UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Regulation and Epigenetic Patterns of Aphid Tissues That are Involved in Symbiosis and Insect-Plant Interactions

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Entomology

by

Do Hyup Kim

June 2019

Dissertation Committee:

Dr. Allison K. Hansen, Chairperson
Dr. Kerry Mauck
Dr. Quinn McFrederick
Dr. Richard Stouthamer

The Dissertation of Do Hyup Kim is approved:

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Chapter 1: This chapter is from the review that has been published as "Novel insights into Insect-microbe interactions – role of epigenomics and small RNAs" in 2016 (*Front. Plant. Sci.* 7:1164; doi: 10.3389/fpls.2016.01164). I would like to thank the co-author, Dr. Margaret Thairu and corresponding author, Dr. Allison Hansen. Also, I would like to thank Dr. Patrick Degnan, Amanda Adams, Katie Frye, and two reviewers for their helpful comments and suggestions on earlier drafts and the University of Illinois at Urbana-Champaign for funding.

Chapter 2: This chapter is the published article, "Key transport and ammonia recycling genes involved in aphid symbiosis respond to host-plant specialization" (2018. *G3.* 8: 2433-2443; doi: 10.1534/g3.118.200297). I thank Dr. Nancy Moran, Dr. Patrick Degnan, Dr. Margaret Thairu, Amanda Adams, Zach Costilow, Danielle Campbell, and Katie Frye for their helpful comments and suggestions on previous versions of this manuscript. I also would like to thank the undergraduate researchers, Alex Riley, Aejin Shon, and Eunji Jeong, who have been working with me on rearing the aphid samples. This work was supported by the USDA AFRI postdoctoral fellowship from the United States Department of Agriculture to Dr. Allison Hansen (2011-67012-30707), the University of Illinois at Urbana-Champaign, and the University of California, Riverside.

Chapter 3: I would like to thank the corresponding author, Dr. Allison Hansen, for the review of the manuscript and funding. Also, I would like to thank Dr. Doris Lagos-Kutz for aphid rearing and providing aphid samples. This work was supported by the University of Illinois at Urbana-Champaign and the University of California, Riverside.

Chapter 4: I would like to thank the corresponding author, Dr. Allison Hansen, for reviewing the manuscript and suggesting helpful ideas. I would like to thank Dr. Kerry Mauck and Dr. Gregory Walker for providing aphid lines and tips on rearing them. This work was supported by the University of California, Riverside.

DEDICATION

First and foremost, I would like to thank my wife, Jisil Lee for her support, love, and encouragement. Although she hates anything that crawls, she has been a perfect wife of an entomology graduate student. I would also like to thank my loving parents for their support both financially and emotionally. All my family members were very supportive and encouraging as well. I would like to thank my sisters for their support. Moreover, I would like to thank my advisor Dr. Allison Hansen for all her support. Although she still may not believe it, I truly think she is the best advisor that a graduate student can possibly have. Also, I am very lucky to have Margaret Thairu as my academic sister. My life in Hansen lab had become more than 50% easier and smoother just because of her. I would like to thank my dissertation committee members, Drs. Richard Stouthamer, Quinn McFrederick, and Kerry Mauck for their advice. There have been so many professors that helped me with precious advice, especially Dr. Patrick Degnan. I have felt very fortunate to have his free advice on every corner of my research. I would also like to thank my cat Choco for his emotional support. Finally, I would like to thank the Entomology Department of both UCR and UIUC.

ABSTRACT OF THE DISSERTATION

The Regulation and Epigenetic Patterns of Aphid Tissues That are Involved in Symbiosis and Insect-Plant Interactions

by

Do Hyup Kim

Doctor of Philosophy, Graduate Program in Entomology
University of California, Riverside, June 2019
Dr. Allison K. Hansen, Chairperson

In this dissertation, the integrated metabolism of three aphid species (*Acyrthosiphon pisum, Aphis glycines,* and *Myzus persicae*) and their specialized endosymbionts was explored. In the first chapter, tissue- and host plant-specific profiles of gene expression and CpG methylation of *A. pisum* were analyzed. Through RNA-Seq and whole genome bisulfite sequencing, I identified key metabolic genes that are differentially expressed and methylated between bacteriocytes and body cells. Moreover, I demonstrated for the first time that key aphid genes involved in the regulation of aphid-*Buchnera* symbiosis are differentially expressed and methylated depending on the aphid's host plant diet, suggesting that DNA methylation may be a key regulatory factor that induces phenotypic variation depending on the host plant diet. In the second aphid system, I empirically and computationally confirmed the functional CpG methylation system in *Aphis glycines.* Also, I showed that lineage-specific genes in *A. glycines* have significantly lower CpG methylation levels compared to evolutionarily conserved genes.

Moreover, five aphid-specific genes were identified to play key roles in insect-plant interactions at the epigenetic level. In the last aphid system, I identified differentially expressed genes in bacteriocytes of *Myzus persicae* compared to its body cells. I demonstrated overall up-regulation of the genes that are involved in amino acid biosynthesis as well as key genes in aphid-*Buchnera* integrated metabolism. I then compared gene expression patterns of *M. persicae* bacteriocytes to those of *A. pisum* bacteriocytes. I found that the two closely related species have very similar gene expression profiles in their bacteriocytes while there are lineage-specific expression signatures in some metabolic genes.

Table of Contents

List of figures

List of Tables

Chapter 1: Introduction

Aphid bacterial endosymbionts and gene expression regulation

Microbial associates that interact with insects can produce a wide array of metabolic products that complement the metabolic needs of their herbivorous hosts (Hansen and Moran, 2014). Consequently, microbes that form persistent but noninvasive associations with their hosts have the potential to provide their hosts with useful novel gene products in a short evolutionary timespan. How these symbiotic microbes and animals reciprocally respond to and regulate shared metabolic processes is a nascent but emerging area of research.

Animals, including insects, can biosynthesize some but not all of the amino acids that are required for building proteins. Food sources that are deficient in those essential amino acids (EAA), such as plant sap, present a nutritional challenge to consumers. Most insect herbivores that feed on a phloem or xylem sap can survive on such nutrient-deficient diet because they harbor nutritional symbionts (Hansen and Moran, 2014). One model system that has been productive for teasing apart the regulatory mechanisms of shared animal-microbe metabolic processes is the pea aphid, *Acyrthosiphon pisum*, a phloem-feeding insect in the order Hemiptera, and its mutualistic endosymbiont bacterium *Buchnera aphidicola*. Aphid and *Buchnera* physiologies are integrated to produce amino acids and this occurs within specialized aphid cells called bacteriocytes. *Buchnera* relies on the aphid for the biosynthesis of aphid-encoded non-essential amino acid pathways, and the aphid relies on *Buchnera* for the biosynthesis of *Buchnera*-encoded essential amino acid pathways. Previous work on this system supports the

prevailing hypothesis that the integrated metabolism is regulated primarily by the aphid host via aphid-encoded transporters (Price *et al.,* 2014) and aphid genes that complement *Buchnera*'s EAA pathways (Hansen and Moran, 2011; Poliakov *et al.,* 2011). Moreover, an aphid-encoded protein of bacterial origin can be transported into *Buchnera* cells and therefore a cross-domain protein translocation system exists for this intimate symbiosis (Nakabachi *et al.,* 2005, 2014).

Gene expression of aphid bacteriocytes has been characterized at the transcriptome and proteome level (Hansen and Moran, 2011; Poliakov *et al.*, 2011). Pathways involved in the amino acid metabolism are especially enriched in bacteriocytes compared to other aphid body cells (Hansen and Moran, 2011; Poliakov *et al.*, 2011). However, the regulatory factors that lead to the development of these tissues and their distinct expression profiles are not well-understood. In other animal systems, the primary regulatory factors that determine a eukaryotic cell's fate and its potential reprograming include histones, DNA methylation, noncoding RNAs, and transcription factors (Peter and Davison, 2015). Work by Braendle *et al.* (2003) identified three transcription factors (Dll, En, and Ubx or Abd-A) that are expressed in temporal patterns during bacteriocyte development in aphid embryos. The timing and expression of this subset of transcription factors is unique compared to any other cell type in insect embryos. Currently, it is unclear how these transcription factors may regulate metabolic processes or if other unknown co-factors are involved during embryonic stages or later during maternal bacteriocyte development. Moreover, it is unknown if chemical marks on histones and/or DNA are important in the regulation and metabolic reprogramming of these symbiotic

cells, especially in response to environmental signals such as host plant nutrients or secondary plant compounds. Therefore, further understanding of how different subsets of host genes turn on and off in bacteriocyte development in response to environmental stimuli is required in order to fully understand how these intimate symbioses evolved, how they are maintained, and how they may ultimately influence host plant interactions.

Gene regulation via epigenomic mechanisms

Under Darwinian natural selection, random genetic mutations in the DNA molecule are inherited from parent to offspring, and an increase in frequency of a given mutation within a population contributes to differential reproductive success (Dobzhansky, 1937). Alternative mechanisms of adaptation, such as Lamarkian inheritance, where an individual can pass down acquired traits that are obtained during its lifetime, have been hotly debated (Pipel and Rechavi, 2015). Evolutionary theory that involves different variations of Lamarkian inheritance have been resurrected multiple times throughout history (Burkhardt, 2013). One controversial variation of Lamarkian inheritance, neo-Lamarkism (Skinner, 2015), has been proposed to explain epigenetic inheritance (Jablonka and Lamb, 2015) and CRISPR-cas immunity in Bacteria and Archaea (Koonin and Wolf, 2009), because these acquired traits are not random but are induced by the environment in a predictable fashion and are inherited through generations.

Epigenetic marks referred to here as chemical marks on DNA and histones are responsible for tissue-specific gene expression in eukaryotes and thus can lead to a

change in an organism's phenotype (Gama-Sosa *et al.*,1983). If the environment induces epigenetic marks in a repeatable way and these marks are inherited across generations then there is potential for epigenetics to play important roles in organismal adaptation in natural populations (Gadjev, 2015), which can then affect the organism's interaction with other organisms. Therefore, we propose that epigenetic mechanisms may be important for the evolution of both insect-plant and insect-microbe interactions.

Patterns of DNA methylation in different phyla

One type of epigenetic mark that is widespread and generally occurs across all domains of life is DNA methylation. DNA methylation involves the enzymatic addition of a methyl group to individual nucleotide bases of DNA in chromosomes by DNA methyltransferases (DNMTs). Now more than ever, DNA methylation has become more tractable to study due to the recent advancement in sequencing technology and bioinformatics. In turn the field of epigenomics is more accessible to researchers for both model and non-model organisms. As such the number of research articles on DNA methylation has been steadily increasing (Romanoski *et al.*, 2015).

