DEVELOPMENT OF GENOMIC RESOURCES TO STUDY APHID RESPONSE TO PLANT DEFENSES AND SYMBIOSIS WITH *BUCHNERA APHIDICOLA*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

John Sawyer Ramsey

February 2010

DEVELOPMENT OF GENOMIC RESOURCES TO STUDY APHID RESPONSE TO PLANT DEFENSES AND SYMBIOSIS WITH *BUCHNERA APHIDICOLA*

John Sawyer Ramsey, Ph.D.

Cornell University, 2010

The green-peach aphid *Myzus persicae* (Hemiptera: Aphididae) is a pest of crop plants worldwide. Aphids are phloem-feeding insects, and they inflict damage through direct effects of feeding, and indirectly by transmission of plant viruses. Development of insecticide resistance in aphids has motivated the development of novel control strategies. New approaches to pest control may be informed by an enhanced understanding of the genes allowing aphids to subsist on the nutritionally unbalanced diet of phloem sap, and to detoxify chemical defenses presented by a wide range of hosts. Certain lineages of *M. persicae* have expanded their host range to include tobacco in agricultural settings, and an objective of this research is to characterize differences between adapted and non-adapted lineages in their ability to accept tobacco as a suitable host in the laboratory, and to tolerate tobacco-specific defenses. Tobacco-adapted aphids are highly resistant to nicotine, a potent neurotoxic plant defense, and may be stimulated to feed on nicotine-containing diets by associating nicotine as a characteristic of a suitable host plant.

A platform for functional genomics studies of *M. persicae* has been established by sequencing expressed genes from a diversity of aphid tissues. Bioinformatic analysis of these sequences allowed for the development of an annotated microarray containing fragments of over 10,000 aphid genes, which has been used to study gene expression changes in the heads of tobacco-adapted aphids in response to nicotine. Several induced genes are associated with aphid salivation and detoxification of plant defenses.

The availability of the whole genome sequence for the closely related pea aphid, *Acyrthosiphon pisum*, has facilitated the annotation of *M. persicae* genes, and allowed for comparison of gene content in the two species. Annotation of pea aphid genes involved in nucleotide metabolism revealed an unexpected loss of purine catabolic capability in the pea aphid. Data from genome analysis of *A. pisum* and its bacterial endosymbiont *Buchnera aphidicola* suggests a shared purine metabolic pathway, which may have led to the loss of key aphid purine salvage genes.

BIOGRAPHICAL SKETCH

John Ramsey's introduction to biological research came in 1997, when he participated in a National Science Foundation summer internship program, Marine Models in Biological Research, at the Woods Hole Marine Biological Laboratory. John joined the laboratory of Dr. Joseph Tsien in the Department of Molecular Biology at Princeton in 1998, where he conducted an undergraduate thesis project entitled *Differential Display Analysis of Changes in Gene Expression During Mouse Hippocampal Development.* John received his bachelor of arts degree in Molecular Biology from Princeton University in June 1999. In 2000-2001, John worked as a research assistant in the laboratory of Dr. Paul Garrity at the Massachusetts Institute of Technology, studying axon pathfinding in the development of the eye-brain complex of *Drosophila melanogaster*. John became interested in plant biology while working at Forget-Me-Not Nursery in Indian, Alaska from 2002 to 2004, and he joined the Cornell University Department of Horticulture as a PhD student in 2004. John transferred into the Cornell Department of Plant Biology in 2005, joining the laboratory of Dr. Georg Jander to study plant-aphid interactions.

I dedicate this thesis to my grandfather, William Spencer; my mother and father, Barbara and James Ramsey; my aunts, Anne Spencer and Gail Everett; my godparents, Ellen and Bob Griffith; my brother, David Spencer Ramsey; my wife, Melody Rose Ramsey; and my children, Lena Louise, Zephyr Green, and Errol Augustine.

ACKNOWLEDGEMENTS

I am grateful for the guidance and support I have received on this thesis from my advisor, Dr. Georg Jander, and from my Special Committee members Dr. Jeffrey Scott and Dr. David Lin, and my field appointed committee member Dr. Randy Wayne. I give thanks to past and present members of the Jander lab, particularly Dr. Jae-Hak Kim and Dr. Martin de Vos, for experimental advice, and for creating a stimulating and supportive community. I thank Dr. Angela Douglas for the opportunity to participate in the pea aphid metabolism genome annotation group, and for assistance in manuscript preparation and grant writing. I acknowledge the many contributions of diverse collaborators, including Dr. Qi Sun for bioinformatics support, and Dr. Cecilia Tamborideguy for assistance in learning the programming languages R and Perl. Support for my first year of graduate work was provided by the Cornell University Presidential Genomics Fellowship (now Presidential Life Sciences Fellowship). The Chemistry Biology Interface Training Grant, funded by the National Institutes of Health through the Cornell Chemistry Department, provided funding and training opportunities in my third and fourth years of graduate study. Funding for this work has been provided by USDA grant number 2005-35604-15446 (to Georg Jander).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Ade: Adenine

Ado: Adenosine

AICAR: Aminoimidazole carboxamide ribonucleotide

AMP: Adenosine monophosphate

APRT: adenine phosphoribosyltransferase

Bt: *Bacillus thuringiensis*

dcSAM: decarboxylated *S*-adenosylmethionine (*S*-adenosylmethioninamine)

deoB: Phosphopentomutase

deoD: Purine nucleoside phosphorylase

EST: Expressed sequence tag

GMP: Guanosine monophosphate

gpt: Xanthine-guanine phosphoribosyltransferase

Gua: Guanine

guaC: GMP reductase

Guo: Guanosine

hpt: Hypoxanthine phosphoribosyltransferase

Hyp: Hypoxanthine

IMP: Inosine monophosphate

Ino: Inosine

pfs: 5'-Methylthioadenosine nucleosidase

PLRV: *Potato leafroll virus*

PRPP: Phosphoribosyl pyrophosphate

purA: Adenylosuccinate synthase

purB: Adenylosuccinate lyase

purH: Phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase

SAM: *S*-adenosylmethionine

SNP: Single nucleotide polymorphism

speD: SAM decarboxylase

speE: Spermidine synthase

Xan: Xanthine

Xao: Xanthosine

XMP: Xanthosine monophosphate

Chapter 1: Introduction

Aphids as crop pests and model biological organisms

The study of aphids is motivated by the need to control those species which are crop pests, and by the desire to understand aspects of their biology which are unique among animals. Management of aphid pests on economically significant crop plants has been frustrated by widespread development of insecticide resistance, and novel aphid control strategies are required to address the limitations of current approaches. Aphids feed from the plant phloem, which they target by probing intercellularly with a needle-like stylet. In the course of feeding, aphids inject saliva into the phloem – salivary components are believed to play a critical role in suppressing plant defenses and preventing the sealing of phloem sieve elements (Miles, 1999). Aphid saliva may also contain plant viruses previously taken up from infected plants, and their efficacy as virus transmitters is central to aphids' pest status (Ng and Perry, 2004). The green-peach aphid, *Myzus persicae*, is the world's most effective aphid virus vector, capable of transmitting more than 100 plant viruses. Aphids represent model biological systems to study virus transmission by insect vectors. Additionally, the mutualistic relationship between aphids and the endosymbiotic bacteria *Buchera* represents a model for the study of symbiosis, and the morphological variation characteristic of different life cycle stages provides a model system to study phenotypic plasticity.

Aphid-*Buchnera* symbiosis

Plant phloem sap contains high levels of sucrose and an unbalanced mixture of amino acids, and aphids rely on their endosymbiotic bacteria, *Buchnera aphidicola*, to compensate for the nutritional deficiencies of their diet. The obligate symbiosis between aphids and *Buchnera* dates to the origins of aphids, 200 million years ago

(Moran *et al*., 1993). Symbionts are maternally transmitted, and are housed in specialized organs, called bacteriocytes, in the aphid body cavity (Baumann *et al*., 1995). Treatment of the aphid with antibiotics to kill the symbionts results in a dramatic reduction of aphid growth rate, and *Buchnera*- free aphids are reproductively sterile. Whole genome sequence data is available for several strains of *Buchnera*, and their genomes are among the smallest sequenced thus far (Shigenobu *et al*., 2000). The strain of *Buchnera* from the pea aphid, *Acyrthosiphon pisum*, has a 641 kb genome, containing only 541 genes (Gil *et al*., 2002). This genome reduction is consistent with the fact that mutualistic bacteria may obtain many of the required molecules for life from their hosts, thereby allowing for the loss of unnecessary genes. *Buchnera* lacks genes for the biosynthesis of non-essential amino acids, while it retains essential amino acid metabolism genes (Zientz *et al*., 2004). This has led to the hypothesis that aphids provision *Buchnera* with non-esssential amino acids such as glutamine and glutamate, while *Buchnera* in turn provides essential amino acids, which are present at vanishingly small quantities in plant phloem sap, to the aphid (Douglas, 1998).

Life cycle variation and phenotypic plasticity

The aphid life cycle is characterized by striking developmental polyphenism, and life cycle variations have been reported between different lineages and species, ranging from cyclical to obligate parthenogenesis (Kanbe and Akimoto, 2009). Cyclically parthenogenetic aphid lineages are composed of asexual females which produce eggs genetically identical to themselves during the spring and summer months. These unfertilized eggs develop within the mother, which then gives live birth to first instar female nymphs. Aphid development is hemimetabolous, marked by incomplete metamorphosis lacking a pupal stage, with adults morphologically similar to nymphs arising after four larval stages. In the fall, in response to environmental cues including

cooler temperatures and shortening day lengths, asexual females parthenogenically produce sexual females and males, which mate and form frost-hardy eggs for overwintering. The female aphids which hatch from these eggs in the spring are called fundatrices, and they may produce billions of descendants through the months of asexual reproduction which follow. In mild climates or in greenhouses, aphids may dispense with seasonal sexual reproduction and reproduce parthenogenetically for many years. Obligately parthenogenetic lineages also occur, which have lost the capacity to complete a full life cycle including production of sexual morphs (Wilson *et al*., 2003).

Host plant range

Variation from generalist to extreme specialist exists between aphid species: aphids such as *M. persicae* feed on thousands of plants from over 40 families, while the host range of *A. pisum* is restricted to plants of the legume family (Fabaceae). Host plant selection by aphids is accomplished by an assessment of plant volatile chemicals, surface waxes, the composition of intercellular polysaccharides encountered during probing, and finally phloem sap itself (Troncoso *et al*., 2005). Host plant chemicals may have deterrent or attractive properties. Feeding by *M. persicae* induces the production of the aphid-deterrent 4-methoxyindol-3-ylmethylglucosinolate in *Arabidopsis thaliana*, and a variety of legume-specific alkaloids have been shown to be deterrent to *A. pisum* (Kim and Jander, 2007; Dreyer *et al*., 1985). On the other hand, host-specific acylated flavonoid glycosides from *Vicia angustifolia* stimulate probing behavior in the *Vicia* (Fabaceae) specialist *Megoura crassicauda*, while glucosinolates, secondary metabolites primarily found in plants of the family Brassicaceae, are important feeding stimulants for the specialist cabbage aphid *Brevicoryne brassicaceae* (Takemura *et al*., 2006; Francis *et al*., 2001).

Adaptation to tobacco by *M. persicae*

Infestations of *M. persicae* on cultivated tobacco, *Nicotiana tabacum*, were first reported in Japan over 60 years ago, and since that time tobacco-feeding lineages of the aphid have been reported across the globe (Margaritopoulos *et al*., 2009). Populations specializing on tobacco have unique morphological and genetic characteristics, and have been granted subspecific status as *M. persicae nicotianae* (Blackman, 1987; Margaritopolous *et al*., 2002). Correlations between trichome density and resistance to *M. persicae* have been established within and between *Nicotiana spp.* (Thurston *et al*., 1966; Goundoudaki *et al*., 2003). Secretions from tobacco glandular trichomes contain large amounts of cembratrieneols (CBTols), which are toxic when applied directly to *M. persicae* (Wang *et al*., 2004). Tobacco varietals naturally high in CBTols experience decreased aphid infestation in the field, and similarly, transgenic *N. tabacum* engineered to produce high levels of CBTols are more resistant to *M. persicae* (Wang *et al*., 2004, Wang *et al*., 2001). Nicotine is the major alkaloid found in secretions from the glandular trichomes of tobacco, suggesting a role for this well-characterized plant defense in aphid resistance (Thurston *et al*., 1966). Tobacco-adapted *M. persicae* lineages can survive fumigation by higher concentrations of nicotine than non-adapted lineages, suggesting that the aphid's host range expansion to include tobacco may be associated with the development of nicotine resistance (Devine *et al*., 1996).

Insect nicotine resistance

Nicotine is one of the oldest known botanical insecticides, and its biosynthesis is stimulated in *Nicotiana spp.* by jasmonic acid signals in response to leaf damage (Ballare, 2001). Nicotine is an agonist of the nicotinic acetylcholine receptor, causing overstimulation of the insect central nervous system. Exposure to nicotine results in

twitching, convulsions, and death in susceptible insects.

The lepidopteran tobacco specialist *Manduca sexta* has developed multiple mechanisms for nicotine resistance. These include the upregulation of cytochrome P450 detoxification enzymes in response to nicotine exposure, and the production of nicotine pumps situated at the blood-brain barrier and the malpighian tubules, which prevent the toxin from reaching its target site and actively stimulate its excretion (Snyder *et al*., 1993a; Murray *et al*., 1994).

Nicotine is produced in plant roots from amino acid precursors, and is transported through the xylem to foliar tissue. Resistance to nicotine may have been a critical adaptation allowing *M. sexta* to specialize on tobacco, and nicotine represents a significant barrier to herbivory to non-adapted insects. Analysis of frass excreted by *M. sexta* feeding on nicotine-containing artificial diets revealed complete metabolism of the toxin to cotinine, in a reaction mediated by a P450 enzyme (Snyder *et al*., 1993b). Unlike in humans, where the enzyme mediating the transformation of nicotine to cotinine is well studied (Cyp2A6, see Nakajima *et al*, 1996), a cytochrome P450 enzyme taking nicotine as a substrate has not yet been identified in insects.

Insecticide Resistance in *M. persicae*

Well-characterized insecticide resistance mechanisms have been reported in *M. persicae*, frustrating agricultural control efforts. Amplification of E4 and FE4 carboxylesterase genes and associated overproduction of the corresponding detoxification enzymes is characteristic of widespread populations of *M. persicae*, conferring broad spectrum resistance to organophosphorus, carbamate, and pyrethroid insecticides (Devonshire *et al*., 1998). Target site resistance to organophosphates and carbamates is also conferred by mutations in acetylcholinesterase genes (MACE alleles) and resistance to pyrethoids has become common in *M. persicae* due to spread

of the *kdr* and *s-kdr* alleles encoding insecticide-resistant voltage-gated sodium channels (Foster *et al*., 2003; Anstead *et al*., 2005). Currently, neonicotinoid insecticides such as imidacloprid have proven more robust to the development of aphid resistance – although there is variation in sensitivity to these agents between aphid populations, neonicotinoids have heretofore retained their effectiveness under field conditions (Nauen and Denholm, 2005). Nonetheless, novel aphid control strategies are needed, and the development of future active agents will be informed by enhancing our understanding of the mechanisms by which plants naturally repel aphids (mechanisms which may be enhanced by plant genetic engineering) and by using the tools of genomics and molecular biology to identify unique elements of aphid biology which may represent novel pest control targets.

Aphid genomics

Considerable advances have been made in aphid genomics in the last several years, including sequencing of nearly 170,000 expressed sequence tags (ESTs) from *A. pisum*, development of a gene expression microarray, and most recently, the completion of whole genome sequencing for the pea aphid (Wilson *et al*., 2006; International Aphid Genomics Consortium, 2009; http://www.aphidbase.com). These resources provide unprecedented opportunities to further our understanding of aphid biology. Although the pea aphid has been the target of many of these early initiatives in aphid genomics, *M. persicae* is the more severe pest. In addition, *M. persicae* is able to feed on the model plant *Arabidopsis thaliana*. Development of genomic resources for *M. persicae*, in tandem with the whole genome sequence and mutant resources available for *Arabidopsis*, will allow the study of plant-aphid biology to be conducted using molecular tools to study both sides of the interaction. By enhancing our knowledge of how plants detect and mount effective responses to aphids, and by

studying the aspects of aphid biology which are central to their role as pests, it will be possible to deepen our basic understanding of plant-insect interactions while advancing novel crop protection strategies.

Research Plan

Aphid cDNA sequencing

In the first stage of my project, I will lay the foundation for future *Myzus persicae* functional genomics studies by sequencing fragments of the aphid transcriptome. cDNA libraries will be constructed from a diverse collection of tissues, including aphids from different lineages characterized by natural variation relative to virus transmission, aphids of different developmental stages, aphids collected from different host plants, and aphid salivary glands, digestive systems, and heads. Sequencing of clones from cDNA libraries will generate a collection of expressed sequence tags (ESTs) which will be assembled into contigs to generate consensus sequences likely to represent unique genes (unigenes). Unigenes will be annotated by comparison with sequence data in Genbank. Oligonucleotides suitable for microarray gene expression studies will be designed to correspond uniquely to each *M. persicae* unigene.

Microarray design and fabrication

A variety of new microarray design platforms have become available in recent years, with drastic improvements in array quality leading to increased confidence in experimental data. Microarrays fabricated by *in situ* oligonucleotide synthesis avoid problems associated with spotting pre-synthesized oligos (such as inconsistent spot size, clogged print tips, and mis-identification of spots) by building specified sequences one base at a time in a specifically defined location on the microarray slide.

I will use Agilent's eArray platform to design 60-mer oligos corresponding to each sequence in our unigene set. For a subset of several thousand unigenes, I will design a second unique 60-mer such that those genes will be represented by two independent probes on the microarray. Confidence in the validity of data on differential gene expression will be enhanced if each of these probes is induced by a given experimental treatment.

The role of nicotine in aphid-tobacco interactions

In order to evaluate whether resistance to nicotine is involved in the adaptation of *M. persicae* to tobacco, I will conduct experiments to assess the effect of nicotine on aphid fecundity in artificial diets and *in planta*. Artificial diet experiments will be performed to determine whether tobacco-adapted aphids can survive and reproduce on diets containing higher concentrations of nicotine than aphids from a non-adapted lineage. Tobacco genotypes varying in nicotine composition will be used to determine the significance of nicotine on aphid performance *in planta*. I will use the aphid microarray which I design to identify genes involved in the response to nicotine by tobacco adapted *M. persicae.*

Pea aphid genome annotation project

I will participate in the international effort to annotate the genome of the pea aphid, as a member of the metabolism and detoxification functional annotation groups. The availability of a well-annotated genome for a closely related aphid species will assist in the annotation of *M. persicae* sequences. While annotating genes associated with particular metabolic pathways, I will integrate EST data from different *A. pisum* cDNA libraries with genomic data to determine whether a given genomic DNA sequence is expressed, and whether it is overrepresented in a particular tissue type. This may

provide clues as to the biological significance of the duplication or deletion of certain sequences from the genome. A comparison of detoxification enzymes between the specialist *A. pisum* and the generalist *M. persicae* may reveal differences relating to the greater diversity of plant toxins likely to be encountered by the generalist herbivore.

Microarray collaborations

As part of the collective effort of the aphid community to share genomic data and resources to the greatest extent possible, I will work to advance microarray collaborations between the Jander lab and others. I will maintain the Jander lab eArray worksite to ensure that other researchers can order microarrays and obtain relevant data on the design of the array. I will also assist visiting researchers in experimental design and sample preparation for microarray experiments carried out here at Cornell.

REFERENCES

Anstead JA, Williamson MS, Denholm I (2005). Evidence for multiple origins of identical insecticide resistance mutations in the aphid *Myzus persicae*. *Insect Biochem Mol Biol* **35(3):**249-256.

Ballare CL (2001). Smart plants or stealthy bugs? *Trends Plant Sci* **6(4):**142.

Baumann P, Baumann L, Lai C-Y, Rouhbakhsh D, Moran N, Clark MA (1995). *Ann Rev Microbiol* **49:**55-94.

Blackman RL (1987). Morphological discrimination of a tobacco-feeding form from *Myzus persicae* (Sulzer) (Hemiptera: Aphididae) and a key to new world *Myzus* (*Nectarosiphon*) species. *Bull Entomol Res* **77:**713-730.

Devine GJ, Harling ZK, Scarr AW, Devonshire AL (1996). Lethal and sublethal effects of imidacloprid on nicotine-tolerant *Myzus nicotianae* and *Myzus persicae*. *Pestic Sci* **48**:57-62.

Devonshire AL, Field LM, Foster SP, Moores GD, Williamson MS, Blackman RL (1998). The evolution of insecticide resistance in the peach-potato aphid, *Myzus persicae*. *Phil Trans R Soc Lond B* **353**:1677-1684.

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

Dreyer DL, Jones KC, Molyneux RJ (1985). Feeding deterrency of some pyrrolizidinem, indolizidine, and quinolizidine alkaloids towards pea aphid (*Acyrthosiphon pisum*) and evidence for phloem transport of indolizidine alkaloid swainsonine. J. Chem. Ecol. **11(8)**:1045-1051.

Foster SP, Kift NB, Baverstock J, Sime S, Reynolds K, Jones JE, Thompson R, Tatchell GM (2003). Association of MACE-based insecticide resistance in *Myzus persicae* with reproductive rate, response to alarm pheromone and vulnerability to

attack by *Aphidius colemani. Pest Manag Sci* **59**:1169-1178.

Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A (2003). Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. *Proc Natl Acad Sci U S A* **99(7)**:4454-4458.

Goundoudaki S, Tsitsipis JA, Margaritopoulos JT, Zarpas KD, Divanidis S (2003). Performance of the tobacco aphid *Myzus persicae* (Hemiptera: Aphididae) on Oriental and Virginia tobacco varieties. *Agric For Entomol* **5**:285-291.

International Aphid Genomics Consortium (2009). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*, in review.

Kanbe T, Akimoto S (2009). Allelic and genotypic diversity in long-term asexual populations of the pea aphid, *Acyrthosiphon pisum*, in comparison with sexual populations. *Mol Ecol* **18(5)**:801-816.

Kim, JH, Jander G (2007). *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. Plant J. **49(6)**: 1008-1019.

Margaritopoulos JT, Kasprowicz L, Mallock GL, Fenton B (2009). Tracking the global dispersal of a cosmopolitan insect pest, the peach potato aphid. *BMC Ecol* **9**:13.

Margaritopoulos JT, Tsitsipis JA, Goudoudaki S, Blackman RL (2002). Life cycle variation of *Myzus persicae* (Hemiptera: Aphididae) in Greece. *Bull Entomol Res* **92(4)**: 309-312.

Miles PW (1999). Aphid saliva. *Biol Rev* **74**:41-85.

Moran NA, Munson MA, Baumann P, Ishikawa H (1993). *Proc Roy Soc Lond B* **253**:167-171.

Murray CL, Quaglia M, Arnason JT, Morris CE (1994). A putative nicotine pump at the blood-brain-barrier of the tobacco hornworm. *J Neurobiol* **25(1)**:23-24.

Nakajima M, Yamomoto T, Nunoya K, Yokoi T, Nagashima K, Inoue K, Funae Y,

Shimada N, Kamataki T, Kuroiwa Y (1996). Role of human cytochrome P4502A6 in C-oxidation of nicotine. *Drug Metab Dispos* **24(11)**:1212-1217.

Nauen R, Denholm I (2005). Resistance of insect pests to neonicotinoid insecticides: current status and future prospects. *Arch Insect Biochem Physiol* **58**:200-215.

Shigenobu S, Watanabe H, Hattori M, Sasaki Y and Ishikawa H (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**: 81-86.

Snyder MJ, Hsu E-L, Feyereisen R (1993a). Induction of cytochrome P-450 activities by nicotine in the tobacco hornworm, *Manduca sexta*. *J Chem Ecol* **19(12)**: 2903-2916.

Snyder MY, Walding JK, Feyereisen R (1993b). Metabolic fate of the allelo-chemical nicotine in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* **24**:837-846.

Takemura M, Kuwahara Y, Nishida R (2006). Feeding responses of an oligophagous bean aphid, *Megoura crassicauda*, to primary and secondary substances in *Vicia angustifolia*. *Entomol Exp Appl* **121**:51-57.

Thurston R, Smith WT, Cooper BP (1966). Alkaloid secretion by trichomes of *Nicotiana* species and resistance to aphids. *Entomol Exp Appl* **9**:428-432.

Troncoso AJ, Vargas RR, Tapia DH, Olivares-Donoso R, Niemeyer HM (2005). Host selection by the generalist aphid *Myzus persicae* (Hemiptera: Aphididae) and its subspecies specialized on tobacco, after being reared on the same host. *Bull Entomol Res* **95**:23-28.

Wang E, Hall JT, Wagner GJ (2004). Transgenic *Nicotiana tabacum* L. with enhanced trichome exudate cembratrieneols has reduced aphid infestation in the field. *Mol Breed* **13**:49-57.

Wang E, Wang R, DeParasis J, Loughrin JH, Gan S, Wagner GJ (2001).

Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. *Nat Biotechnol* **19**:371-374.

Wilson AC, Dunbar HE, Davis GK, Hunter WB Stern DL Moran NA (2006). A dual-genome microarray for the pea aphid, Acyrthosiphon pisum, and its obligate bacterial symbiont, Buchnera aphidicola. *BMC Genomics* **7**:50.

Wilson ACC, Sunnucks P, Hales DF (2003). Heritable genetic variation and potential for adaptive evolution in asexual aphids (Aphidoidea). *Biol J Linn Soc Lond* **79**:115-135.

Zientz E, Dandekar T, Gross R (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol Mol Biol Rev* **68**: 745-770.

Chapter 2. Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design

John S. Ramsey[1], Alex C. C. Wilson[2,3], Martin de Vos[1], Qi Sun[4], Cecilia Tamborindeguy[5], Agnese Winfield[6], Gaynor Malloch[6], Dawn M. Smith[5], Brian Fenton[6], Stewart M. Gray[5], and Georg Jander[1,*]

[1]Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853, USA

[2]Department of Biology, University of Miami, Coral Gables, Florida 33146, USA

[3]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

[4]Cornell Theory Center, Cornell University, Ithaca, NY 14853, USA

[5]USDA/ARS, Plant Protection Research Unit, Ithaca, NY 14853, USA

[6]Scottish Crop Research Institute, Invergowrie, Dundee, UK

* Corresponding author: email: gj32@cornell.edu

Abstract

Background: The green peach aphid, *Myzus persicae* (Sulzer), is a world-wide insect pest capable of infesting more than 40 plant families, including many crop species. However, despite the significant damage inflicted by *M. persicae* in agricultural systems through direct feeding damage and by its ability to transmit plant viruses, limited genomic information is available for this species.

This chapter has been published in BMC Genomics, 2007 Nov 16;8:423

Results: Sequencing of 16 *M. persicae* cDNA libraries generated 26,669 expressed sequence tags (ESTs). Aphids for library construction were raised on *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Brassica oleracea, B. napus,* and *Physalis floridana* (with and without *Potato leafroll virus* infection). The *M. persicae* cDNA libraries include ones made from sexual and asexual whole aphids, guts, heads, and salivary glands. *In silico* comparison of cDNA libraries identified aphid genes with tissue-specific expression patterns, and gene expression that is induced by feeding on *Nicotiana benthamiana*. Furthermore, 2423 genes that are novel to science and potentially aphid-specific were identified. Comparison of cDNA data from three aphid lineages identified single nucleotide polymorphisms that can be used as genetic markers and, in some cases, may represent functional differences in the protein products. In particular, non-conservative amino acid substitutions in a highly expressed gut protease may be of adaptive significance for *M. persicae* feeding on different host plants. The Agilent eArray platform was used to design an *M. persicae* oligonucleotide microarray representing over 10,000 unique genes.

Conclusions: New genomic resources have been developed for *M. persicae*, an agriculturally important insect pest. These include previously unknown sequence data, a collection of expressed genes, molecular markers, and a DNA microarray that can be used to study aphid gene expression.

Background

Insects in the order Hemiptera, which includes all insects that feed exclusively or predominantly on phloem sap, currently represent the most significant challenge for agricultural pest management programs. Although transgenic plants producing *Bacillus thuringiensis* (Bt) toxin have achieved resistance to many devastating lepidopteran pests, these crops remain susceptible to infestation by aphids and other

hemipterans. Reduction in insecticide application, concomitant with the widespread cultivation of Bt crops, has resulted in hemipteran pests being the primary insect threat in major agricultural systems [1]. Aphid feeding causes an alteration of plant source-sink relationships [2], the induction of premature leaf senescence [3], secondary pathogen infection through fungal growth on aphid honeydew, and the transmission of plant viruses [4]. Among these, virus transmission by aphids represents the greatest threat for agricultural crops. *Myzus persicae* (green peach aphid), which is capable of transmitting more than 100 plant viruses, is the world's most versatile aphid viral vector [5, 6]. In particular, *M. persicae* is a very efficient vector of *Potato leafroll virus* (PLRV), which can lead to yield reductions of 40-70% in infected fields [7]. *M. persicae* lineages can vary considerably in their PLRV transmission efficiency [8], suggesting that there are lineage-specific genetic factors that influence this trait.

M. persicae has been found on hundreds of mostly dicotyledonous plant species [6]. Given this broad host range, it is not surprising that differences in host plant utilization among *M. persicae* lineages are quite common, and efforts have been made to identify molecular variation that correlates with host range and reproductive life cycle [9, 10]. The best-studied example of such variation is represented by the typically red-colored lineages that are able to thrive on tobacco [11]. These first appeared on tobacco in Japan more than 60 years ago and have spread to all tobacco-growing regions of the world. In the United States, *M. persicae* has been found on tobacco since at least 1947 [12]. Red strains of *M. persicae* were first reported in the United States in 1985, and by 1987 had become the dominant color morph on tobacco [13, 14]. Based on analysis of multiple morphometric factors Blackman (1987) classified tobacco-feeding strains as a separate species, *M. nicotianae.* However, several lines of evidence now indicate that there is gene flow within a single *M. persicae* species complex: (i) DNA genotyping and GC-MS analysis of cuticular

hydrocarbons did not show systematic differences between tobacco and non-tobacco populations of *M. persicae* [15, 16]; (ii) The same autosomal translocation conferring insecticide resistance is found in both types of aphid lineages [17]; (iii) Tobacco-adapted and non-adapted *M. persicae* lineages are fully cross-fertile in the laboratory [18]; and (iv) There is evidence from Japan and Greece that *M. persicae* from both tobacco and other host plants mingle during sexual reproduction in nature [19, 20].

In the laboratory, several plant species are convenient hosts for rearing and studying *M. persicae*. *Brassica oleracea* (cabbage) is commonly employed as a host plant for maintaining aphid cultures. *Arabidopsis thaliana* (Arabidopsis), a well-developed model genetic organism with a fully-sequenced genome, is readily consumed by *M. persicae,* and *A. thaliana* microarray studies have identified genes that are induced or repressed in response to *M. persicae* feeding [21-24]. *Nicotiana benthamiana,* a wild relative of tobacco [25] also serves as a host plant for some lineages of *M. persicae*. Virally induced gene silencing (VIGS) is particularly effective in *N. benthamiana,* permitting rapid screening of individual plants to study the importance in defense against *M. persicae* and other herbivores [26]. *P. floridana* (downy ground-cherry) serves as a model solanaceous plant for studying the transmission of PLRV by *M. persicae* [27].

So far, only limited sequence information and genetic markers are available for the estimated 313 Mb nuclear genome of *M. persicae* [28-31]. However, recent advances in DNA sequencing make it possible to rapidly acquire information about the coding regions of any genome by building complimentary DNA (cDNA) libraries and sequencing expressed sequence tags (ESTs). Here we describe the creation of such an EST database from 16 sequenced *M. persicae* cDNA libraries and the use of these data to make *in silico* predictions of differentially expressed genes, identify single nucleotide polymorphisms (SNPs) between lineages, and develop probes for an

oligonucleotide microarray to study aphid gene expression.

Results

Phenotypic Characterization of *M. persicae* Lineages

Phenotypic characterization and microsatellite genotyping of lineage 2001-12, which was collected in Scotland, has been described previously as clone Type B in [32]. Other *M. persicae* lineages were collected at five sites in the United States, and were characterized to determine their reproductive life cycle (Supplementary Table 2.1). Of the 46 tested aphid lines, eight were holocyclic, five intermediate, 15 androcyclic, and 17 anholocyclic. One lineage remained unclassified because multiple replicates failed to grow. Seven holocyclic *M. persicae* lineages and one anholocyclic tobacco-adapted lineage were genotyped with seven microsatellite markers, and were determined to be genetically distinct (Table 2.1).

Table 2.1: Life cycle and microsatellite genotype of *M. persicae* lineages

				*Microsatellite fragment sizes (base pairs)						
Lineage	**Host**	**Color**	**Lifecycle**	**M86**	**S17b**	**myz25**	**myz3**	**M40**	**myz2**	**myz9**
F001	squash	green	Holocyclic	117 117	161 161	116 142	121 121	123 135	192 206	195 195
F009	Potato	green	Holocyclic	111 129	165 165	116 118	Null	127 127	188 190	209 223
F012	Potato	green	Holocyclic	129 131	165 165	118 118	121 121	123 123	192 202	193 203
G002	Pepper	green	Holocyclic	91 91	161 161	116 118	123 123	123 127	178 178	203 223
G003	Pepper	green	Holocyclic	Null	161 165	116 142	123 123	Null	192 202	209 209
G006	Pepper	green	Holocyclic	117 117	Null	116 120	121 121	123 133	190 202	221 221
G010	Pepper	green	Holocyclic	117 117	165 167	120 142	121 123	123 123	202 206	195 195
USDA	Tobacco	red	Anholocylic	93 101	165 165	118 118	121 121	115 123	192 202	215 219
2001-12	Potato	red	ND	ND	ND	ND	ND	ND	ND	ND

* Two different fragment sizes indicate heterozygosis at this marker; Null = no amplification; ND = not determined.

Seven holocyclic *M. persicae* lineages (Table 2.1) were tested for their ability to

transmit PLRV acquired from detached virus-infected leaves of *P. floridana* to virus-free *P. floridana* plants.

In three independent trials, aphids of clone G006 transmitted PLRV efficiently, whereas aphids from lineage F001 failed to transmit the virus consistently (Table 2.2). In these experiments, the F001 and G006 clones exhibited similar growth rates and fecundity suggesting that the observed differences in transmission are attributable to differences in the clones' capacity to acquire and/or transmit the virus, rather than to differences in the amount of time spent feeding on the infected or uninfected leaves. In control experiments, aphids from both lineages that were transferred from uninfected detached leaves to uninfected plants failed to transmit PLRV. The USDA aphid lineage, which was used as a positive control, transmitted PLRV with 100% efficiency in these experiments (data not shown).

	% Transmission		
Clone	**Trial #1[a]**	**Trial #2[a]**	**Trial #3[b]**
G002	40	60	-
G003	80	-	-
G006[c]	100	100	100
G010	100	-	-
F001[c]	40	0	10
F009	80	-	-
F012	40	40	17

Table 2.2: PLRV transmission by holocyclic *M. persicae* lineages.

