BIOINFORMATIC ANALYSIS OF PEA APHID SALIVARY GLAND TRANSCRIPTS

by

MATTHEW STEPHEN AKSAMIT

B.S., Kansas State University, 2012

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Biochemistry and Molecular Biophysics Graduate Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Gerald Reeck

Copyright

Abstract

Pea aphids (*Acyrthosiphon pisum*) are sap-sucking insects that feed on the phloem sap of some plants of the family *Fabaceae* (legumes). Aphids feed on host plants by inserting their stylets between plant cells to feed from phloem sap in sieve elements. Their feeding is of major agronomical importance, as aphids cause hundreds of millions of dollars in crop damage worldwide, annually.

Salivary gland transcripts from plant-fed and diet-fed pea aphids were studied by RNASeq to analyze their expression. Most transcripts had higher expression in plant-fed pea aphids, likely due to the need for saliva protein in the aphid/plant interaction.

Numerous salivary gland transcripts and saliva proteins have been identified in aphids, including a glutathione peroxidase. Glutathione peroxidases are a group of enzymes with the purpose of protecting organisms from oxidative damage. Here, I present a bioinformatic analysis of pea aphid expressed sequence tag libraries that identified four unique glutathione peroxidases in pea aphids. One glutathione peroxidase, ApGPx1 has high expression in the pea aphid salivary gland. Two glutathione peroxidase genes are present in the current annotation of the pea aphid genome. My work indicates that the two genes need to be revised.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my major professor, Dr. Gerald Reeck for all of his support over the last four years. First, I would like to thank him for giving me the opportunity to work for him as an undergraduate research assistant, and for taking me on as a Master's student for the past two years. Without his unending support, advice, and encouragement, I would not be where I am in my academic career.

Special thanks goes out to Dr. Chandrasekar Raman, for the aphid salivary gland dissections. His expertise and ability was crucial in the completion of this project.

I would also like to thank the Integrated Genomics Facility at Kansas State University, specifically Dr. Alina Akhunova, Yanni Lun, and Hanquan Liang. Their Bioanalyzer, cDNA library synthesis, and RNASeq work were critical in my research.

Last, I would like to thank two of my friends and colleagues, James Balthazor and Jarrod Bechard. It has been a pleasure working with you two over the last couple of years. We are all in the same boat as graduate students, and your emotional support and scientific input have greatly impacted me.

Dedication

I would like to dedicate this thesis to my parents, Stephen and Debra Aksamit. Over the years, your love and support has driven me to become the man that I am today. Without you, I would be nothing. You've listened to my problems and offered solutions. You've given me way more than I have needed, or ever deserved. You provided me with the opportunity to go off to college and get the education that I have always craved. I would not be me, without you. You have been my greatest advocates over the years, and you will never know how much I appreciate that. I love you, Mom and Dad.

Second, I would like to dedicate this to my family and friends who have gone on before me, specifically my grandpa, Clarence Blaha, my aunt, Julie Blaha, and my good friend, Randon Regnier.

Grandpa, I miss the days we spent fishing together, watched TV or played games, or just browsed the internet for "useless knowledge". I never knew how much that knowledge would pay off in my life.

Aunt Julie, your persistence and encouragement was one of the main reasons that I joined the Marching Band at Kansas State. Joining that organization was one my biggest life changing experiences, and I have met some of my best friends, and traveled many places because of it.

Randon, though we are not blood relation, you and your brothers have always been family to me, and always will be. Your unconditional love to everyone around you made you a special person, and I aspire to love others like you did.

Ad astra per alas fideles, to the stars on the wings of the faithful ones. You three were always faithful to me, and I cannot thank you enough for that. I carry you in my heart, always.

When you walk through the storm,
Hold your head up high,
And don't be afraid of the dark.
At the end of the road,
There's a golden sky,
And the sweet silver song of a lark.

Walk on through the wind,
Walk on through the rain,
Though your dreams be tossed and blown.
Walk on, walk on,
With hope in your heart,
And you'll never walk alone.
You'll never walk alone.

Rodgers and Hammerstein, 1945

With all of you, I know that I will never walk alone.

Chapter 1 - RNASeq Analysis of Pea Aphid Salivary Gland Transcripts

Introduction

The pea aphid (Acyrthosiphon pisum) is a sap-sucking insect that feeds on the phloem sap of some members of the plant family Fabaceae (legumes). Aphids feed on host plants by inserting their stylets between plant cells to feed from phloem sap in the sieve elements. During feeding, aphids are believed to deliver protein effectors in their saliva into their plant host to circumvent plant defense responses and moderate host cell processes, much as in plant-pathogen interactions. Therefore, aphid feeding is of major agronomical importance, as aphids cause hundreds of millions of dollars in crop damage annually.

Numerous saliva protein components and salivary gland enriched transcripts have been identified in aphids. A proteomics study on green peach aphid (Myzus persicae) saliva identified several saliva protein components, including glucose dehydrogenase, α-amylase, α-glucosidase (sucrase), and glucose oxidase (Harmel et al., 2008). The same authors found glucose oxidase activity by an assay that monitored the liberation of hydrogen peroxide from glucose.

Proteomic analysis of saliva from pea aphids identified nine protein components, including regucalcin, M1-zinc metalloprotease, glucose-methanol-choline oxidoreductase, and a dipeptidyl carboxypeptidase, and five non-annotatable proteins. Based on amino acid sequence, one protein, ACYPI009881 was suggested to be a putative sheath protein that could protect aphid mouthparts from plant defense systems (Carolan et al., 2009).

Protein C002 has been identified as a component of pea aphid saliva, and its transcript is salivary gland enriched. Knockdown of the C002 transcript in pea aphids rapidly decreased

aphid survival. Half of the pea aphids injected with siC002-RNA died at 3 days, in comparison to 11 days for siGFP-RNA injections and uninjected aphids (Mutti et al., 2006). Protein C002 was shown to be crucial to pea aphid feeding. C002 knockout pea aphids were subjected to plant feeding, monitored by electrical penetration graph (EPG). C002 knockouts were found to have very short amounts of contact with phloem sap in sieve elements compared to the control, suggesting that C002 is crucial for aphid foraging and feeding. C002 was also found in plants fed on by aphids, indicating that the protein is transferred from the aphid to the plant during feeding (Mutti et al., 2008).

A functional genomics study of EST libraries for M. persicae identified three candidate effector molecules, notated as Mp10, Mp42, and MpC002. Overexpression of Mp10 in Nicotiana benthamiana by agroinfiltration assays was found to induce chlorosis in the plants. Green peach aphids were subjected to feeding on leaf discs from plants that had transient overexpression of 48 candidate effectors, which three of the effectors changed aphid fecundity. Mp10 and Mp42 reduced aphid fecundity, while MpC002 enhanced aphid fecundity (Bos et al., 2010).

Saliva from the Russian Wheat Aphid (Diuraphis noxia) has also been collected and studied via proteomics. Analysis of saliva by mass spectrometry revealed a zinc-binding dehydrogenase, a protein phosphatase, a RNA helicase, and two unknown proteins. In addition to their proteomics work, phosphatase activity in aphid saliva was confirmed by enzymatic assays (Cooper et al., 2010).

A dual transcriptomics and proteomics approach to predict effector molecules in pea aphid salivary glands identified a large list of transcripts that are supported by EST data (Carolan et al., 2011). Proteomic studies of proteins from salivary glands confirmed the presence of

previously identified proteins from other aphid studies, like M1-zinc metalloprotease, glucose dehydrogenase, and C002, but also revealed other possible effector molecules, including Golgi-associated protein R-1, disulfide isomerases, calreticulin, Armet, and trehalase. The transcriptomics portion of this study was from the Reeck lab, with informatics contributed by Professor Doina Caragea of the Department of Computing and Information Sciences at Kansas State University.

Saliva proteins from the English grain aphid (Sitobion avenae) and the rose-grain aphid (Metopolophium dirhodum) have been probed by proteomics as well (Rao et al.,2013). Since the two aphids in that study do not currently have genomes sequenced, proteins found in MS analysis were searched against the proteome for pea aphids. The analysis revealed 12 identifiable proteins in S. avenae, and 7 proteins in M. dirhodum. Three proteins, two paralogs of glucose dehydrogenase and ACYPI009881 (a putative sheath protein identified in previous aphid studies), were found to be in common between S. avenae, M. dirhodum, and A. pisum. In S. avenae, two additional proteins were found that had not been found in previous studies: carbonic anhydrase II, and a β-galactosidase precursor.

Next generation RNA sequencing (RNASeq), also known as whole transcriptome shotgun sequencing, provides a comprehensive look at mRNA amounts in a given sample. Analysis of RNASeq data for two different conditions, such as plant-fed or diet-fed pea aphids, as in this study, provides a view of differential transcript expression in an organism. Next generation sequencing results have not been reported on aphid salivary gland RNA. Only one RNASeq study of aphid RNA has been published, in which the soybean aphid and its associated endosymbionts were studied (Liu et al., 2012).

Here, I present an RNASeq analysis of the salivary gland transcripts of plant-fed and diet-fed pea aphids. A list of 71 transcripts from the above studies that have been suggested or confirmed as salivary gland enriched transcripts, or secreted proteins of aphid saliva were chosen for RNASeq analysis. Relative transcript abundance was determined in both plant-fed and diet-fed pea aphid salivary glands. Each transcript sequences was analyzed for the presence of a signal sequence for secretion and an ER retention signal, and based on these analyses, I propose a list of possible protein components of saliva (marked as green in the Secreted (Prediction) column of Table 1-2).

Materials and Methods

Salivary Gland Dissection

Salivary glands from pea aphids, *Acyrthosiphon pisum* clone LSR1, were dissected from plant-fed and diet-fed insects by Dr. Chandrasekar Raman.

Plant-reared adult, wingless, asexual pea aphids were collected in sterile Petri dishes directly from faba bean plants, *Vicia faba*. Three sets of 50-70 pea aphids were placed in aliquots of 600µL of RNAlater (Qiagen #76106) and dissected over the course of about 1 h. The dissection area and dissection slides were cleaned with a solution of 0.1% DEPC treated water, followed by application of RNAseZap (Sigma-Aldrich #R2020), and subsequent drying of all surfaces. RNALater (100-200µL) was placed on top of dissection slides. An aphid was placed in the RNAlater on the surface of the slide. The aphid was held by forceps at the abdomen, and the antennae were removed with a small needle. A needle with a bent tip was used to remove the exoskeleton from the head of the aphid, exposing the salivary glands. Both pairs of primary and accessory salivary glands from 120 aphids were removed and placed in RNAse/DNAse-free centrifuge tubes containing 50µL of RNAlater.

Diet-fed aphids were reared on Akey-Beck diet (Table 1-1) (Akey and Beck, 1972) for 48h. Approximately 70 aphids were placed in each feeding apparatus, which consisted of a 1oz. container (Dart #100PC) with a piece of Parafilm thinly stretched over the top of the container. The diet (200µL) was spread over the surface of the Parafilm, and another thin layer of Parafilm was stretched over the top of the diet. The containers were then inverted, and placed on top of a piece of yellow paper, to attract the aphids to the diet. The aphids feed by piercing the first layer of Parafilm with their stylets, and sucking the diet from between the two layers of Parafilm. After 48h, the aphids were removed from the feeding apparatus, and salivary glands were dissected as described above.

Salivary Gland RNA Isolation for RNASeq

Dissection of the salivary glands was followed immediately by RNA isolation. After surface sterilization of the lab bench and instruments with RNaseZap, 100µL of QIAzol reagent (Qiagen #79306) was added to the salivary glands from 120 aphids in 50µL of RNAlater. The glands were homogenized with a pestle attached to a rotating tissue homogenizer for 2-3 min until no tissue remained intact. Addition of 900µL of QIAzol followed, and the samples stood for 3 min at room temperature. After 3 min, 1µL of gDNA Eliminator from an RNeasy Kit (Qiagen #74104) was added to reduce genomic DNA contamination from the aqueous phase during phase separation. Chloroform (200µL) was added to the sample, and vortexed several times to ensure even distribution of reagents in the sample. The sample was stored at room temperature for 10 min, and then centrifuged at 12000xg for 15 min at 4°C. The sample formed two distinct layers, a clear, aqueous layer at the top, and a pink, organic layer on the bottom. The clear layer was pipetted into a new, RNAse/DNAse free centrifuge tube, and the organic layer was discarded. Chilled isopropanol (500µL) was added and the sample was allowed to stand for 10 min at room

temperature to precipitate the RNA. The sample was centrifuged at 12000xg for 15 min at 4°C, forming a pellet of RNA. The liquid was removed, and the RNA pellet was washed twice with 500µL of chilled ethanol. The sample was then air dried at room temperature for approximately 15 min to evaporate any excess ethanol. After drying, the RNA pellet was dissolved in 30µL RNase-Free water (Qiagen #129112). To analyze the RNA sample to determine if RNA quality was suitable for RNASeq by the Illumina Mi-Seq platform, 3µL was removed for Bioanalyzer analysis. The remaining 27µL was used in the generation of the cDNA library.

Thermal Cycler Programs for cDNA Library Preparation

The Integrated Genomics Facility at Kansas State University used the following programs in a thermal cycler for the synthesis of the cDNA library:

mRNA denaturation: 65°C for 5 min; hold at 4°C

mRNA elution 1: 80°C for 2 min; hold at 25°C

Elution 2-Frag-Prime: 94°C for 8 min; hold at 4°C

1st strand: 25°C for 10 min; 42°C for 50 minutes; 70°C for 15 min; hold at 4°C

2nd strand: 16°C for 1h; hold at 16°C

End repair: 30°C for 30 min; hold at 4°C

ATAIL70: 37°C for 30 min; 70°C for 5 min; hold at 4°C

Ligation: 30°C for 10 min

PCR: 98°C for 30 s; (15 cycles of) 98°C for 10 s, 60°C for 30 s, 72°C for 30 s, 72°C for 5 min; hold at 10°C

Purification and Fragmentation of mRNA

RNA isolated from pea aphids was sent to the Kansas State University Integrated Genomics Facility for cDNA synthesis in preparation for RNASeq. The synthesis of the cDNA

library was performed using the TruSeq RNA Sample Preparation Kit (Illumina #RS-122-2001). To the total RNA isolated from plant-fed and diet-fed pea aphids, 50µL of magnetic RNA Purification Beads were added to bind poly-A tails of the mRNA, and mixed. The sample was incubated in the thermal cycler under the program *mRNA denaturation*. Once the sample reached 4°C, it was incubated at room temperature for 5 min. The tube was then placed on a magnetic stand for 5 min to isolate the RNA-bound magnetic beads, and the supernatant was discarded. The beads were washed with 200µL of bead washing buffer. The tube was then placed back on the magnetic tube rack for 5 min. The supernatant was removed from the sample and discarded. To the sample, 50µL of elution buffer was added and mixed, followed by incubation in the thermal cycler under the program *mRNA elution 1*. Once the sample reached 25°C, 50µL of bead binding buffer was added and mixed, followed by incubation at room temperature for 5 min. The sample was placed in the magnetic tube stand for 5 min, and the supernatant was removed and discarded. The beads were washed with 200µL of bead washing buffer, followed by 5 min on the magnetic stand. The supernatant was discarded, and 19.5µL of Elute, Prime, Fragment mix was added and mixed. The sample was incubated in a thermal cycler on the program *Elution-2-Frag-Prime* to elute RNA from the beads. When the sample reached 4°C, it was centrifuged briefly. The sample was placed on a magnetic stand for 5 min, and the supernatant was removed and placed in a fresh PCR tube.