The role and patterns of DNA methylation vary widely among the three domains of life (Jeltsch, 2010). Several independent losses of DNA methylation have occurred in various eukaryotic taxa, suggesting that the role of DNA methylation may not be essential for all eukaryotic species (Field *et al.*, 2004; Wion and Casadesús, 2006). For example, in the model systems *Caenorhabditis elegans* (roundworm) and *Drosophila melanogaster* (fruit fly) DNA methylation is not functional because of lineage specific

losses of DNMTs (Goll and Bestor, 2005). Nevertheless, DNMTs are present and DNA methylation is prevalent and functional in a wide diversity of other eukaryotic taxa including plants, vertebrates, and invertebrates (Jeltsch, 2010).

In vertebrates, including humans, cytosine methylation is widespread in the genome, specifically at cytosine-phosphate-guanine (CpG) dinucleotide sites (Bird, 1986). However, CpG-rich regions called CpG islands, which are typically 300- to 3000-base pairs long are located primarily in the promoter regions of vertebrate genes, and are largely un-methylated (Bird, 1986). Methylation of even a single CpG site in a promoter region can significantly inhibit transcription of the downstream genes (Robertson *et al.*, 1995). Such transcriptional inhibition or gene silencing is an important role for DNA methylation as it also helps maintain the integrity of the genome by silencing transposable elements (Zamudio *et al.*, 2015). DNA methylation also has a well-established role in imprinting, such as mammalian X-chromosome inactivation (Augui *et al.*, 2011; Balaton *et al.*, 2015), and differential expression of parental-specific alleles (Reik *et al.*, 1987; Li *et al.*, 1993; Razin and Cedar, 1994). Furthermore, DNA methylation in gene-body regions (e.g., un-translated regions, exons, and introns) can also affect the activity of genes in vertebrate genomes. For example, in human cell lines the inhibition of DNA methylation in gene-body regions resulted in the alternative splicing of exons (Maunakea *et al.*, 2013).

In invertebrates, cytosine methylation also plays an important role in gene regulation. Epigenomic research in invertebrates initially lagged behind, because the two main invertebrate model species in genetics, *C. elegans* and *D. melanogaster*, do not have

active copies of DNMTs. Nevertheless, DNA methylation has been observed in a diversity of other invertebrate species. For example, in the Pacific oyster *Crassostrea gigas*, different levels of CpG methylation have been observed that correlate with gene functions (Gavery and Roverts, 2010). In the Chinese white shrimp, *Fenneropenaeus chinensis*, tissue-specific DNA methylation was observed (He *et al.*, 2015). Furthermore, DNA methylation has been reported in many insect species of various orders including Diptera, Hemiptera, Hymenoptera, Lepidoptera, Coleoptera, Odonata, and Orthoptera (Field *et al.*, 2004; Richards *et al.*, 2008; Walsh *et al.*, 2010; Xiang *et al.*, 2010; Zhang J. *et al.*, 2015; Zhang M. *et al.*, 2015). In general, methylation levels of invertebrate CpG sites are relatively low, ranging from 0.36-20% (Regev *et al.*, 1998), compared to mammalian systems where 60-90% of all CpG dinucleotides are subject to methylation (Suzuki and Bird, 2008).

Methylome studies across different invertebrate taxa have revealed that DNA methylation is often confined to genic regions (promoters, exons, and intros) of the gnome, whereas intergenic regions remain largely unmethylated (Suzuki and Bird, 2008). In hymenopteran genomes, such as parasitoid wasps, ants, and bees, low levels of DNA methylation occur within transposable elements compared to vertebrate genomes. These results suggest that DNA methylation has no or very little association with the repression of transposable elements as shown in vertebrates (Yan *et al.*, 2015). DNA methylation within invertebrate genes has been associated with gene activation and alternative splicing. For example, loss of DNA methylation from multiple CpG sites within the insecticide-detoxifying esterase gene E4 of the green peach aphid *Myzus persicae* was

associated with a reduction of transcription of the esterase gene E4, and thus increased sensitivity to pesticides (Field *et al.*, 1989; Field, 2000). Also, several studies have proposed that DNA methylation is associated with alternative splicing of mRNA transcripts, which leads to behavioral regulation and caste specificity in eusocial insects including bees (Foret *et al.*, 2012; Li-Byarlay *et al.*, 2013), ants (Bonasio *et al.*, 2012), and termites (Terrapon *et al.*, 2014).

Insect DNA methylation and adaptation to variable environments

Mounting evidence from the handful of non-model animal systems that have been studied suggests that environmental cues can trigger the reprogramming of cells through DNA methylation, resulting in the regulation of adaptive traits (Kucharski *et al.*, 2008; Moczek and Snell-Rood, 2008; Alvarado *et al.*, 2015; Table 1-1). As such, differential methylation patterns have the potential to produce an adaptive regulatory response to current environmental conditions. Environmental signals such as diet, stress and anxiety have been shown to alter DNA methylation patterns during an organism's lifetime (Weaver *et al.*, 2004; Jankard and Herman, 2008; Schwenk *et al*., 2013). For example, in the honey bee nutritional cues from royal jelly regulate queen determination via epigenetic mechanisms. Specifically, the gene, *dynactin p62*, is differentially methylated in queens compared to workers, and it is hypothesized to be a key gene in regulating different developmental pathways (Kucharski *et al.*, 2008). In this study, when DNA methylatransferase 3 (*Dnmt3*) is silenced in larvae that feed on protein-rich royal jelly these larvae develop into fertile queens with fully developed ovaries (Kucharski *et al.*,

2008). This result indicates that a nutritional signal can alter epigenetic patterns resulting in caste determination. In another honeybee study, social stimuli of bees were highly correlated with changes in DNA methylation patterns between worker and nurse bees of the same age (Lockett *et al.*, 2012). One particular CpG site in the gene Protein kinase C-binding protein 1 (*PKCbp1*) with variable levels of methylation between worker and nurse bees was strongly correlated with the alternative splicing of its gene product. The direct consequences of this alternative splicing however are unclear (Lockett *et al.*, 2012). DNA methylation also plays a role in caste determination of another social hymenopteran, the carpenter ant *Camponotus floridanus*, by modifying ant body size, a key trait associated with the division of labor (Alvarado *et al.*, 2015).

In the parasitoid wasp *Nasonia vitripennis*, changes in photoperiod are hypothesized to induce genome-wide DNA methylation changes (Pegoraro *et al.*, 2016). When day length is decreased, female *N. vitripennis* wasps induce developmental arrest of their progeny (diapause). This photoperiodic response allows the larvae to survive throughout winter. Knock-down of either DNMT1a or DNMT3 in *N. vitripennis* parents disrupted the photoperiod-induced developmental arrest of their larvae. Although the exact mechanisms are yet to be elucidated, these results suggest that environmentally induced diapause in *N. vitripennis* are linked to DNA methylation.

In addition to hymenopterans, methylation contributes to adaptive regulatory responses to environmental conditions in aphids. In the green peach aphid, *M. persicae*, individuals resistant to an organophosphate pesticide encode a differentially methylated esterase gene that confers the resistant phenotype (Field *et al.*, 1989, 1996; Hick *et al.*,

1996). Biotypes of the Russian wheat aphid, *Diuraphis noxia*, that differ in virulence toward their host plant display differential methylation patterns for key genes that are expressed in their salivary glands, suggesting that methylation may play an important role in this insect's ability to feed on different host plant cultivars (Gong *et al.*, 2012). In *A. pisum*, it has been shown that extreme temperatures may result in variation in DNA methylation patterns, which are correlated to different color phenotypes within genetic clones (Dombrovsky *et al*., 2009). This study revealed that the intensity of methylation in CpG-islands within aphid cuticular genes varied dramatically between three different *A. pisum* color morphs (white, pink, and green). Furthermore, the authors identified correlations between CpG island methylation and growth rate, morph development, and pigmentation of the aphid population by pharmacologically inhibiting the DNA methyltransferases (Dombrovsky *et al*., 2009). In sum, DNA methylation may help drive rapid and precise gene regulation in response to variable environmental conditions.

Role of DNA methylation in symbioses

In the sections above, I detailed several examples of how environmental cues can induce specific DNA methylation patterns in insects, which can result in adaptive gene expression profiles. Although microbial associations are ubiquitous in many insect systems, as of yet, there has not been extensive research on how microbes affect insect host epigenomics. Nevertheless, we predict that epigenomic mechanisms may play major roles broadly in insect-plant and insect-microbe ecology and evolution. For example, insect microbial associations have facilitated numerous host plant niche expansions, and

the diversification of insect lineages (Hansen and Moran, 2014). Moreover, insect symbionts can contribute to a variety of extended insect host phenotypes, which include: defense against viral pathogens, fungal pathogens, and parasitoids (Kaltenpoth *et al.*, 2005; Oliver *et al.*, 2005; Scarborough *et al*., 2005; Scott *et al.*, 2008; Teixeira *et al*., 2008; Vorburger *et al*., 2009), conferring thermal tolerance (Dunbar *et al*., 2007; Brumin *et al*., 2011), facilitating food digestion (Brownlie *et al*., 2009; Salem *et al*., 2012), and manipulating sexual reproduction (Stouthamer and Werren, 1993). Currently, we are still in the discovery phase in identifying specific genetic mechanisms that facilitate these important microbe-induced, insect-extended phenotypes.

To the best of our knowledge the only studies that demonstrate an effect of microbes on insect epigenomics are of *Wolbachia* and mosquitos. *Wolbachia* is an intracellular bacterial symbiont that is both vertically and horizontally inherited in numerous insect species and commonly enhances its transmission through reproductive manipulations (Stouthamer and Werren, 1993). Nevertheless, mosquitos do not have functional DNA methylation like *Drosophila* because they do not encode DNMT1 and DNMT3 (Holt *et al.*, 2002; Nene *et al*., 2007). However, they do encode the methyltransferase, DNMT2, which has substrate specificity for tRNAs (Goll *et al*., 2006), contributes to antiviral defense in *Drosophila* (Durdevic *et al*., 2013), and is involved in random genome methylation patterns (Kunert *et al*., 2003). In one study, when the pathogenic strain of *Wolbachia* (wMelPop) infects the mosquito, *Adedes aegypti*, the mosquito is hypomethylated because *Wolbachia* suppresses DNMT2 (Zhang G. *et al*., 2013). When DNMT2 is overexpressed in mosquito cell lines, *Wolbachia* replication is

inhibited, suggesting that suppression of DNMT2 is for the microbe's survival. Conversely, in *A. aegypti* DNMT2 is induced by the Dengue virus, and this induction promotes virus replication. *In vivo*, this antagonistic interaction ultimately results in *Wolbachia* suppressing the Dengue virus via DNMT2 suppression (Zhang G. *et al*., 2013). In another study, the *Wolbachia* strain wMelPop results in both the methylation and de-methylation of *A. aegypti*'s genome (Ye *et al*., 2013). For the most part, these changes in methylation primarily appear to be random (Ye *et al*., 2013). In the latter study, the direct effect of differential methylation on transcription in wMelPop-infected compared to uninfected mosquitos remains unclear.

