

1) Develop and analyze an EST dataset and then search the database for candidate genes related to novel pesticides and pesticide resistance.

2) Explore for orthologous genes within the EST dataset to allow estimation of divergence times within Acari. A phylogenetic analysis was performed to determine substitution rate and branch distance in order to extrapolate the point of departure from the common ancestor, using both paleontological as well as molecular references.

3) Develop genetic markers for *P. persimilis* and also explore molecular tools to confirmation haplodiploidy in this species of mites. Genetic crosses were performed to observe

inheritance while genetic markers and sequence amplification techniques were developed in parallel for quantifications.

References

Abad-Moyano, R.; T. Pina; F. Ferragut; and A. Urbaneja. 2009. Comparative life-history traits of three phytoseiid mites associated with *Tetranychus urticae* (Acari: Tetranychidae) colonies in clementine orchards in eastern Spain: implications for biological control. Experimental Applied Acarology 47: 121-132.

Adams, M.D.; J.M. Kelley; J.D. Gocayne; M. Dubnick; M.H. Polymeropoulos; H. Xiao; C.R. Merril; A. Wu; B. Olde; R.F. Moreno et al. 1991. Complementary DNA sequencing: Expressed sequence tags and the human genome project. Science 252: 1651-1656.

Al-Tawaha, A.R.M.; N. Odat; A.L. Gzawi; M.M. Al-udatt; M. Turk; and F. Ababneh. 2008. Genomic and chemical approaches to weed control in pasture. American-Eurasian Journalagriculture and environmental science 3: 187-193.

Angus, A.C.; S.T. Ong; and F.T. Chew. 2004. Sequence tag catalogs of dust mite-expressed genomes: utility in allergen and acarologic studies. American Journal of Pharmacogenomics 4: 357-69.

Ayeh, K.O. 2008. Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): emerging molecular maerker tools for improving agronomic traits in plant technology. African Journal of Biotechnology 7: 331-341.

Behura, S.K. 2006. Molecular marker systems in insects: current trends and future avenues. Molecular Ecology 15: 3087–3113.

Bernini, F.; G. Carnevale; G. Bagnoli; and S. Stouge. 2002. An early Ordovician oribatid mite (Acari: Oribatida) from the island of Öland, Sweden. pp 45–47. In: Acarid phylogeny and evolution. Adaptations in mites and ticks. F. Bernini, R. Nannelli, G. Nuzzaci, E. de Lillo (eds). Kluwer, Dordrecht.

Black, W.C. 1993. PCR with arbitrary primers: approach with care. Insect Molecular Biology 2: 1–6.

Boguski, M. S.; T.M. Lowe; C.M. Tolstoshev. 1993. dbEST-database for expressed sequence tags. Nature Genetics 4: 332-333.

Britten, R.J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393-1398.

Church, G.M. 2006. Genomes for all. Scientific American 294: 46-54.

Croft, B.A and H.E. Van de Baan. 1988. Ecological and genetic factors influencing evolution of pesticide resistance in tetranychid and phytoseiid mites. Experimental and applied acarology. 4: 277-300

Cruickshank R.H. 2002. Molecular markers for the phylogenetics of mites and ticks. Systematics Applied Acarology 7: 3–14.

Devonshire, A.L. and L.M. Field. 1991. Gene amplification and insectide resistance. Annual review entomology 36: 1-23.

Donoghue P.C.J and M.J. Benton. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. TRENDS Ecology Evolution 22: 424-431.

Dosse, G. 1958. Uber einige neue Raubmilbenarten (Acarina:Phytoseiidae). Pflanzenschutzber-Berichte 21: 44-61.

Dunlop, J. A. and P.A. Selden. 2009. Calibrating the chelicerate clock: a paleontological reply to Jeyaprakash and Hoy. Experimental and Applied Acarology 48: 183 -197.

Fournier, D.; M. Pralavorio; A. Cuany; and J.B. Berge.1988. Genetic analysis of methidathion resistance in *Phytoseiulus persimilis* (Acari: Phytoseiidae). Journal economic entomology 81: 1008-1013.

Fournier, D; A. Cuany; M. Pralavorio; J. M. Bride; and J. B. Berge. 1987. Analysis of methidathion resistance mechanisms in *Phytoseiulus persimilis* A.H. Pesticide Biochemistry and Physiology 28: 271-278.

Gadagkar, S.R.; M. S. Rosenberg; and S. Kumar. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus. Journal of Experimental Zoology 304B: 64–74.

Gilstrap, F.E. and D.D. Friese. 1985. The predatory potential of *Phytoseiulus persimilis*, *Amblyseius californicus*, and *Metaseiulus occidentalis* (Acarina: Phytoseiidae). International Journal Acarology. 11: 163–168.

Grostal, P and M. Dicke. 1999. Recognising one's enemies: a functional approach to risk assessment by prey. Behavior, ecology and sociobiology 47: 258-264.

Hajek, A.E. 2004. Natural Enemies: An Introduction to Biological Control. Cambridge University Press, Cambridge, UK. 378 pp.

Hartl, D.L. 1971. Some aspects of natural selection in arrhenotokous populations. American zoologist 11: 309-325.

Hatey, F.; G. Tosser-Klopp; C. Clouscard-Martinato; F. Musant; F. Gasser. 1998. Expressed sequence tags for genes: a review. Genetics Selection Evolution. 30: 521-541.

Helle, W.; H.R. Bolland; R.V. van Arendonk; R. De Boer; G.G.M. Schulten; and V.M. Russell. 1978. Genetic evidence for biparental males in haplo-diploid predator mites (Acarina: Phytoseiidae). Genetica 49: 165-171.

Holt, D. C.; K. Fischer; S. J. Pizzutto; B. J. Currie; S. F. Walton; and D. J. Kemp. 2004. A Multigene Family of Inactivated Cysteine Proteases in *Sarcop*tes scabiei. The Journal of Investigative Dermatology 123: 240-241.

Hoy, M.A. 1977. Inbreeding in the arrhenotokous predator *Metaseiulus occidentalis* (Nesbitt) (Acari: Phytoseiidae). International journal of acarology 3: 117-121.

Hoy, M. A. 1987. Developing Insecticide Resistance in Insect and Mite Predators and Opportunities for Gene Transfer. Pp: 125-138. In: Biotechnology in agricultural chemistry. ACS Symposium Series, Vol. 334. H. M. LeBaron; R. O. Mumma; R. C. Honeycutt; J. H. Duesing; J. F. Phillips; and M.J. Haas (eds). American Chemical Society.

Hoy, M. A. 1992. Biological control of arthropods: genetic engineering and environmental risks. *Biological Control* 2:166-170.

Hoy, M.A. 2000. Transgenic arthropods for pest management programs: risks and realities. Experimental Applied Acarology 24: 463-395.

Huelsenbeck, J.P.; F. Ronquist; R. Nielsen; and J.P. Bollback. 2001. Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. Science 294: 2310-2314.

Jarošík V. 1990. *Phytoseiulus persimilis* and its prey *Tetranychus urticae* on glasshouse cucumbers and peppers: key factors related to biocontrol efficiency. Acta Entomologica Bohemoslovaca 87: 414-30.

Kidwell, S.M. and Holland, S.M. 2002. The quality of the fossil record: implications for evolutionary analysis. Annual Review of Ecology and Systematics 33: 561-588.

Khajuria, C.; Y. C. Zhu; S.M. Chen; L. L. Buschman; R. A. Higgins; J. Yao; A.L.B. Crespo; B. D. Siegfried; S. Muthukrishnan; and K. Y. Zhu. 2009. Expressed sequence tags from larval gut of the European corn borer (*Ostrinia nubilalis*): Exploring candidate genes potentially involved in *Bacillus thuringiensis* toxicity and resistance. *BMC Genomics* 10: 286-299.

Ljunggren, E.L.; D. Nilsson; and J. G. Mattsson. 2003. Expressed sequence tag analysis of *Sarcoptes scabiei*. Parasitology 127: 139-145.

Knowles, C.O. 1997. Mechanisms of resistance to acaricides. Pp: 57-77. In: Molecular Mechanisms of Resistance to Agrochemicals vol. 13. V. Sjut (ed). Springer, Berlin Heidelberg.

McMurtry, J. A. and B. A Croft. 1997. Life styles of phytoseiid mites and their roles as biological control agents. Annual Review of Entomology 42: 291-321.

Nachappa, P; D. C. Margolies; J. R. Nechols; and T. Morgan. 2009. Response of a complex foraging phenotype to artificial selection on its component traits. Evolutionary Ecology 24: 631-655.

Nagaraj S.H.; N. Deshpande; R.B. Gasser; and S. Ranganathan. 2007. ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. Nucleic Acids Research 35: W143-W147.

Navajas, M. and B. Fenton. 2000. The application of molecular markers in the study of diversity in acarology: a review. Experimental and Applied Acarology 24: 751-774.

Nelson-Rees, W.A.; M.A. Hoy, M.A.; and R.T. Roush. 1980. Heterochromatinization, chromatin elimination and haploidization in the parahaploid mite *Metaseiulus occidentalis* (Nesbitt) (Acarina: Phytoseiidae). Chromosoma 77: 263–276.

Norton, R.A.; P.M. Bonamo; J.D. Grierson; W.A. Shear. 1988. Oribatid mite fossils from terrestrial Devonian deposit near Gilboa, New York. Journal of Paleontology 62: 259-269.

Norton, R.A.; J. B. Kethley; D.E. Johnston; and B.M. Oconnor. 1993. Phylogenetic perspectives on genetic systems and reproductive modes of mites. Pp. 8–99. In: Evolution and diversity of sex ratio in insects and mites. D.L. Wrensch and M.A. Ebbert (eds). Chapman & Hall, New York.

Oliver, J.H., Jr. 1977. Cytogenetics of mites and ticks. Annual Review of Entomology 22: 407-429.

Oppenoorth, F.J. 1965. Biochemical genetics of insecticide resistance. Annual Review of Entomology 10: 185-206.

Pashley, C; J. R. Ellis; D.E. McCauley; and J.M. Burke. 2006. EST databases as a source for molecular markers: Lessons from *Helianthus*. Journal of Heredity 97: 381–388.

Penman, D.R and R. B. Chapman. 1988. Pesticide-induced mite outbreaks: pyrethroids and spider mites. Experimental and Applied Acarology 4: 265-276.

Perrot-Minnot, M.J.; J. Lagnel; A. Migeon; and M. Navajas. 2000. Tracking paternal genes with DALP markers in a pseudoarrhenotokous reproductive system: biparental transmission

but haplodiploid-like inheritance in the mite *Neoseiulus californicus*. Heredity 84: 702–709.

Presnail, J.K.; A. Jeyaprakash; J. Li; and M.A. Hoy. 1997. Genetic analysis of four lines of *Metaseiulus occidentalis* (Nesbitt) (Acari: Phytoseiidae) transformed by maternal microinjection. Annual Entomological Society America 90: 237-245.

Roderick, G and M. Navajas. 2003. Genetics and evolution in biological control. Nature Reviews Genetics 4: 889 - 899.

Rokas, A; B.L. Williams; N. King; S.B. Carroll. 2003. Genome scale approaches to resolving incongruence in molecular phylogenies. Nature 425: 798–804.

Sabelis, M. W. and C.J. Nagelkerke. 1993. Sex allocation and pseudoarrhenotoky in phytoseiid mites. Pp: 512–541. In: Wrensch, D. L. and Ebbert, M. A. (eds) Evolution and Diversity of Sex Ratio. Chapman & Hall, New York.

Sabelis, M.W. 1985. Predation on spider mites. Pp: 103–129. In: W. Helle and M.W. Sabelis (eds) Spider Mites: Their Biology, Natural Enemies and Control, Vol. 1B. Elsevier, Amsterdam.

Sabelis, M.W. and C.J. Nagelkerke. 1988. Evolution of pseudo-arrhenotoky. Experimental and Applied Acarology 4: 301-318.

Sabelis, M.W., C.J. Nagelkerke, and J.A.J. Breeuwer. 2002. Sex ratio controls in arrhenotokous and pseudo-arrhenotokous mites. Pp: 235 - 253. In: Sex ratios: concepts and research methods. I.C. W. Hardy (ed). Cambridge University Press, Cambridge.

Schulten, G. G. M., R. C. M. Van Arendonk, V. M. Russell, and F. A. Roorda. 1978. Copulation, egg production and se-ratio in *Phytoseiulus persimilis* and *Ambliseius bibens* (Acari:Phyroseiidae). Entomologia Experimentalis and Applicata 24:145-153.

Van Lenteren, J.C. and J. Woets. 1988. Biological and integrated pest control in greenhouses. Annual Review of Entomology 33: 239-269.

Whitten, M.J. and M.A. Hoy. 1999. Genetic improvement and other genetic considerations for improving the efficacy and success rate of biological contro. Pp: 271-296. In: Handbook of Biological control. T.S. Bellows and T.W. Fisher (eds). Academic press, San Diego. CA.

Wysoki, M. and Wirski, E. 1968. Karyotypes and sex-determination of ten species of phytoseiid mites (Acarina: Mesostigmata). Genetica 39: 220-228.

Zhang, Z. Q. and J.P. Sanderson. 1995. Two-spotted spider mite (Acari: Tetranychidae) and *Phytoseiulus persimilis* (Acari: Phytoseiidae) on greenhouse roses: spatial distribution and predator efficacy. Journal Economical Entomology 88: 352–57.

Zucherkandl, E. and L. Paulong. 1962. In: Horizons in Biochemistry. Pp. 189. M. Kasha and B. Pullman (eds). Academic Press, New York.

Zuckerkandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. Journal Theroretical Biology 8: 357-366.

Chapter 2 - Expressed sequence tags from the predatory mite *Phytoseiulus persimilis* Athias-Henriot (Parasitiformes, Mesostigmata, Phytoseiidae) reveal potential new targets for improvement of biocontrol

Abstract

The predatory mite *Phytoseiulus persimilis* (Acari, Phytoseiidae) is one of the most frequently released natural enemies for biological control of spider mites in greenhouse and outdoors crops. The possibility of using transgenic technologies to enhance tolerance and resistance to selective pesticides in predatory mites has been proposed, but awaits a better understanding of resistance mechanisms and identification of underlying genes underlying resistance. In this chapter, I explored a large collection of *P. persimilis* expressed sequence tags (ESTs) to uncover candidate genes involved in pesticide resistance and neurobiology and to develop a resource for future research. A total of 10,256 uniESTs were generated, from which 54% were annotated based on the homology to the sequences in the National Center for Biotechnological Information (NCBI) database. I found sixty-one genes encoding products known to be important in pesticide metabolism and in endocrinology, such as CYP, GST, and ACE homologs. A putative arachnid-specific esterase sub-family was identified from a phylogenetic approach to determine *Pp* ACE homologs from the dataset. Fourteen putative neuropeptides and protein hormones are listed with mature neuropeptides sequence identified. Functions of the neuropeptides and peptide hormones are categorized by their role in physiology in the context of potential application in pest control measures. Two *ccap*-like transcripts were found in the *Pp* uniESTs, which mature sequence of 10 and 9 were identified. This information offers a start in redressing the imbalance in knowledge between pest species and their natural enemies. A better understanding of the physiological pathways underlying pesticide resistance may strengthen biological control strategies, and ultimately allow genetic enhancement of beneficial organisms.

Background

Concern about the use of toxic agrochemicals has stimulated the study of biological control agents as part of pest management strategies (Jensen and Malter, 1995). Increasingly, it has become apparent that pest management may require a combination of biological and pesticide control. However, these tactics may not be compatible because pesticides can negatively affect natural enemies used for biological control. Thus, integration of management methods has become an issue. An example of this involves use of the predatory mite *Phytoseiulus persimilis* (Acari, Phytoseiidae), which is one of the most frequently released natural enemies for biological control of spider mites in greenhouse and outdoors crops (Van Lenteren and Woets, 1988). The preferred prey species for *P. persimilis* is the twospotted spider mite, *Tetranychus urticae* Koch (Walzer et al., 1999), an ubiquitous, generalist herbivore that affects over 150 plant species throughout the world, from food and fiber crops to ornamental plants in greenhouses, gardens, home landscapes and displays (van de Vrie et al., 1972; Jeppson et al., 1975). Acaricides and insecticides have been the preferred method for controlling spider mites, but this has resulted in the rise of pesticide resistant populations (Croft and Van de Baan, 1988). Thus, biological control might be an important component of resistance management programs designed to maintain current and future chemical control options. One approach to integrating biological control into resistance management is to enhance pesticide resistance in predatory mites, which has been of interest since field tolerance to parathion was first documented in *Galendromus* (= *Metaseiulus*) *occidentalis* (Huffaker and Kennett, 1953). Resistance to at least seven active ingredients of pesticides (deltamethrin, demeton-S-methyl, diazinon, dichlorvos, ethion, methidathion, and parathion) has been also been documented in laboratory strains of *P. persimilis* (Whalon et al., 2004-2010; Arthropod Pesticide Resistance Database). Application of transgenic technologies might be used to enhance the resistance and persistence of *P. persimilis* in integrated pest management programs to control spider mite (Hoy, 1995) if mechanisms of resistance and the underlying genes could be identified. Unfortunately, the lack of genomic information about predatory mites is one of the major limitations in the development of genetic enhancement projects (Hoy, 2009).

Arthropods use a variety of means to resist pesticides, including metabolic degradation of toxicants, target site insensitivity, alteration of absorption and allocation of the toxins, and

changes in behavior (Clark and Yamagushi, 2002). Metabolic alteration, either by quantitative or qualitative changes in the enzyme activity toward the toxins, seems to be the most common mechanism conferring pesticide resistance in insects and mites. The major families of enzymes involved in metabolic resistance in insects and mites are the hydrolytic esterases, cytochrome P450 monooxygenases (CYPs), and glutathione S-transferases (GSTs) (Hemingway et al., 1998). Among phytoseiid mites, for instance, resistance to carbaryl (carbamates) in *G. occidentalis* and to methidathion (organophosphate) in *P. pesimilis* has been associated with detoxification via CYP and GST, respectively (Roush and Plapp, 1982; Fournier et al., 1987; 1988). On the other hand, insensibility of ACE has been observed in other species of Acari, *Galendromus* (=*Typhodromus*) *pyri* and *Amblyseius potentillae*, in which resistance to parathion (organophosphate) and propoxur (carbamate) were due to allelic variation in ACE isoforms (Knowles, 1997). Identifying the molecules involved in pesticide resistance in *P. persimilis* would provide insight into plausible mechanisms of resistance that could be exploited in the development of a stable and sustainable predatory mite compatible with pesticide usage.

In addition to the direct neural system targeting represented by most insecticides, (Hammock and Soderlund, 1986), other potential targets in the insect nervous system such as the neuropeptidergic systems are gaining attention for pesticide development (Maule et al., 2002; Fónagy, 2006). Peptide hormones and neuropeptides are major and vital components required for maintaining homeostasis and control of behavioral and physiological events in Metazoa. Identifying acari-specific peptidergic signaling system could provide new opportunities for development of selective acaricides in the long term, and also prevent or diminish the deleterious effects of pesticides in non-target organisms such as natural enemies used in biological control programs.

In this study, I explored a large collection of *P. persimilis* expressed sequence tag (EST) gene sequences to uncover candidate genes involved in pesticide resistance and neurobiology and to develop a resource for future research. I report on the analysis and annotation of the sequences derived from 15,960 complementary DNA (cDNA) clones from the whole body of the predatory mites. The EST data revealed many genes encoding the products known to be important in pesticide metabolism and in endocrinology, such as CYP, GST, and ACE

homologs. I have also conducted a more detailed analysis of sixty-one *Pp* uniESTs belonging to these classes, fourteen of which encode putative neuropeptides and protein hormones.

Material and Methods

cDNA library construction

The data presented in this chapter are from two independent cDNA libraries constructed using mixed age and gender *Phytoseiulus persimilis*. One cDNA library, developed at Kansas State University, was made using about 2000 individual mites provided by Koppert Biological Systems (Romulus, Michigan, USA). The second was prepared at the University of British Columbia from mites reared on site. In both cases, the mite samples contained a mixture of ages and genders, to maximize the representation of biologically relevant mRNAs. Total RNA isolated from both samples by using the Trizol reagent (Invitrogen), and sent to Evrogen (Ru JSC, Moscow, Russia) for cDNA library development. The cDNA library was enriched for full-length DNA by using a SMART kit (Clonetech),, normalized using the DSN normalization method (Zhulidov et al., 2004), and ligated into the pAL16 vector. The ligated products were transformed into *Escherichia coli* (XL1-MR blue, Stratagene), and a total of 15,960 clones (15,000 from UBC and 960 from KSU) were sequenced from both endsby either the Michael Smith Genome Sciences Centre (Vancouver, Canada) or University of California-Riverside Sequencing Facility (Riverside CA, USA)

Sequence analysis

The data obtained from bi-directional sequencing of clones from the two EST libraries (31,673 raw data sequences)) was processed by ArthropodEST (http://bioinformatics.ksu.edu/ArthropodEST/), an EST analysis pipeline, using default settings for sequence trimming, cleaning, vector removal and contaminant screening. The ArthropodEST pipeline includes repeat masking using an arthropod clade as the reference, assembly by using CAP3 (Huang and Madan, 1999), and annotation by the Blast2Go program (Conesa et al., 2005;

GO database January 2010, pipeline version 2.3.5) with the optional parameter for blast search in NCBI BLASTX in the nr database set for <1E-04.

Data mining and analysis of sequence

Transcripts encoding putative homologs of ACE, glutathione S-transferases (GSTs), cytochorme P450s (CYPs), a nicotinic acetylcholine receptor, chitin deacetylases, and chitinases were initially identified by sequence similarity captured in the Balst2Go analysis, considering only those sequences within a BLASTX cutoff of *E*-value of 10E-10. When the sequence similarity alone did not provide strong support for the annotation, additional evidence was sought in the InterProScan (EMBL-EBI; Zdobnov and Apweiler, 2001) and the putative translations of the uniEST sequences were subjected to further search in PANTHER (Protein ANalysis THrough Evolutionary Relationships library version 6; Mi et al., 2005). Each amino acid sequence was searched against statistical models (HMMs) obtained from a set of "training sequences", and a score indicating its relatedness to the model protein is given (close related > E-23, related = E-23 < E-11, distantly related E-11< E-3).

Fragments of *Pp* ACE homologs obtained from conceptual translation of *Pp* uniESTs were aligned to ACE sequences of insects (*Apis mellifera, Bombyx mori, Drosophila melanogaster, Tribolium castaneum*), ticks (*Ixodes scapularis* and *Rhipicephalus microplus*), the nematode *C. elegans*, and human, and also to sequences for insect juvenile hormone esterases and other arthropod esterases and carboxyesterases obtained from blast search hits found in the NCBI non-redundant protein sequence database (October, 2010). For phylogenetic analysis, highly divergent regions were eliminated from the alignment using low stringency options in the Gblocks server, version 0.91b (Castresana, 2000) prior to sequence similarity analyses performed with MEGA4 (Tamura et al., 2007).

Putative *Pp* neuropeptides and protein hormones were identified using stand-alone Blast and queries of arthropod neuropeptide sequences including the *Tribolium* neuropeptides (Li et al., 2008). Predicted signal peptides were identified from the putative transcripts using the SignalIP online tool (Emanuelsson et al., 2007).

Results and Discussion

EST analysis

A total of 31,673 *Pp* EST raw sequences (total length of 23,048,972 bp, average length per sequence read = 728 bp), representing both directional sequences of 15,960 cDNA clones, were cleaned by removing low quality and vector sequences and masking the repeats. A total 30,704 clean sequences with N50 = 824 were obtained after discarding sequence reads smaller than 100 bp in length. Assembly of these sequences produced 6,789 contigs containing 27,237 ESTs. Among the contigs, about 96% contained ≤ 10 EST individual sequences, and only three contigs contained >100 ESTs. Many contigs (54%) contained overlapping ends obtained from forward and reverse sequences of the same clone. A total of 3,467 singlet ESTs (100 -1,013 bp length) remained after the assembly, yielding a total of 10,256 uniESTs.

Gene ontology analysis

The automated Blast2Go-based pipeline analysis of 10,251 uniESTs containing 66% contigs and 34% singlet found 5560 uniESTs that showed significant similarity with pre-existing NCBI database entries at E-value < 1E-04 (NCBI BLASTX, nr database, April, 2010). The E-value distributions for these hits were: ~ 6% < 1E-100, 15.8% ≤9.57E-70, and less than 10% were > 1E-10 (Fig. 2.1). The average length of the uniESTs, with Blast2Go hit, was 1005 bp, ranging between 143 and 3,021 bp.

The taxa with the largest number of top hits generated from Pp uniEST queries belonged to the blacklegged tick, *Ixodes scapularis* (1,920 top hits, Fig. 2.2), due to both its phylogenetic proximity and the large amount of data coming from the *Ixodes* Genome Project (Hill and Wikel, 2005). However, the numerical frequencies of the top hits across various taxa were likely biased by the dramatic differences in the sizes of the data sets available for each species, in addition to the genetic relatedness of the taxa. For example, although the Western predatory mite, *G. occidentalis*, shares a closer evolutionary relatedness to *P. persimilis* within the superorder

Parasitiformes, *G. occidentalis* gene sequences yielded only 132 top hits, making it the number eight top-hit species after four insects, *Amphioxus*, and mammals (Fig. 2.2). Other top-ranked species to which *Pp* sequences were first matched in similarity were insects (Fig. 2.1) belonging to the orders Coleoptera (1 sp.), Hymenoptera (2 spp.), Diptera (5 spp.), Homoptera (1 sp.), and Ptthiraptera (1 sp.).

Figure 3 summarizes 42 Biological Process Gene Ontology (GO) terms extracted from a total of 2,157 *Pp* uniEST sequences. Significant numbers of uniESTs are predicted to be involved in reproduction, development, locomotion, and response to stress. The diversity and abundance of specific ESTs or derived gene products could potentially be used to identify molecular biomarkers for traits such as reproductive rate, foraging capacity and resistance to pesticides, and these could then be deployed in current efforts to target genetic enhancement of phytoseiid mites (Hoy, 1985; 1992; Simoni and Castagnoli, 2010). Gene/PCR based biomarkers have been explored to monitor reproductive quality of mites of the predatory mites in mass rearing facilities (Brian Ellis, unpublished data).

Genes associated with pesticide metabolism

I found 47 uniESTs relevant to pesticide toxicity and metabolism (Table 2.1), of which 79% belonged to the two main enzyme superfamilies involved in the metabolic degradation of pesticides: the cytochrome P450s (CYP), and glutathione S-transferases (GST). Twelve CYP homologs (represented in eight contigs and four singletons) and 20 fragments of GST sequences (12 contigs and eight singletons) were identified in the *Pp* EST data set (Table 1). Aside from their involvement in degradation of xenobiotics (Knowles, 1997), CYPs are also known to be involved in hormonal metabolic pathways among invertebrates and vertebrates, such as in the suppression of juvenile hormone synthesis in insects (Sutherland et al., 1998). The role of CYP in hormonal control of development, especially in molting and metamorphosis, has been suggested as a suitable target for pest control (Spindler-Barth, 1992). I found significant amino acid sequence differences among *Pp* CYPs and *Pp* GSTs to enrich the EST database of *P. persimilis*. However, I did not attempt to assign biological roles to the CYP and GST homologs found in this EST dataset because both groups of enzymes are encoded by complex gene families

whose members carry out diverse metabolic functions, including insecticide detoxification. In addition to this gene family diversity, further complexity may arise from alternative splicing and gene expansion events (Syvanen et al.,1996; Ranson et al., 1998; Christmas et al., 2001). The *Pp* CYP and GST sequences revealed in this study can only be connected to discrete metabolic functions after detailed molecular and biochemical scrutiny.

An unusual set of nine uniESTs was found to share significant similarity to known acetylcholinesterases (ACE) (E-value 2.49E-87 to 6.44E-10) with best matches within the peptide databanks to ACEs from mites or tick species (Table 2.1). In general, however, the amino acid sequences predicted from these uniESTs have low identities to those of the other ACE sequences (Table 2.1), and, indeed, five of the uniESTs were identified as carboxylesterases, rather than ACEs,by the computational prediction analyses (PROSITE-Hidden markovian method). Because ACE is encoded by only a single gene in vertebrates (Massoulié et al., 1993), but by one or two *ace* genes in insects (Weill et at., 2002) and apparently up to three ACEs in the southern cattle tick, *R. microplus* (Temeyer et al., 2004), the multiple different transcripts encoding putative *Pp* ACEs were examined to estimate the number of different genes involved (Appendix A_a). I observed two alignment clusters of *Pp* ACEs within the global alignment of ACE sequences from insects, ticks, *C. elegans* and human. The pattern of similarities of each *Pp* fragment to the reference sequences resolved the matching patterns into two groups within the global alignment; group 1 uniESTs have higher similarity to the first half portion of the ACE reference sequences (PPE014_O05_f, Contig5707, PPE0138_P16_f, PPE013_C07_f, Appendix A_b), while group 2 is aligned to the C-terminal half and each amino acid sequence has relatively low sequence similarity to the reference ACEs (PPE013_C07_r, Contig4735, Contig3716, Contig6677, Contig5966, Appendix A_c). Based on the matching regions of the uniEST to the *ace* genes of other insects, we suggest that six to eight putative ACE transcripts were represented by the nine uniESTs found in this dataset.

Phylogenetic analysis of amino acid sequences revealed that the *Pp* ACEs appear to be related to arachnid esterase genes. These esterases are separated from typical *ace* genes of insect, mammal, and nematode, but also from other carboxylesterases or juvenile hormone esterases. While tick (*Ixodes scapularis*) esterases in this group have been annotated as *aces* in

the computational prediction, the phylogeny (Fig 4.A and B) reveals that this group of esterases is unlikely to belong to the typical *ace* genes. The phylogenetic trees built for each N-terminal and C-terminal matching region of the *ace* sequences revealed a general consensus with the results of a previous study (Yang et al., 2010), particularly for grouping of invertebrate ACE-1 and 2, and nematode ACEs. Addition of more arachnid taxa (*I. scapularis* and *R. microplus*) in this analysis has revealed that these arachnid esterases fall on a branch located in between carboxylesterases and ACE, and are distinct from either of those groups. Since I was unable to identify any other esterases closely related to this group of esterases in blast searches, I would suggest that this group represents an arachnid- specific esterase sub-family. The tree patterns of arachnid-specific esterases were best supported in the N-terminal region of the sequence alignment, while the phylogenetic support for C-terminally aligned sequences was low because of their low levels of sequence conservation. Nevertheless, this relatedness pattern was readily observed in the multiple sequence alignments (Appendix A_b and A_c). The biological function of these arachnid-specific esterases clearly deserves further investigation.

Other potential insecticidal targets are neurotransmitter receptors. Among the *Pp* EST, our Blast2Go analysis detected one putative transcript encoding a portion of a nicotinic acetylcholine receptor (nAChR). This identification was also supported by InterproScan (Hunter et al., 2008) search (SSF63712- Nicotinic receptor ligand binding domain-like), and four transmembrane regions are predicted (TMHMM-InterproScan) within the assembled fragment of 244 amino acids. The nAChRs mediate fast synaptic transmission in insects and are the target of neonicotinoid insecticides (Nauen et al. 2001). The neurotoxic effect of neonicotinoid insecticides is most effective against plant sucking insects because of the systemic action of these insecticides through the plant (Ishaaya et al., 2007). Only mild and short-term miticidal activity of nitenpyram, a common neonicotinoid used to suppress aphids, has been observed in *P. persimilis* after foliage applications (Akayama and Minamida, 1999).

Molecules involved in the chitin synthesis/degradation pathways have also been targeted for pest control (Kramer and Muthukrishnan, 1997). Chitin is an insoluble structural polysaccharide that serves as a key exoskeletal component and in the gut linings in arthropods, and therefore provides a promising target system for control of arthropod pest species by

disrupting their development and feeding behavior (Cohen, 1993; 2001). Chitinases are enzymes of special interest because of their importance in degradation of old cuticle structures during molting. I found uniESTs coding for two putative chitin deacetylase transcripts and three chitinases in the *Pp* EST dataset.

Neuropeptides and protein hormones

A total of 14 putative transcripts encoding putative neuropeptides and protein hormones (Table 2.2, 2.3) were identified using TBLATN search, based on insect and *Ixodes* neuropeptide sequences (Simo and Park, unpublished data). I followed Gäde and Goldsworthy (2003) and Gäde and Hoffmann (2005) in categorizing insect neuropeptides by their physiological effects (Table 2.2) such as energy metabolism, feeding behavior, growth and development, immune system, muscle activity, reproduction, and water and ion balance. Many of the neuropeptides found in the *Pp* ESTs have been annotated for functions related to feeding, water and ion balance, and development, which are the most relevant traits in relation to the foraging efficacy of the predatory mites used as biological control agents.

The number of putative mature peptide(s) was identified from the amino acid sequence (Table 2) using canonical signals for dibasic cleavage sites, while signal peptides and mature peptides were recognized based on similarity to representative precursor sequences, as presented in Table 3. With the exception of bursicon, the start codon (methionine) could be identified from the peptide sequences. Among the putative *Pp* neuropeptide uniESTs, I found two genes encoding putative crustacean cardioactive peptides (CCAP). CCAP is normally encoded by a single copy gene with an absolutely conserved mature peptide sequence (PFCNAFTGCamide) in insects. Two *ccap* transcripts found in the *Pp* uniESTs, named *ccap1* and *2*, encode 108 and 159 amino acid residue polypeptides (25 % similarity) with typical canonical signatures: N-terminal signal peptides and a mature neuropeptide sequence of 10 and 9 residues respectively, differing only by an additional alanine at the N- terminus (Pro-Phe- Cys-Asn-Ala-Phe-Thr-Gly-Cys-Ala-NH2, and Pro-Phe- Cys-Asn-Ala-Phe-Thr-Gly-Cys-NH2, Table 3). One case of gene duplication in CCAP-related genes (M-CCAP1 and M-CCAP2) has been reported in the pond snail, *Lymnaea stagnalis* (Vehovszky et al., 2005). However, that study also demonstrated functional

redundancy of the two *L. stagnalis* CCAPs in feeding modulation. Apparent duplication of a CCAP-encoding gene has been observed also in tick neuropeptide sequences, based on genomic sequence data (Simo and Park, unpublished data), but further functional examination of these peptides will be needed to unveil the exact role of duplicated *ccap* genes in Acari.

Rapid progress in the area of arthropod neuroscience is anticipated to enable novel, target-selective approaches to pest control (Hoffmann and Lorenz, 1998; Maule et al., 2002) (Fónagy, 2006). The present study contributes to our knowledge of *Pp* neuropeptides by providing an initial survey of the expression of neuropeptide-encoding genes in *P. persimilis*. This information, together with the corresponding cDNAs, represent an important platform for the study of physiology and behavior in the Mesostigmata mites, as well as in comparative studies of the evolution of neuropeptide-related physiology. In the future, these resources may also make it possible to genetically enhance predatory mites by increasing development and feeding rates, and thereby generating a more efficient pest control agent.

Conclusions

The analysis of 10,256 uniESTs from an economically important predatory mite, *P. persimilis*, identified large sets of genes involved in metabolite transport, reproduction, development, growth and response to stress. A set of 61 these uniESTs are putatively involved in functions relevant to toxicology and insecticide resistance, and to arthropod endocrinology (neuropeptides and neurohormones). In terms of the use of *P. persimilis* in biological control, knowledge of detoxification mechanisms within the Acari has been limited, and biased toward pest species such as ticks and herbivore mites. My research offers a start in redressing the imbalance in knowledge between pest species and their natural enemies. A better understanding of the physiological pathways underlying pesticide resistance may strengthen biological control strategies, and ultimately allow genetic enhancement of beneficial organisms used in such systems. Together, these approaches are expected to result in reduced pesticide usage and more environmentally sustainable crop production models (Hoy, 2009).

Acknowledgements

I gratefully acknowledge the technical assistance and expert advice of Dr. S. J. Brown, Dr. Dr. Caragea, Dr. Sanjay S. Chellapilla from the Bioinformatics Center at the Kansas State University, Doris Vong and Ian Xhi (UBC), and Dr. L. Simo (KSU). Financial support for this work was provided by the Kansas State University Ecological Genomics Institute and the Department of Entomology, Genome British Columbia, the Investment Agriculture Foundation of British Columbia, Koppert Biological Systems (Koppert BV); BioBest Canada Ltd; Bio-Bee Sde Eliyahu Ltd; Certis BP; SESIL Corp. Biological Systems; Syngenta Bioline; Applied Bionomics Ltd and the Michael Smith Laboratories-UBC.

References

Akayama, A. and I. Minamida. 1999. Discovery of a new systemic insecticide, nitenpyram and its insecticidal properties. Pp.127-148. In: Nicotinoid insecticides and the nicotinic acetylcholine receptor. I. Yamamoto and J.E. Casida (eds.). Srpinger-Verlag. Hong Kong.

Audsley, N; R.J. Weaver; and J.P. Edwards. 2001. In vivo effects of *Manduca sexta* allatostatin and allatotropin on larvae of the tomato moth, *Lacanobia oleracea*. Physiological Entomology 26: 181–188.

Broadie, K.S.; A.W. Sylwester; M. Bate; and N.J. Tublitz. 1990. Immunological, biochemical and physiological analyses of cardioacceleratory peptide 2 (CAP2) activity in the embryo of the tobacco hawkmoth *Manduca sexta*. Development 108: 59-71.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540-552.

Champagne, D. and J.M.C. Ribeiro. 1994. Sialokinin I and II: Vasodilatory tachykinins from the yellow fever mosquito *Aedes aegypti*. Proceedings National Academy Science 91: 138-142.

Christie, A.E.; C. S. Durkin; N. Hartline; P. Ohno; and P. H. Lenz. 2010. Bioinformatic analyses of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel neuropeptide-encoding precursor proteins, including ones from members of several little studied taxa. General Comparative Endocrinology 167: 164-178.

Christmas, P; J. P. Jones; C. J. Patten; D. A. Rock; Y. Zheng; S-M Cheng; B. M. Weber; N. Carlesso; D. T. Scadden; A. E. Rettie; and R. J. Soberman. 2001. Alternative Splicing Determines the Function of CYP4F3 by switching substrate specificity. Journal Biological Chemistry 276: 38166-38172.

Clark, J.M. and I. Yamaguchi. 2002. Scope and status of pesticide resistance. Pp: 1- 22. In: Agrochemical resistance: extent, mechanism, and detection. AGS Symposium series, vol. 808. J.M. Marshall and I. Yamagushi (eds). American chemical society, Massachusetts.

Cohen, E. 1993. Chitin synthesis and degradation as targets for pesticide action. Archives Insect Biochemistry Physiology 22: 245-261.

Cohen, E. 2001. Chitin synthesis and inhibition: a revisit. Pest Management Science 57: 946-950.

Conesa, A.; S. Götz; J. M. García-Gómez; J.a Terol ; M. Talón; and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.

Coast, G.M.; S.G. Webster; K.M. Schegg; S.S. Tobe; and D.A. Schooley. 2001. The *Drosophila melanogaster* homologue of an insect calcitonin-like diuretic peptide stimulates V-ATPase activity in fruit fly Malpighian tubules. Journal Experimental Biology 204: 1795-1804.

Coast, G.M.; I. Orchard; J.E. Phillips; and D.A. Schooley DA, 2002. Insect diuretic and antidiuretic hormones. Advances Insect Physiology 29: 279–409.

Croft, B.A. and H. E. Van De Baan. 1988. Ecological and genetic factors influencing evolution of pesticide resistance in tetranychid and phytoseiid mites. Experimental and applied acarology 4: 277-300.

Dewey, E.M.; S. L. McNabb; J. Ewer; et al. 2004. Identification of the gene encoding bursicon, an insect neuropeptide responsible for cuticle sclerotization and wing spreading. Current Biology 14: 208-213.

Duve, H; N. Audsley; R.J. Weaver; and A. Thorpe. 2000. Triple colocalisation of two types of allatostatin and an allatotropin in the frontal ganglion of the lepidopteran *Lacanobia oleracea* (Noctuidae): innervation and action on the foregut. Cell Tissue Research 300: 153–163.

Emanuelsson, O; S. Brunak; G. von Heijne; and H. Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols 2: 953-971.

Fónagy, A. 2006. Insect neuropeptides and their potential application in pest control. Acta Phytopathologica et Entomologica Hungarica 41: 137-152.

Fournier, D.; A. Cuany; M. Pralavorio; J. M. Bride; and J. B. Berge. 1987. Analysis of methidathion resistance mechanisms in *Phytoseiulus persimilis* A.H. Pesticide Biochemistry and Physiology 28: 271-278.

Fournier, D., M. Pralavorio, A. Cuany; and J.B. Berge. 1988. Genetic analysis of methidathion resistance in *Phytoseiulus persimilis* (Acari: Phytoseiidae). Journal Economic Entomology 81: 1008-1013.

Gäde, G; K.H. Hoffmann; and J.H. Spring. 1997. Hormonal regulation in insects: facts, gaps, and future directions. Physiology Review 77: 963-1032.

Gäde, G. and G.J. Goldsworthy. 2003. Insect peptide hormones: a selective review of their physiology and potential application for pest control. Pest Management Science 59: 1063-1075.

Gäde, G and K.H. Hoffmann. 2005. Neuropeptides regulating development and reproduction in insects. Physiological Entomology 30: 103-121.

Goldsworthy, G.J.; K. Opoku-Ware; and L. Mullen. 2002. Adipokinetic hormone enhances laminarin and bacterial lipopolysaccharide-induced activation of the prophenoloxidase cascade in the African migratory locust, *Locusta migratoria*. Journal Insect Physiology 48: 601-608.

Grauso, M.; E. Culetto; D. Combesa; Y. Fedona; J.-P. Toutant; and M. Arpagaus. 1998. Existence of four acetylcholinesterase genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Federation of European Biochemical Societies Letters 424: 279-284.

Hammock, B.D. and D.M. Soderlund. 1986. Chemical strategies for resistance management. Pp: 111-129. In: Pesticide resistance: Strategies and tactics for management. National Research Council (U.S.). Committee on Strategies for the Management of Pesticide Resistant Pest Populations. National Academy Press, Washington. DC.

Hemingway, J.; N. Hawkes; L.A. Prapanthadara; K.G.I. Jayawardenal; and H. Ranson. 1998. The role of gene splicing, gene amplification and regulation in mosquito insecticide Resistance. Philosophical transactions of the royal Society 353: 1695-1699.

Hill, C.A. and S.K. Wikel. 2005. The *Ixodes scapularis* Genome project: an opportunity for advancing tick research. TRENDS in Parasitology 21: 151-153.

Hoffmann, K.H and M.W. Lorenz. 1998. Recent advances in hormones in insect pest control. Phytoparasitica 26: 323-330.

Holstein, B and C. Cederberg. 1986. Effects of tachykinins on gastric acid and pepsin secretion and on gastric outflow in the Atlantic cod, *Gadus morhua*. American Journal Physiology 250: 309-315.

Honegger, H.W.; D. Market; L.A. Pierce; et al. 2002. Cellular localization of bursicon using antisera against partial peptide sequences of this insect cuticle-slerotizing neurohormone. Journal of Comparative Neurology, 452, 163–167

Hoy, M.A. 1985, Recent advanc es in genetics and genetic improvement of the Phytoseiidae. Annual Review Entomology 30: 345-370.

Hoy, M. A. 1992. Biological control of arthropods: genetic engineering and environmental risks. Biological Control 2:166-170.

Hoy, M. A. 1995. Multitactic resistance management: an approach that is long overdue? Florida Entomologist 78: 443-450.

Hoy, M.A. 2009. The predatory mite Metaseiulus occidentalis: mitey small and mitey large genomes. BioEssays 31: 581–590.

Huang, X and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Research 9: 868-77.

Huffaker, C.B. and C.E. Kennett. 1953. Differential tolerance in parathion of two *Typhlodromus* predatory on cyclamen mite. Journal of Economic Entomology 46: 707-708.

Hunter, S.; R. Apweiler; T. K. Attwood; et al. 2008. InterPro: the integrative protein signature database. Nucleic Acids Research Advance Access online publication. http://nar.oxfordjournals.org/cgi/content/full/gkn785v1

Ishaaya, I.; A. Barazani; S. Kontsedalov; and A.R. Horowitz. 2007. Insecticides with novel modes of action: Mechanism, selectivity and cross-resistance. Entomological Research 37: 148–152.

Jensen, M.H. and A.J. Malter. 1995. Disease and insect control. Pp. 75-76. In: Protected Agriculture: A Global Review. World Bank Publications, Washington, DC.

Jeppson, L.R., H.H Keifer, E.W. Baker. 1975. Mites injurious to economic plants. University of California Press. Berkeley, C.A.

Knowles, C.O. 1997. Mechanisms of resistance to acaricides. pp. 57-77. In: Molecular Mechanisms of Resistance to Agrochemicals V. Sjut (ed). Springer, Berlin.

Kramer, K.J. and S. Muthukrishnan. 1997. Insect chitinases: molecular biology and potential use as biopesticides. Insect Biochemistry Molecular Biology 27: 887-900.

Lange, A.B.; K. K. Chan; and B. Stay. 1993. Effect of allatostatin and proctolin on antennal pulsatile organ and hindgut muscle in the cockroach, *Diploptera punctata*. Archives of Insect Biochemistry and Physiology 24: 79-92.

Li, *B; R. Predel; S. Neupert; F. Hauser; Y. Tanaka; G. Cazzamali; M. Williamson; Arakane Y.; P. Verleyen; L. Schoofs; et al. 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle Tribolium castaneum. Genome Research* 18: 113-122.

Massoulié J; L. Pezzementi; S. Bon; E. Krejci; and F.M. Vallette. 1993. Molecular and cellular biology of cholinesterases. Progress Neuroscience 41: 31-91.

Maule, A.G.; A. Mousley; N.J. Marks; T.A. Day; D.P. Thompson; T.G. Geary; and D.W. Halton;. 2002. Neuropeptide signaling systems - potential drug targets for parasite and pest control. Current Topics Medical Chemistry 2: 733-58.

Mi, H.; B. Lazareva-Ulitsky; R. Loo; A. Kejariwal; J. Vandergriff; S. Rabkin; N. Guo; A. Muruganujan, O. Doremieux; M. J. Campbell; H. Kitano; and P. D. Thomas. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Research 33 (Database Issue): D284-D288.