Synthesis of First Strand cDNA

From the supernatant that contained the fragmented and primed mRNA, 17µL was removed and placed in a new PCR tube, and 1st Strand Master Mix (+SuperScriptII) was added and mixed. The tube was incubated in the thermal cycler under the program *1st Strand*. Following incubation, the tube contained single stranded cDNA.

Synthesis of Second Strand cDNA

Once the above sample reached 4°C, 25µL of 2nd Strand Master Mix was added to the single stranded cDNA and mixed. The tube was incubated in the thermal cycler under the program *2nd Strand*. After incubation, the tube contained double stranded cDNA.

Purification of Double Stranded cDNA

Once the double stranded cDNA reached room temperature, it was transfer to a 1.7mL tube, and 90µL of AmpureXP beads were added to ds cDNA and mixed. The sample was incubated at room temperature for 15 min, and then placed on a magnetic stand for 5 min. The supernatant was removed and discarded, and 200µL of 80% ethanol was added to the tube without disturbing the beads, incubated for 30 s, and then the supernatant was discarded. The washing step was repeated one more time. After the tube was dried for 15 min, 62.5µL of Resuspension buffer was added, and the sample was removed from the magnet and mixed. The tube incubated at room temperature for 2 minutes, and then placed on the magnet for 5 min. A 60µL fraction of purified ds cDNA supernatent was then removed and added to a new tube.

End Repair and Reaction Clean Up

To the purified double stranded cDNA, 40µL of end repair mix was added and mixed, followed by incubation in the thermal cycler under the program *End Repair*. The sample was then transferred to a new tube, and 160µL of AmpureXP beads were added and mixed, followed by incubation at room temperature for 15 min. The sample was then placed on the magnetic stand for 5 min, and then the supernatant was discarded. To the tube, 200µL of 80% ethanol was added without disturbing the beads, and incubated for 30 s. The supernatant was removed, and the ethanol wash was repeated once more. The sample was air dried for 15 min, then 20µL of Resuspension buffer was added and mixed. Incubation at room temperature for 2 min followed,

and the sample was placed in the magnetic stand for 5 min. A 17.5µL fraction of the supernatant was removed and placed in a new PCR tube.

Adenylation of 3' Ends, Adapter Ligation, and Reaction Clean Up

To adenlyate the 3' end of the cDNA library, 12.5µL of A-Tailing Mix was added to the sample and mixed. The sample was incubated in a thermal cycler under the program *ATAIL70*. Once the sample reached 4°C, it was removed.

Adapters must be added to the cDNA library so that it may be sequenced on the Illumina Mi-Seq platform. To ligate the sample, 2.5µL of Resuspension buffer and 2.5µL of ligation mix was added to the tube and then mixed. The tube was incubated in a thermal cycler under the program *Ligation*, and was then removed, and 5µL of Stop ligation buffer was added and mixed.

To clean up the reaction mixture, 42µL of AmpureXP beads were added to the sample and mixed, followed by 15 min of incubation at room temperature. The tube was placed on a magnetic stand and the supernatant was removed and discarded. Without disturbing the beads, 200µL of 80% ethanol was added, incubated for 30 s, and the supernatant was removed and discarded. The ethanol wash was repeated one more time, and the sample was dried for 15 min. Following drying, 62.5µL of resuspension buffer was added, mixed, and then the sample incubated at room temperature for 2 min, and placed on a magnetic stand for 5 min. The supernatant was removed and 50µL was transferred to a new 1.7mL centrifuge tube. AmpureXP beads (50µL) were then added for a second clean up. The sample incubated at room temperature for 15 min, then placed on a magnetic stand for 5 min. The sample was washed with 200µL of ethanol, and incubated for 30 s, and the wash was repeated once more. After drying the sample for 15 min, 22.5µL of Resuspension Buffer was added, incubated for 2 min, and then placed on

the magnetic stand for 5 min. A 20µL sample of the supernatant was transferred to a new PCR tube.

Enrichment of DNA Fragments

The purified, 3' polyadenylated, ligand adapted, cDNA library was then enriched by PCR. To the library, 5µL of PCR primer cocktail and 25µL of PCR master mix was added and mixed. The sample was incubated in the thermal cycler under the program *PCR*.

Clean Up of PCR Products

After the tube is removed from the thermal cycler, 50µL of AmpureXP beads were added and mixed, followed by 15 min of incubation at room temperature, and then placed on a magnetic stand. The supernatant is removed and discarded, and the sample is washed with 200µL of 80% ethanol, and incubated for 30 s. The supernatant is removed, and the ethanol wash is repeated. The sample is air dried for 15 min, and 32.5µL of Resuspension Buffer is added and mixed. After 2 min of incubation at room temperature, the sample is placed on a magnetic stand for 5 min and 30µL of the supernatant was transferred to a new 1.7mL centrifuge tube. This product was the final cDNA library that was then sequenced by RNASeq on the Illumina Mi-Seq platform.

Validation of RNASeq Libraries

Before sequencing, our cDNA libraries were verified by an Agilent Bioanalyzer 2100. DNA dye and DNA gel matrix were equilibrated to room temperature for 30 min, and then 25µL of the dye was added to the DNA gel matrix, and the solution was mixed and spun down at 1500xg for 10 min. The Gel-Dye Mix (9µL) was loaded into a specific well denoted as "G" on a DNA 7500 chip on the priming station.

Mapping of RNASeq Reads

RNA-Seq reads were mapping using the Assemble program in the software package Geneious. The mRNA transcripts for the proteins listed in Table 1-2 were used as a "reference genome". Reads were mapped to the reference genome under Medium-Low Sensitivity, which allows 10 gaps per read, requires 18 consecutive bases to be identical to match a read to the genome, and allows a mismatch percentage for single bases of up to 20%.

Calculation of RPKM and Fold Change Values

Reads per kilobase of exon per million reads mapped values (RPKM) were calculated using the following equation, where R_M is reads mapped to a reference sequence, L_T is the length of the reference transcript, and R_T are the total number of RNA-Seq reads.

$$RPKM = 10^9 * (\frac{R_M}{L_T * R_T})$$

Fold change values (FC) were calculated using the following equation, where $RPKM_P$ is the RPKM for plant fed aphids, and $RPKM_D$ is the RPKM for diet fed aphids.

$$FC = \frac{RPKM_P}{RPKM_D}$$

Results/Discussion

Salivary gland RNA from pea aphids reared on faba beans or Akey-Beck diet (composition of diet found in Table 1-1) was isolated and submitted to the Integrated Genomics Facility at Kansas State University (IGF-KSU) for quality analysis by Agilent 2100 Bioanalyzer. The bioanalyzer profiles for plant-fed (Figure 1-1) and Akey-Beck diet fed (Figure 1-2) show good quality RNA, suitable for RNASeq cDNA library synthesis. The cDNA library was prepared by the IGF-KSU using the procedure described above. Bioanalyzer profiles for the sonicated plant-fed and diet-fed salivary gland cDNA libraries used for sequencing appear in

Figure 1-3 and Figure 1-4, respectively. Sequencing of the cDNA libraries at IGF-KSU on the Illumina Mi-Seq platform generated 39,968,294 paired-end reads for the plant-fed cDNA library, and 43,395,650 paired-end reads for the diet-fed library. All reads were 250 bases in length.

The RNA-Seq reads from plant-fed and diet-fed salivary gland libraries were assembled to mRNA transcripts, using each as a "reference genome". I selected predicted or confirmed proteins of pea aphid saliva. After the reads were mapped, RPKM values for each transcript were calculated. The plant-fed to diet-fed RPKM ratios were calculated. The results of these analyses are found in Table 1-2. Log_2 of the fold change is plotted in Figure 1-5.

The majority of the transcripts studied had higher expression in plant-fed salivary gland libraries. The range of the fold change was 1.12 to 4.69. The five highest fold changes were for unannotated proteins. These results were expected, as aphid feeding on plants is more complex than their feeding on artificial diets. Plants have numerous defense mechanisms that protect them from invading pathogens or insect pests, and many pea aphid saliva proteins may help circumvent these systems.

Several of the transcripts that had higher expression in plant-fed pea aphids have been studied as saliva components. Armet, also called MANF, has been found to have both intracellular and extracellular functions in humans. Armet can be localized to the endoplasmic reticulum, where it is believed to be a member of the unfolded protein response (Apostolou et al., 2008). Armet is also found to be secreted into the plant when aphids feed (Cui and Reeck, personal communication). In the ER, Armet is suspected to be a chaperone protein that assists with protein folding (Lee et al, 2003). The Armet transcript is represented at 3.2 fold higher in plant fed salivary gland libraries in comparison to diet-fed libraries.

C002 is another protein known to be secreted during aphid feeding on plants, and is required for pea aphids to successfully feed (Mutti et al., 2006 and Mutti et al, 2008). The transcript for C002 is 2.7 fold higher in plant-fed aphid salivary glands.

A transcript for a putative sheath protein (NM_001162218) had the highest number of reads in both plant-fed and diet-fed salivary gland libraries. This transcript has 2-fold higher expression in plant-fed insects.

Ten transcripts were found to have higher expression in diet-fed pea aphid salivary glands than in plant-fed pea aphid salivary glands. The transcripts expressed at higher levels in diet-fed pea aphid salivary gland libraries are cadherin, a CLIP-domain serine protease, cathepsin-B and cathepsin-L, juvenile hormone binding protein, multi-copper oxidase 1, sucrase, chorin peroxidase H6, maltase, and RNA helicase.

Interestingly, the transcript level of sucrase was found to be nearly 300 times higher in diet-fed salivary glands versus plant-fed salivary glands. The Akey-Beck diet contains 0.35g/mL sucrose (Akey and Beck, 1974). These results suggest that pea aphids can sense diet composition, and that manipulation of the diet can change gene expression in pea aphid salivary glands. To further explore this hypothesis, it would be interesting to modify the Akey-Beck diet by replacing sucrose with other sugars, such as glucose or trehalose, which may stimulate glucose dehydrogenase or trehalase transcript expression.

The encoded amino acid sequences for all studied transcripts were analyzed to determine the presence of a signal secretion peptide or ER retention signal. The sequences were processed through the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP-3.0/) , and the probabilities that a signal peptide existed were calculated through the Hidden Markov model within the program, the results of which can be seen in Table 1-2. All but three encoded proteins

were predicted to have signal secretion peptide. The lack of a signal secretion peptide in an M1 zinc metalloprotease, unannotated protein 27, and β-galactosidase suggests that these three proteins should not be included in a list of pea aphid saliva components. ER retention signals (KDEL, KEEL, HDEL, KHEL, KEDK, HTEL) were found in nine of the transcripts: two forms of calreticulin (ACYPI002622, ACYPI007677), four forms of protein disulfide isomerases (ACYPI009755, ACYPI005594, ACYPI008926, ACYPI000119), unannotated protein 7 (ACYPI001271), and a protein annotated as Contig_37 (NM_001126135) (Table 1-2). Armet was also found to have an ER retention signal, however, in humans it is known to be both intracellular (localized in the ER as a member of the unfolded protein response), and extracellular. Unpublished studies have shown that Armet is secreted into plants during pea aphid feeding (G.R. Reeck and F. Cui, personal communication). The presence of an ER retention signal in a transcript indicates their expressed proteins may be critical to the salivary gland during feeding; however, they might not be components of pea aphid saliva.

Of the 71 salivary gland transcripts studied, I identify 61as likely components of aphid saliva (marked green in Secreted (Prediction) column in Table 1-2). To further confirm the presence of these proteins in pea aphid saliva, I suggest a top-down proteomics approach. In this approach, saliva secreted into Akey-Beck diet would be collected, and the saliva proteins would be separated from the diet by HPLC through a protein binding C18 column. After separation, column fractions would be analyzed by MALDI-TOF, or ion mobility spectrometry. This study differs from all previous mass spectroscopy approaches, which have all been bottom-up proteomics, in which proteins are isolated by electrophoresis, digested into small fragments by a protease, and identified by tandem mass spectrometry and peptide mass fingerprinting (Carolan et al., 2009, Carolan et al., 2011, Cooper et al., 2010, and Rao et al., 2013). The top-down

approach is recommended because separation of saliva proteins by gel electrophoresis followed by trypsin digestion decreases the sensitivity to detect protein components in saliva. Only proteins that showed visible bands in a gel are subjected to mass spectrometry analysis, so proteins of low concentration could be easily missed.

Figures and Tables

Table 1-1 Composition of the Artificial Diet (Akey and Beck, 1972)

Pea aphids were reared for 48 hours on an artificial diet, referred to as Akey-Beck diet throughout this thesis. The composition of this diet appears below (Akey and Beck, 1972).

1)Amino acids and vitamins 2) Trace metals as chloride salts	From Akey and Beck, 1971	µg/100 ml of diet
	Cu^{2+}	120
	Fe^{3+}	920
	Mn^{2+}	220
	Na^{+}	1000
	Zn^{2+}	400
3) Other		Amount/100mL diet
	Calcium citrate	10 mg
	Cholesterol benzoate	2.5mg
	Magnesium sulfate ($7H_2O$)	242mg
	Potassium phosphate monobasic	250mg
	Sucrose	35g
	pH 7.5, adjusted with KOH	
	Distilled-deionized water to make 100mL of diet	

Table 1-2 Comparative Analysis of Plant-Fed and Diet-Fed Pea Aphid Salivary Gland Libraries by RNASeq

Reads generated by RNASeq were mapped to each individual transcript open reading frame as a "reference genome". RPKM, fold changes, and $\log_2$(fold change) values were calculated as described in the text. Signal peptides and their probabilities were calculated using SignalP 3.0. Transcript names are highlighted in different colors: purple: pea aphid transcripts corresponding to transcripts studied in Russian Wheat aphids (Cui et al., 2012), blue (with white text): transcripts of proteins identified in (Bos et al., 2010), red: pea aphid transcripts of proteins identified in green peach aphid (Harmel et al., 2008), dark green: pea aphid transcripts of proteins identified in Russian Wheat aphids, blue (with black text): pea aphid transcripts of proteins identified in English grain aphid, rose grain aphid, and pea aphid (Rao et al., 2013), light green: pea aphid salivary gland enriched transcripts (Carolan et al., 2011).