Given the paucity of evidence of symbionts affecting invertebrate host epigenetics and studies of insect hosts with functional DNA methylation systems, investigating the effect of symbionts on vertebrate hosts may provide insights into possible ways that symbionts may impact DNA methylation in insect genomes. In general, DNA methylation in vertebrate studies of gut-associated microbes has revealed that host immune responses and microbially derived metabolites affect host DNA methylation. For example, in mice, Takahashi *et al*. (2011) found that the methylation level of the Toll-like receptor 4 gene in intestinal epithelial cells is significantly lower in germ-free mice compared to conventional mice. Moreover, results from this study suggests that this epigenetic modification is elicited by and important for the maintenance of commensal microbes in the gut. In another study, when the human pathogen *Helicobacer pylori* infects the human gut, DNA methylation increases in the promoter regions of the human genes filamin C and thrombomodulin. This results in the silencing of these genes and a

concomitant increase in the risk of gastric cancer (Nakajima *et al.*, 2009). In another human microbiome study, an increase in abundance of two members of the human oral microbiome that belong to Enterobacteriaceae and Tenericutes is associated with the hypermethylation of the promoter regions of the human host gene *MDR1*. Hypermethylation of *MDR1* can result in head and neck squamous cell carcinoma (Bebek *et al*., 2012). In another study on host pathogens, pathogenic viruses including human adenovirus, hepatitis B virus and HIV are known to increase genome-wide levels of methylation of their host by up-regulating DNMT1 (Fang *et al.*, 2001; Burgers *et al*., 2007; Jung *et al*., 2007).

In addition to pathogenic and non-pathogenic gut microbes mediating human immune responses through DNA methylation, microbe-derived metabolites can also influence DNA methylation in humans and ultimately impact expressed phenotypes. For example, nutritional uptake in early postnatal humans modifies the infant's gut microbiome, which in turn affects the epigenetic patterns of the individual (Mischke and Plosch, 2013). This study proposes that changes in the composition of the gut microbiome results in altered profiles of microbe-produced metabolites such as folate and short-chain fatty acids. The same study proposed that an increase in such metabolites may influence the DNA methylation patterns of adjacent intestinal cells, which in turn results in the predisposition to obesity.

Microbial symbionts and pathogens of humans and some insects have been demonstrated to alter patterns of DNA methylation. As such microbial symbionts have the capacity to (radically) alter host phenotypes. I hypothesize that this ability is

widespread in insect symbionts, particularly among co-evolved insect symbionts. These intimate partners may influence methylation of their insect hosts with functional DNA methylation systems by modulating their host's immune responses to microbes. For example, attenuating immune responses so as to permit their intracellular persistence. In addition, these symbionts can encode novel biosynthetic pathways, which may contribute microbially derived metabolites, such as folate, which is a key source of the one carbon group used to methylated DNA. Moreover, in co-evolved insect symbioses, tissue-specific DNA methylation patterns in specialized insect cells that harbor obligate symbionts may facilitate the development and regulation of this long-term symbiotic relationship. Nevertheless, our understanding of the development and regulation of these symbiotic cells in insects is still nascent (Braendle *et al*., 2003; Hosokawa *et al*., 2016). Therefore, by investigating if and how epigenetic modifications affect the regulation of insect-microbe interactions, we will gain a better understanding of key biological mechanisms in symbiosis and evolution in general.

Table 1-1: DNA methylation in various insects and its phenotypic effects.

Species	Common Name	Phenotype	Reference
Acyrthosiphon pisum	Pea aphid	Color morph differentiation	Dombrovsky et al., 2009
Aedes aegypti	Mosquito	Wolbachia infection and gene transcription	Ye et al., 2013
Apis mellifera	Honeybee	Caste determination	Elango et al., 2009; Foret et al., 2012; Herb et al., 2012; Patalano et al., 2012
Apis mellifera	Honeybee	Learning and memory processing	Lockett et al., 2010; Biergans et al., 2012
Bombus terrestris	Bumblebee	Reproductive caste formation	Amarashinghe *et al*., 2014
Bombyx mori	Slikworm	Immune response against bacterial infection	Xiang et al., 2010; Zhang Q. et al., 2015
Componotus floridanus	Florida carpenter ant	Caste determination	Bonansio et al., 2012
Coptotermes formosanus	Subterranean termite	Gene regulation	Glastad et al., 2012
Lucusta migratoria	Migratory locust	Alternative migratory phenotypes	Robinson et al., 2015
Medauroidea extradentata	Stick insect	Gene regulation	Krauss et al., 2009
Myzus persicae	Peach-potato aphid	Overproduction of insecticide detoxifying esterases	Field et al., 1989; Hick et al., 1996; Field et al., 2003
Nasonia vitripennis	Jewel wasp	Photoperiodic response on diapose	Werren et al., 2010; Park et al., 2011; Pegoraro et al., 2016
Nasonia vitripennis	Jewel wasp	Embryo development	Zwier et al., 2012
Nilaparvata lugens	Brown planthopper	Female fecundity	Zhang J. et al., 2015
Onthophagus sp.	Horned Beetle	Nutritional plasticity	Snell-Rood et al., 2012
Pogonomyrmex barbatus	Red harvester ant	Caste determination	Smith et al., 2012
Reticulitermes flavipes	Subterranean termite	Gene regulation	Glastad et al., 2012
Schizaphis graminum	Greenbug aphid	Overproduction of insecticide detoxifying esterases	Ono et al., 1999
Sogatella furcifera	Rice planthopper	Sexual dimorphism	Zhang M. et al., 2015
Sogatella furcifera	Rice planthopper	Wing dimorphism	Zhou et al., 2013
Zootermopsis nevadensis	Dampwood termite	Caste differentiation	Terrapon et al. 2014

Chapter 2: Key Transport and Ammonia Recycling genes Involved in Aphid Symbiosis Respond to Host Plant Specialization

Abstract

Microbes are known to influence insect-plant interactions; however, it is unclear if host-plant diet influences the regulation of nutritional insect symbioses. The pea aphid, Acyrthosiphon pisum, requires its nutritional endosymbiont, *Buchnera*, for the production of essential amino acids. We hypothesize that key aphid genes that regulate the nutritional symbioses respond to host-plant diet when aphids feed on a specialized (alfalfa) compared to a universal host-plant diet (fava), which vary in amino acid profiles. Using RNA-Seq and whole genome bisulfite sequencing, we measured gene expression and DNA methylation profiles for such genes when aphids fed on either their specialized or universal host-plant diets. Our results reveal that when aphids feed on their specialized host-plant they significantly up-regulate and/or hypomethylate key aphid genes in bacteriocytes related to the amino acid metabolism, including glutamine synthetase in the GOGAT cycle that recycles ammonia into glutamine and the glutamine transporter ApGLNT1. Moreover, regardless of what host-plant aphids feed on we observed significant up-regulation and differential methylation of key genes involved in the amino acid metabolism and the glycine/serine metabolism, a metabolic program observed in proliferating cancer cells potentially to combat oxidative stress. Based on our results, we suggest that this regulatory response of key symbiosis genes in bacteriocytes allows aphids to feed on a suboptimal host-plant that they specialize on.

Introduction

When an organism symbiotically lives inside another organism's cells its cellular metabolic processes often become integrated with its hosts'. Archetypes of these ancient cellular integration events are readily observed in eukaryotic cells as mitochondria and plastids (Dyall *et al.,* 2004). Such organelles exhibit complex regulatory mechanisms that control all aspects of cellular processes such as cell division, transport, and metabolism. Similar to the regulation of organelles, simultaneous inter-domain crosstalk of animal host and bacteria exists between host cells and microbial endosymbionts (Zientz *et al.,* 2004). This crosstalk is essential to orchestrate the metabolic needs of both players in the symbiosis. One of the clearest examples of these metabolic integration events can be found within intracellular insect-microbe symbioses (Hansen and Moran, 2014). More than 10% of insect species possess long-term, mutualistic bacteria that provision nutrients to their insect host, and are housed inside of specialized host cells, referred to as bacteriocytes (Sudakaran *et al.,* 2017). Bacteriocytes are adapted to facilitate inter-domain molecular interactions; however, the mechanisms that the host cell uses to regulate, respond to, and control this integrated, symbiotic metabolism is still largely unexplored.

The mutualistic interaction between the pea aphid (*Acyrthosiphon pisum*) and its bacterial endosymbiont, *Buchnera aphidicola,* is one of the best-studied models on nutritional symbioses. In this symbiosis, amino acid pathways of both players are integrated together for the production of essential amino acids (Nakabachi *et al.,* 2005; Wilson *et al*., 2010; Hansen and Moran, 2011; Poliakov *et al*., 2011). This integrated

mixed-domain metabolism ultimately enables aphids to utilize nutrient deficient plant sap as food because like most animals, essential amino acid pathways are not encoded in the aphid's genome (International Aphid Genomics Consortium, 2010). For example, the aphid provides nonessential amino acid inputs to *Buchnera*'s essential amino acid pathways, and then *Buchnera* provides essential amino acids and vitamins to its host (Shigenobu *et al*., 2000; Nakabachi *et al.,* 2005; International Aphid Genomics Consortium, 2010). Several aphid genes, including genes that recycle ammonia into glutamate, complement *Buchnera*'s essential amino acid pathways, and transport non-essential amino acid inputs into bacteriocytes, are predicted to be key aphid genes involved in the regulation of this nutritional symbiosis (Nakabachi *et al*., 2005; Hansen and Moran, 2011; Poliakov *et al*., 2011; Price *et al*,. 2014).

Previous research on this system suggests that when dietary amino acid contents vary, aphids and *Buchnera* collectively adjust amino acid biosynthesis based on the aphid's nutritional requirements (Liadouze *et al.,* 1995; Febvay *et al.,* 1999). Moreover, when aphids feed on an artificial diet that varies only in nonessential amino acid profiles, bacteriocytes rebuild distinct profiles of amino acids that depend on the initial nonessential amino acid input(s) (Haribal and Jander, 2015). Together, these results suggest that bacteriocytes respond to amino acid variation in the aphid's diet. Host plants that pea aphids feed upon in the family *Fabaceae* vary dramatically in free amino acid profiles (Sandström and Pettersson, 1994). In turn, it is unclear how aphid bacteriocytes regulate their key symbiotic genes within the amino acid metabolism and in other cellular processes when a polyphagous pea aphid line feeds on their specialized host plant

compared to other host plants (Hansen and Moran, 2014), which vary in free amino acid profiles (Sandström and Pettersson, 1994).