Nauen, R.; U. Ebbinghaus-Kintscher; A. Elbert; P. Jeschke; and K. Tietjen. 2001. Acetylcholine receptors as sites for developing neonicotinoid insecticides. Pp. 77-105. In: Biochemical Sites Important in Insecticide Action and Resistance. I. Ishaaya (ed). Springer Verlag Berlin, Heidelberg.

Ranson, H; F.H. Collins; and J. Hemingway. 1998. The role of alternative mRNA splicing in generating heterogeneity within the Anopheles gambiae class I GST family. Proceedings National Academy Science 95: 14284-14289.

Richards, K.S. and E. Marder. 2000. The actions of crustacean cardioactive peptide on adult and developing stomatogastric ganglion motor patterns. Journal of Neurobiology 44: 31-44.

Roush, R. T. and F. W. Jr. Plapp. 1982. Biochemical genetics of resistance to aryl carbamate insecticides in the predaceous mite, *Metaseiulus occidentalis*. Journal Economic Entomology 75: 304-307.

Sakai, T.; H. Satake; H. Minakata; and M. Takeda. 2004. Characterization of crustacean cardioactive peptide as a novel insect midgut factor: Isolation, localization, and stimulation of α-amylase activity and gut contraction. Endocrinology 145: 5671-5678.

Simoni, S. and M. Castagnoli, 2010. IPM strategies through specialist and generalist phytoseiids (Acari, Mesostigmata). Pp. 311-325. In: Integrated Management of arthropod pests and insect borne diseases. A. Ciancio, K.G. Mukerji (eds.). Springer Science. NY.

Spindler-Barth, M. 1992. Endocrine strategies for the control of ectoparasites and insect pests. Parasitology Research 78: 89-95.

Stay, B; A. Stephen; S.Tbeb; and W.G. Bandena. 1995. Alatostatins: Identification, primary structures, functions and distribution. Advances in Insect Physiology 25: 267-337.

Stone J.V.; W. Mordue; K.E. Batley; and H.R. Morris. 1976. Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilisation during flight. Nature 263: 207-211.

Sutherland, T.D.; G. C. Unnithan; J. F. Andersen; P. H. Evans; M. B. Murataliev; L. Z. Szabo; E. A. Mash; W. S. Bowers; and R. Feyereisen. 1998. A cytochrome P450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. Proceedings National Academy Science: 95: 12884-12889.

Syvanen, M.; Z. Zhou; J. Wharton; C. Goldsbury; and A. Clark. 1996. Heterogeneity of the glutathione transferase genes encoding enzymes responsible for insecticide degradation in the housefly. Journal Molecular Evolution 43: 236-240.

Tamura, K; J.Dudley; M Nei; and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

Temeyer, K.B.; R.B. Davey; and A.C. Chen. 2004. Identification of a third *Boophilus microplus* (Acari: Ixodidae) cDNA presumptively encoding an acetylcholinesterase. Journal Medical Entomology 41: 259-68.

Van de Vrie, M., J.A. McMurtry; and C.B. Huffaker 1972. Biology, ecology, and pest status, and host plant relations of tetranychids. Hilgardia 41: 343-342.

Van Lenteren, J.C and J. Woets. 1988. Biological and integrated pest control in greenhouses. Annual Review Entomology 33: 239–269.

Vehovszky, A.; H.J. Agricola; C.J.H. Elliott; M. Ohtani; L. Kárpáti; and L. Hernádi. 2005. Crustacean cardioactive peptide (CCAP)-related molluscan peptides (M-CCAPs) are potential extrinsic modulators of the buccal feeding network in the pond snail Lymnaea stagnalis. Neuroscience Letters 373: 200-205.

Walzer, A.; P. Schausberger; and P. Schausberger. 1999. Predation preferences and discrimination between con- and heterospecific prey by the phytoseiid mites *Phytoseiulus persimilis* and *Neoseiulus californicus*. Biocontrol 43: 469-478.

Wei, Z; G.J. Baggerman; R. Nachman; G.J. Goldsworthy; P. Verhaert; A. De Loof; and L. Schoofs. 2000. Sulfakinins reduce food intake in the desert locust, Schistocerca gregaria. Journal Insect Physiology 46: 1259-1265.

Weill, M.; P. Fort; A. Berthomieu; M. P. Dubois; N. Pasteur; and M. Raymond. 2002. A novel acetylcholinesterase gene in mosquitoes codes for the insecticide target and is non-homologous to the ace gene in *Drosophila*. Proceedings Royal Society London 269: 2007-2016.

Whalon, M.E., D. Mota-Sanchez, R. M. Hollingworth, and L. Duynslager. Arthropod pesticide resistance database. 2004-2010. Michigan state University. http://www.pesticideresistance.org/search/1/

Woodhead, A.P.; B. Stay; S.L. Seidel; M.A. Khan; and S.S. Tobe. 1989. Primary structure of four allatostatins: neuropeptide inhibitors of Juvenile Hormone synthesis. Proceeding National Academy Science 86: 5997-6001.

Yang, Z.; J. Chen; Y. Chen; and S Jiang. 2010. Molecular cloning and characterization of an acetylcholinesterase cDNA in the brown planthopper, *Nilaparvata lugens*. Journal of Insect Science 10: 102

Zdobnov, E.M. and R. Apweiler. 2001. InterProScan - an integration platform for the signature recognition methods in InterPro. Bioinformatics. 2001 17: 847-848.

Zhulidov, P.A.; Bogdanova, E.A.; Shcheglov, A.S.; Vagner, L.L.; Khaspekov, G.L.; Kozhemayko, V.B.; Matz, M.V.; Maleshkevitch, E.; Moroz, L.L.; Lukyanov, S.A.;

Shagin, D.A. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Research. 32: e37.

Tables and Figures

Table 2.1. Putative acetylcholinesterase, cytochrome P450, glutathione S- transferase, nicotinic ACE receptor, chitin deacetylase, and chitinase transcripts found in the *Pp* EST data set (Blastx E-value ≤ E-10).

Annotation was confirmed by the top 5 hits with BLASTX in NCBI. Annotation of the first hit in GenBank is provided. Gene descriptor and e-value are taken from the top hit in the NCBInr protein database. Predicted amino acid sequences were used to screen for Protein Functional Analysis using InteProScan. HMM prediction and significant positive value for each amino acid sequence fragment are derived from search within PANTHER V.6. The range of amino acid positions that may contain a given feature signature motif, and the length of the given fragment underneath, are enclosed within brackets.

Sequence ID	Accession number	*Source organism*	E-value	Identities	HMM panther precdition	PANTHER score [a.a. range/total]
Acetylcholinesterase						
PPE0138_P16.f	ABX10450.1	*Tetranychus cinnabarinus*	2.49E-87	149/232 (64%)	Acetylcholinesterase	1.2E-98 [1-219/219 a.a.]
Contig5707	XP_002414015.1	*Ixodes scapularis*	1.44E-56	163/498 (32%)	Carboxylesterase	3.1E-104 [14-505/507 a.a]
Contig4735	XP_002414019.1	*Ixodes scapularis*	2.84E-45	74/323 (22.9%)	Carboxylesterase	2.7E-26 [1-310/313 a.a.]
PPE013_C07.f	AAP92139	*Rhipicephalus microplus*	5.35E-45	117/273 (42%)	Acetylcholinesterase	2.1E-44 [1-150/150 a.a.]
PPE013_C07.r	XP_002402742.1	*Ixodes scapularis*	1.00E-43	90/175 (51%)	Acetylcholinesterase	1.5E-32 [5-139/139 a.a.]
PPE014_O05.f	NP_571921.1*	*Danio rerio*	7.00E-36	97/281 (34%)	Carboxylesterase	9.4E-71 [6-254/257 a.a.]
Contig5966	XP_002414627.1	*Ixodes scapularis*	9.13E-25	94/336 (27%)	Carboxylesterase	1.4E-43 [4-331/335 a.a]
Contig3716	XP_002409707.1	*Ixodes scapularis*	5.59E-15	46/148 (31%)	Acetylcholinesterase	1.10E-14 [1-42 and 61-131/132 a.a]
Contig6677	AAU11285.1	*Rhopalosiphum padi*	6.44E-10	58/242 (23%)	Carboxylesterase	4.60E-16 [1-243/249 a.a.]
Chitin deacetylase						
PPE0139_I21.f	XP_002399997.1	*Ixodes scapularis*	7.00E-117	196/241 (81%)	Low-density lipoprotein receptor	0.000018 [91-127/239 a.a]
Contig2368	XP_002399997.1	*Ixodes scapularis*	9.00E-135	216/243 (88%)	NA	NA
Chitinase						
Contig1449	XP_002199280.1	*Taeniopygia guttata*	4.28E-93	186/371 (50%)	Chitinase	2.2E-102 [37-349/349 a.a]
Contig5189	XP_001952718.1	*Acyrthosiphon pisum*	1.75E-57	124/342 (36%)	Brain chitinase and CHIA	3.6E-80 [14-344/362 a.a]
Contig1344	XP_001973906.1	*Drosophila erecta*	4.21E-12	45/137 (32%)	Dipteran chitinase-related	2.4E-26 [3-124/131 a.a.]
Cytochrome P450						
Contig3726	ABB89144.1	*Neoseiulus womersleyi*	8.00E-78	135/151 (89%)	Cytochrome P450	1.6E-73 [1-199/218 a.a]
Contig77	ABB89142.1	*Neoseiulus womersleyi*	6.00E-71	122/152 (80%)	Cytochrome P450, Subfamily 4C3 (CYPIVC3)	1.1E-71 [12-172/204 a.a]
Contig4501	XP_974280.1	*Tribolium castaneum*	2.20E-70	149/446 (33%)	Cytochrome P450	4.4E-69 [1-432/434 a.a.]
Contig3895	XP_002400112.1	*Ixodes scapularis*	3.19E-66	132/401 (32%)	Cytochrome P450	1.1E-64 [1-341/344 a.a.]
PPE014_B15.f	XP_002435754.1	*Ixodes scapularis*	6.00E-45	97/259 (37%)	Cytochrome P450	1.3E-50 [1-228/229 a.a.]
Contig2745	XP_002403855.1	*Ixodes scapularis*	1.00E-42	88/178 (49%)	Cytochrome P450, SUBFAMILY 44	3.1E-62 [1-170/199 a.a]
ppl-02-H09_065.r	XP_002731620.1	*Saccoglossus kowalevskii*	1.00E-35	88/199 (44%)	Cytochrome P450	6.6E-57 [1-155/162 a.a.]
Contig6506	XP_001865027.1	*Culex quinquefasciatus*	1.13E-24	62/142 (43%)	Cytochrome P450	3.8E-42 [1-134/166 a.a]
PPE0128_F22.f	XP_002404169.1	*Ixodes scapularis*	6.46E-17	54/192 (28%)	Cytochrome P450	1.9E-43 [1-303/303 a.a.]
Contig2403	XP_002413007.1	*Ixodes scapularis*	1.04E-15	40/71 (56%)	Cytochrome P450	2.4E-21 [1-37/70 a.a.]
PPE0121_F05.r	XP_002413001.1	*Ixodes scapularis*	3.59E-14	55/114 (48%)	Cytochrome P450, Subfamily 4C3 (CYPIVC3)	1.4E-37 [1-77/106 a.a]
Contig170	XP_002413007.1	*Ixodes scapularis*	2.17E-11	35/53 (66%)	Cytochrome P450, Subfamily 4C3 (CYPIVC3)	2.5E-31 [2-64/64 a.a.]
Glutathion S-transferase						
Contig530	XP_002434206	*Ixodes scapularis*	7.35E-57	110/232 (47%)	Gluthation S-Transferase class Mu	4.9E-34 [1-137/138 a.a.]
Contig897	XP_002416387	*Ixodes scapularis*	4.00E-56	114/208 (54%)	Gluthation-S-Transderase theta	1.6E-51 [1-207/208 a.a]
Contig1771	ACL51929	*Bombus ignitus*	1.53E-44	96/211 (45%)	Gluthation-S-Transderase theta, GST	2.2E-69 [1-209/209 a.a.]
Contig4121	XP_002401423	*Ixodes scapularis*	1.17E-41	107/305 (35%)	NA	NA
Contig3955	AAK64362	*Galleria mellonella*	3.80E-41	100/215 (46%)	Gluthation-S-Transderase theta, GST	2.9E-71 [1-187/193 a.a]
PPE0124_A22.f	XP_002416387	*Ixodes scapularis*	2.26E-40	93/178 (52%	Gluthation-S-Transderase theta, GST	8.9E-35 [2-178/178 a.a.]
Contig3205	AAK64362	*Galleria mellonella*	7.63E-34	84/190 (44%)	Gluthation-S-Transderase theta, GST	9.1E-53 [1-118/120 a.a.]
Contig1741	XP_002401423	*Ixodes scapularis*	3.52E-32	109/223 (48%	No hit	No hit
Contig1287	AAQ74441	*Haemaphysalis longicornis*	1.26E-31	98/227 (43%)	Gluthation S-Transferase class Mu	2.6E-44 [1-45, 63-214/214 a.a]
Contig6358	XP_002406485	*Ixodes scapularis*	9.00E-30	48/115 (41%)	Glutathione-S-Transferase omega, domain containing	5.2E-24 [1-111/111 a.a.]
Contig34	XP_973541	*Tribolium castaneum*	1.31E-26	57/119 (47%)	Gluthation-S-Transderase theta	1.50E-13 [26-111/119 a.a.]
Contig4404	CAD89618	*Crassostrea gigas*	2.29E-20	106/236 (44%)	Glutathione-S-Transferase, domain containing	3.3E-70 [3-213/215 a.a]
Contig4265	NP_001165912	*Nasonia vitripennis*	2.33E-20	55/140 (39%)	Glutathione-S-Transferase omega, domain containing	3.6E-21 [1-133/135 a.a.]
PPE0127_N17.f	AAQ74441	*Haemaphysalis longicornis*	1.63E-16	45/84 (53%	No hit	No hit
PPE0132_G07.f	CAD89618	*Crassostrea gigas*	1.80E-15	51/150 (34%)	Glutathione-S-Transferase omega, domain containing	6.6E-24 [2-143/143 a.a.]
PPE0116_P12.f	CAD89618	*Crassostrea gigas*	5.37E-15	107/237 (45%)	Glutathione-S-Transferase omega, domain containing	1.4E-67 [3-213/215 a.a]
PPE0115_F09.f	XP_001865046	*Culex quinquefasciatus*	1.00E-14	40/87 (45%)	Gluthation-S-Transderase theta	1.90E-15 [1-86/86 a.a]
PPE016_F24.f	CAD89618	*Crassostrea gigas*	3.20E-12	79/210 (37%)	Glutathione-S-Transferase omega, domain containing	4.9E-55 [12-199/199 a.a.]
PPE0139_J22.f	CAD89618	*Crassostrea gigas*	3.61E-12	78/210 (37%)	Glutathione-S-Transferase omega, domain containing	1.5E-52 [12-198/198 a.a.]
PPE013_K20.f	CAD89618	*Crassostrea gigas*	4.25E-12	78/210 (37%)	Glutathione-S-Transferase omega, domain containing	1.2E-49 [12-164/171 a.a.]
Neuronal nicotinic acetylcholine receptor						
Contig5515	P18845.1*	*Carassius auratus*	3.00E-17	51/175 (29%)	Neurotransmitter gated ion chennel	1.3E-25 [1-99/244 a.a]

***** Exceptions for top hit usage in blastx search of the NCBI nr database. In these cases, we used the second or third hit from the list showing a consensus description or a designation of "unknown".

Table 2.2. Neuropeptide and protein hormone homologs from the Pp EST data set.

The number of mature peptides from each neuropeptide precursor sequence were predicted from amino acid sequences by identifying the peptide processing sites. Functions of the neuropeptides and peptide hormones are categorized by their role in physiology in the context of potential application in pest control measures, following Gäde and Goldsworthy, 2003 and Gäde and Hoffmann, 2005. A hyphen is used to indicate those sequences with no biological activity confirmation in arthropods.

Neuropeptide	# mature peptides	Functions*	Reference
Adipokinetic hormone	1	**Energy metabolism (**mobilize energy substrates into haemolymph) **Reproduction (**inhibition of vitellogenin production) **Immune system**	Stone et al. 1976 Gde et al. 1997 Goldsworthy et al., 2002
Allatostatin A	4	**Growth and development (**inhibition of vitellogenin and juvenile hormone biosynthesis in insects**)** **Feeding behavior** (inhibit gut motility)	Woodhead et al., 1989 Lange et al., 1993 Stay et al. 1995
Allatostatin C	1	**Growth and development (**inhibition of juvenile hormon biosynthesis in insects**)** **Muscle activity** (foregut contraction) **Feeding behavior** (feeding inhibition)	Stay et al. 1995 Duve et al., 2000 Audsley et al., 2001
Calcitonin (CT)-like diuretic hormone 2	1	**Water and ion balance** (diuresis, modulator of cardiac contractions)	Coast et al., 2001 Christie et al. 2010
CCHamide peptide	1	**Unknown in insects**	-
Corticotropin-releasing factor (CRF)-like insect diuretic hormone	1	**Water and ion balance** (diuresis) **Feeding behavior** *(*signals end of the meal - satiation)	Coast et al, 2002
Crustacean cardioactive peptide 1	1	**Feeding behavior** (increase in gut contraction) **Growth and development** (muscle contraction in ecdysis)	Broadie et al., 1990; Richards and Marder, 2000 Sakai et al., 2004
Crustacean cardioactive peptide 2	1	**Feeding behavior** (increase in gut contraction) **Growth and development** (muscle contraction in ecdysis)	Broadie et al., 1990; Richards and Marder, 2000 Sakai et al., 2004
Sulfakinin	2	**Feeding behavior** (antifeedant activity)	Wei et al., 2000 Maestro et al., 2001
Tachykinin	4	**Feeding behavior** (pepsin secretion) **Water and ion balance** (vasodilatory effect)	Holstein and Cederberg, 1986 Champagne and Ribeiro, 1994
Protein hormones			
Bursicon	1	**Growth and development** (cuticle sclerotization and wing spreading)	Honegger et al., 2002 Dewey et al., 2004
Glycoprotein hormone Beta-5	1	**Unknown in insects**	-
Insuline like peptide (1)	1	**Growth and development** (change in organ size)	Brogiolo et al., 2001
Insuline like peptide (2)	1	**Growth and development** (change in organ size)	Brogiolo et al., 2001

Table 2.3. *Phytoseiulus persimilis* neuropeptide and peptide hormone amino acid sequences translated using *Drosophila melanogaster* homologs as reference.

Symbols: Underlined - signaling peptide identified using SignalIP V.3 Bold and italic - putative processing sites. Bold, italic and double underlined - mature peptide. * - stop codon.

A- Neuropetides

1 - Adipokinetic hormone

MNARLLIVAAALVALVLLAQTMPAEA***QITFSKSWQAGKR***ALDDCAQRDLQAINHIKQLIVKEAMNLIQCRGEPLME*

2 - Allatostatin A

MQLRLNSLAAAVPLLSLLVLIAVMSPTVAQDDRAQELTPEEQLIYDVLV***KRPSGGHRYGFGLGKRSPEPMPAQPSFDEPTFSGYKGRQYNFGLGKR***PWPMNDFEEYRK***RKYNFGLGKR***SE*

3 - Allatostatin C

MGYNLKKSLITLFLITLILLINIVSSLAVNRETIKRNTDEQDDEWIAKQKPNVADLAFNSDAGDYLEKLATLFSPRYRSGRWSASSLNSVIPE***KRQIRYHQCYFNPISCFRR***LK*

4 - Calcitonin (CT)-like diuretic hormone 2

MMQTVALIFLIALAGTLASPAPRSDEALQYYYFMQHPPSMEYMLGD***KRSNGMIDFGLVRGMSGVDAAKARLGLKYANDPFGPGRR****

5 - CCHamide peptide

MKASLRLAMRSSVLLVLVAFIFVLSMAEQVSAADAASSFGNDVNRDFNTKRIIALL***RRSSRQPNALAGSCGLYGHSCLGGHGKR***SSPVTAEEVDDDGPMYRIDYDWLQSRT*

6 - Corticotropin-releasing factor (CRF)-like insect Diuretic hormone

MRQLFGCGSALLCLLLVLVAVVSVG***QAAYSQQGQNAYPYPKAFMLHRRAGGMPNLSVVGPLDVLRRKMMLDMMEQRMKSKINANNEFLSRLGKRADAIPAYSALQQVSVFKLPTKPNLHTTSRDKETTIPRHTNR****

7 - Crustacean cardioactive peptide 1

MSAMMMRLITITMIFILSLIELTLTANLVKRESGQQLINSFSGHLD***KRPFCNAFTGCGKK***RSLSIPNYPNLPSSLAMDDSSSASPPSSSSSSYPASSSLLSSSSLSIDEPLAINEWLNYLRMTQKLMEEARSWEILQSRINQFGSIEMGQRQKFIGLKR*

8 - Crustacean cardioactive peptide 2

MQISVSRGALVVFLILAISLVTVQAQPSPNRRVQLQELFGRPMVGGGLR***KRPFCNAFTGCAGKR***TPFALPLGDKQRLRFFRKFNSIDAPDSDEYVIDNANYERELLQE*

9 - Sulfakinin

MKLNLSFLVLTTIVVLLTFAASPSESANLVKRPRVDFKTWLKSIYPQVAEIEETEQNGQE***KRNRDEDYGHLRFGRAGLGDDYGHMRFGRRK****

10 - Tachykinin

MHCGSWCVIALVAAAGLLVLGSEAQYIEGPADDSLVDWRDVNALRDLDM***KRAFHAMRGKKASPFHAMRGKKLKG TNGDINTIIAELRRQIMAGKRGSFGMRGKR***FGPPADLEAEVAAPIIEDTR*

B- Protein hormones

1 - Bursicon (No START methionine identified)

VGRSVKPKHACKQGDLMGRLRPYGGFRPNSFPVAFG*LSLPIPHSVLRRVQCNLSARPEKRFLDATTLFVVSAALLLL RAFPIAAEENCQLKPVIHVIKEPGCQPKPVPSFACHGTCASYVQVSGSKYWQVERSCMCCQEVGEREATRRVYCPDQNPKY KKVITRAPVECMCRPCSTPNEDEIVAQELVAGLTVK*

2 - Glycoprotein hormone Beta-5

MASTSLRLFGFLSPPLWLAGFVLVAGQR***SPLVDAYQFGELATCKVSRYTYMPTRTDSQGRQCWDTLTIHACRGRCES LEFADWRFPYKRSVHSMCVHGARELVKTQLRFCDPDAEEELRDYEYYNALSCSCQICDSSQTSCEGF****

3 – Insulin-like peptide 1

MSGARIGWGLAPVLASVFTVYIGNS***AIVETEGFDVNALPSIRGADNQGMLLGPEALNEWREKFEARSQDDWGRLWN VEKHRRCYSDLQTHMEWFCRSGLSLAKRSGRNANTFGRTAGLHRLPTSLLYALRKVEHEIEDVDLRDDDIYEQNRLGPYGQ PAAQALQLLRKTSGGRRKRKELGIMDECCHQNDGCSWEEYAEYCDVGSRERRRLVPSATV****

4 – Insulin-like peptide 2

MTSTCNRWSVCLLLCVLQVLLP***ERSGASGSEGRRVRMCGKKLSDLTINLCSFVGGLNQRVQRRSERALGSFKRRLRP REVGIVDECCRNPCSLQQLLQYCARPTTKEEESKIMGRFLQVIRIPTKVVSVTRDASENEDAVEMEDEVNAPSQLNTNGAHS SDFIFA****

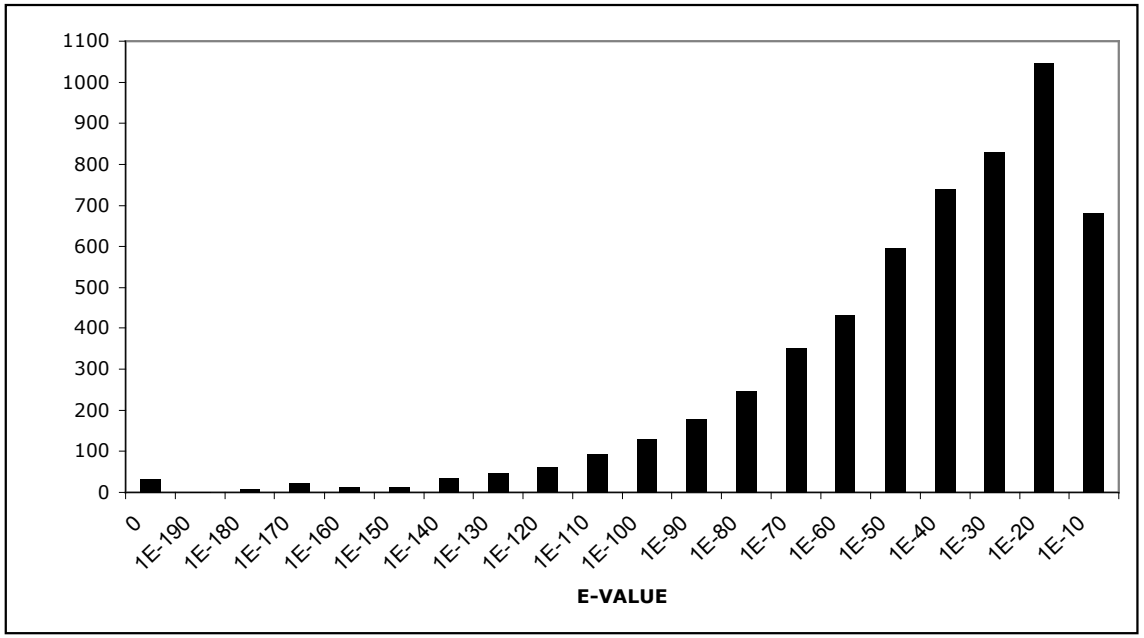


Figure 2.1. Frequency distribution for the top blastx hit E- value of 5560 uniESTs from *Phytoseiulus persimilis* combined data sets.

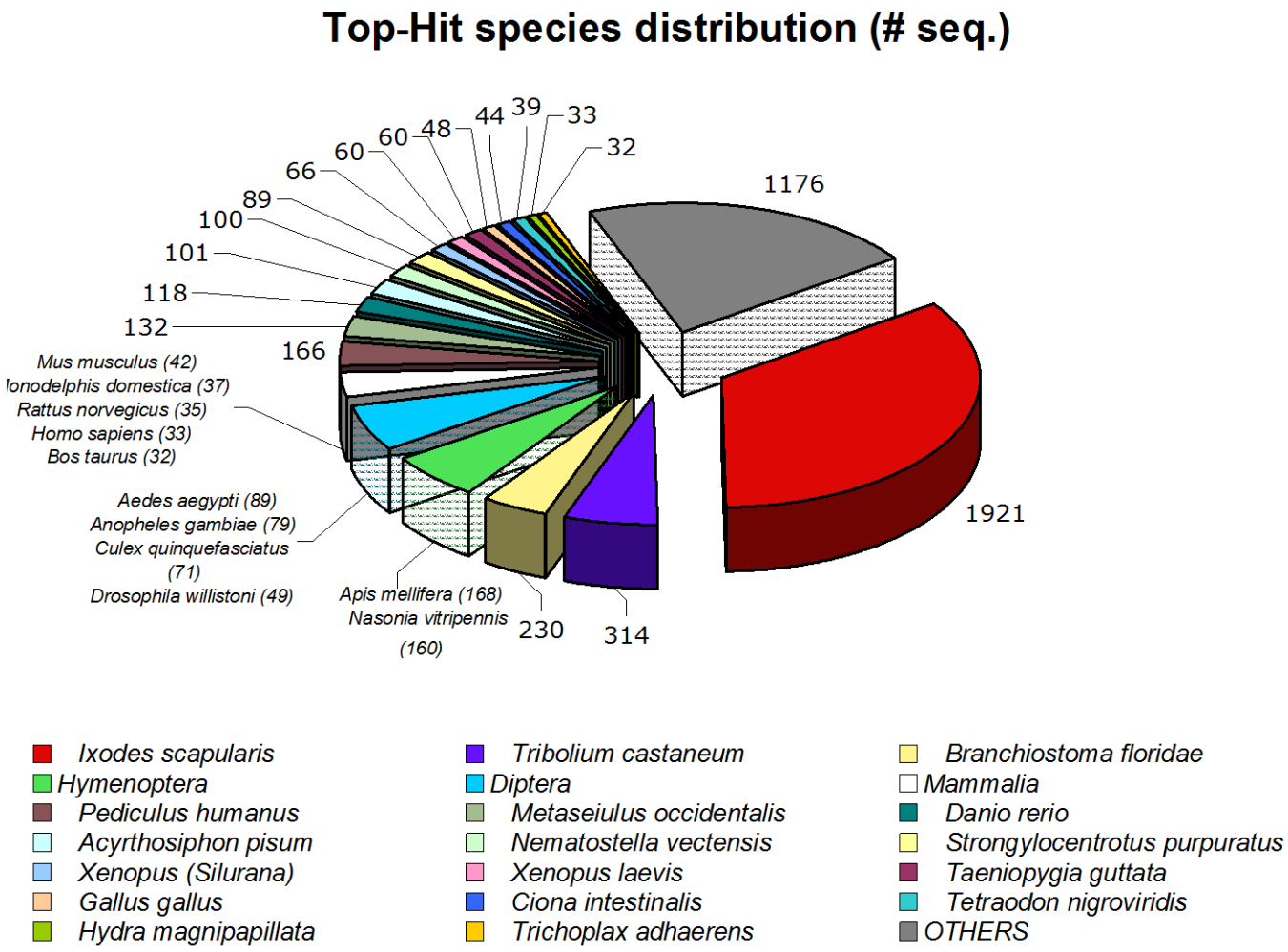


Figure 2.2. Distribution of blast hits to Acari and insect species amongst the top-hit species.

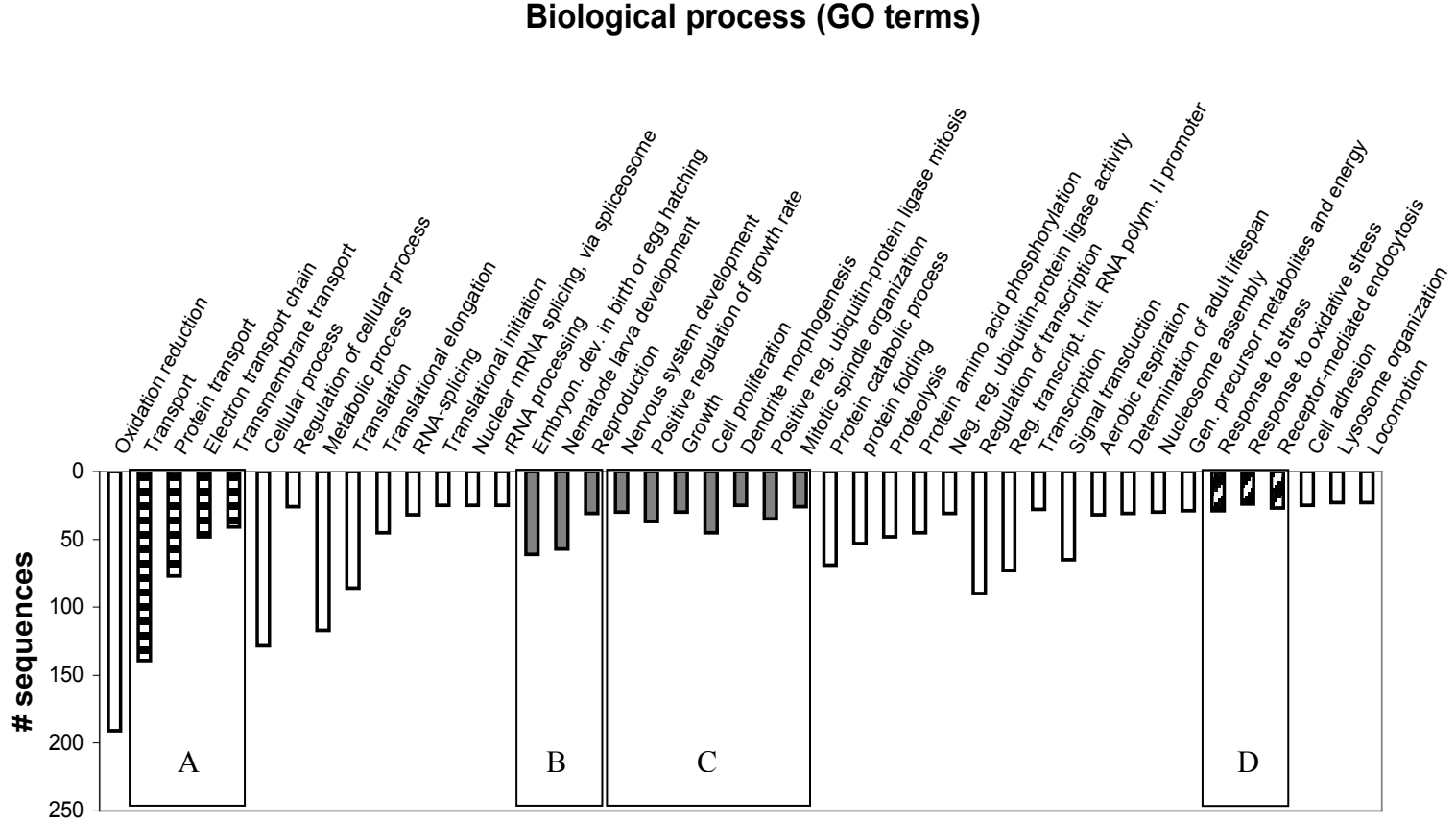
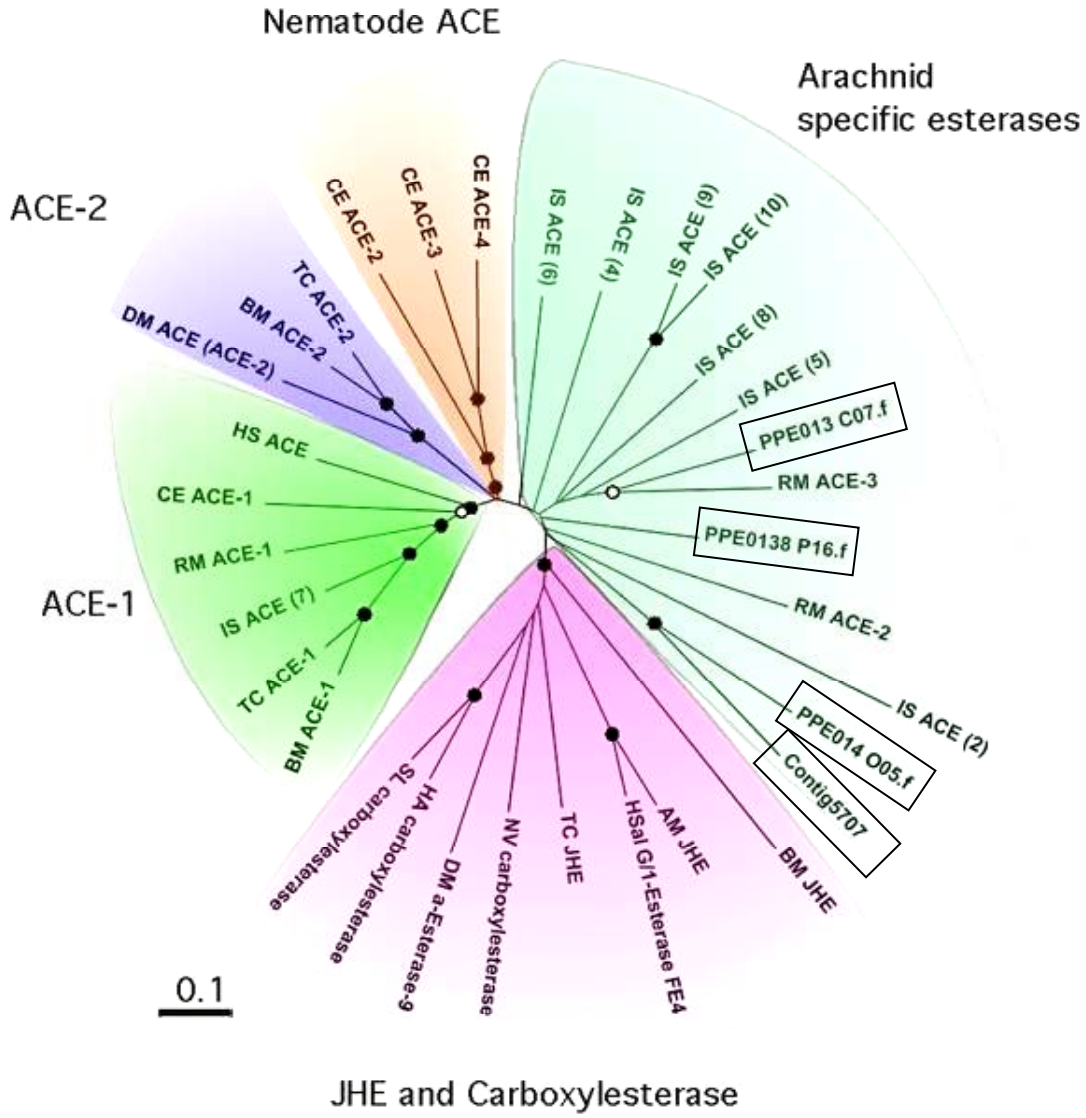


Figure 2.3. Distribution of 2,157 ESTs classified by Biological Process (Blast2GO). Genes involved in A) transport, B) reproduction, C) development and growth, and D) response to stress, indicated by solid bars.

4A


Nematode ACE
Arachnid
specific esterases
ACE-2
ACE-1
CE ACE-2
CE ACE-3
CE ACE-4
TC ACE-2
BM ACE-2
DM ACE (ACE-2)
HS ACE
CE ACE-1
RM ACE-1
IS ACE (7)
TC ACE-1
BM ACE-1
IS ACE (6)
IS ACE (4)
IS ACE (9)
IS ACE (10)
IS ACE (8)
IS ACE (5)
PPE013 C07.f
RM ACE-3
PPE0138 P16.f
RM ACE-2
IS ACE (2)
PPE014 O05.f
Contig5707
BM JHE
AM JHE
HSal G/1-Esterase FE4
TC JHE
NV carboxylesterase
DM a-Esterase-9
HA carboxylesterase
SL carboxylesterase
0.1
JHE and Carboxylesterase

4B

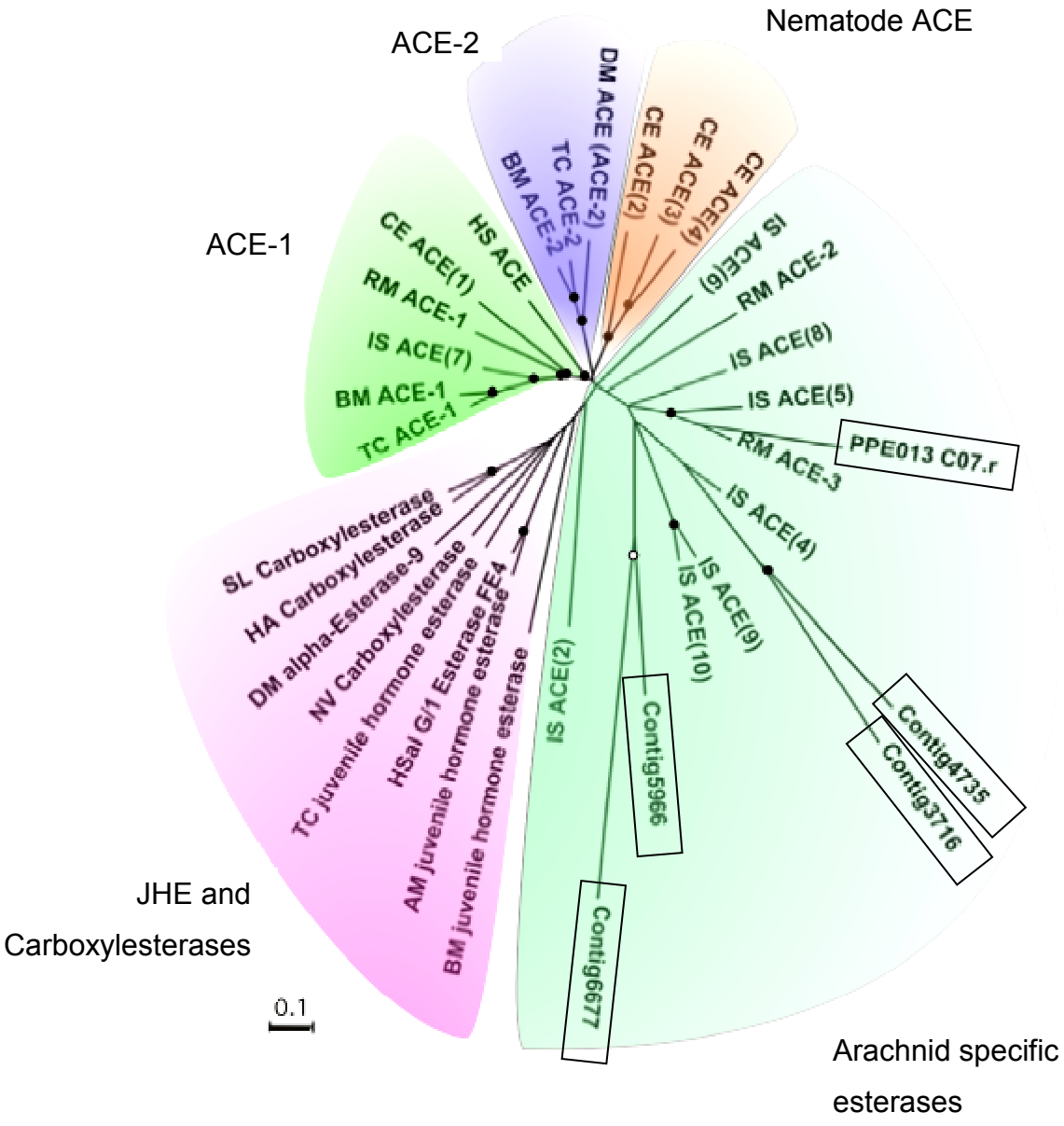

ACE-2
Nematode ACE
ACE-1
DM ACE (ACE-2)
TC ACE-2
BM ACE-2
CE ACE(2)
CE ACE(3)
CE ACE(4)
CE ACE(1)
HS ACE
RM ACE-1
IS ACE(7)
BM ACE-1
TC ACE-1
IS ACE(6)
RM ACE-2
IS ACE(8)
IS ACE(5)
PPE013 C07.r
RM ACE-3
IS ACE(4)
IS ACE(9)
IS ACE(10)
IS ACE(2)
Contig5966
Contig4735
Contig3716
Contig6677
SL Carboxylesterase
HA Carboxylesterase
DM alpha-Esterase-9
NV Carboxylesterase
TC juvenile hormone esterase
HSal G1 Esterase FE4
AM juvenile hormone esterase
BM juvenile hormone esterase
JHE and
Carboxylesterases
0.1
Arachnid specific
esterases

Figure 2.4. Unrooted distance Neighbor-Joining consensus tree showing the phylogenetic relationships of Pp ACEs group 1 (4.A., 229 a.a. positions alignment in Apendix A_d) and group 2 (4.B., 221 a.a. positions alignment in Appendix A-e) in respect to ACEs of insect, ticks, nematode, and human and other esterases and carboxyesterases.

The bootstrap value (1000 replicates) supporting each node is indicated as filled circle when ≥ 60%, open circle between 50%-59%, and none represents percentages below 49%. The evolutionary distances were computed using Poisson correction method and pair wise deletion option in MEGA4. The scale bar indicates a distance of 0.1 amino acid substitutions per position in the sequence. Name of sequences from *P. persimilis* are coding with a suffix Contig (clustered sequences) or PP (singleton) and are enclosed in a rectangle. The proteins or translated sequences correspond in alphabetic order to: AM = *Apis mellifera*, BM = *Bombyx mori*, DM = *Drosophila melanogaster*, CE = *Caenorhabditis elegans*, HSal= *Harpegnathos saltator*, HA = *Helicoverpa armigera*, HS = *Homo sapiens*, IS = *Ixodes scapularis*, NV= *Nasonia vitripennis*, TC = *Tribolium castaneum*, RM = *Rhipicephalus microplus*, SL = *Spodoptera litura*. See accession numbers in Appendix B.

Chapter 3 - A Devonian origin of parasitic mites estimated using expressed sequence tags of the predatory mite *Phytoseiulus persimilis* (Phytoseiidae, Mesostigmata, Acari)

Abstract

In assessing genealogical relationship between organisms, fossil records and biological sequences (nucleotide and amino acid sequences) have been commonly used to estimate ~~the~~ divergence time from a common ancestor. However, time estimation using stratification data frequently does not agree with calculations based on molecular data, which is usually attributed to gaps generated by incomplete fossil data. In the present study, we followed a phylogenomic approach to infer divergence time for Acari and the point of split of two lineages within anactinotrichid mites, *Ixodes scapularis and Phytoseiulus persimilis.* I used expresses sequence tags from the predatory mite *P. persimilis* to pull out 74 orthologous amino acid sequences of twelve invertebrates species: *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus*, *Drosophila melanogaster*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Nassonia vitripennis*, *Acyrthosiphon pisum, Daphnia pulex*, *Ixodes scapularis*, *and Caenorhabditis elegans*. I estimated a similar origin for Chelicerata (578.1 ± 38.2 - 482.2 ± 7.2 Mya) as in other recent studies. However, divergence dating using amino acid sequences suggested a Devonian origin of anactinotrichid mites (487.6 ± 32.2 - 410.1 ± 6.1 Mya) based on four reference dates (two fossil records and two molecular clocks) and four amino acid substitution methods; this estimate is much earlier that those in the current literature. This discrepancy of divergence times may be due to the use of a global clock. Thus, application of a local clock to the Acari branch, or smoothing the evolutionary rate in the phylogenetic analysis, would be recommended to avoid overestimation of the divergence time.