Accession	Transcript Name	Plant Fed Reads Mapped	Plant Fed RPKM	Diet Fed Reads	Diet Fed RPKM	Transcript Length	Fold Change	Signal Peptide	Probability	ER Retention Signal	Secreted (Prediction)	$\log_2$(Fold Change)
ACYPI008617	C002	320652	12155.5	130025	4539.8	660	2.68	Yes	0.82	No	Yes	1.42
ACYPI000733	Dipeptidyl Carboxypeptidase	182432	2392.3	77692	938.3	1908	2.55	Yes	0.962	No	Yes	1.35
ACYPI008001	Armet	4729	225.4	1608	70.6	525	3.19	Yes	0.99	Yes (KEEL)/No	Yes	1.67
ACYPI002891	Cadherin	750	9.3	3370	38.4	2022	0.24	Yes	0.957	No	Yes	-2.05
ACYPI002622	Calreticulin	43878	894.7	23101	433.9	1227	2.06	Yes	0.951	Yes (HDEL)	No	1.04
ACYPI007677	Calreticulin	23518	480.7	10449	196.7	1224	2.44	Yes	0.971	Yes (HDEL)	No	1.29
ACYPI008370	CLIP-domain serine protease	1531	16.3	6597	64.7	2349	0.25	Yes	0.924	No	Yes	-1.99
ACYPI009755	Disulfide isomerase	51048	836.4	19733	297.8	1527	2.81	Yes	0.888	Yes (KDEL)	No	1.49
ACYPI005594	Disulfide isomerase	11282	191.6	4773	74.7	1473	2.56	Yes	0.693	Yes (KHEL)	No	1.36
ACYPI008926	Disulfide isomerase	8670	166.2	4916	86.8	1305	1.91	Yes	0.926	Yes (KEEL)	No	0.94
ACYPI000119	Disulfide isomerase	12605	256.4	4644	87	1230	2.95	Yes	998	Yes (KEEL)	No	1.56
ACYPI000288	Glucose Dehydrogenase	154141	1785.5	81892	873.6	2160	2.04	Yes	0.968	No	Yes	1.03
ACYPI000986	Glucose Dehydrogenase	128339	932.4	70736	473.3	3444	1.97	Yes	0.883	No	Yes	0.98
ACYPI002476	Inositol Monophosphatase	429	11.6	250	6.2	927	1.87	Yes	0.95	No	Yes	0.90
ACYPI004198	Lipophorin precursor	312028	892.4	222263	585.5	8748	1.52	Yes	0.999	No	Yes	0.61
ACYPI009427	M1 zinc metalloprotease	34370	510.9	12528	171.5	1683	2.98	Yes	0.992	No	Yes	1.57
ACYPI002258	M1 zinc metalloprotease	13408	121	7669	63.7	2772	1.90	No	0.199	No	No	0.93
ACYPI003917	Golgi-Associated Protein R-1	266027	3838.5	114806	1525.7	1734	2.52	Yes	0.995	No	Yes	1.33
ACYPI002298	Trehalase	10880	149.5	5944	75.2	1821	1.99	Yes	0.968	No	Yes	0.99
ACYPI45001	Unannotated Protein 10	613137	38449.9	220577	12739.2	399	3.02	Yes	0.969	No	Yes	1.59
ACYPI000852	Unannotated Protein 11	17771	1001.4	5682	294.9	444	3.40	Yes	0.94	No	Yes	1.76
ACYPI55147	Unannotated Protein 12	190375	2945.7	77573	1105.5	1617	2.66	Yes	0.725	No	Yes	1.41
ACYPI001541	Unannotated Protein 13	87733	4461.5	22295	1044.2	492	4.27	Yes	0.996	No	Yes	2.10
ACYPI001606	Unannotated Protein 14	85995	3131.8	36573	1226.7	687	2.55	Yes	0.905	No	Yes	1.35
ACYPI001719	Unannotated Protein 15	198860	6052.9	93344	2616.8	822	2.31	Yes	0.991	No	Yes	1.21
ACYPI003601	Unannotated Protein 16	21741	197.1	16289	136	2760	1.45	Yes	0.595	No	Yes	0.54
ACYPI008667	Unannotated Protein 17	288654	19895.5	113870	7228.6	363	2.75	Yes	0.82	No	Yes	1.46
ACYPI002172	Unannotated Protein 18	174780	5521.4	53810	1565.6	792	3.53	Yes	0.997	No	Yes	1.82
ACYPI001099	Unannotated Protein 19	40836	1968.6	11174	496.1	519	3.97	Yes	0.976	No	Yes	1.99
ACYPI009881	Putative Sheath Protein	2116651	44465.4	1150729	22264.6	1191	2.00	Yes	0.992	No	Yes	1.00
ACYPI56502	Unannotated Protein 20	315882	20581.5	122407	7345.6	384	2.80	Yes	0.824	No	Yes	1.49
ACYPI000558	Unannotated Protein 21	62894	549.8	21396	172.3	2862	3.19	Yes	0.998	No	Yes	1.67
ACYPI009919	Unannotated Protein 22	16632	983.8	6876	374.6	423	2.63	Yes	0.997	No	Yes	1.39
ACYPI001843	Unannotated Protein 23	52413	543	18802	179.4	2415	3.03	Yes	0.999	No	Yes	1.60
ACYPI005818	Unannotated Protein 24	45444	3644.2	26653	1968.5	312	1.85	Yes	0.833	No	Yes	0.89
ACYPI003695	Unannotated Protein 25	16868	1153.1	3902	245.7	366	4.69	Yes	0.96	No	Yes	2.23
ACYPI001887	Unannotated Protein 26	15276	343.1	9986	206.6	1114	1.66	Yes	0.999	No	Yes	0.73

Accession	Transcript Name	Plant Fed Reads Mapped	Plant Fed RPKM	Diet Fed Reads	Diet Fed RPKM	Transcript Length	Fold Change	Signal Peptide	Probability	ER Retention Signal	Secreted (Prediction)	$\log_2$(Fold Change)
ACYPI55148	Unannotated Protein 27	169772	2792.7	65465	991.8	1521	2.82	No	0.407	No	No	1.49
ACYPI43360	Unannotated Protein 28	27433	1694.7	12398	705.4	405	2.40	Yes	0.964	No	Yes	1.26
ACYPI007553	Unannotated Protein 29	92469	4509.9	26822	1204.8	513	3.74	Yes	0.962	No	Yes	1.90
ACYPI000472	Unannotated Protein 3	371047	8454.9	146221	3068.7	1098	2.76	Yes	0.999	No	Yes	1.46
ACYPI001152	Unannotated Protein 30	66726	927.5	58826	753.1	1800	1.23	Yes	0.938	No	Yes	0.30
ACYPI38795	Unannotated Protein 31	15	0.387	8	0.19	969	2.04	Yes	0.975	No	Yes	1.03
ACYPI008224	Me10	369697	19891.9	165989	8225.8	465	2.42	Yes	0.95	No	Yes	1.27
ACYPI000490	Unannotated Protein 5	430231	13747.5	168688	4964.5	783	2.77	Yes	0.978	No	Yes	1.47
ACYPI006346	Unannotated Protein 6	425616	24997.3	190421	10300.5	426	2.43	Yes	0.828	Maybe (RELL)	Maybe	1.28
ACYPI001271	Unannotated Protein 7	98092	3805	30529	1090.7	645	3.49	Yes	0.988	Yes (KEDK)	No	1.80
ACYPI39568	Unannotated Protein 8 (Guo, 2014)	613137	38449.9	220577	12739.2	399	3.02	Yes	0.969	No	Yes	1.59
ACYPI007406	Unannotated Protein 9	268723	11040.1	99378	3760.3	609	2.94	Yes	0.983	No	Yes	1.55
ACYPI23752	Carbonic anhydrase II	18553	559.3	9318	258.7	830	2.16	Yes	0.522	No	Yes	1.11
ACYPI007650	Beta-galactosidase precursor	71	0.9	37	0.4	1923	2.25	No	0.206	No	No	1.17
XM_003240454	Cathepsin B	82453	2022.5	159438	3602	1020	0.56	Yes	0.713	No	Yes	-0.83
NM_001163097	Cathepsin L	8674	211.5	11503	258.3	1026	0.82	Yes	0.747	No	Yes	-0.29
NM_001204960	Juvenile Hormone Binding Protein	5032	165.9	7407	224.9	759	0.74	Yes	0.991	No	Yes	-0.44
XM_003241838	MCO1	167	1.9	424	4.5	2154	0.42	Yes	0.988	No	Yes	-1.24
XM_003240062	Endoribonuclease	955	4.9	752	3.5	4893	1.40	Yes	1	No	Yes	0.49
NM_001126135	Sucrase	36	0.5	10684	138.9	1773	0.004	Yes	0.909	No	Yes	-8.12
NM_001162218	Contig_37	6537	245.6	2610	90.3	666	2.72	Yes	0.981	Yes (HTEL)	No	1.44
BAH72296	Sheath Protein	2131828	15300.6	1168453	7723.9	3486	1.98	Yes	0.665	No	Yes	0.99
XM_001944729	Peptidase	93214	845.9	35254	294.6	2757	2.87	Yes	0.97	No	Yes	1.52
XM_001947380	Chorin Peroxidase H6	237	3.3	885	11.4	1794	0.29	Yes	1	No	Yes	-1.79
XM_001950168	Transmembrane 87B-like	1244	19	565	8	1635	2.38	Yes	0.967	No	Yes	1.25
NM_001135912	Dipeptidyl Carboxypeptidase	161782	2114.8	65347	786.8	1914	2.69	Yes	0.969	No	Yes	1.43
NM_001246005	EMP24 like	3988	127.9	2088	61.6	780	2.08	Yes	0.906	No	Yes	1.05
XP_001948510	Mp42	17169	1060.6	5977	340.1	405	3.12	Yes	0.98	No	Yes	1.64
XM_003246839	Maltase-L	16	0.2	2891	35.5	1875	0.01	Yes	0.95	No	Yes	-7.47
NM_001246102	3-hydroxyacyl-CoA dehydrogenase type-2	193	6	168	4.8	810	1.25	Yes	0.997	No	Yes	0.32
XM_001948922	Unannotated Protein 32	875	19	850	17	1152	1.12	Yes	0.989	No	Yes	0.16
XM_001948134	Zinc binding dehydrogenase	840	16	524	9.2	1311	1.74	Yes	0.999	No	Yes	0.80
XM_001947715	PAMP	453	11.3	376	8.6	1005	1.31	Yes	0.996	No	Yes	0.39
XM_001945136	RNA Helicase	698	8.1	1214	13	2151	0.62	Yes	0.999	No	Yes	-0.68
NM_001162003	ApGPx1	37577	1339.2	51847	1702	702	0.79	Yes	0.905	No	Yes	-0.35

Figure 1-1 Bioanalyzer Profile of Plant-Fed Pea Aphid Salivary Gland RNA

RNA isolated from pea aphid salivary glands reared on faba bean plants was analyzed on an Agilent 2100 Bioanalyzer for determination of RNA quality.

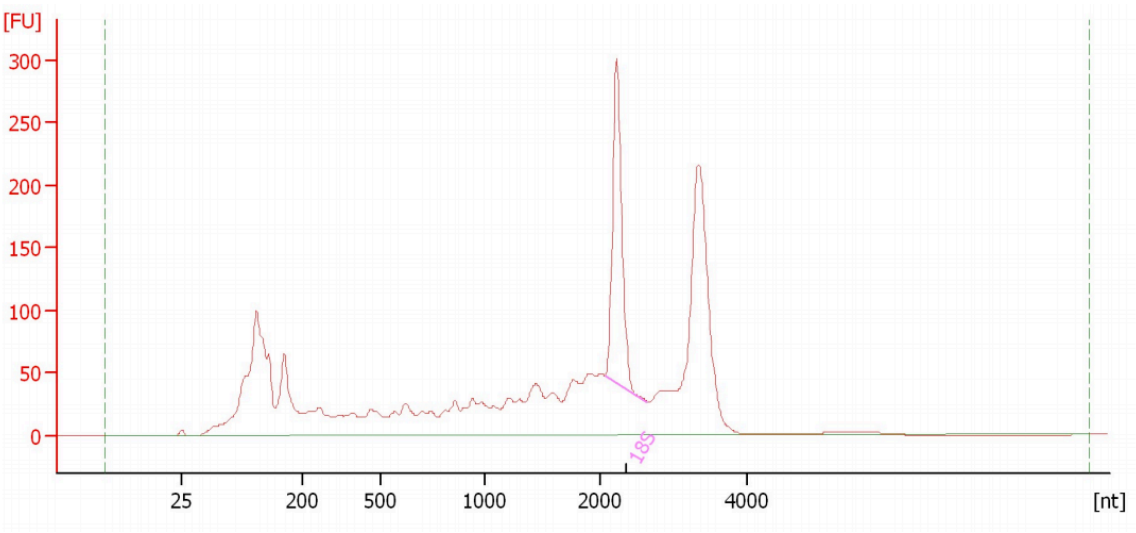


Figure 1-2 Bioanalyzer Analysis of Diet-Fed Pea Aphid Salivary Gland RNA

RNA isolated from pea aphid salivary glands reared on Akey-Beck diet for 48 h was analyzed on an Agilent 2100 Bioanalyzer for determination of RNA quality.

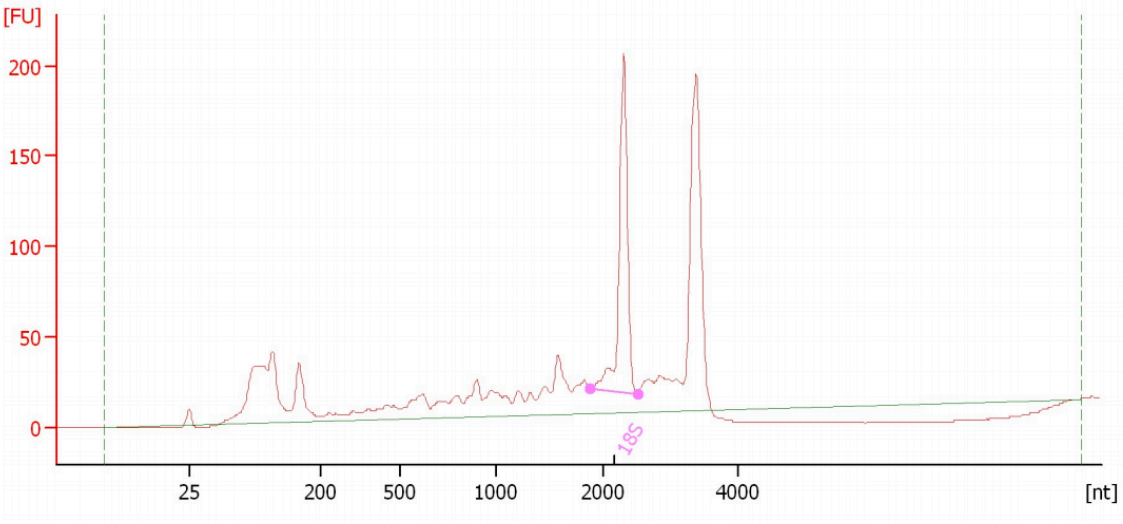


Figure 1-3 Bioanalyzer Profile of Plant-Fed Pea Aphid Salivary Gland cDNA Library

The cDNA library generated from salivary gland RNA isolated from pea aphids reared on faba beans was submitted for Bioanalyzer analysis.