[a]Five plants per experiment; [b]Ten plants per experiment, except six plants for F012; [c]Lineages used for cDNA library construction

cDNA Library Construction and Sequencing

As summarized in Table 2.3, 16 cDNA libraries representing a diversity of tissues and developmental stages were constructed from four aphid lineages (USDA, 2001-12, F001, and G006). Aphids were reared on host plants from the Solanaceae and Brassicacae families, as well as on plants with and without PLRV infection. Since sequencing non-normalized libraries showed a high level of redundancy, normalized

cDNA libraries were created to improve the rate of new gene discovery (Fig. 2.1). Although normalization increased the gene discovery rate, it also precluded making inferences about differential gene expression by comparing EST frequencies between these libraries. Altogether, sequencing of the cDNA libraries produced a total of 26,759 high quality sequencing reads, which have been submitted to GenBank (accession numbers: DW010205-DW015017, EC387039-EC390992, EE570018-EE572264, EE260858-EE265165, ES444641-ES444705, ES217505-ES226848, and ES449829-ES451794).

Table 2.3: Description of source tissue for each of 16 *M. persicae* cDNA libraries.

Library	**Clone**	**Tissue Type**	**Host Plant**	**Normalized**	**ESTs**	***Unigenes**
MpW	USDA	Whole body, asexual females	*Arabidopsis thaliana*	No	4798	2136
MpNB	USDA	Whole body, asexual females	*Nicotiana benthamiana*	No	1020	552
MpG	USDA	Digestive Tract	*Arabidopsis thaliana*	No	750	340
MpH	USDA	Head	*Arabidopsis thaliana*	No	746	259
MpSG	USDA	Salivary Glands	*Brassica oleracea*	Yes	3233	2242
MpGnorm	USDA	Digestive Tract	*A. thaliana/N. benthamiana*	Yes	1807	1542
MpHnorm	USDA	Head	*A. thaliana/N. benthamiana*	Yes	2063	1753
Fenton	2001-12	Whole body, asexual females	*Brassica napus*	No	2019	1196
MpGM	G006	Whole body, males	*Brassica oleracea*	Yes	1437	1219
MpGF	G006	Whole body, sexual females	*Brassica oleracea*	Yes	1388	1227
MpGV	G006	Whole body, PLRV infected asexual females	*Physalis floridana*	Yes	1299	1150
MpGVN	G006	Whole body, PLRV free asexual females	*Physalis floridana*	Yes	866	822
MpFM	F001	Whole body, males	*Brassica oleracea*	Yes	1359	1106
MpFF	F001	Whole body, sexual females	*Brassica oleracea*	Yes	1294	1129
MpFV	F001	Whole body, PLRV infected asexual females	*Physalis floridana*	Yes	1328	1189
MpFVN	F001	Whole body, PLRV free asexual females	*Physalis floridana*	Yes	1262	1164
				All Libraries	**26,669**	**10,341**

*For each library, the number of contigs and singletons is combined to indicate how many unigenes are represented in the library. As many contigs are generated by aligning ESTs from multiple libraries, the total number of unigenes from all libraries is less than the sum of the number of unigenes from each individual library.

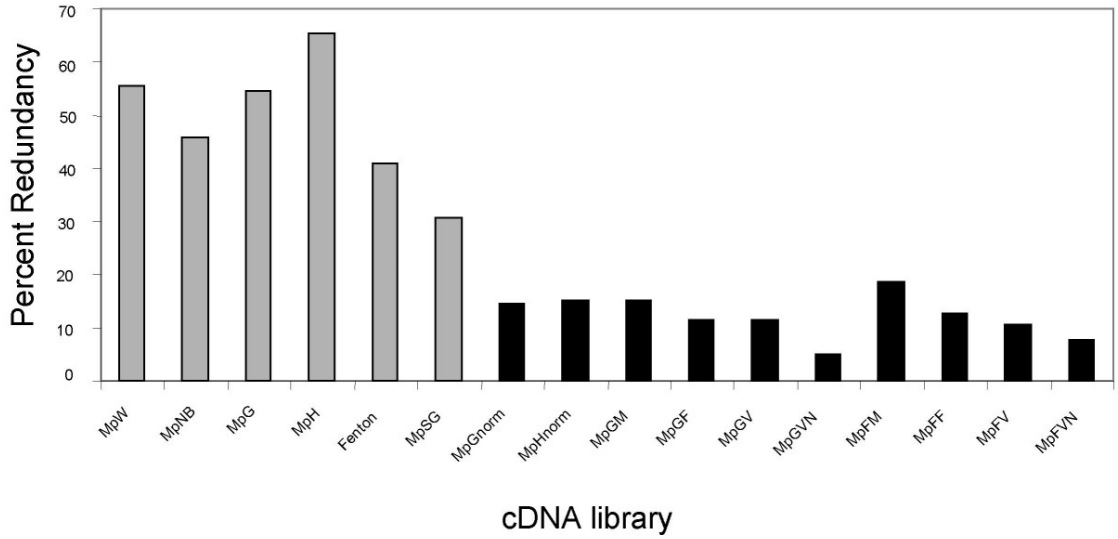


Figure 2.1: Percent redundancy of *M. persicae* cDNA libraries. For each library, percent redundancy = 100 × [1-(#unigenes)/(#ESTs)]. Gray bars represent non-normalized libraries; black bars represent normalized libraries.

Sequence Assembly and Annotation

We identified 3965 contigs and 6376 singletons in the 26,759 high-quality sequences. BlastX (E-value cutoff = 1E-5) was run on all 10,341 unigenes against a database containing all NCBI RefSeq proteins plus the 105 *M. persicae* proteins available at the time in GenBank (January 25, 2007). A one-line annotation was generated for each contig in the following way: (i) if the best Blast hit was for a known *M. persicae* protein, that annotation was used; otherwise (ii) if there was a hit for a *Drosophila melanogaster* protein, that annotation was considered to be most reliable; and (iii) if there were no hits to *D. melanogaster* or *M. persicae* proteins, we used the annotation of the best Blast hit. In addition, the top ten Blast hits are listed for each contig in Supplementary Table 2.2. Information on the ESTs from which each contig is built, including GenBank accession number and source library, is provided (Supplementary Table 2.3). GOslim annotations [33] were tabulated for each contig, and a summary of the molecular function and biological process annotation is provided for the total EST collection and separately for the ESTs from the non-normalized libraries (Fig. 2.2).

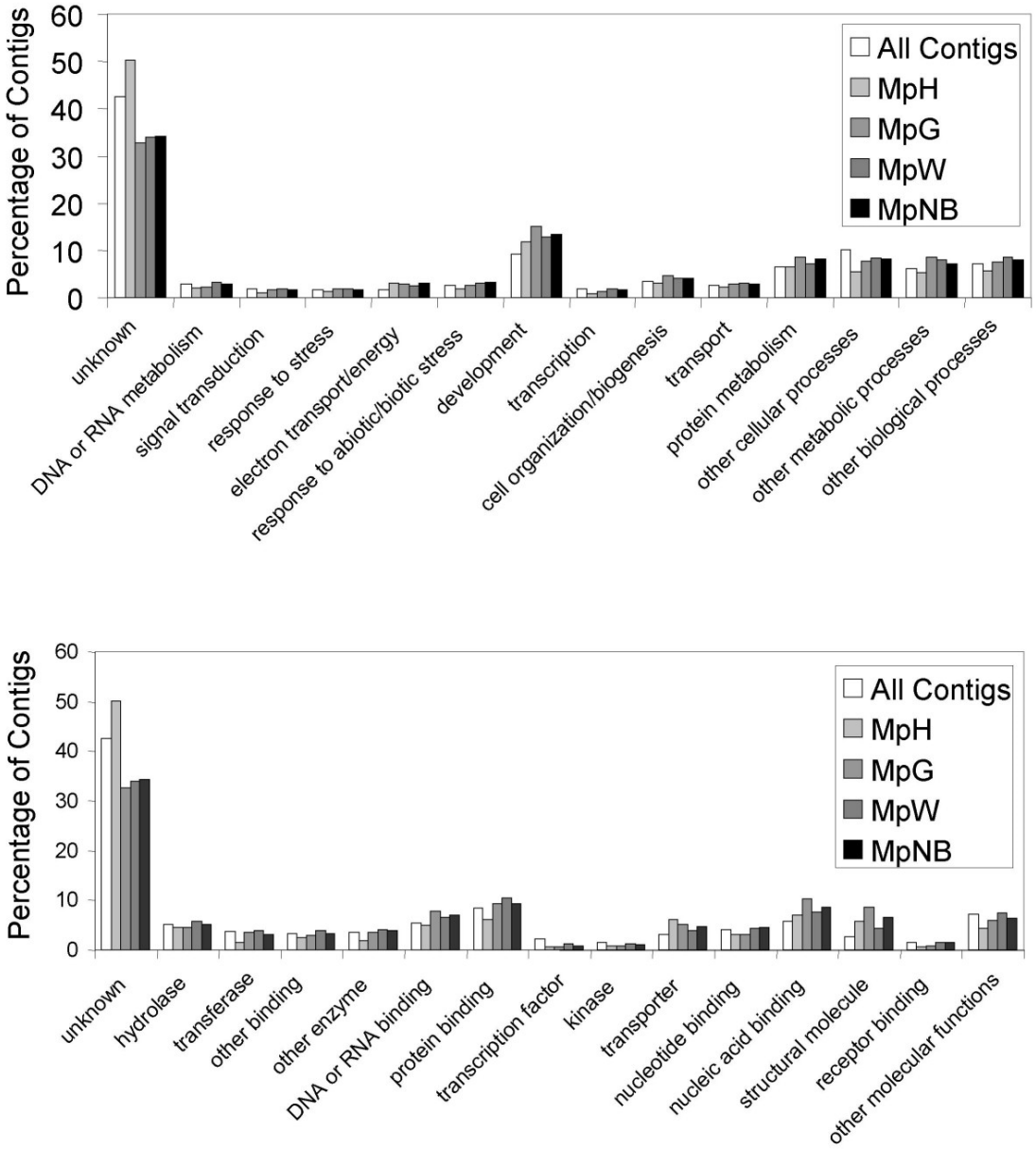


Figure 2.2: (A) Distribution of Gene Ontology biological process annotations, compared between non-normalized libraries. (B) Distribution of Gene Ontology molecular function annotations, compared between non-normalized libraries.

The top twenty contigs in terms of representation in the EST collection can be found in Table 2.4, ranked according to the number of ESTs they contain. Seven of the most highly expressed genes in *M. persicae* have no significant similarity to any

proteins in the RefSeq database [34]. We compared our sequences to the mitochondrial genome of the aphid *Schizaphis graminum* (greenbug) and identified 880 ESTs likely to represent mitochondrial genes. Of these, 491 constitute four contigs that are among the most highly expressed genes in our database (Table 2.4).

Table 2.4: RefSeq annotation for the twenty *M. persicae* contigs composed of the largest number of ESTs.

Rank	Contig ID	Number of ESTs	Gene Description	Predicted Localization
1	1	268	12S small subunit ribosomal RNA gene	Mitochondrial
2	1194	155	unknown protein	Nuclear
3	2319	148	cytochrome c oxidase subunit III	Mitochondrial
4	3100	100	unknown protein	Nuclear
5	3675	67	ATP synthase F0 subunit 6	Mitochondrial
6	3801	60	ribosomal protein, large subunit	Nuclear
7	3543	59	cytochrome c oxidase subunit I	Mitochondrial
8	8	59	tentative cuticle protein	Nuclear
9	130	51	unknown protein	Nuclear
10	375	47	unknown protein	Nuclear
11	1079	43	unknown protein	Nuclear
12	1313	42	ribosomal protein S11	Nuclear
13	844	40	unknown protein	Nuclear
14	1429	39	ribosomal protein L10	Nuclear
15	3408	39	unknown protein	Nuclear
16	1768	37	ribosomal protein L11	Nuclear
17	1881	37	ribosomal protein S24	Nuclear
18	731	37	ribosomal protein S8	Nuclear
19	495	37	muscle LIM protein	Nuclear
20	254	36	cathepsin B	Nuclear

We compared the 10,341 EST contig sequences of *M. persicae* with 17,069 contigs of *Acyrthosiphon pisum*. Using BLASTn (e-value cutoff 1E-10), there were 5513 *A. pisum* contigs with BLAST hits in *M. persicae* contigs, and 5598 *M. persicae* contigs with BLAST hits in *A. pisum* contigs. The low overlap between the contig sequences may be due to the fact that neither EST database represents the whole transcriptome. However, it is likely that many of the contigs in one aphid which do not have homologues in the database of the other aphid species may represent genes

responsible for the adaptation to specific host plants, or for other differences in physiology which have evolved since the species' divergence. There is significant similarity between predicted coding regions in a subset of *M. persicae* and *A. pisum* genes. From more than 5000 shared contig sequences, 1585 have greater than 95% identity in the coding region (coding regions are defined as open reading frames of 50 residues of longer that can be aligned to Uniprot sequences with BlastX values less than 1E-10). These genes are likely to represent cellular housekeeping genes, but may also include genes which are essential for aspects of plant-insect interactions which are specific to, and common among, aphids.

Comparision of *M. persicae* ESTs to other Genomic Data

Similarity searches against other aphid ESTs, performed using the TBlastX program (Evalue cutoff 10E-10), identified 4500 unigenes with no similarities to previously described aphid ESTs. Therefore, these unigenes represent newly described aphid cDNA sequences. However, some of these sequences may arise from untranslated regions of genes, which may not be highly conserved between species. These 4500 unigenes were subsequently compared with the *nr* database (non-redundant NCBI protein and nucleotide database) using TBlastX and BlastX programs. A total of 2423 unigenes had no hits at Evalue <10E-10. Some subset of these may represent *M. persicae*-specific genes. The other 2077 unigenes represent "new-to-aphids" features: genes identified in non-aphid species, which are not represented among the over 80,000 ESTs from five other aphid species that were previously submitted to GenBank.

Blasting the 959 *M. persicae* ESTs previously described by Figueroa *et al.* [31] against our database revealed that 806 of these ESTs match up to 520 of our unigenes at E-values<1E-100. Based on this stringent cut-off value, 153 of these previously

described *M. persicae* ESTs were not found in our data set – however, 109 of these ESTs have matches at E-values<1E-20, indicating that they may represent closely related gene family members.

BlastN (E-value cutoff = 1E-20) of all ESTs against the three *Buchnera aphidicola* genomes available in GenBank, as well as BlastX (percentage identity cutoff 80%) against all Buchnera aphidicola protein sequences in GenPept identified 90 sequences that are almost certainly from the bacterial endosymbiont of *M. persicae, Buchnera aphidicola*. The top Blast hit for five of our unigenes (all singletons) are for *B. aphidicola* proteins, indicating that our filtering failed to remove a small number of contaminating bacterial sequences. No *Escherichia coli* or *Saccharomyces cerevisiae* proteins appeared as a top-ten Blast hit for any of our unigenes giving us confidence that significant contamination from these sources is not a concern.

In silico prediction of differentially expressed genes

We used the previously described R statistic [35] to identify the contigs showing the greatest differences in EST abundance among four of the non-normalized libraries (MpH, MpG, MpNB, MpW). A log likelihood ratio statistic was calculated that estimated the extent to which differences in gene expression correspond to the heterogeneity of the libraries. The twenty top hits of differentially expressed genes are presented with a brief description of the protein, the value of the R statistic, and the abundance of the gene in each of the four libraries (Table 2.5). Among the twenty contigs showing the highest R value, twelve represent genes that are over-expressed in the head library. None of these genes show similarities to published proteins with known function (at E =1E-5), and ten of the twelve genes were found only in head or full body cDNA libraries made from Myzus *persicae* or other aphids. The six contigs representing genes that were most highly expressed in the gut library include two with

no homology to GenBank sequences. Two other contigs show similarity to the lysosomal cysteine protease cathepsin B-N.

Table 2.5: USDA lineage contigs with library-specific expression patterns in non-normalized libraries.

Contig ID	MpG[a]	MpH[b]	MpNB[c]	MpW[d]	Total	R_{tot}[e]	Homology	E-value
1079	5	0	**28**	9	42	34.4	unknown protein	NA
3260	0	1	**18**	2	21	28	unknown protein	NA
8	0	**42**	1	13	56	68.2	unknown protein	NA
3100	0	**35**	0	5	40	66.9	unknown protein	NA
613	0	**28**	1	1	30	57.5	unknown protein	NA
375	0	**17**	0	1	18	35.3	unknown protein	NA
3414	0	**15**	0	0	15	34.2	unknown protein	NA
2753	0	**18**	0	3	21	33.7	unknown protein	NA
3104	0	**20**	1	6	27	31.8	unknown protein	NA
3194	0	**17**	0	3	20	31.6	unknown protein	NA
3319	0	**13**	0	2	15	24.6	unknown protein	NA
2148	0	**10**	0	0	10	22.8	unknown protein	NA
614	0	**9**	0	0	9	20.5	unknown protein	NA
844	0	**10**	0	3	13	17	unknown protein	NA
1196	**23**	0	0	2	25	46.2	glutathione S transferase S1	1E-53
130	**33**	0	3	15	51	46.1	unknown protein	NA
448	**16**	0	0	1	17	33	unknown protein	NA
254	**20**	0	0	11	31	30	cathepsin B-N	0
3427	**16**	0	1	2	19	29	densovirus structural protein	1E-46
256	**9**	0	0	0	9	20.5	cathepsin B-N	1E-161

[a]MpG = digestive tract (gut) library; [b]MpH = head library; [c]MpNB = whole body library, reared on *N.* benthamiana; [d]MpW = whole body library, reared on *A. thaliana*; [e]R_{tot} = likelihood ratio statistic [34].

Contig 3427 shows similarities to a structural protein from densoviruses, which have recently been described as infecting the stomach cells of aphids [36]. Contig 1196, which represents a gene that is more highly expressed in the gut, shows similarity to a glutathione S-transferase (GST). GSTs belong to a large family of proteins implicated in xenobiotic detoxification, and an increase in GST activity has been associated with the adaptation to plant secondary metabolites in *M. persicae* [37]. Two contigs with no homology to known genes have a significant overrepresentation of ESTs from aphids

reared on *N. benthamiana* rather than *A. thaliana*. One of these contigs, number 1079, also contains five ESTs from the digestive tract library and none from the head library, suggesting this gene as a candidate for involvement in aphid response to tobacco-specific defenses.

Prediction of Secreted Salivary Proteins

In order to find aphid proteins involved in the successful infestation of host plants, we have identified cDNA sequences that are predicted to encode for secreted proteins expressed in the salivary glands. These proteins may be required for the establishment of prolonged phloem feeding and suppression of plant defenses. Using stringent criteria (see Materials and Methods) we identified 186 contigs representing sequences expressed in salivary glands. Subsequent *in silico* translation and signal peptide prediction resulted in the identification of 45 *M. persicae* proteins that may be secreted from the salivary glands (Supplementary Table 2.4). These include a total of fifteen proteins that are predicted to possess an anchor sequence (Supplementary Table 2.4), indicating that these proteins remain in the cell membrane upon secretion and might function as receptors or proteins involved in transport. For instance, the protein encoded by contig 515, is a close homologue of tetraspanin 29FA in *D. melanogaster,* where it functions as a cell surface receptor binding protein involved in signal transduction (Mi *et al*., 2003 - personal communication to FlyBase, http://flybase.bio.indiana.edu/).

SNP Identification and Validation

SNPs are effective molecular markers for genetic mapping and can also be used to estimate the level of sequence divergence between lineages. Using the program PolyBayes [38], we identified 12,722 potential SNPs from our EST sequences. Since

we were interested in identifying SNPs that represented differences between, rather than within, lineages, we filtered our list of polymorphisms to include only those SNPs representing a nucleotide difference between two lineages, and exhibiting no apparent heterozygosity within lineages. This resulted in ~800 polymorphisms that can serve as potential molecular markers for differentiating aphid lineages (Supplementary Table 2.5).

Many of the predicted SNPs are represented by only a single EST in one or more lineage, allowing for the possibility that observed sequence differences are artifacts resulting from an error in reverse transcription of the mRNA, PCR amplification of the cDNA, or the sequencing reaction. Therefore, we generated a list of high-confidence polymorphisms in which each of any two lineages was represented by two or more ESTs with the same base at the polymorphic position. This resulted in 167 high-confidence SNPs (Supplementary Table 2.6), from which we selected a small subset to validate by re-sequencing genomic DNA from *M. persicae* lineages USDA, F001, and G006.

Table 2.6: SNP validation by re-sequencing chromosomal DNA

Contig ID	SNP ID	Base Position	Validated
129	1849	230	No
254	7292	1023	Yes
254	7293	1045	Yes
254	7294	1065	Yes
1080	543	850	Yes
3202	11100	409	No
3202	11103	554	No
3285	11751	162	No
3285	11752	204	No
3347	12372	256	No
3713	15695	228	Yes

Eleven SNPs from seven contigs were selected for validation – contigs contained either one, two, or three predicted polymorphisms (Table 2.6). No sequence differences between these lineages were detected when sequencing 158 bp of a control

gene, EF-1α (accession numbers EF660853-EF660855). Five of the 11 tested SNPs were confirmed by resequencing (Table 2.6). Four of these confirmed SNPs represented differences between the two green aphid lineages, F001 and G006, and the red tobacco-adapted USDA lineage. Moreover, three of these SNPs were in the open reading frame of contig 254, which is annotated as a lysosomal cysteine protease cathepsin B-N.

Contig 254 is one of the most highly expressed genes (Table 2.4) and differentially expressed in the aphid digestive tract (Table 2.5). Two of these cathepsin SNPs represent non-synonymous, non-conservative amino acid changes (Fig. 2.3), indicating a possible functional change in enzyme activity.

Figure 2.3: SNPs in contig 254, cathepsin B, a putative gut-specific cysteine protease. Three SNPs were validated by sequencing genomic DNA. Three SNPs in the gene result in one silent change (Leu-Leu), one substitution of a negatively charged amino acid for an aliphatic amino acid (Ala-Glu) and one substitution of an aromatic for an aliphatic amino acid (Leu-Phe) in the green *M. persicae* lineages (G006 and F001) relative to the red USDA lineage.

Structural modeling with the protein fold recognition program Phyre [39] indicates that these residues are in a conserved region in an exposed loop on the surface of the protein. One EST from the second red *M. persicae* lineage (2001-12), shared the three polymorphic nucleotides with the red USDA lineage. Therefore, it may be informative to genotype a wider range of red and green aphids at this locus to determine whether these polymorphisms correlate with differences in host range or life cycle between

lineages. Although it is tempting to speculate that this gut-specific protease is undergoing rapid evolution in order to avoid plant protease inhibitors, the small size of our SNP dataset and the high (>50%) occurrence of false positives prevent us from inferring the significance of changes in this protein arising from variation between these lineages.

Microarray Design

The Agilent eArray platform was used to design a microarray based on our ESTs and an additional 1,121 ESTs from other sources that were available in GenBank (including [31]. Of the total of 10,525 unigenes assembled from these ESTs, we successfully designed 60-mer probes for 10,478 using eArray software. For >95% of the unigenes, three 60-mer probes were designed, corresponding to different regions of the gene. The actual synthesized array consists of one probe group representing all 10,478 unigenes, a second probe group with alternate 60-mers for 4139 of the unigenes, 11 ESTs from *Schizaphis graminum* (greenbug), negative controls corresponding to plant and human specific genes, and positive controls representing insect housekeeping genes. The current slide layout consists of eight arrays of >15,000 elements each, permitting comparison of two treatments with four-fold replication on each slide.

Discussion

Genomic comparisons

Our identification of 26,669 *M. persicae* ESTs (Supplementary Table 3) from 16 cDNA libraries extends previous sequencing efforts for this species [31] 25-fold, and contributes to the rapidly expanding resources that are available for aphid genomics. In addition to the described *M. persicae* data, the GenBank database contains 66,298

ESTs for *Acyrthosiphon pisum* (pea aphid; [40]), 8344 for *Aphis gossypii* (cotton aphid; [41]), 4263 for *Toxoptera citricida* (brown citrus aphid; [42]), 959 for *M. persicae* [31], and 458 for *Rhopalosiphon padi* (bird cherry-oat aphid; [43]). Sequencing of the *A. pisum* genome is ongoing [40], and a comprehensive database for all aphid genomics information has been established (http://www.aphidbase.com; [44]). Functional analysis of aphid genes that are identified by sequencing or expression studies will be facilitated by the recent demonstration that it is possible to silence aphid gene expression by RNA interference [45].

The broad selection of source material for cDNA library construction (Table 3) permitted sequencing of ESTs representing genes expressed at different developmental stages and morphs, as well as genes expressed in response to viral infection, and alternate host plant utilization. In addition, the production of separate libraries from heads, digestive tracts, and salivary glands ensured that genes of special interest to the study of plant-aphid interactions are well-represented in our database. Our comparison of EST frequencies between non-normalized libraries enabled *in silico* prediction of differential gene expression (Table 2.5). Clustering of the ESTs from our first few libraries (Figure 2.1) indicated a high degree of redundancy. We responded to this by normalizing all subsequent libraries, which significantly increased our rate of new gene discovery but eliminated our ability to make inferences about differential expression between libraries. Therefore, it was advantageous to our project to make both types of library, preserving in some cases the natural transcript ratios present in the source tissues, and in others bringing the representation of housekeeping genes more in line with that of rarely expressed transcripts.

Virus-derived genes

Because aphids transmit plant viruses, and are themselves infected by

entomopathogenic viruses, we searched our database for sequences with homology to known viral genes. No ESTs with homology to *Potato leafroll virus* (PLRV) were identified in our database, even in libraries made from aphids feeding on PLRV-infected plants. This absence of PLRV cDNA sequences is consistent with the fact that PLRV does not replicate within the aphids.

Five contigs are annotated as densovirus proteins, including one predicted to be specific to the aphid digestive tract (Table 2.5). All but one of the densovirus ESTs are from the USDA lineage, but this could well be an artifact relating to the fact that the gut cDNA library was made from this aphid strain. A densovirus has been reported to infect the anterior portion of the digestive tract of *M. persicae*, with infected aphids characterized by reduced size, delayed development, and decreased fecundity [36]. Densoviruses represent potential biological pest control agents, and similar viruses from the families *Baculoviridae* and *Tetraviridae* have been commercialized for this purpose.

When the stringency of the BlastX search was reduced to E-value cutoff = 1E-4, one unigene (contig 3464, Supplementary Table 2) has a Dasheen mosaic virus (DsMV) polyprotein as its best hit. DsMV is a non-persistent RNA virus known to be transmitted by *M. persicae* ([46]). Over two-thirds of the 25 ESTs in the contig homologous to the DsMV polyprotein are derived from the salivary gland or head libraries, consistent with the fact that these non-circulative viruses are retained within the mouthparts of their aphid vectors. The relatively large number of DsMV-derived ESTs, which were found in seven different libraries from three lineages (F001, G006, and USDA), is unexpected in light of the fact that this virus should not replicate within the aphids, and none of the plants used for aphid rearing showed obvious signs of viral infection. Furthermore, the host range of DsMV is not known to overlap with the host plants used in these experiments.

Functional significance of annotated unigenes

In cruciferous plants, myrosinase enzymes (β-thioglucosidases, EC 3.2.1.147) initiate the rapid breakdown of glucosinolates into insect-deterrent hydrolysis products during herbivory. However the aphids *Brevicoryne brassicae* (cabbage aphid) and *Lipaphis erysimi* (turnip aphid) have co-opted this defensive system by sequestering plant-derived glucosinolates and producing their own myrosinase as a defense against predators [47-50]. One EST from our database (accession number ES221351, from the G006 lineage) has significant homology to the *B. brassicae* myrosinase gene. Although attempts to measure myrosinase activity in *M. persicae* have been unsuccessful, it is notable that aliphatic rather than indole glucosinolates were used as enzymatic substrates in these experiments [51]. Aliphatic glucosinolates are recovered intact in the honeydew of *M. persicae* on *A. thaliana*, showing that these aphids are able to avoid or inactivate plant myrosinases. In contrast, *A. thaliana* indole glucosinolates are largely broken down within the aphids [52]. Although this glucosinolate breakdown may occur by a non-enzymatic mechanism, it is also possible that *M. persicae* possesses a myrosinase activity that is specific to indole rather than aliphatic glucosinolates.

The genetic mechanisms regulating the cyclically parthenogenetic life cycle characteristic of most aphids are largely unknown. Environmental cues, including shortening days, triggers development of sexual morphs in the autumn [53]. A gene from *Acyrthosiphon pisum*, ApSD1, with similarity to a protein involved in amino acid transport in GABAergic neurons, was overexpressed in pea aphids reared under short photoperiod conditions [54]. We identified one EST, which we had annotated as an amino acid transporter, as being significantly similar to ApSD1. This EST (accession number EC388175) was sequenced from the G006 male library, which is consistent

with a role for this amino acid transporter in the development of winged sexual morphs.

M. persicae has evolved to tolerate plant allelochemicals and insecticides by diverse strategies, including amplification of E4 esterase genes [55], point mutations in insecticide targets [56], and increased activity of glutathione S-transferases in response to glucosinolates in artificial diets [37]. Out of 11 contigs with significant homology to *M. persicae* esterases, two contigs (720 and 3118) from our database are nearly identical to the *M. persicae* E4 esterase (GenBank Accession CAA52648), whereas nine others appear to represent different genes. These nine sequences may have evolved following amplification to acquire novel functions in the hydrolysis of plant secondary metabolites encountered during the expansion of the insect's host range, or in the breakdown of newly developed insecticides. Other potential detoxification genes represented in our database include 24 glutathione S-transferases and 53 cytochrome P450s (Supplementary Table 2).

Among the 168 salivary gland contigs that are predicted to encode secreted proteins (Supplemental Table 2.3), approximately 62% are of unknown function. However, others could have potential function in aphid virulence based on their homology to known proteins. For instance, contig 1300 encodes a protein that belongs to an insect-specific family that includes the yellow proteins of *D. melanogaster*, that are involved in cuticular development and behavior [57], and the major royal jelly proteins of *Apis mellifera* (honeybee). *A. mellifera* proteins from this family are high in essential amino acids and comprise up to 90% of the total protein content of the jelly that is fed to developing larvae [58]. Although major royal jelly proteins are thought to be produced in the cephalic glands of nurse bees [59], another member of this protein family (MRJP 8) was recently identified as a component of the honeybee venom [60]. In *M. persicae,* the homologous protein is less abundant, and ESTs were

only found in the salivary gland and normalized head libraries (MpSG and MpHnorm in Table 2.3). Nevertheless, it is tempting to speculate that the protein has a virulence function in aphids. Two other genes expressed in salivary glands, represented by contigs 2422 and 3025, are predicted to encode secreted proteins that play a role in proteolysis, and therefore could have interesting functions in the interaction between *M. persicae* and its host plants. Contig 2422, which has highest homology to a sequence of unknown function from *D. melanogaster* (GenBank accession NP_611740), encodes a protease-associated domain. Contig 3025 encodes a protein with homology to *Der1*, a gene involved in the degradation of misfolded proteins in yeast [61].

DNA sequence polymorphisms

Comparison of ESTs from the three *M. persicae* lineages identified a large number of potential sequence polymorphisms which were subjected to stringent post-processing to reduce sequencing artifacts. The remaining 167 SNPs, represented by multiple ESTs in more than one aphid lineage (Supplementary Table 2.6), are a good data source for the identification of *M. persicae* genetic markers. Furthermore, as suggested by the cathepsin B-N sequence data (Figure 2.3), these polymorphisms may provide clues about functional divergence of proteins in different *M. persicae* lineages.

However, when we re-sequenced 11 of these SNPs from genomic DNA templates, only about half were confirmed (Table 2.6), suggesting that many potential sequence differences in our EST collection are the result of errors created during reverse transcription, PCR amplification, or sequencing. This highlights the importance of developing effective criteria to select a list of high-confidence SNPs from the large number of polymorphisms predicted by programs such as POLYBAYES, and of validating predicted polymorphisms by re-sequencing of

genomic DNA.

Microarray Development

Given the greater reproducibility of gene expression data collected with oligonucleotide microarrays, as opposed to spotted cDNA microarrays, we decided to develop oligonucleotide microarrays for future studies on *M. persicae* gene expression [62]. The highest quality microarrays currently available are those fabricated by *in situ* oligonucleotide synthesis, a technology pioneered by Agilent. When using such arrays, the number of required technical replicates is reduced because of the high degree of reproducibility between spots, allowing the user to concentrate resources on analyzing biological replicates. In addition, the high cost of purchasing synthesized oligonucleotides makes traditional custom printing of high density arrays at core facilities feasible only if many arrays will be made. There are no up-front costs to design microarrays on Agilent's eArray platform, and the minimum number of slides to order is one.

Transcriptional profiling with microarrays is a powerful technique for identifying genes involved in the response of an organism to its environment. We anticipate that *M. persicae* microarrays can be used to answer a variety of fundamental questions about aphid biology and plant-aphid interactions. Genes critical to the status of this insect as an agricultural pest can be identified by studying expression changes induced by different crop plants and in response to virus infection. Research on aphid genes specifically expressed in salivary glands may identify proteins that prevent clogging of sieve elements or otherwise contribute to the phloem-specific feeding style of aphids. Conversely, these salivary proteins likely also provide phloem-specific cues that allow plants to recognize aphid feeding and mount a defense response. Microarray experiments will allow association of gene expression changes with polyphenism, the

development for morphologically different individuals (*e.g.* winged and unwinged) that are otherwise genetically identical. Analysis of gene expression in aphids feeding on artificial diets or plants with altered amino acid content can identify genes that are critical for the interaction with endosymbiotic *B. aphidicola* bacteria, which synthesize essential amino acids and allow aphids to survive on the otherwise nutritionally imbalanced phloem sap.

The broad host range and differences in host plant preferences among individual lineages of *M. persicae* are some of the more interesting aspects of the biology of this insect. Gene expression differences that underlie within-species variation can be identified by microarray analysis. By sequencing cDNA libraries made from aphids that were raised on both Solanaceae and Cruciferae, we have increased the probability that future microarray experiments performed by ourselves and others will include aphid genes that are expressed only under these particular growth conditions. Evidence for such regulated gene expression comes from our non-normalized libraries, which included two genes that were overrepresented among ESTs from *N. benthamiana* in comparison to *A. thaliana* (Table 2.5). DNA microarray experiments will almost certainly identify additional genes with host plant specific expression patterns. Further research on the function of such differentially expressed genes will illuminate adaptations that have allowed some *M. persicae* lineages to expand their host range to include tobacco. Other *M. persicae* lineages, which show differences in their ability to reproduce on *A. thaliana* (J. Kim and G. Jander, unpublished results), can be studied to identify aphid adaptations for feeding on Cruciferae. In addition, microarray experiments with *M. persicae* feeding on *A. thaliana* will provide the unique opportunity to simultaneously study gene expression changes on both sides of a plant-insect interaction.

Given the broad range of questions that can be addressed by microarray

analysis of *M. persicae* gene expression, the Agilent microarray that we have developed will be of broad interest to aphid researchers. Although the technology necessary for hybridizing and scanning synthesized Agilent arrays is somewhat different from that used for experiments with spotted oligonucleotide arrays, it is available at many universities. The microarrays described here will be made available at cost to other researchers and can be obtained by contacting the corresponding author (G.J.).