Eukaryotic regulons are complex and are orchestrated through a combination of multiple mechanisms, including transcription factors, noncoding RNAs, and epigenetic factors. Among these different layers of gene regulation, the importance of epigenetic factors in influencing gene expression and alternative splicing has only recently begun to be elucidated (Luco *et al.,* 2011; Romanoski *et al.,* 2015). Previously, it has been shown that signals from the environment such as anxiety, stress, and diet can modify DNA methylation, which can subsequently alter gene expression profiles between different tissue types and throughout an organism's development (Feil and Fraga, 2012; Tammen *et al.,* 2013). For example, diet can modify DNA methylation patterns in a diversity of animals including insects, which in turn affects gene expression and subsequently influences organismal phenotypes (Niculescu and Zeisel, 2002; Kucharski *et al.*, 2008; Anderson *et al.*, 2012). The pea aphid is an ideal insect to observe methylation patterns in because it possesses an asexual, clonal, parthenogenetic life stage (Dixon, 1977) with a functional DNA methylation system (Walsh *et al.*, 2010; Pasquier *et al.*, 2014; Mukherjee and Baudach, 2016). Such variation at the epigenomic level but not DNA level within a clonal aphid population may be advantageous if it leads to transient phenotypes that are associated with dynamic ecological factors such as host plant nutrition. Currently it is unknown if the pea aphid has differential methylation in different tissue types and if host plant environment influences methylation patterns in bacteriocytes. Therefore, DNA

methylation may play an important role in the regulation of the aphid-*Buchnera* integrated metabolism, especially in response to different nutritional environments.

Here, using RNA-Seq and whole genome bisulfite sequencing we investigate if key aphid genes involved in the regulation of the aphid-*Buchnera* symbiosis are differentially expressed and methylated between a specialized and a universal host plant diet (i.e. a host plant diet all aphid biotypes can perform well on). If aphid bacteriocytes can alter these key symbiotic genes in response to feeding on different host plant environments, aphids can potentially optimize and/or compensate for specialized plant diets that are otherwise unsuitable in nutrient profiles.

Materials and Methods

Aphid lines and rearing

The LSR1 *A. pisum* strain was used for all aphid host plant trials in this study (International Aphid genomics Consortium, 2010). For host plant trials the LSR1 strain was divided into six independent sub-lines: three on fava (F1, F2, and F3) and three on alfalfa (A1, A2, and A3). Aphid sub-lines were reared at the same conditions as described in Hansen and Moran (2011). Before the start of all trials, three fava sub-lines were reared on the same developmental stage of fava bean (F1, F2, F3; fava bean= 23 ± 2 days after germination (~ 5 whorls)) for over 10 generations, and three alfalfa sub-lines were reared on the same developmental stage of alfalfa (A1, A2, A3; alfalfa= 44 ± 2 days after germination) for over 10 generations. These plant developmental stages were chosen because they correspond with plant developmental ages used previously in pea aphid free

amino acid stylectomy trials, where amino acid profiles varied significantly between fava bean and alfalfa sap (Sandström and Pettersson, 1994).

Aphid mass assessment

Weights of asexual adults (1 day after final instar molt) were measured using a CAHN 29 automatic electrobalance (Cahn, Cerritos, CA) for 20 individuals per sub-line (F1, F2, F3, A1, A2, A3). For statistical analyses, General Linear Models (GLM) was used treating aphid mass as a dependent variable and host plant treatment as a fixed factor with insect line nested within host plant treatment using IBM SPSS Statistics version 23 (IBM SPSS, Armonk, NY). Tukey's multiple comparison post-hoc tests were used to determine aphid mass differences between sub-lines.

Determination of *Buchnera* cell abundance using RT-qPCR

After weight measurements (see above) parthenogenic female adults were preserved in 95% ETOH for *Buchnera* cell abundance measurements using Real Time quantitative PCR (RT-qPCR). Also, several offspring (1^{st} day nymphs) from these adults, reared on plants until the first day of reproduction (pre-reproductive phase of alfalfa aphids = ~10-11 days and 9-10 days for fava aphids), were collected from each sub-line and preserved for *Buchnera* cell abundance measurements. DNA of individual aphids were extracted with the Qiagen DNeasy DNA extraction kit following the manufacture's protocol (Valencia, CA, USA); six individuals per life stage per sub-line were extracted; N=72 individuals total). RT-qPCR was conducted on each aphid individual using the Eppendorf Mastercycler epgradient realplex^2 (Hamburg, Germany). For RT-qPCR reactions KAPA SYBR FAST universal qPCR kit (Wilmington, MA) was used for

concentrations and cycle conditions (2-step qPCR). A standard curve was created for each reaction plate by diluting target gene plasmids for both a single copy gene in *Buchnera* (ATP synthase F0F1 subunit C) and the pea aphid (*Acyrthosiphon pisum* 60 kDa heat shock protein (LOC100168563)). Primer sequences for these *Buchnera* and aphid genes respectively are: BAp_atpE_F1- 5' -CCG CTA GGC AAC CTG ATT TA-3'; BAp_atpE_F1- 5' -CAA TCA TTG GAA TCG CAT CA-3'; and Aphid HS60 F1- 5'- GCC AAG AAG GTA ATG AAC TG-3'; Aphid HS60 R1- 5'- TCA ACA GCA AAG TGT CAT C-3'. The standard curve method for relative quantification (Bookout *et al*., 2006) was used to compute the normalized *Buchnera* expression value for each aphid sample (i.e. normalized to the aphid gene HS60). For statistical analyses, General Linear Models (GLM) was used treating the normalized *Buchnera* expression value as the dependent variable and host plant treatment and life stage (1st instar nymph and adult) as fixed factors, with insect line nested within host plant treatment using IBM SPSS Statistics version 23 (IBM SPSS, Armonk, NY).

TEM imaging of *Buchnera* cells

For transmission electron microscopy (TEM) images, aphids on the last nymphal stage (4th instar) were dissected in buffer and fixed for 2 hours. The tissue was dehydrated in ethanol for 12 hours. Subsequently, bacteriocytes for each host plant treatment were observed using a TEM microscope. For each TEM image, the number of *Buchnera* cells were counted within a unit area (100 μm^2) using Cell Counter plugin of ImageJ version 1.8.0 (Abràmoff *et al*., 2004). The means of each bacteriocyte image (N=7 for fava; N=5 for alfalfa) were compared using a two-sample t-test.

Determination of bacteriocyte abundance

For both host plant treatments, at least 12 aphids from each sub-line were dissected and bacteriocytes were counted using a light microscope (N=45 and 46 aphid individuals for fava and alfalfa treatments, respectively). For statistical analyses GLM was used treating bacteriocyte count as the dependent variable and host plant treatment as a fixed factor, with insect line nested within host plant treatment using IBM SPSS Statistics version 23 (IBM SPSS, Armonk, NY).

RNA-Seq analysis

For RNA-Seq trials, the same six sub-lines (F1, F2, F3, A1, A2, A3) were analyzed. Aphid bacteriocyte and body tissue dissections were conducted as in Hansen and Moran (2011) during the aphid's fourth instar, giving four samples, bacteriocytes of alfalfa feeding aphids (ABAC), body cells of alfalfa feeding aphids (ABODY), bacteriocytes of fava feeding aphids (FBAC), and body cells of fava feeding aphids (FBODY), each with three biological samples. For dissections, one aphid sub-line was dissected at a time for a two-week duration to pool enough material for RNA extractions. All sub-lines were randomly dissected one after the other over a continuous time block of three months, which allowed us to pool approximately 100 aphids to get enough RNA concentration. For each aphid, bacteriocytes were dissected out, and body cells were prepared from the remainder of the dissections by removing any bacteriocytes and embryos. RNA extractions were conducted similar to Hansen and Moran (2011), except RNA > 200 bp was retained for sequencing using Illumina HiSeq 2500 (Illumina, San Diego, CA).

Illumina library preparation and sequencing were conducted by Yale University's Keck Genome Sequencing Center. Libraries were sequenced as paired-end 76-mers using Illumina's pipeline. Three samples were sequenced per lane. Reads for all RNA-Seq samples (12 total: 6 bacteriocyte and 6 corresponding body tissue samples) were submitted to the Gene Expression Omnibus (GEO) database of the National Center for Biotechnology Information (NCBI) (accession no. PRJNA213008). RNA-Seq paired-end data were mapped to *A. pisum*'s genome, version 2.0 (aps_ref_Acyr _2.0_chrUn), with 16, 919 RefSeq genes using HISAT2 v2.1.0 (Pertea *et al.*, 2016). Aligned reads were assembled and quantified using StringTie v1.3.4 (Pertea *et al.*, 2015) following developer's protocol. Raw transcript read counts for each sample were normalized and analyzed using DESeq2 v1.19.38 (Love *et al.*, 2014). Differentially expressed genes between bacteriocytes of both host plant samples were identified using likelihood ratio test based on generalized linear model (~host plant type + tissue type vs. ~host plant type) in DESeq2 (Love *et al.*, 2014). Statistical significance was determined if FDR adjusted p-values were ≤ 0.05. To identify the genes that are differentially expressed in bacteriocytes of different host plants likelihood ratio test was performed between alfalfa bacteriocytes and fava bacteriocytes (ABAC vs. FBAC) based on read counts. Statistical significance was determined if the adjusted p-value ≤ 0.05.

Aphid amino acid pathways and putative enzyme functions were analyzed using the *A. pisum* genome and annotations in the NCBI, Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), BRENDA Enzyme Information System (Chang *et al.*, 2009), EcoCyc (Keseler *et al*., 2009), and AcypiCyc (International Aphid

Genomics Consortium, 2010) databases. Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.*, 2005) was used to determine which KEGG pathways were differentially regulated at the normalized $p < 0.1$ and $p < 0.05$, as described in Hansen and Degnan (2014).

DNA isolation and whole genome bisulfite sequencing

For whole genome bisulfite sequencing trials, the same six sub-lines (F1, F2, F3, A1, A2, A3) were prepared as described above for RNA-Seq trials. Dissections were carried out similar to RNA-Seq trials above, except DNA, instead of RNA, was isolated and extracted with the Master Pure Kit (Epicentre Technologies, Madison, WI) following manufacture guidelines. After DNA was extracted, DNA concentrations of each sample were quantified by Qubit dsDNA HS kit (Invitrogen, Carlsbad, CA). 200 ng of DNA was used for bisulfite conversion with the EZ DNA Methylation-Lightning™ Kit (Zymo Research, Irvine, CA). Genomic DNA was spiked with 0.5 ng of λ phage DNA (New England BioLabs, Ipswich, MA) as a control since it is unmethylated at CpG sites to calculate the rate of false positives from the sodium bisulfite conversion treatment. The reads that mapped to the λ phage were analyzed to control for the background conversion rates. The bisulfite libraries were constructed using the EpiGnome™ Methyl-Seq Kit (Epicentre Technologies, Madison, WI) with the starting material of 50 ng of bisulfite-treated DNA following the manufacturer's guidelines. Strand-specific sequencing was conducted on HiSeq 2500 (Illumina, San Diego, CA) using a TruSeq SBS sequencing kit (Illumina). Reads for all DNA samples (12 total: six bacteriocyte and six corresponding

body tissue samples) were submitted to the Sequence Read Archive of the National Center for Biotechnology Information (NCBI) (accession no. PRJNA339317).