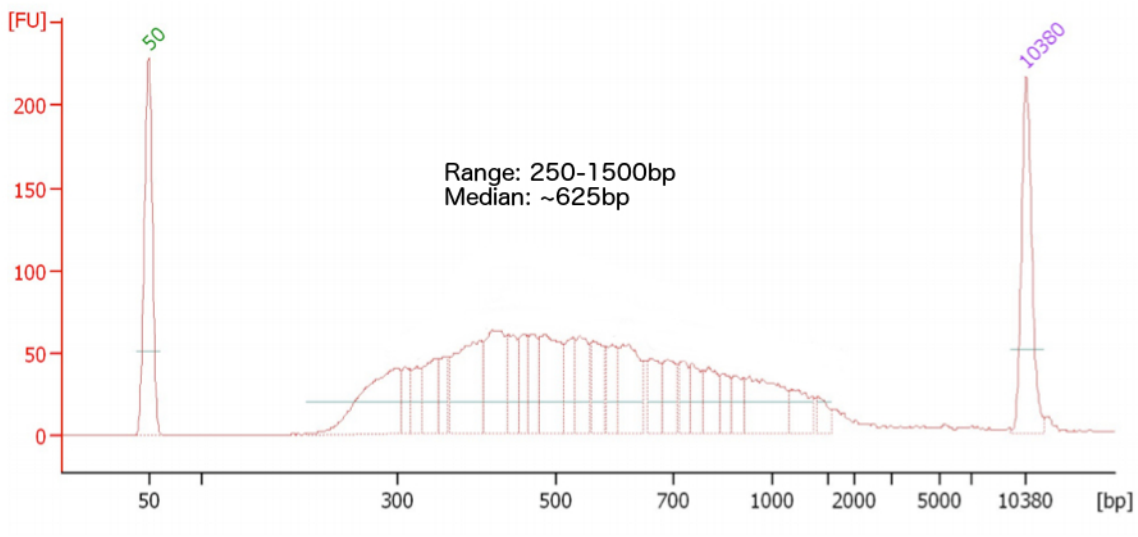


Figure 1-4 Bioanalyzer Profile of Diet-Fed Pea Aphid Salivary Gland cDNA Library

The cDNA library generated from salivary gland RNA isolated from pea aphids reared on Akey-Beck diet for 48 hours was submitted for Bioanalyzer analysis.

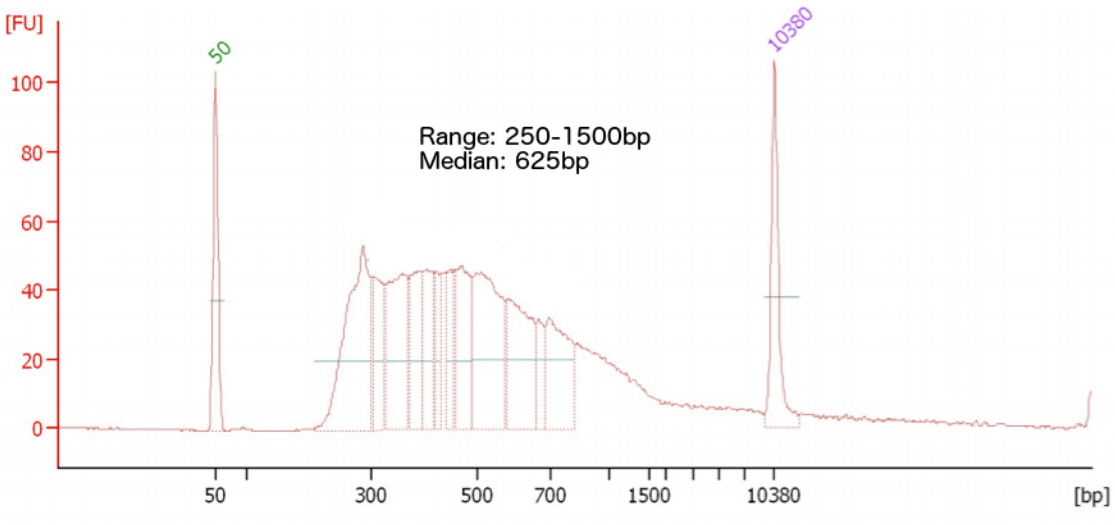
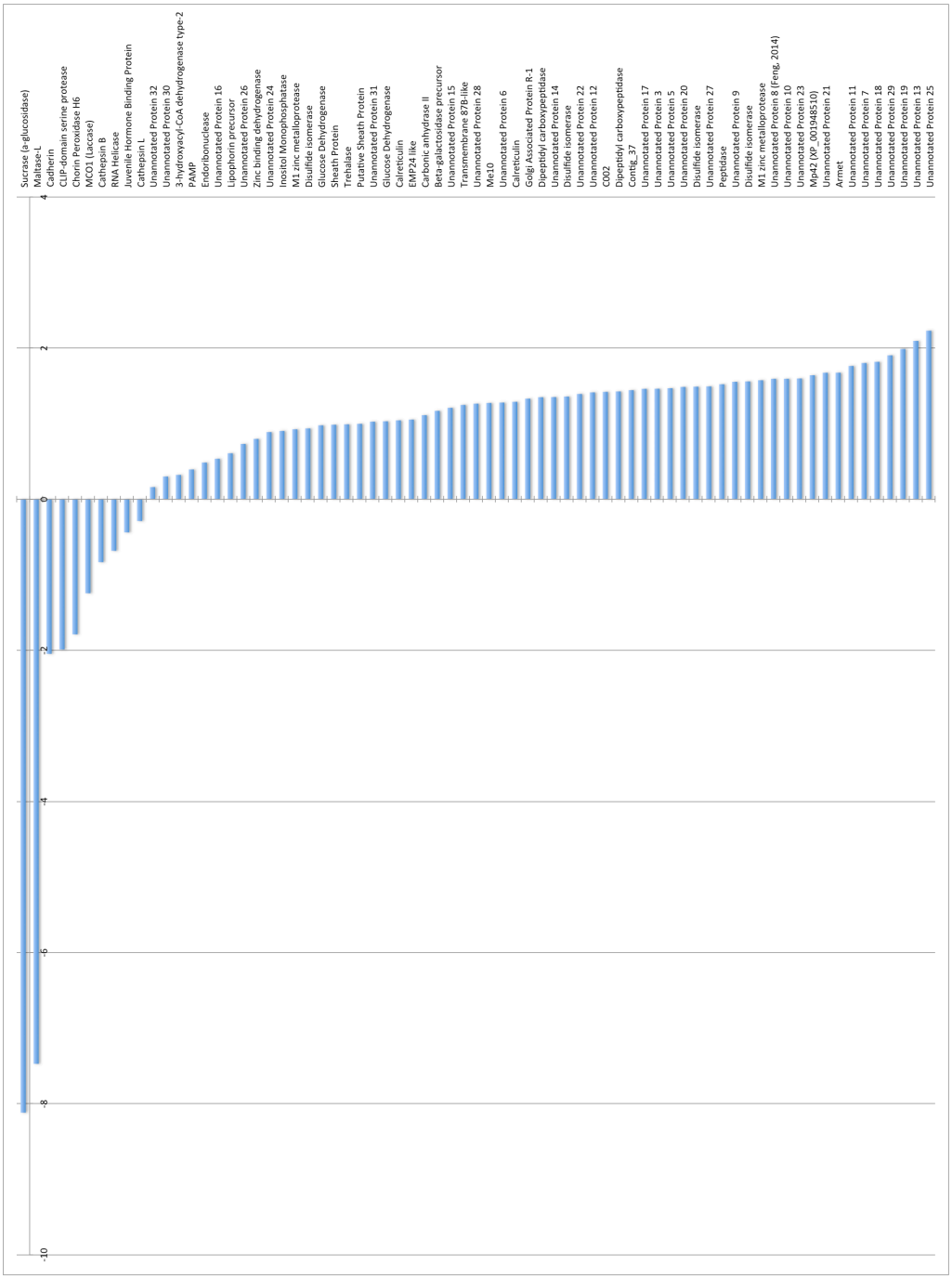


Figure 1-5 Log_2(Fold Change) Plot of Pea Aphid Salivary Gland Transcripts

Log_2 of the fold change (plant-fed RPKM/diet-fed RPKM) was plotted below. Positive values are transcripts of higher expression in plant-fed pea aphid salivary gland libraries, while negative values are expressed at higher levels in diet-fed pea aphid salivary gland libraries.



Chapter 2 - Bioinformatics Analysis of Glutathione Peroxidase

Introduction

Glutathione peroxidases (GPxs) protect organisms from oxidative damage (Brigelius-Flohe et al., 2013). In *Homo sapiens,* eight isozymes of glutathione peroxidase have been discovered, five of which contain selenium (Table 2-1). Glutathione peroxidases are identified in part by a conserved catalytic tetrad consisting of selenocysteine (or cysteine), glutamine, tryptophan, and asparagine. In human GPx1, these are residues U49, Q84, W162, and N163 (Tosatto et al., 2008).

The various human GPx isozymes have different tissue or organ (and intracellular) locations, as well as substrate specificities. In humans, GPx1 is the most abundant isoform (Forgione et al., 2002), located in almost all tissues, and in the cytoplasm, serving to break down hydrogen peroxide. GPx2 is localized to the gastrointestinal system (Chu et al., 1993). GPx4 breaks down phospholipid hydroperoxides into alcohols (Yant et al., 2003), while GPx3 is found to be abundant in plasma (Olson et al., 2010).

The prototypical reaction catalyzed by glutathione peroxidase is as follows, where GSH is monomeric glutathione, GS-SG is oxidized glutathione disulfide:

$$H_2O_2 + 2GSH \rightarrow GS - SG + 2H_2O$$

In the case of GPx4, hydrogen peroxide is replaced by a lipid hydroperoxidase (LHP), and a lipid alcohol is formed as a product:

$$LHP + 2GSH \rightarrow GS - SG + lipid\ alcohol$$

The catalytic mechanism for selenium-containing GPxs has three steps (Prabhakar et al., 2005). In the first step (1), the active GPx with the selenocysteine in selenol form reacts with

hydrogen peroxide to create the selenenic acid form of GPx (E-SeOH) and water, reducing hydrogen peroxide in the process, the crucial step in limiting oxidative damage by hydrogen peroxide.

(1) $(E - SeH) + H_2O_2 \rightarrow (E - SeOH) + H_2O$

Next (2), a molecule of glutathione forms a selenenic acid-glutathione molecule (E-Se-SG), again creating water as a product.

(2) $(E - SeOH) + GSH \rightarrow (E - Se - SG) + H_2O$

In the final step (3), active GPx (E-SeH) is regenerated when GSH is oxidized to form the glutathione disulfide (GS-SG) by glutathione peroxidase.

(3) $(E - Se - SG) + GSH \rightarrow (E - SeH) + GS - SG$

NADPH can reduce the glutathione disulfide in a reaction catalyzed by glutathione reductase to regenerate monomeric glutathione.

Non-selenium dependent GPxs act by a similar mechanism, except that selenocysteine in the active site is replaced by cysteine, and the sulfur acts as the nucleophile instead of selenium.

In silico studies of EST and genomic libraries of invertebrates revealed genes similar to phospholipid glutathione peroxidase (human GPx4) in all genomes surveyed (Bae et al., 2009). In nematodes and platyhelminths, genes similar to human GPx3 and GPx7 were discovered as well. Selenium dependent GPxs were exclusively found in nematode and platyhelminth species. The analysis of Bae et al. (2009) suggested that selection pressure to conserve the selenocysteine codon in glutathione peroxidases seemed to be relaxed during the evolution of the gene.

In *Drosophila melanogaster*, five glutathione peroxidases have been identified at the transcript level (Peroxibase, http://peroxibase.toulouse.infra.fr). However, one *Drosophila* GPx family member is known to have thioredoxin reductase activity, rather than GPx activity

(Missirlis et al., 2005). Therefore, it is critical to note that while putative transcripts may be classified as a member of the glutathione peroxidase family on the basis of their conserved residues, studies at the protein level are required to determine substrate specificites.

A dual proteomics and transcriptomics approach has shown a salivary gland enriched transcript encoding a glutathione peroxidase (referred to as ApGPx1 throughout this thesis) (AphidBase #ACYPI002439-PA). Mass spectrometry analysis of pea aphid salivary gland proteins confirmed the presence of this GPx (Carolan et al., 2011).

Here, I present a bioinformatics analysis that identified three pea aphid glutathione peroxidases. EST evidence supports the presence of all three GPxs, and in conjunction with RNASeq analysis, indicates high production of ApGPx1 in the salivary gland. While current annotations of the pea aphid genome indicated two ApGPx genes, my analysis of pea aphid EST libraries indicate that the annotations should be revised.

Materials and Methods

All materials and methods are as described previously in Chapter 1 of this thesis.

Results

Top Pea Aphid EST Hits to the Eight Human Glutathione Peroxidases Reveal Two Distinct Pea Aphid Glutathione Peroxidase Candidates

The amino acid sequences of the eight *H. sapiens* GPxs (Table 2-1) were run individually, as query sequences, through the NCBI BLAST program tblastn under default parameters, restricting searches to expressed sequence tag libraries of *A. pisum* (taxid: 7029). The top hit for each protein query is shown in Table 2-2. Each encoded the characteristic GPx catalytic tetrad. Multiple sequence alignment of the encoded open reading frame of the hits apparantly revealed three distinct glutathione peroxidase sequences, one best matching human

GPx 2 and 4, one best matching human GPx3, and the last best matching human GPx1, GPx5, GPx7, and GPx8. However, the hits for GPx1, GPx5, GPx7, and GPx8 were incomplete ESTs matching in sequence for hits to GPx2, GPx4, and GPx6. The top hits for human GPx6 and GPx7 were omitted because they matched the top hits for human GPx2 and GPx1, respectively. All of the top hits contained a region of 165 identical amino acid residues that ended at the C-terminus of the open reading frame. The hits for human GPx2, GPx3, and GPx6 had additional residues at the 5' end of the open reading frame, and revealed two unique glutathione peroxidase sequences. The EST FF308178.1 open reading frame (hit for human GPx3) showed an additional sequence 5'-MVNISTSSILFVLVLVVALVFSFYLSFQSKNLSSITNK-3' at the N-terminus of the sequence (immediately before the conserved 165 residue region), while the ESTs FF310179.1 (hit for human GPx2 and GPx6) and CV846368.1 (hit for human GPx4) had the sequence 5'-MGLLFRRLLPSTVVVSSSLIFQSKNLSSITNK-3' before the conserved 165 residues. Therefore, only two unique GPx amino acid sequences and their encoding transcripts from these searches, were identified and considered for further studies.

Identification of Genomic Locations for Pea Aphid Glutathione Peroxidases Suggests Current Genome Annotations Need Revision

The two identified pea aphid GPx family members from the previous section (arbitrarily called ApGPx2a, ApGPx2b, which will be justified later) (Figure 2-2) and ApGPx1, the protein discovered in mass spectrometry studies (Carolan et al., 2011), were used to identify the glutathione peroxidase genes in the current version of the pea aphid genome (pea aphid genome assembly version 2, www.aphidbase.com). Each amino acid sequence was run through the Aphidbase blast server under the program tblastn, default settings, restricted to Genome assembly version 2 scaffolds.