Conclusions

By sequencing and analyzing 26,669 *M. persicae* ESTs, we have generated new genomic resources for this aphid species. Expressed aphid genes, in particular many that show no significant similarity to genes from other organisms, have been identified. Molecular markers that were found by comparing three aphid lineages will be useful not only for genotyping natural isolates, but also future genetic studies with *M. persicae.* The DNA microarray that has been developed will permit further investigation of agriculturally and ecologically relevant transcriptional regulation in *M. persicae.*

To date, the lack of genomic resources for *M. persicae* has stood in stark contrast to the threat posed by this aphid to agricultural systems worldwide. By studying aphid gene expression responses to virus infection, different host plants, and other stresses, it will be possible to obtain a better understanding of this important biological interaction. In addition, our increasing understanding of plant molecular responses to phloem-feeding insects will be complemented by elucidation of the adaptations that allow these insects to establish compatible interactions with their host plants. Further research on *M. persicae* gene expression responses will aid in efforts to

breed crops with increased aphid resistance and will advance ongoing research into aphid ecology, evolution and physiology.

Methods

Aphid Collection, Rearing, and Characterization

M. persicae lineages were collected from five sites in the United States and one in Scotland, as described in Table 2.1 and Supplementary Table 2.1. Aphid lineages were started from a single greenhouse- or field-collected insect. Colonies were reared on *B. oleracea* var. 'Wisconsin Golden Acres' in growth chambers (16:9 h light:dark cycle, 150 μmols $m^{-2}s^{-1}$, at 24°C $\pm$ 1 day, 19°C $\pm$ 1 night, 50% relative humidity). For library construction, aphids from the USDA lineage were reared under the same conditions on *A. thaliana* (land race Columbia-0), *N. benthamiana*, and *B. oleracea* var. 'Wisconsin Golden Acres'. Asexual females of lineages F001 and G006 were reared on PLRV-infected or PLRV-free *P. floridana* (15:9 h light:dark cycle at 24°C). For induction of sexual morphs, aphids were reared on *B. oleracea* var 'Wisconsin Golden Acre' in growth chambers under short day conditions (13:11 h light:dark cycle, 115 μmol/m^2/s light intensity, 18°C $\pm$ 2) in Percival (Perry, IA, USA) Model I36LLVLC8 growth chambers. The 2001-12 lineage was maintained on *B. napus* var. 'Mascot' (16:8 h light:dark cycle at 18°C $\pm$ 2).

M. persicae lineages were genotyped in a single multiplex PCR reaction containing dye-labelled primers (Supplementary Table 2.7) to amplify seven microsatellite loci: M86 and M40 [29] and S17b, myz2, myz3, myz9 and myz25 [30]. One primer for each locus was fluorescently labeled with the following dyes; NED™: S17b-forward and myz25-forward, 6-FAM™: M86-forward, VIC™: myz3-forward, myz9-forward, M40-forward and PET™: myz2-reverse (Applied Biosystems, Foster

City, California, USA). PCR reactions were carried out with 10 µl reaction volumes containing 0.5 units of *Taq* DNA polymerase (Eppendorf, Hamburg, Germany), 50 mM KCl, 10 mM Tris-HCL pH 8.3, 2.0 mM Mg^{2+}, 200 µM of each dNTP, 125 pM of primers myz2f, myz2r, myz9f, myz9r, M40f, M40r, M86f, M86r, myz3f and myz3r, 62.5 pM of primers S17bf, S17br, myz25f and myz25r and approximately 5 ng of template DNA. PCR reactions were run using 5-dye chemistry on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, California, USA) by the Genomic Analysis and Technology Core facility at the University of Arizona. Fragment analysis was completed using STRand software (http://www.vgl.ucdavis.edu/informatics/STRand/index.html).

For life cycle characterization, about ten months after their field collection, three parthenogenetic colonies of each aphid lineage were established on cabbage seedlings in small plexiglass cages. Aphids were raised on these plants as separate lineages for three generations at 20ºC long day (16:8 h light:dark cycle) to remove maternal and grand-maternal effects. Three third-generation (G3) adults were transferred to a new plant and placed at short day conditions (10:14 h light:dark cycle) at 15 ºC. After three days the adults were removed and the juveniles were returned to short day conditions. When the G4 aphids were third or fourth instars, three aphids per lineage were transferred to a new plant and returned to the cabinet under inducing conditions to give birth to the first batch of G5 individuals (G5-1). One week later the three G4 adults were transferred to a new plant to give birth to the G5 batch 2 individuals (G5-2). This processes was repeated a third time to generate the G5 batch 3 progeny. The adult morphs of the three batches of G5 progeny from each lineage were scored. Lineages that produced males and pre-sexual females (gynoparae) and in at least one of the three replicates and no asexual females (vivipara) in any of the three replicates were classified as cyclical parthenogens (holocyclics). Lineages that

produced all three morphs males, gynoparae and vivipara were classified as intermediates. Lineages that produced males in at least one replicate and vivipara but no gynoparae were classified as androcyclics and finally lineages that failed to produce any sexual morphs were classified as obligate parthenogens (anholocyclics). Gynoparae are winged females and were distinguished from alate viviparous aphids by the exclusively production of sexual female progeny (ovipara).

Holocyclic *M. persicae* lineages were tested for their ability to transmit PLRV. A large quantity (>200) of aphids reared on healthy *P. floridana* plants was placed in a dish containing *P. floridana* leaves infected with wildtype PLRV. Aphids were allowed to feed for a 48 hour acquisition period before being transferred to healthy young *P. floridana* plants. After a 5-day transmission period, plants were treated with insecticide (Dibrom 8E; Valent, Walnut Creek, CA, USA) and transferred to the greenhouse. Symptoms of PLRV were observed within three to five weeks.

Tissue Collection

Three separate aphid tissues, digestive tracts, heads and salivary glands were isolated from the USDA red lineage. Digestive tracts and heads were dissected from alate adult asexual females that had been anesthetized by dipping each individual in 70% ethanol. Dissections were performed using a pin embedded in a wooden handle that was inserted dorsally between the head and thorax while holding aphids by their wings with forceps. Aphid heads with the intact digestive tract attached were removed by applying light pressure anteriorly with the pin and posteriorly with the forceps. Following dissection, head and gut tissue were separated and stored individually in RNAlater (Ambion, Austin, TX, USA) at -20°C for up to one month. Salivary gland tissue was dissected from ~400 aphids of different life stages (predominantly fourth instar alates and alate adults) reared on cabbage (*B. oleraceae*). Aphid heads were

detached from their bodies as described above, and then the salivary glands were exposed by removal of the antennae, stylet, and head capsule. Following this both sets of principal and accessory glands were carefully removed from the remaining tissues and stored in RNAlater (Ambion, Austin, TX, USA) solution.

Adult males were obtained after approximately six weeks and adult sexual females after seven weeks following transfer to short day conditions (13:11 h light:dark cycle at 18°C $\pm$ 2). Altogether, 92 sexual females and 128 males of lineage F001, and 81 sexual females and 134 males of lineage G006 were flash frozen and stored at -80 ºC for library construction.

cDNA Library Construction

Total RNA was isolated using RNeasy kits (Qiagen, Valencia, CA, USA) or Tripure reagant (Roche, Indianapolis, IN, USA) and purified mRNA using Oligotex resin (Qiagen, Valencia, CA, USA) or Dynabeads mRNA Purification Kit (Invitrogen, Carlsbad, CA, USA). All cDNA libraries were made from mRNA with the exception of the salivary gland library, (MpSG), which was made from total RNA. Genomic DNA was isolated from flash frozen aphids following the "salting-out" protocol of [63].

Non-normalized libraries: Four libraries were made following the LD PCR protocol from the Creator SMART cDNA Library Construction Kit (Clontech, Mountain View, CA, USA). cDNA generated by reverse transcription was amplified, digested with *Sfi* 1A and *Sfi* 1B, and size fractionated. Double-stranded cDNA was directionally cloned into the pDNR-LIB plasmid vector, and transformed into DH10B competent cells (Invitrogen, Carlsbad, CA, USA). The cDNA for a fifth non-normalized library was generated from mRNA and cloned using a Superscript Plasmid system with Gateway

technology (Invitrogen, Carlsbad, CA, USA). For this library, the size fractionated cDNA was cloned into the pSPORT1 vector cut with *Not*I and *Sal*I and the recombinant plasmids were used to transform ElectroMAX DH10B cells (Invitrogen, Carlsbad, CA, USA).

Normalized libraries: Eleven normalized cDNA libraries were constructed using the TRIMMER direct cDNA Normalization kit (Evrogen, Moscow, Russia) in conjunction with the Creator SMART kit. We generated cDNA by reverse transcription of total RNA (salivary gland library) and mRNA (all other libraries). Double-stranded cDNA for normalization was generated using 15-21 PCR cycles. The double-stranded cDNA was denatured and allowed to re-hybridize under stringent conditions; subsequently the reaction mixture was treated with a duplex specific nuclease [64]. The duplexes corresponded disproportionately to abundant cDNAs, leaving a population of single-stranded cDNA molecules in which the representation of rare transcripts was increased. The remaining single stranded cDNA molecules were subsequently amplified, and library construction proceeded as for non-normalized libraries.

EST Sequencing

Sequencing reactions were performed either on purified plasmids or on PCR-amplified products.

PCR-amplified products: Library aliquots were spread onto selective media and grown overnight at 37°C. Colonies were picked manually into 384 well plates (Genetix, New Milton, Hampshire, UK) containing selective media and grown overnight at 37°C. One μL of liquid culture was used as a template for colony PCR (primer sequences in Supplementary Table 2.7). Colony PCR products were analyzed by gel electrophoresis

to confirm the presence of an insert. PCR products were purified using MinElute 96 UF plates (Qiagen, Valencia, CA, USA) or AMpure (Agencourt Biosciences, Beverly, MA, USA). Sequencing reactions were carried out using ABI PRISM BigDye technology, and sequences were analyzed on the ABI 3730XL automated multicapillary sequencer (Applied Biosystems, Foster City, CA, USA).

Purified plasmids: Library aliquots were spread onto Q-tray vented bioassay plates (Genetix, New Milton, Hampshire, UK) containing selective media and grown for 18 hours at 37°C. Colonies were picked by the Qbot robotic colony manipulator (Genetix, New Milton, Hampshire, UK) into 384-well plates containing selective media and grown for 12 hours at 37°C. Plasmid DNA was purified using SprintPrep 384 HC kits (Agencourt Biosciences, Beverly, MA, USA) and subject to dye-terminator fluorescent DNA sequencing. The sequencing products were purified using CleanSEQ (Agencourt Biosciences, Beverly, MA, USA), and the sequences were analyzed on the ABI 3730XL (Applied Biosystems, Foster City, CA, USA) automated multicapillary sequencer.

Sequence Processing and Annotation

We used Phred [65, 66] to make base calls from sequence traces. Vector and adaptor sequences were identified from each EST using Crossmatch [66], and trimmed along with poly-A tails and low quality sequence (*i.e.* 10 or more bases out of 25 with a quality score below 20). ESTs containing less than 100 bases of quality sequence were discarded. All ESTs were compared to the GenBank nr database using BlastN [67]. Those ESTs for which the best Blast targets were *B. aphidicola* sequences with e-value less that 1E-20 were considered to be endosymbiont contamination and were filtered out. ESTs passing quality tests were clustered using the TribeMCL software (1E-50

and 95% identity for Blast alignment; inflation value 5; [68], and consensus contig sequences were generated using Cap3 [69].

For functional annotation we used the Gene Ontology annotation from the NCBI Gene database (Maglott *et al*., 2005). For classification purposes, we converted all the GO terms in the Gene database to GOslim [33]. BlastX was used to match the contig sequences to the NCBI refseq protein sequences. GOslim annotations of the Refseq hits with BlastX e-value less than 1E-5 were transferred to the query contigs.

SNP Identification and Confirmation

The PolyBayes program [38] was used to identify SNPs. Consensus contig sequences were used as anchor sequences for each alignment, and the cutoff for the SNP probability score was 0.84. Seven PCR primer pairs were used to amplify the 11 predicted SNPs (contigs 129, 254, 1080, 3202, 3285, 3347, 3713: Supplementary Table 2.7) from the F001, G006 and USDA lineages. Primers for elongation factor 1α (EF-1α) were used in control PCR reactions. PCR products were purified using Ampure (Agencourt Biosciences, Beverly, MA, USA). Sequencing reactions were carried out using ABI PRISM BigDye technology (Applied Biosystems, Foster City, CA, USA), and sequences were analyzed on an ABI 3730XL automated multicapillary sequencer. Sequences were aligned using ClustalW [70] to confirm the presence of the putative polymorphism.

***In silico* Prediction of Tissue-Specific Gene Expression**

The method proposed by [35] was used for comparing gene expression profiles from four of the non-normalized cDNA libraries. For each contig, the number of ESTs from each library was counted and the R statistic was calculated. Potential salivary gland specific ESTs were also identified from the normalized salivary gland library. Contigs

considered likely to be expressed in salivary glands contained at least 2 ESTs from the MpSG library, and at least 50% of the ESTs from those contigs were derived from the MpSG library. For each contig, the probability of meeting the selection criteria was calculated using the binomial distribution, the percentage of all ESTs that were in salivary gland libraries (12.1%), and the number of ESTs that contributed to the given contig. Predicted open reading frames from potential salivary gland contigs were translated *in silico* using the ExPASy translator tool [71]. The predicted protein sequence was run through SignalP 3.0 [72], using both the neural network and Hidden Markov model to identify a possible signal peptide in the predicted proteins.

Microarray Design

A 15,000-element expression array containing 60-mers representing the identified *M. persicae* unigene set was designed using eArray (Agilent, Santa Clara, CA, USA). Arrays are being produced at Agilent by *in situ* oligonucleotide synthesis.

Authors' Contributions

ACCW, BF, and GJ collected aphid strains. ACCW, BF, GJ, JSR, MdV, and SG designed experiments. ACCW characterized aphid life cycles and collected sexual morphs. DS and SG conducted virus transmission experiments. AW, MdV, GM, and JSR constructed and sequenced cDNA libraries. ACCW, AW, BF, CT, GJ, GM, JSR, MdV, and QS conducted bioinformatic data analysis. ACCW, GJ, JSR, and MdV wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was funded by USDA grant #2005-35604-15446 to GJ, ACCW, and SG. ACCW was partially supported by NIH Training Grant #1 K12 GM00708 to the

Center for Insect Science, University of Arizona. We thank Nancy Moran for generous advice and helpful comments, Bari Morris and Brandon Smetana for technical assistance with cDNA library construction, Brian Nault and Jim Lamondia for providing aphid lineages, Mark Phillips and John Jones of SCRI for much needed guidance in robotics and computing, Clayton Coffman for pointing out the DsMV homology, and Sara Lewis for assistance with aphid life cycle characterization.

Description of Additional Data Files:

Files available at: http://www.biomedcentral.com/1471-2164/8/423/additional/

Supplementary Table 2.1: Collection and life cycle characterization of aphid lineages, Excel spreadsheet

Supplementary Table 2.2: BlastX comparisons of M. persicae contigs to GenBank DNA sequences, Excel spreadsheet

Supplementary Table 2.3: Full listing of *M. persicae* ESTs, along with the source cDNA libraries, Excel spreadsheet

Supplementary Table 2.4: Secreted proteins in salivary glands, as predicted by SignalP 3.0, Excel spreadsheet

Supplementary Table 2.5: Potential SNPs identified based on sequence differences among M. persicae lineages, Excel spreadsheet

Supplementary Table 2.6: High-confidence SNPs, represented by at least two reads in at least two M. persicae lineages, Excel spreadsheet

Supplementary Table 2.7: PCR primers used in this study, Excel spreadsheet

REFERENCES

1. M Robertson, G Zehnder, M Hammig: **Adaptation of integrated pest management practices by South Carolina cotton growers**. *Journal of Extension* 2005, **43**:6RIB8.
2. C Girousse, B Moulia, W Silk, J Bonnemain: **Aphid infestation causes different changes in carbon and nitrogen allocation in alfalfa stems as well as different inhibitions of longitudinal and radial expansion**. *Plant Physiol.* 2005, **137**:1474-1484.
3. V Pegadaraju, C Knepper, J Reese, J Shah: **Premature leaf senescence modulated by the Arabidopsis *PHYTOALEXIN DEFICIENT4* gene is associated with defense against the phloem-feeding green peach aphid**. *Plant Physiol.* 2005, **139**:1927-34.
4. SM Gray, F Gildow: **Luteovirus-Aphid Interactions**. *Annu Rev Phytopathol* 2003, **41**:539-566.
5. JS Kennedy, MF Day, VF Eastop: **A Conspectus of Aphids as Vectors of Plant Viruses**. London: Commonwealth Institute of Entomology; 1962.
6. RL Blackman, VF Eastop: **Aphids on the World's Crops**. Chichester: Wiley; 2000.
7. T Mowry: **Insecticidal reduction of *Potato leafroll virus* transmission by *Myzus persicae.*** *Annals of Applied Biology* 2005, **146**:81-88.
8. L Terradot, JC Simon, N Leterme, D Bourdin, ACC Wilson, JP Gauthier, Y robert: **Molecular characterization of clones of the *Myzus persicae* complex (Hemiptera: Aphididae) differing in their ability to transmit potato leafroll luteovirus**. *Bull. Ent. Res.* 1999, **89**:355-363.
9. C Vorburger, M Lancaster, P Sunnucks: **Environmentally related patterns of reproductive mode is the aphid Myzus persicae and the predominance of**

two superclones in Victoria, Australia. *Molecular Ecology* 2003, **12**:3493-3504.

10. K Zitoudi, JT Margaritopoulos, Z Mamuris, JA Tsitsipis: **Genetic variation in *Myzus persicae* populations associated with host-plant and life cycle category.** *Ent. Exp. Appl.* 2001, **99**:303-311.
11. RL Blackman: **Morphological discrimination of a tobacco-feeding form from *Myzus persicae* (Sulzer) (Hemiptera: Aphididae), and a key to New World *Myzus* (*Nectarosiphon*) species**. *Bull Ent Res* 1987, **77**:713-730.
12. FE Guthrie, WV Campbell, RL Baron: **Feeding sites of the green peach aphid with respect to its adaptation to tobacco**. *Ann. Ent. Soc. Am.* 1962, **55**:42-56.
13. RM McPherson: **Seasonal abundance of red and green morphs of the tobacco aphid (Homoptera: Aphididae) on flue-cured tobacco in Georgia**. *J Ent Sci* 1989, **24**:531-538.
14. EP Lampert, CA Dennis: **Life history of two color morphs of the green peach aphid (Homoptera: Aphididae) on flue-cured tobacco**. *Tobacco Sci* 1987, **31**:91-93.
15. KM Clements, CE Sorenson, BM Wiegmann, PA Neese, RM Roe: **Genetic, biochemical, and behavioral uniformity among populations of *Myzus nicotinianae* and *Myzus persicae***. *Ent. Exp. Appl.* 2000a, **95**:269-281.
16. JT Margaritopoulos, Z Mamuris, JA Tsitsipis: **Attempted disctimination of *Myzus persicae* and *Myzus nicotinianae* (Homoptera: Aphididae) by random amplified polymorphic DNA polymerase chain reaction technique**. *Ann. Ent. Soc. Am.* 1998, **91**:602-607.
17. LM Field, N Javed, MF Stribley, AL Devonshire: **The peach-potato aphid *Myzus persicae* and the tobacco aphid *Myzus nicotianae* have the same**

esterase-based mechanisms of insecticide resistance. *Insect Mol Biol* 1994, **3**:143-8.

18. RL Blackman, JM Spence: **Electrophoretic distinction between the peach-potato aphid, *Myzus persicae*, and the tobacco aphid, *M. nicotinianae* (Homoptera: Aphididae)**. *Bull Ent Res* 1992, **82**:161-165.
19. T Shigehara, H Takada: **Changes in genotypic composition of *Myzus persicae* (Hemiptera: Aphididae) on tobacco during the past two decades in Japan**. *Bull Ent Res* 2003, **93**:537-544.
20. JT Margaritopoulos, JA Tsitsipis, E Zintzaras, RL Blackman: **Host-correlated morphological variation of *Myzus persicae* (Hemiptera: Aphididae) populations in Greece**. *Bull Entomol Res* 2000, **90**:233-44.
21. M de Vos, VR van Oosten, RM van Poecke, JA van Pelt, MJ Pozo, MJ Mueller, AJ Buchala, JP Metraux, LC van Loon, M Dicke, *et al*: **Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack**. *Mol Plant Microbe Interact* 2005, **18**:923-37.
22. PJ Moran, Y Cheng, JL Cassell, GA Thompson: **Gene expression profiling of *Arabidopsis thaliana* in compatible plant-aphid interactions**. *Arch. Insect Biochem. Physiol.* 2002, **51**:182-203.
23. C Couldridge, HJ Newbury, B Ford-Lloyd, J Bale, J Pritchard: **Exploring plant responses to aphid feeding using a full *Arabidopsis* microarray reveals a small number of genes with significantly altered expression**. *Bull Ent Res* 2007, **in press**.
24. V Pegadaraju: **Molecular insights into Arabidopsis response to *Myzus persicae* Sulzer (green peach aphid)**. Manhattan: Kansas State University; 2005. Ph.D.
25. MT Ruiz, O Voinnet, DC Baulcombe: **Initiation and maintenance of virus-**

induced gene silencing. *Plant Cell* 1998, **10**:937-46.

26. KK Bhattarai, Q Li, Y Liu, SP Dinesh-Kumar, I Kaloshian: **The MI-1-mediated pest resistance requires Hsp90 and Sgt1**. *Plant Physiol.* 2007, **144**:312-23.

27. L Lee, P Palukaitis, SM Gray: **Host-dependent requirement for the Potato leafroll virus 17-kda protein in virus movement.** *Mol Plant Microbe Interact* 2002, **15**:1086-1094.

28. TL Finston, DN Hebert, RB Foottit: **Genome size variation in aphids**. *Insect Biochem Molec Biol* 1995, **25**:189-196.

29. MA Sloane, P Sunnucks, AC Wilson, DF Hales: **Microsatellite isolation, linkage group identification and determination of recombination frequency in the peach-potato aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae)**. *Genet Res* 2001, **77**:251-60.

30. ACC Wilson, B Massonnet, J-C Simon, N Prunier-Leterme, L Dolatti, KS Llewellyn, CC Figueroa, CC Ramirez, RL Blackman, A Estoup, *et al*: **Cross-species amplification of microsatellite loci in aphids: assessment and application**. *Molecular Ecology Notes* 2004, **4**:104-109.

31. CC Figueroa, N Prunier-Leterme, C Rispe, F Sepulveda, E Fuentes-Contreras, B Sabater-Munoz, J-C Simon, D Tagu: **Annotated expressed sequence tags and xenobiotic detoxification in the aphid *Myzus persicae* (Sulzer)**. *Insect Science* 2007, **14**:29-45.

32. B Fenton, G Malloch, JAT Woodford, SP Foster, J Anstead, I Denholm, L King, J Pickup: **The attack of the clones: tracking the movement of insecticide-resistant peach–potato aphids *Myzus persicae* (Hemiptera: Aphididae)**. *Bull Ent Res* 2005, **95**:483-494.

33. MA Harris, J Clark, A Ireland, J Lomax, M Ashburner, R Foulger, K Eilbeck, S

Lewis, B Marshall, C Mungall, *et al*: **The Gene Ontology (GO) database and informatics resource**. *Nucleic Acids Res* 2004, **32**:D258-61.

34. KD Pruitt, T Tatusova, DR Maglott: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic Acids Res* 2007, **35**:D61-5.

35. DJ Stekel, Y Git, F Falciani: **The comparison of gene expression from multiple cDNA libraries**. *Genome Res* 2000, **10**:2055-61.

36. M van Munster, A Dullemans, M Verbeek, J van den Heuvel, C Reinbold, V Brault, A Cleriver, F van der Wilk: **Characterization of a new densovirus infecting the green peach aphid Myzus persicae**. *Journal of Invertebrate Pathology* 2003, **84**:6-14.

37. F Francis, N Vanhaelen, E Haubruge: **Glutathione S-transferases in the adaptation to plant secondary metabolites in the Myzus persicae aphid**. *Arch. Insect Biochem. Physiol.* 2005, **58**:166-74.

38. GT Marth, I Korf, MD Yandell, RT Yeh, Z Gu, H Zakeri, NO Stitziel, L Hillier, PY Kwok, WR Gish: **A general approach to single-nucleotide polymorphism discovery**. *Nat Genet* 1999, **23**:452-6.

39. RM Bennett-Lovsey, AD Herbert, MJE Sternberg, LA Kelley: **Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre**. *Proteins: Structure, Function and Bioinformatics* 2007, **In press**.

40. B Sabater-Munoz, F Legeai, C Rispe, J Bonhomme, P Dearden, C Dossat, A Duclert, JP Gauthier, DG Ducray, W Hunter, *et al*: **Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera)**. *Genome Biol* 2006, **7**:R21.

41. L Lee, WB Hunter, LE Hunnicutt, PM Dang: **An expressed sequence tag**

(EST) cDNA library of *Aphis gossypii* alates. In: *American Phytopathological Society Annual Meeting*; 2005; Austin, TX. P-579.

42. WB Hunter, PM Dang, MG Bausher, JX Chaparro, W McKendree, RG Shatters, Jr., CL McKenzie, XH Sinisterra: **Aphid biology: expressed genes from alate *Toxoptera citricida*, the brown citrus aphid**. *J Insect Sci* 2003, **3**:23.

43. D Tagu, N Prunier-Leterme, F Legeai, J-P Gauthier, A Duclert, B Sabater-Muñoz, J Bonhomme, J-C Simon: **Annotated expressed sequence tags for studies of the regulation of reproductive modes in aphids**. *Insect Biochem Mol Biol* 2004, **34**:809-822.

44. JP Gauthier, F Legeai, A Zasadzinski, C Rispe, D Tagu: **AphidBase: a database for aphid genomic resources**. *Bioinformatics* 2007, **23**:783-4.

45. NS Mutti, Y Park, JC Reese, GR Reeck: **RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum***. *J Insect Sci* 2005, **6**:7 pp, available online insectscience.org/6.38.

46. R Ram, A Joshi, N Verma, S Kulshrestha, G Raikhy, V Hallan, AA Zaidi: **First report of *Dasheen mosaic virus* infecting four ornamental aroids in India**. *Plant Path* 2003, **52**:411.

47. B Pontoppidan, B Ekbom, S Eriksson, J Meijer: **Purification and characterization of myrosinase from the cabbage aphid (*Brevicoryne brassicae*), a Brassica herbivore**. *Eur. J. Biochem.* 2001, **268**:1041-8.

48. AM Jones, M Bridges, AM Bones, R Cole, JT Rossiter: **Purification and characterisation of a non-plant myrosinase from the cabbage aphid *Brevicoryne brassicae* (L.)**. *Insect Biochem. Mol. Biol.* 2001, **31**:1-5.

49. F Francis, G Lognay, JP Wathelet, E Haubruge: **Characterisation of aphid myrosinase and degradation studies of glucosinolates**. *Arch. Insect*

Biochem. Physiol. 2002, **50**:173-82.

50. F Francis, G Lognay, JP Wathelet, E Haubruge: **Effects of allelochemical from first (Brassicaceae) and second (*Myzus persicae* and *Brevicornye brassicae*) trophic levels on *Adalia bipunctata***. *Journal of Chemical Ecology* 2001, **27**:243-256.

51. DB MacGibbon, RM Allison: **Glucosinolate system in the aphid *Brevicoryne brassicae***. *New Zealand J Sci* 1968:444-446.

52. JH Kim, G Jander: ***Myzus persicae* (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate**. *Plant Journal* 2007, **49**:1008-1019.

53. D Tagu, B Sabater-Muñoz, J-C Simon: **Deciphering reproductive polyphenism in aphids**. *Inv Reprod Dev* 2005, **48**:71-80.

54. S Ramos, A Moya, D Martinez-Torres: **Identification of a gene overexpressed in aphids reared under short photoperiod**. *Ins. Bioch. Mol. Bio.* 2003, **33**:289-298.

55. LM Field, SP Foster: **Amplified esterase genes and their relationship with other insecticide resistance mechanisms in English field populations of the aphid, *Myzus persicae* (Sulzer)**. *Pest Manag Sci* 2002, **58**:889-94.

56. D Martinez-Torres, SP Foster, LM Field, AL Devonshire, MS Williamson: **A sodium channel point mutation is associated with resistance to DDT and pyrethroid insecticides in the peach-potato aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae)**. *Insect Mol. Bio.* 1999, **8**:329-346.

57. R Maleszka, R Kucharski: **Analysis of *Drosophila* yellow-B cDNA reveals a new family of proteins related to the royal jelly proteins in the honeybee and to an orphan protein in an unusual bacterium *Deinococcus radiodurans***. *Biochem Biophys Res Commun* 2000, **270**:773-6.

58. J Schmitzova, J Klaudiny, S Albert, W Schroder, W Schreckengost, J Hanes, J Judova, J Simuth: **A family of major royal jelly proteins of the honeybee *Apis mellifera* L**. *Cell Mol Life Sci* 1998, **54**:1020-30.

59. MD Drapeau, S Albert, R Kucharski, C Prusko, R Maleszka: **Evolution of the yellow/major royal jelly protein family and the emergence of social behavior in honey bees**. *Genome Res* 2006, **16**:1385-94.

60. N Peiren, F Vanrobaeys, DC de Graaf, B Devreese, J Van Beeumen, FJ Jacobs: **The protein composition of honeybee venom reconsidered by a proteomic approach**. *Biochim Biophys Acta* 2005, **1752**:1-5.

61. M Knop, A Finger, T Braun, K Hellmuth, DH Wolf: **Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast**. *Embo J* 1996, **15**:753-63.

62. C Brennan, Y Zhang, C Leo, B Feng, C Cauwels, AJ Aguirre, M Kim, A Protopopov, L Chin: **High-resolution global profiling of genomic alterations with long oligonucleotide microarray**. *Cancer Res* 2004, **64**:4744-8.

63. P Sunnucks, DF Hales: **Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae)**. *Molecular Biology and Evolution* 1996, **13**:510-523.

64. PA Zhulidov, EA Bogdanova, AS Shcheglov, LL Vagner, GL Khaspekov, VB Kozhemyako, MV Matz, E Meleshkevitch, LL Moroz, SA Lukyanov, *et al*: **Simple cDNA normalization using kamchatka crab duplex-specific nuclease**. *Nucleic Acids Res* 2004, **32**:e37.

65. B Ewing, P Green: **Base-calling of automated sequencer traces using phred. II. Error probabilities**. *Genome Res* 1998, **8**:186-94.

66. B Ewing, L Hillier, MC Wendl, P Green: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Res* 1998,

8:175-85.

67. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-10.

68. AJ Enright, S Van Dongen, CA Ouzounis: **An efficient algorithm for large-scale detection of protein families**. *Nucleic Acids Res* 2002, **30**:1575-84.

69. X Huang, A Madan: **CAP3: A DNA sequence assembly program**. *Genome Res* 1999, **9**:868-77.

70. JD Thompson, DG Higgins, DJ Gibson: **CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penatilies and weight matrix choice**. *Nuc Acids Res* 1995, **22**:4673-4680.

71. E Gasteiger, A Gattiker, C Hoogland, I Ivanyi, RD Appel, A Bairoch: **ExPASy: The proteomics server for in-depth protein knowledge and analysis**. *Nucleic Acids Res* 2003, **31**:3784-8.

72. JD Bendtsen, H Nielsen, G von Heijne, S Brunak: **Improved prediction of signal peptides: SignalP 3.0**. *J Mol Biol* 2004, **340**:783-95.

Chapter 3: Genomic evidence for complementary purine metabolism in the pea aphid *Acyrthosiphon pisum* and its symbiotic bacterium *Buchnera aphidicola*

John S. Ramsey[1*], Sandy J. MacDonald[2], Georg Jander[1], Atsushi Nakabachi[3], Gavin H. Thomas[2] and Angela E. Douglas[4]

[1] Boyce Thompson Institute, Ithaca, NY 14853, USA

[2] University of York, Department of Biology, York, YO10 5YW, UK

[3] Advanced Science Institute, RIKEN, Wako, Saitama 351-0198, Japan

[4] Cornell University, Department of Entomology, Ithaca, NY 14853, USA

*Corresponding author: email: jsr47@cornell.edu

Abstract

The purine salvage pathway recycles purines to nucleotides, promoting efficient utilization of purine nucleotides. Exceptionally among animals with completely sequenced genomes, the pea aphid lacks key purine recycling genes coding for purine nucleoside phosphorylase and adenosine deaminase, indicating that the aphid cannot metabolize nucleosides to the corresponding purines, or adenosine to inosine. Purine metabolism genes in the symbiotic bacterium *Buchnera* complement aphid genes, and *Buchnera* can meet its nucleotide requirement from aphid-derived guanosine. *Buchnera* demand for nucleosides may have relaxed selection for purine recycling in the aphid, leading to the loss of key aphid purine salvage genes. The coupled purine metabolism of aphid and *Buchnera* could, further, contribute to the dependence of the pea aphid on this symbiosis.

This chapter has been accepted for publication by Insect Molecular Biology

Introduction

Purines are crucially important as precursors of nucleotides that make up DNA and RNA, signaling molecules (e.g. cAMP, GMP) and the central energy currency, ATP. In many insects, purine metabolism is modified to accommodate the high flux from inosine monophosphate (IMP) to uric acid. This modification reflects the role of uric acid as the principal vehicle for the excretion of excess nitrogen in many terrestrial insects, although some terrestrial insects utilize ammonia as a nitrogen excretory compound (Prusch, 1971; Mullins and Cochran, 1972; Harrison and Phillips, 1992). Furthermore, uric acid is synthesized by insects for various functions other than nitrogen excretion, including epidermal pigment, reflectors in the compound eye, dynamic nitrogen store and as an anti-oxidant (Mullins and Cochran, 1972; Hilliker *et al*., 1992; Souza *et al*., 1997; Ninomiya *et al*., 2006).

The purine metabolism of aphids is of particular interest for two reasons. First, uric acid is undetectable in both the tissues and honeydew (excreta) of aphids and ammonia is the principal aphid excretory compound (Sasaki and Ishikawa, 1993; Whitehead *et al*., 1992), raising the possibility that purine metabolism in this group may differ from that of many insects. Second, most aphids of the family Aphididae live in an obligate symbiosis with a vertically transmitted γ-proteobacterium, *Buchnera aphidicola*, from which they derive essential amino acids. Because the *Buchnera* cells are intracellular (restricted to specific insect cells known as bacteriocytes), they derive their total nutritional requirement from the cytoplasm of aphid cells. This requirement is predicted to include certain purines because the genetic capacity of *Buchnera* for purine metabolism is reduced, including proximal truncation of purine biosynthesis and an inability to synthesize GMP (inferred from annotated gene content of *Buchnera*: Shigenobu *et al*., 2000 and http://www.buchnera.org). Furthermore, an *in silico* analysis of the metabolite flux

through the reconstructed metabolic network of *Buchnera* consistently found the net production of adenine (Thomas *et al*., 2009) that, if displayed by *Buchnera in vivo*, would be made available to the aphid. Inspection of the metabolic reconstruction of *Buchnera* revealed that the adenine is derived principally from polyamine metabolism, thereby linking purine and polyamine metabolism of the two organisms. To summarize, these data suggest that the purine metabolism of aphids might have distinctive features related to their relationship with *Buchnera* and to the central role of ammonia, and not uric acid, in aphid nitrogen excretion.