Methylation data analysis

Raw read data from whole genome bisulfite sequencing were trimmed to remove Illumina index sequences using Trimmomatic (Bolger *et al.*, 2014). Methylation read data were aligned to *A. pisum*'s genome, version 2.0 with Bismark as suggested in (Krueger and Andrews, 2011). The methylation level at each CpG site with 10 reads or greater was determined by the number of Cs at a given site in the mapped reads divided by the total number of reads. To standardize the sample size of CpG sites per sample, only shared CpG sites that were ≥ 10 reads per CpG site for all 6 biological replicates were evaluated in this analysis. After standardization, percent methylation (methylated versus unmethylated reads) per sample was calculated by averaging every site-specific methylation percentage within a sample. The paired t-test was used to determine if the percent methylation of CpG sites were significantly different between body cell and bacteriocyte samples (BODY vs. BAC) and between bacteriocytes of alfalfa and fava feeding aphids (ABAC vs. FBAC). To compare how similar CpG methylation profiles are among samples, PCA analysis was conducted with site-specific CpG methylation data for each sample using the methylKit package in R (Akalin *et al.*, 2012). To test the within-group dispersion of CpG methylation profiles for each group, multi-response permutation procedure was used as suggested in (Mielke and Berry, 2003) using vegan package in R (Oksanen *et al.*, 2017). Four groups were defined as bacteriocytes of alfalfa-treated aphids (ABAC), body cells of alfalfa-treated aphids (ABODY),

bacteriocytes of fava-treated aphids (FBAC), and body cells of fava-treated aphids (FBODY). The average within-group distances were calculated using Sorensen distances.

To determine if the percent methylation of CpG sites was significantly different between the different genic regions (exon, intron, and intergenic regions) average percent methylation was compared between each genic region per sample using a paired t-test. To identifydetermine if a specific shared CpG site for a tissue sample (bacteriocyte versus body) is significantly higher (hyper-methylated) or lower (hypo-methylated) in percent methylation between tissue samples (i.e. differential methylation) a two-group comparison was conducted for each host plant treatment (fava and alfalfa), (N=3 aphid sub-lines for each host plant treatment) using the Fisher's exact test within the methylKit statistical package in R (Akalin *et al*., 2012). Methylation levels at a specific cytosine site were determined to be significantly different between bacteriocyte and body tissues if there was ≥ 25 % difference in percent methylation between treatments and the FDR adjusted p-value was $q \leq 0.01$.

To link DNA methylation profiles in aphid tissue treatments to differential gene expression patterns, we combined RNA-Seq data collected from the corresponding aphid tissue samples with the methylation data (N=6 biological replicates) (see methods above). To locate the genes that were both differentially methylated and differentially expressed and/or spliced the TopHat and the CuffDiff2 v2.1.1 pipelines (Trapnell *et al*., 2012) were used. Spliced genes were identified as significantly expressed between bacteriocyte and body tissues for each host plant treatment (N=3) using the FDR adjusted p-values ($q \leq 0.05$), based on the FPKM values generated by CuffDiff2 v2.1.1 (Trapnell *et al*., 2012).

Statistical comparison of FPKM values was conducted as described in Trapnell *et al*. (2012) using Cuffdiff2 v2.1.1, with a false discovery rate criterion for calculating p-values. Statistical significance was determined if $q \leq 0.05$ and when a greater than two-fold change in expression of the bacteriocyte compared to the body occurred in each host plant treatment group. Differentially spliced genes were determined based on Jensen-Shannon divergences of splicing isoforms that were included in Cuffdiff v2.1.1 package (Trapnell *et al*., 2012). Statistical comparison of splicing isoforms was conducted using Cuffdiff2 v2.1.1 (Trapnell *et al*., 2012). Statistical significance was determined if the *p*-values were greater than the false discovery rate (FDR) after the Benjamini-Hochberg correction for multiple testing, as described in Trapnell et al. (2012). Genes that were both differentially methylated and differentially expressed and/or spliced were characterized similar to above for RNA-Seq data using GSEA software *(*Subramanian *et al*., 2005). GSEA (Subramanian *et al*., 2005) was used to determine which KEGG pathways (Kanehisa and Goto, 2000) and gene ontology (GO) terms were both differentially regulated and methylated at the normalized $P \leq 0.05$. Only groups with two or more genes within a KEGG pathway group were analyzed. Significant GSEA groups (GO terms and KEGG pathways) were further analyzed using the *A. pisum* genome and annotations in the NCBI, KEGG (Kanehisa and Goto, 2000), BRENDA Enzyme Information System (Chang *et al*., 2009), and AcypiCyc (International Aphid Genomics Consortium, 2010) databases.

Results

In this study, the pea aphid strain (LSR1), which originated as an alfalfa (*Medicago sativa*) specialist in the field (International Aphid Genomics Consortium, 2010), was used for all trials. Here, the LSR1 strain was divided into six independent sub-lines for all host plant trials; three sub-lines fed on its specialized host plant, alfalfa, and the remaining three sub-lines fed on its 'universal' host plant, fava. The universal host plant fava was chosen because previous studies have indicated that most pea aphid biotypes favor and display higher fitness on their 'universal' host plant fava (*Vicia faba*) compared to the host plant they specialize on in the field (Ferrari *et al.*, 2008; Ferrari *et al.*, 2012; Peccoud *et al.*, 2014). During all trials host plants were of a particular developmental stage where amino acid profiles in sap vary significantly in alfalfa, compared to fava (Sandström and Pettersson, 1994).

Effects of host plant diet on aphid and *Buchnera* phenotype

To investigate if pea aphid LSR1 fitness was significantly greater in aphids feeding on fava compared to their specialist host plant alfalfa we measured adult aphid mass, a surrogate for aphid fitness (Vogel and Moran, 2011). Results indicated that all aphid sub-lines feeding on fava were of similar mass to one another, but aphid mass was significantly greater in aphid sub-lines feeding on fava compared to alfalfa based on Tukey's post-hoc tests ($\chi^2 = 567.017$, d.f. = 1, $p<0.0005$, Fig. 2-1A). A significant aphid line effect was found for host plant treatment ($\chi^2 = 48.005$, d.f. = 1, $p<0.0005$). One aphid sub-line feeding on alfalfa (A2) was significantly greater in mass compared to the other aphid sub-lines (Fig. 2-1A).

Nutritional endosymbionts, such as *Buchnera,* can be regulated in insects at the bacteriocyte and/or symbiont titer level between different insect life stages and morphs (Humphreys and Douglas, 1997; Mira and Moran, 2002; Kono *et al.,* 2008; Nishikori *et al.,* 2009; Stoll *et al.,* 2010; Vigneron *et al.,* 2014; Parkinson *et al.,* 2016; Simonet *et al.,* 2016). To determine if host plant diet affects the number of aphid bacteriocytes and/or *Buchnera* titer we counted bacteriocytes and *Buchnera* cells and then compared them between host plant treatments. The numbers of bacteriocytes in 4th instar aphids were not significantly different between host plant treatments or aphid lines (χ^2=3.522, d.f.=1, p=0.061; χ^2=2.065, d.f.=4, p=0.724, respectively). Average bacteriocyte number per one aphid individual was 68 (95% C.I. 63.8-71.1, N=45) and 63 (95% C.I. 58.9-66.2, N=46) for the fava and alfalfa treatments, respectively. Moreover, *Buchnera* abundance did not significantly differ between host plant treatments based on RT-qPCR (χ^2=2.674, d.f.=1, p=0.102, Fig. 2-1B). Normalized *Buchnera* abundance was higher in first day nymphs compared to first day adults regardless of host plant treatment (χ^2=55.486, d.f.=1, p<0.0005, Fig. 2-1B). Consistent with these findings, the number of *Buchnera* cells within bacteriocytes did not differ significantly between aphids from representative host plant treatments based on TEM images (T=1.275, d.f.=9, p=0.234). Within a unit area (100 μm^2), an average of 21 (N=7) and 24 (N=5) *Buchnera* cells were identified within bacteriocytes of 4th instar fava and alfalfa feeding aphids, respectively (Fig. 2-1C).

Effects of host plant diet on the expression of key aphid symbiosis genes

To investigate if host plant diet affects the expression of key symbiotic genes of aphids at the mRNA level in bacteriocytes, we first conducted RNA-Seq on bacteriocytes

and other aphid body tissues (BAC vs. BODY) (Table 2-1). For the BAC vs. BODY comparison we identified 1,904 genes that were significantly up-regulated between bacteriocytes and body cells for both host plant treatments (Supplemental Dataset S1, Table 2-2). We further identified 4,211 genes that were significantly down-regulated between bacteriocytes and body cells for both host plant treatments (Supplemental Dataset S2, Table 2-2). To determine host plant differences between bacteriocytes (ABAC vs. FBAC) we identified 54 genes that were up-regulated and 101 genes that were down-regulated in bacteriocytes of alfalfa feeding aphids compared to bacteriocytes of fava feeding aphids (Supplemental Dataset S3, Table 2-2).

To characterize the functions of genes differentially expressed in bacteriocytes for each comparison we used GSEA (Subramanian *et al.*, 2005). For the BAC vs. BODY comparison we found 10 KEGG pathways significantly enriched in bacteriocytes compared to body cells for both host plant treatments (Table 2-3a). The top five KEGG pathways in descending order based on the GSEA enrichment score were glycine, serine and threonine metabolism, glyoxylate and dicarboxylate metabolism, phenylalanine metabolism, pentose phosphate pathway, and the nicotinate and nicotinamide metabolism (Table 2-3a). We also identified five KEGG pathways that were significantly down-regulated in bacteriocytes compared to body cells for both host plant treatments (Table 2-3a). Such KEGG pathways in descending order based on the GSEA enrichment score were hippo signaling pathway, notch signaling pathway, other glycan degradation, phototransduction, and neuroactive ligand-receptor interaction.

To identify the pathways that were differentially expressed in bacteriocytes between host plant treatments (ABAC vs. FBAC) we performed GSEA and identified seven KEGG pathways that were significantly enriched in bacteriocytes of alfalfa compared to fava treated aphids. The seven KEGG pathways in descending order based on the GSEA enrichment score were synthesis and degradation of ketone bodies, vitamin B6 metabolism, aminoacyl-tRNA biosynthesis, pyruvate metabolism, Jak-STAT signaling pathway, lipoic acid metabolism, and the butanoate metabolism (Table 2-3b). Aphid and *Buchnera* metabolisms are integrated for the production of amino acids within bacteriocytes. This shared amino acid metabolism is hypothesized to be regulated by the aphid host via transporters, the GS/GOGAT cycle, and genes that complement *Buchnera*'s essential amino acid pathways (Nakabachi *et al.*, 2005; Hansen and Moran, 2011; Poliakov *et al.*, 2011; Price *et al.*, 2014). We examined these aphid genes and found that 22 out of 27 genes were significantly enriched in bacteriocytes relative to body cells for both host plant treatments (BAC vs. BODY; Fig. 2-2). We further identified that two out of eight of these genes (Glutamine synthetase and *ApGLNT1*) were significantly enriched in bacteriocytes of alfalfa feeding aphids when compared to bacteriocytes of fava feeding aphids (ABAC vs. FBAC; Fig. 2-2; Supplemental Dataset S4). Glutamine synthetase (*GS*) is a key enzyme of the GS/GOGAT cycle and recycles ammonia into glutamine (Hansen and Moran, 2011). The transporter *ApGLNT1* imports glutamine into bacteriocytes (Price et al., 2014). Collectively these results suggest that aphid genes that synthesize and transport glutamine, an important amino donor for *Buchnera*'s essential

amino acid pathways, is enriched in bacteriocytes of aphids feeding on their specialized host plant, alfalfa, compared to their universal host plant, fava (Fig. 2-2).