The current version of the pea aphid genome has two genes annotated as glutathione peroxidase, with gene accession numbers ACYPI38240-RA and ACYPI002439-RA (Table 2-3, Figure 2-3). ACYPI38240-RA annotated has 7 exons, labeled Exons 1-7, while ACYPI002439-RA has 12 exons, labeled as Exons 1-12. While both genes have exons 1-7 in common, it is crucial to note that ACYPI38240-RA has an annotation for a shorter exon 7 than ACYPI002439-RA.

My three GPx sequences mapped to the same scaffold in the assembly (Aphidbase #GL35049831) (Table 2-4). ApGPx1 makes use of the 3'-end of exon 7, all of exons 8-11, and the 5'-end of exon 12, where exons refer to the current GPx gene annotations. ApGPx2a and ApGPx2b both use the full lengths of exons 3-6, and the 5'-end of exon 7, however they differ in that ApGPx2a uses a small, unannotated region approximately 200 nucleotides upstream of exon 3 to code for the N-terminal region of the protein, while ApGPx2b uses the 3'-end of exon 2 for its N-terminus. Mapping the ApGPx amino acid sequences to the reference genome indicated that revisions to the gene structure is needed (Figure 2-4). Therefore, EST databases were explored to find all ESTs that encoded each of the pea aphid GPxs.

EST Evidence Indicates Three Pea Aphid Glutathione Peroxidases Exist, and Suggests Salivary Gland Specific Production of ApGPx1

To identify all ESTs that encoded for each of our pea aphid glutathione peroxidases, the amino acid sequences of ApGPx1, ApGPx2a, and ApGPx2b were run through BLAST using tblastn under default settings, restricting searches to pea aphid EST libraries. Table 2-5 summarizes the number of EST hits to each individual ApGPx. The searches identified 74 hits to ApGPx1, 2 hits specific to ApGPx2a, and 27 hits specific to ApGPx2b. EST search results

from ApGPx2a and ApGPx2b revealed 38 overlapping ESTs that could not be distinguished between the two GPxs.

Interestingly, 68 of 74 EST hits for ApGPx1 were from salivary gland EST libraries, while only one EST hit for ApGPx2b was from salivary gland libraries, and neither of the ApGPx2a ESTs were from salivary gland libraries.

RNA-Seq Data Further Supports that ApGPx1 is Highly Expressed in Salivary Glands

I mapped the plant-fed and diet-fed RNASeq reads from the previous chapter of this thesis to the nucleotide sequences for ApGPx1, ApGPx2a, and ApGPx2b. For ApGPx1, the nucleotide sequence encoding the full protein was used as a reference template. The nucleotide sequence encoding the protein sequences unique for ApGPx2a and ApGPx2b (residues boxed in blue and red, respectively in Table 2-6) were used as reference templates. Approximately 38,000 and 58,000 reads mapped to ApGPx1 for plant-fed and diet-fed samples, respectively. The number of reads unique to ApGPx2a and ApGPx2b were approximately 600 and 1,600 in plant-fed samples, and about 900 and 1,800 in Akey-Beck diet-fed samples. To compare transcript expression of ApGPx1, ApGPx2a, and ApGPx2b in plant-fed insects, the first 70 nucleotides of the open reading frame of ApGPx1 was also used as a reference template, resulting in about 21,000 reads mapped. Comparison of RPKM values for the first 70 nucleotides of ApGPx1, and the unique regions of ApGPx2a and ApGPx2b showed that ApGPx1 had about 40-fold higher transcript level than ApGPx2a, and 12-fold higher transcript levels than ApGPx2b. Combined with EST data, these results suggest that ApGPx1 is exclusively produced in the salivary glands. Also, RNASeq data suggests that there is low expression of ApGPx2a and ApGPx2b in salivary glands.

Discussion

I have identified three glutathione peroxidases family members, ApGPx1, ApGPx2a, and ApGPx2b. EST evidence in conjunction with RNASeq analysis indicates that ApGPx1 is highly expressed in pea aphid salivary glands. ApGPx2a and ApGPx2b show some expression in salivary glands based on RNASeq evidence, and they may be expressed at low levels.

The current gene annotation has two pea aphid genes encoding pea aphid glutathione peroxidase (Figure 2-3). Analysis of pea aphid ESTs encoding glutathione peroxidases indicates that revisions to the genome are required. My work indicates that two genes, called *ApGPx1*, *ApGPx2* encode GPx in the pea aphids. The revised gene structure for pea aphid glutathione peroxidase can be seen in Figure 2-3. *ApGPx1* encodes the most abundant salivary gland GPx, ApGPx1. The revised gene structure for *ApGPx1* contains seven exons. Exon 1 codes for the 5' UTR, while the tail of exon 1 and exons 2 through exon 7 encode for the rest of the transcript. The *ApGPx2* gene consists of seven exons, and encodes for two differential spliced transcripts, ApGPx2a and ApGPx2b. Exon 2 encodes the 5' UTR for ApGPx1, while exon 2 through exon 7 code for the open reading frame. The tail of exon 7 encodes the 3'UTR for ApGPx1. ApGPx2b is encoded by the gene *ApGPx2* as well, a gene made up of seven exons, where exon 1 encodes the 5' UTR, the tail of exon 7 encodes the 3'UTR, and exon 3 through exon 7 encode the open reading frame.

The amino acid sequences for ApGPx1, ApGPx2a, and ApGPx2b were run individualy as queries through NCBI Blast program tblastn, restricting to *Hemiptera* nucleotide, protein, and whole genome shotgun contig databases. The search results found eight *Hemiptera* glutathione peroxidases, from potato aphid (Accession #JX134493.1), Hackberry psyllid (Accession #GAOP01093491.1), black-faced leafhopper (Accession #GAQX01031228.1), cowpea aphid (Accession #GAJW01003579.1), assassin bug (Accession #GAHY01001921.1), bean bug

(Accession #GAJX01009604.1), green peach aphid (Accession #EE261230.1), and cotton aphid (Accession #DR396163.1). Glutathione peroxidases from *Drosophila melanogaster* were identified from UniProt (www.uniprot.org). The five unique GPxs are referred to as DmGPx1-5 throughout this thesis, for Uniprot Accession #Q8IRD3, Q8IRD4, Q4V4T9, Q4V6H2, and Q9VZQ8, respectively.

A multiple sequence alignment (Figure 2-4) and phylogenetic tree (Figure 2-5) was created for the eight human, three pea aphid, five *D. melanogaster*, and eight *Hemiptera* glutathione peroxidase amino acid sequences. The catalytic tetrad was identified in all sequences.

Analysis of the amino acid sequences and the phylogenetic tree identified orthologs of ApGPx1 in three species, potato aphid, green peach aphid, and cotton aphid. Orthologs of proteins produced by gene *ApGPx2* were found in hackberry psyllid, black-faced leafhopper, cowpea aphid, assassin bug, and the bean bug. Orthologs of pea aphid glutathione peroxidases were not detected in human or *D. melanogaster*.

Reactive oxygen species are known to elicit defense-signaling pathways as a response to aphid attack (Boyko et al., 2006). Aphid salivary enzymes assist in the degradation of linolenic acid, which in combination with plant peptide hormones trigger the release of hydrogen peroxide (Gatehouse, 2002). The role of glutathione peroxidase in pea aphid saliva is unknown, however, glutathione peroxidase could serve to help circumvent plant defense responses by reducing hydrogen peroxide to water.

Figures and Tables

Table 2-1 The Eight Human Glutathione Peroxidases

Eight human glutathione peroxidases have been identified and studied. ER retention signals were identified by the motif KDEL or KEDL at the C-terminal end of the protein. Signal peptides were predicted by the CBS *SignalP 4.1 Server*. Protein location information was obtained from each individual gene page from NCBI.

Name	Accession	Intracellular Location	Tissue/Organ Location	Function	Selenium Containing	Signal Peptide	ER Retention Signal
GPx1	NP_000572	Cytoplasm	Nearly all mammalian tissues	Detoxification of hydrogen peroxide	Yes	No	No
GPx2	NP_002074	Extracellular	Gastrointestinal tract	Prevention of inflammation in GI tract	Yes	No	No
GPx3	NP_002075	Extracellular	Blood plasma	Detoxification of hydrogen peroxide	Yes	Yes	No
GPx4	NP_001034936	Biological membranes		Phospholipid hydroperoxidase	Yes	Yes	No
GPx5	NP_001500	Extracellular	Epididymis in the male reproductive tract	Protection of spermatozoa membrane from lipid peroxidation	No	Yes	No
GPx6	NP_874360		Embryos and adult olfactory epithelium	Detoxification of hydrogen peroxide	Yes	Yes	No
GPx7	NP_056511	Endoplasmic reticulum		Oxidation of protein disulfide isomerase	No	Yes	Yes
GPx8 (putative)	NP_001008398			Unknown	No	No	Yes

Table 2-2 The Top Pea Aphid EST Hits to Human Glutathione Peroxidases

All eight human GPx protein sequences were run individually as query sequences through the NCBI program *tblastn* under default conditions restricted to pea aphid EST libraries. The top pea aphid EST hit for each human GPx is listed below.

H. sapiens GPx isoform	Length in Amino Acid Residues	*A. pisum* EST Accession	EST Length	Query Coverage (%)	Sequence Similarity (%)
GPx1	203	FF327487.1	716	90	34
GPx2	190	FF310179.1	692	95	30
GPx3	226	FF308178.1	684	83	34
GPx4	227	CV846368.1	753	59	54
GPx5	221	EX646816.1	701	75	38
GPx6	221	FF310179.1	692	92	29
GPx7	187	EX608478.1	716	84	40
GPx8	209	EX642545.1	785	76	39

Table 2-3 Exons of the Two Current Annotated Glutathione Peroxidase Genes

The current pea aphid genome annotation (AphidBase gene consensus version 2) has two annotated genes for glutathione peroxidase. The exons used in annotations appear below.

GeneID	Exon	Start	Stop	Length (bases)
ACYPI38240-RA	1	171664	172020	357
ACYPI38240-RA	2	172098	172400	303
ACYPI38240-RA	3	174282	174407	126
ACYPI38240-RA	4	174475	174620	146
ACYPI38240-RA	5	175648	175794	147
ACYPI38240-RA	6	176783	176868	86
ACYPI38240-RA	7	177449	177760	312
ACYPI002439-RA	1	171767	172020	254
ACYPI002439-RA	2	172098	172400	303
ACYPI002439-RA	3	174282	174407	126
ACYPI002439-RA	4	174475	174620	146
ACYPI002439-RA	5	175648	175794	147
ACYPI002439-RA	6	176783	176868	86
ACYPI002439-RA	7	177449	177974	526
ACYPI002439-RA	8	189301	189531	231
ACYPI002439-RA	9	191317	191459	143
ACYPI002439-RA	10	192235	192381	147
ACYPI002439-RA	11	192465	192550	86
ACYPI002439-RA	12	192887	193168	282

Table 2-4 Pea Aphid Glutathione Peroxidase Mapped to Reference Genome

The three pea aphid GPx amino acid sequences were mapped to the pea aphid reference genome assembly version 2 using *tblastn* on the Aphidbase blast server ((www.aphidbase.com). All three GPxs were mapped to the same scaffold of the genome (GL35049831) but to different regions.

Name	Start	Stop	Annotated Exon #
ApGPx1	177924	177974	Exon 7
ApGPx1	189302	189529	Exon 8
ApGPx1	191319	191460	Exon 9
ApGPx1	192236	192380	Exon 10
ApGPx1	192467	192550	Exon 11
ApGPx1	192888	192935	Exon 12
ApGPx2a	174066	174143	No Exon Annotated
ApGPx2a	174283	174405	Exon 3
ApGPx2a	174477	174621	Exon 4
ApGPx2a	175649	175792	Exon 5
ApGPx2a	176785	176868	Exon 6
ApGPx2a	177450	177482	Exon 7
ApGPx2b	172341	172400	Exon 2
ApGPx2b	174283	174405	Exon 3
ApGPx2b	174477	174621	Exon 4
ApGPx2b	175649	175792	Exon 5
ApGPx2b	176785	176868	Exon 6
ApGPx2b	177450	177482	Exon 7

Table 2-5 ESTs Matching Pea Aphid Glutathione Peroxidases

The three ApGPx amino acid sequences were run individually as query sequences through the NCBI program tblastn, restricting searches to pea aphid EST libraries. Library ID two letter codes indicate the first two letters of accession numbers for each individual EST library. The resulting table is a summary of ESTs and their corresponding libraries that match sequence.

ApGPx	Library Author	ID	Type	Representing ESTs
ApGPx1	Reeck	HS	Salivary Gland	62
ApGPx1	Reeck	DV	Salivary Gland	6
ApGPx1	Richards	EX	Whole Body (2008)	5
ApGPx1	Tagu	FP	Whole Body Males	1
ApGPx2a	Richards	FF	Whole Body (2007)	2
ApGPx2b	Richards	FF	Whole Body (2008)	9
ApGPx2b	Tagu	FP	Whole Body Males	9
ApGPx2b	Richards	EX	Whole Body (2009)	6
ApGPx2b	Hunter	CN	Whole Body	2
ApGPx2b	Reeck	DV	Salivary Gland	1
ApGPx2a/2b	Richards	EX	Whole Body (2007)	17
ApGPx2a/2b	Richards	FF	Whole Body (2008)	16
ApGPx2a/2b	Tagu	FP	Whole Body Males	2
ApGPx2a/2b	Reeck	DV	Salivary Gland	1
ApGPx2a/2b	Tagu	GH	Brain	1
ApGPx2a/2b	Tagu	CV	Antennae	1

Table 2-6 RNASeq Reads Mapping to Pea Aphid Glutathione Peroxidases

RNASeq reads from pea aphid salivary glands were mapped to full length ApGPx1, the first 70 nucleotides encoding ApGPx1, the first 78 nucleotides encoding the unique region of ApGPx2a, and the first 60 nucleotides encoding the unique region of ApGPx2b as described in the text. A) Table of RNASeq results and statistics. B) Sequence alignment showing the unique regions of ApGPx2a and ApGPx2b. The unique amino acid sequence for ApGPx2a and ApGPx2b are boxed in blue and red, respectively.