The complete sequence of the pea aphid genome has recently become available (International Aphid Genomics Consortium, 2009), providing the opportunity to annotate the complement of genes in purine metabolism. An initial automated analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) indicated that the pea aphid has fewer genes of purine metabolism than *Drosophila melanogaster,* but that the two insects have comparable complements of pyrimidine metabolism genes (P.D. Ashton, pers. observ.). The purpose of this study was two-fold: first, to provide a manual annotation of purine metabolism in the pea aphid, linking the aphid gene complement to the genetic capacity of *Buchnera* for purine metabolism; and, second, to investigate the implications of the predicted export of adenine from *Buchnera* (see above) for the purine metabolism of the aphid, including a quantitative analysis by flux balance analysis. Linked to the evidence that the *Buchnera*-derived adenine is generated by polyamine metabolism, this second purpose led to an analysis of the genetic capacity of the aphid for the metabolism of polyamines and its precursor, ornithine, which is a component of the urea cycle in many animals.

Results

Aphid purine metabolism

Manual annotation revealed that the pea aphid has orthologues of the *Drosophila melanogaster* genes mediating the synthesis and degradation of pyrimidines and also purine synthesis, but apparently lacks the genes for three enzymes in the purine salvage pathway by which nucleosides and free bases are recycled to nucleotides (Supplementary Material 3.1). These enzymes are hypoxanthine-guanine phosphoribosyltransferase (HGPRT), also absent from all other insects with completely sequenced genomes, and adenosine deaminase (ADA) and purine nucleoside phosphorylase (PNP), which are present in all other animals with sequenced genomes (Fig. 3.1). Consistent with this conclusion from the genome sequence, the sequences homologous to the genes for HGPRT, ADA and PNP in other animals are absent from publicly available pea aphid ESTs.

In the absence of ADA and PNP, the pea aphid cannot salvage adenosine via inosine and guanosine via guanine, respectively. The pea aphid also lacks any ADA-like growth factors, unlike any other insect with a completely sequenced genome. Nevertheless, the pea aphid does have two genes for adenosine kinase (ACYPI008316 and ACYPI003989), by which adenosine can be recycled to AMP. It also has orthologues for two *D. melanogaster* nucleoside hydrolase genes (ACYPI006571 orthologue of CGCG11158, and ACYPI009203 orthologue of CGCG5418) but these genes are most unlikely to compensate for the absence of PNP because the preferred substrates of these enzymes are inosine and the pyrimidine nucleoside uridine.

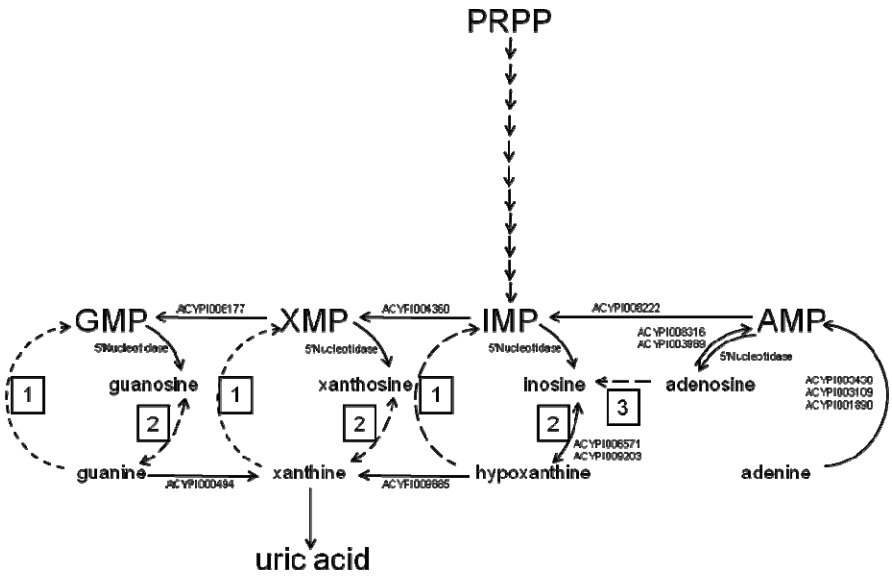


Figure 3.1: The purine salvage pathway and uric acid synthesis. The pea aphid genome apparently lacks genes for (1) hypoxanthine-guanine phosphoribosyltransferase (HGPRT), (2) purine nucleoside phosphorylase (PNP) and (3) adenosine deaminase (ADA). Reactions predicted not to occur in the pea aphid are indicated by dashed lines. ACYPI numbers of other genes coding enzymes in the purine salvage pathway are indicated. The six predicted 5' nucleotidases have the gene IDs ACYPI002452, ACYPI007730, ACYPI003837, ACYPI010172, ACYPI001383, and ACYPI000648.

Futhermore, the 170,000 publicly available pea aphid ESTs include just two ESTs (EX616623 and EX649536) for the nucleoside hydrolase ACYPI006571 and none for ACYPI009203, suggesting that these nucleoside hydrolases are minimally expressed. The broad implication of these findings is that the pea aphid has an incomplete genetic capacity for the purine salvage.

The pea aphid has the genetic capacity to synthesize uric acid from IMP, by means of 5-'nucleotidase (ACYPI002452, ACYPI007730, ACYPI003837, ACYPI010172, ACYPI001383, and ACYPI000648), inosine hydrolase (ACYPI006571 and ACYPI009203), and xanthine dehydrogenase (ACYPI009885), but the genes for urate oxidase and subsequent reactions in uric acid degradation are absent (Fig. 3.1).

Buchnera purine metabolism and shared genetic capacity for purine salvage between the aphid and Buchnera

We then compared the reconstructed pathways of purine metabolism in the pea aphid and *Buchnera*. The *Buchnera* genome codes for fragments of the purine salvage pathway that, when considered in isolation, do not suggest any important physiological function (Thomas *et al*., 2009). However, *Buchnera* has a remarkable complementarity to the genes of the salvage pathway of the aphid. In particular, *Buchnera*, but not the aphid, has the genetic capability to recycle the nucleoside guanosine to a nucleotide and, thereby, to provision fully the purine nucleotide requirement of *Buchnera*. Fig. 3.2 shows the reconstruction of this capability, based on *Buchnera* gene content.

The key *Buchnera* genes involved in recycling aphid-derived purines are *deoDB*, *gpt/hpt*, *guaC* and *purHAB*. The *deoDB* genes code for the terminal two reactions of the purine salvage pathway: the DeoD protein splits purine ribonucleosides to purines and ribose-1-phosphate, and DeoB converts the ribose-1-phosphate to ribose-5-phosphate (Fig. 3.2). Production of this central metabolite allows the *Buchnera* to fully recycle the sugar component of the purine ribonucleoside. The purines released from the DeoD reaction are metabolized to the corresponding nucleotides by phosphoribosyltransferases encoded by *gpt* and *hpt*. The substrate specificities of *E. coli* Hpt and Gpt overlap: Hpt utilises hypoxanthine primarily, but can also use guanine (Guddat *et al*., 2002), while Gpt strongly prefers guanine but can use xanthine and hypoxanthine as substrates (Deo *et al*., 1985).

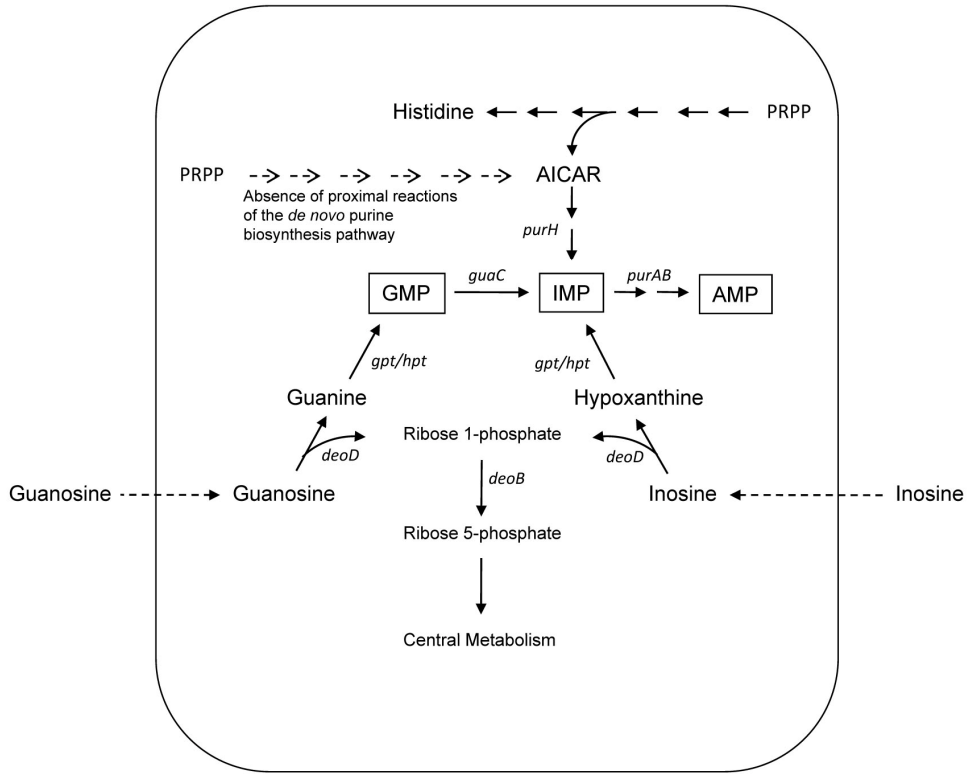


Figure 3.2: Metabolic capabilities of the symbiotic bacterium *Buchnera*, as deduced from its gene content (shown in italics). The proposed uptake of guanosine and inosine by the bacterium is indicated by dashed line.

Our reconstruction of nucleotide synthesis from aphid-derived guanosine (Fig. 3.2) makes IMP a key intermediate metabolite for *Buchnera*, generated from both GMP and the by-product of histidine biosynthetic pathway, AICAR. As DeoD can function with inosine, it is also possible that the IMP pool can be fed by aphid-derived inosine, but this alone is not sufficient for purine provisioning as *Buchnera* cannot make GMP from IMP.

A key prediction of the reconstruction in Fig. 3.2 is that *Buchnera* has high demand for aphid-derived nucleosides, especially guanosine. We identified six pea aphid genes encoding 5'-nucleotidases (legend to Fig. 3.1), two of which (ACYPI002452 and ACYPI007730) were originally detected in the transcriptome

analysis of the bacteriocyte, the aphid cell that houses *Buchnera* (Nakabachi *et al*., 2005). The expression levels of four 5'-nucleotidases in the bacteriocyte were analyzed by real-time quantitative reverse transcription-PCR (Fig. 3.3). Expression of three of the four genes (ACYPI010172, ACYPI002452, and ACYPI007730) was significantly elevated in bacteriocytes relative to the whole body (Mann-Whitney *U*-test; $p<0.01$). The transcript of ACYPI003837 was less abundant in the bacteriocyte than in the whole body (Mann-Whitney *U*-test; $p<0.01$), and the overall expression level of this gene was low (Fig. 3.3).

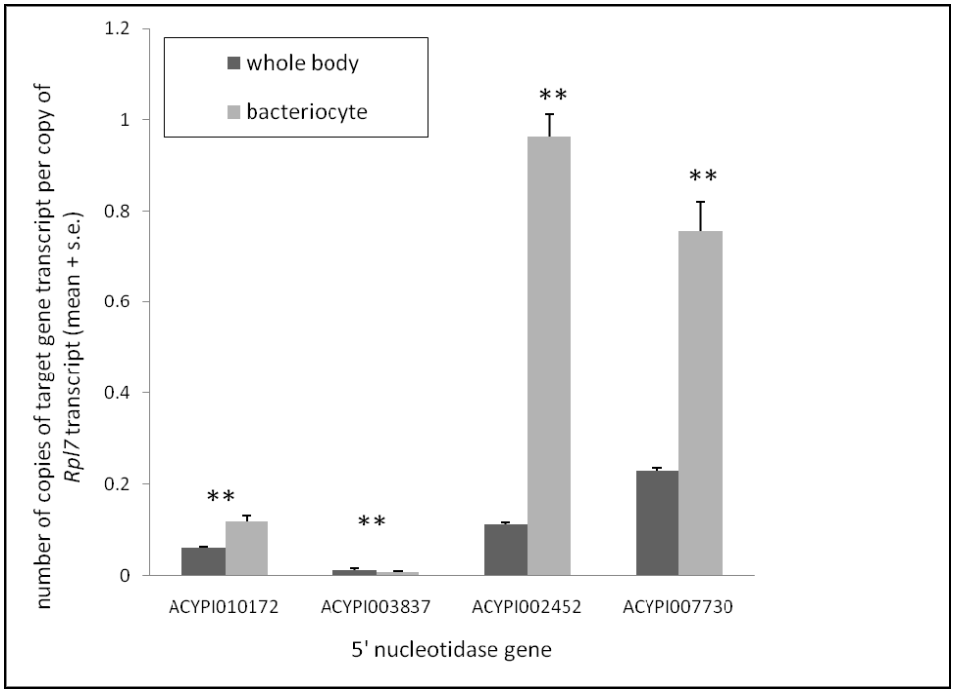


Figure 3.3: Expression of pea aphid 5'-nucleotidase genes in the whole insect and the bacteriocyte; bars, standard errors (n=6). The expression levels are shown in terms of mRNA copies of target genes per copy of mRNA for RpL7. Asterisks indicate statistically significant differences (Mann-Whitney *U* test; **, $p<0.01$).

Shared genetic capacity for purine salvage between the aphid and Buchnera

The metabolic complementarity of the pea aphid and *Buchnera* extends to their capacity to utilize the purine adenine, and generates the potential for the aphid and *Buchnera* to have a shared pathway for the recycling of aphid-derived nucleosides to nucleotides. As described in the Introduction and above, *Buchnera* cells are predicted to produce the purine adenine, and have no genetic capacity to utilize this purine, while the aphid can recycle it to the nucleotide AMP via APRT (Fig. 3.4).

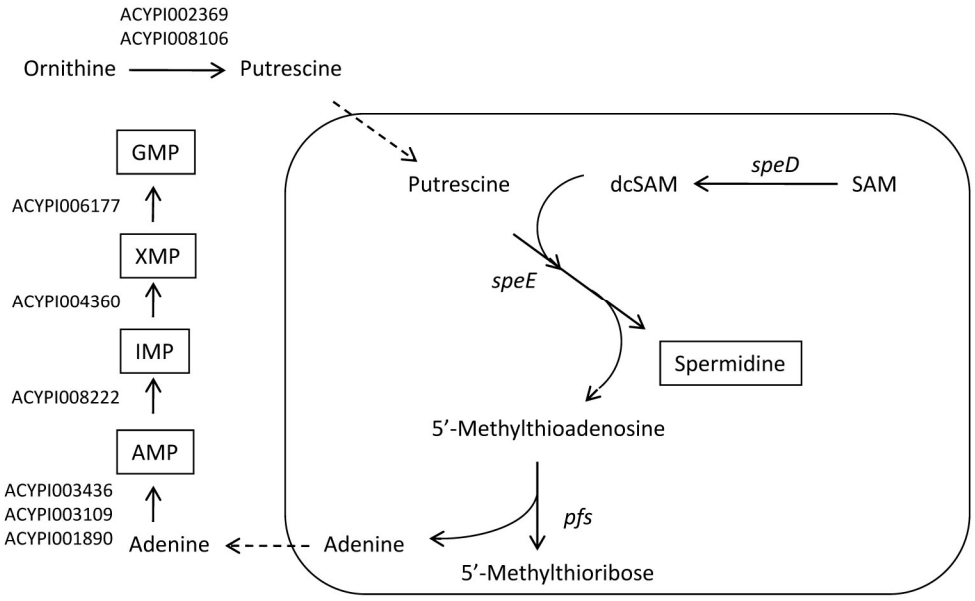


Figure 3.4: Polyamine metabolism in *Buchnera*. The proposed uptake of putrescine and export of adenine by the bacterium is indicated by dashed line. Aphid genes potentially involved in providing putrescine for *Buchnera* spermidine synthesis, and in recycling the byproduct adenine, are indicated by ACYPI number. The metabolic fate of 5'-methylthioribose produced by *Buchnera* is unknown.

We investigated the quantitative importance of adenine recycling by flux balance analysis. Initial analysis revealed that adenine is generated principally as a byproduct of polyamine synthesis. Previous experimental work has shown that the dominant polyamine of *Buchnera* is spermidine (Nakabachi and Ishikawa 2000) at an

estimated concentration of 64 nmol mg^{-1} dry weight of *Buchnera* (Supplementary Material 3.2). When flux through the metabolic network of *Buchnera* was quantified *in silico* by flux balance analysis, 90% of the nucleoside acquired by *Buchnera* was metabolized to nucleotides and 10% was metabolized to adenine and made available for recycling back to the aphid.

This putative recycling of adenine via *Buchnera*-mediated synthesis of spermidine, is dependent on aphid-derived putrescine as a substrate because *Buchnera* lacks the genetic capacity for *de novo* polyamine biosynthesis (Nakabachi and Ishikawa, 2000). The pea aphid has the capacity to synthesize putrescine from ornithine, possessing two genes for ornithine decarboxylase, (ACYPI002369, ACYPI008106; see Supplementary Material 3.1). A second metabolic source of ornithine in various insects is the urea cycle, specifically arginase-mediated breakdown of arginine to ornithine and urea. Other insects with completely sequenced genomes have all the genes in the urea cycle, apart from ornithine transcarbamylase (OTC), which produces citrulline from ornithine and carbamyl phosphate. The pea aphid also lacks this gene and, exceptionally among the insects, the urea cycle genes coding for arginosuccinate synthase, arginosuccinate lyase and arginase, with the implication that it cannot synthesize either ornithine or urea by this route (Supplementary Material 3.1). The pea aphid does, however, retain the gene for nitric oxide synthase (ACYPI001689).

Discussion

The purine salvage pathway functions to recycle nucleosides and free bases to nucleotides. The incomplete genetic capacity for this pathway in the pea aphid is exceptional among animals. The absence of PNP and ADA precludes, first, the flux of purine nitrogen from guanosine and adenosine to uric acid and, second, the salvage of

all nucleosides and purines apart from adenosine and adenine (Fig. 3.1). A possible explanation for the evolutionary loss of these genes arises from the remarkable complementarity between the purine metabolic capabilities of the pea aphid and its symbiotic bacterium *Buchnera*, as illustrated in Fig. 3.2. We specifically propose that aphid-derived guanosine is the principal source of purines for *Buchnera*, on the basis of two considerations. First, *Buchnera*, unlike many other bacteria including its free-living relative *E. coli*, cannot synthesize the carbon skeleton of purines *de novo* from PRPP, but has the genetic capacity to synthesize the total nucleotide complement from guanosine via the genes *deoD*, *gpt, hpt, guaC* and *purAB* (Fig. 3.2). Published microarray analyses (Viñuelas *et al*., 2007) confirm that these *Buchnera* genes are expressed. Second, nucleosides, but not nucleotides, are readily transported across biological membranes (de Koning *et al*., 2005). Although the identity of the putative nucleoside transporters in the bacteriocyte remain to be identified, an indication of the significance of *Buchnera* demand for nucleosides comes from the elevated expression of 5'-nucleotidase genes, especially ACYPI1002452, in the bacteriocyte relative to the insect body (Fig. 3.3).

We hypothesize that the insect in the ancestral symbiosis with *Buchnera* had an intact purine salvage pathway, as occurs in other insects with sequenced genomes. The transfer of guanosine to bacterial cells would compete with uric acid synthesis (Fig. 3.1), raising the possibility that reduced activity of the enzymes PNP and ADA, followed by the loss of the cognate genes, were adaptations of the aphid to ensure supply of nucleosides to the bacteria. The evidence from flux balance analysis of the reconstructed metabolic network of *Buchnera* suggests, further, that approximately 10% of the purine skeleton delivered to the bacteria is recycled via adenine (Fig. 3.4). This would account for the retention of the aphid gene for APRT, and can potentially contribute to the purine economy of the aphid.

There are two important implications of the proposed coupling of purine metabolism between the pea aphid and *Buchnera*. First, the aphid is dependent on its bacterial symbionts for the metabolism of nucleosides. We propose that pea aphids experimentally deprived of their bacteria by antibiotic treatment accumulate nucleosides with deleterious consequences. *Buchnera* is widely recognized to contribute to aphid metabolism and nutrition by their synthesis of essential amino acids, nutrients in short supply in the aphid diet of plant phloem sap (Gündüz and Douglas, 2009), but it has previously been unclear why pea aphids lacking *Buchnera* fail to grow or reproduce when reared on diets with ample essential amino acids (e.g. Prosser and Douglas, 1992). The genomic data obtained in this study suggest that an incomplete capacity for purine metabolism may contribute to dependence of pea aphids on their complement of *Buchnera*.

The second implication of the purine metabolism coupled between the aphid and *Buchnera* is that the aphid is predicted to have very limited capacity to synthesize uric acid, which is a major excretory product of many insects. One great advantage of uric acid as a vehicle for nitrogen excretion is that it can be voided as a solid, so minimizing water loss. This is not relevant for aphids, which utilize the water-rich diet of plant phloem sap and eliminate all waste in liquid honeydew. Consistent with these considerations, uric acid is undetectable in the carcass and honeydew of pea aphids in laboratory culture (Sasaki and Ishikawa, 1992). Nevertheless, the aphid has the gene for xanthine dehydrogenase, which catalyzes uric acid synthesis. Although expression of this gene is minimal in laboratory-reared aphids (as indicated by the few ESTs for this gene), uric acid production may be important under certain field conditions, for example linked to its function as an anti-oxidant or regulator of cation concentrations (see Introduction). The apparent absence of pea aphid genes for urate oxidase and subsequent reactions in uric acid degradation matches the condition

reported in *Tribolium castaneum* and *Apis mellifera*, and differs from *D. melanogaster* and other dipterans with sequenced genomes, which have the capacity to degrade uric acid to allantoin, allantoic acid and in some species to urea (Scaraffia *et al*., 2008).

In conclusion, the pea aphid is unique among the many insects with symbiotic microorganisms in that the genomes of both insect and microbial partners are fully sequenced and annotated. This offers unique opportunities for the construction and testing of specific hypotheses. As an example, our annotation of purine metabolism genes has generated the hypotheses that purine metabolites are exchanged between the aphid and *Buchnera* bacteria, and that imbalance of purine metabolites contributes to the poor performance of aphids deprived of their symbiotic bacteria. We are currently testing these hypotheses experimentally by the metabolite analysis of the symbiotic partners.

Experimental Procedures

Manual annotation of genes

Refseq protein IDs for *Drosophila* genes involved in nucleotide and urea cycle metabolism were retrieved from the KEGG pathway database and FlyBase. The corresponding *Drosophila* protein sequences were retrieved in fasta format from NCBI using Batch Entrez. These protein sequences were used as blast queries against: 1) the *A. pisum* merged Glean/Refseq protein database (blastp, evalue cutoff = 1E-3) and 2) the *A. pisum* whole genome sequence (tblastn, evalue cutoff = 1E-3). The *A. pisum* gene corresponding to the best Glean/Refseq protein hit was taken as the *A. pisum* orthologue of each *Drosophila* gene. If no hits to a *Drosophila* protein were found in the set of predicted *A. pisum* proteins or in the whole genome sequence, the *A. pisum* genome is predicted to lack an orthologue of the corresponding *Drosophila* gene.

Flux balance analysis

The published *in silico* metabolic reconstruction of *Buchnera*, iGT196 (Thomas *et al.*, 2009), was modified to match the purine metabolism predicted from this study: first, the adenosine exchange reaction EX_adn was removed, thereby limiting the source of external nucleoside to guanosine (as in Fig. 3.4); and, second, the phosphopentomutase reaction PPM mediating conversion of ribose-1-phosphate to ribose-5-phosphate and coded for by *deoB* was added, allowing the *Buchnera* to metabolize the sugar component of the nucleoside through central metabolism via the pentose phosphate pathway. In addition, the stoichiometric coefficient for spermidine in the biomass reaction, VGRO, was increased from 0.007 (the value derived from the data of Reed *et al* (2003) based on their metabolic reconstruction of *Escherichia coli* K-12) to 0.05, to match the empirically determined concentration of spermidine in *Buchnera* cells [see Nakabachi & Ishikawa, (2001) and Supplementary Material 3.2]. Flux balance analysis was carried out using the COBRA toolbox software (Becker *et al.*, 2007) running in MATLAB, and the network was optimised for maximal biomass while releasing essential amino acids at empirically determined rates (as in Thomas *et al*., 2009). The biomass production of the modified model was 5.20, as compared to 5.21 for iGT196.

Real-time quantitative RT-PCR

Strain ISO, a parthenogenetic clone of the pea aphid that is free from secondary symbionts, was used for the analysis. The insects were reared on *Vicia faba* at 15°C in a long-day regime of 16 hr light and 8 hr dark. RNA was isolated from whole bodies and bacteriocytes of 12-15 day-old parthenogenetic apterous adults using TRIzol reagent, followed by RNase-free DNase I treatment. Each whole body sample and bacteriocyte sample were derived from one individual and a batch of bacteriocytes that

were collected from about ten individuals, respectively. First-strand cDNAs were synthesized using pd(N)6 primer and PrimeScript reverse transcriptase (Takara). Quantification was performed with the LightCycler instrument and FastStart DNA Master^{PLUS} SYBR Green I kit (Roche), as described previously (Nakabachi *et al*., 2005). The running parameters were: 95º for 10 min, followed by 45 cycles of 95ºC for 10s, 55ºC for 5s, and 72ºC for 4s. The primers used are shown in Table 1.

Table 3.1: Primers used in real-time quantitative PCR.

Gene	Forward primer (5' – 3')	Reverse primer (5' – 3')
ACYPI010172	1306F AAAACTACGAGACGATGGCGTTA	1396R GTCTTCCACTTCCCTGGCAAT
ACYPI003837	475F ATTGGGCGTACAGCCGTAGA	555R ATAGGAATCCGCTATCAGGTTTCC
ACYPI002452	810F TTGCCAGAAATCCCTTCGGTTG	895R CCAGAATGTCCCAACGCTATG
ACYPI007730	384F CAACACGCTGTTCAATTTGCCAG	465R CCTCGAGTAGTTTGGCGAAG

Results were analyzed using the LightCycler software version 3.5 (Roche), and relative expression levels were normalized to mRNA for the ribosomal protein RpL7. Statistical analyses were performed using the Mann-Whitney *U* test.

Authors' Contributions

JSR and SJM annotated aphid genes. GHT annotated purine metabolism genes in *Buchnera*. GHT and SJM conducted flux balance analysis. AN performed real time quantitative PCR experiments. AED, GJ, JSR, and SJM analyzed data from aphid and *Buchnera* genome annotations. JSR, SJM, GHT, and AED wrote the paper.

Acknowledgements

We thank Dr Jeff Scott, Cornell University, for his comments on the manuscript. This study was supported in part by ANR-BBSRC SysBio (grant number: BB/F005342/1) (GHT), Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (AN), USDA grant number 2005-35604-

15446 (GJ and JSR) and the Sarkaria Institute of Insect Physiology and Toxicology (AED).

Description of Additional Data Files

Supplementary Table 3.1. *Acyrthosiphon pisum* genes involved in purine metabolism and the urea cycle. (**Excel spreadsheet provided on CD**)

Supplementary Material 3.2. Calculation of the stoichiometric coefficient of spermidine for flux balance analysis of *Buchnera.* (**see Appendix**)

REFERENCES

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O. and Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* **2:** 727-738.

Deo, S.S., Tseng, W.C., Saini, R., Coles, R.S., and Athwal, R.S. (1985) Purification and characterization of *Escherichia coli* xanthine-guanine phosphoribosyltransferase produced by plasmid pSV2gpt. *Biochim Biophys Acta* **839(3):** 233-239.

Guddat, L.W., Vos, S., Martin, J.L., Keough, D.T., de Jersey, J. (2002) Crystal structures of free, IMP-, and GMP-bound *Escherichia coli* hypoxanthine phosphoribosyltransferase. *Protein Sci* **11(7):** 1626-1638.

Gündüz, E.A. and Douglas, A.E. (2009). Symbiotic bacteria enable insect to utilise a nutritionally-inadequate diet. *Proc R Soc Lond B Biol Sci* **276:** 987-991.

Harrison, J.F. and Phillips, J.E. (1992). Recovery from acute haemolymph acidosis in unfed locusts. II. Role of ammonium and titratable acid excretion. *J Exp Biol* **165:** 97-110.

Hilliker, A.J., Duyf, B., Evans, D. and Phillips, J.P. (1992). Urate null rosy mutants of *Drosophila melanogaster* are hypersensitive to oxygen stress. *Proc Natl Acad Sci USA* **89:** 4343-4347.

International Aphid Genomics Consortium (2009). Genome Sequence of the Pea Aphid *Acyrthosiphon pisum. PLoS Biol,* in review**.**

de Koning, H.P., Bridges, D.J., and Burchmore, R.J.S. (2005) Purine and pyrimidine transport in pathogenic protozoa: From biology to therapy. *FEMS Microbiol Rev* **29:** 987-1020.

Mullins, D.E. and Cochran, D.G. (1972). Nitrogen excretion in cockroaches – uric acid is not a major product. *Science* **177:** 699.

Nakabachi, A. and Ishikawa, H. (2000). Polyamine composition and expression of

genes related to polyamine biosynthesis in an aphid endosymbiont, *Buchnera*. *Appl Environ Microbiol* **66(8):** 3305-3309.

Nakabachi, A. and Ishikawa, H. (2001). Expression of host *S*-adenosylmethionine decarboxylase gene and polyamine composition in aphid bacteriocytes. *Insect Biochem Mol Biol* **31:** 491-496.

Nakabachi A., Shigenobu S., Sakazume N., Shiraki T., Hayashizaki Y., Carninci P., Ishikawa H., Kudo T., Fukatsu T. (2005). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium *Buchnera. Proc Natl Acad Sci USA* **102(15):** 5477-5482.

Ninomiya, Y., Tanaka, K. and Hayakawa, Y. (2006). Mechanisms of black and white stripe pattern formation in the cuticle of insect larvae. *J Insect Physiol* **52:** 638-645.

Prosser, W.A. and Douglas, A.E. (1992). A test of the hypotheses that nitrogen is upgraded and recycled in an aphid (*Acyrthosiphon pisum*) symbiosis. *J Insect Physiol* **38:** 93-99.

Prusch, R.D. (1971). The site of ammonia excretion in the blowfly larva, *Sarcophaga bullata*. *Comp Biochem Physiol* **39A:** 761-767.

Reed, J.L., Vo, T.D., Schilling, C.H. and Palsson, B.O. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* **4:** R54.

Sasaki, T. and Ishikawa, H. (1993). Nitrogen recycling in the endosymbiotic system of the pea aphid *Acyrthosiphon pisum*. *Zoolog Sci* **10:** 779-785.

Scaraffia, P.Y., Tan, G., Isoe, J., Wysocki, V.H., Wells, M.A. and Miesfeld, R.L. (2008). Discovery of an alternate pathway for urea synthesis in adult *Aedes aegypti* mosquitoes. *Proc Natl Acad Sci USA* **105:** 518-523.

Shigenobu, S., Watanabe, H., Hattori, M., Sasaki, Y. and Ishikawa, H. (2000).

Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407:** 81-86.

Souza, A.V.G., Petretski, J.H., Demasi, M., Bechara, E.J.H. and Oliveira, P.L. (1997). Urate protects a blood-sucking insect against hemin-induced oxidative stress. *Free Radic Biol Med* **22:** 209-214.

Thomas, G.H., Zucker, J., Sorokin, A., Goryanin, I. and Douglas, A.E. (2009). A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Syst Biol* **3:** 24.

Viñuelas, J., Calevro, F., Remond, D., Bernillon, J., Rahbé, Y., Febvay, G., Fayard, J.-M. and Charles, H. (2007). Conservation of the links between gene transcription and chromosomal organization in the highly reduced genome of *Buchnera aphidicola* APS. BMC Genomics **8,** 143.

Whitehead, L.F., Wilkinson, T.L. and Douglas, A.E. (1992). Nitrogen recycling in the pea aphid (*Acyrthosiphon pisum*) symbiosis. *Proc R Soc Lond B Biol Sci* **250:** 115-117.

Chapter 4. Comparative analysis of detoxification enzymes in *Acrythosiphon pisum* and *Myzus persicae*

John S. Ramsey[1+], Dean S. Rider[2+], Thomas K. Walsh[3], Martin De Vos[1], Karl Gordon[4], Lalit Ponnala[5], Simone Macmil[6], Bruce A. Roe[6], and Georg Jander[1*]

[1]Boyce Thompson Institute for Plant Research, Ithaca, NY, USA

[2]Department of Veterinary Pathobiology, Texas A&M University, College Station, TX, USA

[3]CSIRO Entomology, Centre for Environment and Life Sciences (CELS), Floreat Park, WA 6014, Australia

[4]CSIRO Entomology, Black Mountain Laboratories, Clunies Ross Street, Black Mountain, Acton ACT 2601, Australia

[5]Biotechnology Center, Cornell University, Ithaca, NY, USA

[6]Department of Chemistry and Biochemistry, University of Oklahoma, OK, USA

[+]John S. Ramsey and Dean S. Rider contributed equally to this research.

* Corresponding author: email: gj32@cornell.edu

Abstract

Herbivorous insects use detoxification enzymes, including cytochrome P450 monooxygenases, glutathione *S*-transferases, and carboxy/cholinesterases, to metabolize otherwise deleterious plant secondary metabolites. Whereas *Acrythosiphon pisum* (pea aphid) feeds almost exclusively from the Fabaceae, *Myzus persicae* (green peach aphid) feeds from hundreds of species in more than forty plant families.

This chapter has been accepted for publication by Insect Molecular Biology

Therefore, *M. persicae* as a species would be exposed to a greater diversity of plant secondary metabolites than *A. pisum*, and has been predicted to require a larger complement of detoxification enzymes. A comparison of *M. persicae* cDNA and *A. pisum* genomic sequences is partially consistent with this hypothesis. There is evidence of at least 40% more cytochrome P450 genes in *M. persicae* than in *A. pisum*. In contrast, no major differences were found between the two species in the numbers of glutathione *S*-transferases, and carboxy/cholinesterases. However, given the incomplete *M. persicae* cDNA data set, the number of identified detoxification genes in this species is likely an underestimate.

Introduction

In their long co-evolution with insect herbivores, plants have developed a variety of defenses to keep from being eaten (Rosenthal and Berenbaum, 1991). These include physical barriers such as spines, tough bark, and sticky sap, as well as numerous distasteful or toxic compounds that are often unique to particular plant genera or families. Defense against herbivory is likely the primary function for many of the several hundred thousand different plant secondary metabolites found in nature (Becerra, 2007; Bino *et al*., 2004). Nevertheless, most, likely all, land plants are fed upon by at least one of the several hundred thousand herbivorous insect species (Schoonhoven *et al*., 1998). This suggests that insects, as part of an evolutionary arms race with their host plants, have developed efficient mechanisms to avoid or detoxify plant secondary metabolites.