Another major subset of genes that were enriched significantly in bacteriocytes compared to other body cells for both host plant treatments (BAC vs. BODY) belonged to the glycine/serine metabolism (Fig. 2-3, Supplemental Dataset S4). Specifically, the genes for serine biosynthesis (D-3-phosphoglycerate dehydrogenase; *PHGDP*, phosphoserine aminotransferase 1; *PSAT1*, phosphoserine phosphatase; *PSPH*) were up-regulated significantly in bacteriocytes compared to body cells. Also, serine hydroxymethyltransferase (*SHMT*), which converts serine to glycine, and the bifunctional purine biosynthesis protein (*PURH*) were up-regulated significantly higher in bacteriocytes compared to body cells. The glycine/serine metabolism also relies on the maintenance of the cofactor tetrahydrofolate (*THF*). In the bacteriocytes of both treatments *THF* was maintained by the significant up-regulation of genes in the one carbon pool by folate metabolism, and the production of 5,10-methenyl-*THF* through the Glycine Cleavage System (Fig. 2-3, Supplemental Dataset S4). Among such genes, *PHGDP* and *PURH* were significantly higher in bacteriocytes of alfalfa compared to fava feeding aphids (ABAC vs. FBAC; Fig. 2-3).

For both host plant treatments (BAC vs. BODY), gluconeogenesis instead of glycolysis appears to be occurring in bacteriocytes compared to body cells as indicated by the up-regulation of malate dehydrogenase (*MDH*) and phosphoenolpyruvate carboxykinase (*PEPCK*) (Fig. 3). In turn, instead of glucose, alternative energy substrates such as extracellular and *Buchnera* derived pyruvate may provide the carbon backbone to

fuel the glycine/serine metabolism. For example, pyruvate transporters were significantly up-regulated in bacteriocyte cells compared to body cells (monocarboxylate transporter; *MCT* and mitochondrial pyruvate carrier; *MPC*) (Fig. 2-3).

Another key pathway that is important in the aphid-*Buchnera* symbiosis involves the production of uracil. *Buchnera* is unable to produce its own uracil (Shigenobu *et al.*, 2000) and therefore it depends on the host for uracil biosynthesis in this integrated metabolism. The uracil salvage pathway, especially pseudouridine kinase, was found to be significantly enriched in bacteriocytes of aphids feeding on alfalfa compared to fava (ABAC vs. FBAC; Supplemental Dataset S4).

Effects of host plant diet on DNA methylation profiles

Diet cues can alter DNA methylation patterns within and adjacent to invertebrate genes (Kucharski *et al.*, 2008). Moreover, these DNA methylation marks have been associated with active genes and alternative splicing (Yan *et al.*, 2015). To determine if host plant diet affects the methylation of key symbiotic genes in aphid bacteriocytes we used whole-genome bisulfite sequencing (Table 2-4). The average percentages of CpG methylation were 3.4% and 4.3% for bacteriocytes and body cells, respectively (Fig. 2-4A). Bacteriocytes had a significantly lower percent of CpG methylation compared to body cells (BAC vs. BODY; paired t-test, $t=13.47$, $df=5$, $p<0.0001$) (Fig. 2-4A). The average percentages of CpG methylation were 3.8% and 3.0% for bacteriocytes of alfalfa compared to fava feeding aphids, respectively (Fig. 2-4B). Bacteriocytes of alfalfa feeding aphids had a significantly higher percent of CpG methylation compared to

bacteriocytes of fava feeding aphids (ABAC vs. FBAC; t=5.18, df=4, p=0.0066) (Fig. 2-4B).

To investigate if there was a difference in percent methylation within and outside genic regions percent methylation within the exon, intron, and the intergenic regions were determined. For both host plant diets there was not a significant difference in percent methylation between bacteriocyte and body cells for the exon, intron, or the intergenic regions (BAC vs. BODY; paired t-test, t=1.76, df=2, p=0.22) (Fig. 2-4C). In contrast, when comparing between host plant diets percent methylation in the exon region was significantly higher in bacteriocytes of alfalfa feeding aphids compared to bacteriocytes of fava feeding aphids (ABAC vs. FBAC; paired t-test, t=10.34, df=4, p=0.0005) (Fig. 2-4D). The percent of methylation in the intron and intergenic regions were not significantly different in bacteriocytes of alfalfa feeding aphids compared to bacteriocytes of fava feeding aphids (ABAC vs. FBAC; paired t-test, t=2.40, df=2, p=0.14 for introns; t=0.57, df=2, p=0.63 for intergenic) (Fig. 2-4D).

For all bacteriocyte and body tissue samples percent methylation within the exon regions was significantly higher compared to the intron regions (paired t-test, t=12.05, df=11, p<0.001) (Fig. 2-4C). Within bacteriocytes of both alfalfa feeding aphids and fava feeding aphids, percent methylation levels of exon regions were significantly higher than those of intron regions (paired t-test, t=8.50, df=5, p<0.001) (Fig. 2-4D). Also, percent methylation was significantly higher in the intron regions compared to the intergenic regions (for all samples; paired t-test, t=19.03, df=11, p<0.001) (Fig. 2-4C) (within bacteriocytes; t=17.02; df=5; p<0.001) (Fig. 2-4D).

In order to visually determine how shared CpG sites differ between all samples in percent methylation when aphids feed on their specialized compared to universal host plant diet we conducted a Principal Component Analysis (Fig. 2-5). To test if methylated CpG profiles were significantly different between bacteriocyte, body cell, and host plant treatments we used multi-response permutation procedure (MRPP) (Mielke and Berry, 2003). We found that the four *a priori* groups: bacteriocytes of alfalfa feeding aphids (ABAC), body cells of alfalfa feeding aphids (ABODY), bacteriocytes of fava feeding aphids (FBAC), and body cells of fava feeding aphids (FBODY) were significantly different from one another in CpG methylation profiles ($p < 0.001$; A=0.09356). In the PCA ordination, all body samples clustered tightly together in ordination space away from bacteriocyte samples. Using MRPP, we found that methylated CpG profiles of bacteriocyte samples were significantly different compared to body cell samples (BAC vs. BODY; $p < 0.005$; A=0.08245), with a significantly higher dispersion of within-group differences for bacteriocyte samples (delta=0.2759) compared to body cell samples (delta=0.1772). In accord to the PCA ordination MRPP results indicate that bacteriocyte cells have significantly different methylation profiles compared to body cells, which have more similar distributions to one another, regardless of host plant treatment (Fig. 2-5). Nevertheless, body cells are a mixture of different aphid cell types and therefore we cannot exclude the possibility that more abundant host cell types mask host plant differences of less abundant cell types. In contrast, among bacteriocyte samples (ABAC vs. FBAC), fava samples are more heterogeneous in CpG profiles compared to alfalfa samples, with a higher dispersion of within-group differences for FBAC (delta=0.2810)

compared to ABAC (delta=0.2645) (Fig. 2-5). These results indicate that methylation profiles are specific to aphid cell type and host plant diet, especially for bacteriocytes in the aphid's specialized host plant treatment, alfalfa.

A total of 3,474 CpG sites were significantly differentially methylated between bacteriocytes and body cells of both host plant treatments (BAC vs. BODY) (Table 2-5a). Between bacteriocytes of alfalfa feeding aphids compared to fava feeding aphids a total of 294 CpG sites were differentially methylated significantly (ABAC vs. FBAC). For both comparisons, differential CpG methylation was primarily confined to the gene body regions (82% for BAC vs. BODY; 78% for ABAC vs. FBAC) (Table 2-5b) as revealed in other non-mammal animals, which may contribute to gene activation and/or alternative splicing (Hunt *et al.*, 2013).

To link the patterns of differential DNA methylation with differential gene expression, we identified 441 genes that were both differentially methylated and differentially expressed significantly between bacteriocyte and body samples for both alfalfa and fava feeding aphids (BAC vs. BODY) (Table 2-5). All 441 genes were up-regulated and hypomethylated in bacteriocytes relative to body cells. We also identified 702 genes that are both differentially methylated and differentially spliced between bacteriocyte and body samples of both alfalfa and fava feeding aphids (BAC vs. BODY) (Table 2-5; Supplemental Dataset S4). Furthermore, we identified three genes that were both differentially methylated and differentially expressed significantly between bacteriocytes of alfalfa compared to fava feeding aphids (ABAC vs. FBAC; Table 2-5). All three genes were down-regulated and hyper-methylated in alfalfa compared to fava

feeding aphid bacteriocytes. These genes were the serine/arginine repetitive matrix 1 gene (LOC100160294), the proton-coupled amino acid transporter 4-like (LOC100159667), and the broad-complex core protein isoforms 1/2/3/4/5-like gene (LOC100167015). We also identified three genes that were both differentially methylated and differentially spliced between bacteriocyte samples from alfalfa and fava feeding aphids (ABAC vs. FBAC; Table 2-5): anoctamin-1-like (LOC100167803), formin-binding protein 1-like (LOC100166693), and tumor protein D54-like (LOC100164449). Using GSEA (Subramanian *et al.,* 2005), we found five pathways that were both significantly differentially methylated (hypo-methylated) and up-regulated in bacteriocytes compared to body cells. No pathways were differentially methylated and significantly down-regulated in bacteriocytes compared to body cells (BAC vs. BODY; Table 2-3c). These five pathways were the metabolic pathways, lysosome, protein processing in endoplasmic reticulum, selenocompound metabolism, and tryptophan metabolism (Table 2-3c). Genes that were both differentially methylated and differentially spliced between bacteriocyte and body samples (BAC vs. BODY) belong to 16 KEGG pathways (Table 2-6). These pathways include the key glutamine transporter *ApGLNT1,* the bulk movement into cells and digestion (15 genes), the degradation, processing, and transport of RNAs (47 genes), four different signaling pathways including one involved in the immune response (Jak-Stat) (37 genes), protein processing and degradation (20 genes), and the biosynthesis of amino acids (6 genes) (Fig. 2-2).

Fourteen key aphid genes associated with the integrated aphid-*Buchnera* symbiosis and the glycine/serine metabolism were both differentially expressed and

differentially methylated between the bacteriocytes and body cells (BAC vs. BODY) (Figs. 2-2, 2-3; Supplemental Dataset S4). For example, glutamate synthase, an important enzyme of the GS/GOGAT cycle that converts glutamine to glutamate was both hypo-methylated, differentially spliced, and up-regulated in bacteriocytes of both fava and alfalfa feeding aphids compared to body cells (Fig. 2-3). In addition, an active glutamine transporter (*ApGLNT1*) that was previously characterized to be important for the regulation of *Buchnera*'s essential amino acid biosynthesis pathways (Price *et al.,* 2014) was significantly hypo-methylated and up-regulated in bacteriocytes compared to body cells (Fig. 2-2). Also, two enzymes (*PURH* and cytoplasmic C-1-tetrahydrofolate synthase) in the one carbon pool by folate pathway were hypo-methylated and up-regulated in the bacteriocytes compared to the body cells (Fig. 2-3).