A)

Pea Aphid GPx Name	Plant-fed Reads Mapped	Plant-fed RPKM	Diet-fed Reads Mapped	Diet-fed RPKM	Fold Change
ApGPx1	37,634	1339.8	51847	1708	0.79
ApGPx1 (first 70 nucleotides)	21324	7621.7	8656	2849.5	2.67
ApGPx2a (unique-blue box)	596	191.2	924	272.9	0.70
ApGPx2b (unique-red box)	1585	660.9	1802	692.1	0.95

B)

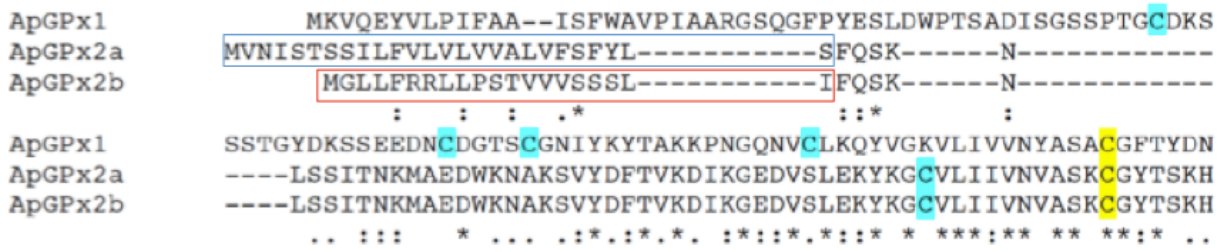


Figure 2-1 Sequence Alignment of the Translated Top Pea Aphid EST Hits to Human GPxs

The translated sequences for the top hit of pea aphid ESTs to Human GPxs were aligned via ClustalOmega. Residues highlighted in yellow make up the catalytic tetrad. The top hit for human GPx2 and GPx6 were identical, and therefore combined in the alignment. Hits for human GPx5, GPx7, and GPx8 were incomplete ESTs encoding ApGPx2b. Asterisks (*) indicate positions that have a single, fully conserved residue.

```
FF327487.1_GPx1_Hit                 RRLLPSTVVVSSSLIFQSKNLSSITNKMAEDWKNAKSVY
FF310179.1_GPx2/6_Hit             MGLLFRRLLPSTVVVSSSLIFQSKNLSSITNKMAEDWKNAKSVY
FF308178.1_GPx3_Hit           MVNISTSSILFVLVLVVALVFSFYLSFQSKNLSSITNKMAEDWKNAKSVY
CV846368.1_GPx4_Hit               MGLLFRRLLPSTVVVSSSLIFQSKNLSSITNKMAEDWKNAKSVY
EX646816.1_GPx5_Hit                                                 MAEDWKNAKSVY
EX608478.1_GPx7_Hit                                                 MAEDWKNAKSVY
EX642545.1_GPx8_Hit                                                 MAEDWKNAKSVY
                                                                    ************

FF327487.1_GPx1_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
FF310179.1_GPx2/6_Hit         DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
FF308178.1_GPx3_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
CV846368.1_GPx4_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
EX646816.1_GPx5_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
EX608478.1_GPx7_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
EX642545.1_GPx8_Hit           DFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKHYKELIELDEKYRD
                              ************************************************

FF327487.1_GPx1_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
FF310179.1_GPx2/6_Hit         KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
FF308178.1_GPx3_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
CV846368.1_GPx4_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
EX646816.1_GPx5_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
EX608478.1_GPx7_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
EX642545.1_GPx8_Hit           KGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGNDAHP
                              ************************************************

FF327487.1_GPx1_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKKLE
FF310179.1_GPx2/6_Hit         LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKNLE
FF308178.1_GPx3_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPFGNKVFLF
CV846368.1_GPx4_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDK>
EX646816.1_GPx5_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKNLE
EX608478.1_GPx7_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKNLE
EX642545.1_GPx8_Hit           LWKYLKSKQGGLLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKILE
                              ****************************

FF327487.1_GPx1_Hit           KY>
FF310179.1_GPx2/6_Hit         KYL>
FF308178.1_GPx3_Hit           KLFK>
CV846368.1_GPx4_Hit
EX646816.1_GPx5_Hit           KYL>
EX608478.1_GPx7_Hit           KYL>
EX642545.1_GPx8_Hit           KYL>
```

Figure 2-2 Sequence Alignment of Pea Aphid GPxs

Sequences for the four Pea Aphid glutathione peroxidases were aligned via clustalW. Residues highlighted in yellow are strictly conserved residues of the catalytic tetrad. Cysteine residues are highlighted in light blue. Asterisks (*) indicated positions that have a single, fully conserved residue. Colons (:) indicate conservation between groups of strongly similar properties. Periods (.) indicate conservation between groups of weakly similar properties.

```
ApGPx1           MKVQEYVLPIFAA--ISFWAVPIAARGSQGFPYESLDWPTSADISGSSPTGCDKS
ApGPx2a      MVNISTSSILFVLVLVVALVFSFYL-----------SFQSK------N------------
ApGPx2b          MGLLFRRLLPSTVVVSSSL-----------IFQSK------N------------
                      :   :  :  .*                ::*       :
ApGPx1       SSTGYDKSSEEDNCDGTSCGNIYKYTAKKPNGQNVCLKQYVGKVLIVVNYASACGFTYDN
ApGPx2a      ----LSSITNKMAEDWKNAKSVYDFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKH
ApGPx2b      ----LSSITNKMAEDWKNAKSVYDFTVKDIKGEDVSLEKYKGCVLIIVNVASKCGYTSKH
                  .. :::   * ... .:*.:*.*. :*::*.*::* * ***:** ** **:* ..

ApGPx1       VCTLSEFAQKYRKCGLEILVFPSNDFLQNIGGNIAA--EELANNHPEFEVFSEICVNGRA
ApGPx2a      YKELIELDEKYRDKGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGND
ApGPx2b      YKELIELDEKYRDKGLKILGFPCNQFGGQEPGDADSICSFTAKQNVKFDIFEKIDVNGND
               * *: :***. **:** **.*:*  :  *:  :  .  *::. :*::*.:* ***.

ApGPx1       QHPVYRFLKYKLPGPFNTKTIKWNFTKFVVDRNGCPVQRYEATDSFKDIEELVQELLKDQ
ApGPx2a      AHPLWKYLKSKQGG-LLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKNLEKYL>
ApGPx2b      AHPLWKYLKSKQGG-LLIDSIKWNFTKFIVDKNGQPVERHAANVSPLGLEKNLEKYL>
              **:::** *  * :  .:********:**:** **:*: *. *   :*: ::: *

ApGPx1       CC>
ApGPx2a
ApGPx2b
```

Figure 2-3 The Current and Revised Gene Structures of Pea Aphid GPxs

The pea aphid genome assembly version 2 (www.aphidbase.com has two current gene annotations, represented by A) ACYPI002439-RA and B) ACYPI38240. C) Proposed gene structure of *ApGPx1*. D) Proposed gene structure of *ApGPx2*. Color codes: Green: 5'UTR and 3'UTR. Red or Pink: protein coding region. Black: Exons.

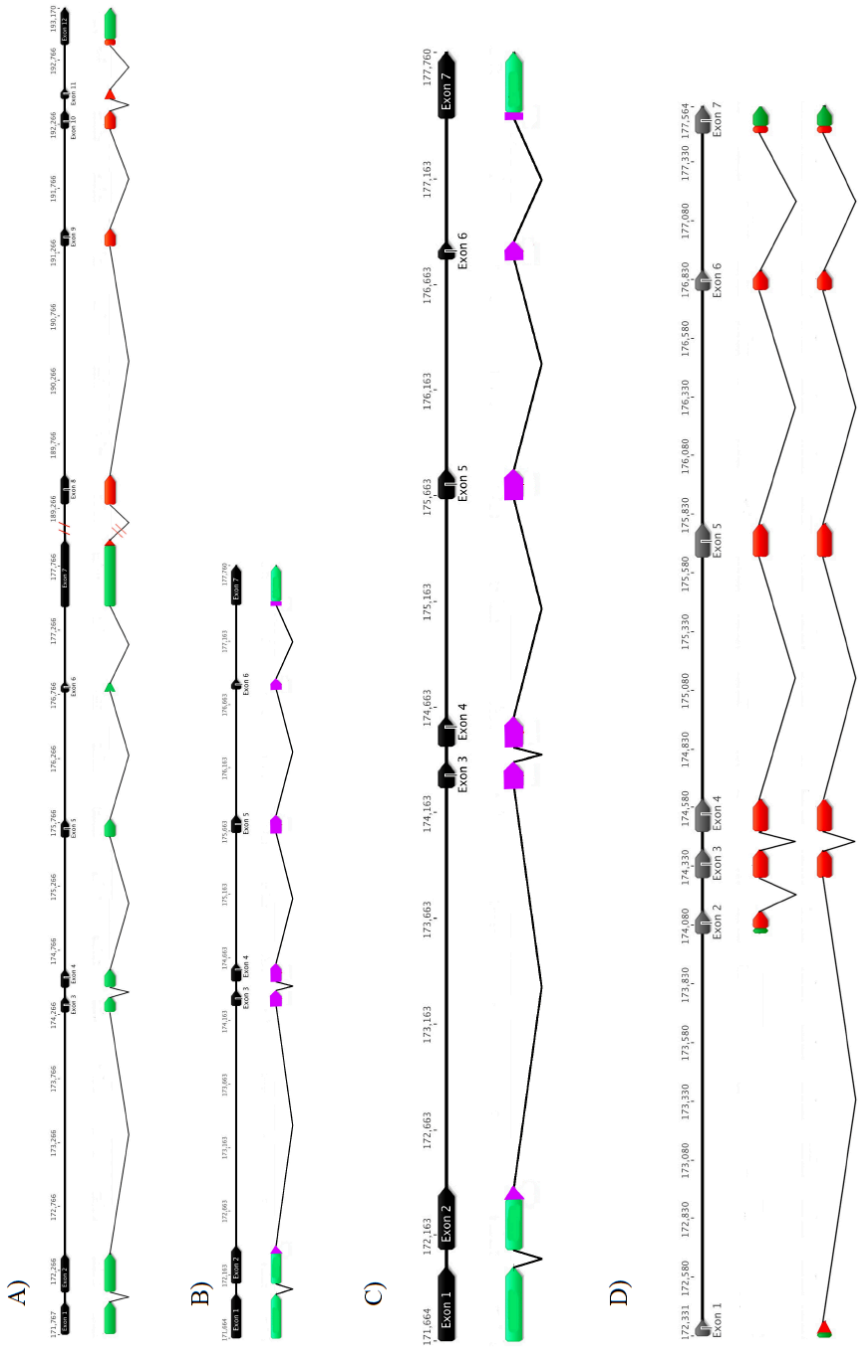


Figure 2-4 Sequence Alignment of Human, Pea Aphid, Drosophila, and Hemiptera GPxs

Sequences for human, pea aphid, *D. melanogaster*, and *Hemiptera* glutathione peroxidases were aligned by Clustal Omega. Residues highlighted in yellow are strictly conserved residues of the catalytic tetrad. Asterisks (*) indicated positions that have a single, fully conserved residue. Colons (:) indicate conservation between groups of strongly similar properties. Periods (.) indicate conservation between groups of weakly similar properties. Hs: Human, Ag: cotton aphid, Mp: green peach aphid, Ap: pea aphid, Me: potato aphid, Dm: *D. melanogaster*, Ct: bean bug, Pv: hackberry psyllid, Ac: cowpea aphid, Gn: black-faced leafhopper, Rp: assassin bug.

```
HsGPx5       ------------------------------------------------------------      0
HsGPx3       ------------------------------------------------------------      0
HsGPx6       ------------------------------------------------------------      0
HsGPx1       ------------------------------------------------------------      0
HsGPx2       ------------------------------------------------------------      0
HsGPx7       ------------------------------------------------------------      0
HsGPx8       ------------------------------------------------------------      0
AgGPx        ------------------------------------------------------------      0
MpGPx        ------------------------------------------------------------      0
ApGPx1       ------------------------------------------------------------      0
MeGPx        ------------------------------------------------------------      0
DmGPx3       ------------------------------------------------------------      0
DmGPx4       ------------------------------------------------------------      0
HsGPx4       ------------------------------------------------------------      0
DmGPx2       ------------------------------------------------------------      0
DmGPx1       ------------------------------------------------------------      0
DmGPx5       ------------------------------------------------------------      0
CtGPx        ------------------------------------------------------------      0
PvGPx        MPRESLPRQAKTAAVNELKSLREEAITKPKAKKEVKKKIEKAIKKQTPVKKSPKVKATSK      60
ApGPx2a      ------------------------------------------------------------      0
ApGPx2b      ------------------------------------------------------------      0
AcGPx        ------------------------------------------------------------      0
GnGPx        ------------------------------------------------------------      0
RpGPx        ------------------------------------------------------------      0

HsGPx5       ------------------------------------------------------------      0
HsGPx3       ------------------------------------------------------------      0
HsGPx6       ------------------------------------------------------------      0
HsGPx1       ------------------------------------------------------------      0
HsGPx2       ------------------------------------------------------------      0
HsGPx7       ------------------------------------------------------------      0
HsGPx8       ------------------------------------------------------------      0
AgGPx        ------------------------------------------------------------      0
MpGPx        ------------------------------------------------------------      0
ApGPx1       ------------------------------------------------------------      0
MeGPx        ------------------------------------------------------------      0
DmGPx3       ------------------------------------------------------------      0
DmGPx4       ------------------------------------------------------------      0
HsGPx4       ------------------------------------------------------------      0
DmGPx2       ------------------------------------------------------------      0
DmGPx1       ------------------------------------------------------------      0
DmGPx5       ------------------------------------------------------------      0
CtGPx        ------------------------------------------------------------      0
PvGPx        GKNKKEEDDAKIEDEEEEEKEEEDEKNNKNEEDEEKDEKDEDEVDNEKDEKEEVIEKEDI     120
ApGPx2a      ------------------------------------------------------------      0
ApGPx2b      ------------------------------------------------------------      0
AcGPx        ------------------------------------------------------------      0
GnGPx        ------------------------------------------------------------      0
RpGPx        ------------------------------------------------------------      0
```