Insect herbivores are often broadly classified as either specialists or generalists. Whereas specialists consume only a small number of closely related plant species, generalists tend to forage more widely on a variety of host plants. It is estimated that

90% of herbivorous insects are specialists that forage on three or fewer plant families (Bernays and Graham, 1988), and must therefore negotiate a relatively limited array of plant defenses. In comparison to these more abundant specialists, generalist herbivores are almost certainly exposed to a greater diversity of plant defensive chemicals.

Insect responses to plant secondary metabolites can include avoidance of the most well-defended tissue types, target site insensitivity, rapid passage of toxins through the gut, efflux pumps and direct metabolic detoxification. Cytochrome 450 monooxygenases (P450s) constitute the largest and most functionally diverse class of insect detoxification enzymes (Li *et al*., 2007). Members of the CYP3 clade have been implicated in the oxidative detoxification of furanocoumarins, alkaloids, numerous other plant secondary metabolites and synthetic insecticides (Feyereisen, 1999; Mao *et al*., 2006; Scott, 1999; Snyder and Glendinning, 1996). The CYP4 clade has been implicated in pheromone metabolism (Maïbèche-Coisne *et al*., 2004). Members of the CYP2 clade and mitochondrial targeted P450s contribute to hormone, sterol, and fatty acid metabolism (Feyereisen, 1999; Feyereisen, 2006).

Similar to P450s, carboxyl/cholinesterases (CCEs) can function broadly in xenobiotic detoxification. CCEs from aphids and other insects have been shown to hydrolyze both plant secondary metabolites, organophosphates, and other man-made insecticides (Field, 2000; Li *et al*., 2007). Other members of the CCE superfamily have important neurological and developmental functions, or are involved in pheromone processing (Oakeshott *et al*., 1999).

Glutathione S-transferase (GST) enzymes, which occur in all eukaryotic organisms, function by conjugating xenobiotics and endogenously activated compounds to the thiol group of reduced glutathione, thereby targeting them for more rapid excretion or degradation (Li *et al*., 2007). In insects, GSTs have been associated with resistance to insecticides, including DDT, spinosad, diazinon and nitenpyram,

which target the nervous system, as well as lufenuron and dicyclanil, which cause larval lethality during life stage transitions (Enayati *et al*., 2005; Low *et al*., 2007). GST enzyme activity in the generalist *Myzus persicae* (green peach aphid) increases upon ingestion of isothiocyanates, a class of toxic secondary metabolites found in the Brassicaceae (Francis *et al*., 2005) suggesting that GSTs are involved in detoxification. Comparison of the legume specialist *Acyrthosiphon pisum* (pea aphid) and generalist *Aulacorthum solani* (foxglove or potato aphid) showed broad differences in GST enzyme activities that may reflect their respective host plant preferences (Francis *et al*., 2001).

Whereas specialist insect herbivores tend to have highly efficient detoxification mechanisms that target the predictable set of secondary metabolites in their favored host plants, generalists that feed on a wide variety of plants would need a less specific array of constitutive or inducible detoxification enzymes. Therefore, it is often assumed that generalist herbivores must possess a greater diversity of detoxification enzymes than specialists. Research on Lepidoptera in the genus *Papilio* (swallowtail butterflies) showed that the generalists possess a broader array of P450s for detoxification of furanocoumarins (Li *et al*., 2002; Mao *et al*., 2007). However, since the only sequenced lepidopteran genome is that of *B. mori* (Mita *et al*., 2004), an extreme host plant specialist, it is difficult to assess the actual genetic diversity of P450s and other detoxification enzymes relative to the number of host plants that can be consumed by particular larvae of a particular lepidopteran species.

The *A. pisum* genome (International Aphid Genomics Consortium, 2009) together with a large *M. persicae* expressed sequence tag (EST) collection (Ramsey *et al*., 2007), allows direct comparison of the xenobiotic detoxification enzymes in two related insect species with different feeding habits. Both are classified in the tribe Macrosiphini within the aphid sub-family Aphidinae (von Dohlen *et al*., 2006) and are

about 95% identical at the DNA sequence level. However, *A. pisum* is specialized for feeding on legumes (Fabaceae) and *M. persicae* is a broad generalist that feeds from hundreds of species in more than forty plant families (Blackman and Eastop, 2000).

Legumes, which can be consumed by both *A. pisum* and *M. persicae,* have aphid-deterrent secondary metabolites. For instance, a hemiterpene glucoside was identified as an aphid-deterrent in *Vicia hirsuta* (Ohta *et al*., 2005) and low saponin and phenolic content in alfalfa was associated with improved *A. pisum* performance (Golawska and Lukasik, 2008). Although phloem feeders likely encounter a smaller repertoire of secondary metabolites than chewing insects feeding on the same plants, *M. persicae* almost certainly ingests secondary metabolites that a legume specialist like *A. pisum* would not normally encounter. This would include glucosinolates in Brassicaceae (Kim *et al*., 2008) and alkaloids in the Solanaceae. Tobacco-adapted strains of *M. persicae* are reported to have a nine-fold greater resistance to nicotine in artificial diet (Nauen *et al*., 1996) and nicotine vapors in air (Devine *et al*., 1996), suggesting routine exposure to this alkaloid when *M. persicae* are feeding from tobacco plants.

Given its more diverse feeding habits, *M. persicae* as a species would need to avoid or inactivate a greater variety of plant defenses than *A. pisum*. Although there are defensive proteins (Walz *et al*., 2004) and other barriers to phloem feeding that would not be specifically targeted by detoxifying enzymes (Walling, 2008), secondary metabolites definitely have important defensive functions in many plant species. Here we compare the predicted P450s, GSTs, and esterases produced by *M. persicae* and *A. pisum* to test the hypothesis that a broad generalist insect herbivore would have a greater abundance and functional diversity of xenobiotic detoxification enzymes than a specialist.

Results and Discussion

M. persicae *cDNA sequencing, assembly, and annotation*

To allow large-scale comparisons with the recently sequenced *A. pisum* genome (International Aphid Genomics Consortium, 2009), an existing expressed sequence tag (EST) collection with >10,000 *M. persicae* unigenes produced by Sanger sequencing (Ramsey *et al*., 2007) was augmented by DNA sequencing using Roche GS-FLX (454) technology. Sequencing of cDNA libraries prepared from tobacco-adapted and non-adapted isolates of *M. persicae* produced 142,600 ESTs, with a median read length (N50) of 235 bp and a mean read length of 200 bp. The SeqClean program (http://www.tigr.org/tdb/tgi/software/) was used to remove polyA tails and primer sequences, and any sequence less than 30 bp long after cleaning was discarded. This resulted in a set of 118,756 sequences, with an N50 of 173 bp and a mean length of 166 bp. Removal of 39,795 reads that aligned to an existing *M. persicae* Sanger-sequencing unigene set (Ramsey *et al*., 2007) with a blastn e-value less than 1e-10 resulted in 78,961 sequences were clustered using the MCL algorithm (Enright *et al*., 2002; Van Dongen, 2008). Of these, 34,781 sequences could not be clustered with any other sequences, whereas the remaining 44,180 sequences yielded 9201 clusters. The sequences in each cluster were assembled into contigs using the CAP3 program (Huang and Madan, 1999). The default CAP3 parameters, b=20 and d=200, for the sequence quality score and number of allowable differences, respectively were used to determine whether two given sequences would be assembled into one contig. This yielded a total of 7710 contigs and 5341 singlets. These two non-overlapping datasets, the 34,781 unclustered sequences (Supplemental Table 4.1) and the 13,051 contigs and

singletons (Supplemental Table 4.2), represent a significant expansion of existing cDNA sequence available for *M. persicae*. A FASTA file containing the sequences of these 47,832 unigenes can be downloaded from AphidBase (http://www.aphidbase.com/aphidbase/downloads).

Both clustered and unclustered sequence datasets were compared to the *D. melanogaster* RefSeq protein set, the merged set of *A. pisum* Glean and RefSeq proteins, and the *A. pisum* genomic scaffolds. Approximately 10% of the 454 sequences (4890) had a hit to a *D. melanogaster* RefSeq protein (blastx, e-value ≤ 1e-3). These sequences were annotated by parsing GenBank records to retrieve gene descriptions (Supplemental Table 4.2). More than 25% of the 454 sequences (12,371) had a match to *A. pisum* Glean and/or RefSeq predicted proteins. GenBank records containing automated annotation of *A. pisum* proteins were parsed to annotate *M. persicae* orthologs according to these gene descriptions (Supplemental Tables 4.1 and 4.2).

Over 25% of the 454 unigenes (13,531) had a hit in a BLAST comparison against the *A. pisum* genomic scaffolds (tblastx, e-value ≤ 1E-4), but not to any Glean or RefSeq protein data sets (International Aphid Genomics Consortium, 2009). This discrepancy likely results from *M. persicae* sequences aligning to untranslated regions of the *A. pisum* genome. A higher stringency e-value cutoff was used for this tblastx search than for the blastx against RefSeq proteins, but it is possible that differences between these two algorithms are responsible for the increased number of 454 unigenes having positive hits against the genome. It is also possible that the *A. pisum* Glean and RefSeq protein sets are incomplete, and that some of the *M. persicae* sequences with a match in the *A. pisum* genome but not the Glean/RefSeq set are aligning to coding regions of genes that were missed by gene prediction programs.

The *M. persicae* unigene set, consisting of >10,000 Sanger sequencing contigs

(Ramsey *et al*., 2007) combined with 47,832 additional sequences representing expressed genes identified by 454 sequencing, was compared to the *A. pisum* genome sequence (International Aphid Genomics Consortium, 2009) to assess the relative abundance of xenobiotic detoxification enzymes in a specialist and a broad generalist aphid species.

Cytochrome P450s

The *A. pisum* genome encodes 83 predicted P450s (Supplemental Table 4.3). At least 58 loci exhibit both good sequence homology to other insect P450s and contain a complete P450 domain. Additionally, 25 P450-related loci in the *A. pisum* genome may have incomplete P450 domains. The presence of partial domains could result from gene assembly and annotation problems (*e.g.* aberrant exon prediction), or these loci could represent actual pseudogenes. The majority of pea aphid P450 sequences are related to members of the CYP3 clade, followed by those related to the CYP4 clade (Table 4.1).

Table 4.1. Cytochrome P450 genes

Clade	***Drosophila melanogaster***	***Apis mellifera***	***Acyrthosiphon pisum***	***Myzus persicae***[1]
CYP2	6	8	10	3
CYP3	36	28	33	63
CYP4	32	4	32	48
Mitochondrial	11	6	8	1
total	**85**	**46**	**83**	**115**

[1]Numbers based on *M. persicae* EST data

Of the 83 identified *A. pisum* P450s, 51 show evidence of expression through ESTs in public data sets. There is evidence for P450 evolution through tandem duplications, with several scaffolds in the *A. pisum* genome assembly possessing two or more P450 loci that are closely related to one another.

Analyses of other sequenced arthropod genomes have identified 143 P450s in *Tribolium castaneum* (red flour beetle), 106 in *Anopheles gambiae* (malaria mosquito), 86 in *Bombyx mori* (silk moth), 85 in *Drosophila melanogaster* (fruit fly), 46 in *Apis mellifera* (honeybee) and 75 in *Daphnia pulex* (Baldwin *et al*., 2009; Claudianos *et al*., 2006; Li *et al*., 2005; Richards *et al*., 2008; http://drnelson.utmem.edu/CytochromeP450.html). Therefore, *A. pisum* appears to have a fairly typical number of P450s. However, if one assumes that some of the 83 putative P450 loci are pseudogenes with incomplete P450 domains, then the *A. pisum* P450 complement would be toward the lower end of sequenced insect genomes. This relatively low number of P450s may be particularly interesting in light of mounting evidence that the pea aphid has a tendency to accumulate duplicated genes over time (International Aphid Genomics Consortium, 2009). It has been suggested that the social organization of the bee hive, which may shield the *A. mellifera* queen and larvae from environmental exposure to toxins, has permitted a relative loss of P450 loci in this species (Claudianos *et al*., 2006). However, this is certainly not the case in *A. pisum*.

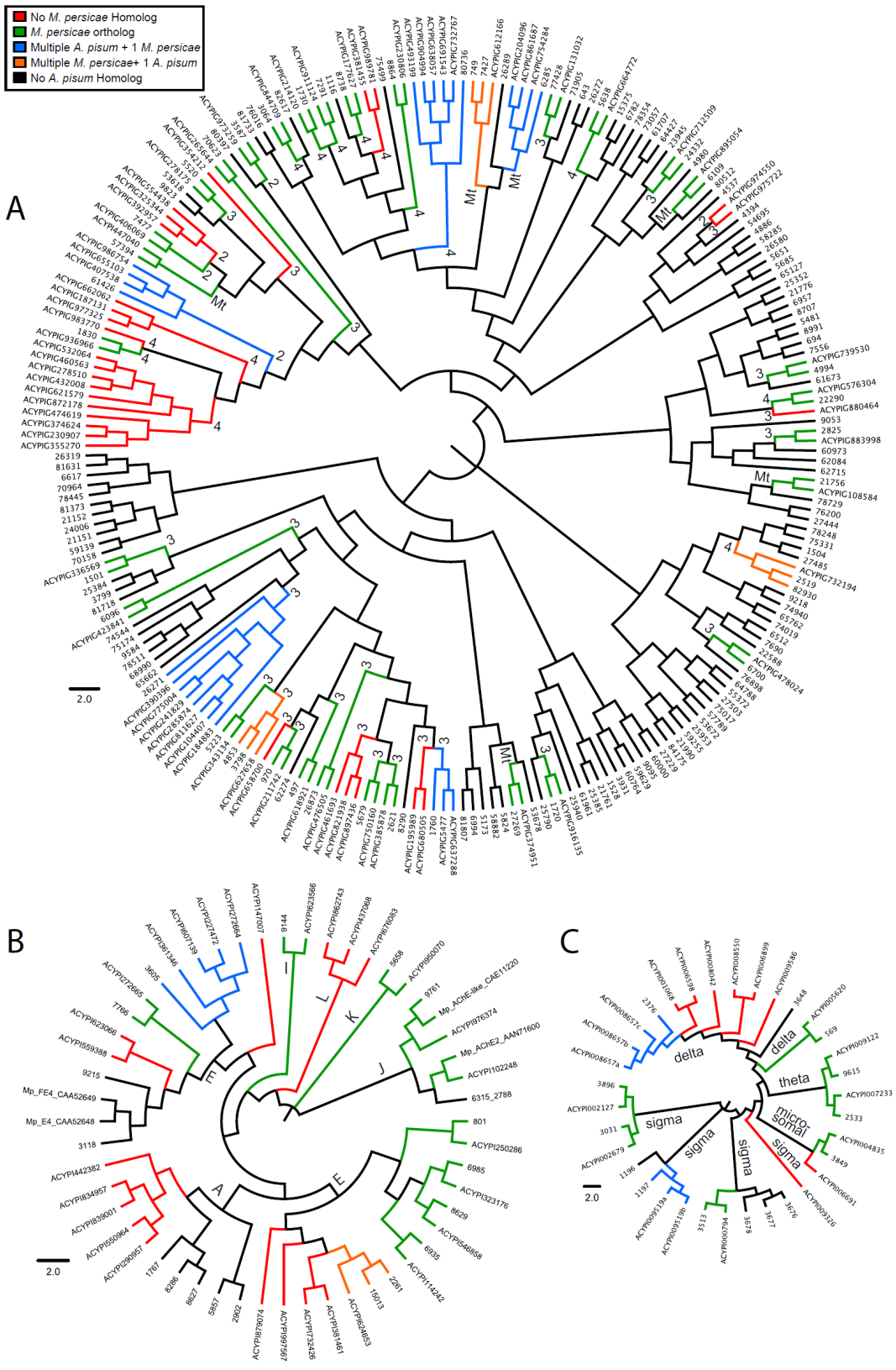
A subset of *A. pisum* P450s that are predicted to be involved in 20-hydroxy-ecdysone biosynthesis (orthologs of the fruit fly loci *disembodied* and *shade*) appear to have undergone duplications in the *A. pisum* genome (Supplemental Table 4.3). This is in contrast to evidence from holometabolous insects, where it has been suggested that the presence of only single genes for these enzymes represents a structural and evolutionary constraint (Rewitz *et al*., 2007). Duplications of *shade* were not observed

among the *M. persicae* ESTs, suggesting that this observation may be specific to the *A. pisum* genome.

Analysis of *M. persicae* ESTs identified more than 150 P450-related sequences. Unique genes, as opposed to allelic variants of the same P450 gene, were identified based on the contig assembly parameters described above. Many of the DNA sequences could not be confidently assigned as P450s based on translations, perhaps representing the non-coding regions, and most of the EST contigs represented incomplete coding regions. However, based on significant homology to the *A. pisum* P450s (e-value ≤ 1E-4), at least 115 of the *M. persicae* EST contigs represent expressed P450 loci. Neighbor joining trees based on Clustal alignments (Larkin *et al*., 2007) of either DNA sequences or protein translations indicated that 30% to 50% of *A. pisum* P450 loci have an ortholog in *M. persicae* (Figure 4.1A).

Although the P450 loci from the pea aphid genomic sequence could be readily assigned to clades through the construction of phylograms, the addition of fragments of P450 loci from the *M. persicae* EST clusters increased the level of 'noise' in the phylograms. This is evident in Figure 4.1A, where clade assignments for the pea aphid P450s (labeled nodes in the figure) are somewhat dispersed. However, potential *M. persicae* orthologs were identified using phylograms of DNA sequences from the *M. persicae* EST clusters and the predicted pea aphid DNA sequences. Putative orthologs are indicated in the supplementary table 4.3. Sequences were considered putatively orthologous if they were paired with a pea aphid gene and had bootstrap support greater than 50%. Although there is considerable uncertainty in the specific assignment of the *M. persicae* sequences at the clade level (Table 4.1), it is clear that some of the pea aphid sequences group separately from some *M. persicae* sequences, indicating divergence in the P450 complement between the two species.

Figure 4.1. Comparison of cytochrome P450 (A), esterase (B), and GST (C) sequences from *A. pisum* and *M. persicae*. Predicted protein coding genes from *A. pisum* were compared to EST sequences from *M. persicae* using the Clustal alignment algorithm and a neighbor-joining tree. *A. pisum* sequences with putative orthologs in *M. persicae* are highlighted with green branches and those with no relatives present in the *M. persicae* EST data are indicated with red branches. In some cases, clear orthologs were not established due to many *A. pisum* sequences being grouped with a single *M. persicae* sequence (blue branches) or a single *A. pisum* sequence being grouped with many *M. persicae* sequences (orange branches). *M. persicae* sequences for which there is no clear *A. pisum* homolog are shown with black branches. *A. pisum* genes are represented by ACYPI numbers, and *M. persicae* genes are represented by contig numbers from the current analysis (http://www.aphidbase.com/aphidbase/downloads, files: MyzusSanger, Myzus454), with the exception of previously studied esterases that are included as GenBank identifiers. The scale bars are proportional to the number of amino acid changes between different proteins.


No M. persicae Homolog
M. persicae ortholog
Multiple A. pisum + 1 M. persicae
Multiple M. persicae+ 1 A. pisum
No A. pisum Homolog
A
B
C
2.0
delta
theta
sigma
micro-somal
Mt
Mp_AChE-like_CAE11220
Mp_AChE2_AAN71600
Mp_FE4_CAA52649
Mp_E4_CAA52649

It is impossible to determine whether the remaining *A. pisum* loci have *M. persicae* orthologs, because the available EST data likely represent only a fraction of all *M. persicae* P450s. On the other hand, more than 100 of the *M. persicae* P450 ESTs have no clear match in the *A. pisum* genome, suggesting that there has been an expansion of this gene family in *M. persicae* or, conversely, a contraction in *A. pisum*. Most species in the Aphidinae (von Dohlen *et al*., 2006) are more specialized than *M. persicae* in their feeding habits, but the number of P450 enzymes encoded in their genomes is as yet unknown.

Carboxyl/cholinesterases

There are 30 members of CCE superfamily in the *A. pisum* genome (Supplemental Table 4.4), compared to the 24, 35, and 51 that have been identified in *A. mellifera*, *D. melanogaster* and *A. gambiae,* respectively (Table 4.2; Claudianos *et al*., 2006). All of the *A. pisum* genes appear functional, though several lack EST support and a few are truncated, likely due to errors with the genome assembly. In comparison, there are 19 to 23 identifiable CCEs in the *M. persicae* EST unigene set, 13 with putative *A. pisum* homologs and 6 to 10 esterase-like genes that do not have any obvious homologs in the *A. pisum* genome (Figure 4.1B). This suggests that there is diversification of enzyme functions in the esterase gene family, at least for those that are not involved in basal metabolism that would likely be common to all aphids.

Known CCEs can be divided into 13 clades (Ranson *et al*., 2002), seven of which are represented in *A. pisum* and *M. persicae* (Table 4.2). Clades without identifiable *A. pisum* or *M. persicae* homologs are the Diptera-specific clades B and C, integument esterases (D), dipteran juvenile hormone esterases (F), lepidopteran juvenile esterases (G) and the glutactin like esterases (H). Thirteen of the *A. pisum* esterase genes are found in small clusters of 2 to 6 genes per scaffold. This is

particularly apparent in clade E, where there appears to be an expansion in *A. pisum*. However, this level of duplication is less than that found in the Diptera, where 8 to 10 CCEs may cluster together (Campbell *et al*., 2003). In *A. pisum*, the largest CCE cluster consists of 6 genes that share no more than 60% amino acid identity to each other (Figure 4.1B), suggesting fairly ancient duplication events.

Table 4.2. Carboxyl/cholinesterase genes

	Clade	*Drosophila melanogaster*	*Apis mellifera*	*Acyrthosiphon pisum*	*Myzus persicae*[1]
Dietary	A	0	8	5	5
	B	2	0	0	0
	C	11	0	0	0
Pheromone and hormone processing	D	3	1	0	0
	E	3	3	18	12
	F	2	0	0	0
	G	0	1	0	0
Neuro and develop-mental	H	4	0	0	0
	I	2	2	1	1
	J	1	2	2	3
	K	1	1	1	1
	L	4	5	3	0
	M	2	1	0	0
	Total	**35**	**24**	**29**	**22**

[1]Numbers based on *M. persicae* EST data

Esterases in clades A-C are involved in the detoxification of xenobiotics. *A. pisum* or *M. persicae* only possess esterases in clade A, which also contains an esterase linked to organophosphate resistance in *Anisopteromalus calandrae*, a parasitic wasp (Zhu *et al*., 1999). It has been suggested that the eight clade A esterases in *A. mellifera* represent an order-specific radiation within the Hymenoptera

(Claudianos *et al*., 2006). However, given a similar radiation in *A. pisum* and *M. persicae,* it would seem that an order-specific radiation in clade A is not unique to *A. mellifera*. Comparing *A. pisum* and *M. persicae* shows that they both have five members of clade A. However, the absence of direct homologs suggests an ancient radiation event or independent expansions of this gene family (Figure 4.1B).

A. pisum shows a reduction in diversity in CCEs involved in hormone and pheromone processing (clades D-H) with only clade E having any members at all (Table 4.2). However, within clade E, *A. pisum* esterase genes have undergone a considerable expansion to 18 genes, compared to three in *D. melanogaster* and *A. mellifera* and five in *A. gambiae* (Claudianos *et al*., 2006). When the 18 *A. pisum* clade E CCEs are aligned with CCEs from other species they form a monophyletic clade. However when aligned with *M. persicae* they split into two sub-clades, perhaps a reflection of the diversity within clade E in aphids (Figure 4.1B). The 11 - 12 *M. persicae* unigenes in clade E suggest that this expansion is not unique to *A. pisum*. Clade E enzymes are thought to be largely involved with pheromone and hormone processing in insects, and it is interesting that this gene family has been expanded in *A. pisum* and *M. persicae* (Table 4.2). Although, aphid alarm and sex pheromones have been identified (Dawson *et al*., 2005; Hatano *et al*., 2008; Verheggen *et al*., 2008), aphid pheromone communication is almost certainly less complex than that of honeybees (Slessor *et al*., 2005). Therefore, the relative expansion of the aphid CCE clade E family likely serves a different function.

In organophosphate-resistant *M. persicae*, the E4 esterase gene is often found in clusters of up to 80 virtually identical copies, and is associated with another esterase, FE4 (Field and Devonshire, 1998). Although the closest *A. pisum* homolog of *M. persicae* E4 esterase (ACYPI623066) is also clustered with other esterase genes, there is only one copy in the sequenced genome (Figure 4.1B). ACYPI623066 and

ACYPI559388, the *A. pisum* genes most similar to *M. persicae* FE4 esterase (Figure 4.1B), also generated the most hits in comparisons to the *M. persicae* transcript data (Supplemental Table 4.4).

M. sexta AAG42021	LKWVQRNAHFFGGRPDDVTLM**GQSAG**AAATHILSLSKAADGLF
B. mori AAL55240	LRWVKRNARAFGGNPDNVTLA**GQSAG**AAAAHLLTLSKATEGLV
H. virescens AAC38822	LRWVQRNAKNFGGDPSDITIA**GQSAG**ASAAHLLTLSKATEGLF
D. melanogaster CG8425	LRWVQRNIRFFGGDPQRVTIF**GQSAG**GVAAHMHLLSPRSHGLF
A. pisum ACYPI381461, 1st	LKWVQKNIDKFGGDPKKVTLF**GQSAG**SASVGLHLLSPMSKGLF
A. pisum ACYPI381461, 2nd	TLNDPLNAVKVCFILLKVTLF**GQSAG**SASVGLHLLSKMSKGLF
A. pisum ACYPI929836	LQWVKKNIDKFGGNRRKVTLF**GQSAG**SASVGLHLLSKMSKGLF

Figure 4.2: Alignment of known juvenile hormone esterase (JHE) sequences with the potential pea aphid JHEs. Almost all known JHE enzymes possess a GQSAG motif (in bold) in the nucleophillic elbow of the active site.

Clades F and G contain validated juvenile hormone esterases (JHE) for Diptera and Lepidoptera respectively. Given their phylogenetic specificity, it is perhaps not surprising that *A. pisum* and *M. persicae* do not have members of these clades. However ACYPI381461 and ACYPI929836 in clade E are candidate aphid JHEs. The predicted coding sequence of ACYPI381461 contains two GQSAG nucleophilic elbow motifs, whereas ACYPI929836 has one. The GQSAG motif is present in the active site in all functionally validated JHEs, though there is usually only one in each protein. All *A. pisum* motifs align with the JHEs from *D. melanogaster* and lepidopteran species (Figure 4.2). Other analyses show that ACYPI381461 is somewhat longer than the other aphid esterases in clade E and does not have an obvious *M. persicae* homolog.

With the exception of acetylcholine esterase (ACHE, clade J), members of the neurological/developmental group (Table 4.2, clades I-M) tend to be non-catalytic and are involved in cell-cell interactions. Seven members of this group were identified in *A. pisum* (5 in *M. persicae*; Table 4.2), less than the 10 to 12 identified in *A. mellifera*, *D. melanogaster* and *A. gambiae* (Claudianos *et al*., 2006). Like *A. mellifera* and *A.*

gambiae, *A. pisum* has two AChEs, one of which one (ACYPI102248) has a likely neurological function. In contrast, *M. persicae* EST data show three potential AChEs. Other CCEs found in aphids include the neuroligins (clade L), conserved structural proteins involved in synapse formation, gliotactin (clade K), which is thought to be a structural protein, and clade I, which has unknown function (Table 4.2).

Glutathione S-*transferases*

Analysis of the *A. pisum* genome identified 20 putative members of the GST superfamily (Figure 4.1C; Supplemental Table 4.5). This is fewer than the number of loci identified in *Drosophila* (38) and *Anopheles* (31) but more than *A. mellifera* (10), which has an reduced number of genes encoding detoxification enzymes (Claudianos *et al*., 2006; Ranson *et al*., 2001). cDNA sequences provide gene expression evidence for 15 of the 20 likely *A. pisum* GSTs.

Table 4.3: Glutathione *S*-transferase genes

Class	***Drosophila melanogaster***	***Apis mellifera***	***Acyrthosiphon pisum***	***Myzus persicae***[1]
Delta	11	1	10	8
Epsilon	14	0	0	0
Omega	5	1	0	0
Sigma	1	4	6	8
Theta	4	1	2	2
Zeta	2	1	0	0
Microsomal	1	2	2	2
Total	**38**	**10**	**20**	**21**

[1]Numbers based on *M. persicae* EST data

Insects generally harbor six different classes of GSTs (Chelvanayagam *et al*., 2001). However, although the *A. pisum* genome encodes two microsomal GSTs and GSTs in the delta (10), theta (2), and sigma (6) classes, it apparently lacks the epsilon, omega

and zeta classes (Table 4.3).

The delta and epsilon GST classes are found uniquely in insects (Ranson *et al*., 2001) and have been implicated in insecticide resistance. However, *A. pisum* has fewer members of these two GSTs classes than *D. melanogaster* and *A. gambiae*.

EST data from the generalist aphid *M. persicae* show at least 15 and a maximum 21 of GST-like genes (Figure 4.1C; Table 4.3). *M. persicae* homologs were found for all except two of the predicted *A. pisum* genes. Eight *M. persicae* genes are obviously homologous to genes identified in *A. pisum*, whereas others seemed to have diversified (Figure 4.1C). As in the case of *A. pisum,* the epsilon, omega and zeta GST classes are lacking in *M. persicae* (Table 4.3), suggesting that these three GST classes may be absent from aphids in general. Analysis of cDNA libraries from specific tissue types (Ramsey *et al*., 2007) suggests that one GST (contig 1196) is specifically expressed in gut tissue, whereas another (contig 3648) is over-represented in the salivary glands of *M. persicae.*

A. pisum EST data show that several alternative splicing variants that are apparently derived from a single GST gene (Figure 4.1C; Supplemental Table 4.5). Although the *M. persicae* EST data also suggest GST alternative splicing, these do not occur for the same gene as in *A. pisum*. Such alternative splicing could be an evolutionary strategy to broaden the spectrum of effectiveness in detoxification enzymes, and might thus allow aphids to expand their host range or to effectively metabolize xenobiotics such as plant secondary metabolites and insecticides (Ranson *et al*., 2001). Further analysis of the *A. pisum* genome showed the presence of enzymes that are potentially involved in the degradation of conjugated glutathione, including eight gamma-glutamyl transpeptidases and 34 aminopeptidases (Supplemental Table 4.5). These *A. pisum* genes along with their

M. persicae EST homologs (Supplemental Table 4.5), indicate that aphids have the complete pathway for degradation of xenobiotics via conjugation to glutathione.

Conclusions

The *A. pisum* genome and large-scale sequencing of *M. persicae* ESTs have partially made it possible to determine whether a broad generalist insect herbivore has a greater diversity of detoxification enzymes than a specialist. There is apparently no great expansion of the GSTs and CCEs in *M. persicae* relative to *A. pisum*, and the number of interspecies differences in the presence and absence of specific genes are likely to be similar those found for other aphid gene families. In the case of the CCEs, a relatively large fraction of these proteins are involved in basal metabolic functions that are likely to be the same or similar in *A. pisum* and *M. persicae.* Although it is difficult to estimate exact gene numbers from the EST data, the P450 gene family is at least 40% larger in *M. persicae* than in *A. pisum.* If one assumes similar gene representation in EST libraries produced from the two species (115 *vs*. 51 unique genes), then *M. persicae* may encode twice as many P450s as *A. pisum.* This expansion may reflect the different host ranges of the two aphid species, only Fabaceae for *A. pisum* and 40 different plant families for *M. persicae*, and is consistent with the hypothesis that a phloem-feeding generalist insect herbivore would require a greater number of detoxification enzymes than a specialist.

Experimental Procedures

RNA isolation and cDNA synthesis:

Total RNA was isolated using the RNeasy kit (Qiagen, USA) from 100-200 aphids, a mixture of adults and all larval stages. RNA samples were prepared from two different lineages of *M. persicae*: tobacco-adapted *M. persicae* feeding on tobacco, and non-adapted *M. persicae* feeding on cabbage. mRNA was purified using Oligotex (Qiagen, USA) and precipitated to increase the concentration to ~100 ng/uL. cDNA synthesis and normalization were performed using a modification of the Creator SMART cDNA synthesis protocol (Clontech, USA) in conjunction with the Trimmer Direct cDNA normalization kit (Evrogen, Moscow, Russia). A modified reverse transcription primer, 5'-AAGCAGTGGTATCAACGCAGAGTGGCCGAGGTTTTGTTTTTTTTTCTTTTTTTTTTVN-3',
was used. To avoid potential problems associated with 454 sequencing of homopolymeric stretches, adaptor-ligated poly-T primers for reverse transcription were modified to disrupt the poly-T tract every several bases. The presence of the SMART IV oligo (Clontech, USA) in the reverse transcription reaction established the 3' end of the reverse transcription product as the reverse complement of the 5' adaptor. Therefore, the subsequent second strand synthesis and amplification (16 cycles) was accomplished with one primer (5' PCR primer, 5'-AAGCAGTGGTATCAACGCAGAGT-3'), which anneals to the identical 3' ends of complementary strands.

cDNA normalization

Normalization of cDNA by duplex-specific nuclease treatment was performed to increase the relative abundance of low-expression transcripts (Zhulidov *et al*., 2004). Normalized cDNA was subjected to 2 rounds of amplification: 20 cycles in the first round and 12 cycles in the second round. cDNA was purified and sequenced with a Roche GS-FLX system.

Sequence assembly and analysis

Primer and adaptor sequences were trimmed from sequences, and any trimmed sequences less than 30 bp were discarded. In an effort to identify sequences not represented in previous Sanger EST sequencing of *M. persicae*, 454 sequences were aligned to an existing unigene set (Ramsey *et al*., 2007). 454 sequences with over 95% overlap with an existing unigene were discarded. The remaining sequences were clustered with MCL software (Enright *et al*., 2002), and consensus contigs were formed from each cluster using CAP3 software (Huang and Madan, 1999). Contigs and singletons were BLASTed against the *M. persicae* Sanger unigene set, and sequences with more than 75% overlap with the Sanger unigene were discarded, as were sequences less than 100 bp in length.

Annotation of 454 unigenes

Annotation of *M. persicae* sequences relied on BLAST (Altschul *et al*., 1997; Tatusova and Madden, 1999) to identify the most similar sequences among *Drosophila* and *A. pisum* RefSeq proteins, as well as *A. pisum* genomic scaffolds. BioPerl modules and custom Perl scripts were used to parse GenBank files and extract gene descriptions for top BLAST hits.

Phylogenetic trees

Protein sequences for *A. pisum* were downloaded from AphidBase and/or GenBank. *M. persicae* ESTs were obtained from GenBank and converted into protein sequences. Trees are built using the amino acid sequence and neighbor-joining using FigTree (version1.2.1.; Andrew Rambaut; http://tree.bio.ed.ac.uk/software/figtree/).

Acknowledgements

This work was funded by USDA grant 2005-35604-15446 to GJ and by the CSIRO Office of the Chief Executive postdoctoral fellowship scheme. The authors would also like to thank Qi Sun for help with sequence assemblies, and Alexandre dos Santos Cristino and John Oakeshott for their comments and advice in the preparation of this manuscript.