Introduction

Fossil records and biological sequences (nucleotide and amino acid sequences) have been frequently used to estimate the divergence time of related groups from a common ancestor in

determining genealogical relationship between organisms. Use of the biological sequences for dating evolutionary events has become increasingly prominent since it was first suggested by Zuckerkandl and Pauling (1962, 1965), and later refined by Sarich and Wilson (1969), regarding the concept of regularity in albumin change (Tobias, 1995). The increasing availability of molecular data improved the resolving power of inferences in phylogenetic studies by increasing the number of homologous characters to compare (Delsuk et al., 2005). Since the late 70's, numerous studies using DNA, RNA, and protein sequences have challenged the evolutionary theories and the timing of divergences among animal phyla (Wray et al., 1996; Ayala et al., 1998; Gaunt and Miles, 2002; Hedges et al. 2004). As result, there is an on-going debate regarding the deeper origins of animal phyla, ~800 to 1,200 million years ago (Mya) (Blair and Hedges, 2004), much earlier than what the fossil evidence indicated regarding the Cambrian explosion (~520 Mya). With little fossil evidence to confirm pre-Cambrian origins of animal taxa, this debate might remain unresolved but for a number of molecular technology (i.e. whole genome, expressed sequence tags) and paleontology tools (i.e. X-ray tomography, ion probe mass spectrometry) that are being developed to query the genomic and fossil records in ways that was not possible before.

The molecular clock hypothesis has become the basis for modern phylogenetic approaches. However, conflict between fossil evidence and molecular clock-based estimates has often troubled the matching of the evolutionary history and ecological roles. Divergence time estimated by stratification data frequently does not agree with molecular inference, which is usually attributed to gaps generated by missing fossil data (Morris, 1993) and the difficulty to assume a constant evolutionary rate between lineages (Britten, 1986; Hassani, 2006). As such, paleontological records are best used to set the minimum age constraint for dating (Marshall, 2008). A recent revision in paleontological records of chelicerate organisms illustrates the challenge to divergence dating using molecular data compared to the fossil records. Dunlop and Selden (2009) provided updated paleontological information that claimed an earlier time of divergence for some groups of chelicerates than that inferred from a molecular clock by Jeyaprakash and Hoy (2009). Within Acari, the oldest fossil record, belonging to Anactinotrichida (Parasitiformes *s. l.*) mite families, could only date the Ixodidae ticks back to the Cretaceous era (~100 Mya, Poinar and Brown, 2003; Poinar and Buckley, 2008). However,

using a molecular clock from mitochondrial DNA, the origin of this tick family is estimated to date from circa 241 Mya (Jeyaprakash and Hoy, 2009). It is interesting that fossil records evidence a much later appearance of parasitic mites compared to some acariform (= Actinotrichida) mite families from the Devonian (410 Mya; Hirst, 1923), whereas in Jeyaprakash and Hoy's molecular inference suggested origin of parasitic mites ~336 Mya, barely ~60 Mya diverged from the acariform mites. Morphological and molecular data suggest both Actinotrichida and Anactinotrichida are natural groups, but are not able to fully recover Acari as a monophyletic clade. Dulop and Alberti in the review about affinity between mites and ticks (2007) suggested that pronounced mite-specific characters and incompatible nomenclature have hindered the morphological comparison between mites and other arachnids; hence, not possible to resolve totally the phylogeny within Arachnida. The appealing effort by Jeyaprakash and Hoy (2009) to elucidate divergence times using a molecular clock within Acari potentially offers insight into their ecological relationships and evolution, especially in the search for the hosts of the ancient parasitic mite forms.

In the present study, I followed a molecular clock approach to infer divergence time of two lineages within anactinotrichid mites, order Ixodida and order Mesostigmata represented by *Ixodes scapularis and Phytoseiulus persimilis* respectively, using currently accepted insect clocks and fossil records. I used expresses sequence tags from the predatory mite to pull out 74 orthologous amino acid sequences using an orthologous database OrthoDB. The analysis yielded a dating of the parasitic life style over 100 Mya earlier than that reported by Jeyaprakash and Hoy (2009).

Material and Methods

Divergence times between two anactinotichid mites and that of Chelicerata (arachnids) - Mandibulata (crustacean and insects) were estimated using a molecular clock approach, based on the amino acid sequence alignment concatenated from 74 orthologous genes (Fig. 3.1). The phylogenetic relationships were determined using several insect species (*Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus*, *Drosophila melanogaster*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Nassonia vitripennis*, *Pediculus humanus*, *Acyrthosiphon pisum)*, a

crustacea (*Daphnia pulex*), a tick (*Ixodes scapularis*), and a nematode (*Caenorhabditis elegans*), for which annotated protein databases were available as a part of their genome sequence projects. I also used an expressed sequence tags (ESTs) dataset from the predatory mite *P. persimilis* (10,256 unique sequences). I initially considered all fourteen invertebrate species for ortholog search and alignment, but selected only thirteen species to generate the phylogenetic tree; the human louse was excluded when I generated the tree because, in preliminary observations using Neighbor - Joining method for phylogenetic inference, *P. humanus* was clustered next to the hymenoptera clade with low support (47% from 1000 bootstrap sampling), whereas all other nodes were highly supported (80-100% bootstrap value). The evolutionary relationships among insects remained the same after this exclusion.

Gene ontology analysis

The dataset of 10,256 uniESTs from *P. persimilis* (31,673 sequences in the raw data by sequencing both senses of the cDNA library) was yielded from ArthropodEST, an EST analysis pipeline (http://bioinformatics.ksu.edu/ArthropodEST/), by selecting default settings for sequence trimming, cleaning, vector and contaminant screening. The processes in the ArthropodEST pipeline includes repeat mask using arthropod clade as the reference, assembly by using CAP3 (Huang and Madan, 1999), and annotation by Blast2Go program (Conesa et al., 2005; GO database January 2010, pipeline version 2.3.5) with the optional parameter for blast search in NCBI BLASTX in nr database for <1E-04. For more details of cDNA library and assembly and processing of ESTs see Chapter 2 in this document.

Homologous groups and searching parameters

I screened an initial set of 325 *P. persimilis* uniESTs from sequences having high similarities with known genes in the NCBI protein database in the blastx search (E- value < 10E-100 top hit). I screened the UniESTs that are single copy and highly conserved among the aforementioned taxa to filter the set of data that represent one-to-one orthologies. OrthoDB online search options (Kriventseva et al., 2008; last version updated by 9/25/2009) was used to search for pre-determined orthologous groups within the database, by entering the query amino acid sequences of *P. persimilis* to pull out one-to-one orthologous protein groups. The parameter

chosen to search for single-copy ortholog used was "all except one" option (one species could have more than one copy). As *C. elegans* is not included in the OrthoDB database, its orthologs were pulled out from WormBase (April-May 2010), using blastp search with the mite sequence for single and top hit with e-value < E -60. From the 325 P. *persimilis* selected UniEST sequences, 216 UniEST were found not to have a one-to-one orthology group or had ambiguity in the orthology (i.e. more than one orthologous sequences at least in one species by OrthoDB classification) and were therefore excluded from the next step for orthology confirmation. In rare cases in which OrthoDB and Wormbase output lacked the orthology in one or more taxa, these were recovered by manual searches of the longest top-hit sequence in the GenBank. Appendix A lists the accession number for the amino acid sequence of eleven arthropod species retrieved from OrthoDB (database orthologous group reference number and sequence reference number), the accession number for *C. elegans* from WormBase, and that from fifteen sequences retrieved from GenBank.

Sequence alignment and concatenation

Screening the one-to-one orthologs were extended to the test of individual trees for all 14 species using MEGA4 (Tamura et al., 2007) and applying Neighbor-Joining method for Poisson amino acid substitution. Aligned sequences were manually cleaned to remove the highly diverging region or large gaps. In order for a more stringent screening for true ortholog groups, I excluded the gene fragments from the final analysis when individual gene trees clustering which *P. persimilis* is grouped with *C. elegans*, but not to *I. scapularis*, with over 60% bootstrapping support. Seventy-four orthologous groups, out of 109 examined, were used for dating divergence of Acari. *P. humanus* was not included into the phylogeny used to infer divergence time. Appendix B lists the accession number for the amino acid sequence of eleven arthropod species retrieved from OrthoDB (database orthologous group reference number and sequence reference number), and the accession number for *C. elegans* from WormBase. Fifteen sequences were retrieved from GenBank as they were not found in these databases.

The concatenated amino acid alignments were cleaned using Gblocks (Castresana, 2000; sequence (Supplemental data 1), which removes noisy region of sequences for better

phylogenetic analysis with the option of allowing less strict flaking positions. The parameters used for the Gblock included in the alignment were: minimum number of sequences used for a conserved position = 8, minimum number of sequences for a flanking position = 8, minimum number of sequences for a number of contiguous nonconserved positions = 8, minimum length of a block = 10, not allow gap positions, and allow use of similarity matrices for less strict definition of conserved positions. GeneBank Accession number for the ortholog amino acid sequences are in Appendix B.

Phylogenetic analysis and estimating divergence times

The phylogenetic analysis was performed with 72 genes from thirteen taxa, with a total 18,416 amino acids positions that contained 10,106 variable sites (6,059 parsimony informative sites). The consensus tree topology inferred from the assumed global substitution rate supported the current phylogenetic relationship among the considered arthropod clades and the nematode.

Four different amino acid substitution methods (MEGA 4, see Tamura et al., 2007) were used in the phylogenetic analysis and the branch distances were calculated for substitution rate in the methods Poisson correction, Equal input, PAM, Jones-Taylor-Thornton (JTT). A gamma parameter of 0.3 was used to correct variation of substitution rate among sites, which gave the tree topology with the highest bootstrap value congruent to current phylogeny (Nei and Kumar, 2000). Branch lengths were used to calculate the divergence time between chelicerates (Arachinida) and mandibulates (Crustacea and Insecta) within the arthropod lineage, and that between Ixodid mites (ticks) and Mesostigmata mites (*I. scapularis* and *P. persimilis*). Branch lengths, based on amino acid substitution rates, were obtained from two distance methods (Neighbor-Joining and Minimum evolution) and a discrete method (Maximum parsimony). A global clock was assumed in our phylogenetic analysis running with linear regression model in MEGA4 (Li and Graur, 1991; Sanderson, 1997). Four reference calibration points (Table 3.1) were chosen from among peer-reviewed literature: two were molecular clocks inferred from cytochrome oxidase I (COX I) from a same reference (Gaunt and Miles, 2002) and two were paleontological inference from the oldest fossil records.

These calibration points were narrowed by the arthropod species with best genome data available used in our study. The reference times inferred from COX I (basal Brachycera-Nematocera and basal Lepidoptera- Diptera) were determined under the experimental setting in Gaunt and Miles (2002) with calibration reference set at the Blattaria (cockroaches) - Orthoptera (crickets and locusts) divergence. These two reference points refer to: the divergence between dipteran suborders Nematocera (long-horn flies represented by *Anopheles*, *Culex* and *Aedes* lineages) and Brachycera (represented by *Drosophila* lineage) and the basal node of Lepidoptera and Diptera. The third reference point was taken from the oldest known beetles from the early Permian (Artinskian) deposits on Obora (see citations in Ponomarenki, 2002). And the fourth calibration reference used here refers to the oldest Brachycera flies contemporary to Nematocera representatives in fossil records.

Results and Discussion

A Devonian split between the lineages of ixodid ticks and phytoseiid mites, superorder Anactinotrichida, was highly supported by a robust phylogenetic inference (>84% in 1000 bootstrap sampling among neighbor-joining, minimum evolution, and maximum parsimony; Fig. 3.2). The tree topology obtained using thirteen species is congruent with currently recognized arthropod molecular phylogenies (Giribet et al., 2001, Meusemann et al., 2010), suggesting a monophyletic relationship of Insecta clade as a sister group to Crustacea, represented here by *D. pulex* (Fig. 3.2; see review in Telford et al., 2008), and the clustering of Insecta - Crustacea (Mandibulata) as the sister clade of Arachnida (Chelicerata). The divergence time estimated between these two parasitic lineages (node marked by blue sphere in Fig. 3.3) by four reference times ranged from 487.6 ± 32.2 to 410.1 ± 6.1 Mya (Table 3.2), using four different models of amino acid substitution. All of these inferences pre-dated by over 100 Mya that recently estimated by a molecular dating study using mitochondrial genome data (336 ± 26 Mya; Jeyaprakash and Hoy, 2009). Although overestimation of divergence time using mitochondrial genes has been a general concern for studying deep phylogenetic relationship, Jeyaprakash and Hoy's analysis using mitochondrial genes underestimated divergence compared to the results from my analysis using 74 gene fragments. However, both molecular studies suggest a much earlier rise of the parasitic behavior among anactinotrichid mites, over 300 Mya, than might be

supported by any fossil record (Dunlop and Seldon, 2009). An earlier origin of ixodid lineage suggested by molecular data is intriguing, since the time frame given to the hypothesis of ancestral reptilian host habitat for the tick is decoupled. Due to the parasitic behavior of currently known ticks, reptiles were assumed to be the earliest host as lizards were populating the terrestrial environments by 276 ± 54 Mya (molecular estimations, Kumar and Hedges 1998; Hedges and Kumar 2003). However, the hypothetical coevolution of reptiles and parasitic mites was never challenged because the fossil record of Anactinotrichida is poor, both in abundance and diversity along the stratification data.

Only ten families within Anactinotrichida have been recorded in the literature (Dunlop and Seldom, 2009), and the oldest fossil record from parasitic mites only goes back to the late Cretaceous (145.5 ± 4 to 65.5 ± 0.3 Mya). Current paleontological records do not address a Devonian origin of anactinotrichid mites with a parasitic lifestyle; the oldest available record for the suborder is a hard tick fossil, *Cornupalpatum burmanicum* n.g, n.sp., which has been dated barely to 100 Mya (Poinar and Brown, 2003; Poniard and Buckle, 2008); this probably favored the understanding of ixodid tick associations with monotremes and marsupials (Filippova in Klompen et al., 1996 and literatures therein). A large gap of 300 million years remains between oldest tick and the oldest Acari fossil records identified from Devonian stratification samples (c. 410 Mya; Hirst (1923); Dubinin (1962)). It is more likely that the fossil record is missing rather than a recent split of the parasitic clade. In order to obtain support for the observed divergence time of parasitic mites, divergence of Chelicerata - Mandibulata was examined (node marked with red sphere in Fig. 3.3; Table 3.3); it was general congruency to currently accepted estimations, which supports the robustness of the present analysis.

Despite an early divergence time for parasitic mites suggested by the present methodology, it was slightly more conservative in the estimate of the origin of Chelicerata compared to currently available dating from molecular clocks. This estimation using 74 gene fragments yielded a range of 578.1 ± 38.2 to 482.2 ± 7.2 Mya (Table 3.3). The oldest time estimations of the split between chelicerata and mandibulata lineages were inferred from the molecular reference points of the higher dipterans. Estimates using this reference point were also the closest to previous molecular inferences; the last common ancestor of cheliceratan and

crustacean hemocyanins was close to 600 Mya (Burmester, 2001), and the divergence of insects from crustaceans was around 666 ± 58 Mya, inferred from concatenated sequence of nine nuclear and 15 mitochondrial genes (Pisani et al., 2004). In the later study, their molecular clock analysis was confirmed and supported by the fossil evidence in the split of millipedes and centipedes ~442 ± 50 Mya (review by Dunlop and Selden, 2009). Arthropods divergence may have served as a lower constraint in dating the origin of parasitic mites. The proximity of the upper bound (482.2 ± 7.2 Mya) of the cheliceratan-crustacean split to the estimated divergence of anactinotrichid may suggest an overestimation of the divergence of the later clade. This phenomenon is likely the product of a larger number of substitutions found among the parasitic mites, resulting in a long branch length.

The longer branch of the Acari lineages, especially that of *P. persimilis* (Fig. 3.2 and 3.3), suggested a higher evolutionary rate of the genes in the predatory mites, which may lead to overestimation of the divergence times using different clocks. Indeed, among Acari high nucleotide substitution rate and gene arrangement has been documented (mitochondrial DNA, Navajas and Fenton, 2000), and a significantly different evolutionary rates between lineages has been observed, i.e. higher rate in Mesostigmata than in other parasitiforms (Murrel et al., 2005; Klompen et al., 2007). These characteristics of molecular composition in Acari have usually caused problems in sequence alignment and conflictive phylogeny due to long-branch attraction effect (Murrel et al., 2005; Dabert et al., 2010). However, in this study, congruent phylogeny was observed using 74 gene fragments. Longer branch observed among the Acari taxa could not to be attributed to the sequence quality in the *P. persimilis* UniESTs; we exclude this possibility with our high stringency quality controls by manual examinations of the original chromatogram and removal of poorly aligned region by using Gblocks. Therefore, the most likely explanation for the higher substitution rate observed within acari may be associated with the rapid evolution in Mesostigmata and possibly in all anactinotrichid lineages. In order to overcome the possible problem in dating the branching point based on the branch length, a local clock (a specific evolutionary rate for each branch) or smoothing the evolutionary rate methods would need to be applied (Welch and Lindell, 2005; Yang and Rannala, 2006).

References

Ayala, J. A., A. Rzhetsky, and F. J. Ayala. 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proceedings National Academy Science. 95: 606–611.

Bernini, F.; G. Carnevale; G. Bagnoli; and S. Stouge. 2002. An early Ordovician oribatid mite (Acari: Oribatida) from the island of Öland, Sweden. pp 45-47. In: F. Bernini, R. Nannelli, G. Nuzzaci, E. de Lillo (eds) Acarid phylogeny and evolution. Adaptations in mites and ticks. Kluwer, Dordrecht.

Blagoderov, V.A., E.D. Lukashevich ED; and M.B. Mostovski. 2002. Order Diptera Linné, 1758. The True Flies. P.p. 227-241. In: History of Insects. A.P. Rasnitsyn, D.L.J. Quicke (eds). Kluwer Academic Publishers, Dordrecht.

Blair, J.E. and S. Blair Hedges. 2005. Molecular clocks do not support Cambrian explosion. Molecular Biology and Evolution 22: 387-390.

Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393-1398

Burmester, T. 2001. Molecular evolution of the arthropod hemocyanin superfamily. Molecular Biology Evolution 18: 184–195.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540-552.

Conesa, A.; S. Götz; J. M. García-Gómez; J.A. Terol ; and M. Talón; and M. Robles. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Dabert, M.; W. Witalinski; A. Kazmierski; Z. Olszanowski; and J. Dabert. 2010. Molecular phylogeny of acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and long-branch attraction artifacts. Molecular phylogeny and evolution 56: 222-241.

Delsuk, F; H. Brinkmann, and H. Pjilippe. 2005. Phylogenomics and the reconstruction of the tree of life. Nature Review. Genetics 6: 361-375.

Dunlop, J.A. and P.A. Selden. 2009. Calibrating the chelicerate clock: a paleontological reply to Jeyaprakash and Hoy. Experiment and Applied Acarology 48: 183-97.

Dubinin, V.B. 1962. Class Acaromorpha: mites or gnathosomic chelicerate arthropods. pp 447–473. In: B.B. Rodendorf (ed) Fundamentals of palaeontology. Academy of Sciences of the USSR, Moscow. (In Russian).

Gaunt, M.W. and M.A. Miles. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Molecular Biology and Evolution 19: 748-61.

Giribet, G.; G. D. Edgecombe and W. C. Wheeler. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413: 157-161.

Hassani, A., 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading events of multiple changes in pattern and rates of substitution. Molecular Phylogenetic Evolution 38: 100-116.

Hirst, S. 1923. On some arachnid remains from the Old Red Sandstone (Rhynie Chert Bed, Aberdeenshire). Annals and Magazine of Natural History 12: 455-474.

Hedges, S.B. and S. Kumar. 2003. Genomic clocks and evolutionary timescales. Trends Genetics. 19: 200-206.

Hedges, S. B., J. E. Blair, M. L. Venturi, and J. L. Shoe. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol. Biol. 4:2.

Huang, X. and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Research 9: 868-77.

Jeyaprakash, A. and M.A. Hoy. 2009. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. Experiment and Applied Acarology 47: 1-18.

Klompen, J. S. H.; W. C. Black IV; J. E. Keirans; and J. H. Oliver, Jr . 1996. Evolution of ticks. Annual Review Entomology 41: 141-161

Klompen, H.; M. Lekveishvili, and W.C. Black IV. 2007. Phylogeny of parasitiform mites (Acari) based on rRNA, Molecular Phylogenetics and Evolution 43: 936-951.

Kriventseva, E. V.; N. Rahman; O. Espinosa; and E.M. Zdobnov. 2008. *OrthoDB.The Hierarchical Catalog of Eukaryotic Orthologs* Nucleic Acids Research 36:D271-5.

Krivolutsky, A. and A.Y. Druk. 1986. Fossil oribatid mites. Annual Review of Entomology 31:533-545.

Kumar, S. and S.B. Hedges.1998. A molecular timescale for vertebrate evolution. Nature 392: 917-920.

Li, W.H. and D. Graur. 1991. Fundamentals of molecular evolution.1st edition. Sinauer Associates. Sunderland, Massachusetts.

Marshall, C.R. 2008. A simple method for bracketing absolute divergence times of molecular

phylogenies using multiple fossil calibration points. American Naturalist 171: 726-742.

Meusemann, K.; B. M. von Reumont; S. Simon; F. Roeding; S. Strauss; P. Kück; I. Ebersberger; M. Walzl, G. Pass; S. B.; V. Achter; A. von Haeseler; T. Burmester; H. Hadrys; J. W. Wägele; and B. Misof. A Phylogenomic Approach to Resolve the Arthropod Tree of Life. Molecular Biology and Evolution 27: 2451-2464.

Morris, S. C. 1993. The fossil record and the early evolution of the Metazoa. Nature 361: 219 - 225.

Murrell, A.; S.J. Dobson; D.E. Walter; N.J.H. Campbell; R. Shao; and S.C. Barker. 2005. Relationships among the three major lineages of the Acari (Arthropoda : Arachnida) inferred from small subunit rRNA: paraphyly of the parasitiformes with respect to the opilioacariformes and relative rates of nucleotide substitution. Invertebrate Systematics 19: 383-389.

Navajas, M and B. Fenton. 2000. The application of molecular markers in the study of diversity in acarology: a review. Experimental Applied Acarology 24: 751-774.

Nei, M. and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Pisani, D.; L.L. Poling; M. Lyons-Weiler; S.B. Hedges. 2004. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. BMC Biololy 2: 1–10.

Poniard, G.O.Jr and A.E. Brown. 2003. A new genus of hard ticks in Cretaceous Burmese amber (Acari: Iodide: Ixodidae). Systematics Parasitology 54:199-205.

Poniard, G.O.Jr and R. Buckley. 2008. *Compluriscutula vetulum* (Acari: Iodide: Ixodidae), a new genus and species of hard tick from Lower Cretaceous Burmese amber. Proceedings of Entomological Society 110: 445-450.

Ponomarenko, A.G. 2002. Superorder Scarabaeidea Laicharting, 1781. Order Coleoptera Linné, 1758. The Beetles. Pp. 1640-176. In: History of Insects. A.P. Rasnitsyn and D.L.J. Quicke (eds). Kluwer Academic Publishers, Dordrecht.

Sanderson, M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology Evolution 14: 1218-1231.

Sarich, V.M. and A. Wilson. 1969. A molecular time scale for human evolution. Proceedings of the National Academy of Sciences 58:1088-1093.

Tamura, K; J. Dudley; M. Nei; and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

Telford, M.J.; S. J. Bourlat; A. Economou; D. Papillon; and O. Rota-Stabelli. 2008. The evolution of the Ecdysozoa. Philosophical Transactions of the Royal Society B 363: 1529 -1537.

Tobias, P.V. 1995. The bearing of fossils and mitochondrial DNA on the evolution of modern human, with a critique of the "mitochondrial eve" hypothesis. South African Archaeological Bulletin 50: 155-167.

Welch, J.J and B. Lindell. 2005. Molecular dating when rates vary. Trends Ecology Evolution 20: 320-327.

Wray, G. A., J. S. Levinton, and L. H. Shapiro. 1996. Molecular evidence for deep Precambrian divergences among metazoan phyla. Science 274: 568–573.

Yang, Z. and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23: 212-226.

Zuckerkandl, E. and L. Pauling, L. 1965. Molecules as documents of evolutionary history. Journal of Theoretica Biology 8: 357-366.

Zuckerkandl, E. and L. Pauling. 1962. Molecular disease, evolution and genic heterogeneity. PP: 189-225. In: Horizons in biochemistry. M. Kasha and B. Pullman (eds). Academy Press, New York.

Table and Figures

Table 3.1. Reference calibration points for molecular clock inferences: two molecular clocks inferred from cytochrome oxidase I (COX I) and two paleontological records based on the oldest fossil known.

Reference calibration points	Time (Mya)	Evidence type	Reference
1) Basal Brachycera - Nematocera	**282.8 - 247.7**	COX I (a.a. clock)	Gaunt and Miles, 2002
2) Basal Lepidoptera - Diptera	**351.4 - 338**	COX I (a.a. clock)	Gaunt and Miles, 2002
3) Oldest Brachycera - Nematocera	**245 - 237**	Fossil	Blagoderov et al., 2002
4) Oldest known beetles	**284.4 - 275.6**	Fossil	Ponomarenko, 2002

Table 3.2. Divergence time inferences for split between P. persimilis and I. scapularis estimated using four calibration time points, and four amino acid substitution models.

Two molecular clock inferences from amino acid sequence COX I and two fossil dates were used (see Material and method). Estimated time in million years ago (Mya) ± Mya for top and bottom limits. Four amino acid substitution models were used for phylogenetic inference and lineage distance using concatenated sequences of 74 orthologous protein fragments. The most distant and closest estimated dates from each calibration time (column) are marked in bold italic and bold, respectively.

	Molecular calibration		***Fossil calibration***	
a.a. substitution models/reference	**Brachycera/ Nematocera**	**Basal Lepidoptera - Diptera**	**Brachycera/ Nematocera**	**Oldest beetles**
Poisson correction	478.1± 31.6	439.6 ± 8.5	434.4 ± 7.2	415.6 ± 6.2
Equal input	487.1 ± 32.2	443.4 ± 8.6	442.6 ± 7.3	***421.2 ± 6.3***
PAM (Dayhoff)	***487.6 ± 32.2***	***445.3 ± 8.6***	***443.0 ± 7.3***	420.2 ± 6.3
JTT	**477.7 ± 31.6**	**434.6 ± 8.4**	**434.1 ± 7.2**	**410.1 ± 6.1**

Table 3.3. Divergence time inferences for Chelicerata-Mandibulata using four calibration time points and four amino acid substitution models.

Two molecular clock inferences from amino acid sequence COX I and two fossil dates were used (see Material and method). Estimated time in million years ago (Mya) ± Mya for top and bottom limits. Four amino acid substitution models were used for phylogenetic inference and lineage distance using concatenated sequences of 74 orthologous protein fragments. The most distant and closest estimated dates from each calibration time (column) are marked in bold italic and bold, respectively.

	Molecular calibration		***Fossil calibration***	
a.a. substitution models/reference	**Brachycera/ Nematocera**	**Basal Lepidoptera - Diptera**	**Brachycera/ Nematocera**	**Oldest beetles**
Poisson correction	563.4 ± 37.3	518 ± 10	511.9 ± 8.5	489.6 ± 7.3
Equal input	***578.1 ± 38.2***	***526.1 ± 10.2***	***525.2 ± 8.7***	***499.8 ± 7.5***
PAM (Dayhoff)	567.2 ± 37.5	518 ± 10	515.4 ± 8.5	488.9 ± 7.3
JTT	**561.8 ± 37.2**	**511.1 ± 9.9**	**510.4 ± 8.5**	**482.2 ± 7.2**

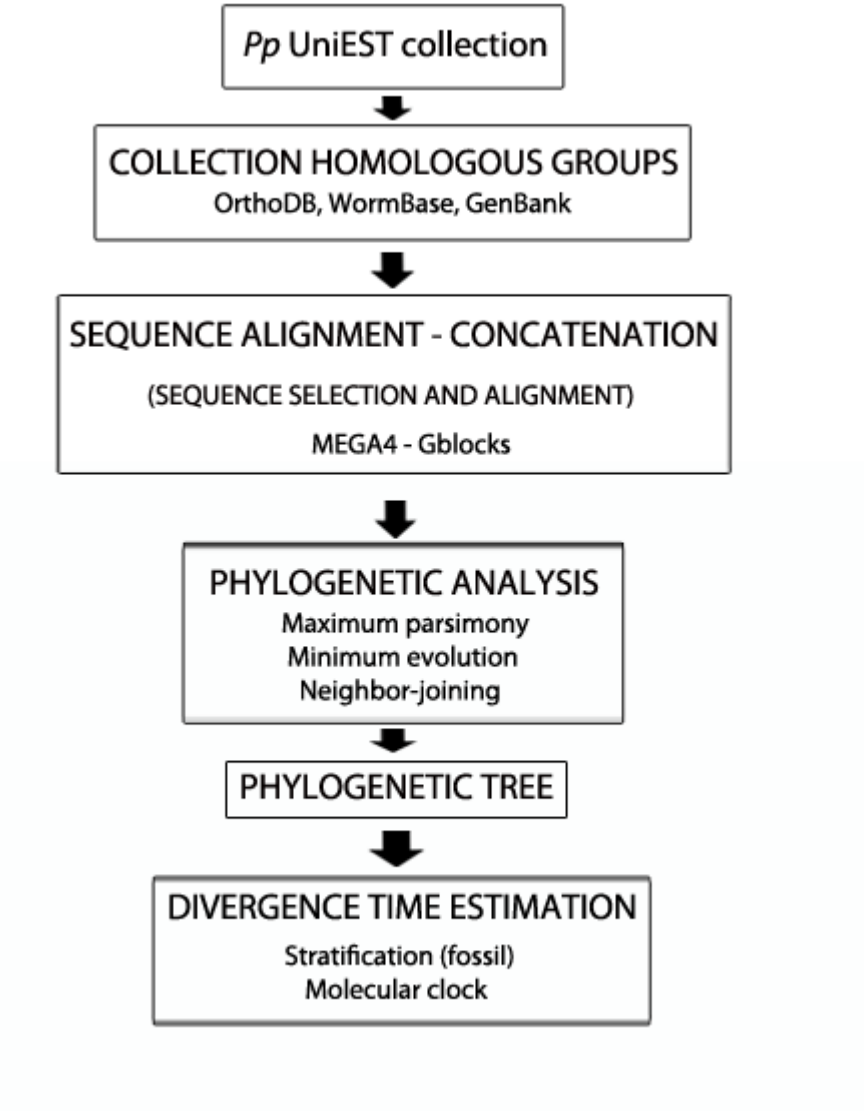
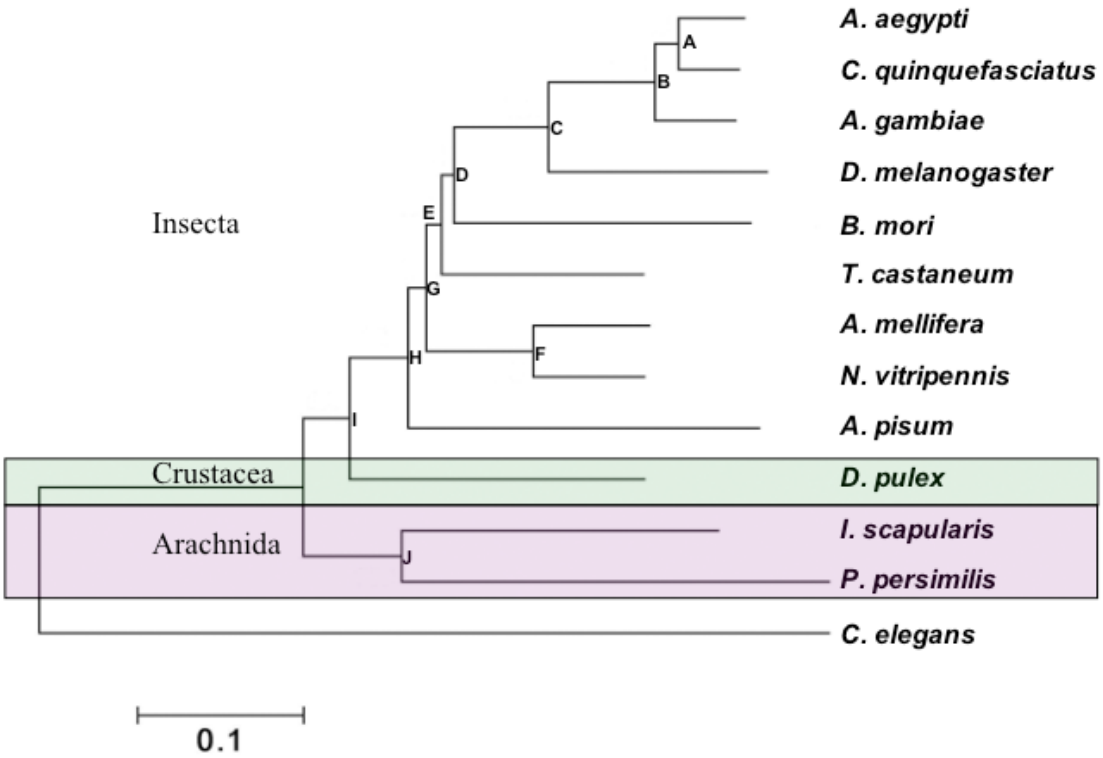


Figure 3.1. Work flow chart for Acari divergence time inference. Major steps in block capital letters, tools and methods in lower case.



METHOD/NODE	A	B	C	D	E	F	G	H	I	J
NEIGHBOR - JOINNING	100	100	100	100	99	100	99	100	100	100
MINIMUM EVOLUTION	100	100	100	93	95	100	84	100	100	100
MAXIMUM PARSIMONY	100	100	100	96	88	100	97	100	97	99

Figure 3.2. Unrooted phylogenetic tree representing the evolutionary history of 12 arthropod species (classes Insecta, Crustacea, Arachnida) and the nematode, *C. elegans*.

Tree inference using Neighbor-Joining method, JTT amino acid substitution model, gamma distribution (shape parameter = 0.3), and 1000 bootstrap repetition sampling. Support value of each node (A-J) from the Neighbor-joining, Minimum evolution, and Maximum parsimony inference methods are presented below. Bar= 0.1 a.a substitution/site/Ma.

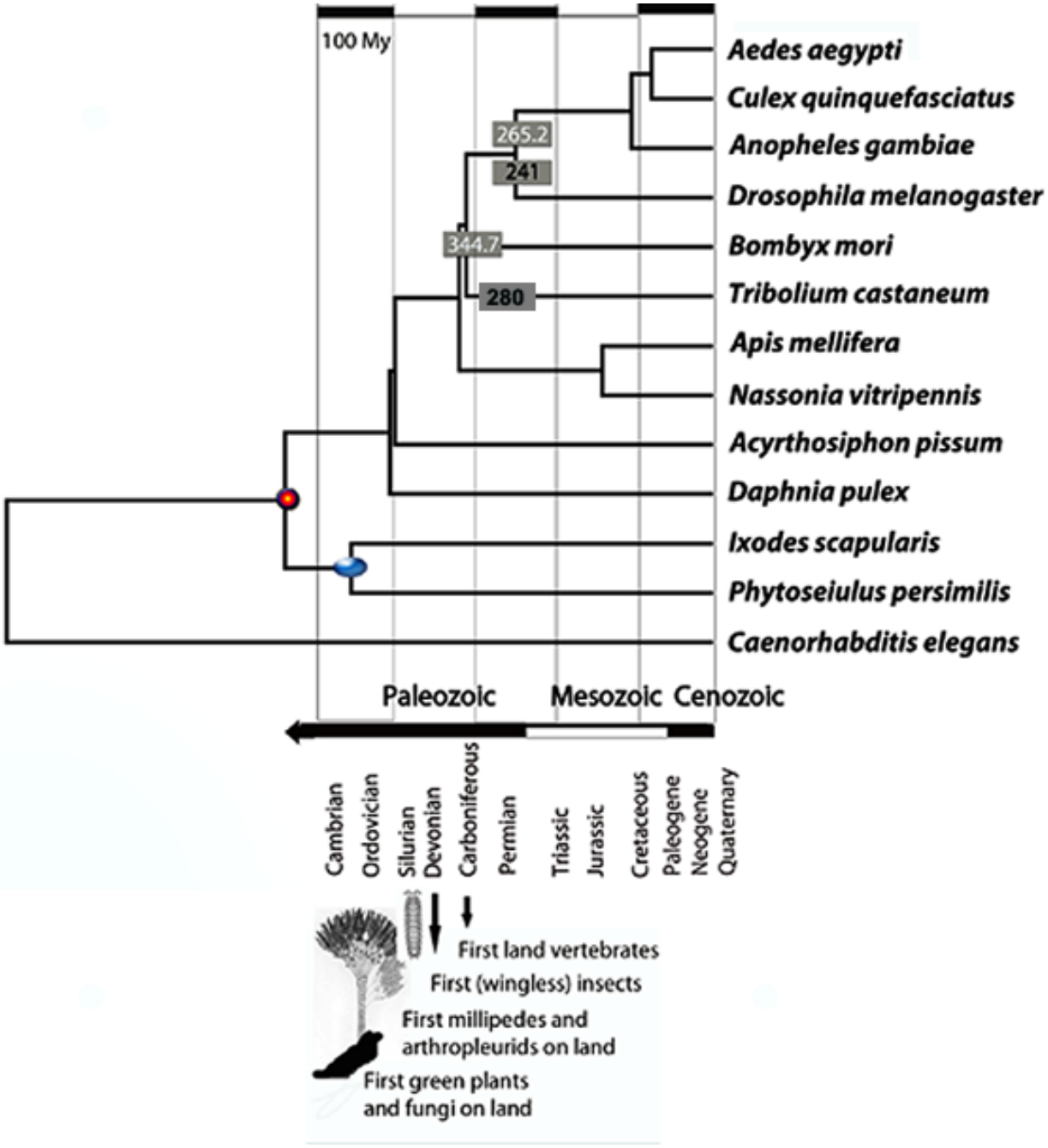

100 My
Aedes aegypti
Culex quinquefasciatus
Anopheles gambiae
Drosophila melanogaster
Bombyx mori
Tribolium castaneum
Apis mellifera
Nassonia vitripennis
Acyrthosiphon pissum
Daphnia pulex
Ixodes scapularis
Phytoseiulus persimilis
Caenorhabditis elegans
265.2
241
344.7
280
Paleozoic
Mesozoic
Cenozoic
Cambrian
Ordovician
Silurian
Devonian
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Paleogene
Neogene
Quaternary
First land vertebrates
First (wingless) insects
First millipedes and arthropleurids on land
First green plants and fungi on land

Figure 3.3. Evolutionary relationship of 13 taxa used to infer the divergence time of Anactinotrichida (blue oval in node, data in Table 3.2) and Chelicerata - Mandibulata (red sphere in node, dates in Table 3.3).

The four calibration clocks used (see Material and Methods) are in shaded boxes, in million years (Ma), with white numbers denoting the references from fossil records and in black number are those inferred from molecular data. Four major events in which first evidence of terrestrial representative of plant and fungi, arthropod, winged insect, and vertebrate were found, are marked in the geographic periods from Paleozoic era. Phylogenetic tree was built with Neighbor-Joining method, JTT a.a. substitution, 1000 bootstrap repetitions, and gamma distribution of 0.3.

Chapter 4 - Haplodiploidy investigated by PCR-based molecular markers in predatory mite *Phytoseiulus persimilis* Athias-Henriot (Acari: Mesostigmata)

Summary

Parahaploidy is a process, generalized among the phytoseiid mites, by which haplodiploidy occurs. In it, a biparental contribution of chromosomes (i.e., mating) is required but the paternal chromosome set is usually eliminated (PGL) from male offspring during early embryonic development, such that the paternal genome is not present in males. However, genetic studies in two phytoseiid species suggested diploid males with PGL only in the germ cells. Haploid males in the predatory mite *Phytoseiulus persimilis* has been supported through cytological and genetic evidences. In the present study, EST-derived makers were used to observe haplodiploid condition in *P. persimilis* and determine paternal genome inheritance to daughters. Unlike to previous studies, the haploid condition was observed, from two male mites, from the first filial generation within a single family examined. PGL is most likely during the early development stage in males of *P. persimilis*, therefore paternal alleles were not observed later in the adult stage. Haploid males carry only maternal genome and the diploid daughters inherited from both parents observed from twenty-three grandaughters. Independent assortment of all five markers is suggested by genotype of twenty-three F2 females from a backcrossed couple. However, larger sample size would be required for higher detection power.

Background

Genetic systems of sex determination in insects and mites often depart from the conventional diplodiploid system, with unusual mitotic events during embryonic development (White, 1973). This may include chromosome elimination or inactivation, usually during early embryonic development, and could be either in somatic cells or germ cells. Several families of homopteran insects (reviewed in Nur, 1980), such as the coffee berry borer *Hypothenemus hampei* (Scolytidae) (Brun et al., 1995; Borsa and Kjellberg, 1996), and in the predatory mites of

the family Phytoseiidae (Nelson-Rees et al., 1980; Schulten, 1985) are known to have sexual reproduction but the functional set of chromosome is of maternal origin. What is not known is whether paternal genome is inactivated or eliminated in the males.

In the predatory mite family Phytoseiidae (Order Acari or Acarina), heterochromatization has been reported and generally supported paternal genome elimination in males (Hansell et al., 1964; Hartl and Brown, 1970; Nelson-Rees et al., 1980; Schulten, 1985). The haploid condition of males has been generalized through karyotype observations of several species within the family (Sabelis, 1985; Norton et al., 1993). However genetic evidence does not corroborate uniparental contribution in males; biparental genetic contribution in phytoseiid males was first suggested through radiation studies (Helle et al., 1978; Hoy, 1979) and was supported by genetic studies in *Typhlodromus pyri* (Perrot-Minnot and Navajas; 1995) and in *Neoseiulus californicus* (Perrot-Minnot at al., 2000). The later study reported that the sperm produced by diploid males carried only the maternal genome, because in genetic cross experiments paternal genes inherited by the granddaughters were exclusively of the grandmother's origin. A lower amplification profile of the paternal loci observed in male offspring of *N. californicus* compare to the female offspring strongly suggested a selective elimination of paternal genome among male tissues. This feature may resemble the haploid/diploid mosaic male embryo observed in the mealybug, *Pseudococcus affinis* (Nur, 1990), in which paternal genome becomes activated (euchromatization) in poliploid cells, and activation is largely determined by the genome of maternal origin present in the cell.

Phytoseiulus persimilis is a Phytoseiid mite that is commercially reared and released for biological control of spider mites in field and greenhouse crops. Study of their genetic system has been intense and described possible parahaploidy. Mating is required in this species for egg production (Hoy, 1979; Toyoshima et al., 2000) and karyotyping by chromosome counting during early embryonic development suggested haploid males (n= 4) and diploid females (2n= 8) (Wysoki & Swirski, 1968; Olivier, 1977; Schulten et al., 1978). However, diploid cells have been observed, if only rarely, from the karyotyped nerve tissue of male mites, suggesting that heterochromatization is not homogeneous in all tissues of the embryo (Sabelis and Nagelkerke, 1993). Thus, discrepancies for haploid state of males in phytoseiid mites presents confusion. To

help resolve this question, we developed molecular markers by using the sequence variations found in Expressed Sequence Tag (EST), which served as markers for potential genetic variations observed in commercial populations of *P. persimilis*. The markers were used to examine the haploid condition of several *P. persimilis* males.

Material and Methods

Origin and maintenance of *P. persimilis*

Predatory mites from two commercial populations were used to study male genetic contribution in *P. persimilis* using EST-derived genetic markers. Difference in the allelic composition between populations was expected to enhance the probability for genetic marker discovery. In order to have genetically homogeneous populations for the crossing scheme described below, four isogenic predatory mite colonies were established from an in-house *P. persimilis* population (originally supplied by Koppert Biological Systems, Ann Arbor, Michigan, U.S.A.), reared in the laboratory for three years. Single-time-mated females were used to build up isogenic populations by allowing the offspring to interbreed for two months (at least eight generations). A second predatory mite population was purchased from Biotactics Inc. (California, U.S.A). All predatory mites were fed ad libitum with two spotted spider mites, *Tetranychus urticae*, reared on lima bean plant (*Phaseolus lunatus)*. Mite populations were maintained in the lab at room temperature (22-24 °C), 55-60 % RH and 16:8 L: D photoperiod.

Single mite DNA extraction and whole genome amplification (WGA)

The predatory mite *P. persimilis* is a small size organism, which length ~ 323-338 um (Denmark and Schicha, 1983) does provide small amount of genetic material for subsequent experimental assays. Therefore, I used the technique of whole genome amplification (WGA) for single mites developed by Konakandla et al. (2006) to obtain sufficient amount of DNA for individual genotyping. WGA procedure was followed immediately after DNA extraction and purification from single mites using E.Z.N.A.® Mag-Bind® Tissue DNA Kit from Omega Bio-Tec, with modification (Appendix C) of the original Mag-Bind Tissue DNA Protocol (2008).

The amount of reagent used at each step has been reduced to ten times less from the original protocol in order to optimize the volume for the small amount of tissue from a single mite. A whole mite is submerged in 8 µl of lyses solution and ground with a heat pulled glass rod. A 10 µl of Lysis solution was used to rinse the rod and mixed with 0.05 mg of proteinase K for overnight incubation at 55 °C in a rotation chamber. The duration of incubation time and separation times (when samples are loaded on the magnetic devise) are doubled from the suggested by the manufacturer. Final 5 µl elution was subjected to the WGA.

WGA from single mite DNA was performed using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Additional amount (25 U) of phi-29 DNA polymerase (EPICENTRE® Biotechnologies Cat. No. PP040110, 0.1 µg/µl (100 U/µl)) was added to the WGA reaction, which was used to enhance amplification yield from a small amount DNA, for more detail information and reference in Konakandla et al., 2006. Incubation time for WGA is extended to four hours at 30 °C, and heat inactivation at 65 °C for 20 minutes to stop the reaction prior to downstream applications and storage. Around 15 µl of WGA product is obtained and 1 µl of this is used to run on 1% agarose gel to confirm the quantity and quality of the amplified DNA, compared to a standard.