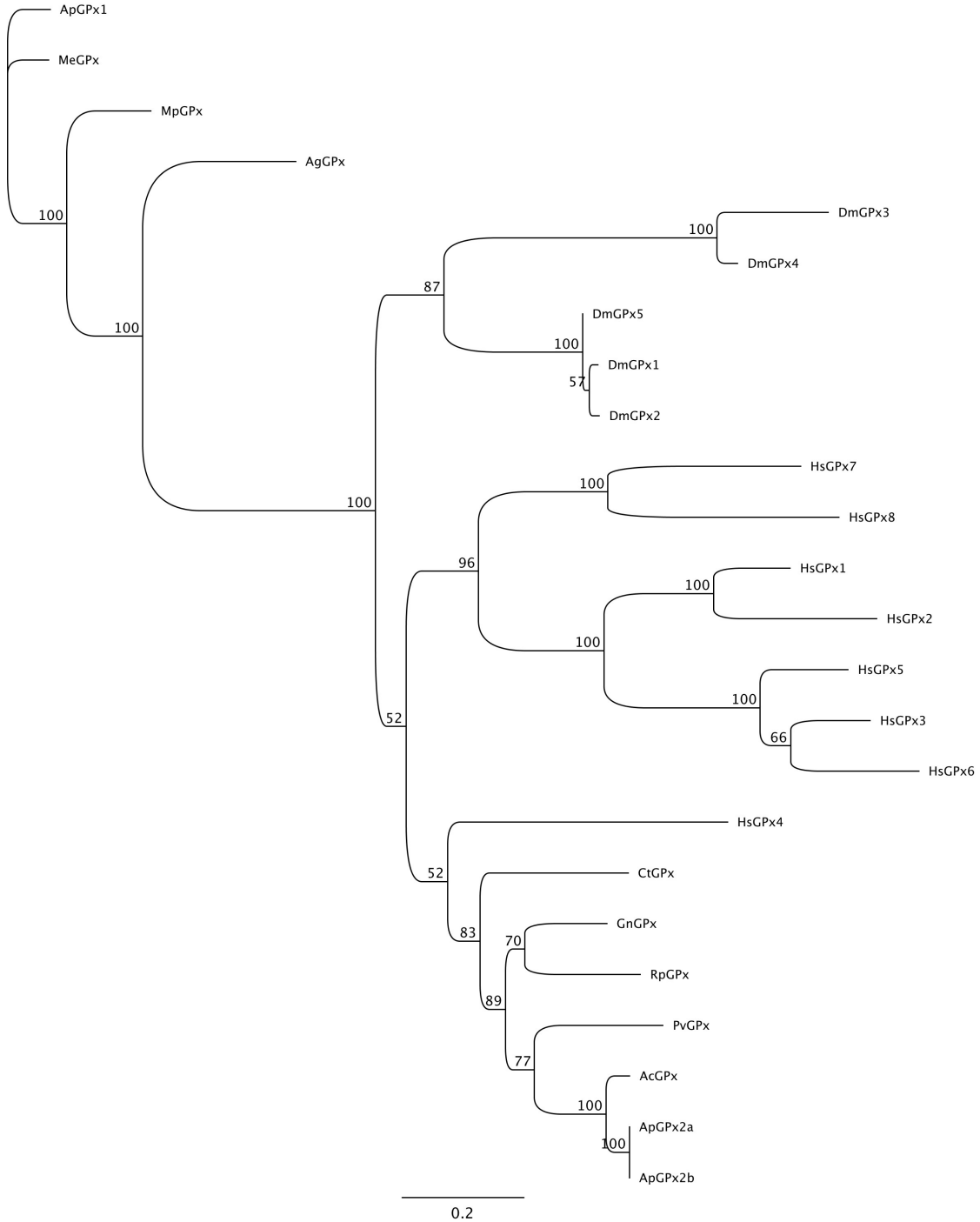
```
HsGPx5      ------------------------------------------------------------      0
HsGPx3      ------------------------------------------------------------      0
HsGPx6      ------------------------------------------------------------      0
HsGPx1      ------------------------------------------------------------      0
HsGPx2      ------------------------------------------------------------      0
HsGPx7      ------------------------------------------------------------      0
HsGPx8      ----------------------------------------------MEPLAAYPLKC-      11
AgGPx       ------------------------------------------------------------      0
MpGPx       ------------------------------------------------------------      0
ApGPx1      ------------------------------------------------------------      0
MeGPx       ------------------------------------------------------------      0
DmGPx3      ------------------------------------------------------------      0
DmGPx4      ------------------------------------------------------------      0
HsGPx4      ------------------------------------------------------------      0
DmGPx2      ------------------------------------------------------------      0
DmGPx1      ---------------------------------------MSL----RQFQNISRQALRCY      17
DmGPx5      ------------------------------------------------------------      0
CtGPx       ------------------------------------------------------------      0
PvGPx       TAKKSESNEDKDDKPEEEKKEEGEEEKKEESNSKKEEVENDIADDEPAVKKLKKDTPDI-     179
ApGPx2a     ------------------------------------------------------MVNI-      4
ApGPx2b     ------------------------------------------------------------      0
AcGPx       ------------------------------------------------------------      0
GnGPx       ------------------------------------------------------------      0
RpGPx       ------------------------------------------------------------      0

HsGPx5      --------------------MTTQLRVVHLLPLLLA--------------------------      16
HsGPx3      --------------------MARLLQASCLLSLLLA--------------------------      16
HsGPx6      --------------------MFQQFQASCLVLFFLV--------------------------      16
HsGPx1      ------------------------------------------------------------      0
HsGPx2      ------------------------------------------------------------      0
HsGPx7      ------------------------------MVAA-TVAAAWL-------------------      11
HsGPx8      ------------------SGPRAKVFA---VLLSIVLCTVTLF-------------------      33
AgGPx       ----------------------MNTQEYIILIFA-VSFWMAPISADMSQA----YDPFDWQ      34
MpGPx       ----------------------MKIQECILLIFATISFWAEPFSAVNCQQGYSNPLNLNWP      39
ApGPx1      ----------------------MKVQEYVLPIFAAISFWAVPIAARGSQ-GFP-YESLDWP      37
MeGPx       ---------------------------------------------RVSQ-GFP-YESLDWP      14
DmGPx3      ------------------------------------------------------------      0
DmGPx4      ------------------------MFDKEFLFPGLLVAV-ALVVV----------------      20
HsGPx4      ------------------------MSLGRLCRLLKPALLCGALAA-----------------      21
DmGPx2      ------------M----AG-RS-IV----HFFL-------GSVAIA---------------      17
DmGPx1      SMRRTPGPVLELS----RGQRQ-CLRLC-TIMLPVSCAATPMNAIS---------------      57
DmGPx5      ------------------------------------------------------------      0
CtGPx       ------------------------------------------------------------      0
PvGPx       TESKSNGDVGKKEESDENGDKNEEVTSEEDVKKVQ------SNGKD---------------     219
ApGPx2a     STSS-----------------I-LFVLVLVVALVF------SF-YL---------------      25
ApGPx2b     --MG-----------------L-LFRRLLPSTVVV------SS-SL---------------      19
AcGPx       ------------------------------------------------------------      0
GnGPx       ------------------------------MFSTV------KN-IL---------------      9
RpGPx       --MR-----------------Y-YLLIVSLILLAT------SI-IL---------------      19

HsGPx5      -----------------CFVQTSPKQEKMKMDCHKDEKGTIYDYEAIALNKNEYVSFKQY      59
HsGPx3      -----------------GFVSQSRGQEKSKMDCHGGISGTIYEYGALTIDGEEYIPFKQY      59
HsGPx6      -----------------GFAQQTLKPQNRKVDCNKGVTGTIYEYGALTLNGEEYIQFKQF      59
HsGPx1      --------------------------MCAARLAAAAAAAQSVYAFSARPLAGGEPVSLGSL     35
HsGPx2      ---------------------------------MAFIAKSFYDLSAISL-DGEKVDFNTF     26
HsGPx7      --------------------------LLWAAACAQQEQDFYDFKAVNI-RGKLVSLEKY      43
HsGPx8      --------------------------LLQLKFLKPKINSFYAFEVKDA-KGRTVSLEKY      65
AgGPx       TSRSMMDISKSTSFEYDKF---GFEN-SNSDEYDYILYENVYDYTVQNL-DGQEVCLRKY     89
MpGPx       TSADIL---EPSPFGYDK--------SSFESNSDEESCCNLYNFVVKRP-NGQDVSLKQY     87
ApGPx1      TSADIS---GSSPTGCDKSSSTGYDKSSEEDNCDGTSCGNIYKYTAKKP-NGQNVCLKQY     93
MeGPx       TMPDIS---GSSPTGYDKSSSSGYDKSSDEDNCNEDSCGNIYKFTARKP-NGQNVCLQQY     70
DmGPx3      ----------------------MASDDTRSHRSRYVWESSATRYIRR------PC--VAY     30
DmGPx4      ----------------------LQTRSRLQQDLQDMRWRLTIHALTVRDT-FGNPVQLDTF     58
HsGPx4      ----------------------PGLAGTMCASRDDWRCARSMHEFSAKDI-DGHMVNLDKY     59
DmGPx2      ----------------LGSYIYFTMQIDMSANGDYKNAASIYEFTVKDT-HGNDVSLEKY     60
DmGPx1      ----------------SAAQHSTAAAIDMSANGDYKNAASIYEFTVKDT-HGNDVSLEKY     100
DmGPx5      ---------------------------MSANGDYKNAASIYEFTVKDT-HGNDVSLEKY     31
CtGPx       ----------------------RSFSIMSDASEINWKEASSVYDFTVKDI-KGNDVPLKDY     38
```

```
PvGPx        ----------------T-TASTNENSESKKM--ATEEPSTVYDFTVKNI-KGEDVPLSKY     259
pGPx2a      ----------------S-FQSKNLSSITNKMAEDWKNAKSVYDFTVKDI-KGEDVSLEKY      67
ApGPx2b      ----------------I-FQSKNLSSITNKMAEDWKNAKSVYDFTVKDI-KGEDVSLEKY      61
AcGPx        -------------------QSKSLSSTTNKMAEDWKNAKSVYDFTVKDI-KGEDVSLEKY      40
GnGPx        ----------------G-EALSSFSS--NMGNEDWKNAKSVYDFTVKDI-KGEEVSLEKY      49
RpGPx        ----------------H-YQLTGFYSTEHTMAESEKDASSVYDFTVKDI-AGNDISLEKY      61

HsGPx5       VGKHILFVNVATYCGLT-AQYPELNALQEELKP-YGLVVLGFPCNQFGKQEPGDNKE-IL     116
HsGPx3       AGKYVLFVNVASYUGLT-GQYIELNALQEELAP-FGLVILGFPCNQFGKQEPGENSE-IL     116
HsGPx6       AGKHVLFVNVAAYUGLA-AQYPELNALQEELKN-FGVIVLAFPCNQFGKQEPGTNSE-IL     116
HsGPx1       RGKVLLIENVASLUGTTVRDYTQMNELQRRLGP-RGLVVLGFPCNQFGHQENAKNEE-IL      93
HsGPx2       RGRAVLIENVASLUGTTTRDFTQLNELQCRF-P-RRLVVLGFPCNQFGHQENCQNEE-IL      83
HsGPx7       RGSVSLVVNVASECGFTDQHYRALQQLQRDLGP-HHFNVLAFPCNQFGQQEPDSNKE--I     100
HsGPx8       KGKVSLVVNVASDCQLTDRNYLGLKELHKEFGP-SHFSVLAFPCNQFGESEPRPSKE--V     122
AgGPx        AGQVLIIVNYASTCGFTFENVCNLSKLSEKYRR-QGLTILMFPSNDFFQNIAGNTAAEML     148
MpGPx        AGKVLIILNYASGCGFTQDNVCTLTEFSNKYRA-CGLEILVFPSNDFSQNFGGNTAAQIF     146
ApGPx1       VGKVLIVVNYASACGFTYDNVCTLSEFAQKYRK-CGLEILVFPSNDFLQNIGGNIAAEEL     152
MeGPx        VGKVLIIVNYASACGFTYDNVCTLSDFAQRYRK-CGLEILVFPSNDFLQNFGGDIAAEEF     129
DmGPx3       RQ-------HCLQMRTYFISYDGLRYLLEEYED-QGLRILNFPCNQFGGQMPESDGQEML      82
DmGPx4       AGHVLLIVNIASKCGLTLSQYNGLRYLLEEYED-QGLRILNFPCNQFGGQMPESDGQEML     117
HsGPx4       RGFVCIVTNVASQUGKTEVNYTQLVDLHARYAE-CGLRILAFPCNQFGKQEPGSNEE--I     116
DmGPx2       KGKVVLVVNIASKCGLTKNNYEKLTDLKEKYGE-RGLVILNFPCNQFGSQMPEADGEAMV     119
DmGPx1       KGKVVLVVNIASKCGLTKNNYEKLTDLKEKYGE-RGLVILNFPCNQFGSQMPEADGEAMV     159
DmGPx5       KGKVVLVVNIASKCGLTKNNYEKLTDLKEKYGE-RGLVILNFPCNQFGSQMPEADGEAMV      90
CtGPx        EGKVLLIVNVASKCGLTSNNYKELVNLDEKYRN-EGLRILAFPCNQFARQEPGSAEE--I      95
PvGPx        RGNVLLIVNVASKCGYTSKHYKELVELDEKYRE-RGLRILAFPCNQFGGQEPGSPDQ--I     316
ApGPx2a      KGCVLIIVNVASKCGYTSKHYKELIELDEKYRD-KGLKILGFPCNQFGGQEPGDADS--I     124
ApGPx2b      KGCVLIIVNVASKCGYTSKHYKELIELDEKYRD-KGLKILGFPCNQFGGQEPGDADS--I     118
AcGPx        KGFVLIIVNVASKCGYTSKHYKELVELDEKYHD-KGLRILGFPCNQFGGQEPGDAES--I      97
GnGPx        KGHVLLIVNVASQCGLTKDNYKELVELDEKYRESKGLRILAFPCNQFGGQEPGTNAD--I     107
RpGPx        RGHVLIIVNVASQCGLTSTNYKELVELDEKYRETKGLRILAFPCNQFGSQEPGSRED--I     119
.            :  :         . :* **.*:*  .         .

HsGPx5       PGLKYVRPGGGFVPSFQLFEKGDVNGEKEQKVFSFLKHSCPHPSE----ILGTFKSISWD     172
HsGPx3       PTLKYVRPGGGFVPNFQLFEKGDVNGEKEQKFYTFLKNSCPPTSE----LLGTSDRLFWE     172
HsGPx6       LGLKYVCPGSGFVPSFQLFEKGDVNGEKEQKVFTFLKNSCPPTSD----LLGSSSQLFWE     172
HsGPx1       NSLKYVRPGGGFEPNFMLFEKCEVNGAGAHPLFAFLREALPAPSDDATALMTDPKLITWS     153
HsGPx2       NSLKYVRPGGGYQPTFTLVQKCEVNGQNEHPVFAYLKDKLPYPYDDPFSLMTDPKLIIWS     143
HsGPx7       ESFARR----TYSVSFPMFSKIAVTGTGAHPAFKYLAQTS---------------------     136
HsGPx8       ESFARK----NYGVTFPIFHKIKILGSEGEPAFRFLVDSS---------------------     158
AgGPx        AR----------SHPEFEVFSQICVNGKDTHPFYRFLKYK---LP----------------G     181
MpGPx        AD----------NHPQFEVFAEICVNGRAQHPMYRFLKNK---LP----------------G     179
ApGPx1       AN----------NHPEFEVFSEICVNGRAQHPVYRFLKYK---LP----------------G     185
MeGPx        AN----------NHPEFEVFSEICVNGRSQHPLYRFLKNK---LP----------------G     162
DmGPx3       DHLRRE-----GANIGHLFAKIDVKGAQADPLYKLLTRH----------------------     116
DmGPx4       DHLRRE-----GANIGHLFAKIDVKGAQADPLYKLLTRH----------------------     151
HsGPx4       KEFAAG-----YNVKFDMFSKICVNGDDAHPLWKWMKIQP--KG----------------K     154
DmGPx2       CHLRDS-----KADIGEVFAKVDVNGDNAAPLYKYLKA----KQ----------------T     155
DmGPx1       CHLRDS-----KADIGEVFAKVDVNGDNAAPLYKYLKA----KQ----------------T     195
DmGPx5       CHLRDS-----KADIGEVFAKVDVNGDNAAPLYKYLKA----KQ----------------T     126
CtGPx        CSFAEK-----KNAKFDFFEKINVNGKNAHPLWQYLKERKKRKK----------------S     135
PvGPx        CEFTKK-----KNVSFDLFEKVNVNGDQAHPLWNFLKQ----KQ----------------G     352
ApGPx2a      CSFTAK-----QNVKFDIFEKIDVNGNDAHPLWKYLKS----KQ----------------G     160
ApGPx2b      CSFTAK-----QNVKFDIFEKIDVNGNDAHPLWKYLKS----KQ----------------G     154
AcGPx        CSFTAKKFYCKKNVKFDIFDKVDVNGNDAHPLWKYLKS----KQ----------------G     138
GnGPx        CTFVSK-----YSAKFDFFDKIDVNGDNAHPLWKYLKH----KQ----------------G     143
RpGPx        VCFMRQ-----KNASFEMFDKIEVNGSNAHPLWKYLKS----KQ----------------G     155
.. :  : *      :  :

HsGPx5       PVKVHDIRWNFEKFLVGPDGIPVM------------------RWSHRATVSSVKTDILAY     214
HsGPx3       PMKVHDIRWNFEKFLVGPDGIPIM------------------RWHHRTTVSNVKMDILSY     214
HsGPx6       PMKVHDIRWNFEKFLVGPDGVPVM------------------HWFHQAPVSTVKSDILEY     214
HsGPx1       PVCRNDVAWNFEKFLVGPDGVPLR------------------RYSRRFQTIDIEPDIEAL     195
HsGPx2       PVRRSDVAWNFEKFLIGPEGEPFR------------------RYSRTFPTINIEPDIKRL     185
HsGPx7       ---GKEPTWNFWKYLVAPDGKVVG------------------AWDPTVSVEEVRPQITAL     175
HsGPx8       ---KKEPRWNFWKYLVNPEGQVVK------------------FWKPEEPIEVIRPDIAAL     197
AgGPx        AFKSKSIKWNFTKFIIDRNGCPVK------------------RYSTKDSFQDIEECIQQL     223
MpGPx        AFNTKTIKWNFTKFVVDRNGCPIQ------------------RFAATDGFTDLEELIQEL     221
ApGPx1       PFNTKTIKWNFTKFVVDRNGCPVQ------------------RYEATDSFKDIEELVQEL     227
```

```
MeGPx        ACNAKPIKWNFTKFVVDRNGCPVQ------------------RYAATDSFKDIEDLVQEL     204
DmGPx3       ---QHDIEWNFVKFLVDRKGNIHK------------------RYGAELEPVALTDDIELL     155
DmGPx4       ---QHDIEWNFVKFLVDRKGNIHK------------------RYGAELEPVALTDDIELL     190
HsGPx4       GILGNAIKWNFTKFGHRLSTVPHRQERLRGEALRTHGGAPGDREGPAPLF--LAPQVCGP     212
DmGPx2       GTLGSGIKWNFTKFLVNKEGVPIN------------------RYAPTTDPMDIAKDIEKL     197
DmGPx1       GTLGSGIKWNFTKFLVNKEGVPIN------------------RYAPTTDPMDIAKDIEKL     237
DmGPx5       GTLGSGIKWNFTKFLVNKEGVPIN------------------RYAPTTDPMDIAKDIEKL     168
CtGPx        GLLGSAIKWNFTKFLIDKEGQPVE------------------RFGPKDSFEKIDESVSKH     177
PvGPx        GTLFDAIKWNFTKFIVDKNGIPVE------------------RHAATTSAASLSPNIEKY     394
ApGPx2a      GLLIDSIKWNFTKFIVDKNGQPVE------------------RHAANVSPLGLEKNLEKY     202
ApGPx2b      GLLIDSIKWNFTKFIVDKNGQPVE------------------RHAANVSPLGLEKNLEKY     196
AcGPx        GLLIDSIKWNFTKFIVDKDGQPVE------------------RHAANVSPLGLEKSLEKY     180
GnGPx        GTLGDSIKWNFSKFIVDKNGQPVE------------------RFAPTTPPHKLVSSLEKY     185
RpGPx        GTLVDNIKWNFTKFIIDKNGQPVE------------------RHGPMTNPSKLLSSLEKY     197
*** *:    .                                 :   :

HsGPx5       LKQFKTK--------      221
HsGPx3       MRRQAALGVKRK---      226
HsGPx6       LKQFNTH--------      221
HsGPx1       LSQGPSCA-------      203
HsGPx2       LKVAI----------      190
HsGPx7       VRKLILLKREDL---      187
HsGPx8       VRQVIIKKKEDL---      209
AgGPx        LMDQSC---------      229
MpGPx        LKNQNC---------      227
ApGPx1       LKDQCC---------      233
MeGPx        LKDQCC---------      210
DmGPx3       LGR------------      158
DmGPx4       LGR------------      193
HsGPx4       ARAPAHALGAFHRHS      227
DmGPx2       L--------------      198
DmGPx1       L--------------      238
DmGPx5       L--------------      169
CtGPx        L--------------      178
PvGPx        L--------------      395
ApGPx2a      L--------------      203
ApGPx2b      L--------------      197
AcGPx        F--------------      181
GnGPx        W--------------      186
RpGPx        W--------------      198
```