Authors' Contributions

DSR, GJ, JSR, KG, MDV, and TKW annotated aphid detoxification genes. BAR and SM performed Roche 454 sequencing of *Myzus persicae* cDNA. LP assembled *M. persicae* cDNA sequences. JSR annotated *M. persicae* unigenes. JSR, GJ, DSR, MDV, and TKW wrote the paper.

Description of Additional Data Files (Excel spreadsheets provided on CD)

Supplementary Table 4.1: Best blast hit of unclustered M. persicae unigenes against *A. pisum* Glean/Refseq proteins. Evalue cutoff (blastx) 1e-3. Genbank gene descriptions are not available for Glean proteins.

Supplementary Table 4.2: Best blast hit of clustered M. persicae unigenes against

A. pisum Glean/Refseq proteins. Evalue cutoff (blastx) 1e-3. Genbank gene descriptions are not available for Glean proteins.

Supplementary Table 4.3: Predicted cytochrome P450s in the *A. pisum* genome and M. persicae orthologs

Supplementary Table 4.4: Carboxy/cholinesterases encoded in the *A. pisum* genome

Supplementary Table 4.5: Glutathione S-transferases and enzymes involved in glutathione metabolism

REFERENCES

Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W and Lipman, DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-402.

Baldwin, WS, Marko, PB and Nelson, DR (2009) The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*. *BMC Genomics* **10**: 169.

Becerra, JX (2007) The impact of herbivore-plant coevolution on plant community structure. *PNAS* **104**: 7483-7488.

Bernays, EA and Graham, M (1988) On the evolution of host specificity in phytophagous arthropods. *Ecology* **69**: 886-892.

Bino, RJ, Hall, RD, Fiehn, O, Kopka, J, Saito, K, Draper, J, Nikolau, BJ, Mendes, P, Roessner-Tunali, U, Beale, MH, Trethewey, RN, Lange, BM, Wurtele, ES and Sumner, LW (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* **9**: 418-425.

Blackman, RL and Eastop, VF (2000) *Aphids on the World's Crops*. Wiley, Chichester. 466 pp.

Campbell, PM, de, QRGC, Court, LN, Dorrian, SJ, Russell, RJ and Oakeshott, JG (2003) Developmental expression and gene/enzyme identifications in the alpha esterase gene cluster of Drosophila melanogaster. *Insect Mol Biol* **12**: 459-71.

Chelvanayagam, G, Parker, MW and Board, PG (2001) Fly fishing for GSTs: a unified nomenclature for mammalian and insect glutathione transferases. *Chem Biol Interact* **133**: 256-260.

Claudianos, C, Ranson, H, Johnson, RM, Biswas, S, Schuler, MA, Berenbaum, MR, Feyereisen, R and Oakeshott, JG (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol*

Biol **15**: 615-36.

Dawson, GW, Griffiths, DC, Merritt, LA, Mudd, A, Pickett, JA, Wadhams, LJ and Woodcock, CM (2005) Aphid semiochemicals - A review and recent advances on the sex pheromones. *J Chem Ecol* **16**: 1573-1561.

Devine, GJ, Harling, ZK, Scarr, AW and Devonshire, AL (1996) Lethal and sublethal effects of imidacloprid on nicotine-tolerant *Myzus nicotinianae* and *Myzus persicae*. *Pest Sci* **48**: 57-62.

Enayati, AA, Ranson, H and Hemingway, J (2005) Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* **14**: 3-8.

Enright, AJ, Van Dongen, S and Ouzounis, CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575-84.

Feyereisen, R (1999) Insect P450 enzymes. *Annual Review of Entomology* **44**: 507-33.

Feyereisen, R (2006) Evolution of insect P450. *Biochem Soc Trans* **34**: 1252-5.

Field, LM (2000) Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochem J* **349**: 863-8.

Field, LM and Devonshire, AL (1998) Evidence that the *E4* and *FE4* esterase genes responsible for insecticide resistance in the aphid *Myzus persicae* (Sulzer) are part of a gene family. *Biochem J* **330 (Pt 1)**: 169-73.

Francis, F, Haubruge, E, Gasper, C and Dierickx, PJ (2001) Glutathione S-transferases of *Aulacorthum solani* and *Acyrthosiphon pisum*: partial purification and characterization *Comp Biochem Physiol B* **129**: 165-171.

Francis, F, Vanhaelen, N and Haubruge, E (2005) Glutathione S-transferases in the adaptation to plant secondary metabolites in the *Myzus persicae* aphid. *Arch Insect Biochem Physiol* **58**: 166-74.

Golawska, S and Lukasik, I (2008) Acceptance of low-saponin lines of alfalfa with

varied phenolic concentrations by pea aphid (Homoptera: Aphididae). *Biologia* **64**: 377-382.

Hatano, E, Kunert, G, Bartram, S, Boland, W, Gershenzon, J and Weisser, WW (2008) Do aphid colonies amplify their emission of alarm pheromone? *J. Chem. Ecol.* **34**: 1149-52.

Huang, X and Madan, A (1999) CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868-77.

International Aphid Genomics Consortium (2009) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biology,* in review.

Kim, JH, Lee, BW, Schroeder, FC and Jander, G (2008) Identification of indole glucosinolate breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). *Plant Journal* **54**: 1015-1026.

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ and Higgins, DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-8.

Li, B, Xia, Q, Lu, C, Zhou, Z and Xiang, Z (2005) Analysis of cytochrome P450 genes in silkworm genome (*Bombyx mori*). *Sci China C Life Sci* **48**: 414-8.

Li, W, Petersen, RA, Schuler, MA and Berenbaum, MR (2002) CYP6B cytochrome p450 monooxygenases from *Papilio canadensis* and *Papilio glaucus*: potential contributions of sequence divergence to host plant associations. *Insect Mol Biol* **11**: 543-51.

Li, X, Schuler, MA and Berenbaum, MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology* **52**: 231-53.

Low, WY, Ng, HL, Morton, CJ, Parker, MW, Batterham, P and Robin, C (2007)

Molecular evolution of glutathione *S*-transferases in the genus *Drosophila*. *Genetics* **177**: 1363-75.

Mao, W, Berhow, MA, Zangerl, AR, McGovern, J and Berenbaum, MR (2006) Cytochrome P450-mediated metabolism of xanthotoxin by *Papilio multicaudatus*. *J. Chem. Ecol.* **32**: 523-36.

Mao, W, Schuler, MA and Berenbaum, MR (2007) Cytochrome P450s in *Papilio multicaudatus* and the transition from oligophagy to polyphagy in the Papilionidae. *Insect Mol Biol* **16**: 481-90.

Maïbèche-Coisne, M, Nikonov, AA, Ishida, Y, Jacquin-Joly, E and Leal, WS (2004) Pheromone anosmia in a scarab beetle induced by in vivo inhibition of a pheromone-degrading enzyme. *Proc Natl Acad Sci U S A* **101**: 11459-64.

Mita, K, Kasahara, M, Sasaki, S, Nagayasu, Y, Yamada, T, Kanamori, H, Namiki, N, Kitagawa, M, Yamashita, H, Yasukochi, Y, Kadono-Okuda, K, Yamamoto, K, Ajimura, M, Ravikumar, G, Shimomura, M, Nagamura, Y, Shin, IT, Abe, H, Shimada, T, Morishita, S and Sasaki, T (2004) The genome sequence of silkworm, *Bombyx mori*. *DNA Res* **11**: 27-35.

Nauen, R, Strobel, J, Tietjen, K, Yuichi, O, Erdelen, C and Elbert, A (1996) Aphicidal activity of imidacloprid against a tobacco feedins strain of *Myzus persicae* (Homoptera: Aphididae) from Japan closely related to *Myzus nicotinianae* and highly resistant to carbamates and organophosphates. *Bull Ent Res* **86**: 165-171.

Oakeshott, JG, Claudianos, C, Russell, RJ and Robin, GC (1999) Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. *Bioessays* **21**: 1031-42.

Ohta, N, Mori, N, Kuwahara, Y and Nishida, R (2005) A hemiterpene glucoside as a probing deterrent of the bean aphid, *Megoura crassicauda*, from a non-host

vetch, *Vicia hirsute*. *Phytochemistry* **67**: 584-588.

Ramsey, JS, Wilson, AC, de Vos, M, Sun, Q, Tamborindeguy, C, Winfield, A, Malloch, G, Smith, DM, Fenton, B, Gray, SM and Jander, G (2007) Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* **8**: 423.

Ranson, H, Claudianos, C, Ortelli, F, Abgrall, C, Hemingway, J, Sharakhova, MV, Unger, MF, Collins, FH and Feyereisen, R (2002) Evolution of supergene families associated with insecticide resistance. *Science* **298**: 179-81.

Ranson, H, Rossiter, L, Ortelli, F, Jensen, B, Wang, X, Roth, CW, Collins, FH and Hemingway, J (2001) Identification of a novel class of insect glutathione *S*-transferases involved in resistance to DDT in the malaria vector *Anopheles gambiae*. *Biochem J* **359**: 295-304.

Rewitz, KF, O'Connor, MB and Gilbert, LI (2007) Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation. *Insect Biochem Mol Biol* **37**: 741-53.

Richards, S, *et al*. (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949-55.

Rosenthal, GA and Berenbaum, MR (1991) *Herbivores: Their Interactions With Secondary Plant Metabolites, Second Edition, Vol. I: The Chemical Participants.* Academic Press, San Diego.

Schoonhoven, LM, Jermey, T and van Loon, JJA (1998) *Insect-Plant Biology: From Physiology to Evolution*. Chapman & Hall, London.

Scott, JG (1999) Cytochromes P450 and insecticide resistance. *Insect Biochem Mol Biol* **29**: 757-77.

Slessor, KN, Winston, ML and Le Conte, Y (2005) Pheromone Communication in the Honeybee (*Apis mellifera* L.) *J Chem Ecol* **31**: 2731-2745.

Snyder, MJ and Glendinning, JI (1996) Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. *J Comp Physiol [A]* **179**: 255-61.

Tatusova, TA and Madden, TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247-50.

Van Dongen, S (2008) Graph clustering via a discrete uncoupling process. *Siam Journal on Matrix Analysis and Applications* **30**: 121-141.

Verheggen, FJ, Mescher, MC, Haubruge, E, Moraes, CM and Schwartzberg, EG (2008) Emission of alarm pheromone in aphids: a non-contagious phenomenon. *J. Chem. Ecol.* **34**: 1146-8.

von Dohlen, CD, Rowe, CA and Heie, OE (2006) A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. Molecular *Mol Phylogenet Evol* **38**: 316-329.

Walling, LL (2008) Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiol* **146**: 859-66.

Walz, C, Giavalisco, P, Schad, M, Juenger, M, Klose, J and Kehr, J (2004) Proteomics of curcurbit phloem exudate reveals a network of defence proteins. *Phytochemistry* **65**: 1795-1804.

Zhu, YC, Dowdy, AK and Baker, JE (1999) Differential mRNA expression levels and gene sequences of a putative carboxylesterase-like enzyme from two strains of the parasitoid *Anisopteromalus calandrae* (Hymenoptera: Pteromalidae). *Insect Biochem Mol Biol* **29**: 417-25.

Chapter 5: Induction of detoxification and salivation-associated gene expression by nicotine in *Myzus persicae*

John S. Ramsey[1], Pooja Sarkar[2], Yi-Ran Xin[1], Georg Jander[1]

[1] Boyce Thompson Institute, Ithaca, NY 14853, USA

[2]Cornell University, Department of Chemistry, Ithaca, NY 14853, USA

Abstract

Barriers to host range expansion by generalist insect herbivores include plant defensive metabolites. Lineages of *Myzus persicae* that have expanded their host range to include tobacco are characterized by increased resistance to nicotine. Red aphids of the tobacco-adapted lineage used in this study can survive and reproduce on liquid diets containing nicotine at concentrations approximately 10-fold higher than diets lethal to green aphids of a non-adapted lineage. Red aphid fecundity is increased by moderate amounts of dietary nicotine, suggesting that exposure to this plant toxin may improve aphid performance. Compared to populations that have been reared exclusively on cabbage, green aphids have significantly higher fecundity on tobacco when they have been reared on tobacco for at least three generations. We used *M. persicae* gene expression microarrays to identify transcripts up-regulated in aphid heads in response to nicotine. Induced expression was found for genes previously associated with insect responses to plant defenses and insecticides, as well as genes encoding aphid salivary proteins.

Introduction

The generalist aphid *Myzus persicae* (Hemiptera: Aphididae) feeds on more than 400

plant species from over 40 families, including several economically significant crop plants (Blackman and Eastop, 2000). Lineage-specific adaptations conferring resistance to plant defenses associated with new host plants have been documented in *M. persicae* populations. For instance, artificial diet experiments showed that Australian lineages of *M. persicae* which have expanded their host range to include narrow-leafed lupine (*Lupinus angustifoluis*) are more resistant to the lupine-specific alkaloid lupanine than non-adapted lineages (Cardoza *et al*., 2006). Resistance to fumigation by the tobacco alkaloid nicotine has been reported in tobacco-adapted lineages of *M. persicae*, with nicotine-resistant clones also displaying tolerance to the neonicotinoid insecticide imidacloprid (Devine at al., 1996). An understanding of *M. persicae* genes involved in nicotine resistance will shed light on mechanisms of host range expansion by this polyphagous herbivore, and may inform future crop protection strategies.

Nicotine resistance in the lepidopteran tobacco specialist *Manduca sexta* (Lepidoptera: Sphingidae) is conferred enzymatically by cytochrome P450 monooxygenases (Synder *et al*, 1993), and physically by nicotine pumps in the malpighian tubules and the blood -brain barrier, which facilitate excretion and prevent the delivery of the toxin to its biological target, the nicotinic acetylcholine receptor (Murray *et al*., 1994). *M. sexta* larvae prefer nicotine non-producing mutants of *Nicotiana attenuata* over wild type plants in choice experiments, and grow faster on the mutant plants (Steppuhn *et al*., 2004). Natural variation in nicotine resistance between populations of *Drosophila melanogaster* (Diptera: Drosophilidae) was used to identify quantitative trait transcripts (QTT) for nicotine resistance (Passador-Gurgel *et al*., 2007). Nicotine treatment increased expression of γ-aminobutyric acid (GABA) biosynthesis genes and decreased expression of glutamate biosynthesis genes in resistant *D. melanogaster* populations, suggesting a role for the modulation of amino

acid neurotransmitter levels in the response to nicotine.

Since the emergence of *M. persicae* as a pest on tobacco over 60 years ago, tobacco-feeding populations have been documented across the globe (Margaritopoulos *et al*., 2009). Variation in color between aphid lineages is common: whereas non-adapted lineages of *M. persicae* are typically green or yellow, tobacco-adapted *M. persicae* lineages are generally dark red in color. There are, however, many exceptions to this over-simplified classification. Recent surveys in Japan, where tobacco-feeding populations were originally identified, showed an increase in the number of green clones feeding on tobacco relative to previous studies (Shigehara and Takada, 2003).

Recent expansion of genome and transcriptome sequence data for *M. persicae* and the closely related pea aphid, *Acyrthosiphon pisum*, have paved the way for functional genomics studies, which have the potential to identify aphid genes involved in resistance to plant defenses and insecticides (International Aphid Genomics Consortium, 2009, Ramsey *et al*., 2007). The goal of this study was to compare the response to nicotine in artificial diets and plants between tobacco adapted and non-adapted aphid lineages. Furthermore, a *M. persicae* microarray was used to identify genes induced by acute nicotine exposure in heads of tobacco-adapted aphids.

Results

Nicotine resistance in tobacco-adapted *M. persicae*

The red aphid used in this study was derived from a population adapted to tobacco, whereas the green aphid originated from a non-tobacco producing region, and only colonized tobacco in our growth facility when caged on plants in a no-choice context. Aphid fecundity on artificial diets was determined by counting the number of progeny from four adult aphids in three days. While the fecundity of green and red *M. persicae* is comparable when feeding on control diets, red aphids were able to survive and reproduce on diets containing nicotine at concentrations approximately 10-fold higher

than levels lethal to aphids of the green lineage (Fig. 5.1).

Aphid relative fecundity is the ratio of the number of progeny from aphids feeding on nicotine-containing diets to the number of progeny from aphids feeding on control diets. Aphid fecundity data were transformed into measures of relative aphid fecundity, in order to identify the nicotine concentration for the red and green lineage at which the number of progeny is half that of aphids feeding on control diets.

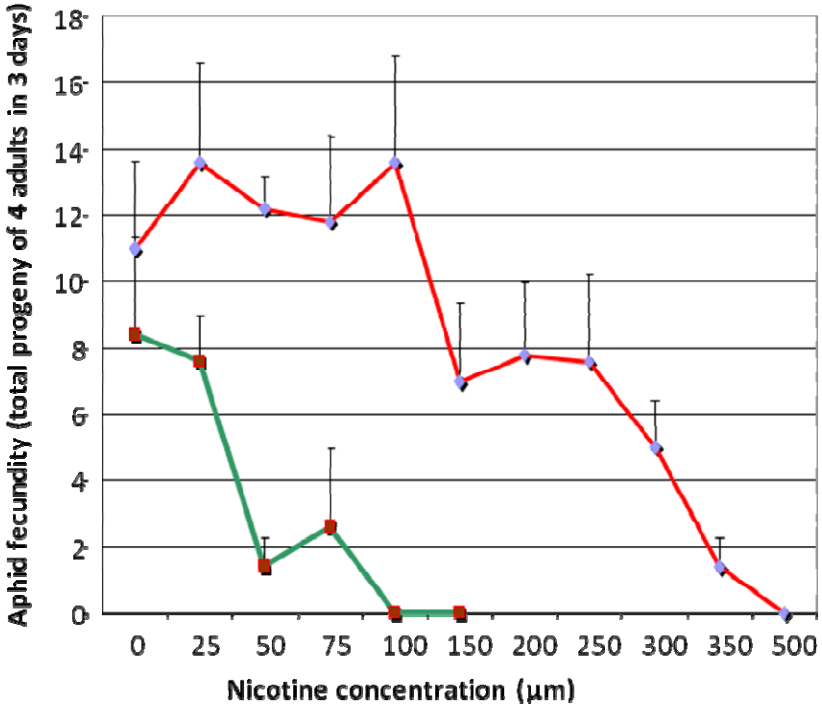


Figure 5.1. Fecundity of red and green aphids on nicotine-containing artificial diets. Red line represents red aphid fecundity at increasing nicotine concentrations, green line represents green aphid fecundity. Total progeny from four adult aphids was counted after three days. For each concentration of nicotine, five replicate diet cups were set up, each containing four adult aphids. Mean + standard error of n=5.

The half maximal inhibitory concentration (IC_{50}) of nicotine was measured as the concentration at which aphid fecundity (total progeny from 4 adult aphids over three days) on nicotine-containing diets is half that of fecundity on control diets. This was

determined by linear regression analysis to be 20 µM for green aphids, and 330 µM for red, tobacco-adapted aphids.

Moderate amounts of nicotine in artificial diet increase fecundity of tobacco-adapted aphids.

Initial artificial diet bioassays revealed a consistent increase in the fecundity of red aphids on diets containing nicotine at levels up to 100 µM. Subsequent bioassays comparing red aphid fecundity on control and 100 µM nicotine diets showed a significant increase in the number of progeny from aphids feeding on diets containing 100 µM nicotine (Fig. 5.2). This suggests that nicotine is improving aphid performance on artificial diet, either by acting as a feeding stimulant or through transformation into something useful by the aphids. Nicotine is produced in plants from amino acid precursors, and the aphid may catabolize nicotine into these nutritionally significant precursors as part of its detoxification response. This hypothesis is unlikely, however, as artificial diet contains 20 amino acids at millimolar levels.

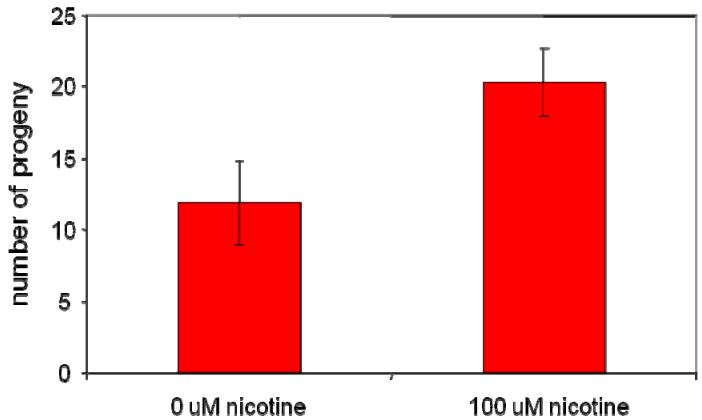


Figure 5.2. When 100 µM nicotine is added to artificial diet, the number of offspring produced by tobacco-adapted *M. persicae* is increased. Number of progeny represents offspring produced by four adult aphids in three days. Mean +/- standard error of N = 10, $P < 0.05$ by 2-tailed Student's t-test.

Strikingly, the concentration of nicotine which enhances red aphid fitness is lethal to green aphids.

Prior host effects on the performance of green and red aphids on tobacco

The green *M. persicae* strain used in this study is derived from a lineage that was not adapted to infest tobacco in the field. In the growth chamber, green aphids were able to survive and reproduce on *N. tabacum*, but their fecundity is significantly lower than red aphids (Fig. 5.3a).

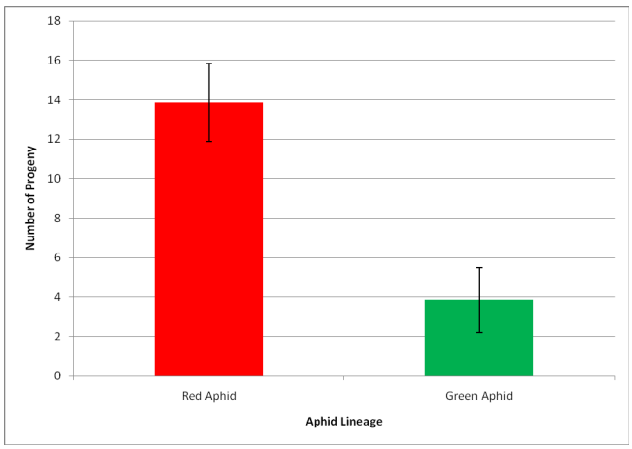


Figure 5.3a. Comparison of fecundity of green and red aphids on NC95 (high nicotine) tobacco. Single adult aphids were caged on four-week old tobacco plants, and progeny were counted after six days. Mean +/- standard error of N=7, $P < 0.01$ by 2-tailed Student's t-test.

Green aphids were reared on *N. tabacum* genotype NC95 (high nicotine line) in isolation, with no choice of alternate host plant, for several generations. Adult green aphids derived from populations established on NC95 for at least three generations were caged on four-week old NC95 plants, as were green aphids derived from a population that had never been exposed to tobacco, and had been reared for several generations on cabbage. Progeny from a single adult green aphid transferred to

tobacco, with either cabbage or tobacco as their prior host, were counted after six days. Green aphids reared on tobacco displayed significantly greater fecundity than green aphids with cabbage as their prior host (Fig 5.3b).

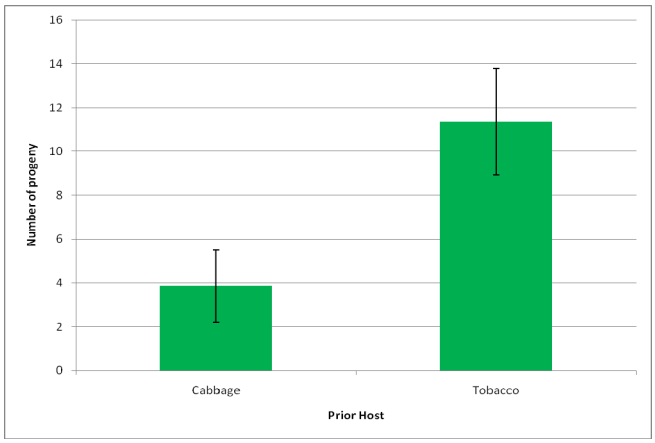


Figure 5.3b. Green aphid performance on NC95 tobacco depends on prior host. Progeny from single adult aphids transferred to tobacco from either cabbage or tobacco were counted after six days. Mean +/- standard error of N=7, $P<0.05$ by 2-tailed Student's t-test.

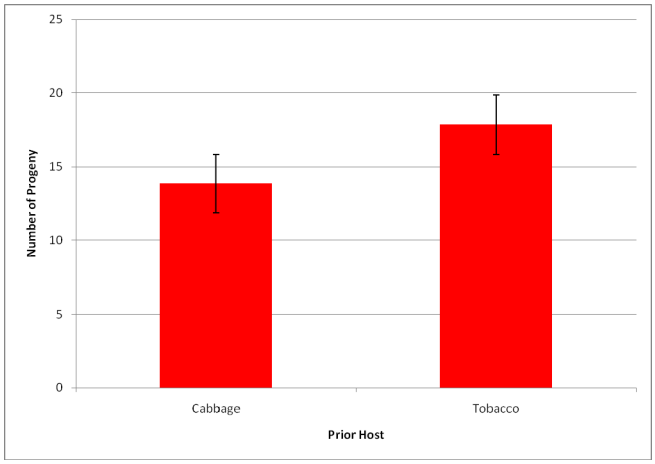


Figure 5.3c. Red aphid performance on NC95 tobacco was compared between insects previously reared on cabbage or tobacco. Progeny from single adults were counted after six days on tobacco. No significant difference was observed in number of progeny based on prior host. Mean +/- standard error of N=7, $P>0.05$ by 2-tailed Student's t-test.

When the same experiment was conducted with red aphids, however, there was

no prior host effect: the number of progeny from red aphids which had been reared continuously for several generations on cabbage was not significantly different from that of red aphids reared on tobacco (Fig. 5.3c).

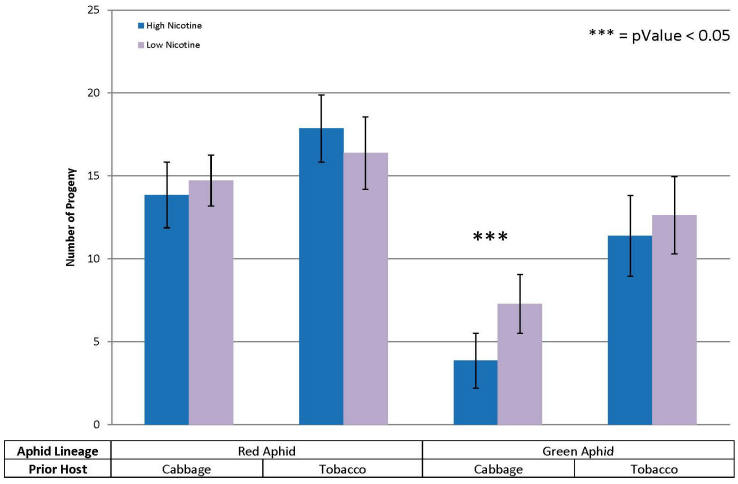


Figure 5.4. Fecundity of red and green aphids on high nicotine and low nicotine tobacco lines – prior host of aphids was either cabbage or tobacco. Fecundity is given as the number of progeny from a single adult aphid over six days. No significant difference in fecundity was observed between red aphids feeding on different tobacco genotypes. Green aphids with no prior experience on tobacco produce significantly more progeny when feeding on low nicotine lines. Green aphids reared for three generations on tobacco reproduce at similar levels on high and low nicotine lines. Mean +/- standard error of N=7. *** $P<0.05$ by 2-tailed Student's t-test.

Fecundity of red and green aphids on high (NC95) and low (LAFC53) nicotine-containing lines of *N. tabacum* was compared (Kinnersley and Dougall, 1980). Red aphids performed equally well on both tobacco genotypes, regardless of their prior host. Whereas green aphids reared on cabbage had significantly more progeny on LAFC53 than on NC95, this difference was erased when the aphids were reared on tobacco before being transferred to NC95 or LAFC53 (Fig. 5.4). Green aphids that have adapted to tobacco as a host may have acquired some measure of nicotine tolerance: contrary to their cabbage-reared counterparts, tobacco-reared green aphids

experience equal fecundity when feeding on high and low nicotine lines. However, green aphids reared continuously on tobacco are not able to tolerate significantly higher levels of nicotine in artificial diets than cabbage-reared green aphids (data not shown).

Nicotine-induced gene expression changes in heads of tobacco-adapted aphids

The microarrays used in this study consist of 60-mer probes designed to correspond to 10,295 unique genes derived from sequencing and assembly of *M. persicae* ESTs (Ramsey *et al*., 2007). These probes are identified on the microarray by the number of the contig which they represent, followed by "a", *e.g.* 3615a. An additional set of probes represents alternate 60-mers designed based on 4,322 of the total 10,295 unigenes. These probes are identified by the number of the contig which they represent, followed by "b", *e.g*. 3615b. In the cases where one probe meets the criteria for differential expression but a second probe designed against the same aphid gene does not, the P-value and fold change for the second probe were inspected manually. In all of these cases the second probe has a P-value below the 0.05 cutoff, but the fold change is less than two, and was therefore not selected as differentially expressed according to our criteria. There were no examples where one probe indicated the gene was up-regulated and the second probe indicated the reverse.

The unigenes represented on the array were originally annotated by comparison with the *D. melanogaster* RefSeq protein set. The genome of the pea aphid, *Acyrthosiphon pisum*, has recently been sequenced and annotated (International Aphid Genomics Consortium, 2009). This has enabled us to improve our annotation of *M. persicae* genes by comparing this unigene set to the pea aphid RefSeq protein set (Supplementary Table 5.1).

In order to learn more about the genes involved in nicotine resistance in

tobacco-adapted aphids, microarrays were used to identify genes induced by nicotine in red aphid heads. A total of 130 genes were identified as differentially expressed between tissue from control and 100 μM nicotine diets, all of which were up-regulated (Supplementary Table 5.2). A total of 746 genes were differentially expressed between tissue from control and 250 μM nicotine diets, of which 305 were down-regulated and 441 were up-regulated (Supplementary Table 5.3). Up-regulated genes include those involved in xenobiotic detoxification, and genes encoding proteins previously identified in aphid saliva.

Raw and normalized microarray data from these experiments have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al*., 2002) and are accessible through GEO Series accession number GSE18569 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18659).

Xenobiotic detoxification

Three types of genes implicated in the detoxification of xenobiotics by insects are cytochrome P450 monooxygenases (P450s), carboxylesterases, and glutathione-S-transferases (GSTs). P450s and GSTs comprise large gene families, members of which may act on either endogenous or xenobiotic substrates. Our microarray contains probes designed against 56 contigs annotated as P450s and 24 contigs annotated as GSTs, and analyses of changes in gene expression in response to xenobiotics represents a way to cut through the complexity of these gene families and identify genes relevant specifically to nicotine detoxification.

Three P450s were induced in aphid heads by to 100 μM nicotine, and a unique set of eight P450s was induced by 250 μM nicotine (Table 5.1). Of these 11 P450 genes, seven belong to the CYP6AX1 class, which has been implicated in the response of the Hemipteran *Nilaparvata lugens* to allelochemicals in resistant rice cultivars

(Yang *et al*., 2007). The up-regulation of aphid CYP6AX1 in response to nicotine is consistent with the role of members of the CYP6 family in insect response to plant toxins and resistance to insecticides (reviewed in Scott *et al*., 1998).

Table 5.1. *M. persicae* detoxification genes induced by nicotine in artificial diet experiments.

ProbeID	Nicotine Concentration	Gene Description: Gene ID of *A. pisum* ortholog	Fold Change
1501a	100 µM	cytochrome P450 CYP6AX1 protein: ACYPI000639	2.11
1504b	100 µM	cytochrome P450: ACYPI002079	2.07
8290a	100 µM	cytochrome P450: ACYPI010012	2.37
9053a	250 µM	cytochrome p450: ACYPI001913	2.31
970a	250 µM	cytochrome P450 CYP6AX1: ACYPI008473	2.11
970b	250 µM	cytochrome P450 CYP6AX1: ACYPI008473	2.54
497a	250 µM	cytochrome P450 CYP6AX1: ACYPI008843	2.58
497b	250 µM	cytochrome P450 CYP6AX1: ACYPI008843	3.29
5173a	250 µM	cytochrome P450 CYP6AX1: ACYPI008843	2.92
5223a	250 µM	cytochrome P450 CYP6AX1: ACYPI003528	2.12
8290a	250 µM	cytochrome P450: ACYPI010012	3.6
9095a	250 µM	cytochrome P450 CYP6AX1: ACYPI003528	2.23
3931b	250 µM	cytochrome p450: ACYPI005477	2.47
720a	250 µM	carboxylesterase; esterase FE4: ACYPI000631	3.03
720b	250 µM	carboxylesterase; esterase FE4: ACYPI000631	3.16
3118a	250 µM	carboxylesterase; esterase FE4: ACYPI000631	2.26
3118b	250 µM	carboxylesterase; esterase FE4: ACYPI000631	2.29
9215a	250 µM	carboxylesterase; esterase FE4: ACYPI000631	3.73
1196b	100 µM	glutathione S-transferase-like protein: ACYPI009519	2.56
4744a	100 µM	glutathione S-transferase-like protein: ACYPI009519	3.78
1196a	100 µM	glutathione S-transferase-like protein: ACYPI009519	3.68
7666a	250 µM	glutathione S-transferase-like protein: ACYPI009519	2.18
3031a	250 µM	glutathione S-transferase-like protein: ACYPI002679	2.32
3031b	250 µM	glutathione S-transferase-like protein: ACYPI002679	2.23
3676b	250 µM	glutathione S-transferase-like protein: ACYPI000794	2.11
3677a	250 µM	glutathione S-transferase-like protein: ACYPI000794	3.48
3677b	250 µM	glutathione S-transferase-like protein: ACYPI000794	2.82
3678a	250 µM	glutathione S-transferase-like protein: ACYPI000794	2
3678b	250 µM	glutathione S-transferase-like protein: ACYPI000794	2.15
4715a	250 µM	glutathione S-transferase-like protein: ACYPI000794	2.18
8694a	250 µM	glutathione S-transferase-like protein: ACYPI000794	2.31

Three *M. persicae* contigs annotated as FE4 carboxylesterases, represented by five unique microarray probes, were induced by 250 μM nicotine, yet displayed no significant change in response to 100 μM nicotine (Table 5.1). These three contigs all had the same *A. pisum* gene as their best blast hit. This suggests that either our contig assembly criteria are too stringent, and these sequences ought to be considered to represent the same gene, or that there has been expansion of this gene family in *M. persicae* relative to *A. pisum*. The latter hypothesis is unlikely, as a comparative analysis of detoxification genes found that these species contain approximately the same number of carboxylesterase sequences (Ramsey *et al*., 2009).

Expression of several aphid glutathione-S-transferases also increased in response to both concentrations of nicotine, with three *M. persicae* genes up-regulated at 100 μM and seven genes, represented by 10 probes (orthologous to three *A. pisum* genes) up-regulated by 250 μM nicotine.

Curiously, the P450 and GST genes induced by the lower dose of nicotine are not among the genes induced by the higher dose. It may be expected that the response initiated by 100 μM nicotine would be simply enhanced upon exposure to 250 μM nicotine, with genes induced by the lower dose displaying a higher level of up-regulation by the higher dose, but this appears not to be the case. This suggests that the detoxification response to different levels of nicotine may be fundamentally different.