PCR amplification, DNA purification and sequencing

Polymerase Chain Reaction (PCR) for DNA amplification was performed from both population genomic DNA and WGA DNA from single mites using 20 µl of total reaction: 2 µl of 10X Econotaq PCR buffer with MgCl, 0.2 M of each oligonucleotide primers (Table 4.1), 0.2 mM of dNTPs, and 0.5 Units of Taq polymerase. The PCR protocol for all candidate markers was as follow: denaturation DNA at 94 °C for 3 minutes, followed by a first set for 25 cycles of 94 °C for 30 s, 55 °C for 1 min, 72 °C for 30 s, followed by a second set of 12 cycles of 94 °C for 30 s, 55 °C for 1 min, 72 °C for step extensions of 15 s + 5 s for elongation. The size of PCR products were determined on a 0.8 % polyacrylamide gel and visualized using UV light after ethidium bromide staining.

PCR products were cleaned up using ethanol precipitation procedure. Resuspended DNA in water was quantified and used a s the template for sequencing in one of following three facilities: Sequencing and Genotyping facility at the University of California – Riverside (sequencing for G0 and F1 generations); Retrogen Inc. (California, U.S.A.) and the Sequencing Core facility at the Kansas State University.

Genetic crosses

Genetic crossese were performed in two consecutive mite generations to observe inheritance:

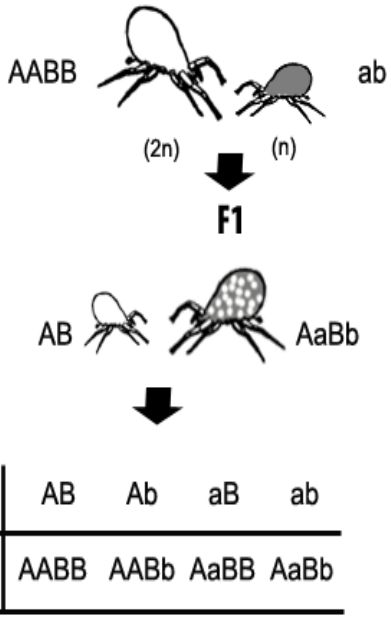
Parental pairing: Virgin mites of both sexes for the genetic crosses (see scheme below) were obtained by collecting eggs of similar age from isogenic colonies and single mites were reared in separate glass vials (2.3 x 9.5 cm). Two to four virgin females per each of the four isogenic colonies were paired with males from Biotactics Inc. population. Mating success and subsequent oviposition rate were determined for each couple. Only ten of these couples were isolated for breeding. Males were left with the female for 3-4 days after mating and then collected for extraction of DNA. Females were also collected after three days of oviposition. Eggs from the respective couples were collected and transferred into separate vials for rearing and sexing. Mites from this generation were considered as the first filial generation, or F1.

Backcross to maternal population: Four to six F1 mites per parental pair were backcrossed to the respective maternal isogenic population as following. Virgin hybrid males (F1M) and hybrid female (F1F) were paired to the corresponding virgin (five-day-old) mites form their maternal population. Males were allowed to stay with the paired female for up to eight days to ensure mating and sperm replenishment in order to achieve the maximum number of offspring for the second filial generation or F2 (70 eggs per pair as suggested by Denmark and Schicha, 1983). Despite of the capacity of *P. persimilis* females to produce all possible offspring with one single mating event (Rasmy and Hussein, 1996), re-mating seldom occurred with the same male from few hours to a few days after first mating. Females were transferred into a new glass vial for oviposition every other day, to prevent mating with their sons. A faster development was observed in males compared to females, turning into adult from eggs in three

days at room temperature. This step was crucial also to prevent sib-sib mating when kept in small clutches prior to sexing. These females were allowed to lay eggs until dead or stop oviposition for four consecutive days. Nymphs were transferred into individual vials and reared to adult for sexing purpose. All living offspring from F2 were collected and stored at -20 ^{o}C until DNA extraction for genotyping.

Selecting a family to follow up the F2: One great challenge in the experimental design for this experiment is the fast developmental time of the mites. During the experiment, *P. persimilis* females developed from eggs to adult from five to seven days, half time in males, at room temperature. In order to narrow the number of mite families being tested, genotyped the parental pairs first to select a single G0 family with the most polymorphic sites (Family 5). Then, two male F1 and one female F1 were also genotyped by sequencing to observe paternal inheritance. Offspring from the backcrossed F1 female were used to verify Mendelian segregation of each pair of genes. Chi-square test was performed to test the null hypothesis for equal chance of segregation of pairs of loci with twenty-three female mites of F2 generation. Under the haplodiploid genetic system, in which backcrossed female is heterozygote (AaBb) and male from maternal population is AB, the expected ratio for each genotype in assorted allele segregation will be ¼ AABB, ¼ AABb, ¼ AaBB, and ¼ AaBb, as indicated in the scheme below:

Genetic crosses:



1- Parental pairing

Koppert's female

mated to Biotactic's

2- F1 Daughter backcross

male from maternal population

3- F2 - generation

Ho: Pair of loci segregate Independently

¼ AABB, ¼ AABb

¼ AaBB, ¼ AaBb

Molecular marker development from EST data

A database of Expressed Sequence Tags (EST) derived from sequencing ~15,000 cDNA clones made from the predatory mite *P. persimilis* (pool of multiple stages and both sexes) was analyzed. This database was facilitated by Dr. Brian Ellis' *P. persimilis* genome project from the University of British Columbia, Canada. EST sequences were clustered by similarity into contigs using 80% minimum match and 20 bp of minimum overlap with the "dirty data" algorithm for contig assembly in the Sequencher V.4.6 (Gencode). In order to identify candidate molecular markers revealed in the EST, we ranked each contig for identifying the contigs carrying large numbers of clustered single nucleotide polymorphism (SNP) and insertion/deletions (INDEL) by examining raw sequence data to consider the sequence quality. In order to determine the marker candidacy, the polymerase chain reaction (PCR) approach was used determine the putative polymorphisms among experimental populations. Primer pairs for Polymerase Chain Reaction (PCR) for DNA amplification were designed in the conserved

regions flanking the clustered polymorphic regions, targeting for the lengths in the range of 100 bp to 300 bp based on the EST sequence data. Primer design was assisted by online softwares Primer3: WWW primer tool (Rozen and Skaletsky; 1998) and IDT OligoAnalyzer 3.1 (IDT Integrated DNA Technologies, Inc). Fifty sets of primer pairs (Appendix E) were designed and optimized for reaction. Initial optimization efforts used genomic DNA pooled from each of three *P. persimilis* population: two recently purchased from commercial insectaries Koppert Biological System (K population, Michigan, U.S.A.) and Biotactics Inc. (California, U.S.A.), and an in-house population, originally obtained from Koppert, that had been maintained in the lab for three years. Genomic DNA was extracted from a pool of 500 adult mites from each population using protocol for animal tissue of DNeasy Blood and Tissue Kit (Qiagen). Three µl of the PCR product was used to confirm the amplification on a 1% agarose gel electrophoresis. Subsequently, all primer sets were tested again for single mite DNA obtained after whole genome amplification.

A total of 434 contigs, ranging from 2 to 37 EST sequences assembled, was obtained from the database of 29,552 EST sequences from a cDNA library of the whole *P. persimilis* mites (Koppert). Gene ontology annotation was obtained from 193 contigs (average length 1,239 bp, ranging from 573 to 3,022 bp), in which 150 displayed high sequence similarities to the known sequences in the NCBI database (e-value ranging from 1E-20 to 1E-172). A group of 50 top polymorphic contigs with high blast score values were chosen to design oligonucleotide primers for PCR amplification. Primer sets were designed to amplify DNA sequences ranging from 147 to 318 nucleotides, and flanking the maximum number of polymorphisms (SNP, Indels, and small repeats) projected from the EST data. Genomic DNA of three populations was used to screen amplification rate among populations for marker candidacy. Four out of fifty primer sets did not yield DNA amplification within the range of PCR conditions tested. Ten out of fifty EST-derived candidates produced an amplicon much longer (> 400 bp) than the predicted from the EST sequence's length; this suggested that intronic region of the genome has been amplified using the flanking primer set. Single mite WGA-DNA samples from Koppert and Biotactics populations were used to screen the remaining candidate markers. We attempted a number of diagnostic genotyping assays: Restriction Fragment Length Polymorphism (RFLP), Degenerate Gradiente Gel Electrophoresis (DGGE.), and PCR amplifications of specific alleles

(PASA) techniques. None of these techniques were individually successful to identify polymorphisms in all candidate markers. But each of these techniques were used to screen all candidate markers. The migration patterns on agarose or polyacrylamide gels after electrophoresis were used to score polymorphisms and successfully narrow down from fifty to six candidate markers observed among the experimental populations. The chosen EST-derived markers were screened by sequencing the purified amplicons from each sampled individual of the genetic cros

Genotyping the sequence results

Sequencing results were analyzed using Sequencher v. 4.6 to genotype the mites in crosses. I examined manually the contigs and sequence chromatograms to avoid artifacts produced by erroneous events. Sequencing errors were confronted by comparing the results from both 5'and 3' senses. Contigs clustering multiple sequences of high allelic variation and multiple polymorphic sites were the most deterministic criteria used to choose reliable candidate markers. Manual examination of sequences and their chromatogram was necessary for zygocity analysis and for ambiguous base calling. Some markers contained small introns. To determine the intron/exon structure in the sequenced fragments, consensus EST and consensus sequenced marker fragments were assembled into contig with the Large Gap algorithm for 80% similarity and 20 bases of minimum overlap (Sequencher v.4.6.), and GT-AG canonical sequences for splicing donor and acceptor, respectively, were identified.

RESULTS

Whole genome amplification (WGA)

The success of WGA was largely related to the amount of initial DNA template form single mites like observed in Konakandla et al. (2006). Single mite DNA extraction was performed using magnetic beads (E.Z.N.A.® Mag-Bind ®Tissue DNA Kit) then exposed to a longer incubation time for tissue lyses and, extension in the standing time for washes during

DNA purification steps. The reaction volume for single mite WGA was reduced (half or fourth part) from the suggested in standard protocol and, additional amount of phi 29-polymerase was added to reach the efficiency of Genomi-phi WGA kit (see protocols in Appendix C). The incubation time for WGA was extended to 4 hours at 30 °C with a significant increase in yield compared to standard protocol from the kit.

One microliter of amplified product from each sample was used for 1% agarose gel electrophoresis. A smear was produced showing large sizes of amplification product after whole genome amplification. Most common sizes of the amplification product observed form the gel were over one thousand base pairs (Appendix E). Genomi-phi polymerase mix was essential for an efficient amplification but very low efficient amplification was also observed when solely phi-29 polymerase was used. It is possible that, other than just polymerase, essential components such as oligonucleotides have been included in the brand mix. Sample 3 in figure of Appendix 4.C represented the most common amplification product seen through the research.

Mite crossing and reproduction

Ten pairs of the mixed population of *P. persimilis* were chosen to produce offspring for backcross to their respective isogenic maternal populations. The backcross was performed using siblings of both sexes and observed for reproduction capacity, determined by eggs produced per days of oviposition. Egg production was quantified for seventeen F1 families until female terminated oviposition, either died or collected after oviposition ceased for more than three days. The longest oviposition period was 21 days (n=1) followed by 20 days (n=6) before being collected by lack of egg production. The reproductive capacity per mite ranged from 3.05 to 3.94 eggs per day similar to that of 3.7 eggs per day during 22 days reported by Takafuji and Chant (1976) at 25 °C. These data suggest that crossing *P. persimilis* from two populations did not affect their reproductive capacity. See Appendix F for data per family.

Candidate markers

Six EST-derived markers were selected among candidate markers after genotyping the eight parental pairs (G0). Oligonucleotide primer sets for the final candidate

markers are listed in Table 4.1. Figure 4.1 summarized the allele distribution and quantification among six candidate loci in each parental pair. Polymorphisms were unevenly distributed among the eight families. Two to four informative sites were found among the candidate loci and up to four allele types were observed within candidate locus among examined mites.

The family derived from the pair #5 (family 5) was chosen to observe the genetic system in the filial generations due to be the family with largest polymorphic sites detected in the EST-derived markers. Another molecular marker used in this study is *ppfor* (*Phytoseiulus persimilis* foraging gene), a partial sequence of a putative ortholog of the foraging gene (Appendix G). The primer set amplifies a DNA sequence of ~ 670 base pairs including ~240 base pairs of the intronic region (Konakandla et al., 2006) from which a high degree of polymorphisms was found within and among populations.

Table 4.2 summarizes the information for the molecular markers chosen. Blast searches against public gene sequence databases have found strong similarity of *P. persimilis* EST sequences to protein sequences of predicted function in insects and tick species. Four of the six EST-derived markers (ppM11, ppM21, ppM30, ppM45) yielded amplicons slightly longer (~100 bp) than the expected size deducted from EST data (Fig. 4.1). Presence of intronic region within the amplified sequence was confirmed by comparing the sequenced fragment to the EST consensus sequence (Appendix H). The intronic regions were determined by splicing donor and acceptor motifs GT and TA, respectively, in the amplicons from ppM11, ppM21, ppM30, and ppM45. All the polymorphisms observed in these candidate markers were in the putative introns.

Uniparental inheritance in males

Hemizygocity was observed in F1 males (n=2) carry only maternal alleles in the progeny (F1) of the Family 5, which suggest uniparental inheritance (Fig. 4.2). One daughter per family was examined to observe heterozygocity through biparental contribution. Haploid male is being suggested by absence of paternal alleles in the informative markers (ppM11, ppM30, ppM44, ppM45, and ppfor). Candidate markers ppM21 and ppM34 are heterozygotes in the mother and share an allele type with the father (Fig. 4.3). Therefore, inheritance from paternal contribution

would not be informative by sequence analysis with ppM21 and ppM34. Chromatograms were examined to determine single nucleotide polymorphism (SNP) in all but ppM40 marker, in which insertion/deletion was observed (Fig. 4.4)

No detectable recombination among the marker loci

In order to determine independent support of uniparental inheritance from each of the five loci examined, twenty- three F2 females were genotyped with five markers, and the frequency distribution of the genotype within each marker is listed in Table 4.3. Pairs of loci were tested for independent allele segregation through diploid phases, mother into daughters. Haploid male condition has been assumed in the homozygote genotype of the father, so an equal proportion for each genotype was expected. No significant difference was observed for independent assortment of each pair of loci at the upper limit of 95% of confidence in the Chi-square tests ($p>0.05$ observed in all, see Table 4.4). There is no evidence to reject the null hypothesis for an independent inheritance between loci, therefore suggests an unlike linkage existing among the loci examined in this study. A graph summarizing the statistical power for the degree of linkage at respective sample size is given in Figure 4.5. At the sample size of 23 female mites, which is the sample size in this study, it is only possible to detect 30% of the linkage, within 95% of confidence. It would require around a hundred mites examined by pair of loci to reach the probability of $p<0.05$ to observe a linkage with the recombination frequency of 40% - 50%. Thus, we conclude that the five loci we examined in this study are farther each other than 30% recombination units (Appendix I).

Discussion

Haploidy in male *P. persimilis* has been observed in two sons and the biparental inheritance in females support the parahaploid condition in phytoseiid mites generalized through karyotype examinations (Norton et al., 1993). Haplodiploid condition in phytoseiid mites has been first suggested by observations in inheritance of parathion resistance in *P. persimilis* (Helle

et al. 1978). Subsequently, cytological confirmations of heterochromatization of one set of chromosome strongly suggest the haploid condition in males (Nelson-Rees et al., 1980). However, the genetic study by Perrot-Minnot and collaborators (2000) in *Neoseiulus californicus* using DALP markers evidenced the diploid condition in males; in which paternal genome is retained in the somatic tissues but selectively eliminated in sperm production. Their result supported one possible model in PGL for parahaploid condition suggested in *Metaseiulus* (= *Typhlodromus* or *Galendromus*) *occidentalis* (Nelson-Rees et al., 1980) in which diploid condition in both sexes has been observed in early embryo and followed by chromosome elimination in 24-48 hour-old eggs. The result from the present study supports PGL in males of *P. persimilis* but no evidence of diploid condition in somatic cells was found. But, a general agreement on PGL in the germ line has meet from the genetic studies. And, the variable evidences using molecular tools from three species of phytoseiid mites may support the questioning on the stability of parahaploidy in mites (Norton et al., 1993).

At least four hypotheses regarding function of paternal chromosome have been exposed in parahaploid species of mites and coccids (summarized in Perrot-Minnot et al., 2000). Based upon all evidences from the phytoseiid mites, specially considering the PGL suggested in the present study, retention of paternal genome in males is likely to have a function during the embryonic development and possibly related to the integrity of first cell divisions during embryonic male development before its elimination as suggested by Johanowicz and Hoy (1998). *P. persimilis* differs from *N. californicus* in lack of retention of paternal genome in adult males and so discards the similarity of parahaploid system to coccids in which heterochromatized chromosomes could be reactivated to express the paternal genome (Nur, 1990). By the same term, it is unlikely that heterochromatized chromosomes would have a residual function for dosage compensation, "bulk" effect, as suggested in Nelson-Rees, 1962; Hoy 1979; Schulten, 1985. Masking of deleterious mutations and recombination repair have been proposed by Sabelis and Nagelkerke (1993) as possible advantages of paternal genome retention in early embryo in mites. Both ideas have not been demonstrated in the parahaploid mite species yet. A meiotic-like chromosome pairing was observed in *M. occidentalis* by Nelson-Reed et al. (1980) before reductional division and PGL in one day old embryo. However, no evidence so far exists to support recombination during the chromosome pairing. The maternal inheritance in

males and haplodiploid system evidenced from diverse studies in *P. persimilis* suggest the genetic system's resemblance to thelytoky. Sabelis and Nagelkerke (1988) suggest that parahaploid is an adaptative genetic system under certain conditions pertinent to phytoseiid mites, therefore reversion of parahaploidy to diploidy or evolution toward arrhenotoky or thelytoky is unlikely. Cytogenetic and molecular tools have not been able to determine more uniform features to describe parahaploid y in phytoseiid mites. It is possible that using a more systematic approach to unify methods could unveil the differences between species.

General considerations

Expressed sequence tags (ESTs) have been an efficient yet inexpensive approach to generate molecular markers for this study. Detecting single nucleotide polymorphisms and tandem repeats using ESTs could generate large sets of candidate markers and tandem repeats in the coding region of the genome. These markers are becoming more reliable than the usual polymorphism markers such as RAPID (low reproducibility) and RFLP (expensive and labor intensive). With the fast renewing DNA sequencing technologies, determine polymorphism has become an efficient and cost effective method for genetic studies in non-model species (Parkinson and Blaxter, 2009; Bai et a., 2010). This study has been pursued to understand the limitations in the experimental design and molecular marker development. Through this study, I have realized the need for balance between sample size and limitations in rearing and maintenance of *P. persimilis*. A different strategy should be approached to increase the significance of the evidence for PGL in *P. persimilis* males, possibly considering isofemale families and pooling sister groups mated to a single male to test for F2 individuals.

Rreferences

Altschul, S. F.; T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Avise, J.C. 1994. Molecular Markers, Natural History, and Evolution. Chapman & Hall. New York.

Bai, X; W. Zhang; L. Orantes; T-H. Jun; O. Mittapalli; M. A. R. Mian; and A. P. Michel1. 2010. Combining Next-Generation Sequencing Strategies for Rapid Molecular Resource Development from an Invasive Aphid Species, *Aphis glycines.* PLoS One 5(6): e11370.

Blommers-Schlösser, R. and Blommers L. 1974. Karyotypes of eight species of phytoseiid mites (Acarina:Mesostigmata). From Madagascar. Genetica 45: 145-138.

Borgia, G. 1980. Evolution of haplodiploidy: models for inbred and outbred systems. Theoretical Population Biology 17: 103-128.

Bottema, C. D. K. and S.S. Sommer. 1993. PCR amplification of specific alleles: Rapid detection of known mutations and polymorphisms. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 288: 93-102.

Bouck, A. and T. Vision. 2006. The molecular ecologist's guide to expressed sequence tags. Molecular ecology 16: 907–924.

Brown S.W. and U. Nur. 1964. Heterochromatic chromosomes in coccids. Science 145: 130-136.

Brown, S.W. and W.A. Nelson-Rees. 1961. Radiation analysis of a lecanoid genetic system. Genetics 46. 983-1007.

Brun, L. O.; J. Stuart; V. Gaudichon; K. Aronstein; and R.H. Ffrench-Constant. 1995. Functional haplodiploidy: a mechanism for the spread of insecticide resistance in an important international insect pest. Proceeding National Academy Science USA 92: 9861-9865.

Chant, D.A. 1961. An experiment in biological control of *Tetranychus telarius* (L.) (Acarina: Tetranycidae) in a greenhouse, using *Phytoseiulus persimilis* Athias-Henriot (Phytoseiidae). Canadian Entomology 93:437-443.

Denmark, H.A. and E. Schicha. 1983. Revision of the genus *Phytoseiulus* Evans (Acarina: Phytoseiidae). International Journal of Acarology 9: 27-35.

Hansell, R.J.G.; M. Mollison; and W.L. Putman. 1964. A cytological demostration of arrhenotoky in three mites of the family Phytoseiidae. Chromosoma 15:562 567.

Hartl, D.L. and S.W. Brown. 1970. The origin of male haploid genetic systems and their expected sex ratio. Theoretical Population Biology 1: 165-190.

Helle, W., H.R. Bolland; R. van Arendonk; R. de Boer; G.G.M. Schulten; and V.M Russell. 1978. Genetic evidence for biparental males in haplodiploid predator mites (Acarina: Phytoseiidae). Genetica 49: 165-171.

Hoy, M.A. 1979. Parahaploidy of the "arrhenotokous" predator *Metaseiulus occidentalis* (Acarina: Phytoseiidae) demonstrated by X-irradiation of males. Entomologia Explerimentalis Applicata 26: 97-104.

Johanowicz, D. L. and M.A. Hoy. 1998. Experimental induction and termination of non-reciprocal reproductive incompatibilities in a parahaploid mite. Entomologia Experimentalis Applicata 87: 51–58.

Konakandla, B; Y. Park; and D. Margolies. 2006. Whole genome amplifcation of Chelex-extracted DNA from a single mite: a method for studying genetics of the predatory mite Phytoseiulus persimilis. Experimental Applied Acarollogy 40: 241–247.

Nelson-Rees, W.; M. A. Hoy; R.T. Roush. 1980. Heterochromatization, chromatin elimination and haploidization in the parahaploid mite *Metaseiulus occidentalis* (Nesbitt) (Acarina: Phytoseiidae). Chromosoma 77: 263 276.

Norton, R. A.; J. B. Kethley; D. E. Johnston; and B. M. O'Connor. 1993. Phylogenetic perspectives on genetic systems and reproductive modes of mites. Pp. 8-99. In: Evolution and diversity of sex ratio in insects and mites. D.L. Wrensch and M.A Ebbert. (Eds.). Chapman & Hall. New York.

Nur, U. 1980. Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera). Pp. 97 177. In: R. L. Blackman, G. M. Hewitt and M. Ashburner (eds.), Insect cytogenetics. Blackwell Scientific Publications, Oxford, U.K.

Nur, U. 1990. Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera). Development: 29–34.

Oliver, J. H. 1977. Cytogenetics of mites and ticks. Annual Review Entomology 22: 407-429.

Parkinson, J. and M. Blaxter. 2009. Expressed sequence tags: an overview. Methods Molecular Biology 533: 1-12.

Perrot-Minnot, M.-J. and M. Navajas. 1995. Biparental inheritance of RAPD markers in males of the pseudoarrhenotokous mite *Typhlodromus pyri*. Genome 38: 838-844.

Perrot-Minnot, M.-J.; J. Lagnel; A. Migeon; and M. Navajas. 2000. Tracking paternal genes with DALP markers in a pseudoarrhenotokous reproductive system: biparental transmission

but haplodiploid-like inheritance in the mite *Neoseiulus californicus*. Heredity 84: 702-709.

Rasmy, A. H. and H. E. Hussein. 1996. Effect of mating on egg production in two species of predatory mites, *Agistemus exsertus* Gonzalez and *Phytoseiulus persimilis* Athias - Henriot. Anzeiger für Schädlingskunde 69: 88-89.

Rozen, S. and H. J. Skaletsky.1998. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Sabelis, M. W. and C.J. Nagelkerke. 1993. Sex allocation and pseudoarrhenotoky in phytoseiid mites. Pp. 512–541. In: Wrensch, D. L. and Ebbert, M. A. (eds) Evolution and Diversity of Sex Ratio. Chapman & Hall, New York.

Sabelis, M.W.,1985. Sex allocation. Pp. 83-94. In: W. Helle and M.W. Sabelis (eds) Spider mites, their biology, natural enemies and control. World Crop Pests Vol. 1B, Elsevier, Amsterdam.

Schulten, G. G. M., R. C. M. Van Arendonk, V. M. Russell, and F. A. Roorda. 1978. Copulation, egg production and se-ratio in Phytoseiulus persimilis and Ambliseius bibens (Acari:Phyroseiidae). Entomologia Experimentalis and applicata 24:145-153.

Takafuji, A. and D.A. Chant. 1976. Comparative studies of two species of predacious phytoseiid mites (Acarina: Phytoseiidae), with special reference to their responses to the density of their prey. Research population Ecology 17: 255-310.

Toyoshima, S.; M. Nakamura; Y. Nagaham; and A. Hiroshi. 2000. Process of egg formation in the female body cavity and fertilization in male eggs of *Phytoseiulus persimilis* (Acari: Phytoseiidae). Experimental and Applied Acarology 24: 441–451.

White, M. J. D. 1973. Animal Cytology and Evolution, 3rd ed. Cambridge University Press, Cambridge.

Wysoki, M. 1972. Further studies on karyotypes and sex determination of phytoseiid mites (Acarina:Mesostigmata). Genetica 44: 139-145.

Wysoki, M. and Wirski, E. 1968. Karyotypes and sex-determination of ten species of phytoseiid mites (Acarina: Mesostigmata) Genetica 39: 220-228.

Tables and figures

Table 4.1. Oligonucleotide primer sequences of candidate markers.

Marker	Forward oligo sequence	Reverse oligo sequence
ppM11	*5'-TATCACAGGAACGTCAACTTT-3'*	*5'-AGGCTCAAAGATGCTGAAT-3'*
ppM21	*5'-CCGACGAGGACAAATTGTT-3'*	*5'-CAAATTCCAGAGTTCAAGAACG-3'*
ppM30	*5'-ATCTGCTCGTGATGCATGTC-3'*	*5'-TCTGGCTTACAAGTTCTTCCAT-3'*
ppM34	5'- AGCTGGGAATGTATTCAGTC -3'	*5'-GATTACCTCTACGTGGGTTT-3'*
ppM44	*5'-CGCAACAGTCGTTTTTCTGG-3'*	*5'-TTGCGGTGGTGTAACTGAT-3'*
ppM45	*5'-AGCTCCATCGAATTTGAGAT-3'*	*5'-GCAGCGTTATGATTTTCACA-3'*
Ppfor	*5'-AGGGTCGTCGCCACCATCGG-3'*	*5'-GCGAAGCCGAAATCGACGAG-3'*

Table 4.2. Characteristics of candidate markers.

Contig sequence length and number of the sequences clustered into contig, sequence description from gene ontology analysis (GO) and its minimum E-value, the expected sequence length from contig prediction, and predicted amplicon composition (only exon or intron + exon) through electrophoresis results and sequence alignment compared to its uniEST. The markers ppM34 and ppM44 had no intronic region in the amplified fragments: ppM11- 1 SNP, ppM21 – 3 SNP, ppM30- 1 SNP, ppM44- 1 SNP and 1 INDEL, ppM45- 1 SNP. Three nucleotide differences, 1 SNP in exon and other 2 localized in the no coding region, were found in *Ppfor* marker between the chosen parental pair (Family 5).

Marker (locus)	Contig length (bp)/ # seq. clustered	GO - seq. description	Blastx e-value	Expected size (bp)	Amplicon Composition
ppM11	1205/4	Vacuolar ATP synthase subunit e	8e-52	220†	Intron + exon
ppM21	933/21	Von hippel-lindau binding protein 1	2e-44	226†	Intron + exon
ppM30	1442/4	Translocation protein 1	1e-59	176†	Intron + exon
ppM34	772/10	G protein gamma subunit	5e-10	267	Exon
ppM44	1646/4	H3 family 3b	1e-69	296	Exon
ppM45	816/6	Rwd domain containing 1	4e-19	296†	Intron + exon
Ppfor	679*	Foraging gene	4e-56	679	Intron + exon

***** *Ppfor* consensus sequence from Konakandla et al. 2006.

† The expected size derived from EST sequence was not observed after amplification due to presence of intro

Table 4.3. Genotype distribution for all markers among 23 examined F2 females. Allele types are expressed as A and B in all markers to ease the follow up of allelic segregation from F1 patents to F2 females.

Individual ID	ppM11	ppM30	ppM44	ppM45	*Ppfor*
F1F	AB	AB	AB	AB	AB
Male	B	B	B	B	B
F2-1	BB	AB	AB	AB	AB
F2-2	BB	BB	BB	BB	AB
F2-3	BB	AB	BB	AB	BB
F2-4	BB	AB	AB	BB	BB
F2-5	AB	AB	AB	AB	AB
F2-6	AB	AB	BB	BB	BB
F2-7	AB	BB	AB	AB	AB
F2-8	AB	AB	BB	AB	BB
F2-9	BB	AB	AB	BB	AB
F2-10	BB	AB	AB	AB	BB
F2-11	BB	-	AB	BB	BB
F2-12	BB	AB	BB	BB	BB
F2-13	AB	BB	BB	BB	BB
F2-14	BB	AB	-	BB	AB
F2-15	AB	AB	BB	AB	AB
F2-16	BB	AB	BB	BB	BB
F2-17	AB	AB	AB	BB	AB
F2-18	BB	AB	AB	AB	BB
F2-19	AB	AB	BB	AB	BB
F2-20	BB	AB	AB	AB	BB
F2-21	BB	BB	BB	AB	AB
F2-22	BB	BB	AB	AB	BB
F2-23	BB	AB	AB	AB	-
F2 Homozygote	**15**	**5**	**10**	**10**	**13**
F2 Heterozygote	**8**	**17**	**12**	**13**	**9**
Total F2	**23**	**22**	**22**	**23**	**22**

Table 4.4. Probabilities (p-value) of Chi-square test for equal chance of diallelic segregation of pairs of loci with twenty - two or twenty - three female mites in F2 generation.

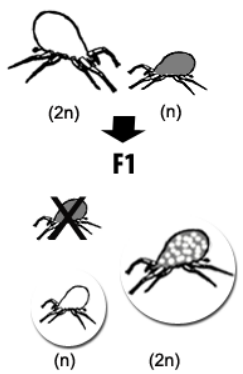
Under the haplodiploid genetic system, in which backcrossed female is heterozygote (AaBb) and male from maternal population is AB, the expected ratio for each genotype in assorted allele segregation will be ¼ AABB, ¼ AABb, ¼ AaBB, and ¼ AaBb.

Marker/loci	M30	M44	M45	Mppfor
M11	0.85	0.22	0.78	0.51
M30		0.51	0.96	0.37
M44			0.43	0.66
M45				0.94

Figure 4.1. Allele type distribution and quantification among six candidate loci within the

	Family 2		Family 3		Family 4		**Family 5**		Family 6		Family 7		Family 8		Family 9	
	F2	M2	F3	M3	F4	M4	**F5**	**M5**	F6	M6	F7	M7	F8	M8	F9	M9
ppM11	A	B	_	A	A	B	**A**	**B**	_	B	A	A	_	_	A	B
ppM21	A	A	B	B	C	B	**D**	**A**	A	B	C	B	A	B	B	B
ppM30	A	B	_	_	_	C	**A**	**C**	_	B	B	_	A	_	B	D
ppM34	A	C	B	D	B	D	**B**	**C**	C	D	C	D	C	D	B	D
ppM44	A	A	B	B	A	A	**A**	**B**	A	C	B	A	A	A	A	_
ppM45	A	C	B	C	A	C	**A**	**C**	C	C	C	C	A	C	A	C
# INFORMATIVE SITES	**4**		**2**		**4**		**6**		**3**		**3**		**3**		**4**	

pairs (F= female, M= male) of eight (families 2-9) G0 families of the predatory mite. Highest number of informative sites was used to determine the family (#5, in bold) used to study paternal genome inheritance. Underscore sign indicates no PCR amplification using the standard protocol.



IND./Loci	ppM11	ppM21	ppM30	ppM34	ppM44	ppM45	ppfor
Father	11A	21A	30A	34A	44A	45A	forA
Mother	11BB	21AB	30BB	34AB	44BB	45BB	forBB
F1M1	**11B**	**21A**	**30B**	**34A**	**44B**	**45B**	**forB**
F1M2	**11B**	**-**	**30B**	**-**	**44B**	**45B**	**-**
F1F	**11AB**	**21AA**	**30AB**	**34AA**	**44AB**	**45AB**	**forAB**

Figure 4.2. Genotyping Family 5, homozygote parental pair (Father - Mother) and three F1 offspring.

A- Two haploid sons (F1M) with maternal inheritance (white male in drawing) and 1 diploid daughter (F1F) with biparental inheritance (dotted female). Allele types distribution among six EST-derived candidate loci and ppfor marker. Five markers used were informative and two markers (ppM21 and ppM34) were not (in gray). F1 males only carried maternal inheritance but not from father as show in drawing.

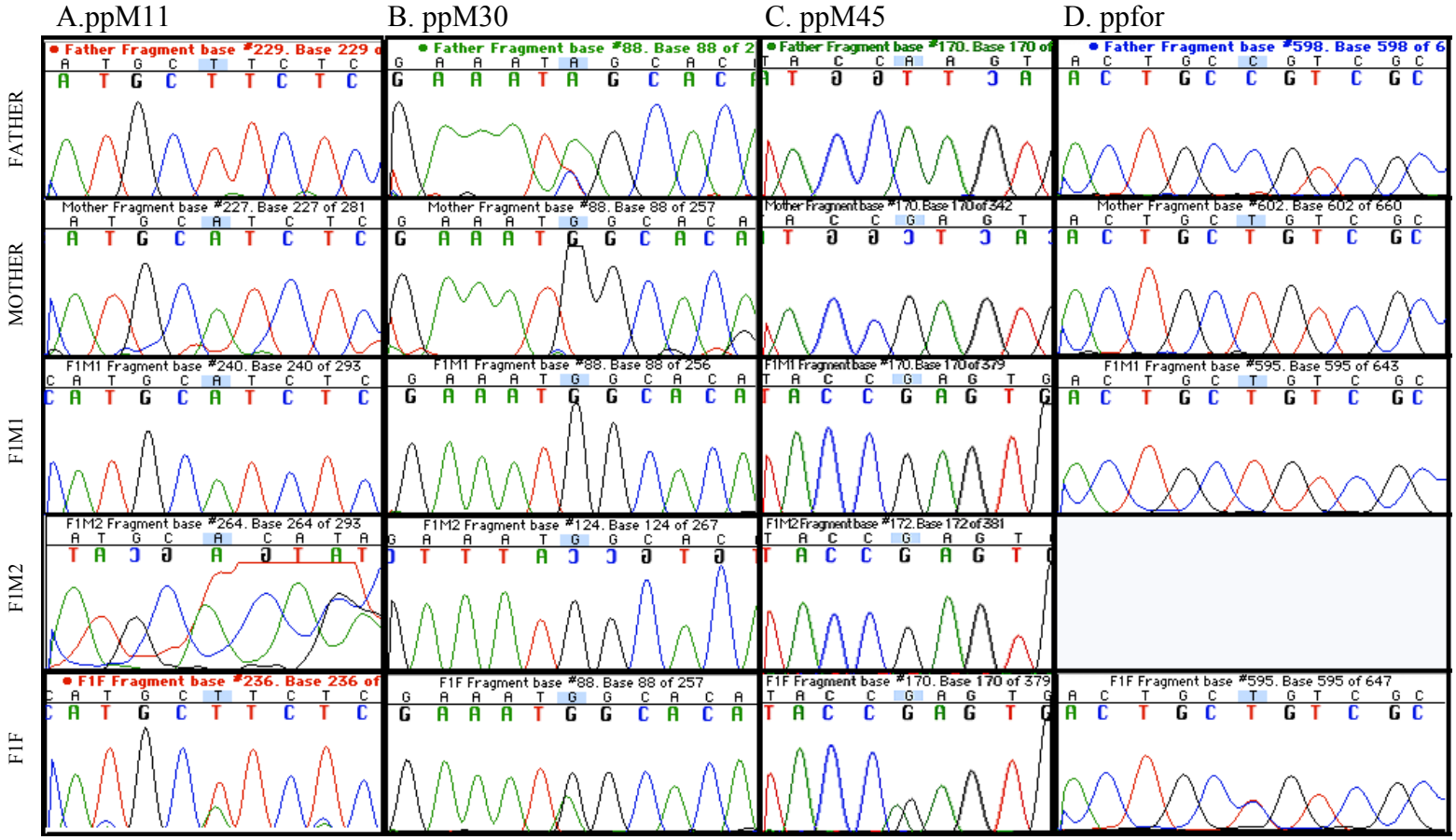


Figure 4.3. Genotyping Family 5, homozygote parental pair (Father - Mother) and three F1 offspring.

B- Chromatogram fragment of the DNA read franking the informative sites are showing the SNP variations in both parents and three progeny (two sons and a daughter) from ppM11, ppM30, ppM45, and ppfor.

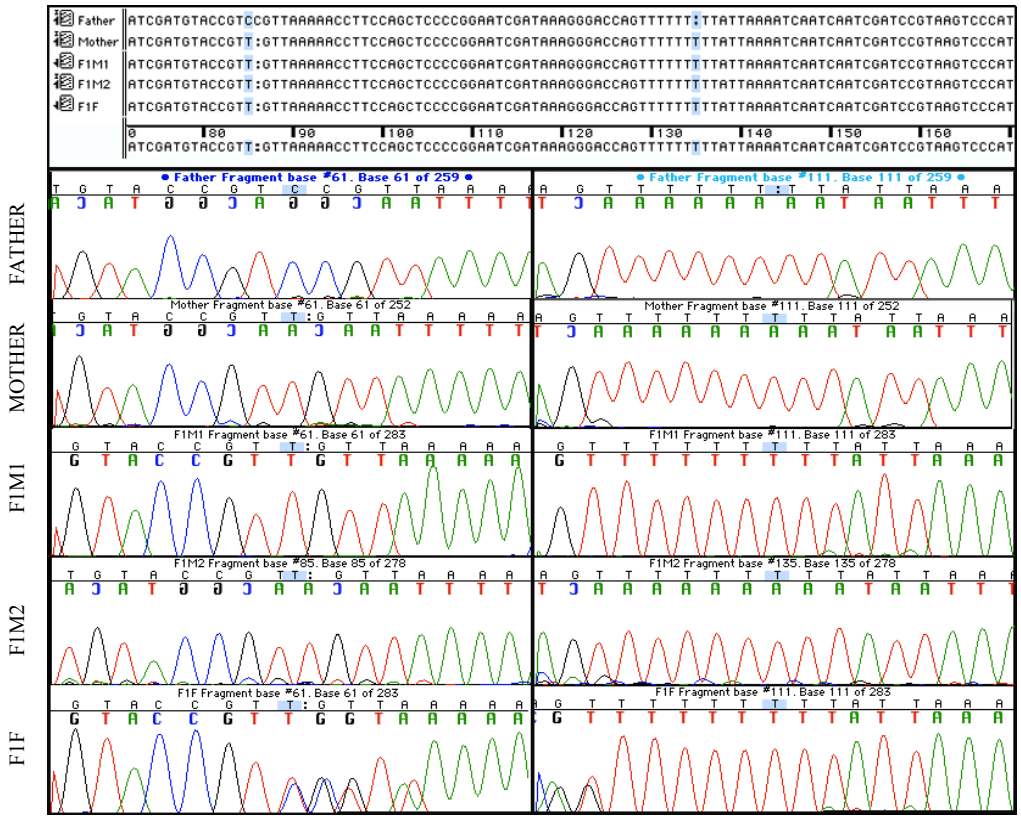


Figure 4.4. Genotyping Family 5, homozygote parental pair (Father - Mother) and three F1 offspring.

C- Chromatogram fragment of the DNA read franking the informative sites are showing the INDELs in marker ppM44 from both parents and three progeny (two sons and a daughter). Male and Female G0 (Father and Mother) are homozygotes. Allele in Father has an insertion of cytosine (I) and later deletion of one thymine (II) within the range of 50 nucleotides. This allele is inherited only by the daughter (F1F) and not by any of two sons examined (F1M1 and F1M2).

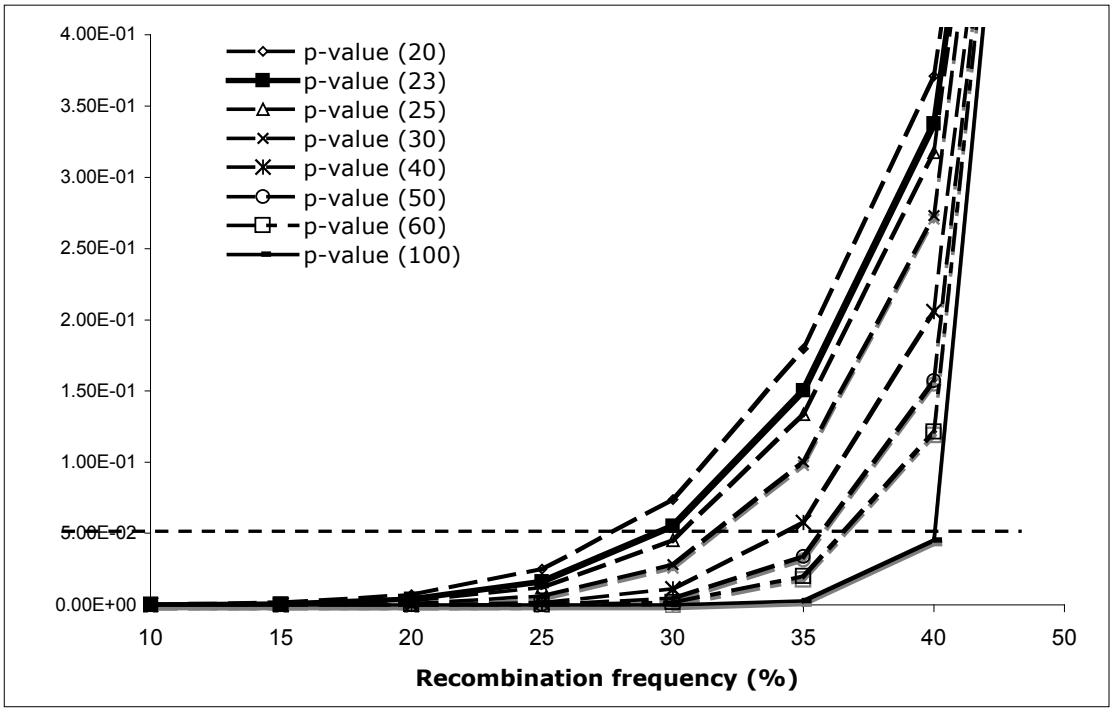


Figure 4.5. Estimation of the statistic power at sample number used to calculate linkage rate between pair of loci, at the level of confidence of P < 0.05 (horizontal dashed line).

At the sample size of 23 individuals, the statistic power is only able to detect 30% of recombination frequency (solid line with close boxes). The recombination frequencies ranges from 10% to 50%, representing 0% as no linkage between pair of loci and 50% as totally independent segregation or no linkage.

Chapter 5 - Conclusions and future direction

Expresses Sequence Tags (EST) are the most cost effective approach for transcriptome exploration (Nagaraj et al., 2007). As there are increases in high-throughput sequencing technology and its analysis methods through bioinformatics, the production of ESTs become accessible for almost all type of organisms (Church, 2006; Nagaraj et al., 2007). I have utilized this approach to target biological and evolutionary questions in *Phytoseiulus persimilis*, an organism with potential interest for genome sequencing due its role in applied entomology.

P. persimilis is a well-known predatory mite and one of the most popular biological control agents for spider mites. Because of its importance in agriculture, information on basic biology and ecology, including developmental rates, mating behavior, sex ratios, oviposition behavior, feeding habits, predator-prey interactions, prey consumption and preference, tolerance or resistance to pesticides, and interactions with other predatory mites and with insect predators, is readily available. However, there remains an incredibly large gap in genetic and genomic information on this species, as well as on beneficial species in general compared to pest organisms. This has limited development of alternative methods for pest control using beneficials. The paucity of genomic information available has held back this area of development in predatory mites. The Expressed Sequence Tag approach, a universal and cost-effective method to obtain large amount of genetic and physiological information, has made possible the *Pp* uniESTs dataset from which information about sixty genes involved in physiology of detoxification, endocrinology and behavior was obtained in the present study. Moreover, a large set of genes involved in reproduction, development, locomotion and molecules transportation may be targeted for mite enhancement and monitoring quality in mite reproduction at raring facilities; Implementing mite quality monitoring using biomarkers was first suggested by Dr. B. Ellis, a collaborator from the University of British Columbia.

ESTs have also proven to be an efficient approach to sample genomes for phylogenetic inferences, providing a larger number of characters to compare among distally related organisms. With the advent of phylogenomics, large set of biological sequences has become available from

genomic approaches, such as ESTs, which have been considered promising in resolving the deeper phylogenetic relationships within metazoan (Baurain et al., 2007). In this study, a more ancient evolution history of the mite clade (Acari) was estimated using a phylogenetic analysis assuming a global clock model. The divergence dating using amino acid sequences suggested a Devonian origin of anactinotrichid mites (487.6 ± 32.2 - 410.1 ± 6.1 Ma), about one hundred million years earlier than the most recent estimations using molecular phylogeny (Jeyaprakash and Hoy, 2009). However, using the same dataset, the origin for Chelicerata is similar to earlier study by Burmester (2001). Differences in mutation rates between lineages has been a particularly serious concern in molecular phylogeny and dating and being under constant scrutiny (Sanderson, 1997; Shultz and Churchill, 1999; Smith and Peterson, 2002; Welch and Bromham, 2005). Indeed, among Acari, high nucleotide substitution rate and gene arrangement has been documented (Navajas and Fenton, 2000), and significantly different evolutionary rates between lineages has been estimated, i.e. higher rate in Mesostigmata than in other parasitiforms (Murrel et al., 2005; Klompen et al., 2007). The discrepancy of the mites' divergence time with a previous study by Jeyaprakash and Hoy may be due to my use of a global clock. Thus, application of a local clock to the Acari branch, or smoothing the evolutionary rate in the phylogenetic analysis (see review in Welch and Bromham, 2005), would be appropriate to avoid overestimation of divergence time. Some more versatile software programs, such as Phylobayes (Lartillot et al., 2009), utilize Markov Chain Monte Carlo (MCMC) model in the sampling methods, which allows greater flexibility in combining phylogenetic models into analysis (Larget and Simon, 1999). This would be the next step to explore in support of the finding regarding an earlier divergence of the parasitic life form in Acari.