Figure 2-5 Phylogenetic Tree of Human, Pea Aphid, Drosophila, and Hemiptera GPxs

A phylogenetic tree was constructed using the software program Geneious, which utilized the Jukes-Cantor tree building method. The scale bar below represent a distance of 0.2. Distance values appear in the computational matrix below. Bootstrap values appear on the phylogenetic tree at each clade.



	MeGPx	MpGPx	AgGPx	DmGPx5	DmGPx3	DmGPx4	HsGPx4	DmGPx1	DmGPx2	HsGPx7	HsGPx8	CtGPx	HsGPx1	HsGPx2	HsGPx5	HsGPx3	HsGPx6	PvGPx	GnGPx	RpGPx	AcGPx	ApGPx2a	ApGPx2b
ApGPx1	0.14	0.3	0.54	1.01	1.4	1.26	1.24	1.03	1.03	1.36	1.42	1.08	1.34	1.48	1.44	1.47	1.55	1.14	1.04	1.1	1.08	1.08	1.08
MeGPx		0.3	0.54	1	1.4	1.25	1.24	1.03	1.03	1.36	1.42	1.08	1.34	1.48	1.43	1.47	1.55	1.13	1.04	1.1	1.08	1.08	1.08
MpGPx			0.51	0.98	1.37	1.23	1.21	1	1	1.33	1.39	1.05	1.31	1.45	1.41	1.44	1.52	1.11	1.01	1.07	1.05	1.05	1.05
AgGPx				0.96	1.36	1.22	1.2	0.99	0.99	1.32	1.38	1.04	1.3	1.44	1.4	1.43	1.51	1.09	1	1.06	1.04	1.04	1.04
DmGPx5					0.85	0.7	0.91	0.03	0.03	1.03	1.09	0.75	1.01	1.15	1.11	1.14	1.22	0.8	0.71	0.77	0.75	0.75	0.75
DmGPx3						0.21	1.31	0.88	0.88	1.43	1.49	1.15	1.41	1.55	1.5	1.54	1.62	1.2	1.11	1.17	1.15	1.15	1.15
DmGPx4							1.16	0.73	0.73	1.28	1.34	1	1.26	1.4	1.36	1.39	1.47	1.06	0.97	1.02	1	1	1
HsGPx4								0.93	0.94	1.16	1.23	0.75	1.15	1.29	1.24	1.28	1.36	0.81	0.71	0.77	0.75	0.75	0.75
DmGPx1									0.03	1.05	1.12	0.77	1.04	1.18	1.13	1.17	1.25	0.83	0.74	0.79	0.78	0.78	0.78
DmGPx2										1.06	1.12	0.78	1.04	1.18	1.13	1.17	1.25	0.83	0.74	0.8	0.78	0.78	0.78
HsGPx7											0.69	1	1.03	1.17	1.12	1.16	1.24	1.06	0.97	1.02	1	1	1
HsGPx8												1.07	1.09	1.23	1.19	1.22	1.3	1.12	1.03	1.09	1.07	1.07	1.07
CtGPx													0.99	1.13	1.08	1.12	1.19	0.54	0.45	0.5	0.48	0.48	0.48
HsGPx1														0.39	0.7	0.74	0.81	1.04	0.95	1.01	0.99	0.99	0.99
HsGPx2															0.84	0.88	0.96	1.18	1.09	1.15	1.13	1.13	1.13
HsGPx5																0.32	0.4	1.14	1.05	1.1	1.08	1.08	1.08
HsGPx3																	0.34	1.17	1.08	1.14	1.12	1.12	1.12
HsGPx6																		1.25	1.16	1.21	1.2	1.2	1.2
PvGPx																			0.42	0.48	0.36	0.36	0.36
GnGPx																				0.32	0.37	0.37	0.37
RpGPx																					0.42	0.42	0.42
AcGPx																						0.08	0.08
ApGPx2a																							0

References

Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guillonneau, F., De Pauw, E., Haubruge, E., and Francis, F. (2008). Identification of aphid salivary proteins: a proteomic investigation of Myzus persicae. *Insect molecular biology*, *17*(2), 165-174.

Carolan, J. C., Fitzroy, C. I., Ashton, P. D., Douglas, A. E., & Wilkinson, T. L. (2009). The secreted salivary proteome of the pea aphid Acyrthosiphon pisum characterised by mass spectrometry. *Proteomics*, *9*(9), 2457-2467.

Bos, J. I., Prince, D., Pitino, M., Maffei, M. E., Win, J., & Hogenhout, S. A. (2010). A functional genomics approach identifies candidate effectors from the aphid species Myzus persicae (green peach aphid). *PLoS genetics*, *6*(11), e1001216.

Mutti, N. S., Park, Y., Reese, J. C., & Reeck, G. R. (2006). RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum. *Journal of Insect Science*, *6*(38), 1-7.

Mutti, N. S., Louis, J., Pappan, L. K., Pappan, K., Begum, K., Chen, M. S., Park, Y., Dittmer, N., Marshall, J., Reese, J., & Reeck, G. R. (2008). A protein from the salivary glands of the pea aphid, Acyrthosiphon pisum, is essential in feeding on a host plant. *Proceedings of the National Academy of Sciences*, *105*(29), 9965-9969.

Cooper, W. R., Dillwith, J. W., & Puterka, G. J. (2010). Salivary proteins of Russian wheat aphid (Hemiptera: Aphididae). *Environmental entomology*, *39*(1), 223-231.

Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., Cui, F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J, Reeck, G. R., Wilkinson, T. L., & Edwards, O. R. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (Acyrthosiphon pisum): a dual transcriptomic/proteomic approach. *Journal of proteome research*, *10*(4), 1505-1518.

Rao, S. A., Carolan, J. C., & Wilkinson, T. L. (2013). Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins.*PloS one*, *8*(2), e57413.

Liu, S., Chougule, N. P., Vijayendran, D., & Bonning, B. C. (2012). Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. *PloS one*, *7*(9), e45161.

Akey, D. H., & Beck, S. D. (1972). Nutrition of the pea aphid,< i> Acyrthosiphon pisum</i>: Requirements for trace metals, sulphur, and cholesterol. *Journal of insect physiology*, *18*(10), 1901-1914.

Cui, F., Michael Smith, C., Reese, J., Edwards, O., & Reeck, G. (2012). Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. *Insect Science*, *19*(4), 429-440.

Apostolou, A., Shen, Y., Liang, Y., Luo, J., & Fang, S. (2008). Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced cell death.*Experimental cell research*, *314*(13), 2454-2467.

Lee, A. H., Iwakoshi, N. N., & Glimcher, L. H. (2003). XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Molecular and cellular biology*, *23*(21), 7448-7459.

Brigelius-Flohé, R., & Maiorino, M. (2013). Glutathione peroxidases. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1830*(5), 3289-3303.

Tosatto, S. C., Bosello, V., Fogolari, F., Mauri, P., Roveri, A., Toppo, S., Flohe, L., Ursini, F., & Maiorino, M. (2008). The catalytic site of glutathione peroxidases. *Antioxidants & redox signaling*, *10*(9), 1515-1526.

Blankenberg, S., Rupprecht, H. J., Bickel, C., Torzewski, M., Hafner, G., Tiret, L., Smieja, M., Cambien, F., Meyer, J., & Lackner, K. J. (2003). Glutathione peroxidase 1 activity and cardiovascular events in patients with coronary artery disease. *New England Journal of Medicine*, *349*(17), 1605-1613.

Chu, F. F., Doroshow, J. A., & Esworthy, R. S. (1993). Expression, characterization, and tissue distribution of a new cellular selenium-dependent glutathione peroxidase, GSHPx-GI. *Journal of Biological Chemistry*, *268*(4), 2571-2576.

Yant, L. J., Ran, Q., Rao, L., Van Remmen, H., Shibatani, T., Belter, J. G., Motta, L., Richardson, A., & Prolla, T. A. (2003). The selenoprotein GPX4 is essential for mouse development and protects from radiation and oxidative damage insults. *Free Radical Biology and Medicine*, *34*(4), 496-502.

Olson, G. E., Whitin, J. C., Hill, K. E., Winfrey, V. P., Motley, A. K., Austin, L. M., Deal, J., Cohen, H., & Burk, R. F. (2010). Extracellular glutathione peroxidase (Gpx3) binds specifically to basement membranes of mouse renal cortex tubule cells.*American Journal of Physiology-Renal Physiology*, *298*(5), F1244.

Prabhakar, R., Vreven, T., Morokuma, K., & Musaev, D. G. (2005). Elucidation of the mechanism of selenoprotein glutathione peroxidase (GPx)-catalyzed hydrogen peroxide reduction by two glutathione molecules: a density functional study. *Biochemistry*, *44*(35), 11864-11871.

Bae, Y. A., Cai, G. B., Kim, S. H., Zo, Y. G., & Kong, Y. (2009). Modular evolution of glutathione peroxidase genes in association with different biochemical properties of their encoded proteins in invertebrate animals. *BMC evolutionary biology*, *9*(1), 72.

Missirlis, F., Rahlfs, S., Dimopoulos, N., Bauer, H., Becker, K., Hilliker, A., Phillips, J. P., & Jäckle, H. (2003). A putative glutathione peroxidase of Drosophila encodes a thioredoxin peroxidase that provides resistance against oxidative stress but fails to complement a lack of catalase activity. *Biological chemistry*, *384*(3), 463-472.

Boyko, E. V., Smith, C. M., Thara, V. K., Bruno, J. M., Deng, Y., Starkey, S. R., & Klaahsen, D. L. (2006). Molecular basis of plant gene expression during aphid invasion: wheat Pto-and Pti-like sequences are involved in interactions between wheat and Russian wheat aphid (Homoptera: Aphididae). *Journal of economic entomology*, *99*(4), 1430-1445.

Gatehouse, J. A. (2002). Plant resistance towards insect herbivores: a dynamic interaction. *New phytologist*, *156*(2), 145-169.