DISCUSSION

In this study, we demonstrated for the first time that key aphid genes involved in the regulation of the aphid-*Buchnera* symbiosis are differentially expressed, spliced, and methylated when aphids feed on a specialized compared to a universal host plant diet. Our data indicate that this regulatory and epigenetic response to distinct host plant types that vary in amino acid profiles may play a significant role in modulating the aphid-*Buchnera* amino acid metabolism when aphids feed on their specialized compared to universal host plant diet. Moreover, this regulatory response in combination with lower aphid fitness when aphids feed on their specialized compared to universal host plant diet is consistent with the pea aphid engaging in a compensatory metabolic response when it

specializes on a less suitable host plant. We also identified key aphid genes and pathways involved in the aphid-*Buchnera* symbiosis that are differentially expressed, spliced, and methylated in both host plant diets. These results collectively suggest that DNA methylation may play both a conserved (maintenance methylation) and an environmentally induced (*de novo* methylation) regulatory role in bacteriocytes when aphids feed on host plant diets that vary in amino acid profiles.

Here we reveal that instead of regulating bacteriocyte or *Buchnera* cell number the aphid-*Buchnera* integrative metabolism modulates patterns of bacteriocyte DNA methylation and gene expression in response to its specialized host plant diet, alfalfa, when compared to its universal host plant diet, fava. Results from our study that provide evidence for this finding include the following: ***1)*** One aphid enzyme (Glutamine synthetase; *GS*) in the GS/GOGAT cycle was enriched in bacteriocytes of alfalfa feeding aphids compared to fava feeding aphids (Fig. 2-2). The GS/GOGAT cycle is hypothesized to play a key role with *Buchnera* in sustaining aphids on a nitrogen-limited diet, because *GS* recycles waste ammonia for the production of glutamine (Hansen and Moran, 2011). ***2)*** The transporter *ApGLNT1* is significantly up-regulated in the bacteriocytes of alfalfa feeding aphids compared to fava feeding aphids (Fig. 2-2). Interestingly this transporter was significantly hypo-methylated only in bacteriocytes compared to body cells of both alfalfa and fava feeding aphids. This transporter imports glutamine into bacteriocytes and is inhibited by arginine produced by *Buchnera*. In turn, this transporter may play a key role in regulating *Buchnera*'s essential amino acid metabolism (Price *et al.,* 2014), by promoting essential amino acid biosynthesis in alfalfa

feeding aphids***. 3)*** The vitamin B6 pathway was significantly enriched in bacteriocytes of alfalfa feeding aphids compared to fava feeding aphids (Table 2-3b). Vitamin B6 is an essential cofactor in animals and microbes and plays an important role in the amino acid and carbohydrate metabolism and singlet oxygen resistance (John, 1995; Daub and Ehrenshaft, 2000). Collectively, both the aphid and *Buchnera* do not encode the entire vitamin B6 biosynthesis I on II pathways, however *Buchnera* still encodes serC and thrC and the aphid encodes the enzymes 2.6.1.52, 2.7.1.35, and 1.4.3.5., which make up the majority of the pathway. Both enzymes 2.6.1.52 and 2.7.1.35 are up-regulated significantly in bacteriocytes compared to body tissues 15X and 8X respectively suggesting that there is a demand for vitamin B6 biosynthesis in aphid bacteriocytes. ***4)*** The uracil salvage pathway was found to be significantly enriched in bacteriocytes of aphids feeding on alfalfa compared to fava. These results suggest that more uracil potentially for both *Buchnera* and/or aphid mRNA biosynthesis is needed for the maintenance of alfalfa compared to fava bacteriocytes. ***5)*** Also six genes were both differentially expressed/spliced and methylated between bacteriocytes of alfalfa compared to fava bacteriocytes and are involved in amino acid transport**,** protein kinase activity, calcium activated chloride channel activity, gene regulation via epigenomic interactions utilizing the POZ zinc finger domain, and unknown function. More information on how these genes help regulated the integrated aphid-*Buchnera* metabolism when feeding on different host plants is needed. Collectively, these results suggest that when aphids feed on their suboptimal, specialized host plant, alfalfa, key

aphid genes involved in the regulation of the integrative amino acid metabolism are enriched in bacteriocytes.

The observed aphid regulatory changes are indicative of a compensatory response of aphids feeding on alfalfa that require more essential amino acids compared to aphids on fava. For example, previous amino acid concentration data shows a limitation of essential amino acids in alfalfa sap compared to fava sap (Sandström and Pettersson, 1994). Specifically, Sandström and Pettersson (1994) revealed that total amino acid concentrations in sap obtained from aphid stylets are relatively similar between alfalfa and fava, however concentrations of nine essential amino acids (arginine, isoleucine, leucine, lysine, phenylalanine, tryptophan, tyrosine, valine, and histidine) are lower in alfalfa sap compared to fava sap. Our data coincides with these amino acid concentration data (Sandström and Pettersson, 1994), because we revealed an up-regulation of key aphid genes and pathways that are important in provisioning amino donors, co-factors, and energy to fuel the integrative amino acid metabolism with *Buchnera* for the production of essential amino acids.

In this study, we also observed that regardless of host plant treatment 1,143 genes were differentially expressed/spliced and methylated in aphid bacteriocytes compared to body cells. These genes included glutamine synthase, the other gene involved in the GS/GOGAT cycle (Fig. 2-2), and several pathways involved in cell signaling, immune function, and the regulation of RNA and protein biosynthesis and degradation. One of the most distinctive and highly enriched pathways identified within bacteriocytes compared to body cells was the glycine/serine metabolic profile (Fig. 2-3). Interestingly,

to the best of our knowledge this metabolic profile has only been identified previously as a metabolic hallmark in cancer cells (Perroud *et al.,* 2006; Jain *et al.,* 2012; Amelio *et al.,* 2014; Locasale, 2013). Unlike some cancer cells, however, expression profiles of bacteriocytes in our study reveal that gluconeogenesis instead of glycolysis is occurring (Fig. 2-3); *i.e.* the Warburg effect, which is the phenomenon that the cellular energy production heavily on aerobic glycolysis (Vander Heiden *et al.,* 2009) is not occurring. This pattern of carbohydrate utilization in pea aphid bacteriocytes was identified previously (Poliakov *et al.*, 2011). An alternative energy source instead of glucose, such as pyruvate, may be metabolized to fuel the glycine/serine metabolism. This model of using pyruvate as the carbon skeleton for the glycine/serine metabolism has been proposed previously in breast cancer cells (Diers *et al.,* 2012). Similar to breast cancer cells we also observed the up-regulation of the pyruvate transporters MCT and MPC and the enzymes that utilize the carbon skeleton of pyruvate into gluconeogenesis in bacteriocyte cells (Fig. 2-3). Currently, it is unclear why cancer cells display this distinctive glycine/serine enriched metabolic profile, however several hypotheses have been proposed, such as reducing oxidative stress from cells that are metabolically very active (Jain *et al.,* 2012; di Salvo *et al.,* 2013; Maddocks *et al.,* 2013; Labuschagne *et al.,* 2014). Investigating the regulatory mechanisms behind the differences between metabolically active and prolific cancer cells and metabolically active, yet not prolific aphid bacteriocytes can become a key factor in elucidating the significance of how and why the glycine metabolism is adopted in both malignant cancer cells and symbiotic cells.

The genome-wide CpG methylation patterns of pea aphid showed higher levels of methylation in gene bodies especially in alfalfa feeding bacteriocytes (Fig. 2-4B), which is consistent with methylation patterns of other insect species with functional methylation systems (Suzuki *et al.,* 2007; Feng *et al.,* 2010; Zemach *et al.,* 2010; Provataris *et al.,* 2018). Bacteriocytes from both host plant treatments revealed significantly lower percent CpG methylation levels compared to the body cells of pea aphid (Fig. 2-4A). These results suggest that more exon and intron regions within bacteriocytes are hypo-methylated compared to body cells. Host plant treatments also influenced hypo- *vs.* hyper-methylation patterns in our study, as hundreds of genes were differentially methylated between bacteriocytes and body cells and between bacteriocytes depending on what host plant they fed upon. In one study (Huh *et al.,* 2013) it has been hypothesized that methylation inside of gene bodies plays an important role in reducing transcriptional noise (Huh *et al.,* 2013). Differential methylation within gene bodies has also been associated with alternative gene splicing (Shukla *et al.,* 2011). Here we observed hundreds of genes that were both differentially spliced and methylated within bacteriocytes compared to body cells and between bacteriocytes from alfalfa compared to fava feeding aphids. Genes that are hyper-methylated may also result in higher gene expression levels in insects. For example, a recent study suggested that gene body methylation of the lysosomal alpha-mannosidase (*LAM*) gene of *Apis melifera* increased the expression level of *LAM* (Wedd *et al.,* 2016). In contrast to this latter study, all genes that were both differentially expressed and methylated in this current study were hypo-

methylated and up-regulated, which is more in line with what is observed in vertebrates but within the promoter regions (Jones, 2012).

The effect of host plant treatment on DNA methylation patterns inside of pea aphid bacteriocytes was evident in our study. For example, our results reveal that the CpG methylation profiles of fava feeding aphid bacteriocytes were more heterogeneous between biological replicates compared to alfalfa feeding bacteriocytes (Fig. 2-5). Such heterogeneity may have come from the relaxed nutritional constraints of the 'universal' host plant, fava, compared to the specialized host plant, alfalfa. Alternatively, environmental cues from its specialized host plant, alfalfa, may induce host plant specific methylation profiles. In this study we did identify gene candidates to examine further that were both differentially expressed/spliced and methylated in bacteriocytes compared to body cells and between bacteriocytes in a host plant specific manner. If host plant specific patterns of DNA methylation induce tissue and host plant specific gene expression profiles in bacteriocytes this may be a key regulatory factor that induces phenotypic variation of the integrative metabolism in response to host plant diet.

In summary, our findings indicate that when aphids feed on their specialized host plant, alfalfa, key aphid genes that are involved in the integrative metabolism with *Buchnera* are differentially expressed/spliced, and some of these are differentially methylated. Future studies are needed to investigate how and if these methylated sites can influence the regulation of the bacteriocyte, be inherited, and ultimately drive host plant specialization broadly in sap-feeding insect-nutritional symbioses. We hypothesize that

this host plant induced metabolic modification to the aphid's integrative metabolism may ultimately allow aphids to utilize host plant diets that were once unsuitable.