Phytoseiulus persimilis ESTs were also used successfully to develop and screen for molecular markers in different mite populations. Single nucleotide polymorphisms and short repeats were easily observed and determined for diverse use such as the study of paternal genome loss in males predatory mites. Five EST-derived markers were useful in determining haploid condition in adult males of *P. persimilis* and revealing that paternal contribution was lost during early developmental stage. My observations using ESTs differ from previous observations using RAPD (random amplification of polymorphic DNA) markers (Perrot-Minnot, 1995) and DALP (direct amplification of length polymorphism) markers (Perrot-Minnot, 2000). The

preliminary observations of paternal genome loss in the predatory mite have helped to identify biological and methodology constrains in the approach. Examination of more male individuals and a larger sample size may be required to confirm my conclusions. As the development of high-throughput sequencing technology makes more cost effective the EST approach, the evolutionary process in genome loss shall be unveiled in the near future.

References

Baurain, D.; H. Brinkmann; H. Philippe. 2007. Lack of Resolution in the Animal Phylogeny: Closely Spaced Cladogeneses or Undetected Systematic Errors? Molecular biology and evolution 24: 6-9.

Burmester, T. 2001. Molecular evolution of the arthropod hemocyanin superfamily. Molecular Biology Evolution 18: 184–195.

Church, G.M. 2006. Genomes for all. Scientific American 294: 46-54.

Hoy, M.A. 2009. The predatory mite *Metaseiulus occidentalis*: mitey small and mitey large genomes. BioEssays 31: 581-590.

Jeyaprakash, A. and M.A. Hoy. 2009. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. Experiment and Applied Acarology 47: 1-18.

Klompen, H.; M. Lekveishvili, and W.C. Black IV. 2007. Phylogeny of parasitiform mites (Acari) based on rRNA, Molecular Phylogenetics and Evolution 43: 936-951.

Larget, B. and D. L. Simon. 1999, Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. Molecular Biology and Evolution 16: 750-759.

Lartillot, N.; T. Lepage; and S. Blanquart. 2009. PhyloBayes 3. A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics Applications Note 25: 2286–2288.

Murrell, A.; S.J. Dobson; D.E. Walter; N.J.H. Campbell; R. Shao; and S.C. Barker. 2005. Relationships among the three major lineages of the Acari (Arthropoda : Arachnida) inferred from small subunit rRNA: paraphyly of the parasitiformes with respect to the opilioacariformes and relative rates of nucleotide substitution. Invertebrate Systematics 19: 383-389.

Nagaraj S.H.; N. Deshpande; R.B. Gasser; and S. Ranganathan. 2007. ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. Nucleic Acids Research 35: W143-W147.

Navajas, M and B. Fenton. 2000. The application of molecular markers in the study of diversity in acarology: a review. Experimental Applied Acarology 24: 751-774.

Perrot-Minnot, M.-J. and M. Navajas. 1995. Biparental inheritance of RAPD markers in males of the pseudoarrhenotokous mite *Typhlodromus pyri*. Genome 38: 838-844.

Perrot-Minnot, M.-J.; J. Lagnel; A. Migeon; and M. Navajas. 2000. Tracking paternal genes with DALP markers in a pseudoarrhenotokous reproductive system: biparental transmission but haplodiploid-like inheritance in the mite *Neoseiulus californicus*. Heredity 84: 702-709.

Sanderson, M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology Evolution 14: 1218-1231.

Schultz, T. R. and G. A. Churchill. 1999. The role of subjectivity in reconstructing ancestral character states: A Bayesian approach to unknown rates, states, and transformation asymmetries. Systematic Biology 48: 651-664.

Smith, A.B. and K.J. Peterson. 2002. Dating the time of origin of major clades: molecular clocks and the fossil record. Annual Review Earth Planet Sciences 30: 65-88.

Welch, J.J and L. Bromham. 2005. Molecular dating when rates vary. Trends Ecology Evolution 20: 320-327.

Appendix A - Pp UniESTs encoding ACE-like arachnid specific esterases.

A_a- Nine amino acid sequences from conceptual translation of Pp UniESTs encoding ACE-like arachnid specific esterases.

>PPE013_C07.f

IVELVTGRIEGVPATEQSVHKFLGIPFAEPPIGDLRFKHPVAVKSWSPKIVKADTK
PFPCLQGPLYINSNLTIETTNSTEDCLYLNVWTPDDCVAIGACQSKKSVMVFIYGGTYTF
GSSGWNMYDAEQLALRGDVVVVTFNYRVGPLGFL

>PPE013_C07.r (stop codon removed for sequence alignment)

RQAAEMVGDLAMVCPTKYFAEEAAAQNLSVHYYEFDFRSSFSTWPD*VGTTHG
EEIPFVFGHPMSGLEPNATAQDKDMSAEIIGIWTDFAKTGLTPTKVRGTPWPAYTEKSQT
YLRFGQESTGIGRGPNERMCNSWRKYL

>PPE014_O05.f

TSVHGNPVVQGPLGIITGKFQTVLDAEVESFLGIPYAKPPVGNRRFALPETFGTVG
NLSATEYSRQCPQPYLRENEPGRWHLDEDCLYLNVFRKRGTAETEGKAVIAIIHGGGYII
GTASESYVSPASLVAFGDVIVVSFNYRLGIFGFADMKELAPGNLGLYDQRLALQWIQDH
IGGFGGNPEKVTVIGVSAGSMSLSAQIITPIDEKNLFQSAVLDAGVFAGFDEDSES
SFTRVKKIAKKLGCPVGSSEMLDCLR

>PPE0138_P16.f

SSGSIRGTSVDFRGVKVYQFLGIPFAEPPLNELRFQKPVPKKPWNGVLSVNKWGS
ACMQPVFPGFNTELHLSEDCLILNVFTTDAAFQDKQNGKKNSLRPVMVWIHGGDFNFG
SANTASQYDGTPITGLKDVIIVSINYRLSSLGFLHLPEAGVPGNMGLWDQQLALKWVKD
NIEHFGGDPNRVTIFGESAGSMSVSAHIVSPHSKGLFKNAIIQSGSI

>Contig3716

ANRGVKVYRYLWEHKPSTSYWPKWSGATHCDDVPFTMGSQFDIGNKAEKSKQ
ASEGLVRYMQTPITEAEKQLIKDSLKMIGDFVKTGNPSRPDGSQWPLYTAEKREVVQIG
TTSFSDISLQNSRCPIWTDFI

>Contig4735

KRAVMQSGSPFSFVPRNTKDQGETLFRSLASYTDCMSVHFNSTLRYDDVLQRM
KKQPFENIIAASEKFNGKGVNSFFPVMGEEFIPMNPKDSLLLKRFSNVDLLVTTTKSEGA
YFLQHFLSPFTNVADADKINPGEIVFYLRVFLSALLGGKPTASLNELAQNTEPETREEKV
QFLKNISAVIGDYPYLCATTDFGVEYANPRHNVYHMQYDHRPWFLLHPTWFPSTHGDD
IMFWLGSVYKLKERTRADERVADELMAILTAFAGKGTPQTRGKLTWPRINQGGYYMK
VGSEVTELLRTPAANCRQWSRHYPQ

>Contig5707

LFAAIAVGGVVGGSPVVEVSTGTVSGKVETHRGAEVESFLGMPYAKPPVGELRF
ALPETVAPFGDLNATAFSASCPQSGDRPPLTFPESEDCLYLNIFRKTGAQQDSKKPVLFVI
HGGAFRVGAASQGFYNGLPIAAIGDVIVVTINYRLGVLGFADMKDLAPGNLGLFDQLLA
LEWVHENIAAFGGDPDRVTILGVSAGSMSVSALVNTPLVRGRNLFKQAVMDAGVTSRT
VVMSQELSLGRVKKIAAKVSCETEGEKMLACLRGANATQLTDASFDREFSPILTFAPTV
DGKFIAAEPSRDIQQNSDKFVDVRMIVGVAKDEGTLFAALYPAANTIKTESEFLDLAKDI
SRGFLYPLDFDDQVVRDAVTQTYFSKSADRQSDAAEFVADGSFVCPTNAFVKSYAKTH
ENVFVYNFEKVMKRKYLKFGPEKLGAYHFSPFANMFGAFLTMSEEELGGPLDPEDEQF
MVDSMNLLVDFVNSDEVPKFRGVTWPNYSEGEGILTINDT

>Contig5966

GHQGHLDQVLVLNWVSQNIRYFGGDPNRVTLFGVSAGSFSISWHLLTGFSAGLF
HAAVIDAGVLTHTHAETRRDHVSRAQKMVRTSICKSLKEFGDRATKEQRRRILNCLMKI
DANELVELQQRYASSQTYAFRPTFNNQEYLPRSPTCMTNEEAFSVNVPIIIGDATNEGLF
LLPRKVHDPLPSFSSFDEVLEWDIDILGGHHSAQAPNYGPNNSTIAEIYNNETNDREPEA
AALFNAATQIIGDGLFVCPVMNFADRYSSVQANVYFYRWERVRVDETYPHERADGAY
HGLMFYTGVGSQYLYLQGQMADADKSYIENTIKMIADFASNPTGS

>Contig6677

RNIPASSLLPKSNSDARGNLSFIPTTDGAFLPRDVEEYVAKNPAELRKVRTIVGYS
LDEGSMFVRLLDPKFDFSTARPRDEILDYCGKISETFDFPFNASKRETREKIGQLYVDDNS
GNAFKAVSSFIADGWFKCPINTFIRSYSRHNDKVFAYQFERRLNRPYFKLFDPKVLGAFH
YSPYLHFGGAILLDGGTVNEGDKQFSLDAMSMISKFSKSDGPLMFRDVGWPPFSESGEV
FIFKETPTVAKEL

A_b -Alignment of *Pp* amino acid sequences PPE0138_P16.f, PPE013_C07.f, PPE014_O05.f, and Contig5707 (group 1), to reference ACEs from three insect species (*Bombyx mori* (BM ACE1-2: NP_001037380.1, NP_001108113.1, NP_496963.1, NP_496962.1), *Drosophila melanogaster* (DM ACE: NP_476953.1), *Tribolium castaneum* (TC ACE1-2: EFA04156.1, EEZ99262.1), the cattle tick *Rhipicephalus microplus* (RM ACE1-3: CAA11702.1, CAB93511.1, AAP92139.1), the nematode *Caenorhabditis elegans* (CE ACE1-4: NP_510660.1, NP_491141.1), and *Homo sapiens* (HS ACE: NP_000656.1).

Numbering of the amino acid residues is from the N-terminus of mature proteins. Position of identical amino acid residues is indicated by asterisks (*) and of conservative substitutions by dots.

```
TCACE-2           -------MGSN-------------------------------------------------- 4
BMACE-2           MINYGKIVFTK-------------------------------------------------- 11
DMACE             ------------------------------------------------------------
TCACE-1           --MTGAWAACLLVILLPSCIPSPHRGRHHPPEP---------------------------- 31
BMACE-1           --MRVVLAALTALAARTLAGPHEHRARHHAPAPPQPYHGHGEAVRYNPELDTILPRLEDH 58
CEACE-1           ------------------------------------------------------------
RMACE-1           ----MDPEQDMLHENLASCH---------------------------------------- 16
HSACE             ----MRPPQCLLHTPSLASP---------------------------------------- 16
CEACE-3           ------------------------------------------------------------
CEACE-4           ------------------------------------------------------------
CEACE-2           ------------------------------------------------------------
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           ----------------MYSRIVVLLLYGYALRSAHARCYIFQPGDPIQDTDEKIYLAQPF 44
PPE014_O05.f      ------------------------------------------------------------
Contig5707        ------------------------------------------------------------
RMACE-2           ------------------------------------------------------------
```

```
TCACE-2           ------LVVVVVVVVVVVVASLSASARAYSWPSEETTTRPPQARDFHSDPLVVETTSGLVR 58
BMACE-2           ------LLLCVLMSGTFARSWANHHDTTTSTTQTTPTTSPVPKNIHNDPLIVETKSGLIK 65
DMACE             -------MAISCRQSRVLPMSLPLPLTIPLPLVLVLSLHLSGVCGVIDRLVVQTSSGPVR 53
TCACE-1           -------HAEAYHMSRDPFDPHRDSEEFRRDAPDDKREFTRRDSEDDPLVIQTKKGKVRG 84
BMACE-1           ETSSKRASDAETSSKRTKYEERFYSNHERAAELMADEPVSEKGDEEDPLVIRTRKGKVRG 118
CEACE-1           ------------------------------MRNSLLFFIFLPSTILAVDLIHLHDGSPLFG 31
RMACE-1           --------LTLLALLVCGGVVLRCLSIEPEEDASNRVEDQDAEDPVETVVVETAWGPVKG 68
HSACE             --------LLLLLLWLLGGGVG-------------------AEGREDAELLVTVRGGRLR 49
CEACE-3           ------------------MRRRRVLLLLLSTTTFLRVSTAQQDEPKASVVEVQTKLGTVR 42
CEACE-4           ------------------------MKPKLAFFAFFIFITVFIDSVQAVHPVVLETKLGDIK 37
CEACE-2           -----------MRAPVIGRHLTYHVFCQFALVTLFIVRRIEPRSIVRGDHVVHTPLGTIR 49
PPE0138_P16.f     ------------------------------------------------------SSGSIR 6
PPE013_C07.f      ----------------------------------------------------IVELVTGRIE 10
RMACE-3           NVTIPLEAGIPTTAQPASTGKDQLSLVVPPDDIQSLSTTTSRPISTEESPIVETNSGPVQ 104
PPE014_O05.f      ------------------------------------------TSVHGNPVVQGPLGIIT 17
Contig5707        ----------------------------------LFAAIAVGGVVGGSPVVEVSTGTVS 25
RMACE-2           -----------MYVRVSLVFASAWIIACLGYTETREARGIAVLEDGASPVVQIHAGKLRG 49

TCACE-2           GKAKTVLGREVHVFTGIPFAKPPIEQLRFRKPVPIDPWHGILD-ATKLPNSCYQERYEYF 117
BMACE-2           GYAKTVMGREVHIFTGIPFAKPPLGPLRFRKPVPIEPWHGVLE-ANLMPNSCYQERYEYF 124
DMACE             GRSVTVQGREVHVYTGIPYAKPPVEDLRFRKPVPAEPWHGVLD-ATRLSATCVQERYEYF 112
TCACE-1           ISLTAATGKKVDAWLGIPYAQKPLGNLRFRHPRPAEKWEGVMN-TTSQPNSCVQIIDTVF 143
BMACE-1           ITLTSATGKKVDAWFGIPYAQKPMGDLRFRHPRPVEDWGDEILNTTTLPHSCVQIVDTVF 178
CEACE-1           EEVLSQTGKPLTRFQGIPFAEPPVGNLRFKKPKPKQPWRIPLN-ATTPPNSCIQSEDTYF 90
RMACE-1           FIAQSPLGKPVRVFYGIPYAKPPTGKRRFDRAESIEEPWTDVLDATVKPNSCFQVLDTLY 128
HSACE             GIRLKTPGGPVSAFLGIPFAEPPMGPRRFLPPE-PKQPWSGVVDATTFQSVCYQYVDTLY 108
CEACE-3           GTESDHGNKRVRSFLGVPFAEPPINEHRFKKPTPARPWNGTISANTLSPACFQGRDS-YD 101
CEACE-4           GTEFFFLSKKIRTFFGVPFAEPAVEDFRFRKPREKKQWRGLYDATKPANACFQTRDN-YN 96
CEACE-2           GVGQTFDGAKVSAFLGVPYAKPPIGSRRFKMAEMIDRWSGELEARTLAKTCYLTIDS-AF 108
PPE0138_P16.f     GTSVDFRGVKVYQFLGIPFAEPPLNELRFQKPVPKKPWNGVLSVNKWGSACMQPVFPGFN 66
PPE013_C07.f      G--VPATEQSVHKFLGIPFAEPPIGDLRFKHPVAVKSWSPKIVKADTKPFPCLQGPLYIN 68
RMACE-3           GRRVYAANRTLYQFVGIPFAEPPVGPLRFRNPVPVKPWS-SVYQATKKPFPCLQTDFYIN 163
PPE014_O05.f      GKFQTVLDAEVESFLGIPYAKPPVGNRRFALP----------ETFGTVGNLSATEYSRQC 67
Contig5707        GKVETHRGAEVESFLGMPYAKPPVGELRFALP----------ETVAPFGDLNATAFSASC 75
RMACE-2           AKRVVLGEKFAYAFTGVPYAKPPVGELRYQKPESAQPWV---EEVKDATVTPPSCMQGNV 106
                                  : *:*:*: .    *:  .

TCACE-2           PGFEGEEMWNPNTNISEDCLYLNIWVPQRLRIRHHGEKLPQDRPK--------------- 162
BMACE-2           PGFEGEEMWNPNTNISEDCLYLNIWVPQHLRVRHHQDKPLAERPK--------------- 169
DMACE             PGFSGEEIWNPNTNVSEDCLYINVWAPAKARLRHGRGANGGEHPNGKQADTDHLIHNGNP 172
TCACE-1           GDFPGATMWNPNTPLNEDCLYVNVVVP--KPRPT-------------------------- 175
BMACE-1           GDFPGAMMWNPNTDMQEDCLYINIVTP--RPRPK-------------------------- 210
CEACE-1           GDFYGSTMWNANTKLSEDCLYLNVYVPGKVDPNK-------------------------- 124
RMACE-1           GNFSGSTMWNANTEMSEDCLKLNVWAPGPPTSSGG------------------------- 163
HSACE             PGFEGTEMWNPNRELSEDCLYLNVWTPYPRPTS--------------------------- 141
CEACE-3           PTFWGSEMWNANTPVSEDCLYVNIWAPADAYN---------------------------- 133
```

```
CEACE-4           TSFWGSEMWNANTQISEDCLYLNIWAPADAYN---------------------------- 128
CEACE-2           PQFPGAEMWNPPGAISEDCLNMNIWVPEDHD----------------------------- 139
PPE0138_P16.f     TELHLSEDCLILNVFTTDAAFQDKQNGKKNSL---------------------------- 98
PPE013_C07.f      SNLTIETTNS-----TEDCLYLNVWTP-DDCVAIGACQS--------------------- 101
RMACE-3           SNVTIPTANS-----SEDCLYLNVWTPSRECVLGKFSCV--------------------- 197
PPE014_O05.f      PQPYLRENEPGRWHLDEDCLYLNVFRKRGTAETEG------------------------- 102
Contig5707        PQSGDRP--PLTFPESEDCLYLNIFRKTGAQQDSK------------------------- 108
RMACE-2           FSPRNLLWLPYDHQKSEDCLYLNVWTPRLN-TSAG------------------------- 140
                                *.   :

TCACE-2           ------VPVLVWIYGGGYMSGTS-TLDIYDADIIAATSDVIVASMQYRVGAFGFLYLSKY 215
BMACE-2           ------VPILVWIYGGGYMSGTA-TLDLYKADIMASTSDVIVASMQYRVGAFGFLYLNKY 222
DMACE             QNTTNGLPILIWIYGGGFMTGSA-TLDIYNADIMAAVGNVIVASFQYRVGAFGFLHLAPE 231
TCACE-1           -----SAAVMVWVFGGGFYSGTN-TLEVYDHNILVSEENIILVSMQYRVASLGFLYFG-T 228
BMACE-1           -----NAAVMLWVFGGGFYSGTA-TLDVYDPKILVSEEKVVYVSMQYRVASLGFLFFD-T 263
CEACE-1           -----KLAVMVWVYGGGFWSGTA-TLDVYDGRILTVEENVILVAMNYRVSIFGFLYMN-R 177
RMACE-1           ----RPLAVLVWIYGGGFYSGTS-TLDVYDARTLVSEENVVVVSMNYRVASLGFLSFG-N 217
HSACE             -----PTPVLVWIYGGGFYSGAS-SLDVYDGRFLVQAERTVLVSMNYRVGAFGFLALPGS 195
CEACE-3           ------LTVLVWLFGGGFWYGSP-SLLLYDGKELATRGNVIVVNINYRVGPFGYLFLD-H 185
CEACE-4           ------LTVMVWFFGGGFYSGSP-SLSIYDGKALTSTQNVIVVNINYRLGPFGFLYLG-H 180
CEACE-2           ------GSVMVWIYGGGFFSGTP-SLDLYSGSVFAAKEHTIVVNVNYRLGPFGFLYFGDD 192
PPE0138_P16.f     ------RPVMVWIHGGDFNFGSANTASQYDGTPITGLKDVIIVSINYRLSSLGFLHLP-E 151
PPE013_C07.f      -----KKSVMVFIYGGTYTFGSS-GWNMYDAEQLALRGDVVVVTFNYRVGPLGFL----- 150
RMACE-3           -----PKTVIVYIYGGTFSFGSS-GWDWYDGKEFVARGDVVMVSMNYRVGPMGFFHSGTT 251
PPE014_O05.f      ------KAVIAIIHGGGYIIGTA-SESYVSPASLVAFGDVIVVSFNYRLGIFGFADMKEL 155
Contig5707        ------KPVLFVIHGGAFRVGAA-SQGFYNGLPIAAIGDVIVVTINYRLGVLGFADMKDL 161
RMACE-2           ------LPVMAWIHGGGFQEGSA-AIPLDDGTYLAAFGNVVVVTIAYRLQSFGFLYDETS 193
                        .::  ..** :  *:       .   :.     : . . **:  :*:

TCACE-2           FP-RGSEEAPGNMGMWDQALAIRWIKENAAAFGGDPDLITLFGESAGGGSVSILLLSPVT 274
BMACE-2           FS-PGSEEAPGNMGLWDQQLAIRWIKENARAFGGDPELITLFGESAGGGSVSLHMLSPEM 281
DMACE             MPSEFAEEAPGNVGLWDQALAIRWLKDNAHAFGGNPEWMTLFGESAGSSSVNAQLMSPVT 291
TCACE-1           PD------VPGNAGLFDQMMALQWVRDNIAAFGGNPNNITLFGESAGAVSVSLHLLSPLS 282
BMACE-1           AD------VPGNAGLFDQLMALQWVKDNIGYFGGNPHNITLFGESAGAVSVSLHLLSPLS 317
CEACE-1           PE------APGNMGMWDQLLAMKWVHKNIDLFGGDLSRITLFGESAGAASVSIHMLSPKS 231
RMACE-1           ET------LPGNAGLYDQYMALKWVQENVAAFGGDPDRVTLFGESAGAASAGLHVLSPLS 271
HSACE             RE------APGNVGLLDQRLALQWVQENVAAFGGDPTSVTLFGESAGAASVGMHLLSPPS 249
CEACE-3           ED------VPGNMGMLDQQLALYWIRDHIFSFGGNPARISLVGESAGAASIVAHLIAPAS 239
CEACE-4           PD------APGNMGLLDQQLALHWVRQNIVSFGGNPDKVAVFGQSAGAASIVAHLIAPGS 234
CEACE-2           SP------IQGNMGLMDQQLALRWVHENIGAFGGDRSRVTLFGESAGSASTTAHLFAPNS 246
PPE0138_P16.f     AG------VPGNMGLWDQQLALKWVKDNIEHFGGDPNRVTIFGESAGSMSVSAHIVSPHS 205
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           HS-------SGNAGLHDQLLAMKWVKQNIRNFGGDPDDVTLVGQSAGAISIGLHLVSPLS 304
PPE014_O05.f      AP--------GNLGLYDQRLALQWIQDHIGGFGGNPEKVTVIGVSAGSMSLSAQIITPID 207
Contig5707        AP--------GNLGLFDQLLALEWVHENIAAFGGDPDRVTILGVSAGSMSVSALVNTPLV 213
RMACE-2           AP--------GNMGLHDQQLALKWIQENIAAFGGNPGEVTLFGWSAGGISTGFHLISPGS 245
```

```
TCACE-2           -KGLARRGILQSGTMNAPWSYMSGERAQQIGKVLVEDCGCNVS---LLETRPHEVIDCMR 330
BMACE-2           -KGLFKRGILQSGTLNAPWSWMTGERAQDIGKVLIDDCNCNSS---LLAKDPSLVMDCMR 337
DMACE             -RGLVKRGMMQSGTMNAPWSHMTSEKAVEIGKALINDCNCNAS---MLKTNPAHVMSCMR 347
TCACE-1           -RNLFSQAIMESGSATAPWAIISREESILRGLRLAEAVGCPHE-----RHELSAVIDCLK 336
BMACE-1           -RNLFSQAIMQSGAATAPWAIISREESILRGIRLAEAVHCPHS-----RSDLAPMIECLR 371
CEACE-1           -APYFHRAIIQSGSATSPWAIEPRDVALARAVILYNAMKCGNMSLI--NPDYDRILDCFQ 288
RMACE-1           -ESLFHRVILQSGSPAVPWGFQDRDKARQSAKKLATALRAP-------DSLDQETLDSLR 323
HSACE             -RGLFHRAVLQSGAPNGPWATVGMGEARRRATQLAHLVGCPPGGTGGNDTELVACLRTRP 308
CEACE-3           -KGLFQNGILQSGSLDNKWSMDSPKRAKQKSTALADLVGCNQTK--ITDQTACLRNTPAQ 296
CEACE-4           -RGLFKNAILQSGSLENTWAINSPFRAKQKSEKLLELVGCNKTT--VDTSMACLRLVSPE 291
CEACE-2           -HKYFRNIIAKSGSIINSWASATPPTMLDLSFRLAKKVNCSSPD--MNAIVKCLRSVPAH 303
PPE0138_P16.f     -KGLFKNAIIQSGSI---------------------------------------------- 219
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           -KGLFKRIIMESGSPYFRIADNTREGPHKVEKLARALQCARNDMT---IESHMAEMVECL 360
PPE014_O05.f      -EKNLFQSAVLDAGVFAG-FDEDSESSFTRVKKIAKKLGCPVGS---------SEMLDCL 256
Contig5707        RGRNLFKQAVMDAGVTSRTVVMSQELSLGRVKKIAAKVSCETEG---------EKMLACL 264
RMACE-2           -QTLFKRAIVQSAAVTKKGRARDKTEMLEYSQKFAANFGCYGGDS---AANASQDIADCM 301


TCACE-2           AVEAKTISLQQWNSYSGILGFPSTPTVDGVFMPKHPMDMLAEGDYEDMEILVGSNQDEGT 390
BMACE-2           GVDAKTISVQQWNSYTGILGFPSAPTVDGIFLPKDPDTMMKEGNFHNSEVLLGSNQDEGT 397
DMACE             SVDAKTISVQQWNSYSGILSFPSAPTIDGAFLPADPMTLMKTADLKDYDILMGNVRDEGT 407
TCACE-1           KKDPIDLVNNEWG-TLGICEFPFVPVIDGAFLDESPTRALANKNFKKTNILMGSNTEEGY 395
BMACE-1           KKNADELVNNEWG-TLGICEFPFVPIIDGSFLDEMPVRSLAHQNFKKTNILMGSNTEEGY 430
CEACE-1           RADADALRENEWAPVREFGDFPWVPVVDGDFLLENAQTSLKQGNFKKTQLLAGSNRDESI 348
RMACE-1           CERPEDIVNNETN-SGGVVDFPFVPVADGVFLPDTPQALTDKGSFARNISVMLGSNANEG 382
HSACE             AQVLVNHEWHVLP-QESVFRFSFVPVVDGDFLSDTPEALINAGDFHG-LQVLVGVVKDEG 366
CEACE-3           LLIDNIWNVG---LNFLEFPFAIVSKDQNFFKHLDGFIALREGTYSTDVNLMFGINHDEG 353
CEACE-4           QLSLSTWNIS---LTYLEFPFVIVSRDKHFFGHLDAHAALREGDFNRDVNLMIGMNKDEG 348
CEACE-2           LVQAEADNISGDIGPPMTFAYVPVSSDANFFQG-DVFQKLANKQFKKDVNIIFGSVKDEG 362
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           RKIDGKELLIMSNTIFGVHALTFFPVFGDDIIPDDPYLMMEQKKFHKADLLIGNNLDEGS 420
PPE014_O05.f      R----------------------------------------------------------- 257
Contig5707        RGANATQLTDASFDREFSPILTFAPTVDGKFIAAEPSRDIQQNSDKFVDVRMIVGVAKDE 324
RMACE-2           RTVNASLIVAVEATFVGSGSGKFEPIYGDEFLPIEPRMADFPGDKDVMIGQTANEGSNIL 361


TCACE-2           YFLLYD--------FIDFFEKDGPSFLQRDK--------YHDIIDTIFKNMSRLERDAIV 434
BMACE-2           YFLLYD--------FLDYFEKDGPSFLQREK--------FLEIVDTIFKDFSKIKREAIV 441
DMACE             YFLLYD--------FIDYFDKDDATALPRDK--------YLEIMNNIFGKATQAEREAII 451
TCACE-1           YFIIYY--------LTELFRKEENVYVNRQE--------FLRAVTELNPYFNAISRQAIV 439
BMACE-1           YFILYY--------LTELFPKEENVGISREQ--------FLQAVRELNPYVNDVARQAII 474
CEACE-1           YFLTYQ--------LPDIFPVADFFTKTDFIKDRQLWIKGVKDLLPRQILKCQLTLAAVL 400
RMACE-1           SWFLQY--------FFGFPVTDETPEVTKEN--------FTAVLEALDPSLEQTPIAEIM 426
HSACE             SYFLVY--------GAPGFSKDNESLISRAE--------FLAGVRVGVPQVSDLAAEAVV 410
```

```
CEACE-3           NFWN-----------IYNLAKFFDKQSVKPGLDRDEFHECVDTAFAVQPELVRTAAKYVY 402
CEACE-4           NYWN-----------IYQLPQFFDKADPPE-LNRTEFDFLIDRTFSIQPDIIRSAAKYIY 396
CEACE-2           TYWLPYYMSLPKYGFAFNHTISAEDPHNRALITRDHYEESMRAFMPYFAGSKLVLNAFMN 422
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           YFVFYL--------FGRALDLEQAHKITKYE--------VDLYVSYCLQMLLRKNVSPIR 464
PPE014_O05.f      ------------------------------------------------------------
Contig5707        GTLFAA------------------LYPAANTIKTESEFLDLAKDISRGFLYPLDFDDQVV 366
RMACE-2           YTTFRD------------------TFSEALPPR-KINKAEMIHFLGSLYKLSLSDIEKLQ 402


TCACE-2           FQYTDWEHVNDGYLNQKMVGDVVGDYFFICPTNDFAELAAERGMKVYY-----------Y 483
BMACE-2           FQYTDWEEITDGYLNQKMIADVVGDYFFVCPTNYFAEILADAGVGVYY-----------Y 490
DMACE             FQYTSWEGNPG-YQNQQQIGRAVGDHFFTCPTNEYAQALAERGASVHY-----------Y 499
TCACE-1           FEYTNWLNPDDPVSNRDSLDKMVGDYHFTCNVNEFAHRYAETGNTVYM-----------Y 488
BMACE-1           YEYTDWLNPEDPVKNRNALDKMVGDYHFTCGVNEFAHRYAETGNNVYT-----------Y 523
CEACE-1           HEYEPQDLPVTPRDWINAMDKMLGDYHFTCSVNEMALAHTKHGGDTYY-----------Y 449
RMACE-1           KTYTAGEIPSTAADILKALDSIVGDYHFTCPVVRWADTFARAGIPVYQ-----------Y 475
HSACE             LHYTDWLHPEDPARLREALSDVVGDHNVVCPVAQLAGRLAAQGARVYA-----------Y 459
CEACE-3           SDPKCTDPKKKTDFYTEQVNQMVGDYFFTCDSIWFAHNYPKMAGNQSN--------VFVY 454
CEACE-4           SDPNCTDHGRKTRFYAGQMNQIVGDYFFSCDSLWLADQFRSIPRVKSSSPQRKPGKVFVY 456
CEACE-2           SYEHVSTSNVPEERYRDGVARFLGDLFFTCSLIDFADLISDNIFGNVY----------MY 472
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           HYYLSHIGERENIKALQKAAEAVGDFAIICPTKYFAESFASMNNKVHY-----------Y 513
PPE014_O05.f      ------------------------------------------------------------
Contig5707        RDAVTQTYFSKSADRQSDAAEFVADGSFVCPTNAFVKSYAKTHENVFVYNF-----EKVM 421
RMACE-2           KEYMGEIGDYDYDALRQALAETKGDTHVKCGAINTACKLANATANAQSGKE-----VHFY 457


TCACE-2           FFTHRTSTSLWGEWMGVMHGDEIEYVFGHPLNMSLQ----FNSRERELSLKIMQAFARFA 539
BMACE-2           YFTHRTSTSLWGEWMGVMHGDEMEYVFGHPLNMSLQ----YHSRERDLAAHIMQSFTQFA 546
DMACE             YFTHRTSTSLWGEWMGVLHGDEIEYFFGQPLNNSLQ----YRPVERELGKRMLSAVIEFA 555
TCACE-1           YYKHRTVANPWPSWTGVMHADEINYVFGEPLNPTKS----YTAQEVDLSKRIMRYWANFA 544
BMACE-1           YYKHRSKNNPWPSWTGVMHADEINYVFGEPLNPGKN----YSPEEVEFSKRLMRYWANFA 579
CEACE-1           YFTHRASQQTWPEWMGVLHGYEINFIFGEPLNQKRFN---YTDEERELSNRFMRYWANFA 506
RMACE-1           VFARRSSQNPWPQWTGVIHGEEVPFVFGEPLNDTYC----YSEEDKTLSRRIMRYWANFA 531
HSACE             VFEHRASTLSWPLWMGVPHGYEIEFIFGIPLDPSRN----YTAEEKIFAQRLMRYWANFA 515
CEACE-3           YFDQPSSANPWPKWTGVMHGYEIEYVFGVPLHNTTAG---YTKEEMDVSEKVIDFWTTFA 511
CEACE-4           HFTQSSSANPWPKWTGAMHGYEIEYVFGIPLSYSKN----YKRREQIFSRKIMQFWASFA 512
CEACE-2           YFTYRSSANPWPKWMGVMHGYEIEYAFGQPYWRPHLYDQTHLEDEKRLSSIIMQIWANFA 532
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           YFTHRPSFSTWPGWVGPTHGDEVFFVMGLPFSSPTIA----TDQERELSKLMIETWTTFA 569
PPE014_O05.f      ------------------------------------------------------------
Contig5707        KRKYLKFGPEKLGAYHFSPFANMFGAFLTMSEEELGG--PLDPEDEQFMVDSMNLLVDFV 479
RMACE-2           ELNYVSACVKKQPWFGMTHGDELPLVFGRVFERQGG-----CAGDMDYSRNIMKLWSDFA 512
```

```
TCACE-2           ATGKPVTDDV------NWPLYTKDQPQYFIFNADKNG-----IGKGPRATACAFWNDFLP 588
BMACE-2           LTGKPHEPDE------KWPLYSRSSPHYYTYTAVGPSGPA--GPRGPRASACAFWNDFLN 598
DMACE             KTGNPAQDGE------EWPNFSKEDPVYYIFSTDDKIEK---LARGPLAARCSFWNDYLP 606
TCACE-1           KTGNPSQSPNGVWTPTFWPPHTAFGREFLTLDVNSTA-----TGRGPRLKQCAFWKKYLP 599
BMACE-1           RSGNPSLNPNGEMTKIHWPVHTAFGREYLSLAVNSSS-----VGRGLRVKQCAFWQKHLP 634
CEACE-1           KTGDPNKNEDGSFTQDVWPKYNSVSMEYMNMTVESSYPSMKRIGHGPRRKECAFWKAYLP 566
RMACE-1           KTGNPNLPEDGSPGSTIRWPERTDSLKRHLVLDVNES-----VGWAHRQTYCDFWENVRR 586
HSACE             RTGDPNEPRD--PKAPQWPPYTAGAQQYVSLDLRPLE-----VRRGLRAQACAFWNRFLP 568
CEACE-3           NTG--VPSLRKRAVGTTQKIKWDRYDGTDHTTWMNIKTG-SFRMIQEIKKVECDLWRNAK 568
CEACE-4           KNG--TPKLRVLKN----SEHWPEFNEQNHYRWMQLRSGSNIRPIKPKKQVECQFWRRVK 566
CEACE-2           NTGRTDSFWPQYNKIERKAIELGETTLQGKHRIISDVHGGFCRMIDEAKAFVKQKNANDC 592
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           KTGKVPDVAS-----KPWPEFTAKNQVYMELNPKRYT-----YGRGPNEKNCNFWKSYLK 619
PPE014_O05.f      ------------------------------------------------------------
Contig5707        NSDEVPKFRG-----VTWPNYSEGEGILTINDT---------------------------- 507
RMACE-2           KGRSPVGFQG-----KEWPKFTADSRSFMKLTATGSEVFTFNNEPRCKILKELKLY---- 563

TCACE-2           KLRDNPALKGRGISSILAGHKLKALPANWLCFCVGGDIQGKSVAGYKISRDRSAPESSNI 648
BMACE-2           KLNELERAPCDGAVTGPYSSVAGTALPVTLLTTLAITIAL---------------------- 638
DMACE             KVRSWAGTCDGDSGSASISPRLQLLGIAALIYICAALRTKRVF------------------ 649
TCACE-1           QLQQQTSELLNQPPRQNCTDAASSLRWSRDGAAGLLMVSTVAALLAGPF------------ 648
BMACE-1           QLMAATNKPEPPPKNCTNSVPSLWPSRNTLGFNVIATAALTGTALFKYTI------------ 683
CEACE-1           NLMAAVADVGDPYLVWKQQMDKWQNEYITDWQYHFEQYKRYQTYRQSDSETCGG------ 620
RMACE-1           NRTPPVPSC------------------------------------------------------ 595
HSACE             KLLSATDTLDEAERQWKAEFHRWSSYMVHWKNQFDHYSKQDRCSDL--------------- 614
CEACE-3           DMEYSAYKEELATTSSSTLTQYTIYLILLSAFQLVFNFF----------------------- 607
CEACE-4           DTEYTAYLTQEYSSSSLTTYSYWLLLYIPLFIFQIF------------------------- 602
CEACE-2           RTTRKSASTEDLTSSSSTTYLFSIIVYLSILISYISL------------------------ 629
PPE0138_P16.f     ------------------------------------------------------------
PPE013_C07.f      ------------------------------------------------------------
RMACE-3           P------------------------------------------------------------ 620
PPE014_O05.f      ------------------------------------------------------------
Contig5707        ------------------------------------------------------------
RMACE-2           ------------------------------------------------------------

TCACE-2           SSPVHAHTTIDLILLREYGLHFFRRKRLRLWIYPHAQRRHVSSFRKCLFSKTRRL 703
BMACE-2           ---------------------------------------------------------
DMACE             ---------------------------------------------------------
TCACE-1           ---------------------------------------------------------
BMACE-1           ---------------------------------------------------------
CEACE-1           ---------------------------------------------------------
RMACE-1           ---------------------------------------------------------
```

```
HSACE             ----------------------------------------------------
CEACE-3           ----------------------------------------------------
CEACE-4           ----------------------------------------------------
CEACE-2           ----------------------------------------------------
PPE0138_P16.f     ----------------------------------------------------
PPE013_C07.f      ----------------------------------------------------
RMACE-3           ----------------------------------------------------
PPE014_O05.f      ----------------------------------------------------
Contig5707        ----------------------------------------------------
RMACE-2           ----------------------------------------------------
```

A_c -Sequence alignment of five *Pp* ACE-like sequences, PPE013_C07.r, Contig3716, Contig5966, Contig4735, and Contig6677 (group 2), to reference ACEs from three insect species (*Bombyx mori* (BM ACE1-2: NP_001037380.1, NP_001108113.1, NP_496963.1, NP_496962.1), *Drosophila melanogaster* (DM ACE: NP_476953.1), *Tribolium castaneum* (TC ACE1-2: EFA04156.1, EEZ99262.1), the cattle tick *Rhipicephalus microplus* (RM ACE1-3:

CAA11702.1, CAB93511.1, AAP92139.1), the nematode *Caenorhabditis elegans* (CE ACE1-4: NP_510660.1, NP_491141.1), and *Homo sapiens* (HS ACE: NP_000656.1).