Aphid salivation-associated genes

Proteomic studies of aphid saliva in *M. persicae* and *A. pisum* have identified a number of salivary proteins (Harmel *et al*., 2008, Carolan *et al*., 2009). Genes for three proteins identified in the saliva of *M. persicae* - α-amylase, α-glucosidase, and glucose dehydrogenase - are significantly up-regulated in response to both 100 and 250 μM nicotine (Table 5.2).

Table 5.2. Aphid genes associated with salivation induced by nicotine in artificial diet.

ProbeID	Nicotine Concentration	Gene Description: Gene ID of *A. pisum* ortholog	Fold Change
3682a	100 µM	alpha-amylase: ACYPI007122	11.23
3682b	100 µM	alpha-amylase: ACYPI007122	5.5
3682a	250 µM	alpha-amylase: ACYPI007122	11.31
3682b	250 µM	alpha-amylase: ACYPI007122	6.4
3987a	100 µM	alpha-amylase: ACYPI007122	9.71
3987b	100 µM	alpha-amylase: ACYPI007122	8.57
3987a	250 µM	alpha-amylase: ACYPI007122	12.12
3987b	250 µM	alpha-amylase: ACYPI007122	8.81
2868a	100 µM	alpha-amylase: ACYPI009042	3.45
2868b	100 µM	alpha-amylase: ACYPI009042	3.55
2868a	250 µM	alpha-amylase: ACYPI009042	4.53
2868b	250 µM	alpha-amylase: ACYPI009042	4.56
3812a	100 µM	alpha-amylase: ACYPI009042	3.38
3812a	250 µM	alpha-amylase: ACYPI009042	4.62
3812b	250 µM	alpha-amylase: ACYPI009042	5.85
8554a	100 µM	alpha-glucosidase: ACYPI001718	3.73
3806b	100 µM	alpha-glucosidase: ACYPI001718	3.13
3435b	100 µM	alpha-glucosidase: ACYPI005549	4.05
3435a	250 µM	alpha-glucosidase: ACYPI005549	8.11
3435b	250 µM	alpha-glucosidase: ACYPI005549	5.09
6126a	250 µM	alpha-glucosidase: ACYPI008059	2.42
4022a	250 µM	S1 sucrase: ACYPI000002	3.45
4022b	250 µM	S1 sucrase: ACYPI000002	5.65
7717a	250 µM	S1 sucrase: ACYPI000002	3.13
933a	100 µM	sucrase: ACYPI001436	14.72
933b	100 µM	sucrase: ACYPI001436	6.19
933a	250 µM	sucrase: ACYPI001436	14.6
933b	250 µM	sucrase: ACYPI001436	5.81
10020a	250 µM	sucrase: ACYPI001436	3.63
10020a	100 µM	sucrase: ACYPI001436	2.11
688a	250 µM	glucose dehydrogenase: ACYPI000113	2.6
688b	250 µM	glucose dehydrogenase: ACYPI000113	2.82
1461a	250 µM	trehalase: ACYPI002298	2.34
1461b	250 µM	trehalase: ACYPI002298	2.29

Several sucrase genes and one trehalase gene are also induced by nicotine in aphid heads – although the corresponding proteins were not identified in proteomics studies (Harmel *et al*., 2008, Carolan *et al*., 2009), these enzyme activities have been reported

to be present in aphid saliva (Miles, 1999). Sucrase and α-amylase genes are among those most induced by nicotine out of all differentially expressed genes, with fold-changes of 12 to 14. Unlike the case with detoxification-associated genes induced by nicotine, aphid salivation-associated genes are typically induced both by low and high doses of nicotine. The up-regulation of these genes is generally slightly greater in response to the higher dose.

An additional nicotine-induced gene potentially associated with aphid salivation is a 1,4-beta-xylanase (contig 3525, orthologous to *A. pisum* gene ACYPI005168), which is predicted to be involved in the degradation of hemicellulose, a major component of plant cell walls. This enzyme activity is typically associated with microbes from insect guts, where it plays an essential role in hydrolyzing cell wall polysaccharides resistant to degradation (Brennan *et al*., 2004). Aphid xylanase may play a role in clearing a path for the stylet through cell wall hemicellulose as it probes intercellularly to initiate feeding in the plant phloem.

Discussion

Nicotine is synthesized in roots and transported via the xylem to foliar tissues in tobacco. Although nicotine levels in phloem sap have not been measured, to our knowledge, for any tobacco species, tobacco-feeding aphids are likely to come into contact with nicotine through periodic ingestion of xylem sap and through exposure to alkaloid secretions by glandular trichomes (Thurston *et al*., 1966). In this study we have demonstrated that a tobacco-adapted lineage of *Myzus persicae* is able to tolerate significantly higher levels of nicotine in artificial diets than a lineage with no prior experience on tobacco, suggesting that the development of nicotine resistance has accompanied the expansion of the aphid's host range to include tobacco. This high level of nicotine resistance is not, however, essential for the adaptation of aphids to

tobacco. The green, non-adapted lineage used in this study experienced similar fecundity to the red aphid on plants of the high nicotine NC95 *N. tabacum* genotype after being reared on tobacco in the lab for several generations, but this exposure to tobacco did not significantly enhance the green aphid's tolerance to nicotine in artificial diets. This suggests that there are other factors besides nicotine which are acting as barriers to herbivory for aphids not adapted to tobacco. *M. persicae* is able to overcome these barriers in a matter of generations when caged on tobacco in a no-choice context in the lab. The concentration of nicotine which tobacco-feeding aphids are exposed to is not known. Our observation that a green aphid clone can adapt to tobacco as a host plant in the lab, without gaining tolerance to nicotine in artificial diet, suggests that tobacco-feeding aphids encounter nicotine at concentrations less than the 10 μM level which the green aphid in this study can tolerate with no deleterious consequences.

Gene expression changes in heads of tobacco-adapted aphids in response to nicotine include representatives of the three main classes of insect detoxification genes, cytochrome P450s, carboxylesterases, and glutathione S-transferases. While it may be possible to undermine the aphid detoxification response by silencing the genes implicated in an effective response to plant defenses, it will likely to be more effective to focus on identifying the signaling and regulatory factors which induce the expression of these genes upon perception of the plant-derived signal. In addition to the genes highlighted in Table 5.1 and 5.2, there are a number of other genes, many with no known function, whose expression is regulated in the aphid by nicotine exposure (see supplementary table 5.2 and 5.3). It will be valuable to perform cluster analyses to find suites of genes coregulated by different plant toxins, and to incorporate time course studies into microarray experimental designs to parse the different stages of the aphid's perception of and response to plant defensive chemicals.

The up-regulation of large numbers of aphid salivation-associated genes in response to nicotine may be related to the role of aphid saliva in suppressing plant defenses. The aphid may respond to the perception of plant defenses by increasing salivation, in an effort to interfere with the continued establishment of an effective plant response to herbivore attack. Electrical penetration graph studies of changes in aphid feeding behavior in response to nicotine may shed light on this question.

We have expanded our *M. persicae* unigene database by 454 transcriptome sequencing (Ramsey *et al*., 2009). Based on this new sequence data (available at www.aphidbase.com/aphidbase/downloads, file name Myzus454) we have designed an additional probe set of 30,517 60-mers. Using these in conjunction with the sequences represented on the array employed in this study we have designed a 4x44K Agilent gene expression microarray which can be obtained by contacting GJ.

Materials and Methods

Plant growth and insect rearing

Plant growth conditions

Seeds from the *Nicotiana tabacum* lines NC95 and LAFC53 were kindly provided by Dr. Ramsey Lewis (North Carolina State University). Seeds were sown in MetroMix 360 (SunGro). After two weeks, seedlings were transplanted in Cornell mix (Landry *et al*., 1995) with Osmocote fertilizer (Scotts). Cabbage (*Brassica oleracea* var. Wisconsin Golden Acres; Seedway) was sown directly in Cornell mix with osmocote. Plants were reared in Conviron growth chambers (16:8 h light:dark cycle, 2000 μmols $m^{-2}s^{-1}$, at 23°C, 50% relative humidity) in 20 x 40 cm nursery flats.

Insect rearing

A tobacco-adapted red lineage of *M. persicae* was obtained from S. Gray (USDA Plant Soil and Nutrition Laboratory, Ithaca, NY, USA). A green lineage of *M. persicae* not adapted to tobacco was obtained from the Boyce Thompson Institute greenhouse, where it was feeding from cabbage. Aphids were raised on tobacco or cabbage (16:8 h light:dark cycle, 150 μmols $m^{-2}s^{-1}$, at 24°C ± 1 day, 19°C ± 1 night, 50% relative humidity).

Aphid artificial diet and tobacco fecundity assays

Nicotine toxicity assays

Nicotine (S-(-)-isomer, pur. >99%, Fluka) was kindly provided by R. Halitschke (Cornell University). Liquid diet for aphid culture was prepared from a mixture of water, amino acids, and sucrose (Kim and Jander, 2007). Diet was sterilized by passing through a 0.20 micron syringe filter (Corning) and stored as 10 mL aliquots at -80°C. A 100 mM nicotine stock solution was prepared in methanol and stored at -80°C. Stock solution was used to prepare aliquots of artificial diet containing varying concentrations of nicotine for use in toxicity assays. Four adult aphids of the red or green lineage were transferred from *B. oleracea* into vented hard plastic cups (30 ml volume) for artificial diet experiments (Ramsey and Jander, 2008). Liquid diet was suspended between two layers of Parafilm stretched over the rim of the cup. Surviving adults and offspring were counted after 3 days. Linear regression analysis was performed using R (R Development Core Team, 2005).

Aphid fecundity on tobacco

Single adult aphids of the green or red lineage were transferred from either *N. tabacum* or *B. oleracea* onto the center of a four-week old NC95 or LAFC53 *N. tabacum* plant.

Individual plants were enclosed in cages and aphid progeny were counted after 6 days.

Aphid microarrays

M. persicae gene expression microarray slides with design ID 017179 (Ramsey *et al*., 2007) were obtained from Agilent (http://earray.chem.agilent.com). One slide consists of eight microarrays, each containing 15,744 features. Four arrays compared gene expression between heads of aphids feeding on control and 100 μM nicotine containing diets, and four arrays compared gene expression between heads of aphids feeding on control and 250 μM nicotine containing diets. For each comparison, three arrays represented biological replicates, and the fourth array was a technical replicate.

Heads were dissected from red aphids 24 hours after they were transferred from *B. oleracea* onto artificial diets containing 0, 100, or 250 μM nicotine. Aphids observed to be actively feeding from liquid diet were collected individually with a paint brush, immersed in 70% ethanol for 1-2 seconds, and placed in distilled water on a microscope slide. Aphid heads were sliced off with razor blades (Electron Microscopy Sciences) and immediately transferred by forceps into a microcentrifuge tube on dry ice. Each sample for subsequent experiments consisted of 15 aphid heads.

Aphid RNA was extracted using the Ribopure kit (Ambion). Tissue was homogenized using a 21 guage needle attached to a 1 mL syringe, following the addition of 0.5 mL Tri Reagant (Ambion) to the sample. RNA quantitation was performed using the Nanodrop (Thermo Scientific). Agilent bioanalyer analysis was conducted by the Cornell Microarray Core Facility to confirm RNA quality.

The Amino Allyl MessageAmp II aRNA Amplification kit (Ambion) was used to prepare RNA samples for array hybridization. RNA spike-ins (Two-Color RNA Spike-In kit, Agilent) were added to each sample. Reverse transcription was performed

using a T7 oligo dT primer. Following second strand synthesis and purification of cDNA, *in vitro* transcription was carried out with amino-allyl modified UTP added to the reaction mix. The resulting amplified RNA (aRNA) including amino-allyl modified nucleotides was stored at -80°C until immediately prior to hybridization. At that time, a dye coupling reaction with the dyes Alexa fluor 555 and 647, which conjugate with the modified UTP moieties on the RNA molecules, was performed. Purified aRNA was fragmented at 60°C following Agilent's Two-Color Microarray-Based Gene Expression Analysis protocol. Fragmentation was terminated by addition of 2x GEx Hybridization Buffer HI-RPM, and samples were placed on ice and immediately loaded onto the array.

Hybridization, wash, and scan of microarrays were performed according to the Agilent Two-Color Microarray-Based Gene Expression Analysis. Agilent scanner settings for 8x15K slide formats were used, and data was extracted using Agilent Feature Extraction Software.

Microarray data analysis was carried out using the limma package within R (Smyth, 2004; Smyth and Speed, 2003; Smyth, 2005). Within array normalization of was performed by the loess method without background subtraction. Linear modeling of arrays was performed with the lmFit function, and statistical analysis was performed using the eBayes function. Genes with a P-value ≤ 0.05 and a fold change ≥ 2 or ≤ 0.5 were considered to be differentially expressed.

Authors' Contributions

GJ and JSR designed experiments. JSR performed aphid artificial diet experiments. JSR and YX performed assays of aphid fecundity on tobacco. JSR and PS performed microarray experiments. JSR analyzed microarray data. JSR wrote the paper.

Acknowledgements

This work was funded by USDA grant 2005-35604-15446 to GJ. Additional funding for YX came from Carolyn Sampson in support of educational outreach at the Boyce Thompson Institute. The authors thank the Cornell Microarray Core Facility for technical assistance.

Description of Additional Data Files (Excel spreadsheets provided on CD)

Supplementary Table 5.1: Comparison of *Myzus persicae* cDNA sequences to *Acyrthosiphon pisum* merged Glean/RefSeq proteins.

Supplementary Table 5.2: Microarray probes significantly differentially expressed between heads of red aphids feeding on control and 100 μM nicotine diets.

Supplementary Table 5.3: Microarray probes significantly differentially expressed between heads of red aphids feeding on control and 250 μM nicotine diets.

REFERENCES

Blackman RL, Eastop VF (2000). *Aphids on the world's crops: an identification and information guide*. Chichester, John Wiley & Sons, 466 p.

Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, Hernandez M, Keller M, Li K, Palackal N, Sittenfeld A, Tamayo G, Wells S, Hazlewood GP, Mathur EJ, Short JM, Robertson DE, Steer BA (2004). Unusual microbial xylanases from insect guts. *Appl Environ Microbiol* **70(6)**:3609-3617.

Cardoza YJ, Wang SF, Reidy-Crofts J, Edwards O (2006). Phloem alkaloid tolerance allows feeding on resistant *Lupinus angustifolius* by the aphid *Myzus persicae*. *J Chem Ecol* **32**:1965-1976.

Carolan JC, Fitzroy CI, Ashton PD, Douglas AE, Wilkinson TL (2009). The secreted salivary proteome of the pea aphid Acyrthosiphon pisum characterised by mass spectrometry. *Proteomics* **9(9)**:2457-2467.

Devine GJ, Harling ZK, Scarr AW, Devonshire AL (1996). Lethal and sublethal effects of imidacloprid on nicotine-tolerant *Myzus nicotianae* and *Myzus persicae*. *Pestic Sci* **48**:57-62.

Edgar R, Domrachev M, Lash AE (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30(1):**207-210.

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

International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*, in review.

Kim, JH, Jander G (2007). *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. *Plant J* **49(6)**: 1008-1019.

Kinnersley AM, Dougall DK (1980). Correlation between the nicotine content of tobacco plants and callus cultures. *Planta* **149**:205-206.

Landry LG, Chapple CC, Last RL (1995). Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol* **109:** 1159–1166.

Margaritopoulos JT, Kasprowicz L, Mallock GL, Fenton B (2009). Tracking the global dispersal of a cosmopolitan insect pest, the peach potato aphid. *BMC Ecol* **9**:13.

Miles PW (1999). Aphid saliva. *Biol Rev* **74**:41-85.

Murray CL, Quaglia M, Arnason JT, Morris CE (1994). A putative nicotine pump at the blood-brain-barrier of the tobacco hornworm. *J Neurobiol* **25(1)**:23-24.

Passador-Gurgel G, Hsieh W, Hunt P, Deighton N, Gibson G (2007). Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. *Nat Genet* **39**:264-268.

Ramsey JS, Jander G (2008). Testing Nicotine Tolerance in Aphids Using an Artificial Diet Experiment. *JoVE*. **15**. http://www.jove.com/index/details.stp?id=701 doi: 10.3791/701

Ramsey, JS, Wilson, AC, de Vos, M, Sun, Q, Tamborindeguy, C, Winfield, A, Malloch, G, Smith, DM, Fenton, B, Gray, SM and Jander, G (2007) Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* **8**: 423.

Ramsey JS, Rider DS, Walsh TK, de Vos M, Gordon K, Ponnala L, Macmil S, Roe BA, Jander G. (2009) Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. (*Insect Mol Biol*, in press).

R Development Core Team (2005). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

Scott JG, Liu N, Wen ZM (1998). Insect cytochromes P450: diversity, insecticide resistance, and tolerance to plant toxins. *Comp Biochem Physiol C* **121**:147-155.

Shigehara T, Takada H (2003). Changes in genotypic composition of *Myzus persicae* (Hemiptera: Aphididae) on tobacco during the past two decades in Japan. *Bull Entomol Res* **93(6)**: 537-544.

Smyth GK (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3, No. 1, Article 3.

Smyth GK, Speed TP (2003). Normalization of cDNA microarray data. *Methods* **31**: 265-273.

Smyth GK (2005). Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pp. 397-420.

Snyder MJ, Hsu E-L, Feyereisen R (1993). Induction of cytochrome P-450 activities by nicotine in the tobacco hornworm, *Manduca sexta*. *J Chem Ecol* **19(12)**: 2903-2916.

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004). Nicotine's defensive function in nature. *PloS Biol* **2**:1074-1080.

Thurston R, Smith WT, Cooper BP (1966). Alkaloid secretion by trichomes of *Nicotiana* species and resistance to aphids. *Entomol Exp Appl* **9**:428-432.

Yang Z, Zhang F, He Q, He G (2005). Molecular dynamics of detoxification and toxin-tolerance genes in brown planthopper (*Nilaparvata lugens* Stal., Homoptera: Delphacidae) feeding on resistant rice plants. *Arch Insect Biochem Physiol* **59**:59-66.

Chapter 6: Microarray Collaborations

The *Myzus persicae* microarray which I have designed has been made available to the international aphid community. This has laid the foundation for the establishment and extension of collaborations between the Jander lab and researchers in diverse fields of biology interested in discovering aphid genes associated with response to insecticides, drought stress, and feeding on *Arabidopsis thaliana* trehalose metabolism mutants. Manuscripts describing the results of these studies are currently being prepared with our collaborators: in this chapter I will describe the background and rationale of these experiments, and outline the experimental design which we implemented.

Aphid Response to Insecticides

Pirimicarb (2-dimethylamino-5,6-dimethylpyrimidin-4-yl dimethylcarbamate) is an aphidicide used on a broad range of crops. Pirimicarb exerts its activity through inhibition of acetylcholinesterase, an enzyme which terminates synaptic transmission at neuromuscular junctions by the degradation of acetylcholine. Chemical control of aphids has been frustrated by the multiple resistance mechanisms which have spread through aphid populations. The amplification of genes encoding insecticide-detoxifying esterases confers broad spectrum resistance to organophosphates, carbamates, and pyrethroids in *M. persicae* (Devonshire *et al*., 1998). Mutations in the insecticide target sites acetylcholinesterase (MACE) and voltage-gated sodium channels (*kdr* and *s-kdr*) confer resistance to carbamates and pyrethroids, respectively (Foster *et al*., 2003, Anstead *et al*., 2005). A collaboration with the laboratory of Dr. Christian Figueroa at the Universidad Austral de Chile was initiated to investigate the

transcriptional response of three *M. persicae* lineages to spraying by pirimicarb.

Aphid lineages for this study were selected based on biochemical assays of their esterase activity (following Devonshire *et al*., 1992) and by their genotype at the MACE, *kdr,* and *s-kdr* loci (Table 6.1).

Table 6.1: Esterase levels, and genotype at MACE, *kdr*, and *s-kdr* loci, of *M. persicae* lineages used in this study. Genotype nomenclature: ss = homozygous/susceptible; **rs** = heterozygous/resistant. Esterase activity is stated as units absorbance (620 nm) per aphid equivalent, measuring cleavage of α-naphthyl acetate by aphid homogenate *in vitro* by conjugation of the resulting aldehyde with Fast Blue B colorant solution.

Lineage	MACE	kdr	s-kdr	Esterases
S (susceptible; no mutations)	ss	ss	ss	0.150
RS (resistant by a single mutated gene)	ss	**rs**	ss	0.207
RM (resistant by multiple mutated genes)	**rs**	**rs**	ss	0.194

Aphids were sprayed with 20 ppm pirimicarb using a Potter Tower (Burkard, UK); after treatment aphids were transferred to sweet pepper plants (Olivares-Donoso *et al*., 2007) to feed for 30 hours. Aphid RNA was extracted in Chile by Andrea Silva, and transported in a stabilization solution of ethanol and sodium acetate. Upon arrival at the Boyce Thompson Institute, aphid RNA was precipitated and subjected to Agilent Bioanalyzer analysis. Samples were prepared for microarray analysis as described in Chapter 6. The experimental design used in this study is described in Table 6.2.

Table 6.2: Experimental design for *M. persicae* insecticide treatment microarray. Designation of A, B, and C represents biological replicates for each sample type

Microarray ID	Sample Description (Alexa Fluor 555)	Sample Description (Alexa Fluor 647)
1	RM lineage A, water spray	RM lineage A, 20 ppm pirimicarb spray
2	RM lineage B, 20 ppm pirimicarb spray	RM lineage B, water spray
3	RS lineage A, water spray	RS lineage A, 20 ppm pirimicarb spray
4	RS lineage B, 20 ppm pirimicarb spray	RS lineage B, water spray
5	S lineage A, water spray	S lineage A, 20 ppm pirimicarb spray
6	S lineage B, 20 ppm pirimicarb spray	S lineage B, water spray
7	S lineage C, water spray	S lineage C, 20 ppm pirimicarb spray
8	RM lineage A, water spray	RM lineage A, 20 ppm pirimicarb spray

One 8x15K Agilent microarray slide was hybridized, washed, and scanned at the

Cornell Microarray Core Facility. Data analysis was conducted using R as described in Chapter 5. The experimental design of this study was based on a block design, using dye swaps, and genes differentially expressed between pirimicarb treated and control samples were identified for each of the three *M. persicae* genotypes used in this study. This work was funded in part by PBCT-REDES R-01 and FONDECYT grants to Dr. Christian Figueroa, and CONICYT and DID-UACH PhD grants to Andrea Silva.

Aphid Response to Drought Stress

The high sugar content of aphid phloem sap represents a challenge to aphid osmoregulation. Dietary osmotic pressure is significantly higher than the osmotic pressure of aphid hemolymph, and assimilation of sugars and formation of oligosaccharides represent adaptations of the aphid to reduce the osmotic pressure of ingested fluids, thereby reducing water loss (Wilkinson *et al*., 1997). These osmoregulatory challenges may be increased in aphids feeding on plants exposed to drought conditions, which may lead to an increase in phloem osmotic pressure (Cernusak *et al*., 2003). Studies on the performance of the cereal aphids *Rhopalosiphum padi* and *Sitobion avenae* on drought-stressed plants indicate that aphid fecundity is significantly reduced in response to drought (Pons and Tatchell, 1995). Feeding on drought-stressed barley seedlings resulted in an increase in sieve element salivation in *R. padi* (Ponder *et al*., 2001). This raised the possibility that modulation of phloem osmotic pressure may be effected by aphid salivary compounds, and implicates increased salivation as a component of the aphid response to abiotic stress.

The feeding response of *M. persicae* to droughted *Brassica nigra* was assessed by electrical penetration graph (EPG) analysis. Aphid responses to drought stress included a significant decrease in phloem sap ingestion, which was accompanied by a

reduction in honeydew production (Laura Vickers, unpublished data). Phloem sap of drought-stressed plants contains greater concentrations of amino acids than control plants, raising the possibility that aphids need to ingest less in order to acquire sufficient quantities of the nutritionally-limited components of their diet (Girousse, 1996; Tully and Hanson, 1979). In addition to a reduced rate of ingestion, EPG measurements revealed increased salivation in *M. persicae* feeding on drought-stressed *B. nigra* (Laura Vickers, unpublished data).

Laura Vickers visited the Jander laboratory in May 2009 to extend her physiological studies with microarray analyses of aphid gene expression changes in response to drought stress. She shipped aphids on dry ice from England, and upon their arrival at the Boyce Thompson Institute she extracted total RNA, and carried out microarray experiments under my guidance as described in Chapter 6. She set up her experiments in a block design, using a dye swap, as described in Table 6.2. Microarray data analysis was performed using R, as described in Chapter 5.

Table 6.3: Experimental design for *M. persicae* drought response microarray. Designation of A, B, C, and D represents biological replicates for each sample type. Mild drought refers to aphids which were collected from *Brassica nigra* plants subjected to 3-6 days of total water withdrawal. Extreme drought refers to aphids which were collected from *Brassica nigra* plants subjected to 6-9 days of total water withdrawal. Plants were aged at 20 days under lab conditions before onset of droughting regime.

Microarray ID	Sample Description (Alexa Fluor 555)	Sample Description (Alexa Fluor 647)
1	control A	mild drought A
2	mild drought B	control B
3	control C	mild drought C
4	mild drought D	control D
5	control A	extreme drought A
6	extreme drought B	control B
7	control C	extreme drought C
8	extreme drought D	control D

Aphid response to feeding on *Arabidopsis thaliana* trehalose metabolism mutants

Trehalose is the major hemolymph component of *Myzus persicae* (Moriwaki *et al*., 2003), and trehalose and its metabolic precursor trehalose-6-phosphate also play a significant role as plant signaling molecules (Paul *et al*, 2008). Twenty-two genes encoding putative trehalose metabolizing enzymes have been found in the *Arabidopsis* genome, including seven trehalose-6-phosphate synthases, which produce trehalose-6-phosphate from glucose-6-phosphate and UDP-glucose (Ramon and Rolland, 2007). One of these, TPS11, is induced by *M. persicae* feeding, and aphid populations are larger on *tps11* mutant plants than on col-0 wild-type *Arabidopsis*, suggesting that TPS11 is required for plant defense against the aphid. (Vijay Singh, Joe Louis, and Jyoti Shah, unpublished data).

In an effort to further understand how aphids respond to *Arabidopsis* defenses, aphids reared on either wild-type or mutant plants for time periods varying from six to 48 hours were collected. Aphids were shipped from our collaborator, Dr. Jyoti Shah (University of North Texas), on dry ice. I proceeded to prepare samples for microarray analysis as described in Chapter 5 (See Table 6.3 for experimental design).

Table 6.4: Experimental design for *M. persicae tps11* microarray. Aphids were reared either on col-0 wild-type *Arabidopsis* or *tps11* mutant *Arabidopsis*.

Microarray ID	Sample Description (Alexa Fluor 555)	Sample Description (Alexa Fluor 647)
1	6 hours feeding, *tps11*, replicate A	6 hours feeding, col-0, replicate A
2	6 hours feeding, col-0, replicate B	6 hours feeding, *tps11*, replicate B
3	12 hours feeding, *tps11*, replicate A	6 hours feeding, col-0, replicate A
4	12 hours feeding, col-0, replicate B	12 hours feeding, *tps11*, replicate B
5	24 hours feeding, *tps11*, replicate A	24 hours feeding, col-0, replicate A
6	24 hours feeding, col-0, replicate B	24 hours feeding, *tps11*, replicate B
7	48 hours feeding, *tps11*, replicate A	48 hours feeding, col-0, replicate A
8	48 hours feeding, col-0, replicate A	48 hours feeding, *tps11*, replicate B

As *Arabidopsis tps11* mutants are hypothesized to have a reduced capacity to mount

an effective defense against *M. persicae*, aphid genes involved in responding to anti-herbivore defenses in the plant may be expressed at low levels in aphids feeding on mutant plants compared to wild type. Identification of aphid genes characterized by altered expression between plant mutants with lesions in defensive signaling responses may pave the way for a better understanding of the mechanisms by which *M. persicae* recognizes and responds to plant defenses.

REFERENCES

Anstead JA, Williamson MS, Denholm I (2005). Evidence for multiple origins of identical insecticide resistance mutations in the aphid *Myzus persicae*. *Insect Biochem Mol Biol* **35(3)**:249-256.

Cernusak LA, Arthur DJ, Pate JS, Farquhar GD (2003). Water relations link carbon and oxygen isotope discrimination to phloem sap sugar concentrations in *Eucalyptus globules*. *Plant Physiol* **131**:1544-1554.

Devonshire AL, Devine GJ, Moores GD (1992). Comparison of microplate esterase assays and immunoassay for identifying insecticide resistant variants of *Myzus persicae. Bull Entomol Res* **82(4)**:459-463.

Devonshire AL, Field LM, Foster SP, Moores GD, Williamson MS, Blackman RL (1998). The evolution of insecticide resistance in the peach-potato aphid, *Myzus persicae*. *Phil Trans R Soc Lond B* **353**:1677-1684.

Foster SP, Kift NB, Baverstock J, Sime S, Reynolds K, Jones JE, Thompson R, Tatchell GM (2003). Association of MACE-based insecticide resistance in *Myzus persicae* with reproductive rate, response to alarm pheromone and vulnerability to attack by *Aphidius colemani. Pest Manag Sci* **59**:1169-1178.

Girousse C, Bournoville R, Bonnemain JL (1996). Water deficit-induced changes in concentrations in proline and some other amino acids in the phloem sap of alfalfa. *Plant Physiol 111*:109-113.

Moriwaki N, Matsushita K, Nishina M, Kono Y (2003). High concentrations of trehalose in aphid hemolymph. *Appl Entomol Zool* **38(2)**:241-248.

Olivares-Donoso R, Troncoso AJ, Tapia DH, Aguilera-Olivares D and Niemeyer HM (2007). Contrasting performances of generalist and specialist Myzus persicae (Hemiptera:Aphididae) reveal differential prevalence of maternal effects after host transfer. *Bull Entomol Res* **97**:61-67.

Paul MJ, Primavesi LF, Jhurreea D, Zhang Y (2008). Trehalose Metabolism and Signaling. *Annu Rev Plant Biol* **59**:417-441.

Ponder KL, Pritchard J, Harrington R, Bale JS (2001). Feeding behaviour of the aphid *Rhopalosiphum padi* (Hemiptera: Aphididae) on nitrogen and water-stressed barley (Hordeum vulgare) seedlings. *Bull Entomol Res* **91**:125-130.

Pons X, Tatchell GM (1995). Drought stress and cereal aphid performance. *Ann Appl Biol* **126**:19-31.

Ramon M, Rolland F (2007). Plant development: introducing trehalose metabolism. *Trends Plant Sci* **12(5)**:185-188.

Tully RE and Hanson AD (1979). Amino acids translocated from turgid and water-stressed barley leaves I. Phloem exudation studies. *Plant Physiol* **64**:460-466.

Wilkinson T, Ashford D, Pritchard J, Douglas A (1997). Honeydew sugars and osmoregulation in the pea aphid *Acyrthosiphon pisum*. *J Exp Biol* **200(15)**: 2137-2143.

Chapter 7: Future Prospects

Continuing collaborations will provide an opportunity to confirm hypotheses generated from microarray experiments showing transcriptional changes in *Myzus persicae* in response to nicotine, and to provide experimental evidence for the significance of purine nucleotide recycling between *Acyrthosiphon pisum* and its endosymbiont *Buchnera aphidicola*. In this concluding chapter, I will describe the experiments which are underway to identify the products of nicotine metabolism in *M. persicae*, and to determine whether physiological changes in salivation and honeydew production are associated with aphid feeding from nicotine-containing diets. Finally, I will describe the experiments which I plan to carry out during my postdoctoral fellowship in the laboratory of Dr. Angela Douglas, in the Cornell Entomology Department, to study purine metabolism as a target for the control of aphid pests.

Synthesis of ^{14}C nicotine for the study of aphid nicotine metabolism

A collaboration between the Jander laboratory and the laboratory of Dr. Tadhg Begley (Texas A&M University, Department of Chemistry) is underway to synthesize radiolabeled nicotine for use in aphid feeding studies. Honeydew, the sugary excreta produced by aphids, will be collected from aphids feeding on ^{14}C nicotine-containing diets by lining the inside of insect feeding cups with aluminum foil. Foil will be removed after 24 hours of feeding, and honeydew will be extracted with methanol and subjected to analysis by high performance liquid chromatography (HPLC) coupled to a radiochemical detector. HPLC peaks exhibiting β-particle emission characteristic of ^{14}C decay will be further characterized by mass spectrometry, enabling the direct identification of products of aphid nicotine metabolism.

Physiological measurements of aphid salivation and honeydew production in response to nicotine

One of the most striking results from my artificial diet experiments on the effect of nicotine on aphid fecundity was that moderate amounts of nicotine increase fecundity in tobacco-adapted aphids (see Figure 5.2). We hypothesize that nicotine may be recognized by the aphid as a characteristic of a suitable host plant, thereby serving as a feeding stimulant. Production of honeydew can be measured as a proxy for diet consumption using a honeydew clock (Wilkinson and Douglas, 1995). Honeydew can be quantitated by visualizing the number of drops excreted onto a circle of filter paper treated with 0.1% bromophenol blue. Honeydew drops change in color from yellow to blue on the pretreated filter paper and can be readily counted, allowing an assessment of differences in honeydew production between aphids feeding on control and nicotine-containing diets.

Microarray analysis of gene expression changes in aphid heads in response to dietary nicotine revealed the upregulation of a number of salivation-associated genes (See Table 5.2). Salivation in response to the detection of a plant defense may be related to the role that aphid saliva is believed to play in suppressing plant defenses. The laboratory of Dr. Jeremy Pritchard (University of Birmingham) has extensive experience in physiological studies of aphids, using techniques such as the electrical penetration graph to record distinct features of the different feeding activities of aphids, including cell penetration, salivation, and ingestion (Tjallingii, 1978). A collaboration between the Pritchard and Jander laboratories has been established to compare the waveforms associated with aphid feeding on control and nicotine-containing diets, as a direct test of the hypothesis stemming from our microarray data that exposure to nicotine induces salivation in *Myzus persicae.*

Purine metabolism as a novel target to control aphid pests

As a postdoctoral researcher in Dr. Angela Douglas's lab, I will pursue a combination of molecular and biochemical strategies to test the hypothesis of a shared purine recycling pathway between *Acyrthosiphon pisum* and *Buchnera aphidicola*, stemming from my work annotating nucleotide metabolism genes for the pea aphid genome project. The strategy which Dr. Douglas and I will pursue is described below, in the project narrative for the postdoctoral fellowship grant which we recently submitted to the United States Department of Agriculture (June 2009).

Introduction

Aphids as pests

Over 250 species of aphids (superfamily Aphidoidea) feed on agricultural and horticultural crops worldwide, and they represent major pests in many agricultural sectors. Their pest status is a direct result of the aphid feeding habit, the consumption of plant phloem sap via needle-like stylets. Some aphids, especially those that inject toxins into the plant, can cause direct damage, and many are vectors for viruses that cause plant disease. Altogether, an estimated 50% of all known insect-borne plant viruses are transmitted by aphids, including representatives of 19 of 70 recognized virus genera. These aphid-borne viruses are of major economic importance, especially in cereal and potato crops. Finally, aphid honeydew deposited onto the plant surface can depress plant yield, especially when colonized by sooty molds, and cause cosmetic damage that reduces the economic value of crops.

In conventional agriculture, chemical control by carbamates, pyrethroids, neonicotinoids and others are crucial for aphid pest control. Cultural control (i.e. management of physical or biological environment of the crop to reduce aphid damage) by sowing in trap crops, maintaining non-crop refuge habitats, relay

cropping, intercropping etc. and biological control by parasitoids, predators and entomopathogenic fungi also play an play an important role in the management of aphid pests, especially as contributory elements in IPM programs.

The aphid-*Buchnera* symbiosis as a novel target for aphid pest management

The basis for this application is that the aphid symbiosis with the bacterium *Buchnera aphidicola* has potential as a target for novel aphid pest control strategies. Virtually all aphids of the family Aphididae, including all pests, possess this bacterium. The aphid is absolutely dependent on *Buchnera* for sustained growth and reproduction. When the bacteria are eliminated by antibiotic treatment, the aphids grow very slowly, at rates 10-20% of untreated individuals, and the *Buchnera*-free adults are reproductively sterile. Of course, antibiotics are not acceptable as aphicides because of their high cost and environmental reasons, but a more specific and cost-effective route to target the symbiosis would have great potential as a route to suppress aphid populations. In summary, there is potential to disrupt the aphid-*Buchnera* symbiosis in multiple ways: by interference with the function of the *Buchnera* cells or by disrupting its interactions with the aphid host, including nutritional interactions and vertical transmission.

Insight into the aphid-*Buchnera* symbiosis from genomics and systems biology: unusual features of purine metabolism in aphids

The genomes of both the pea aphid *Acyrthosiphon pisum* and its *Buchnera* (known as *Buchnera* APS) have been completely sequenced and annotated. This provides remarkable new insight into the biology of the association, generates new and very precise hypotheses of symbiosis function for experimental testing, and creates opportunities to identify novel pest control strategies.

In the last year, I have contributed to the annotation of the pea aphid genome by the International Aphid Genomics Consortium, including the annotation of metabolism genes. I established that the pea aphid genome contains orthologs to all *Drosophila* pyrimidine metabolism genes, but that its gene complement for purine metabolism is reduced relative to other insects with completely sequenced genomes. Specifically, the pea aphid lacks the genes for three enzymes in the purine salvage pathway: adenosine deaminase (ADA) and purine nucleoside phosphorylase (PNP), which are present in all other animal (including insect) genomes that have been sequenced, and hypoxanthine-guanine phosphoribosyltransferase (HGPRT), which is absent from all other insects with completely sequenced genomes (see Figure 3.1).

The purine salvage pathway is crucial for recycling of purines generated during metabolic turnover, cell death, tissue remodeling etc. and genetic lesions in the purine recycling pathway in other animals are associated with metabolic dysfunction, including human disease, e.g. Lesch-Nyhan syndrome. The reduced genetic capacity of the pea aphid for purine salvage is, consequently, most exceptional. A possible explanation for the incomplete capacity of the pea aphid for purine salvage comes from analysis of the genome of its bacterial symbiont, *Buchnera* APS. The complete metabolic network of *Buchnera* APS has been reconstructed, using the homology between *Buchnera* and *E. coli* genes. The network comprises just 289 reactions and 238 metabolites, just 30% and 40%, respectively, of the 972 reactions and 609 metabolites in the metabolic network of *E. coli* K12 (Thomas *et al*. 2009). In other words, the metabolic capabilities of *Buchnera* APS are reduced and highly fragmented in comparison to the related freeliving *E. coli*. This metabolic reduction of *Buchnera* APS includes purine metabolism, which appears nonfunctional when considered in isolation.

Crucial to this application, the fragmentary capabilities of the pea aphid and *Buchnera* APS for purine metabolism are complementary. When considered together, they generate functional metabolic pathways (see Figure 3.2). Specifically, *Buchnera* APS has the genetic capability to provision its nucleotide requirement through recycling of aphid derived nucleosides, especially guanosine. The critical *Buchnera* genes in this pathway are *deoDB, gpt/hpt, guaC, and purHAB* (see Figure 3.2), all of which are expressed (Viñuelas *et al*. 2007). These considerations have led to the hypothesis that the transfer of guanosine to *Buchnera* APS cells would compete with the aphid purine salvage pathway, resulting in reduced activity of the enzymes PNP and ADA, and the eventual loss of the cognate genes. In other words, the loss of the genes for PNP and ADA is an adaptation for supplying nucleosides to *Buchnera*.

The metabolic network of *Buchnera* APS has been applied to investigate the flux of metabolites *in silico* by flux balance analysis (Thomas *et al*. 2009). One output (also known as an overflow metabolite) of the network is the purine adenine, produced as a byproduct of the reactions involved in the synthesis of the polyamine spermidine (see Figure 3.4). *Buchnera* APS contains very high concentrations of spermidine, which has been proposed to stabilize its DNA. Although *Buchnera* APS cannot metabolize adenine, the aphid can salvage it, i.e. convert it to the nucleotide AMP, via the enzyme APRT. Thus, a proportion of the nucleoside delivered to the *Buchnera* APS is recycled by the aphid. Flux balance analysis was applied to estimate the amount of purine recycled in this way: it is 10% of the guanosine transferred to the *Buchnera* APS.

This analysis illustrates the power of combining genome annotation and systems level modeling. The annotation revealed interactions between the aphid and *Buchnera* APS that had not previously been considered; and the modeling provided a quantitative estimate of the significance of the putative interactions. The key limitation

is that the entire exercise is *in silico*, requiring experimental testing on the living system.

A key prediction is that 5'nucleotidase and APRT are expressed in the bacteriocytes (the aphid cells bearing *Buchnera*). The pea aphid genome contains six predicted 5'-nucleotidase genes and three predicted APRT genes. The publicly available EST libraries for pea aphid bacteriocytes include members of each of these gene families, and 454 pyrosequencing of cDNA from bacteriocytes have demonstrated that the 5' nucleotidase gene ACYPI002452 and the APRT gene ACYPI003109 are strongly expressed in these cells.

The protein IMP dehydrogenase (IMPDH, see Figure 3.1) is the rate-limiting step in GTP biosynthesis in animals. Highly sensitive inhibitors IMPDH have been developed as chemotherapeutic agents, blocking DNA synthesis in cancer cells through depletion of GTP (Catapano *et al*., 1995). Given the predicted reliance of *Buchnera* on aphid guanosine, inhibition of aphid IMPDH is expected to have deleterious consequences on the symbiosis.

The basis of aphid dependence on *Buchnera*

The potential of the aphid-*Buchnera* symbiosis as a target for novel pest control strategies is founded on the dependence of aphids on the symbiosis. The basis for this dependence is not fully understood. As considered above, Buchnera provides the insect with essential amino acids. One would, therefore, expect that antibiotic-treated aphids (which lack functional *Buchnera*) would perform well on diets with a complete amino acid content. They do not. Their growth rates on complete diets are, at most, half of those of untreated aphids. The implication is that although essential amino acid provisioning is required for aphid to thrive on a diet of plant phloem sap, it is not the sole basis for aphid dependence on *Buchnera*. *Buchnera* has also been implicated in

vitamin nutrition of aphids, but the loss of this capability would not be sufficient to account for the rapid decline in performance of antibiotic-treated aphids. The conclusion from genome annotation and systems analysis of metabolic flux that the purine metabolic capabilities of the pea aphid and *Buchnera* APS are inter-linked raises the possibility that the aphid requires *Buchnera* for purine metabolism, including the sustained production of nucleotides for DNA, RNA and ATP synthesis. In other words, the poor performance of antibiotic-treated aphids may be attributed in large part to disruption of the symbiotic interactions in purine metabolism; and purine metabolism is predicted to be a valuable target for aphid pest management strategies.

Specific objectives of the proposed research

I propose to investigate the coupling of aphid-*Buchnera* purine metabolism as a target for novel aphid pest control strategy.

Objective 1: **Determine the biochemical properties of recombinant APRT and 5' nucleotidase, the key enzymes in purine metabolism in bacteriocytes (the aphid cells containing *Buchnera*)**. I will produce recombinant enzymes and analyze their biochemical properties (activity, substrate specificity, response to inhibitors etc.) *in vitro*. This will validate the accuracy of the genome annotation and provide quantitative data essential for the future development of active agents.

Objective 2: **Quantify the flux of nucleosides, especially guanosine, into *Buchnera*, and the efflux of adenine from *Buchnera***.

These experiments will test the predicted metabolite exchange between the aphid and its symbionts, and will be conducted by radiotracer methods.

Objective 3: **Assess the impact of specific enzyme inhibitors on the symbiosis**.

Inhibitors of aphid purine metabolic enzymes will be fed or injected into aphids, and the viability and abundance of *Buchnera*, and the lifespan, growth rate and fecundity

of the aphid will be determined. These experiments will provide proof of principle for aphid purine metabolism as a novel target.

Rationale and Significance

The rationale and significance of this project are two-fold: the need for novel strategies against aphid pests; and the opportunities afforded by genomics and systems level biology to advance our fundamental understanding of insect pests.

Aphids are major pests. There is a widely recognized demand for novel aphid control strategies, especially in the context of increasing aphid resistance to current insecticides and heightened environmental and public health concerns about insecticide use. This need is heightened by the absence of any naturally-occurring *Bt* toxin effective against aphid pests, and the concern among some authorities that the effective control of 'chewing' insect pests, especially lepidopterans, by *Bt* may be increasing the pest status of aphids and other sap-feeding pests through ecological release (e.g. Faria *et al*. 2007). Current chemical control strategies against insect pests generally, and aphids in particular, are dominated by nervous system targets. These achieve rapid knock-down of the pest, but pose risks for non-target organisms, including biological control agents, and for public health. The great advantage of the aphid-*Buchnera* symbiosis as a target is that it is predicted to have unique features without parallel in non-symbiotic systems (Douglas 2007); and the purine metabolism, the focus of this study, is unique at least among animals with completely sequenced genomes.

The second basis for this project relates to the leverage obtained by genomics and systems level biology. Prior to my annotation of the pea aphid purine genes, there was no indication that the purine metabolism of aphids might be unusual, apart from the observation that the capacity of the *Buchnera* symbiont for purine metabolism

appeared to be fragmented (Zientz *et al*. 2004). The genome annotation and linked flux balance analysis represent a substantial development in our understanding of these pest insects, but only in the sense that they have generated very specific hypotheses of function. They are meaningful and useful only in so far as they are validated by experimental research, the core purpose of this project.

The full power of the *in silico* analyses can be obtained when the experimental data are fed back to refine the modeling, yielding additional hypotheses for experimental testing. This iterative process of experimental and *in silico* analysis has enormous potential for understanding the function of the aphid-*Buchnera* relationship, including the impacts of antagonists of purine metabolism. This process is made possible for this project through the collaborative link to the METNET consortium with whom I have already collaborated on purine metabolism (Ramsey *et al*., 2009). My mentor, Dr. Angela Douglas, initiated and was the original coordinator of the METNET program and, since her move to the USA, has become project partner for the consortium.

Approach

I will use a combination of biochemical and molecular techniques to test the hypotheses derived from my analysis of nucleotide metabolism in the pea aphid and *Buchnera* genomes.

In Objective 1, I will characterize the biochemical properties of the dominant 5' nucleotidase and APRT enzymes in the aphid bacteriocytes. This requires the pure enzyme, which I will generate by heterologous expression of the aphid cDNA in insect cell culture using baculovirus vectors. These enzymes represent promising targets for aphid control through disruption of the symbiosis, and quantitative data on their

activity, substrate specificity, and response to inhibitors will be essential for future development of control agents.

In Objective 2, I will directly quantify the flux of nucleosides, especially guanosine, into *Buchnera*, and the efflux of adenine from *Buchnera*, using radiotracer analysis of isolated *Buchnera* preparations. This will test the predicted metabolite exchange between the aphid and its symbiont.

In Objective 3, I will use commercially available small molecule inhibitors of IMP dehydrogenase and assess the effect of inhibition of this enzyme on the symbiosis, by analysis of *Buchnera* viability and abundance, and on the lifespan, growth rate, and fecundity of the aphid. IMP dehydrogenase is the rate-limiting step in guanine nucleotide biosynthesis. This will allow an assessment of the effects on the symbiosis of depletion of these metabolites, which cannot be synthesized independently by *Buchnera*.

Objective 1: **Determine the properties of the key enzymes APRT and 5' nucleotidase in bacteriocytes, the aphid cells containing *Buchnera***

We will characterize the biochemical properties of APRT and 5'-nucleotidase using recombinant proteins expressed in insect cells.

Methods

The Bac-to-Bac TOPO Expression System (Invitrogen, USA) will be used to express recombinant aphid proteins in insect cells. Baculovirus-mediated expression has the advantage over an *E.* coli expression system of increasing the likelihood of production of a functional protein, subjected to the post-translational modifications which may be necessary for its activity. The baculovirus system is capable of producing large amounts of recombinant proteins, and following pilot studies I will scale-up my

expression and purification methods to generate 10 mg of protein for biochemical analysis.

The two aphid genes will be cloned in parallel into an expression vector carrying a C-terminal 6xHis tag and transformed into an *E. coli* strain containing a baculovirus shuttle vector (bacmid). Transposition of the donor plasmid with the bacmid is facilitated by a helper plasmid also carried by the bacterium, producing a recombinant bacmid, which will be purified and used to transfect Sf9 insect cells. Recombinant His-tagged protein will be purified by chromatography from insect cell culture, and characterized according to its activity, substrate specificity, and response to inhibitors. Specifically, the activity of APRT will be determined by using high pressure liquid chromatography (HPLC) to measure the amount of AMP produced from adenine and 5-phosphoribosyl 1-pyrophosphate (PRPP). The activity of 5'-nucleotidase will be determined by using HPLC to measure production of the purine nucleoside from its cognate nucleotide substrate.

Expected outcomes

Baculovirus technology is capable of producing 1 mg of recombinant protein per 100 mL of cell culture. The likelihood that the recombinant proteins will be functional is much greater than if a bacterial expression system were used, justifying the additional work which insect cell culture will entail. I expect that adenine will be the only substrate which APRT is capable of utilizing, supporting our conclusion from genome annotation that the aphid is unable to independently recycle guanine, hypoxanthine, or xanthine. Given the predicted reliance of *Buchnera* on guanine nucleotides from the aphid, I expect that the 5'-nucleotidase to be cloned in my study, predicted by EST evidence to be highly expressed in the bacteriocyte, will accept guanosine as its preferred substrate.

Analysis and Interpretation of Data

Biochemical assays will be performed on purified proteins to measure their specific activity. The rate of product formation will be determined for a given amount of protein by directly measuring product formation by HPLC. Performing biochemical assays with different purine (APRT) and different purine mononucleotides (5'-nucleotidase) substrates will indicate the substrate specificity of each protein. In the event that the bacteriocyte-enriched 5'-nucleotidase has a preference for GMP as a substrate, this will support our hypothesis of the transport of guanosine from the aphid to *Buchnera.*

Pitfalls and Limitations

Expertise in recombinant protein expression in insect cell culture in the collaborator's lab (Gary Blissard) will be instrumental in ensuring the success of this objective. In the most unlikely event that the baculovirus system fails to produce a functional protein, I will first test whether the His-tag is interfering with protein activity. The cloning vector used for these experiments contains a Tobacco Etch Virus (TEV) recognition site which allows the removal of the 6xHis tag from the recombinant fusion protein using AcTEV protease (Invitrogen, USA). Cleavage of the His tag following protein purification will remove any deleterious effect which the tag may have on protein activity. In the event of further difficulties with baculovirus-mediated protein expression, heterologous expression of protein will be carried out in *Pichia pastoris.*

A limitation of these experiments is that they will be done with recombinant proteins rather than with native proteins purified from aphid tissue. There are three principal reasons why I will produce recombinant proteins rather than attempt to purify proteins directly from aphid tissue.

1) Aphids are small, and the amount of starting material required to purify sufficient protein for biochemical analysis will be impractical to collect.
2) I specifically predict based on transcript data that one of the six predicted 5'-nucleotidases from the aphid genome is most likely to play a role in a shared purine metabolic pathway between the pea aphid and *Buchnera*. It is unlikely that this one protein could be purified from aphid tissue at the exclusion of the other five. With the availability of whole genome sequence, however, I will be able to design unique primers which can be used to clone only our gene of interest.
3) Prior work in the mentor's laboratory (Price *et al*., 2007; Shakesby *et al*., 2009) has resulted in the successful expression of recombinant aphid proteins in heterologous systems, in sufficient quantity for biochemical characterization.

Objective 2: Quantify the flux of guanosine into *Buchnera* and efflux of adenine from *Buchnera*, primarily by radiotracer analysis of isolated *Buchnera* preparations

Objective 2 provides a direct test of the prediction from gene annotation studies that the symbiotic bacterium *Buchnera* takes up nucleosides, especially guanosine, and releases adenine (Fig. 3 and 4). The methods have been established in relation to research on the amino acid relations of aphids, and they can be applied to study the purine relations with very little modification.

Methods

The experimental material is *Buchnera* cells isolated directly from aphids. These bacteria cannot be maintained in long-term culture, but viable and metabolically active cells are readily obtained by protocols adapted from methods to isolate organelles (Harrison *et al*. 1989; Whitehead & Douglas 1992). Protocols and buffers have been

established that minimize mechanical and osmotic damage to the *Buchnera* cells. Briefly, aphids are crushed in a pestle and mortar, and the filtrate is passed through cheese-cloth, treated with detergent (0.008% Nonidet-P40) and washed with centrifugation, to obtain an enriched *Buchnera* preparation. The remaining contamination is predominantly lipid droplets, and the *Buchnera* cells are substantially purified from these and other contaminants by Percoll density gradient centrifugation. The coherent band of *Buchnera* cells will be aspirated from the gradient, washed and used immediately for metabolic studies. A convenient scale for isolations uses ca. 50 mg aphid material (25 final instar larvae) as starting material, yielding 0.5-1 x 10^7 *Buchnera* cells, but the protocol can readily be scaled, to yields up to ten-fold greater than this. The viability of the isolated *Buchnera* cells will be checked by BAC-Light bacterial viability stain (Invitrogen), which routinely give > 95% viability. Preparations with inferior viability occur rarely and they will not be used in experiments. The *Buchnera* cell density will be determined by hemocytometer counts (Wilkinson *et al*. 2001). Viability of *Buchnera* cells remains high (>95%) and oxygen consumption is stable for at least 6 hours in isolation (Whitehead & Douglas, 1993; Douglas). Even so, the experiments will be conducted immediately after isolation to ensure that the bacteria are in prime physiological condition.

For uptake and metabolism experiments, the *Buchnera* cells will be incubated in culture medium on a vacuum filter manifold, and the reaction will be terminated at different times after addition of radiolabelled precursor, e.g. ^{14}C-guanosine, by rapid vacuum filtration. The amount of radioactivity retained by the cells will be determined by liquid scintillation counting. All experiments will include zero-time values, which will be subtracted from all values, and radioactivity-free samples as the negative control. Additional controls will include heat-killed cells and cells incubated with excess non-radioactive substrate. Long-term (60-90 minutes) uptake experiments will

be conducted first with a range of substrate concentrations, to establish the overall pattern of uptake, including the duration of linear uptake. Subsequent experiments will use time courses over the linear range to examine uptake kinetics, specifically to quantify how the initial uptake rate varies with substrate concentration. On the basis of previous research (e.g. Whitehead and Douglas 1993), these timescales will be less than 5 minutes: the vacuum manifold is ideally suited for transport studies over time courses as short as 30 seconds. Subsequent experiments will check the specificity of the transporter(s) by competition experiments, in which uptake of the radiolabelled compound by the *Buchnera* cells is quantified in the presence of 2-to-5-fold excess of unlabelled compound, with control samples containing no competing compound.

I will also determine the identity of compounds released by isolated *Buchnera* preparations, particularly targeting whether adenine is released. I will adopt a pulse-chase design, i.e. the cells will be incubated with ^{14}C-guanosine (pulse), and then transferred to non-radioactive medium (chase). The reaction will be terminated by centrifugation, and the ^{14}C content of the cells and medium will be quantified by liquid scintillation counting: the amount released will be expressed in proportionate terms [i.e. (^{14}C in medium) / (^{14}C in medium + in cells)]. I will design the duration of the pulse and chase and the concentration of exogenous guanosine precursor to give readily-detectable release, on the expectation that 10% of the guanosine taken up is exported as adenine (see Introduction). I will then seek to identify the ^{14}C-compounds in the medium and cells. The data for the cells represent a crucial control: if the released compounds have the same identity and relative composition to those in the cell, then the release likely results from lysis or damage to the cells. The ^{14}C-labelled compounds of interest are the low-molecular weight compounds, which will be recovered as the 'TCA-soluble fraction', i.e. soluble in cold trichloroacetic acid, from both the medium and cells. After neutralization, the purines in this fraction will be

analyzed by reverse-phase HPLC with UV detection and liquid scintillation counting of the fractionated eluate (Munagala and Wang, 2003). Peak identification and quantification will be by comparison with standards.

Expected results

Buchnera is predicted to be able to utilize exogenous guanosine, and possibly other nucleosides, including inosine and adenosine. Based on previous studies of other metabolites, I expect sustained uptake of guanosine at rates that vary with substrate concentration, and that the uptake will show saturation kinetics, i.e. that the uptake rate plateaus at high substrate concentrations. This would be indicative of a carrier-mediated uptake system. If this obtains, I will be able to determine the Michaelis-Menten kinetic constants, K_m and V_{max}, which are established indices of the affinity and capacity of the transporter for the substrate. The competition experiments will identify overlap in transporter activity between the radiolabelled substrate and competing non-radioactive substrate. Generally, inhibition is maximal for the cognate compound (e.g. inhibition of ^{14}C-guanosine uptake by unlabelled guanosine). Other compounds that share the transporter(s) used by the radiolabelled compound are identified by significant reduction in ^{14}C uptake relative to the control (with no competing compound).

Buchnera is predicted to release adenine, as a metabolic product of guanosine. These experiments will provide the basis to test the prediction from flux balance analysis *in silico* that the exported adenine accounts for 10% of the guanosine taken up. I expect that the complexity of ^{14}C-labelled compounds in the medium will be considerably lower than in the cells: adenine may be the sole ^{14}C compound released, and its concentration in the medium may be considerably greater than that in the cells. These results would provide spectacular confirmation of the predictions of flux

balance analysis. I will pay particular attention to comparing the identity and complexity of ^{14}C-labelled compounds in the medium and in the cells, to check that non-specific release from damaged cells does not contribute to the observed export.

Analysis and Interpretation of data

The uptake of nucleosides and release of adenine (or other purines) will be quantified as mol per cell per unit time, and the kinetic constants for uptake will be determined by the Lineweaver-Burk plot or another transformation, with confidence limits. These empirical data will inform the flux balance model of *Buchnera* metabolism, specifically the identity of purine compounds that are utilized (i.e. inputs to *Buchnera* metabolism) and their relative contribution to the total requirement of the *Buchnera* for host-derived purines. The impact of these experimentally-validated inputs on *Buchnera* metabolism *in silico* will be assessed, with particular focus on biomass production, polyamine synthesis and adenine release. If the effects of varying the purine inputs on *Buchnera* metabolism *in silico* are great, additional experiments to validate the *in silico* results will be required. I have considered the value of this interplay between experimental data and *in silico* models previously in relation to troubleshooting. As indicated here, working between the modeling and experiments will also play a vital role in the overall analysis and interpretation of *Buchnera* purine metabolism.

Pitfalls and Limitations

Troubleshooting problems with the release experiments will be helped by expertise in the host laboratory, which has extensive experience of working with nutrient release from symbiotic microorganisms. The chief predicted problem is nonspecific release from damaged cells, and I have included controls that will enable me to identify this:

viability stain with BAC-Light, comparison of profile of ^{14}C in released compounds and cell contents. The likely cause of this problem is mechanical or osmotic damage during preparation and incubation. Support and advice from colleagues in the host laboratory will help me optimize my technique. A different possible problem is poor release, possibly arising from suboptimal composition of the culture medium. Empirical testing of multiple media would be an inefficient route to assess this issue. Instead, I will use the *in silico* tools developed for *Buchnera*, specifically flux balance analysis (Thomas *et al*. 2009), to explore how varying the inputs (i.e. medium composition) is predicted to influence adenine export, and then modify the medium accordingly. This could lead to an iterative development between empirical data and model design, to solve these core problems in *Buchnera* metabolism.

In vivo effects of aphid IMP dehydrogenase inhibitors on the aphid-*Buchnera* symbiosis

Methods

IMP dehydrogenase (IMPDH) is the rate limiting enzyme in *de novo* GTP biosynthesis. IMPDH activity is enhanced in tumor cells, and specific inhibitors of the enzyme, including tiazofurin and ribivirin, have been developed as chemotherapeutic agents (Catapano *et al*., 1995). Given the predicted reliance of *Buchnera* on aphid guanine nucleotides, we will determine the effects of GTP depletion on the symbiosis *in vivo* by adding these commercially available IMPDH inhibitors to aphid artificial diets. The impact of the inhibitors will be assayed by analysis of *Buchnera* viability and abundance, and lifespan, growth rate and fecundity of the aphid. Aphids will be reared on liquid diets containing sucrose, amino acids, and varying concentrations of inhibitor. Guanine, guanosine, and GTP will be added to artificial diets containing

toxic levels of inhibitor, and the ability of these compounds to mitigate inhibitor toxicity will be assessed.

Buchnera abundance will be determined by quantitative real-time PCR for the *Buchnera* gene *dnaK* using the TaqMan assay, as used routinely in the Douglas lab (e.g. Douglas *et al*. 2007; Chandler *et al*. 2008). Values of threshold cycle (*C*T: the cycle at which a significant increase in fluorescence occurs) below 40 were taken as a positive result. Standard curves will be generated with serial dilutions of known concentrations of template, and *dnaK* copy number was estimated from the standard curve. The data will be normalised to total DNA content per sample, to control for any variation in efficiency of DNA extraction. *Buchnera* viability will be determined by the BacLight live/dead stain, as also used in objective-2.

Aphid growth rates will be determined as weight increase over fixed timescales (2-7 day-old) of larval development, quantified on a microbalance, to give the relative growth rates; and fecundity will be determined from daily counts of offspring production over the full adult lifespan. These methods are applied routinely in the Douglas laboratory.

Expected Outcomes

Inhibition of aphid IMPDH is expected to have deleterious effects on aphid fecundity, live span, and growth rate, due to the essential role of this enzyme in guanine nucleotide biosynthesis. Because genome analysis suggests that *Buchnera* has no capacity for GTP biosynthesis independent from the aphid, depletion of the aphid guanine nucleotide pool will result in less guanosine transferred to the symbiont, resulting in compromised viability of the bacterium.

Analysis and Interpretation of Data

Dose-response curves will be produced to quantify the effect of the IMPDH inhibitors tiazofurin and and ribivirin on the pea aphid and *Buchnera.* IC_{50} values will be determined for each compound using linear regression analysis.

Based on the hypothesis that guanosine is transported from the aphid to *Buchnera,* it may be expected that IMPDH inhibitor toxicity to *Buchnera* will be rescued by addition of guanosine to aphid artificial diets. Lacking orthologs of known purine salvage genes, however, the pea aphid is incapable of synthesizing guanine nucleotides from guanosine. Therefore the addition of guanosine to IMPDH inhibitor-containing diets is not expected to rescue the direct effects of guanine nucleotide depletion on the aphid. If aphids are able to tolerate IMPDH inhibitors when guanosine is added to aphid diets, this will indicate either that aphids possess a novel mechanism for purine salvage, or that guanine nucleotides salvaged by *Buchnera* are transferred back to the aphid.

The addition of guanine to aphid artificial diets is also not expected to reverse the toxicity of IMPDH inhibitors. The pea aphid contains a gene for guanine deaminase, which can synthesize xanthine from guanine. In the absence of HGPRT (see Fig. 3.1), however, the aphid has no way to make GMP or XMP from guanine or xanthine, respectively. Addition of GMP to aphid diets containing IMPDH inhibitors is expected to reverse deleterious effects of inhibitor on both the aphid and *Buchnera*.

Pitfalls and limitations

A primary pitfall of aphid artificial diet experiments is the possibility that the inhibitors added to aphid diets may not be soluble in the solution of water, sucrose, and amino acids. In my prior experience, aphids can tolerate small (up to 1%) levels of methanol in their diets. To facilitate solubility of inhibitors, I will make stock

solutions of the compounds dissolved in methanol, to be added to diets at varying concentrations. There is also the possibility that inhibitors added to aphid diets will have antifeedant effects, and observed deleterious phenotypes will be due to starvation rather than actual effects of the chemical. This effect can be assessed by quantitation of honeydew produced by aphids feeding on control and inhibitor-containing diets – if the aphids are eating less on the treated diet they will also be excreting less (Karley *et al*., 2002). This antifeedant effect can be avoided by directly injecting inhibitors into the aphid body cavity using a nanoliter microinjector. This approach has been used in the mentor's lab to inject radioisotopes into aphids (Wilkinson *et al*., 2001).

An additional challenge of these experiments will be to discriminate between direct and indirect effects of IMPDH inhibitors on *Buchnera*. Viability and abundance of *Buchnera* may be directly affected by the inhibition of IMPDH and the consequence disruption of guanosine supply, or they may experience indirect effects because of the effect of the inhibitor on the aphid. *Buchnera* relies on a healthy aphid to support the many aspects of the symbiosis. These different scenarios can be deconvoluted by comparing the effect on *Buchnera* of a toxin known to directly target the aphid nervous system with the inhibitors used in this objective.

The primary limitation of this objective is that the inhibitors considered are not viable pest control agents themselves, due to their activity against the human ortholog of the aphid enzyme. However, they represent convenient chemical tools to test our hypotheses of the role of purine metabolism in the aphid-*Buchnera* symbiosis. Should these experiments prove successful, they will lead the way for future efforts to identify unique characteristics of aphid IMPDH which can be used to design inhibitors orthogonal to the human enzyme.

REFERENCES

Catapano CV, Dayton JS, Mitchell BS, Fernandes DJ (1995). GTP depletion induced by IMP dehydrogenase inhibitors blocks RNA-primed DNA synthesis. *Mol Pharmacol* **47**: 948-955.

Chandler SM, Wilkinson TL and Douglas AE (2008). Impact of plant nutrients on the relationship between a herbivorous insect and its symbiotic bacteria. *Proc R Soc Lond B* **275**: 565-570.

Douglas AE (2007). Symbiotic microorganisms: untapped resources for insect pest control. *Trends Biotechnol* **25**: 338-342.

Faria CA, Wackers FL, Pritchard J, Barrett DA, and Turlings TCJ (2007). High susceptibility of Bt maize to aphids enhances the performance of parasitoids of lepidopteran pests. *PLoS One* **2(7)**: e600.

Harrison CP, Douglas AE and Dixon AFG (1989). A rapid method to isolate symbiotic bacteria from aphids. *J Invertebr Pathol* **53**, 427-428.

Munagala NR and Wang CC (2003). Adenosine is the primary precursor of all purine nucleotides in Trichomonas vaginalis. *Mol Biochem Parasitol* **127(2)**: 143-149.

Price DRG, Karley AJ, Ashford DA, Isaacs HV, Pownall ME, Wilkinson HS, Gatehouse JA and Douglas AE (2007). Molecular characterisation of a candidate gut sucrase in the pea aphid Acyrthosiphon pisum. *Insect Biochem Mol Biol* **37**: 307-317.

Ramsey JS, MacDonald SJ, Jander G, Nakabachi A, Thomas GH, and Douglas AE (2009). Genomic evidence for complementary purine metabolism in the pea aphid Acyrthosiphon pisum and its symbiotic bacterium Buchnera aphidicola. *Insect Mol Biol*, in press.

Shakesby AJ, Wallace IS, Isaacs HV, Pritchard J, Roberts DM and Douglas AE (2009). A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem Mol Biol* **39**: 1-10.

Thomas GH, Zucker J, Macdonald AJ, Sorokin A, Goryanin I and Douglas AE (2009). A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. *BMC Syst Biol* **3**: 24.

Tjallingii, WE (1978). Electronic recording of penetration behaviour by aphids. *Entomol Exp Appl* **24**:521-530.

Viñuelas J, Calevro F, Remond D, Bernillon J, Rahbé Y, Febvay G, Fayard J-M and Charles, H (2007). Conservation of the links between gene transcription and chromosomal organization in the highly reduced genome of Buchnera aphidicola APS. *BMC Genomics* **8**: 143.

Whitehead LF and Douglas AE (1993). A metabolic study of Buchnera, the intracellular bacterial symbionts of the pea aphid, Acyrthosiphon pisum. *J Gen Microbiol* **139**: 821-826.

Wilkinson TL, Adams D, Minto LB, and Douglas AE (2001). The impact of host plant on the abundance and function of symbiotic bacteria in an aphid. *J Exp Biol* **204**:3027-3038.

Zientz E, Dandekar T, Gross R (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol Mol Biol Rev* **68**:745-770.

APPENDIX

Supplementary Material 3.2. **Calculation of the stoichiometric coefficient of spermidine for flux balance analysis of *Buchnera***

1. Nakabachi and Ishikawa (2001) determined the spermidine concentration in bacteriocytes (the aphid cells bearing *Buchnera*) of 10 day old aphid bacteriocytes at 70 nmol mg^{-1} protein.
2. *Buchnera* occupy 60% of bacteriocyte volume (Whitehead and Douglas, 1993).
3. From (1) and (2), the spermidine concentration in *Buchnera* cells is calculated as 117 nmol mg^{-1} protein, on the assumption that all bacteriocyte spermidine is in the *Buchnera* cells.
4. Based upon experimental measurements of *E. coli* K-12 biomass composition (Neidhart, 1987), protein comprises 55% of bacterial dry weight.
5. From (3) and (4), the concentration of spermidine in *Buchnera* is 64.35 nmol mg^{-1} dry weight or 16.4 μg mg^{-1} dry weight (MW of spermidine = 255).
6. Based upon experimental measurements of *E. coli* K-12 biomass composition (Neidhart, 1987), t he total nucleotide content of *Buchnera* is 23.6% bacterial dry weight equivalent to 236 μg mg^{-1} dry weight.
7. From (5) and (6), *Buchnera* is calculated to contain 0.07 units spermidine per unit total nucleotide.
8. The sum of the nucleotide stoichiometric coefficients from the iGT196 biomass equation is 0.72 (GTP=0.203, CTP=0.126, UTP=0.136, dCTP=0.013297, dGTP=0.013297, dTTP=0.036803, dATP=0.036803, ATP=0.153, AMP=0.001).
9. From (7) and (8), the stoichiometric coefficient for spermidine is calculated as 0.05.

REFERENCES

Nakabachi, A. and Ishikawa, H. (2001). Expression of host S-adenosylmethionine decarboxylase gene and polyamine composition in aphid bacteriocytes. *Insect Biochem Mol Bio* **31**, 491-496.

Neidhardt, F. C. (1987). Chemical composition of *Escherichia coli*, pp. 3–6. In: F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (eds), *Escherichia coli* and *Salmonella typhimurium:* Cellular and molecular biology. American Society for Microbiology, Washington, DC.

Whitehead LF and Douglas AE (l993). Populations of symbiotic bacteria in the parthenogenetic pea aphid (*Acyrthosiphon pisum*) symbiosis. *Proc R Soc Lond B Biol Sci* 254, 29-32.