Numbering of the amino acid residues is from the N-terminus of mature proteins. Position of identical amino acid residues is indicated by asterisks (*) and of conservative substitutions by dots.

```
TCACE-1          --MTGAWAACLLVILLPSCIPSPHRGRHHPPEP---------------------------- 31
BMACE-1          --MRVVLAALTALAARTLAGPHEHRARHHAPAPPQPYHGHGEAVRYNPELDTILPRLEDH 58
CEACE-1          ------------------------------------------------------------
RMACE-1          ----MDPEQDMLHENLASCH---------------------------------------- 16
HSACE            ----MRPPQCLLHTPSLASP---------------------------------------- 16
TCACE-2          -------MGSN------------------------------------------------- 4
BMACE-2          MINYGKIVFTK------------------------------------------------- 11
DMACE            ------------------------------------------------------------
CEACE-3          ------------------------------------------------------------
CEACE-4          ------------------------------------------------------------
CEACE-2          ------------------------------------------------------------
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          --MYSRIVVLLLYGYALRSAHARCYIFQPGDPIQDTDEKIYLAQPFN------------- 45
RMACE-2          ------------------------------------------------------------
Contig3716       ------------------------------------------------------------
Contig5966       ------------------------------------------------------------
Contig4735       ------------------------------------------------------------
Contig6677       ------------------------------------------------------------

TCACE-1          -------HAEAYHMSRDPFDPHRDSEEFRRDAPDDKREFTRRDSEDDPLVIQTKKGKVRG 84
BMACE-1          ETSSKRASDAETSSKRTKYEERFYSNHERAAELMADEPVSEKGDEEDPLVIRTRKGKVRG 118
CEACE-1          ----------------------------MRNSLLFFIFLPSTILAVDLIHLHDGSPLFG 31
RMACE-1          --------LTLLALLVCGGVVLRCLSIEPEEDASNRVEDQDAEDPVETVVVETAWGPVKG 68
HSACE            --------LLLLLLWLLGGGVG-------------------AEGREDAELLVTVRGGRLR 49
TCACE-2          ------LVVVVVVVVVVVASLSASARAYSWPSEETTTRPPQARDFHSDPLVVETTSGLVR 58
BMACE-2          ------LLLCVLMSGTFARSWANHHDTTTSTTQTTPTTSPVPKNIHNDPLIVETKSGLIK 65
DMACE            -------MAISCRQSRVLPMSLPLPLTIPLPLVLVLSLHLSGVCGVIDRLVVQTSSGPVR 53
CEACE-3          ------------------MRRRRVLLLLLSTTTFLRVSTAQQDEPKASVVEVQTKLGTVR 42
CEACE-4          ----------------------MKPKLAFFAFFIFITVFIDSVQAVHPVVLETKLGDIK 37
CEACE-2          -----------MRAPVIGRHLTYHVFCQFALVTLFIVRRIEPRSIVRGDHVVHTPLGTIR 49
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          -VTIPLEAGIPTTAQPASTGKDQLSLVVPPDDIQSLSTTTSRPISTEESPIVETNSGPVQ 104
RMACE-2          -----------MYVRVSLVFASAWIIACLGYTETREARGIAVLEDGASPVVQIHAGKLRG 49
Contig3716       ------------------------------------------------------------
Contig5966       ------------------------------------------------------------
Contig4735       ------------------------------------------------------------
Contig6677       ------------------------------------------------------------

TCACE-1          ISLTAATGKKVDAWLGIPYAQKPLGNLRFRHPRPAEKWEGVMN-TTSQPNSCVQIIDTVF 143
BMACE-1          ITLTSATGKKVDAWFGIPYAQKPMGDLRFRHPRPVEDWGDEILNTTTLPHSCVQIVDTVF 178
CEACE-1          EEVLSQTGKPLTRFQGIPFAEPPVGNLRFKKPKPKQPWRIPLN-ATTPPNSCIQSEDTYF 90
RMACE-1          FIAQSPLGKPVRVFYGIPYAKPPTGKRRFDRAESIEEPWTDVLDATVKPNSCFQVLDTLY 128
HSACE            GIRLKTPGGPVSAFLGIPFAEPPMGPRRFLPPE-PKQPWSGVVDATTFQSVCYQYVDTLY 108
TCACE-2          GKAKTVLGREVHVFTGIPFAKPPIEQLRFRKPVPIDPWHGILD-ATKLPNSCYQERYEYF 117
BMACE-2          GYAKTVMGREVHIFTGIPFAKPPLGPLRFRKPVPIEPWHGVLE-ANLMPNSCYQERYEYF 124
DMACE            GRSVTVQGREVHVYTGIPYAKPPVEDLRFRKPVPAEPWHGVLD-ATRLSATCVQERYEYF 112
CEACE-3          GTESDHGNKRVRSFLGVPFAEPPINEHRFKKPTPARPWNGTIS-ANTLSPACFQGRDSYD 101
CEACE-4          GTEFFFLSKKIRTFFGVPFAEPAVEDFRFRKPREKKQWRGLYD-ATKPANACFQTRDNYN 96
```

```
CEACE-2          GVGQTFDGAKVSAFLGVPYAKPPIGSRRFKMAEMIDRWSGELE-ARTLAKTCYLTIDSAF 108
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          GRRVYAANRTLYQFVGIPFAEPPVGPLRFRNPVPVKPWSSVYQ------ATKKPFPCLQT 158
RMACE-2          AKRVVLGEKFAYAFTGVPYAKPPVGELRYQKPESAQPWVEEVKDATVTPPSCMQGNVFSP 109
Contig3716       ------------------------------------------------------------
Contig5966       ------------------------------------------------------------
Contig4735       ------------------------------------------------------------
Contig6677       ------------------------------------------------------------

TCACE-1          GDFPGATMWNPNTPLNEDCLYVNVVVP--KPRPT--------------------------- 175
BMACE-1          GDFPGAMMWNPNTDMQEDCLYINIVTP--RPRPK--------------------------- 210
CEACE-1          GDFYGSTMWNANTKLSEDCLYLNVYVPGKVDPNK--------------------------- 124
RMACE-1          GNFSGSTMWNANTEMSEDCLKLNVWAPGPPTSSGG-------------------------- 163
HSACE            PGFEGTEMWNPNRELSEDCLYLNVWTPYPRPTS---------------------------- 141
TCACE-2          PGFEGEEMWNPNTNISEDCLYLNIWVPQRLRIRHHGEKLPQDRPK---------------- 162
BMACE-2          PGFEGEEMWNPNTNISEDCLYLNIWVPQHLRVRHHQDKPLAERPK---------------- 169
DMACE            PGFSGEEIWNPNTNVSEDCLYINVWAPAKARLRHGRGANGGEHPNGKQADTDHLIHNGNP 172
CEACE-3          PTFWGSEMWNANTPVSEDCLYVNIWAPADAYN----------------------------- 133
CEACE-4          TSFWGSEMWNANTQISEDCLYLNIWAPADAYN----------------------------- 128
CEACE-2          PQFPGAEMWNPPGAISEDCLNMNIWVPEDHD------------------------------ 139
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          DFYINSNVTIPTANSSEDCLYLNVWTPSRECVLGKFS------------------------ 195
RMACE-2          ---RNLLWLPYDHQKSEDCLYLNVWTPRLNTSAG--------------------------- 140
Contig3716       ------------------------------------------------------------
Contig5966       ------------------------------------------------------------
Contig4735       ------------------------------------------------------------
Contig6677       ------------------------------------------------------------

TCACE-1          -----SAAVMVWVFGGGFYSGTNTLEVYDHNILVSEENIILVSMQYRVASLGFLYFG-TP 229
BMACE-1          -----NAAVMLWVFGGGFYSGTATLDVYDPKILVSEEKVVYVSMQYRVASLGFLFFD-TA 264
CEACE-1          -----KLAVMVWVYGGGFWSGTATLDVYDGRILTVEENVILVAMNYRVSIFGFLYMN-RP 178
RMACE-1          ----RPLAVLVWIYGGGFYSGTSTLDVYDARTLVSEENVVVVSMNYRVASLGFLSFG-NE 218
HSACE            -----PTPVLVWIYGGGFYSGASSLDVYDGRFLVQAERTVLVSMNYRVGAFGFLALPGSR 196
TCACE-2          ------VPVLVWIYGGGYMSGTSTLDIYDADIIAATSDVIVASMQYRVGAFGFLYLSKYF 216
BMACE-2          ------VPILVWIYGGGYMSGTATLDLYKADIMASTSDVIVASMQYRVGAFGFLYLNKYF 223
DMACE            QNTTNGLPILIWIYGGGFMTGSATLDIYNADIMAAVGNVIVASFQYRVGAFGFLHLAPEM 232
CEACE-3          ------LTVLVWLFGGGFWYGSPSLLLYDGKELATRGNVIVVNINYRVGPFGYLFLD-HE 186
CEACE-4          ------LTVMVWFFGGGFYSGSPSLSIYDGKALTSTQNVIVVNINYRLGPFGFLYLG-HP 181
CEACE-2          ------GSVMVWIYGGGFFSGTPSLDLYSGSVFAAKEHTIVVNVNYRLGPFGFLYFGDDS 193
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          ---CVPKTVIVYIYGGTFSFGSSGWDWYDGKEFVARGDVVMVSMNYRVGPMGFFHSGTTH 252
RMACE-2          ------LPVMAWIHGGGFQEGSAAIPLDDGTYLAAFGNVVVVTIAYRLQSFGFLYDETSA 194
Contig3716       ------------------------------------------------------------
Contig5966       ------------------------------------------------------------
Contig4735       ------------------------------------------------------------
Contig6677       ------------------------------------------------------------

TCACE-1          D------VPGNAGLFDQMMALQWVRDNIAAFGGNPNNITLFGESAGAVSVSLHLLSPLSR 283
BMACE-1          D------VPGNAGLFDQLMALQWVKDNIGYFGGNPHNITLFGESAGAVSVSLHLLSPLSR 318
CEACE-1          E------APGNMGMWDQLLAMKWVHKNIDLFGGDLSRITLFGESAGAASVSIHMLSPKSA 232
RMACE-1          T------LPGNAGLYDQYMALKWVQENVAAFGGDPDRVTLFGESAGAASAGLHVLSPLSE 272
HSACE            E------APGNVGLLDQRLALQWVQENVAAFGGDPTSVTLFGESAGAASVGMHLLSPPSR 250
TCACE-2          P-RGSEEAPGNMGMWDQALAIRWIKENAAAFGGDPDLITLFGESAGGGSVSILLLSPVTK 275
BMACE-2          S-PGSEEAPGNMGLWDQQLAIRWIKENARAFGGDPELITLFGESAGGGSVSLHMLSPEMK 282
DMACE            PSEFAEEAPGNVGLWDQALAIRWLKDNAHAFGGNPEWMTLFGESAGSSSVNAQLMSPVTR 292
CEACE-3          D------VPGNMGMLDQQLALYWIRDHIFSFGGNPARISLVGESAGAASIVAHLIAPASK 240
CEACE-4          D------APGNMGLLDQQLALHWVRQNIVSFGGNPDKVAVFGQSAGAASIVAHLIAPGSR 235
CEACE-2          P------IQGNMGLMDQQLALRWVHENIGAFGGDRSRVTLFGESAGSASTTAHLFAPNSH 247
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          S-------SGNAGLHDQLLAMKWVKQNIRNFGGDPDDVTLVGQSAGAISIGLHLVSPLSK 305
RMACE-2          P--------GNMGLHDQQLALKWIQENIAAFGGNPGEVTLFGWSAGGISTGFHLISPGSQ 246
Contig3716       ------------------------------------------------------------
Contig5966       ---------GHQGHLDQVLVLNWVSQNIRYFGGDPNRVTLFGVSAGSFSISWHLLTGFSA 51
Contig4735       ------------------------------------KRAVMQSGSPFSFVPRNTKDQGE 23
Contig6677       --------------------------------RNIPASSLLPKSNSDARGNLSFIPTTDG 28
```

```
TCACE-1          NLFSQAIMESGSATAPWAIISREESILRGLRLAEAVGCPHE------RHELSAVIDCLKK 337
BMACE-1          NLFSQAIMQSGAATAPWAIISREESILRGIRLAEAVHCPHS------RSDLAPMIECLRK 372
CEACE-1          PYFHRAIIQSGSATSPWAIEPRDVALARAVILYNAMKCGNMSL-I--NPDYDRILDCFQR 289
RMACE-1          SLFHRVILQSGSPAVPWGFQDRDKARQSAKKLATALRAP--------DSLDQETLDSLRC 324
HSACE            GLFHRAVLQSGAPNGPWATVGMGEARRRATQLAHLVGCPPGGT-GGNDTELVACLRTRPA 309
TCACE-2          GLARRGILQSGTMNAPWSYMSGERAQQIGKVLVEDCGCNVS----LLETRPHEVIDCMRA 331
BMACE-2          GLFKRGILQSGTLNAPWSWMTGERAQDIGKVLIDDCNCNSS----LLAKDPSLVMDCMRG 338
DMACE            GLVKRGMMQSGTMNAPWSHMTSEKAVEIGKALINDCNCNAS----MLKTNPAHVMSCMRS 348
CEACE-3          GLFQNGILQSGSLDNKWSMDSPKRAKQKSTALADLVGCNQT-----KITDQTACLRNTPA 295
CEACE-4          GLFKNAILQSGSLENTWAINSPFRAKQKSEKLLELVGCNKT-----TVDTSMACLRLVSP 290
CEACE-2          KYFRNIIAKSGSIINSWASATPPTMLDLSFRLAKKVNCSSP-----DMNAIVKCLRSVPA 302
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          GLFKRIIMESGSPYFRIADNTREGPHKVEKLARALQCARND----MTIESHMAEMVECLR 361
RMACE-2          TLFKRAIVQSAAVTKKGRARDKTEMLEYSQKFAANFGCYGG----DSAANASQDIADCMR 302
Contig3716       ------------------------------------------------------------
Contig5966       GLFHAAVIDAGVLTHTHAETRRDHVSRAQKMVRTSICKSLKEFGDRATKEQRRRILNCLM 111
Contig4735       TLFRSLASYTDCMSVHFNSTLRYDDVLQR---------------------------MK 54
Contig6677       AFLPRDVEEYVAKN------------------------------------------- 42

TCACE-1          KDPIDLVNNEWG-TLGICEFPFVPVIDGAFLDES--PTRALANKNFKKTNILMGSNTEEG 394
BMACE-1          KNADELVNNEWG-TLGICEFPFVPIIDGSFLDEM--PVRSLAHQNFKKTNILMGSNTEEG 429
CEACE-1          ADADALRENEWAPVREFGDFPWVPVVDGDFLLEN--AQTSLKQGNFKKTQLLAGSNRDES 347
RMACE-1          ERPEDIVNNETN-SGGVVDFPFVPVADGVFLPDT--PQALTDKGSFARNISVMLGSNANE 381
HSACE            QVLVNHEWHVLP-QESVFRFSFVPVVDGDFLSDT--PEALINAGDFHG-LQVLVGVVKDE 365
TCACE-2          VEAKTISLQQWNSYSGILGFPSTPTVDGVFMPKH--PMDMLAEGDYEDMEILVGSNQDEG 389
BMACE-2          VDAKTISVQQWNSYTGILGFPSAPTVDGIFLPKD--PDTMMKEGNFHNSEVLLGSNQDEG 396
DMACE            VDAKTISVQQWNSYSGILSFPSAPTIDGAFLPAD--PMTLMKTADLKDYDILMGNVRDEG 406
CEACE-3          QLLIDNIWNVG---LNFLEFPFAIVSKDQNFFKHLDGFIALREGTYSTDVNLMFGINHDE 352
CEACE-4          EQLSLSTWNIS---LTYLEFPFVIVSRDKHFFGHLDAHAALREGDFNRDVNLMIGMNKDE 347
CEACE-2          HLVQAEADNISGDIGPPMTFAYVPVSSDANFFQG-DVFQKLANKQFKKDVNIIFGSVKDE 361
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          KIDGKELLIMSNTIFGVHALTFFPVFGDDIIPDD--PYLMMEQKKFHKADLLIGNNLDEG 419
RMACE-2          TVNASLIVAVEATFVGSGSGKFEPIYGDEFLPIEP------RMADFPGDKDVMIGQTANE 356
Contig3716       ------------------------------------------------------------
Contig5966       KIDANELVELQQRYASSQTYAFRPTFNNQEYLPR-SPTCMTNEEAFSVNVPIIIGDATNE 170
Contig4735       KQPFENIIAASEKFNGKGVNSFFPVMGEEFIPMN--PKDSLLLKRFSNVDLLVTTTKSEG 112
Contig6677       ----------PAELRKVRTIVGYSLDEGSMFVRLLDPKFDFSTARPRDEILDYCGKISET 92

TCACE-1          YYFIIYY--------LTELFRKEENVYVNRQE--------FLRAVTELNPYFNAISRQAI 438
BMACE-1          YYFILYY--------LTELFPKEENVGISREQ--------FLQAVRELNPYVNDVARQAI 473
CEACE-1          IYFLTYQ---------LPDIFPVADFFTKTDFIKDRQLWIKGVKDLLPRQILKCQLTLAAV 399
RMACE-1          GSWFLQY---------FFGFPVTDETPEVTKEN--------FTAVLEALDPSLEQTPIAEI 425
HSACE            GSYFLVY---------GAPGFSKDNESLISRAE--------FLAGVRVGVPQVSDLAAEAV 409
TCACE-2          TYFLLYD---------FIDFFEKDGPSFLQRDK--------YHDIIDTIFKNMSRLERDAI 433
BMACE-2          TYFLLYD---------FLDYFEKDGPSFLQREK--------FLEIVDTIFKDFSKIKREAI 440
DMACE            TYFLLYD---------FIDYFDKDDATALPRDK--------YLEIMNNIFGKATQAEREAI 450
CEACE-3          GNFWN-----------IYNLAKFFDKQSVKPGLDRDEFHECVDTAFAVQPELVRTAAKYV 401
CEACE-4          GNYWN-----------IYQLPQFFDKADPPE-LNRTEFDFLIDRTFSIQPDIIRSAAKYI 395
CEACE-2          GTYWLPYYMSLPKYGFAFNHTISAEDPHNRALITRDHYEESMRAFMPYFAGSKLVLNAFM 421
PPE013_C07.r     ------------------------------------------------------------
RMACE-3          SYFVFYL---------FGRALDLEQAHKITKYE--------VDLYVSYCLQMLLRKNVSPI 463
RMACE-2          GSNILYT--------TFRDTFSEALPPRKINKA--------EMIHFLGSLYKLSLSDIEKL 401
Contig3716       ------------------------------------------------------ANRGVK 6
Contig5966       GLFLLPR----------KVHDPLPSFSSFDEVLEWDIDILGGHHSAQAPNYGPNNSTIAE 220
Contig4735       AYFLQHF---------LSPFTNVADADKINPGE-----IVFYLRVFLSALLGGKPTASLNE 159
Contig6677       FDFPFN-------------------------------------------------ASKRET 104

TCACE-1          VFEYTNWLNPDDPVSNRDSLDKMVGDYHFTCNVNEFAHRYAETGNTVYM----------- 487
BMACE-1          IYEYTDWLNPEDPVKNRNALDKMVGDYHFTCGVNEFAHRYAETGNNVYT----------- 522
CEACE-1          LHEYEPQDLPVTPRDWINAMDKMLGDYHFTCSVNEMALAHTKHGGDTYY----------- 448
RMACE-1          MKTYTAGEIPSTAADILKALDSIVGDYHFTCPVVRWADTFARAGIPVYQ----------- 474
HSACE            VLHYTDWLHPEDPARLREALSDVVGDHNVVCPVAQLAGRLAAQGARVYA----------- 458
TCACE-2          VFQYTDWEHVNDGYLNQKMVGDVVGDYFFICPTNDFAELAAERGMKVYY----------- 482
BMACE-2          VFQYTDWEEITDGYLNQKMIADVVGDYFFVCPTNYFAEILADAGVGVYY----------- 489
DMACE            IFQYTSWEGNPG-YQNQQQIGRAVGDHFFTCPTNEYAQALAERGASVHY----------- 498
CEACE-3          YSDPKCTDPKKKTDFYTEQVNQMVGDYFFTCDSIWFAHNYPKMAGNQSN--------VFV 453
CEACE-4          YSDPNCTDHGRKTRFYAGQMNQIVGDYFFSCDSLWLADQFRSIPRVKSSSPQRKPGKVFV 455
```

```
CEACE-2          NSYEHVSTSNVPEERYRDGVARFLGDLFFTCSLIDFADLISDNIFGNVY----------M 471
PPE013_C07.r     ----------------RQAAEMVGDLAMVCPTKYFAEEAAAQNLSVHY----------- 32
RMACE-3          RHYYLSHIGERENIKALQKAAEAVGDFAIICPTKYFAESFASMNNKVHY----------- 512
RMACE-2          QKEYMGEIGDYDYDALRQALAETKGDTHVKCGAINTACKLANATANAQSGKE-----VHF 456
Contig3716       VYRYLWEHKPSTSYWPKWSGATHCDDVPFTMGSQFDIGNKAEKSKQASEG---------- 56
Contig5966       IYNNETNDREPEAAALFNAATQIIGDGLFVCPVMNFADRYSSVQANVYF----------- 269
Contig4735       LAQNTEPETREEKVQFLKNISAVIGDYPYLCATTDFGVEYANPRHNVYH----------- 208
Contig6677       REKIGQLYVDDNSGNAFKAVSSFIADGWFKCPINTFIRSYSRHNDKVFA----------- 153
                                          *

TCACE-1          YYYKHRTVANPWPSWTGVMHADEINYVFGEPLNPTKS----YTAQEVDLSKRIMRYWANF 543
BMACE-1          YYYKHRSKNNPWPSWTGVMHADEINYVFGEPLNPGKN----YSPEEVEFSKRLMRYWANF 578
CEACE-1          YYFTHRASQQTWPEWMGVLHGYEINFIFGEPLNQKRFN---YTDEERELSNRFMRYWANF 505
RMACE-1          YVFARRSSQNPWPQWTGVIHGEEVPFVFGEPLNDTYC----YSEEDKTLSRRIMRYWANF 530
HSACE            YVFEHRASTLSWPLWMGVPHGYEIEFIFGIPLDPSRN----YTAEEKIFAQRLMRYWANF 514
TCACE-2          YFFTHRTSTSLWGEWMGVMHGDEIEYVFGHPLNMSLQ----FNSRERELSLKIMQAFARF 538
BMACE-2          YYFTHRTSTSLWGEWMGVMHGDEMEYVFGHPLNMSLQ----YHSRERDLAAHIMQSFTQF 545
DMACE            YYFTHRTSTSLWGEWMGVLHGDEIEYFFGQPLNNSLQ----YRPVERELGKRMLSAVIEF 554
CEACE-3          YYFDQPSSANPWPKWTGVMHGYEIEYVFGVPLHNTTAG---YTKEEMDVSEKVIDFWTTF 510
CEACE-4          YHFTQSSSANPWPKWTGAMHGYEIEYVFGIPLSYSKN----YKRREQIFSRKIMQFWASF 511
CEACE-2          YYFTYRSSANPWPKWMGVMHGYEIEYAFGQPYWRPHLYDQTHLEDEKRLSSIIMQIWANF 531
PPE013_C07.r     YEFDFRSSFSTWPDVGTTHGEEIPFVFGHPMSGLEPN----ATAQDKDMSAEIIGIWTDF 88
RMACE-3          YYFTHRPSFSTWPGWVGPTHGDEVFFVMGLPFSSPTI----ATDQERELSKLMIETWTTF 568
RMACE-2          YELNYVSACVKKQPWFGMTHGDELPLVFGRVFERQGG-----CAGDMDYSRNIMKLWSDF 511
Contig3716       ---------------------------LVRYMQTPIT------EAEKQLIKDSLKMIGDF 83
Contig5966       YRWERVRVDETYPHERADGAYHGLMFYTGVGSQYLYLQG-QMADADKSYIENTIKMIADF 328
Contig4735       MQYDHRPWFLLHPTWFPSTHGDDIMFWLGSVYKLKER-----TRADERVADELMAILTAF 263
Contig6677       YQFERRLNRPYFKLFDPKVLGAFHYSPYLHFGGAILLDGGTVNEGDKQFSLDAMSMISKF 213
                                                           :      :     *

TCACE-1          AKTGNPSQSPNGVWT-PTFWPPHTAFGREFLTLDVNSTA-----TGRGPRLKQCAFWKKY 597
BMACE-1          ARSGNPSLNPNGEMT-KIHWPVHTAFGREYLSLAVNSSS-----VGRGLRVKQCAFWQKH 632
CEACE-1          AKTGDPNKNEDGSFT-QDVWPKYNSVSMEYMNMTVESSYPSMKRIGHGPRRKECAFWKAY 564
RMACE-1          AKTGNPNLPEDGSPG-STIRWPERTDSLKRHLVLDVNES-----VGWAHRQTYCDFWENV 584
HSACE            ARTGDPNEPRD--PK-APQWPPYTAGAQQYVSLDLRPLE-----VRRGLRAQACAFWNRF 566
TCACE-2          AATGKPVTDDV-------NWPLYTKDQPQYFIFNADKNG-----IGKGPRATACAFWNDF 586
BMACE-2          ALTGKPHEPDE-------KWPLYSRSSPHYYTYTAVGPSGPA--GPRGPRASACAFWNDF 596
DMACE            AKTGNPAQDGE-------EWPNFSKEDPVYYIFSTDDKIEK---LARGPLAARCSFWNDY 604
CEACE-3          ANTG--VPSLRKRAVGTTQKIKWDRYDGTDHTTWMNIKTG-SFRMIQEIKKVECDLWRNA 567
CEACE-4          AKNG--TPKLRVLKN----SEHWPEFNEQNHYRWMQLRSGSNIRPIKPKKQVECQFWRRV 565
CEACE-2          ANTGRTDSFWPQYNKIERKAIELGETTLQGKHRIISDVHGGFCRMIDEAKAFVKQKNAND 591
PPE013_C07.r     AKTGLTPTKVRG-----TPWPAYTEKSQTYLRFGQESTG-----IGRGPNERMCNSWRKY 138
RMACE-3          AKTGKVPDVAS------KPWPEFTAKNQVYMELNPKRYT-----YGRGPNEKNCNFWKSY 617
RMACE-2          AKGRSPVGFQGK------EWPKFTADSRSFMKLTATGSE-----VFTFNNEPRCKILKEL 560
Contig3716       VKTGNPSRPDGS------QWPLYTAEKREVVQIGTTSFS------DISLQNSRCPIWTDF 131
Contig5966       ASNPTGS----------------------------------------------------- 335
Contig4735       AGKGTPQTRGKLTWPRINQGGYYMKVGSEVTELLRTPAAN-------------CRQWSRH 310
Contig6677       SKSDGPLMFRDVG------WPPFSESGEVFIFKETPTVAKEL------------------ 249


TCACE-1          LPQLQQQTSELLNQPPRQNCTDAASSLRWSRDGAAGLLMVSTVAALLAGPF--------- 648
BMACE-1          LPQLMAATNKPEPPKNCTNSVPSLWPSRNTLGFNVIATAALTGTALFKYTI--------- 683
CEACE-1          LPNLMAAVADVGDPYLVWKQQMDKWQNEYITDWQYHFEQYKRYQTYRQSDSETCGG---- 620
RMACE-1          RRNRTPPVPSC------------------------------------------------- 595
HSACE            LPKLLSATDTLDEAERQWKAEFHRWSSYMVHWKNQFDHYSKQDRCSDL------------ 614
TCACE-2          LPKLRDNPALKGRGISSILAGHKLKALPANWLCFCVGGDIQGKSVAGYKISRDRSAPESS 646
BMACE-2          LNKLNELERAPCDGAVTGPYSSVAGTALPVTLLTTLAITIAL------------------ 638
DMACE            LPKVRSWAGTCDGDSGSASISPRLQLLGIAALIYICAALRTKRVF--------------- 649
CEACE-3          KDMEYSAYKEELATTSSSTLTQYTIYLILLSAFQLVFNFF-------------------- 607
CEACE-4          KDTEYTAYLTQEYSSSSLTTYSYWLLLYIPLFIFQIF------------------------ 602
CEACE-2          CRTTRKSASTEDLTSSSSTTYLFSIIVYLSILISYISL---------------------- 629
PPE013_C07.r     L---------------------------------------------------------- 139
RMACE-3          LKP-------------------------------------------------------- 620
RMACE-2          KLY-------------------------------------------------------- 563
Contig3716       I---------------------------------------------------------- 132
Contig5966       ------------------------------------------------------------
Contig4735       YPQ-------------------------------------------------------- 313
Contig6677       ------------------------------------------------------------
```

```
TCACE-1           ------------------------------------------------------------
BMACE-1           ------------------------------------------------------------
CEACE-1           ------------------------------------------------------------
RMACE-1           ------------------------------------------------------------
HSACE             ------------------------------------------------------------
TCACE-2           NISSPVHAHTTIDLILLREYGLHFFRRKRLRLWIYPHAQRRHVSSFRKCLFSKTRRL 703
BMACE-2           ------------------------------------------------------------
DMACE             ------------------------------------------------------------
CEACE-3           ------------------------------------------------------------
CEACE-4           ------------------------------------------------------------
CEACE-2           ------------------------------------------------------------
PPE013_C07.r      ------------------------------------------------------------
RMACE-3           ------------------------------------------------------------
RMACE-2           ------------------------------------------------------------
Contig3716        ------------------------------------------------------------
Contig5966        ------------------------------------------------------------
Contig4735        ------------------------------------------------------------
Contig6677        ------------------------------------------------------------
```

A_d -Aligned 229 amino acid positions from *Pp* ACEs group 1 (21% of 1058 a.a.) after low stringent trimming with Gblocks (allow smaller final blocks, gap positions within the final blocks, and less strict flanking positions) used for phylogenetic analysis.

See reference sequence information in Appendix A_f. Boxshade v.3.21 shading with black background for concensus and gray background for similar residues in the fraction of 0.5 aligned sequences.

```
PPE014_O05_f     1 VQGPLG----VESFLGIPYAKPPVGNRRFALPTFGTVNLSATEYSR-QCPQEDCLYLNVFR-AVIAIIHGGGYIIGVSPA
Contig5707       1 VEVSTG----VESFLGMPYAKPPVGELRFALPTVAPFDLNATAFSA-SCPQEDCLYLNIFR-PVLFVIHGGAFRVGYNGL
PPE0138_P16_f    1 ---SSG----VYQFLGIPFAEPPLNELRFQKPPKKPWVLSVNKWGS-ACMQEDCLILNVFT-PVMVWIHGGDFNFGYDGT
PPE013_C07_f     1 VELVTG----VHKFLGIPFAEPPIGDLRFKHPAVKSWKIVKADTKPFPCLQEDCLYLNVWTPSVMVFIYGGTYTFGYDAE
DM ACE           1 VQTSSGPVRGVHVYTGIPYAKPPVEDLRFRKPPAEPWGVLDATRLSATCVQEDCLYINVWAPPILIWIYGGGFMTGYNAD
RM ACE-1         1 VETAWGPVKGVRVFYGIPYAKPPTGKRRFDRAIEEPWVLDATVK-PNSCFQEDCLKLNVWAPAVLVWIYGGGFYSGYDAR
RM ACE-2         1 VQIHAGKLRGAYAFTGVPYAKPPVGELRYQKPSAQPWEVKDATVTPPSCMQEDCLYLNVWTPPVMAWIHGGGFQEGDDGT
RM ACE-3         1 VETNSGPVQGLYQFVGIPFAEPPVGPLRFRNPPVKPWSVYQATKKPFPCLQ-DCLYLNVWTPTVIVYIYGGTFSFGYDGK
```

```
IS (2)           1 VTGQVMGLPVTTAFLGIPFADNTGGTRRFKKPIYHGWGVFNATYTRKPCCQ-DCLHLNIWVPPVMFWVYGGGFVFGYDGR
IS (4)           1 ILTSTGQVQGVRAFLGIPYAEPPTGDLRFQKPPKRQWGILNATSLPALCSQEDCLYLNVFTPPVIVSIHGGGFSSGSDSS
IS (5)           1 VKTKSGLVRGIDAFLGVPFAEPPVGALRFKPPPAKPWREFRALFLPPSCIQ-DCLYLNVWAPAVMLFFHGGAFFFGYDGS
IS (6)           1 VPTSLGLVAGAKAFLGIPFAEPPVGPKRFKRPRKKPWGILNVIQLPRPCFQEDCLYLNIWVPPVMVWIHGGSYRIGYDGT
IS (7)           1 VHTTKGPIRGVDVFYGIPYAQPPVGRYRFRHPPTDPWGVLDATVKPSSCYQEDCLTLNVWVPAVLVWIYGGGFYSGYDGR
IS (8)           1 TYTRHGLVRGVDAYYGIPYAEPPLAQLRFRKPPPSSWGIYYATKKRPPCSQ-DCLHVNVWTSAVMVFLHGGGFQFGYDGR
IS (9)           1 VATSSGLISGVDAFYGIPYAKPPVGDLRFRKPPADPWGTYKATTKPMACNQ-DCLYMNVWRPPVVVFIYGGGFQWGYDGA
IS (10)          1 NGILSELNEGVDAFYGIPYAKPPVGDLRFRKPPAEPWGTYEATTKPTACNQ-DCLYVNVWRPPVVVFIHGGGFQWGYDAA
BM ACE-1         1 VETKSGLIKGVHIFTGIPFAKPPLGPLRFRKPPIEPWGVLEANLMPNSCYQEDCLYLNIWVPPILVWIYGGGYMSGYKAD
BM ACE-2         1 IRTRKGKVRGVDAWFGIPYAQKPMGDLRFRHPPVEDWEILNTTTLPHSCVQEDCLYINIVTPAVMLWVFGGGFYSGYDPK
TC ACE-1         1 IQTKKGKVRGVDAWLGIPYAQKPLGNLRFRHPPAEKWGVMNTTSQPNSCVQEDCLYVNVVVPAVMVWVFGGGFYSGYDHN
TC ACE-2         1 VETTSGLVRGVHVFTGIPFAKPPIEQLRFRKPPIDPWGILDATKLPNSCYQEDCLYLNIWVPPVLVWIYGGGYMSGYDAD
CE ACE-1         1 ----LIHLHDLTRFQGIPFAEPPVGNLRFKKPPKQPWIPLNATTPPNSCIQEDCLYLNVYVPAVMVWVYGGGFWSGYDGR
CE ACE-2         1 VQTKLGTVRGVRSFLGVPFAEPPINEHRFKKPPARPWGTISANTLSPACFQEDCLYVNIWAP-VLVWLFGGGFWYGYDGK
CE ACE-3         1 LETKLGDIKGIRTFFGVPFAEPAVEDFRFRKPEKKQWGLYDATKPANACFQEDCLYLNIWAP-VMVWFFGGGFYSGYDGK
CE ACE-4         1 VHTPLGTIRGVSAFLGVPYAKPPIGSRRFKMAMIDRWGELEARTLAKTCYLEDCLNMNIWVP-VMVWIYGGGFFSGYSGS
HS ACE           1 VTVRGGRLRGVSAFLGIPFAEPPMGPRRFLPPPKQPWGVVDATTFQSVCYQEDCLYLNVWTPPVLVWIYGGGFYSGYDGR
BM JHE           1 AQTESGWVCGYASFRGVPYAKQPVGELRFKELPAEPWDYLDATEEGPVCYQ--CIYANIHVPPILVFIHGGGFAFGYGPE
TC JHE           1 VYTKYGSVIGYMSFKGIPFAKPPVGDLRFKAPPPEPWFSINGTKDAPFCIQ-DCLYLNVYVPPVMVFIHWGGFFAGYNGP
AM JHE           1 VKTPLGAIKGYEAYEGIPYALPPVGKFRFKAP-IPAWGELSATKFGFPCLQEDCLYLNVYVPPVIFWIHGGAFQFGGAKY
DM esterase      1 VSTTYGPIKGYFSFERIPFAKPPVGELRYKAP-PEVWEVRSCTSQGPKPLQEDCLYLNVYTKPVMVWIYGGGFQFGSPDY
HSal esterase    1 IKTPLGGLKGYEAYEGIPFALPPVGELRFKPP-VTPWGELSATKLSSACIQEDCLYLNVYVPPVLFWIHGGAFQFGGAKY
NV COesterase    1 VNTRSGRLRGYCTFKGIPYAKPPVGELRFKDP-AEPWGVRDATEYGTICGQDDCLYLNVYVRPVMVWIHGGAFMFGGPDY
HA COesterase    1 VRVSDGLLEGYFSFRGIPYAQPPLGDLRFKAP-PTPWNVRSAKEFGNNCLQEDCLYLNVYTPPVMVWIHGGGFVSGGPKF
SL COesterase    1 VRVNEGLLEGFYSFKGIPYAEPPVGDLRFKAP-PKAWGVRSAKEFGPKCYQEDCLYLNVYTPPVMFWIHGGGFFCGGPEF

PPE014_O05_f    75 SLVAFGDVIVVSFNYRLGIFGFAPGNLGLYDQRLALQWIQDHIGGFGGNPEKVTVIGVSAGSMSLSAQIITPEKNLFQSA
Contig5707      75 PIAAIGDVIVVTINYRLGVLGFAPGNLGLFDQLLALEWVHENIAAFGGDPDRVTILGVSAGSMSVSALVNTPGRNLFKQA
PPE0138_P16_f   72 PITGLKDVIIVSINYRLSSLGFLPGNMGLWDQQLALKWVKDNIEHFGGDPNRVTIFGESAGSMSVSAHIVSP-KGLFKNA
PPE013_C07_f    77 QLALRGDVVVVTFNYRVGPLGFL------------------------------------------------------------
DM ACE          81 IMAAVGNVIVASFQYRVGAFGFLPGNVGLWDQALAIRWLKDNAHAFGGNPEWMTLFGESAGSSSVNAQLMSPTRGLVKRG
RM ACE-1        80 TLVSEENVVVVSMNYRVASLGFLPGNAGLYDQYMALKWVQENVAAFGGDPDRVTLFGESAGAASAGLHVLSPSESLFHRV
RM ACE-2        81 YLAAFGNVVVVTIAYRLQSFGFL-GNMGLHDQQLALKWIQENIAAFGGNPGEVTLFGWSAGGISTGFHLISPSQTLFKRA
RM ACE-3        80 EFVARGDVVMVSMNYRVGPMGFFSGNAGLHDQLLAMKWVKQNIRNFGGDPDDVTLVGQSAGAISIGLHLVSPSKGLFKRI
IS (2)          80 YIAGFGNLIVVVPNYRVGSFGFLPGNMGMHDVIAAYRWVRDHIGSFGGDRENIILAGQSAGSIISGLLMISPPLTLFSKA
IS (4)          81 ELSVRGDLVTVAIAYRLGAFGFLPGNMGLYDQVMALRWVKENIQSFGGDPDKITLMGPSAGSVAVGVHILSPSRGLFHRA
IS (5)          80 NLAALGDVVVVTANYRLGPFGFMPGNQGLYDQNLAMRWVRDNIRYFGGDEEQVTLFGQSAGAISIGYHLASPSKGLFKRV
IS (6)          81 ILSDYGRVVVVSFNYRLGALGFLPGNMGLWDQYAALRWVNENIASFGGDPSRVTLFGESVGGASSGMLAQSPCRGLIRRI
IS (7)          81 SLVAEERLVLVSMNYRVASLGFLPGNAGLFDQLMALQWIQDNIAAFGGNPRNVTLFGESAGAVSVSMHLLSPSRDLFSQA
IS (8)          80 YLALLGDVVVVVPNYRLNVFGFLPGNMGMYDQILALKWVQKNIAMFGGDPNRVTLFGQSAGSVSTGYHLLSPSRGLFRRV
IS (9)          80 NFVALSDVIFVSFNHRLSMMGFLPGNLGFWDQLLVLKWVQQNIGRFGGNPQDVTLLGHSAGAVSAGLHAVSQSKGLFHRL
IS (10)         80 NFVALSDVIFVSFNHRLSMMGFLPGNLGFWDQLLALKWVRRNIARFGGNPKDVTLAGHSAGAVSAGIHAVSQSKGLFHRL
BM ACE-1        81 IMASTSDVIVASMQYRVGAFGFLPGNMGLWDQQLAIRWIKENARAFGGDPELITLFGESAGGGSVSLHMLSPMKGLFKRG
BM ACE-2        81 ILVSEEKVVYVSMQYRVASLGFLPGNAGLFDQLMALQWVKDNIGYFGGNPHNITLFGESAGAVSVSLHLLSPSRNLFSQA
TC ACE-1        81 ILVSEENIILVSMQYRVASLGFLPGNAGLFDQMMALQWVRDNIAAFGGNPNNITLFGESAGAVSVSLHLLSPSRNLFSQA
TC ACE-2        81 IIAATSDVIVASMQYRVGAFGFLPGNMGMWDQALAIRWIKENAAAFGGDPDLITLFGESAGGGSVSILLLSPTKGLARRG
CE ACE-1        77 ILTVEENVILVAMNYRVSIFGFLPGNMGMWDQLLAMKWVHKNIDLFGGDLSRITLFGESAGAASVSIHMLSPSAPYFHRA
CE ACE-2        80 ELATRGNVIVVNINYRVGPFGYLPGNMGMLDQQLALYWIRDHIFSFGGNPARISLVGESAGAASIVAHLIAPSKGLFQNG
CE ACE-3        80 ALTSTQNVIVVNINYRLGPFGFLPGNMGLLDQQLALHWVRQNIVSFGGNPDKVAVFGQSAGAASIVAHLIAPSRGLFKNA
CE ACE-4        80 VFAAKEHTIVVNVNYRLGPFGFLQGNMGLMDQQLALRWVHENIGAFGGDRSRVTLFGESAGSASTTAHLFAPSHKYFRNI
HS ACE          81 FLVQAERTVLVSMNYRVGAFGFLPGNVGLLDQRLALQWVQENVAAFGGDPTSVTLFGESAGAASVGMHLLSPSRGLFHRA
BM JHE          79 YLVTRN-VVVITFNYRLNFFGFFPGNNGLRDMVTLLRWVKRNARAFGGNPDNVTLAGQSAGAAAAHLLTLSKTEGLVSRA
TC JHE          80 EYIMDKDVILVTFNYRLGVFGFLPGNFGLKDQVMALKFVHENIECFGGDNNRVTIFGQSAGSGSVNLHLISPSRGLFQQA
```

```
AM JHE          80 LMDS--DVIFVTINYRLGILGFLPGNMGLKDQSMALRWVSENIEWFGGNPKRITLIGLSAGGASVHYHYLSPSAGLFQGG
DM esterase     80 LLRE--DVVVISINYRLGPLGFLPGNAGLKDQVLALRWVKANCSRFGGDSANITIFGDSAGSASVHYMMITETHGLFHKA
HSal esterase   80 LMDH--DVIFVTINYRLGPLGFLPGNMGLKDQNMALRWVSENIQRFGGDPKQVTLCGMSAGGASVHYHYLSRSAGLFRGG
NV COesterase   80 LLKK--DIVLVTCNYRVGVLGFLPGNQGLKDQIMVLKWVQENIANFGGDANNVTLFGESAGASSVHYIALSDSRGLFHKT
HA COesterase   80 LVRH--GVILVTINYRLEVLGFLPGNAGMKDQVAALRWVNKNIANFGGDPNNVTIFGESAGGVSVSYQVISPSKGLFKRA
SL COesterase   80 LVRH--GVILVTINYRVDVLGFLPGNAGMKDQVQALRWVNKNIASFGGDPNNVTIFGESAGGGSVSYHLISPSKGLFKRA

PPE014_O05_f   155 VLDAGVFAG--IAKKLGC----------------------------------------------------
Contig5707     155 VMDAGVTSRT-IAAKVSCVADGSFVCPTNAFVKSYAVFVYNFEKV------KYLKFGPEFGAFLTMSEE
PPE0138_P16_f  151 IIQSGSI-------------------------------------------------------------
PPE013_C07_f       -------------------------------------------------------------------
DM ACE         161 MMQSGTMNAPWLINDCNCVGDHFFTCPTNEYAQALAVHYYYFTHRVLHGDEIEYFFGQPVIEFAKTGNP
RM ACE-1       160 ILQSGSPAVPWLATALRAVGDYHFTCPVVRWADTFAVYQYVFARRVIHGEEVPFVFGEPWANFAKTGNP
RM ACE-2       160 IVQSAAVTKKGFAANFGCKGDTHVKCGAINTACKLAVHFYELNYVMTHGDELPLVFGRVWSDFAKGRSP
RM ACE-3       160 IMESG--SPYFLARALQCVGDFAIICPTKYFAESFAVHYYYFTHRPTHGDEVFFVMGLPWTTFAKTGKV
IS (2)         160 YLMSGSVFTLLIAKEANCFTDVLFRCPMESMGKEFDVFFQEFWPKASHADDVFMLFGYPLAAFSRDG--
IS (4)         161 IMQSG--SPFSLATALGCFGHLQIVCPTLEFGKRLRVYMYQFSHRTTHGDDIPFSLGSMISSFSHTGIP
IS (5)         160 IMQSG--SPYWAAVKLNCMGDFAFLCPLMYFAEDYAVHFYEFAHRVAHFEEIPFVFGYTWTHFAKTGEV
IS (6)         161 IMQSG--TPRWLARNVGCAGHFFFDCPTVFMADTLALWAYRFQHRVTHTDEIQFVFGVPWITFAHTGYP
IS (7)         161 IMQSGTATAPWLAEALHCVGDYHFTCSVSEWAHHYAVYVYYFTHRVIHGEEIAFLFGEPWANFAKTGNP
IS (8)         160 IMQSG--TPYWLAMGLGCVGDFSLICPSQLFAEAFAVYFYLFNHRVTHGDELGFVFGFPWSTFAKNGYP
IS (9)         160 IMQSS--SPLSIAGKLGCVGDLLMNCATDLFAEVTSTHRYVFEHRVAHADDLFKLLSSLLDIFCTTRKP
IS (10)        160 IMQSS--TPLSIAGKLNCVGDVLMNCPTDLFAEVTSTHRYVFDHRVAHADEIPFALGSLSKLLSSVRYT
BM ACE-1       161 ILQSGTLNAPWLIDDCNCVGDYFFVCPTNYFAEILAVYYYYFTHRVMHGDEMEYVFGHPFTQFALTGKP
BM ACE-2       161 IMQSGAATAPWLAEAVHCVGDYHFTCGVNEFAHRYAVYTYYYKHRVMHADEINYVFGEPWANFARSGNP
TC ACE-1       161 IMESGSATAPWLAEAVGCVGDYHFTCNVNEFAHRYAVYMYYYKHRVMHADEINYVFGEPWANFAKTGNP
TC ACE-2       161 ILQSGTMNAPWLVEDCGCVGDYFFICPTNDFAELAAVYYYFFTHRVMHGDEIEYVFGHPFARFAATGKP
CE ACE-1       157 IIQSGSATSPWLYNAMKCLGDYHFTCSVNEMALAHTTYYYYFTHRVLHGYEINFIFGEPWANFAKTGDP
CE ACE-2       160 ILQSGSLDNKWLADLVGCVGDYFFTCDSIWFAHNYPVFVYYFDQPVMHGYEIEYVFGVPWTTFANTG-V
CE ACE-3       160 ILQSGSLENTWLLELVGCVGDYFFSCDSLWLADQFRVFVYHFTQSAMHGYEIEYVFGIPWASFAKNG-T
CE ACE-4       160 IAKSGSIINSWLAKKVNCLGDLFFTCS----LIDFAVYMYYFTYRVMHGYEIEYAFGQPWANFANTGRT
HS ACE         161 VLQSGAPNGPWLAHLVGCVGDHNVVCPVAQLAGRLAVYAYVFEHRVPHGYEIEFIFGIPWANFARTGDP
BM JHE         158 ILMSGAGTSTFLFSILGVCTDSYYKYPAMKLAEKRSVFLYQFSYDAGHADDLTYVLKVNVTNFMRCSAP
TC JHE         160 ISQSGAALDLWLAAFTGCYGDRAFHYGIYQTVILQLIWMYNFNYKVSHCDDLLYLFKSPWTNFAIYGNP
AM JHE         158 ISISGTALNCW-------ASDRFFITDGEKAARMQAVWFYYYTYKVCHADDAYMVVDTPWVSFVNNGVP
DM esterase    158 ICMSGNTLSPW-------VSHEYFWFPIYRTVLSRLTYLYRFDFDTCHGDDLSYLFYNSWTHFAACGNP
HSal esterase  158 ISVSGTALNCW-------AGDRFFVADSEKAARMQAVWYYYYTYRVSHGDDVYLIVDTPWVSFATNGVP
NV COesterase  158 ILQSGVSCNTW-------YTDLQFVEGIHRVVKAQVTYFYRFTYDASHFDEVQYLFNMQWVNFAKCGSP
HA COesterase  158 IAQSGVSVGYW-------YGMDIFSFPTVWWIRLIAAYLYKFTVKVAHSDDLSYLFS-SWVNFAKYGDP
SL COesterase  158 IAQSGASTSPW-------VSMQMFGYPTMQWVRHCALYVYKFSCKTSHADDIAYLFNPKWTNFAKYGNP
```

A_e -Aligned 221 amino acid positions from *Pp* ACEs group 2 alignment (20% of 1058 a.a.) after low stringent trimming with Gblocks (allow smaller final blocks, gap positions within the final blocks, and less strict flanking positions) for phylogenetic analysis.

See reference sequence information in Appendix A_f. Boxshade v.3.21 shading with black background for concensus and gray background for similar residues in the fraction of 0.5 aligned sequences.

```
PPE013_C07_r     1 ---------------------------------------------------------------------------
Contig4735       1 ---------------------------------------------------------------------------
Contig3716       1 ---------------------------------------------------------------------------
Contig5966       1 ---------------------------------------------------------------------------
Contig6677       1 ---------------------------------------------------------------------------
DM ACE           1 VQTSSGPVRGYTGIPYAKPPVEDLRFRKPGVLDATRLSATCVQEDCLYINVWAPPILIWIYGGGFMTGVIVASFQYRVGA
RM ACE-1         1 VETAWGPVKGFYGIPYAKPPTGKRRFDRAVLDATVK-PNSCFQEDCLKLNVWAPAVLVWIYGGGFYSGVVVVSMNYRVAS
RM ACE-2         1 VQIHAGKLRGFTGVPYAKPPVGELRYQKPEVKDATVTPPSCMQEDCLYLNVWTPPVMAWIHGGGFQEGVVVVTIAYRLQS
```

```
RM ACE-3          1 VETNSGPVQGFVGIPFAEPPVGPLRFRNPSVYQATKKPFPCLQ-DCLYLNVWTPTVIVYIYGGTFSFGVVMVSMNYRVGP
IS ACE(2)         1 VTGQVMGLPVFLGIPFADNTGGTRRFKKPGVFNATYTRKPCCQ-DCLHLNIWVPPVMFWVYGGGFVFGLIVVVPNYRVGS
IS ACE(4)         1 ILTSTGQVQGFLGIPYAEPPTGDLRFQKPGILNATSLPALCSQEDCLYLNVFTPPVIVSIHGGGFSSGLVTVAIAYRLGA
IS ACE(5)         1 VKTKSGLVRGFLGVPFAEPPVGALRFKPPREFRALFLPPSCIQ-DCLYLNVWAPAVMLFFHGGAFFFGVVVVTANYRLGP
IS ACE(6)         1 VPTSLGLVAGFLGIPFAEPPVGPKRFKRPGILNVIQLPRPCFQEDCLYLNIWVPPVMVWIHGGSYRIGVVVVSFNYRLGA
IS ACE(7)         1 VHTTKGPIRGFYGIPYAQPPVGRYRFRHPGVLDATVKPSSCYQEDCLTLNVWVPAVLVWIYGGGFYSGLVLVSMNYRVAS
IS ACE(8)         1 TYTRHGLVRGYYGIPYAEPPLAQLRFRKPGIYYATKKRPPCSQ-DCLHVNVWTSAVMVFLHGGGFQFGVVVVVPNYRLNV
IS ACE(9)         1 VATSSGLISGFYGIPYAKPPVGDLRFRKPGTYKATTKPMACNQ-DCLYMNVWRPPVVVFIYGGGFQWGVIFVSFNHRLSM
IS ACE(10)        1 NGILSELNEGFYGIPYAKPPVGDLRFRKPGTYEATTKPTACNQ-DCLYVNVWRPPVVVFIHGGGFQWGVIFVSFNHRLSM
BM ACE-1          1 IRTRKGKVRGWFGIPYAQKPMGDLRFRHPEILNTTTLPHSCVQEDCLYINIVTPAVMLWVFGGGFYSGVVYVSMQYRVAS
BM ACE-2          1 VETKSGLIKGFTGIPFAKPPLGPLRFRKPGVLEANLMPNSCYQEDCLYLNIWVPPILVWIYGGGYMSGVIVASMQYRVGA
TC ACE-1          1 IQTKKGKVRGWLGIPYAQKPLGNLRFRHPGVMNTTSQPNSCVQEDCLYVNVVVPAVMVWVFGGGFYSGIILVSMQYRVAS
TC ACE-2          1 VETTSGLVRGFTGIPFAKPPIEQLRFRKPGILDATKLPNSCYQEDCLYLNIWVPPVLVWIYGGGYMSGVIVASMQYRVGA
CE ACE-1          1 ----LIHLHDFQGIPFAEPPVGNLRFKKPIPLNATTPPNSCIQEDCLYLNVYVPAVMVWVYGGGFWSGVILVAMNYRVSI
CE ACE-2          1 VHTPLGTIRGFLGVPYAKPPIGSRRFKMAGELEARTLAKTCYLEDCLNMNIWVP-VMVWIYGGGFFSGTIVVNVNYRLGP
CE ACE-3          1 VQTKLGTVRGFLGVPFAEPPINEHRFKKPGTISANTLSPACFQEDCLYVNIWAP-VLVWLFGGGFWYGVIVVNINYRVGP
CE ACE-4          1 LETKLGDIKGFFGVPFAEPAVEDFRFRKPGLYDATKPANACFQEDCLYLNIWAP-VMVWFFGGGFYSGVIVVNINYRLGP
HS ACE            1 VTVRGGRLRGFLGIPFAEPPMGPRRFLPPGVVDATTFQSVCYQEDCLYLNVWTPPVLVWIYGGGFYSGTVLVSMNYRVGA
BM JHE            1 AQTESGWVCGFRGVPYAKQPVGELRFKELDYLDATEEGPVCYQ--CIYANIHVPPILVFIHGGGFAFGVVVITFNYRLNF
TC JHE            1 VYTKYGSVIGFKGIPFAKPPVGDLRFKAPFSINGTKDAPFCIQ-DCLYLNVYVPPVMVFIHWGGFFAGVILVTFNYRLGV
AM JHE            1 VKTPLGAIKGYEGIPYALPPVGKFRFKAPGELSATKFGFPCLQEDCLYLNVYVPPVIFWIHGGAFQFGVIFVTINYRLGI
DM esterase       1 VSTTYGPIKGFERIPFAKPPVGELRYKAPEVRSCTSQGPKPLQEDCLYLNVYTKPVMVWIYGGGFQFGVVVISINYRLGP
HSal esterase     1 IKTPLGGLKGYEGIPFALPPVGELRFKPPGELSATKLSSACIQEDCLYLNVYVPPVLFWIHGGAFQFGVIFVTINYRLGP
NV COesterase     1 VNTRSGRLRGFKGIPYAKPPVGELRFKDPGVRDATEYGTICGQDDCLYLNVYVRPVMVWIHGGAFMFGIVLVTCNYRVGV
HA COesterase     1 VRVSDGLLEGFRGIPYAQPPLGDLRFKAPNVRSAKEFGNNCLQEDCLYLNVYTPPVMVWIHGGGFVSGVILVTINYRLEV
SL COesterase     1 VRVNEGLLEGFKGIPYAEPPVGDLRFKAPGVRSAKEFGPKCYQEDCLYLNVYTPPVMFWIHGGGFFCGVILVTINYRVDV


PPE013_C07_r      1 ----------------------------------------------------------------------------
Contig4735        1 ------------------------------------------------------------KRAVMQSGSPFSFVLASYTDCR
Contig3716        1 ----------------------------------------------------------------------------
Contig5966        1 -----GHQGHLDQVLVLNWVSQNIRYFGGDPNRVTLFGVSAGSFSISWHLLTG-AGLFHAAVIDAGVLTHTHICKSLKEC
Contig6677        1 ----------------------------------------------------------------------------
DM ACE           81 FGFLPGNVGLWDQALAIRWLKDNAHAFGGNPEWMTLFGESAGSSSVNAQLMSPTRGLVKRGMMQSGTMNAPWLINDCNCC
RM ACE-1         80 LGFLPGNAGLYDQYMALKWVQENVAAFGGDPDRVTLFGESAGAASAGLHVLSPSESLFHRVILQSGSPAVPWLATALRAS
RM ACE-2         81 FGFL-GNMGLHDQQLALKWIQENIAAFGGNPGEVTLFGWSAGGISTGFHLISPSQTLFKRAIVQSAAVTKKGFAANFGCC
RM ACE-3         80 MGFFSGNAGLHDQLLAMKWVKQNIRNFGGDPDDVTLVGQSAGAISIGLHLVSPSKGLFKRIIMESG--SPYFLARALQCC
IS ACE(2)        80 FGFLPGNMGMHDVIAAYRWVRDHIGSFGGDRENIILAGQSAGSIISGLLMISPPLTLFSKAYLMSGSVFTLLIAKEANCC
IS ACE(4)        81 FGFLPGNMGLYDQVMALRWVKENIQSFGGDPDKITLMGPSAGSVAVGVHILSPSRGLFHRAIMQSG--SPFSLATALGCC
IS ACE(5)        80 FGFMPGNQGLYDQNLAMRWVRDNIRYFGGDEEQVTLFGQSAGAISIGYHLASPSKGLFKRVIMQSG--SPYWAAVKLNCC
IS ACE(6)        81 LGFLPGNMGLWDQYAALRWVNENIASFGGDPSRVTLFGESVGGASSGMLAQSPCRGLIRRIIMQSG--TPRWLARNVGCC
IS ACE(7)        81 LGFLPGNAGLFDQLMALQWIQDNIAAFGGNPRNVTLFGESAGAVSVSMHLLSPSRDLFSQAIMQSGTATAPWLAEALHCC
IS ACE(8)        80 FGFLPGNMGMYDQILALKWVQKNIAMFGGDPNRVTLFGQSAGSVSTGYHLLSPSRGLFRRVIMQSG--TPYWLAMGLGCC
IS ACE(9)        80 MGFLPGNLGFWDQLLVLKWVQQNIGRFGGNPQDVTLLGHSAGAVSAGLHAVSQSKGLFHRLIMQSS--SPLSIAGKLGCC
IS ACE(10)       80 MGFLPGNLGFWDQLLALKWVRRNIARFGGNPKDVTLAGHSAGAVSAGIHAVSQSKGLFHRLIMQSS--TPLSIAGKLNCC
BM ACE-1         81 LGFLPGNAGLFDQLMALQWVKDNIGYFGGNPHNITLFGESAGAVSVSLHLLSPSRNLFSQAIMQSGAATAPWLAEAVHCC
BM ACE-2         81 FGFLPGNMGLWDQQLAIRWIKENARAFGGDPELITLFGESAGGGSVSLHMLSPMKGLFKRGILQSGTLNAPWLIDDCNCC
TC ACE-1         81 FGFLPGNMGMWDQALAIRWIKENAAAFGGDPDLITLFGESAGGGSVSILLLSPTKGLARRGILQSGTMNAPWLVEDCGCC
TC ACE-2         81 LGFLPGNAGLFDQMMALQWVRDNIAAFGGNPNNITLFGESAGAVSVSLHLLSPSRNLFSQAIMESGSATAPWLAEAVGCC
CE ACE-1         77 FGFLPGNMGMWDQLLAMKWVHKNIDLFGGDLSRITLFGESAGAASVSIHMLSPSAPYFHRAIIQSGSATSPWLYNAMKCC
CE ACE-2         80 FGFLQGNMGLMDQQLALRWVHENIGAFGGDRSRVTLFGESAGSASTTAHLFAPSHKYFRNIIAKSGSIINSWLAKKVNCC
CE ACE-3         80 FGYLPGNMGMLDQQLALYWIRDHIFSFGGNPARISLVGESAGAASIVAHLIAPSKGLFQNGILQSGSLDNKWLADLVGCC
CE ACE-4         80 FGFLPGNMGLLDQQLALHWVRQNIVSFGGNPDKVAVFGQSAGAASIVAHLIAPSRGLFKNAILQSGSLENTWLLELVGCC
HS ACE           81 FGFLPGNVGLLDQRLALQWVQENVAAFGGDPTSVTLFGESAGAASVGMHLLSPSRGLFHRAVLQSGAPNGPWLAHLVGCC
```

```
BM JHE          79 FGFFPGNNGLRDMVTLLRWVKRNARAFGGNPDNVTLAGQSAGAAAAHLLTLSKTEGLVSRAILMSGAGTSTFLFSILGVV
TC JHE          80 FGFLPGNFGLKDQVMALKFVHENIECFGGDNNRVTIFGQSAGSGSVNLHLISPSRGLFQQAISQSGAALDLWLAAFTGCC
AM JHE          81 LGFLPGNMGLKDQSMALRWVSENIEWFGGNPKRITLIGLSAGGASVHYHYLSPSAGLFQGGISISGTALNCW-------C
DM esterase     81 LGFLPGNAGLKDQVLALRWVKANCSRFGGDSANITIFGDSAGSASVHYMMITETHGLFHKAICMSGNTLSPW-------F
HSal esterase   81 LGFLPGNMGLKDQNMALRWVSENIQRFGGDPKQVTLCGMSAGGASVHYHYLSRSAGLFRGGISVSGTALNCW-------C
NV COesterase   81 LGFLPGNQGLKDQIMVLKWVQENIANFGGDANNVTLFGESAGASSVHYIALSDSRGLFHKTILQSGVSCNTW-------F
HA COesterase   81 LGFLPGNAGMKDQVAALRWVNKNIANFGGDPNNVTIFGESAGGVSVSYQVISPSKGLFKRAIAQSGVSVGYW-------F
SL COesterase   81 LGFLPGNAGMKDQVQALRWVNKNIASFGGDPNNVTIFGESAGGGSVSYHLISPSKGLFKRAIAQSGASTSPW-------F

PPE013_C07_r     1 -----------GDLAMVCPTKYFAEEAAVHYYEFDFRTTHGEEIPFVFGHPWTDFAKTGLT
Contig4735      23 MK---KNSFFPGDYPYLCATTDFGVEYAVYHMQYDHRSTHGDDIMFWLGSVRTRADERVAD
Contig3716       1 ----------------------------AVYRYLWEHKATHCDDVPFTMGSQITEAEKQLIK
Contig5966      75 LMKIDAYAFRPGDGLFVCPVMNFADRYSVYFYRWERVFYTGVGSQYLY---DKSYIENTIK
Contig6677       1 -RNIPALSFIPADGWFKCPINTFIRSYSVFAYQFERRFHYSPYLHFGGAILVNEGDKQFSL
DM ACE         161 MRSVDAFPSAPGDHFFTCPTNEYAQALAVHYYYFTHRVLHGDEIEYFFGQPVIEFAKTGNP
RM ACE-1       160 LRCERPFPFVPGDYHFTCPVVRWADTFAVYQYVFARRVIHGEEVPFVFGEPWANFAKTGNP
RM ACE-2       160 MRTVNAGKFEPGDTHVKCGAINTACKLAVHFYELNYVMTHGDELPLVFGRVWSDFAKGRSP
RM ACE-3       158 LRKIDGLTFFPGDFAIICPTKYFAESFAVHYYYFTHRPTHGDEVFFVMGLPWTTFAKTGKV
IS ACE(2)      160 LRARTSQIFSPTDVLFRCPMESMGKEFDVFFQEFWPKASHADDVFMLFGYPLAAFSRDG--
IS ACE(4)      159 LRSKDVDAFFPGHLQIVCPTLEFGKRLRVYMYQFSHRTTHGDDIPFSLGSMISSFSHTGIP
IS ACE(5)      158 LQEKNALTYHPGDFAFLCPLMYFAEDYAVHFYEFAHRVAHFEEIPFVFGYTWTHFAKTGEV
IS ACE(6)      159 LQKVSAFTFQPGHFFFDCPTVFMADTLALWAYRFQHRVTHTDEIQFVFGVPWITFAHTGYP
IS ACE(7)      161 LRRQDPFAFVPGDYHFTCSVSEWAHHYAVYVYYFTHRVIHGEEIAFLFGEPWANFAKTGNP
IS ACE(8)      158 VRSKSRNYMIPGDFSLICPSQLFAEAFAVYFYLFNHRVTHGDELGFVFGFPWSTFAKNGYP
IS ACE(9)      158 LRKIDVQFFSPGDLLMNCATDLFAEVTSTHRYVFEHRVAHADDLFKLLSSLLDIFCTTRKP
IS ACE(10)     158 LRQIDAQFFSPGDVLMNCPTDLFAEVTSTHRYVFDHRVAHADEIPFALGSLSKLLSSVRYT
BM ACE-1       161 LRKKNAFPFVPGDYHFTCGVNEFAHRYAVYTYYYKHRVMHADEINYVFGEPWANFARSGNP
BM ACE-2       161 MRGVDAFPSAPGDYFFVCPTNYFAEILAVYYYYFTHRVMHGDEMEYVFGHPFTQFALTGKP
TC ACE-1       161 MRAVEAFPSTPGDYFFICPTNDFAELAAVYYYFFTHRVMHGDEIEYVFGHPFARFAATGKP
TC ACE-2       161 LKKKDPFPFVPGDYHFTCNVNEFAHRYAVYMYYYKHRVMHADEINYVFGEPWANFAKTGNP
CE ACE-1       157 FQRADAFPWVPGDYHFTCSVNEMALAHTTYYYYFTHRVLHGYEINFIFGEPWANFAKTGDP
CE ACE-2       160 LRSVPAFAYVPGDLFFTCS----LIDFAVYMYYFTYRVMHGYEIEYAFGQPWANFANTGRT
CE ACE-3       160 LRNTPAFPFAIGDYFFTCDSIWFAHNYPVFVYYFDQPVMHGYEIEYVFGVPWTTFANTG-V
CE ACE-4       160 LRLVSPFPFVIGDYFFSCDSLWLADQFRVFVYHFTQSAMHGYEIEYVFGIPWASFAKNG-T
HS ACE         161 LRTRPAFSFVPGDHNVVCPVAQLAGRLAVYAYVFEHRVPHGYEIEFIFGIPWANFARTGDP
BM JHE         159 EKLNEAVVETPTDSYYKYPAMKLAEKRSVFLYQFSYDAGHADDLTYVLKVNVTNFMRCSAP
TC JHE         160 LRKIEAEKQTDGDRAFHYGIYQTVILQLIWMYNFNYKVSHCDDLLYLFKSPWTNFAIYGNP
AM JHE         154 LRYRPAFGPVTSDRFFITDGEKAARMQAVWFYYYTYKVCHADDAYMVVDTPWVSFVNNGVP
DM esterase    154 LKNAKGFSFGPSHEYFWFPIYRTVLSRLTYLYRFDFDTCHGDDLSYLFYNSWTHFAACGNP
HSal esterase  154 LRTRPAFGAVVGDRFFVADSEKAARMQAVWYYYYTYRVSHGDDVYLIVDTPWVSFATNGVP
NV COesterase  154 LRSIKCFPFGPTDLQFVEGIHRVVKAQVTYFYRFTYDASHFDEVQYLFNMQWVNFAKCGSP
HA COesterase  154 LKVQPAVYFSVGMDIFSFPTVWWIRLIAAYLYKFTVKVAHSDDLSYLFS-SWVNFAKYGDP
SL COesterase  154 FKNQPKINFTVSMQMFGYPTMQWVRHCALYVYKFSCKTSHADDIAYLFNPKWTNFAKYGNP
```

A_f - Gene name, amino acid length and accession number of esterases and carboxyesterases used for sequence alignment and phylogenetic analysis performed with *Pp* ACE -like, arachnid specific esterases.

Table A.1. Gene name, amino acid length and accession number of esterases and carboxyesterases for *Pp* ACE-like esterases analysis.

Species	Abbreviation in figures	Gene name (NCBI)	peptide length (# a.a.)	NCBI/GenBank accession #
Harpegnathos saltator	HSal esterase	esterase FE4	564	EFN85479.1
Caenorhabditis elegans	CE ACE-1	abnormal Acetylcholinesterase family member (ace-1)	620	NP_510660.1
Caenorhabditis elegans	CE ACE-2	abnormal Acetylcholinesterase family member (ace-2)	629	NP_491141.1
Caenorhabditis elegans	CE ACE-3	abnormal Acetylcholinesterase family member (ace-3)	607	NP_496963.1
Caenorhabditis elegans	CE ACE-4	abnormal Acetylcholinesterase family member (ace-4)	602	NP_496962.1
Drosophila melanogaster	DM ACE (ACE-2)	acetylcholine esterase, isoform A	649	NP_476953.1
Homo sapiens	HS ACE	acetylcholinesterase	614	NP_000656.1
Ixodes scapularis	IS ICE(2)	acetylcholinesterase (2)	493	XP_002400739.1
Ixodes scapularis	IS ICE(4)	acetylcholinesterase (4)	547	XP_002414019.1
Ixodes scapularis	IS ICE(5)	acetylcholinesterase (5)	558	XP_002414015.1
Ixodes scapularis	IS ICE(6)	acetylcholinesterase (6)	538	XP_002414013.1
Ixodes scapularis	IS ICE(7)	acetylcholinesterase (7)	623	XP_002413212.1
Ixodes scapularis	IS ICE(8)	acetylcholinesterase (8)	534	XP_002413109.1
Ixodes scapularis	IS ICE(9)	acetylcholinesterase (9)	617	XP_002412990.1
Ixodes scapularis	IS ICE(10)	acetylcholinesterase (10)	654	XP_002412917.1
Tribolium castaneum	TC ACE-1	acetylcholinesterase ace-1	648	EFA04156.1
Tribolium castaneum	TC ACE-2	acetylcholinesterase ace-2	703	EEZ99262.1
Rhipicephalus microplus	RM ACE-1	acetylcholinesterase 1	595	CAA11702.1
Rhipicephalus microplus	RM ACE-2	acetylcholinesterase 2	563	CAB93511.1
Rhipicephalus microplus	RM ACE-3	acetylcholinesterase 3	620	AAP92139.1
Bombyx mori	BM ACE-1	acetylcholinesterase type 1	683	NP_001037380.1
Bombyx mori	BM CAE-2	acetylcholinesterase type 2	638	NP_001108113.1
Drosophila melanogaster	DM esterase	alpha-Esterase-9, isoform B	554	NP_524258.1
Helicoverpa armigera	HA carboxylesterase	carboxylesterase	597	ABQ42338.1
Spodoptera litura	SL carbosylesterase	carboxylesterase	537	ABE01157
Apis mellifera	AM JHE	juvenile hormone esterase	567	NP_001011563.1
Bombyx mori	BM JHE	juvenile hormone esterase	570	NP_001037027.1
Nasonia vitripennis	NV carboxylesterase	carboxylesterase clade A, member 9	530	NP_001165951.1
Tribolium castaneum	TC JHE	juvenile hormone esterase	588	NP_001180223.1

Appendix B - *P. persimilis* uniEST reference number and reference/accession number of 74 orthologous groups conformed by amino acid sequence of eleven arthropod species retrieved from OrthoDB (database orthologous group reference number and sequence reference number), and the accession number for *Caenorhabditis elegans* from WormBase.

Species name is abbreviated to the first letter of genus and species epithet, for example: AA = *Aeders aegypti*. Accession number for protein sequences retrieved from GenBank are found in *Aedes aegypti* (ABF18332.1, ABF18250, ABF18250), *Bombyx mori* (NP_001036892.1, NP_001040257, NP_001040387.1), *Tribolium castaneum* (XP_969827.2), *Apis mellifera* (XP_001120515), *Nassonia vitripennis* (NP_001153445.1, NP_001153430.1, NP_001153324.1, NP_001153324.1), *Acyrthosiphon pisum* (XP_001949197, NP_001153808), and *Ixodes scapularis* (XP_002415425).

Table B.1. *P. persimilis* uniEST reference number and reference/accession number of 74 orthologous groups.

			ORTHOLOG PROTEIN ACCESSION NUMBER												
Pp #	**contig/singlet name**	**BLAST2GO GENE NAME**	**OrthoDB #**	**AA**	**AG**	**CQ**	**DM**	**BM**	**TC**	**AM**	**NV**	**AP**	**DP**	**IS**	**CE**
P1	Contig4172	spectrin alpha	EOG9SN1WH	AAEL015065-PA	AGAP006686-PA	CPIJ014790-PA	FBgn0250789	BGIBMGA007948-PA	TcGLEAN_00749	GB18557-PA	NV15886-PA	ACYPI004133-PA	JGI_V11_206384	ISCW000012-PA	WP:CE07373
P2	Contig5943	splicing factor 3b	EOG95MNCB	AAEL003605-PA	AGAP000178-PA	CPIJ009000-PA	FBgn0031266	BGIBMGA006851-PA	TcGLEAN_12382	GB16777-PA	NV15361-PA	ACYPI003176-PA	JGI_V11_328912	ISCW009360-PA	WP:CE03641
P3	Contig33	rab gdp-dissociation inhibitor	EOG9FTW6W	AAEL012904-PA	AGAP010025-PA	CPIJ015309-PA	FBgn0004868	BGIBMGA001119-PA	TcGLEAN_07497	GB13730-PA	NV18770-PA	ACYPI009608-PA	JGI_V11_207615	ISCW017892-PA	WP:CE14944
P4	Contig5052	s-adenosylhomocysteine hydrolase	EOG9MPHWG	AAEL008341-PA	AGAP000719-PA	CPIJ008116-PA	FBgn0014455	BGIBMGA011168-PA	TcGLEAN_14564	GB14324-PA	NP_001153324.1	ACYPI003908-PA	JGI_V11_299913	ISCW015944-PA	WP:CE17154
P5	Contig6202	proteasome (macropain) 26s 1	EOG9RJG85	AAEL012095-PA	AGAP003216-PA	CPIJ011552-PA	FBgn0015282	BGIBMGA004908-PA	TcGLEAN_09675	GB15218-PA	NV15169-PA	ACYPI009001-PA	JGI_V11_306388	ISCW018294-PA	WP:CE28229
P6	Contig4053	proteasome (macropain) 26s 1	EOG9RJG85	AAEL012095-PA	AGAP003216-PA	CPIJ011552-PA	FBgn0015282	BGIBMGA004908-PA	TcGLEAN_09675	GB15218-PA	NV15169-PA	ACYPI009001-PA	JGI_V11_306388	ISCW018294-PA	WP:CE28229
P7	Contig6173	metalloprotease m41 ftsh	EOG9DZ21Q	AAEL008565-PA	AGAP006949-PA	CPIJ006578-PA	FBgn0036702	BGIBMGA010088-PA	TcGLEAN_02463	GB16186-PA	NV17214-PA	ACYPI005429-PA	JGI_V11_194014	ISCW009180-PA	WP:CE34400
P8	Contig2335	heat shock 70kda protein 9	EOG93R3V0	ABF18332.1	AGAP010876-PA	CPIJ004454-PA	FBgn0001220	NP_001036892.1	TcGLEAN_00487	GB19860-PA	NV18519-PA	ACYPI004693-PA	JGI_V11_313359	ISCW017192-PA	WP:CE08631
P9	Contig3547	inosine 5 -phosphate dehydrogenase 2	EOG9JM7WP	AAEL009273-PA	AGAP011133-PA	CPIJ011687-PA	FBgn0003204	BGIBMGA001035-PA	TcGLEAN_05099	GB16346-PA	NV18927-PA	ACYPI004360-PA	JGI_V11_218652	ISCW001657-PA	WP:CE30188
P10	Contig4308	transport protein sec61 subunit alpha 2-not use kept he number	EOG9N30VX	AAEL010716-PA	AGAP009182-PA	CPIJ007723-PA	FBgn0086357	BGIBMGA012687-PA	TcGLEAN_10557	GB11642-PA	NV16578-PA	ACYPI000055-PA	JGI_V11_308832	ISCW020077-PA	WP:CE14954
P11	Contig4919	ribosomal protein l3	EOG9VT647	ABF18250	AGAP003556-PA	CPIJ007488-PA	FBgn0020910	BGIBMGA013567-PA	TcGLEAN_05395	GB15261-PA	NV11608-PA	ACYPI009439-PA	JGI_V11_301006	ISCW023076-PA	WP:CE05598
P12	Contig5164	glycerol-3-phosphate dehydrogenase	EOG9NS3HV	AAEL003873-PB	AGAP004437-PA	CPIJ003608-PA	FBgn0022160	BGIBMGA003049-PA	TcGLEAN_14609	GB12526-PA	NV18749-PA	ACYPI001475-PA	JGI_V11_187868	ISCW023707-PA	WP:CE14180
P13	Contig1978	ribosomal protein l3	EOG9VT647	ABF18250	AGAP003556-PA	CPIJ007488-PA	FBgn0020910	BGIBMGA013567-PA	TcGLEAN_05395	GB15261-PA	NV11608-PA	ACYPI009439-PA	JGI_V11_301006	ISCW023076-PA	WP:CE05598
P14	Contig5499	cop9 constitutive photomorphogenic homolog subunit 5	EOG9B5PBH	AAEL014747-PA	AGAP002880-PA	CPIJ012314-PA	FBgn0027053	BGIBMGA002790-PA	TcGLEAN_11816	GB12145-PA	NV19042-PA	ACYPI006786-PA	JGI_V11_109433	ISCW019527-PA	WP:CE06722
P15	Contig430	electron-transferring-flavoprotein dehydrogenase	EOG92BXGC	AAEL007707-PA	AGAP009510-PA	CPIJ002460-PA	FBgn0032237	NP_001040257	TcGLEAN_08177	GB18727-PA	NV16898-PA	XP_001949197	JGI_V11_308915	ISCW007624-PA	WP:CE20820
P16	PPE0125_K20.f	cytidylyltransferase	EOG9V176H	AAEL001940-PA	AGAP007175-PA	CPIJ009904-PA	FBgn0010350	BGIBMGA012695-PA	TcGLEAN_14743	GB15128-PA	NV16172-PA	ACYPI004171-PA	JGI_V11_315618	ISCW022031-PA	WP:CE26907
P17	Contig3784	tubulin gamma-1	EOG9JWVM8	AAEL013064-PA	AGAP007834-PA	CPIJ005487-PA	FBgn0260639	BGIBMGA013500-PA	TcGLEAN_14883	GB18755-PA	NV16082-PA	ACYPI004513-PA	JGI_V11_301492	XP_002415425	WP:CE00224
P18	PPE0116_C14.f	protein phosphatase-5	EOG91VK95	AAEL005080-PA	AGAP004151-PA	CPIJ009620-PA	FBgn0005777	BGIBMGA004807-PA	TcGLEAN_08963	GB12422-PA	NV18433-PA	ACYPI000637-PA	JGI_V11_196725	ISCW013839-PA	WP:CE36619
P19	Contig2874	nadh dehydrogenase ubiquinone flavoprotein 1 precursor	EOG9V43FW	AAEL007681-PA	AGAP010039-PA	CPIJ006281-PA	FBgn0031771	BGIBMGA006011-PA	TcGLEAN_14119	GB17095-PA	NV11489-PA	ACYPI000112-PA	JGI_V11_53209	ISCW005985-PA	WP:CE02132
P20	Contig247	beta	EOG995ZZV	AAEL013461-PA	AGAP002356-PA	CPIJ003742-PA	FBgn0052528	BGIBMGA000683-PA	TcGLEAN_00078	GB17089-PA	NV16836-PA	NP_001153808	JGI_V11_39424	ISCW002088-PA	WP:CE18268
P21	Contig4414	isocitrate dehydrogenase	EOG9HHP8V	AAEL000454-PA	AGAP002728-PA	CPIJ002859-PA	FBgn0027291	BGIBMGA007160-PA	TcGLEAN_04962	GB19422-PA	NV19148-PA	ACYPI002445-PA	JGI_V11_222251	ISCW004116-PA	WP:CE34018
P22	Contig5973	succinate- gdp- beta subunit	EOG92RDF0	AAEL005552-PA	AGAP004352-PA	CPIJ005779-PA	FBgn0029118	BGIBMGA007824-PA	TcGLEAN_05166	XP_001120515	NV13707-PA	ACYPI009250-PA	JGI_V11_306482	ISCW015999-PA	WP:CE04242
P23	Contig2454	prp6 pre-mrna splicing factor 6 homolog	EOG9DV5TJ	AAEL010387-PA	AGAP005640-PA	CPIJ016885-PA	FBgn0036828	BGIBMGA008911-PA	TcGLEAN_05794	GB14450-PA	NV16313-PA	ACYPI002837-PA	JGI_V11_303888	ISCW022841-PA	WP:CE28858
P24	Contig3492	developmentally regulated gtp binding protein 2	EOG9Q58W5	AAEL012250-PA	AGAP002185-PA	CPIJ013949-PA	FBgn0038723	BGIBMGA007131-PA	TcGLEAN_05461	GB19395-PA	NV14395-PA	ACYPI004640-PA	JGI_V11_187388	ISCW023585-PA	WP:CE36094
P25	Contig4622	upf0027 protein c22orf28	EOG9CC47P	AAEL003336-PA	AGAP008147-PA	CPIJ004874-PA	FBgn0032781	BGIBMGA006828-PA	TcGLEAN_10635	GB14373-PA	NV11757-PA	ACYPI001460-PA	JGI_V11_302483	ISCW024552-PA	WP:CE23663
P26	Contig6495	vacuolar protein sorting 26 homolog b	EOG9Q8546	AAEL014323-PA	AGAP006139-PA	CPIJ011277-PA	FBgn0014411	BGIBMGA000120-PA	TcGLEAN_15215	GB14744-PA	NV13952-PA	ACYPI008834-PA	JGI_V11_309382	ISCW018536-PA	WP:CE28684
P27	Contig6479	usp39 protein	EOG94TPF7	AAEL001827-PA	AGAP008381-PA	CPIJ001879-PA	FBgn0030969	BGIBMGA007676-PA	TcGLEAN_03290	GB10090-PA	NV15065-PA	ACYPI002195-PA	JGI_V11_187802	ISCW016690-PA	WP:CE20660
P28	Contig5676	ras-related gtp binding a	EOG9G7C4P	AAEL001188-PA	AGAP002991-PA	CPIJ007526-PA	FBgn0037647	BGIBMGA003355-PA	TcGLEAN_10924	GB17040-PA	NV13473-PA	ACYPI009223-PA	JGI_V11_301559	ISCW011735-PA	WP:CE02361
P29	Contig5266	fructose -bisphosphate aldolase	EOG99GKP1	AAEL005766-PB	AGAP002564-PA	CPIJ009571-PA	FBgn0000064	BGIBMGA013021-PA	TcGLEAN_14998	GB19460-PA	NV16284-PA	ACYPI007027-PA	JGI_V11_188180	ISCW011371-PA	WP:CE30646

P30	Contig4973	pyruvate dehydrogenase beta	EOG9MPHWH	AAEL004338-PA	AGAP010421-PA	CPIJ016430-PA	FBgn0039635	BGIBMGA002750-PA	TcGLEAN_11159	GB17238-PA	NV13940-PA	ACYPI005282-PA	JGI_V11_327265	ISCW009219-PA	WP:CE27647
P31	Contig4522	pelota	EOG995ZZX	AAEL007854-PA	AGAP008269-PA	CPIJ005823-PA	FBgn0011207	BGIBMGA006754-PA	TcGLEAN_01682	GB10750-PA	NV13121-PA	ACYPI009932-PA	JGI_V11_223147	ISCW013073-PA	WP:CE01059
P32	Contig4861	s-adenosylhomocysteine hydrolase	EOG9MPHWG	AAEL008341-PA	AGAP000719-PA	CPIJ008116-PA	FBgn0014455	BGIBMGA011168-PA	TcGLEAN_14564	GB14324-PA	NP_001153324.1	ACYPI003908-PA	JGI_V11_299913	ISCW015944-PA	WP:CE17154
P33	Contig2081	mitogen-activated protein kinase kinase 1	EOG9894TR	AAEL012723-PA	AGAP001103-PA	CPIJ015801-PA	FBgn0010269	BGIBMGA001458-PA	TcGLEAN_04803	GB13922-PA	NV21219-PA	ACYPI000770-PA	JGI_V11_46211	ISCW005428-PA	WP:CE25437
P34	Contig3930	wd40 protein	EOG9Q8557	AAEL001211-PA	AGAP009700-PA	CPIJ005918-PA	FBgn0032030	BGIBMGA005930-PA	TcGLEAN_14755	GB13929-PA	NV14890-PA	ACYPI003455-PA	JGI_V11_300247	ISCW012635-PA	WP:CE04150
P35	Contig4287	proteasome (macropain) 26s non- 12	EOG9SJ5M8	AAEL014325-PA	AGAP005535-PA	CPIJ018130-PA	FBgn0028690	BGIBMGA007268-PA	TcGLEAN_06260	GB11260-PA	NV10361-PA	ACYPI004078-PA	JGI_V11_198904	ISCW002593-PA	WP:CE02632
P36	Contig3839	fructose- -bisphosphatase	EOG9S4PNR	AAEL001158-PB	AGAP009173-PA	CPIJ010409-PA	FBgn0032820	BGIBMGA001399-PA	TcGLEAN_01209	GB17912-PA	NV11497-PA	ACYPI002694-PA	JGI_V11_227042	ISCW005292-PA	WP:CE21023
P37	Contig1040	short-chain dehydrogenase	EOG9DR9K2	AAEL002416-PA	AGAP012513-PA	CPIJ009733-PA	FBgn0039537	BGIBMGA007118-PA	TcGLEAN_05801	GB17223-PA	NV14165-PA	ACYPI000192-PA	JGI_V11_203975	ISCW018577-PA	WP:CE02490
P38	Contig996	cyclin-dependent kinase 7	EOG9D534B	AAEL001038-PA	AGAP002646-PA	CPIJ018296-PA	FBgn0015617	BGIBMGA001550-PA	TcGLEAN_03871	GB12759-PA	NV12779-PA	ACYPI005789-PA	JGI_V11_213779	ISCW020730-PA	WP:CE30320
P39	Contig6301	atp-binding cassette sub-family f member 2	EOG9JT0C1	AAEL010977-PA	AGAP002693-PA	CPIJ002890-PA	FBgn0030672	BGIBMGA002004-PA	TcGLEAN_13884	GB18938-PA	NV10139-PA	ACYPI009003-PA	JGI_V11_304799	ISCW009589-PA	WP:CE18971
P40	Contig3197	arp1 actin-related protein 1 homolog centractin alpha	EOG9JT0CN	AAEL003383-PA	AGAP002127-PA	CPIJ001227-PA	FBgn0011745	BGIBMGA012151-PA	XP_969827.2	GB10819-PA	NV18912-PA	ACYPI006251-PA	JGI_V11_203872	ISCW022123-PA	WP:CE35675
P41	Contig4892	fructose- -bisphosphatase (1)	EOG9S4PNR	AAEL001158-PB	AGAP009173-PA	CPIJ010409-PA	FBgn0032820	BGIBMGA001399-PA	TcGLEAN_01209	GB17912-PA	NV11497-PA	ACYPI002694-PA	JGI_V11_227042	ISCW005292-PA	WP:CE21023
P42	Contig6783	general transcription factor iib	EOG9KPTJ2	AAEL005368-PB	AGAP009558-PA	CPIJ006843-PA	FBgn0004915	BGIBMGA006014-PA	TcGLEAN_10900	GB18567-PA	NV12842-PA	ACYPI000165-PA	JGI_V11_325996	ISCW023433-PA	WP:CE27396
P43	Contig5474	cyclin-dependent kinase 5	EOG9S1TFD	AAEL008648-PA	AGAP005772-PA	CPIJ000065-PA	FBgn0013762	BGIBMGA004018-PA	TcGLEAN_09740	GB11941-PA	NV14994-PA	ACYPI001995-PA	JGI_V11_299705	ISCW021518-PA	WP:CE21213
P44	PPE0122_D17.f	succinate dehydrogenase	EOG9C5CRB	AAEL010608-PA	AGAP010429-PA	CPIJ006167-PA	FBgn0017539	BGIBMGA009000-PA	TcGLEAN_06271	GB17439-PA	NV12395-PA	ACYPI007401-PA	JGI_V11_228809	ISCW000555-PA	WP:CE03917
P45	PPE0110_G20.f	phosphoglycerate kinase 1	EOG96HGHK	AAEL004988-PA	AGAP008802-PA	CPIJ009280-PA	FBgn0250906	BGIBMGA007681-PA	TcGLEAN_15540	GB11056-PA	NV10148-PA	ACYPI007869-PA	JGI_V11_299795	ISCW015616-PA	WP:CE13100
P46	Contig4135	phosphatidylinositol class k	EOG905S77	AAEL010465-PA	AGAP004301-PA	CPIJ014600-PA	FBgn0023545	BGIBMGA001918-PA	TcGLEAN_09622	GB18506-PA	NV15190-PA	ACYPI002167-PA	JGI_V11_326467	ISCW000202-PA	WP:CE40776
P47	Contig454	ras-related gtp binding a	EOG9G7C4P	AAEL001188-PA	AGAP002991-PA	CPIJ007526-PA	FBgn0037647	BGIBMGA003355-PA	TcGLEAN_10924	GB17040-PA	NV13473-PA	ACYPI009223-PA	JGI_V11_301559	ISCW011735-PA	WP:CE02361
P48	PPE0124_C03.f	smu-1 suppressor of mec-8 and unc-52 homolog (elegans	EOG99PB5F	AAEL010896-PA	AGAP001414-PA	CPIJ000644-PA	FBgn0038666	BGIBMGA013454-PA	TcGLEAN_01051	GB19636-PA	NV17415-PA	ACYPI005047-PA	JGI_V11_304030	ISCW014661-PA	WP:CE15742
P49	Contig1653	eukaryotic translation initiation factor subunit 1 alpha	EOG97WNVZ	AAEL013675-PA	AGAP011190-PA	CPIJ011194-PA	FBgn0004925	BGIBMGA004302-PA	TcGLEAN_10161	GB18739-PA	NV15947-PA	ACYPI009395-PA	JGI_V11_306457	ISCW017360-PA	WP:CE29373
P50	Contig3190	dolichyl-diphosphooligosaccharide -protein glycosyltransferase	EOG9TQMG9	AAEL002174-PA	AGAP006383-PA	CPIJ005168-PA	FBgn0014868	BGIBMGA002814-PA	TcGLEAN_10678	GB18042-PA	NV12354-PA	ACYPI005607-PA	JGI_V11_319543	ISCW017658-PA	WP:CE01081
P51	Contig5566	26s proteasome subunit s9	EOG9B5PBS	AAEL012419-PA	AGAP001440-PA	CPIJ015971-PA	FBgn0028689	BGIBMGA014136-PA	TcGLEAN_12621	GB18440-PA	NP_001153445.1	ACYPI005455-PA	JGI_V11_58294	ISCW012119-PA	WP:CE25011
P52	Contig1957	ribosomal protein l4	EOG91ZFHP	AAEL009994-PA	AGAP002306-PA	CPIJ003890-PA	FBgn0003279	BGIBMGA001800-PA	TcGLEAN_13168	GB15503-PA	NV10263-PA	ACYPI008165-PA	JGI_V11_205304	ISCW007478-PA	WP:CE07669
P53	PPE0127_B20.f	chaperonin containing subunit 4	EOG9T78B2	AAEL007702-PA	AGAP009505-PA	CPIJ002450-PA	FBgn0032444	BGIBMGA010666-PA	TcGLEAN_07791	GB16452-PA	NV10040-PA	ACYPI003279-PA	JGI_V11_212503	ISCW007627-PA	WP:CE02262
P54	PPE015_J15.f	arp2 actin-related protein 2 homolog	EOG9KSPRM	AAEL011750-PA	AGAP000985-PA	CPIJ000683-PA	FBgn0011742	BGIBMGA000082-PA	TcGLEAN_00144	GB13841-PA	NV12348-PA	ACYPI007128-PA	JGI_V11_209234	ISCW000900-PA	WP:CE06111
P55	Contig4486	prohibitin	EOG90S0NS	AAEL009345-PA	AGAP009323-PA	CPIJ007073-PA	FBgn0002031	BGIBMGA006158-PA	TcGLEAN_13727	GB18647-PA	NV10743-PA	ACYPI000080-PA	JGI_V11_306004	ISCW001788-PA	WP:CE26775
P56	Contig2232	sorting nexin 6	EOG9KWK0B	AAEL003758-PA	AGAP004487-PA	CPIJ005342-PA	FBgn0032005	BGIBMGA011186-PA	TcGLEAN_00458	GB13777-PA	NV16639-PA	ACYPI007100-PA	JGI_V11_207004	ISCW000872-PA	WP:CE41070
P57	Contig4158	vesicular mannose-binding lectin	EOG9HX56R	AAEL010584-PA	AGAP004407-PA	CPIJ001653-PA	FBgn0039160	BGIBMGA010839-PA	TcGLEAN_12578	GB14056-PA	NV17786-PA	ACYPI007683-PA	JGI_V11_209797	ISCW012623-PA	WP:CE23946
P58	Contig2574	splicing factor subunit 4	EOG9933RK	AAEL013795-PA	AGAP001538-PA	CPIJ014149-PA	FBgn0015818	BGIBMGA005282-PA	TcGLEAN_03731	GB12347-PA	NV50219-PA	ACYPI000210-PA	JGI_V11_40110	ISCW010601-PA	WP:CE36374
P59	Contig4149	deoxyhypusine synthase	EOG947G0J	AAEL011245-PA	AGAP006531-PA	CPIJ002993-PA	FBgn0035854	BGIBMGA000138-PA	TcGLEAN_02446	GB18573-PA	NV17388-PA	ACYPI001561-PA	JGI_V11_228087	ISCW022659-PA	wP:CE19037
P60	Contig5747	synaptic glycoprotein sc2	EOG9BVS21	AAEL008740-PA	AGAP010714-PA	CPIJ005584-PA	FBgn0035471	BGIBMGA003493-PA	TcGLEAN_01699	GB10614-PA	NV11303-PA	ACYPI001763-PA	JGI_V11_199063	ISCW002618-PA	WP:CE20509
P61	Contig5596	pseudouridylate synthase 1	EOG9K9BKS	AAEL007354-PA	AGAP002409-PA	CPIJ019666-PA	FBgn0038811	BGIBMGA008689-PA	TcGLEAN_04829	GB13048-PA	NV10940-PA	ACYPI009981-PA	JGI_V11_12957	ISCW006672-PA	WP:CE20158
P62	Contig2170	ribosomal protein s3a	EOG92BXGQ	AAEL013158-PA	AGAP003532-PA	CPIJ008584-PA	FBgn0017545	BGIBMGA011416-PA	TcGLEAN_15539	GB16500-PA	NV15382-PA	ACYPI010127-PA	JGI_V11_303071	ISCW008069-PA	WP:CE00664

P63	Contig5775	soluble nsf attachment protein	EOG97PXCR	AAEL011776-PA	AGAP003192-PA	CPIJ013203-PA	FBgn0250791	BGIBMGA010068-PA	TcGLEAN_13571	GB15709-PA	NV18645-PA	ACYPI001188-PA	JGI_V11_188735	ISCW015531-PA	WP:CE38494
P64	PPE012_E14.f	elongation protein 3 homolog	EOG92NJ6N	AAEL006400-PA	AGAP008300-PA	CPIJ004187-PA	FBgn0031604	BGIBMGA007665-PA	TcGLEAN_04103	GB17808-PA	NV13198-PA	ACYPI002840-PA	JGI_V11_303123	ISCW007494-PA	WP:CE40344
P65	Contig3061	branched-chain-amino-acid mitochondrial precursor (placental protein 18)	EOG98SH1J	AAEL007909-PA	AGAP000011-PA	CPIJ015408-PA	FBgn0030482	BGIBMGA001237-PA	TcGLEAN_00403	GB14457-PA	NV15828-PA	ACYPI008372-PA	JGI_V11_300660	ISCW015784-PA	WP:CE03457
P66	Contig1164	t-complex protein 1 subunit theta	EOG9DJK48	AAEL012746-PA	AGAP010588-PA	CPIJ005864-PA	FBgn0033342	BGIBMGA011508-PA	TcGLEAN_01073	GB13033-PA	NV10556-PA	ACYPI002247-PA	JGI_V11_226650	ISCW012934-PA	WP:CE44228
P67	Contig5439	member ras oncogene family	EOG954955	AAEL006091-PB	AGAP011363-PA	CPIJ004642-PA	FBgn0015797	BGIBMGA003597-PA	TcGLEAN_01600	GB11363-PA	NV16743-PA	ACYPI001698-PB	JGI_V11_128059	ISCW010707-PA	WP:CE07541
P68	Contig6177	gtp-binding nuclear protein ran	EOG9M65R2	AAEL009287-PA	AGAP007699-PA	CPIJ008595-PA	FBgn0020255	BGIBMGA006751-PA	TcGLEAN_12793	GB13869-PA	NV14356-PA	ACYPI004024-PA	JGI_V11_305970	ISCW023654-PA	WP:CE16194
P69	Contig503	lon peptidase mitochondrial	EOG9WQ17Z	AAEL006474-PA	AGAP010451-PA	CPIJ008800-PA	FBgn0036892	BGIBMGA007194-PA	TcGLEAN_00570	GB15408-PA	NV13391-PA	ACYPI003409-PA	JGI_V11_306039	ISCW000683-PA	WP:CE16894
P70	ppl-01-A08_064.f	casein kinase alpha 1 polypeptide	EOG9PG67K	AAEL012094-PA	AGAP005569-PA	CPIJ011540-PA	FBgn0000258	BGIBMGA003474-PA	TcGLEAN_10853	GB11423-PA	NP_001153430.1	ACYPI002006-PA	JGI_V11_223674	ISCW007259-PA	WP:CE17321
P71	Contig4571	mitochondrial glutamate	EOG908NFN	AAEL004124-PA	AGAP003208-PA	CPIJ013684-PA	FBgn0039223	BGIBMGA001669-PA	TcGLEAN_14580	GB18315-PA	NV12021-PA	ACYPI000349-PA	JGI_V11_301164	ISCW023734-PA	WP:CE37822
P72	Contig6270	actin related protein 2 3 subunit 41kda	EOG9RXZ5X	AAEL007546-PA	AGAP008908-PA	CPIJ010029-PA	FBgn0001961	BGIBMGA004645-PA	TcGLEAN_00050	GB14874-PA	NV12043-PA	ACYPI007750-PA	JGI_V11_300535	ISCW011129-PA	WP:CE24672
P73	PPE0126_O22.r	proteasome subunit alpha type-3	EOG9DFPV5	AAEL003871-PA	AGAP001973-PA	CPIJ003586-PA	FBgn0023175	NP_001040387.1	TcGLEAN_08015	GB12732-PA	NV18836-PA	ACYPI002276-PA	JGI_V11_306433	ISCW005463-PA	WP:CE36420
P74	Contig4475	electron-transfer-flavoprotein beta polypeptide	EOG96Q70K	AAEL005056-PA	AGAP004653-PA	CPIJ003654-PA	FBgn0039697	BGIBMGA004824-PA	TcGLEAN_08707	GB16132-PA	NV11586-PA	ACYPI007381-PA	JGI_V11_305476	ISCW012010-PA	WP:CE29774

Appendix C - Adjusted protocol for DNA extraction and purification from single mite and Whole Genome Amplification.

C_a - Single mite DNA extraction protocol (modified from E.Z.N.A.® Mag-Bind ®Tissue DNA Kit)

1. Place 1 mite into 0.6 ml PCR tube, add 18 µl Buffer TL. Mince the mite into small fragments.

2. Add **2 µl** of **Proteinase K** at 25 mg/ml solution, shake the tube to mix, and incubate overnight at 55 ° C with gentle shaking or rotation of tubes.

3. Add **20 µl Buffer MSL** and mix by tapping the tube. Incubate at 70 ° C for 30 minutes to 2 hours.

4. Place the sample on bench for 5 minutes to bring the sample to room temperature.

5. Add **26 µl absolute ethanol** followed with **2 µl Mag-Bind magnetic particles**. Mix thoroughly by tapping the tube. Incubate at room temperature for 20 minutes.

6. Place the tube on magnetic separation device for 15-20 minutes.

7. Completely aspirate the cleared supernatant by pipetting.

8. Remove the tube or plate containing the magnetic particles from the magnetic separation device. Add **50 µl of MP Buffer.**

9. Resuspend magnetic particles pellet by tapping the tube (break the completely the pellet). Incubate 15-20 minutes at room temp. Shake few times during incubation.

10. Place the tube on magnetic separation device for 15-20 minutes.

11. Completely aspirate the cleared supernatant by pipetting.

12. Remove the tube from the magnetic separation device. Add **50 µl SPM Buffer**.

13. Resuspend magnetic particles and incubate 10 minutes at room temp. Shake few times during incubation. Place the tube on magnetic separation device for 15-20 minutes.

14. Completely aspirate the cleared supernatant by pipetting.

15. Remove the tube from the magnetic separation device. Add **50 µl SPM Buffer.**

16. Resuspend magnetic particles.

17. Place the tube on the magnetic separation device. Completely aspirate the cleared supernatant by pipetting.

19. Leave the tube to air dry on the magnetic device for 10 minutes or until dry.

20. Remove the tube from magnetic device. Add **5 µl of Elution Buffer to** elute DNA from magnetic particles.

21. Resuspend magnetic particles. Incubate 10-15 minutes (optional at 70 °C). Close tight the lid to avoid evaportion.

22. Place the tube onto a magnetic device for 15-20 minutes. Transfer the cleared supernatant to a new tube.

C_b - Whole genome amplification (WGA) from single mite DNA using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare).

Additional amount (25 U) of phi-29 DNA polymerase (EPICENTRE® Biotechnologies Cat. No. PP040110) is added to the WGA reaction. Reagents cut down to half reaction from original protocol for WGA using single mite extracted DNA.

Two-step single mite WGA protocol:

1- HEAT DENATURATION OF TEMPLATE IN SAMPLE BUFFER:

Mix the whole amount of DNA extracted from single mite (~5 µl) with 4.5 µl of the **Sample Buffer.** Heat at 95 °C for 3 minutes, and inmediately cool on ice to 4 °C.

2- PREPARATION OF AMPLIFICATION REACTION:

Combine 4.5 µl of **Reaction Buffer** with 0.25 µl of **Enzyme Mix** and 0.25 µl of Phi-29 DNA plymerase (0.1 µg/µl (100 U/µl)), on ice. Add the mixture to the cooled sample.

3- Incubate at 30 °C for 4 hours.

4- Heat the sample to 65 °C for 20 minutes to inactivate the enzyme.

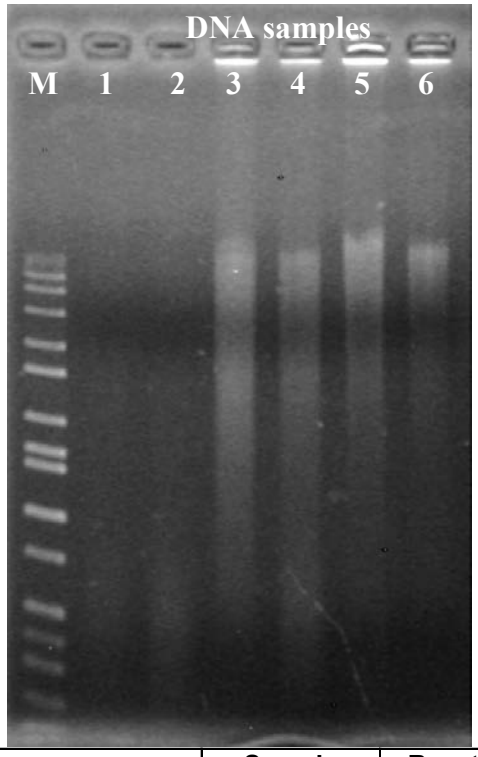
5- Cool the sample to 4 °C or storage at - 20 °C.

Appendix D - Fifty pairs of oligonucleotide sequences derived from EST sequences and used as primer sets for PCR amplification of DNA of the predatory mite *Phytoseiulus persimilis.*

Name	Forward oligo sequence	Reverse oligo sequence
ppM1	5'- TCAGTGACATACAAACCACCTC -3'	5'- GATGCTAGGGAACACTATCAGG -3'
ppM2	5'- TCTTGCCTCATTCTTTCCAAA -3'	5'- GCCTAAGAAAACGGAAAAATACG -3'
ppM3	5'- CGAGAGAAGACGGGATAAGAT -3'	5'- GGAATCGCTGAGAAACTTGA -3'
ppM4	5'- TGATGTGTGTGGTAGAAAGAA -3'	5'- GACTTATACATGGCAACAAGC -3'
ppM5	5'- AATGGCGAGTTGATTAGTAGAA -3'	5'- ACGTGAAGGAGATCGAAAAATA -3'
ppM6	5'- TCGTCTACGTTGACTTTCTTCTTC -3'	5'- TCGAGAAGGGAGACAAGGAT -3'
ppM7	5'- GTCTCAAAGCAAAGTCCATTCG -3'	5'- TACGCTACGATTTCCAA GCT -3'
ppM8	5'- GCGGTTTCTCTTTCTCTGTTTC -3'	5'- ATCTTAGAAGCAATTCCGAACG -3'
ppM9	5'- GAAGGGCATTGGTCGTATTATAG -3'	5'- GAGAGATCGCCAGAAGAAGAT -3'
ppM10	5'- ATGATGACGAAATGAACACTTT -3'	5'- TCCAGGGAATATTTATCGCAT -3'
ppM11	5'- TATCACAGGAACGTCAACTTT -3'	5'- AGG CTCAAAGATGCTGAATAG -3'
ppM12	5'- AGCTTGTCAATCGAACGAATT -3'	5'- GCAAGGAGAAAACAGACCC -3'
ppM13	5'- AGATGGGTTCCTGGCGTAA -3'	5'- CAAACCAGAACACCAGCCAA -3'
ppM14	5'- GAGGGCTCGGCTCAATTAT -3'	5'- ATCGCTTTGCGTTCCTAATG -3'
ppM15	5'- GCAATGCTTTAAACATGGACA -3'	5'- GAAGTCACAAACCAAACTGATT -3'
ppM16	5'- CGGCATCGAGAAAGAGTGTT -3'	5'- GTTGCTCATTCGACTTGTTGTG -3'
ppM17	5'- GGCTAATTTGATGAGTTTCGC -3'	5'- TAGCACATCATTTGGTGGAC -3'
ppM18	5'- CCGTTCGTTACCTTTATTGTGA -3'	5'- GGGAAAAGAACAGACCGAC -3'
ppM19	5'- TTCTTACGCAGAGCAGAA CATTG -3'	5'- GTTTACCACGAGGTTGGCG -3'
ppM20	5'- TCCTAGTTGGTGTTTTTGCAG -3'	5'- ACGCACAACATTGATGTACC -3'
ppM21	5'- CCGACGAGGACAAATTGTT -3'	5'- CAAATTCCAGAGTTCAAGAACG -3'
ppM22	5'- CGGTGTTTCCGTATTTGTGA -3'	5'- GACCCAACCCTCGAAAAAG -3'
ppM23	5'- GATTCACCCCC AGACACACT -3'	5'- ACCCCGCACATTGTGTAAAT -3'
ppM24	5'- GGGGATTCAGGATTTGGACATC -3'	5'- TGCAGCTACCATGACCTCA -3'
ppM25	5'- CGGATTGCATTTCACAGATTCT -3'	5'- CGATGAGGCATTCCAGAGAT -3'
ppM26	5'- GAACATTTATTCGAGAGGTTGG -3'	5'- TCAAGTTGTCGTGTCTCAA -3'
ppM27	5'- AAT TTTCCGTTTCCTTCGTC -3'	5'- CCAAGGAAACTCGAAAAACC -3'
ppM28	5'- TGCTTATCCCAATCCCTCA -3'	5'- GGAGACCCATACGAGAAAAAT -3'
ppM29	5'- GCGAATGTAAACGTTGGAGTC -3'	5'- TCGACGCAGCAAGTTTGGTT -3'
ppM30	5'- ATCTGCTCGTGATGCATGTC -3'	5'- TCTGGCTTACAAGTTCTTCCAT -3'
ppM31	5'- TCAACACTTCTGCGATTCGT -3'	5'- CCAGAGCAGGTGTACAATGAA -3'
ppM32	5'- TTTGCCTGACGAAAGGTGTG -3'	5'- TGCTTCAACAGATTCCAGTTCG -3'
ppM33	5'- CAATACAACTCTTACATGCAGT -3'	5'- TCGTTGTTGTGCATAGGAT -3'
ppM34	5'- AGCTGGGAATGTATTCAGTC -3'	5'- GATTACCTCTACGTGGGTTT -3'
ppM35	5'- ACACCTTTGGGACCGAATT -3'	5'- AGGATAGAAAGATGCCATTCCG -3'

Appendix E - Whole genome amplification product using predatory mites of different life stages and gender (1-6 DNA samples).

Adult female used are in gravid state and is about five times larger than an adult male. Nymph stages are similar to the size of adult male. Sample 3 served as a control for cutting into half of reaction from protocol of Genomi-phi WGA kit. Samples in lanes 4 to 6 were testing to reduce the units of Genomiphi polymerase in half WGA reaction. Supplementation of polymerase from Repli-phi did not result in change of WGA as shown in samples 4-6. M = 2 µl of HI-LO DNA marker/mass ladder (~ 935ng/10ul).



Mite stage	Sample Buffer (µl)	Reaction Buffer (µl)	Genomi-phi polymerase (µl)	Repli-phi polymerase (0.1 µg/µl) (µl)	Amplification
1- Late female nymph	4.5	4.5	-	0.5	**Low**
2- Adult female	4.5	4.5	-	0.5	**Low**
3- Late female nymph	4.5	4.5	0.5	-	**High**
4- Adult female	2.25	2.25	0.25	0.5	**High**
5- Nymph	4.5	4.5	0.25	0.25	**High**
6- Adult male	4.5	4.5	0.25	0.25	**High**

Figure E.1. Agarose gel electrophoresis of single mite WGA products for protocol standarization.

Appendix F - Adult mites from filial generation two (F2) were sexed and collected for genotyping.

The following is a summary of the information gathered from second offspring generation (F2). Eggs from each female were isolated every two days to avoid remating with sons. Total number of eggs was considered to calculate the reproductive capacity (eggs/day) through the survived period of time. Other than accidental death during transfer to new feeding/oviposition vials, females were observed producing eggs for a maximum of twenty days. All F2 mites, sexed in adult stage, were recovered from each family using female or male to backcross to their respective isogenic maternal population. The reproductive capacity of the mites from the backcrosses, estimated per day, is similar between individuals within and between isogenic populations. No significant difference was observed between isogenic populations ($F = 0.16$, $df = 1/3$, $p = 0.92$). And, there was not significant effect of using Esther sex for backcrossing ($F = 0.96$, $df = 1/1$, $p = 0.34$), calculated regardless of the genetic origin. Twenty-three fmales from the backcrossed lineage 2-2-5 was used for genotyping.

Table F.1. Reproductive capacity (eggs/day) of backcrossed pairs (males and females backcrossed) and the sex of offspring (F2 collected) reared to adult for determination.

Isogenic population	Code and sex for backcross	F2 females	F2 males	F2 Collected	Total*	Rep. capacity (eggs/day)
1	1-5-6 female	32	14	46	64	3.37
1	1-5-7 female	33	23	56	73	3.47
1	1-5-2 male	28	18	46	69	3.83
1	1-5-5 male	37	7	44	63	3.65
1	1-5-10 male	28	18	46	71	3.55
2	2-4-1 female	14	6	20	31	3.7
2	2-4-2 female	26	20	46	71	3.94
2	2-4-3 male	35	13	48	64	3.05
2	2-4-5 male	29	11	40	66	3.3
2	2-5-3 female	47	9	56	70	3.5
2	2-5-4 male	8	24	32	49	3.5
2	2-5-5 male	44	8	52	78	3.9
4	4-2-1 female	46	5	51	71	3.9
4	4-2-4 female	10	11	21	28	3.5
4	4-2-2 male	21	10	31	42	3.23
4	4-2-3 male	58	7	65	75	3.75
4	4-2-5 male	21	19	40	58	3.22

* total eggs including not collected for use (mated females, dead individuals, lost, not hatched)

Appendix G - *Phytoseiulus persimilis* foraging gene, *Ppfor,* a fragment of a putative orthologue gene to foraging gene in *Drosophila*, used as molecular marker in the present study.

G_a - *Ppfor* sequence amplified from specific primers consists of approximately 680 bp which contains 2 exons (capital letters) and intronic regions (underlined small caps). Forward and reverse primer sites (20 nucleotides each) are marked in bold and double underlined at each end of the sequence.

Forward primer site

agggtcgtcgccaccatcggcgtgggcggcttcggcaggg*TGAGTAGCCTTTAGGTTTATCTCGGTGGAACT CCGCGTTTCGCTCTTCATTCTGTTGCGGTGTGCTTCCCTCGAGCGCCTGCGCCACGACCCGATGTC ATTGAACTCACGCGAAATTTCCGCCGCTCAG*gttgagctcgtccagcccgtcaacgatccgagccgatcgttcgctctcaaagt catgaagaaagcgcagattgtggaaactcgacagcagcagcacatcatgtccgaaaagcagattctcgtcgagacgaactgtgatttcgttatcaa gctgtacaaaacgttcaaagacacgaagtatctctacatgcttctcgaagcctgcctcggaggagaactctggacgatcctgagggaccgcggga atttcgacgactccacaacgcgcttctacacagcttgtgttctggaggcgttcgattatctgcattctcgcaatatcatttaccgggatttgaagcctgaaa acatgctcctcgaccaacgcggctacatcaag*GTGAGCAGAATCGAAACGTCCCCTCGCTACGGTCCGGTAGGCAT CCGAAGCTTTCTCGGTGAATCATTCCACTGCCGTCGCCCTGAGTGAGTTCCTCGTTTATTTTCCCTAG **Ctcgtcgatttcggcttcgc***

Reverse primer site

G_b -Ppfor marker amplicon aligned to the consensus foraging gene fragment PPFOR, from which primer set was designed.

Poor quality portion of the amplicon was trimmed out (not shown) for sequence alignment. Aligned nucleotide positions are marked by a star below the sequences. Primer par is in bold and double underlined on the EST consensus sequence.

```
PPFOR           TGACGAAGTCCTGAGTAGTCCCCAGCGACCTC**AGGGTCGTCGCCACCATCGG**CGTGGGCG 60
ppfor            ------------------------------------------------------------CCG 3
                                                                              **
PPFOR           GCTTCGGCAGGGTGAGTAGCCTTTAGGTTTATCTCGGTGGAACTCCGCGTTTCGCTCTTC 120
Ppfor           GCCGCCGCTGTGTGTGT------CAGACTTATCTCGCTGGCATTCCGCGGTTCGCTATTC 57
                **  * ** * *** **       **  ******** *** * ****** ****** ***
PPFOR           ATTCTGTTGCGGTGTGCTTCCCTCGAGCGCCTGCGCCACGACCCGATGTCATTGAACTCA 180
Ppfor           ATTCTGTTGCGGTGTGCTTCCCTCGAGCGCCTGCGCCACGACCCGATGTCATTGAACTCA 117
                ************************************************************
PPFOR           CGCGAAATTTCCGCCGCTCAGGTTGAGCTCGTCCAGCCCGTCAACGATCCGAGCCGATCG 240
Ppfor           CGCGAAATTTCCGCCGCTCAGGTTGAGCTCGTCCAGCCCGTCAACGATCCGAGCCGATCG 177
                ************************************************************
PPFOR           TTCGCTCTCAAAGTCATGAAGAAAGCGCAGATTGTGGAAACTCGACAGCAGCAGCACATC 300
Ppfor           TTCGCTCTCAAAGTCATGAAGAAAGCGCAGATTGTGGAAACTCGACAGCAGCAGCACATC 237
                ************************************************************
PPFOR           ATGTCCGAAAAGCAGATTCTCGTCGAGACGAACTGTGATTTCGTTATCAAGCTGTACAAA 360
ppfor           ATGTCCGAAAAGCAGATTCTCGTCGAGACGAACTGTGATTTCGTTATCAAGCTGTACAAA 297
                ************************************************************
PPFOR           ACGTTCAAAGACACGAAGTATCTCTACATGCTTCTCGAAGCCTGCCTCGGAGGAGAACTC 420
Ppfor           ACGTTCAAAGACACGAAGTATCTCTACATGCTTCTCGAAGCCTGCCTCGGAGGAGAACTC 357
                ************************************************************
PPFOR           TGGACGATCCTGAGGGACCGCGGGAATTTCGACGACTCCACAACGCGCTTCTACACAGCT 480
Ppfor           TGGACGATCCTGAGGGACCGCGGGAATTTCGACGACTCCACAACGCGCTTCTACACAGCT 417
                ************************************************************
PPFOR           TGTGTTCTGGAGGCGTTCGATTATCTGCATTCTCGCAATATCATTTACCGGGATTTGAAG 540
Ppfor           TGTGTTCTGGAGGCGTTCGATTATCTGCATTCTCGCAATATCATTTACCGGGATTTGAAG 477
                ************************************************************
PPFOR           CCTGAAAACATGCTCCTCGACCAACGCGGCTACATCAAGGTGAGCAGAATCGAAACGTCC 600
Ppfor           CCTGAAAACATGCTCCTCGACCAACGCGGCTACATCAAGGTGAGCAGAATCGAGACGTCC 537
                ************************************************** ******
PPFOR           CCTCGCTACGGTCCGGTAGGCATCCGAAGCTTTCTCGGTGAATCATTCCACTGCCGTCGC 660
Ppfor           CCTCGCTACGGTCCGGTAGGCATCCGAAGCTTTCTCGGTGAATCATTCCACTGCTGTCGC 597
                ************************************************** *****
PPFOR           CCTGAGTGAGTTCCTCGTTTATTTTCCCTAG**CTCGTCGATTTCGGCTTCGC**GAAGAAGCT 720
Ppfor           CCTGAGTGAGTTCCTCGTTTATTTTCCCTAGCTCGTCGATTTCGGCTTT---------- 646
                ********************************************
```

Appendix H - Sequence alignments of candidate molecular marker sequences (bottom) to their source EST consensus sequence (top) using ClustalW2: ppM11, ppM21, ppM30, ppM34, ppM44, ppM45.

Poor quality portion in either ends of the amplicon was trimmed out (not shown) for sequence alignment. Aligned nucleotide positions are marked by a star below the sequences. Primer par is in bold and double underlined on the EST consensus sequence.

1. **ppM11** fragment sequence:

```
CL533Contig1_M11_      CATACAGAGCGGACTTATTCATAGTTTATTCATCTTCGATCGCGTCGCCTTCAAATCGGT
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      TGGGGCCATTCAAACTGCCTTCAAGCACGACATGTTCGGCTGAGGCCTGACATTGCCGCC
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      ACGGCTATTCCATAACGTTCCTCAATCACAACAACATGTACTCGATTTCCTCAAGACCAA
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      CATACCGACAATCGGCTTGAAAATGTTCARAAGTCTGTTTCTCATCAGCTACAGTCTCGG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      ACTGAGTCCAATAGAGCCTGAATCGTCTCTCACTGTTGCTCATTTTCTCGTCTTTCATTC
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      AACTTGAACGGGGCGTCCAACGAGCTCGGAGGGAGAACATTCAACGTAACAATCGATAAA
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      CTTTGATAAAATGTAAGGGAGCAAATTCCACAGTGGAAACGAATCGTTGTGTGCGAAGTG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      TCCAACTGTGCGACTTGCGGAATGACTATACCGGACCATGCCATTGCTTCCGAGTAGTTT
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      CTGATCGAGGGACGATTCTAGAATCTAGTCCTTGAACTTCCTGTTGGGGTTGGAGCCGAA
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      AAGTTGGTTCCGGATCTGCGGCATCATTTGAAACGAAACCTGTTCAAGCCTTTTCTCCAG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      AGTGTTGCCGACCATGATTTTGCCGTTGCGGGCGTAGAGTTCCACTCCTCCTCCACTCCG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      ATCGCTCAGGAAGTTATCCTTGTCGACCTGAATTGTGCACTTCTTGCCAGCTTTCTGGGT
ppM11                  ------------------------------------------------------------

CL533Contig1_M11_      GAATTGTTGCAAAACCCCGTTTTTCACCTGATCTATCACAGGAACGTCAACTTTGCGGCA
ppM11                  ----------------------------------TTATCACAGGAACGTCAACTTTGCGGCA
                                                          ***************************
```

```
CL533Contig1_M11_      GCGGACGACAACCTT-----------------------------------------------
ppM11                  GCGGACGACAACCTTCTGGAGATGCATGGAACAGATCGTAGATAGGCATTGAGTGTATCG
                       ***************
CL533Contig1_M11       ------------------------------------------TTCTTCGATGAGCTGCAG
ppM11                  AGAGAATTGCACTGCCTGTCAGAAGTCGGATTGAGGACTCACTTCTTCGATGAGCTGCAG
                                                                 ******************
CL533Contig1_M11       AAGCCCCTGCAGGAGGAGTTTCTCCACGAGCCCACGGTACTGATCCGGCTTTTCAGTGAC
ppM11                  AAGCCCCTGCAGGAGGAGTTTCTCCACGAGCCCACGGTACTGATCCGGCTTTTCAGTGAC
                       ************************************************************
CL533Contig1_M11       AGCGCTGAGCTGCTTGCGAGTTTCCTCGAGAACCGTCTCGATGCGATCTTCCTTTGC-TT
ppM11                  AGCGCTGAGCTGCTTGCGAGTTTCCTCGAGAACCGTCTCGATGCGATCTTCCTTTGCAHT
                       ********************************************************  *
CL533Contig1_M11       TCAGAACTCTGAGACGAGAGCTATTCAGCATCTTTGAGCCTTGAATTTTTCTCTGGAGCT
ppM11                  TCAGARCTCTGAGACGAGAGCTATTCAGCATCTTTGAGCCTA------------------
                       ***** ***********************************
CL533Contig1_M11       CAATCTGCTTCTCCCGTTTCTCGTATTCATCGTTGATCTTCTGCCTCGCCTCAGTGACCT
ppM11                  ------------------------------------------------------------

CL533Contig1_M11       GTTGCCCCTTTGAGAGATTGAATTCTTCTTCAGCTTTGGAGTCGATTTCGTCGGCTTTTT
ppM11                  ------------------------------------------------------------

CL533Contig1_M11       CGTGAGCTTCCTGCTCGATGAAAGCTAGCATGTGCTGGATCTGTTTCTGGACGTAGGAGG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11       GATCCGAAGCGGCCATCGTGCAGAGTCTTGGAAACGTTCTCACGCGGGTCGGAAAAAGGG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11       ACTTCGTTCGCTGACGAATCTACTCAAATGAGCTTTACTCCCTTTCTGGGTACCGCTCTG
ppM11                  ------------------------------------------------------------

CL533Contig1_M11       TTGGTGTGTCCCC
ppM11                  -------------
```

2. **ppM21** fragment sequence

```
CL116Contig1_M21       CGTTGGCNCCGTGAGAAAGMGTTTTATTTCCACTTAGCCTACGAAATCAGGCCAATTGCA
ppM21                  ------------------------------------------------------------

CL116Contig1_M21       AAGAATGGACGTCTTTCCGAGAAAACTGTTTACTTGAAACATAATTTATCACCCCATTGA
ppM21                  ------------------------------------------------------------

CL116Contig1_M21       CACGTTAAGCTGTACACTGCAACAAGCCCTAGTCATCGCCCCCCGACTTCGCAGTCTGGA
ppM21                  ------------------------------------------------------------

CL116Contig1_M21       CCTTCTTCACGTTCCAGTTGTGTACACGAGCCATGTTCACTTCTGTGGTTGTGATCTGGT
ppM21                  ------------------------------------------------------------

CL116Contig1_M21       CTCGAAGAAACTCCATGTCCGAAACAACTTCTTTGAGGATCCCTTCCGACGAGGACAAAT
ppM21                  ------------------------------------------------------------

CL116Contig1_M21       TGTTTGAGAGCAGCTTCTCGGCTTCGTCGAGCTCATATTCGAGCATCACGTTAGCTCCGA
ppM21                  ----------------CTCG-CTTCGTCG-GCTC-TATTCGAGCATCACGTTAGCTCCGA
                                       **** ******** **** ************************
CL116Contig1_M21       GCCATAAGCACACTTTTTCCGTGGGAGGGATCCTAGCTTTGGCGAAAACGTCGTCAGAGA
ppM21                  GCCATAAGCACACTTTTTCCGTGGGAGGGATCCTAGCTTTGGCGAAAACGTCGTCAGAGA
                       ************************************************************
CL116Contig1_M21       GAAGAAAAGTTGTATTGATATCTTCTGAAGCT-----------------------------
ppM21                  GAAGAAAAGTTGTATTGATATCTTCTGAAGCTTCCAAGCTTCCTAAGGGTTCAAACACTC
                       ********************************
CL116Contig1_M21       ------------------------------------------------------------
ppM21                  CACACTTTTTCCGTGGGAGGGATCCTAGCTTTGGCGAAAACGTCGTCAGAGAGAAGAAAA

CL116Contig1_M21       ------------------------------------------------------------
ppM21                  GCTGTTATTGGAGGTATGGGCCATAAAATGTCAACTTCAGCTGCATAGATACTTCAGAGT

CL116Contig1_M21       -------------------------------------------TCCTTTTTAGCCTTCAA
ppM21                  TCGGTAGCCTTCTGAAGGAAGCAACTTGCGGGAGCGGGACTCATCCTTTTTAGCCTTCAA
```

```
                                                                  *****************
CL116Contig1_M21      CATCTTCAAAAGACCCAGAGCGTTCTTGAACTCTGGAATTTGAGAGGTGACCTTGGCTTT
ppM21                 CATCTTCAAAAGACCCAGAGCGTTCTTGAACTCTGGAATTTGAG----------------
                      ****************************************
CL116Contig1_M21      CCGTTGCTGTAGGGATGTGTCCATGAAGTTATACTTGCTCAACATCTCTCCGAGGTCTCG
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      CAACGCGTGCTCAGCCGTGGGATGTTGTTCTTTCTTGAGAAACTCCTCAACGTTTTCCAC
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      GAACTCGGCTTTGGGAATCCCTCCATGAGGTTTCCGTGGTTTATTTTCCTCTGACATGTC
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      TCGTGACACTGTGTACCTTCAGGAAGTCTTCAGTCGATGTCACCGACGTCAATTGCCCGA
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      GTGCCCACTTGAGGATCGAGTTTATGAAAGTAATTCAAAGAAACCGTCAGCCGGCAACAG
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      TCAGGCCAGCGCTGTCTACGGGAAGCACAACCAAATAGAGAAGTAGAGGCCGGGCACTCC
ppM21                 ------------------------------------------------------------

CL116Contig1_M21      TCACGCATCAGCATGAAAGCGTCTGGCTGGCTGCACAGTCCCCC
ppM21                 ---------------------------------------------
```

3. **ppM30** fragment sequence

```
CL718Contig1_M30      AGGGACGTATAATTGAGACCAATGCATTGACGTCGCATGGTTCCTCGGTTTTATTCAATG
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GTGACATTAACATAGCACAGAGATTCAAATGCGGTAGGTCGATGCACGGATGGATGAATC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      AACAAAACAATCTGATTCATAGGTCTCCCCTCTGGGATGATGTTTTTATATTCTATGGAT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      AGTAACCGTCATGAAGGATGTTCGCGGTAAGAAAATACTCTCAAGAATCCCCTGACTCCT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GTAGCCATCGTGGGATCACTCCACCGGAAAATGAGATACAAGAACAACGAACTTCATCAC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      CTCGACGAGAGCGTTGGTCGGAGTTTCACTGGTCCTTCTTGACATCGCCAGCCTCAAGAA
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      TCTCGAACTCGCTCTCTTCGGACCCTGTTTTTTCTTCTTTAGACGAGGATGTTGAGGGCT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      TAGATTTTGGAATCGGCTCCTTGGTTTTTTTCTTCGACCCAGAGGCTTTGGCATCACTCT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      TCGACGCTTTCATACCATCGCCACAGAACTCGTAACTGTATAACGGAACGAAGCTCTCGA
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      AGAACCCACAATCTTCCGTCAAATTCGGCAGAATCCATAGTCTGTGCTTGCCAAGCGATA
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      CCACAAAAATCAAGATAAACAAGATTTGTCTGACGACAGCTAAGGAAAGAATGACGCCCA
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GAAATCCTGCCGCAGCCACGCTGAGATAGTAGACCCCGTCACGCATTTTACGTGGCCACA
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GCGGGAACATGCAGATAGCGATTGCAGCGACCACGAACAAGCTGCCGAGCACCCAAGTCT
ppM30                 ------------------------------------------------------------
```

```
CL718Contig1_M30      TCAAGGGGATGGGATCGTACAACCAGACGTACGGCTCGTTCGCATCTACGAAGATCTGCT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      CGTGATGCATGTCCAACTTGAATTTTTTCTTCCCTTTCTCGGCGCGCGGAGTCGCTTTAG
ppM30                 ---------------------------TTTTCTCCTTTCTCGGCGCGCGGA-TCGCTTTAG
                                                 ** *  ****************** *********
CL718Contig1_M30      GGGTAACGTTAGACTCGGTGTCCTCCTCCGCGCCTTCCGGATCCTGCT------------
ppM30                 GGGTAACGTTAGACTCGGTGTCCTCCTCCGCGCCTTCCGGATCCTGCTGAAATGGCACAA
                      *********************************************

CL718Contig1_M30      ------------------------------------------------------------
ppM30                  TTTCTCTCAATGCCTCGGATTAGATGTGAAGTCCACGCAATAATCCTCATGGGGTGAACT

CL718Contig1_M30      ---------------------------------------TTTTCTTAGACTTGACCTCCAT
ppM30                  GGATCAGAGAAGGATTGATTTTCGAGTTCTGAATACCTTTTTCTTAGACTTGACCTCCAT
                                                             **********************

CL718Contig1_M30      CGCGATTTTCCGCGCACGATGGAAGAACTTGTAAGCCAGAAGCTCTTCCATCAGGTACGC
ppM30                 CGCGATTTTCCGCGCACGATGGAAGAACTTGTAAGCCAGAA------------------
                      ***************************************

CL718Contig1_M30      AGCCTGACTTCTCGAATGAATCTCGAGCTTTTCCTTCAATGTCGACTCCATGAGCGAATC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GATCGCTTTCCGCCCGATGAAATAGTTGACCTTGTGTCCCAGGAGCGAGGTGCTCTTCTC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      TGGAAGTTGACTCTTCAGAAACTTAGCCACGTCATAAAGCTTTTTGGGCACCTCTGGCTT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      GGCGGGAGCCTCCACCACATCTCTTTTCTTCGCCATCTGGTAGAAGTCGACTGTCACAGT
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      CGAAAGTTGTGAGTGGTTGGAACTTTATTCGACGGAATGGAAGATGCTACGGCGACAGAC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      AGAAAGAATCACTCGTTCTTCACACGGAGCACTACGACCACAGTAGCAGCACCACACGAC
ppM30                 ------------------------------------------------------------

CL718Contig1_M30      CAGCAGCAAGTCAGCCGAAGTCTGTGCTATCACAACACTCTAGCAACCCCCC
ppM30                 ----------------------------------------------------
```

4. **ppM34** fragment sequence

```
CL60Contig1_M34       GACGTTCAACTTGGGCTTTATTTCGATGGTTCTGAGAATCGTTTCCGCGAGCACGCGTTG
ppM34                 ------------------------------------------------------------

CL60Contig1_M34       CCGTGCTTCCTCGCAGCAGAATAAAAAGCTCATTACACAGGTTCAATGAATTTCATTCCT
ppM34                 ------------------------------------------------------------

CL60Contig1_M34       TCGCAAAAGAGCCACGTCACATATTCGCCTGGGATTCCCAACTTTTGCTCTCATCAAACC
ppM34                 ------------------------------------------------------------

CL60Contig1_M34       ACTGAGAGGCTCATGTTCATGCCTTGATTGGCATGTTTCGATAGATGCGTAGGTGCACTC
ppM34                 ------------------------------------------------------------

CL60Contig1_M34       TCGCAGCTGGGAATGTATTCAGTCATCTTCAATGCAGTGTCCCGTACCGAAATCAGCAGC
ppM34                 ---------------------------GCCGAGAC-GTGTATMGTACCGAAATC-GCAGC
                                                  * *  * ****   *********** *****
CL60Contig1_M34       GGTTATCGGATCGATGAGATATTCCTTGACCCTGACAGATGGAACTTCCCCTTTCTGGGA
ppM34                 GGTTATCGGATCGATGAGATATTCCTTGACCCTGACAGATGGAACTTCCCCTTTCTGGGA
                      ************************************************************

CL60Contig1_M34       TGGGTTAATTTTTTCCCGACCCATCTCAAAGAAACGTTCCTTGAAACCGGCATCGGGGAA
ppM34                 TGGGTTAATTTTTTCCCGACCCATCTCAAAGAAACGTTCCTTGAAACCGGCATCGGGGAA
                      ************************************************************

CL60Contig1_M34       CAGTGGAGCAGTTCTCCTAGAGCATGCTGCAAGCGCTCTTCTCCCGGAAAGGGTTTAATT
ppM34                 CAGTGGAGCAGTTCTCCTAGAGCATGCTGCAAGCGCTCTTCTCCCGGAAAGGGTTTAATT
                      ************************************************************

CL60Contig1_M34       TCTGATTCGGGAAACCCACGTAGAGGTAATCCTCATTTTTACGCTCAGTCACATAGGCCA
ppM34                 TCTGATTCGGGAAACCCACGTAGAGGTAATC----------------------------
```

```
                    ******************************
CL60Contig1_M34_     GAAGATCCTCGATCGCGCGTGACGCCGGAATTCGCTGGTGGTTCGCTTCGCGTCGCAGCT
M34sequen            ------------------------------------------------------------

CL60Contig1_M34     GCTCCACCTGCTGCCTGAGCTGCTGCAGCGACGACATTCTGTGATTCGAATTGCAATCCA
ppM34               ------------------------------------------------------------

CL60Contig1_M34     ATGCGAGGCCGTTGAGGAAACGATTCAGTCGCAACAAGTAGTTGTTTTGTTTCGAACCGT
ppM34               ------------------------------------------------------------

CL60Contig1_M34     CAGACACTCTGTCCCCCGTCTKCTCGCAATGGTGACGGACGAAGGATCCCCC
ppM34               ---------------------------------------------------
```

5. **ppM44** fragment sequence

```
CL731Contig1_M44     GGGGGTAGGATCTTGACTTTATTGGATAATTCGTGAGTTCTTGTTAGTTATCGAGATAAT
ppM44                ------------------------------------------------------------

CL731Contig1_M44     GTATTATGATCGCGTGGTAGTAATTCAGTTTACAAATGCTCACATTCCAGGCGCATACAC
ppM44                ------------------------------------------------------------

CL731Contig1_M44     AGAAATATCCAATTTTGAATCATCAGCCTGGAAGCCGACTTTCAGATGAAGCTTGCACGC
ppM44                ------------------------------------------------------------

CL731Contig1_M44     ATGCACGTTCGATCACACATGGCGGCAGTCGTGATGCGAATGAACCTTGTTCGCTTTTTC
ppM44                ------------------------------------------------------------

CL731Contig1_M44     CATATTCCAATCTTGATCTTCACCAAGTATCAGACCGATTGAAGTGGGCATTTGAGCATG
ppM44                ------------------------------------------------------------

CL731Contig1_M44     AATTTTCAGGATTTAACCTCATCCACTGTTCCGACTCCCCGCAACAGTCGTTTTTCTGGA
ppM44                ----------------------------------------TCGCAACAGTCGTTTTTCTGGA
                                                             **********************
CL731Contig1_M44     TGATCTTCGCATTTCTCACGTGCATGATGCATCGTCGGCTTCGCCCATCGATGTACCGTC
ppM44                TGATCTTCGCATTTCTCACGTGCATGATGCATCGTCGGCTTCGCCCATCGATGTACCGTT
                     ************************************************************

CL731Contig1_M44     CGTTAAAAACCTTCCAGCTCCCCGGAATCGATAAAGGGACCAGTTTTTTTT-ATTAAAAT
ppM44                -GTTAAAAACCTTCCAGCTCCCCGGAATCGATAAAGGGACCAGTTTTTTTTTATTAAAAT
                      ********************************************** ********

CL731Contig1_M44     CAATCAATCGATCCGTAAGTCCCATCATAGACTCGATACATACATCTGGCATATTGTTAG
ppM44                CAATCAATCGATCCGTAAGTCCCATCATAGACTCGATACATACATCTGGCATATTGTTAG
                     ************************************************************

CL731Contig1_M44     TACT--CGTGGGTTTTGTGGGGATATCCCCGTTGTCGGGTGATCCTTTTTTTGGTTGATT
ppM44                TATCTACGTGGGTTTT----------------------------------------------
                     **    **********

CL731Contig1_M44     TTCATAGAAGAATTTACTCATCAGTTACACCACCGCAAGCCGGACGGTAGCAAGAATGAT
ppM44                ------------------------------------------------------------

CL731Contig1_M44     AGTGTGTACAGGGCGGCTCATGAGAAAATTGTATATTGTAATAAGATGCAGCGATGCCAG
ppM44                ------------------------------------------------------------

CL731Contig1_M44     CTGACCAGCTACCCCTGAGGGCGCCATGACCCGAAAGTCAAGAGGTCTTACATCTTCGCA
ppM44                ------------------------------------------------------------

CL731Contig1_M44     CACAGGTTTCTACAAAGAAAACACTTGTCTCTCGGCGTCTGTGCTCGACGTATTAGTTCC
ppM44                ------------------------------------------------------------

CL731Contig1_M44     AGCCGAATTGGAGATAAGCAATACCGATGCTTGTTCCTGAATCCTAAGTATCGCGAGCCT
ppM44                ------------------------------------------------------------

CL731Contig1_M44     AATCTATCGCTCATGTTATGGATTACCACCTGATTGGGTCAGGATCGCAATTACAGAGGA
ppM44                ------------------------------------------------------------

CL731Contig1_M44     GTGGGTTCGAGTAGGATTAACCTCCGTTGAATAGCGGTATACAGGCATTTGGGTTTTGTC
ppM44                ------------------------------------------------------------
```

```
CL731Contig1_M44      CCGAAAATTTCTGGTTTTGTTTCCACCTTCGTTTGGGTGAGAATTCCGATTCCTTGGGGC
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      TCGTTTTGTCGGGCCTTTCTTTCGAAGCCGAAGTAGCCGTATATTATTAGACGTTTGATG
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      AGACAAATAGAGATTTAAGCACGCTCGCCGCGAATTCTACGAGCGAGTTGGATATCCTTG
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      GGCATGATTGTGACACGCTTGGCGTGAATGGCGCACAAGTTGGTGTCCTCGAAGAGTCCG
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      ACCAAGTAGGCCTCTGAAGCCTCCTGCAGAGCGCCGATGGCAGCGCTCTGGAAACGGAGA
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      TCAGTCTTGAAATCCTGGGCGATCTCACGAACGAGACGCTGGAACGGGAGCTTACGAATC
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      AGAAGCTCAGTGGATTTCTGGTAACGACGGATTTCACGAAGGGCGACGGTTCCTGGCCTA
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      TACCTATGAGGTTTCTTGACTCCTCCGGTGGAAGGCGCCGATTTCCTGGCAGCTTTGGTG
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      GCAAGCTGCTTACGGGGAGCTTTTCCTCCAGTAGATTTACGAGCGGTTTGCTTCGTACGG
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      GCCATTGCTCACTTGTTAGTTCGCTCGTTCGGTTAACTTAGTAGGAACACCACACACAGA
ppM44                 ------------------------------------------------------------

CL731Contig1_M44      CTCCTGTACCGTTCGCTGCGGAATCCCCC
ppM44                 -----------------------------
```

6. **ppM45** fragment sequence

```
CL331Contig1_M45       AGAAAATAATCAACACTCTATTCATCTTCCAGCTCTGCCAACTCGTCGTCTTGGAAAAGA
ppM45                  ------------------------------------------------------------

CL331Contig1_M45       GTCTCATCCACGTCAACGTCGTCAGCTCCATCGAATTTGAGATCGCTTTCCGCCAAGGCC
ppM45                  ---------------------------------------------TACTGCGCCGAG-CC
                                                                      **** ** **
CL331Contig1_M45       TTGTCTCTGATAAACATCTGCTTCCCGGTAAGTTTTCCCTTGATCAGGTCCTCCTTGGGA
ppM45                  TTGTCTCTG--AAAAATCTGCTTCCCGGTAAGTTTTCCCTTGATCAGGTCCTCCTTGGGA
                       *********  ***.**********************************************
CL331Contig1_M45       TTGAGCACTCTCATTTCCTCGTCGAACTTCTCCTTCCATTGTGTGAATGATTCGACCGTC
ppM45                  TTGAGCACTCTCATTTCCTCGTCGAACTTCTCCTTCCATTGTGTGAATGATTCGACCGTC
                       ************************************************************
CL331Contig1_M45       ACCTTC------------------------------------------------------
ppM45                  ACCTTCGTACCTGCCGGAGGATCAGAAGCTTACCGAGTGTAAGCCGTTGGCCGCTTCGTC
                       ******
CL331Contig1_M45        ------------------------------------------------GTACCTTCAAGCTT
ppM45                   GTGTTCAGCGTTTATCCAGACACATACACGACCGCCATCGCCTTCACGTACCTTCAAGCTT
                                                                         **************
CL331Contig1_M45       CTTCTTCAGTTCTTCGTCCTCGATCTCGCGTAGTTTCCGCTTCTTGAGTTTCTCTGCTTCC
ppM45                  CTTCTTCAGTTCTTCGTCCTCGATCTCGCGTAGTTTCCGCTTCTTGAGTTTCTCTGCTTCC
                       ************************************************************
CL331Contig1_M45       ATTTCTTTCCGGAGATGTTCTTGCGCTGCGTTTACCAATGTGAAAATCATAACGCTGCCCA
ppM45                  ATTTCTTTCCGGAGATGTTCTTGCGCTGCGTTTACCAATGTGAAAATCATA-CGCTGC---
                       ********************************************** *****
CL331Contig1_M45       CCCAAACGTTGGCCTCGTCCTGCATGGCCTCCAGGAGAGCTTCATCATCGCACTCGATGTC
ppM45                   ------------------------------------------------------------

CL331Contig1_M45       GGGTATGGCATCGGGGTAGGTTTCCGGTAAAGTAAATTTGACAGTCGCCTCCGCGGTTTTC
ppM45                  ------------------------------------------------------------

CL331Contig1_M45       TCTCGGTCATCTTCATCCTCGACTTTGACATCGATCTCGAGAGCCAAGTTCGGAACCTCTT
```

```
ppM45                 ------------------------------------------------------------

CL331Contig1_M45      CGAGAACTCGGAGTTCTGCTGGATAAATAGATTGGAGAGCTTCCAGTTCCTGTGCTTGCTC
ppM45                 ------------------------------------------------------------

CL331Contig1_M45      CTCTTTGTAGTCCATGGCCGTCACTTATCTGTTCCGCAGCTGACTCCTAGTGTTCTCCCGA
ppM45                 ------------------------------------------------------------

CL331Contig1_M45      AAAACGAAGTTGTCTCAGCTGGCACAGAGTCGCTACGGCATCATCTGTTTATCTTGCCGGTC
ppM45                 ------------------------------------------------------------

CL331Contig1_M45      CACGTTTTTCTCATGGCCTGTGGTAAATCGAATGCTTCTTCCACCGCTAAAATCAACACCGC
ppM45                 ------------------------------------------------------------

CL331Contig1_M45      CCCCC
ppM45                 -----
```