Figure Legend:

Figure 2-1. Effects of a specialized host-plant diet on aphid and *Buchnera* phenotype. Different letters above bars indicate significant differences between aphid sub-lines within each sub-figure (A, B, C) (Tukey's multiple comparison post-hoc test $p<0.05$) (A) Aphid mass of 1st day adults. N=20 aphid individuals per aphid sub-line. (B) *Buchnera* abundance of 1st day nymphs and 1st day adults measured by a single copy *Buchnera* gene with RT-qPCR and normalized by a single copy aphid gene. N=6 aphid individuals per aphid sub-line (C) *Buchnera* cell density of 4th instar nymphs measured by the number of *Buchnera* cells per unit area (100 μm^2) using TEM.

Figure 2-2. Host-plant effects on differential expression and methylation of key aphid genes that complement and regulate the integrated aphid-*Buchnera* amino acid metabolism. Gene boxes are annotated with either E.C. numbers, Genbank LOC numbers, and/or gene names. Genes are significantly up-regulated in bacteriocytes compared to body cells (BAC vs. BODY) and enriched in alfalfa compared to fava bacteriocytes (ABAC vs. FBAC) if adjusted p-value$\leq$0.05 and normalized read counts are 50% higher. Genes are differentially methylated significantly if there was $\geq$10% difference in percent methylation and FDR corrected p-value $\leq$ 0.01.

Figure 2-3. Differential gene expression and methylation of the Glycine/Serine metabolism in aphid bacteriocytes. Gene boxes are annotated with either E.C. numbers, Genbank LOC numbers, and/or gene names. '*H*' denotes the glycine cleavage system H protein (LOC100169052). Genes are significantly up-regulated in bacteriocytes compared to body cells (BAC vs. BODY) and in alfalfa compared to fava bacteriocytes

(ABAC vs. FBAC) if adjusted p-value ≤0.05 and normalized read counts are 50% higher. Genes are differentially methylated significantly if there was ≥10% difference in percent methylation and FDR corrected p-value ≤ 0.01.

Figure 2-4. Percent CpG methylation level for bacteriocyte and body samples from each host-plant treatment. Different letters on bars indicate significant difference between each group and within each subfigure (A-D) (paired t-test $p < 0.05$). BAC and BODY denote bacteriocytes and body cells, respectively. Each sample has 6 biological replicates from both alfalfa and fava feeding aphids. ABAC and FBAC denote bacteriocytes of alfalfa feeding aphids and fava feeding aphids, respectively. Each sample has 3 biological replicates. (A) Average methylation levels of BAC and BODY. (B) Average methylation levels of ABAC and FBAC. (C) Average methylation levels of genic regions of BAC and BODY. (D) Average methylation levels of genic regions of ABAC and FBAC.

Figure 2-5. Principal Component Analysis of CpG methylation profiles for each bacteriocytes and body sample. F1, F2, and F3 denote 3 biological replicates of aphid sub-lines within the fava treatment. A1, A2, A3 denote 3 biological replicates of aphid sub-lines within the alfalfa treatment. PC1 explains 93.1% of total variance with standard deviation of 3.34. PC2 explains 1.5% of total variance with standard deviation of 0.43.

Figure 2-1.

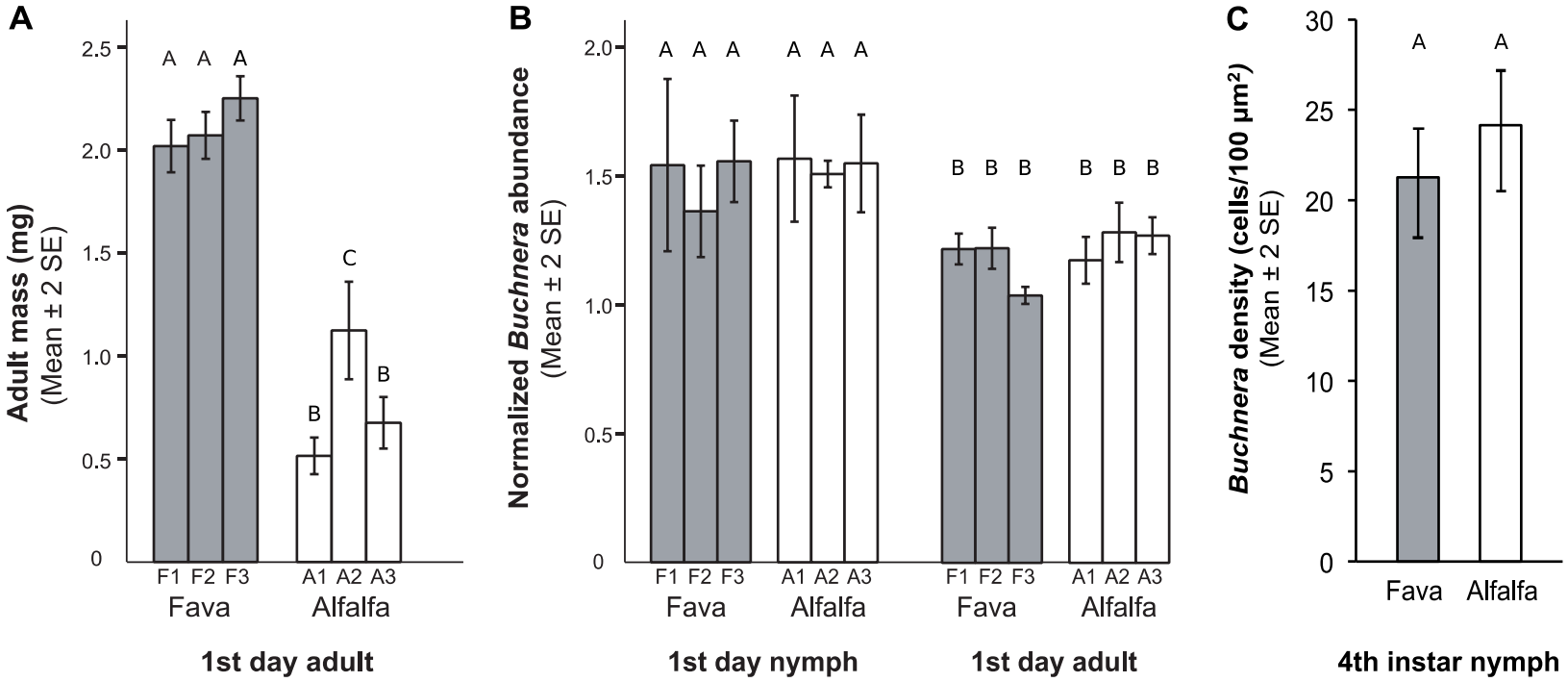
A
Adult mass (mg)
(Mean ± 2 SE)
2.5
2.0
1.5
1.0
0.5
0
A A A
B
C
B
F1 F2 F3
A1 A2 A3
Fava
Alfalfa
1st day adult
B
Normalized Buchnera abundance
(Mean ± 2 SE)
2.0
1.5
1.0
0.5
0
A A A A A A
B B B B B B
F1 F2 F3
A1 A2 A3
Fava
Alfalfa
1st day nymph
1st day adult
C
Buchnera density (cells/100 µm²)
(Mean ± 2 SE)
30
25
20
15
10
5
0
A A
Fava
Alfalfa
4th instar nymph

Figure 2-2.

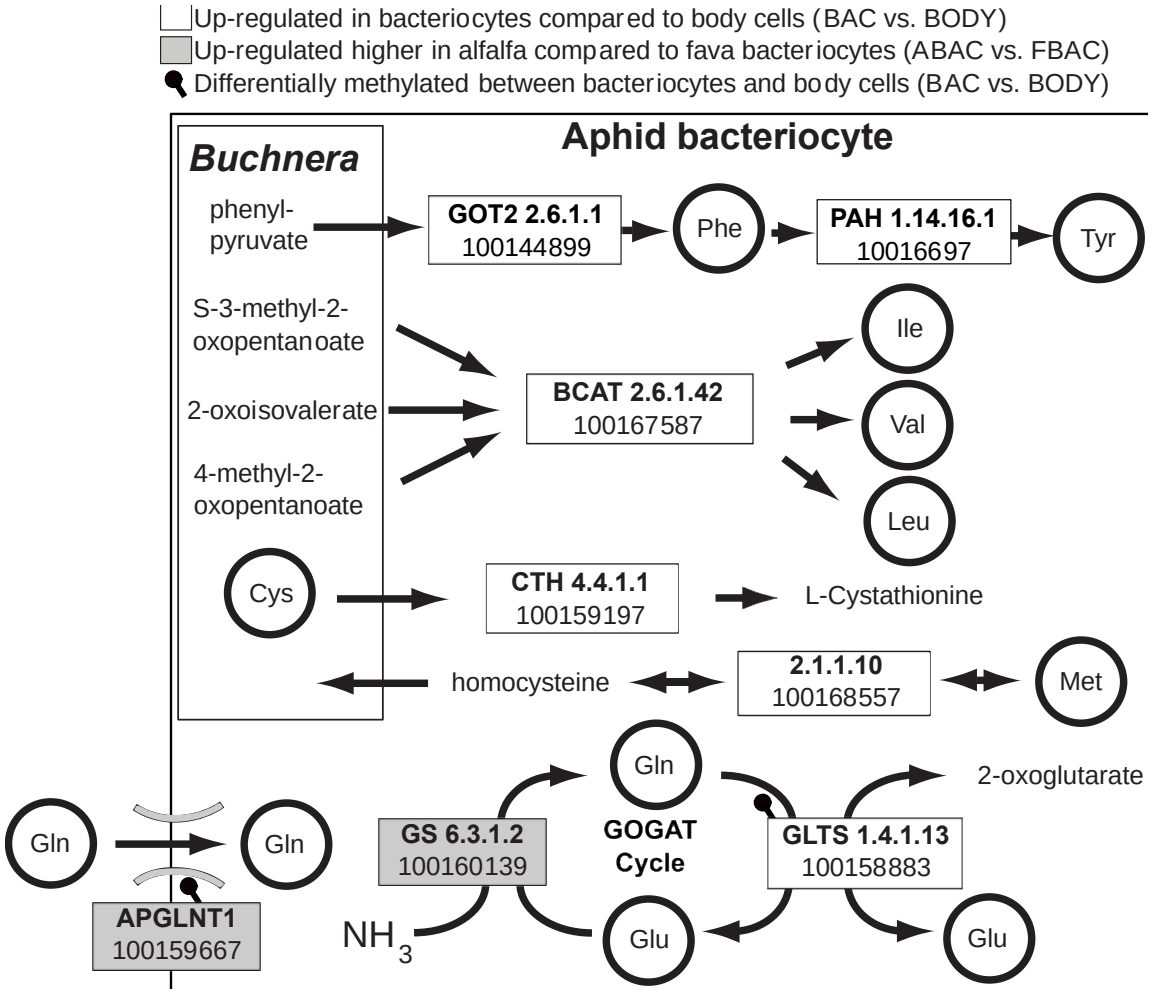
Up-regulated in bacteriocytes compared to body cells (BAC vs. BODY)
Up-regulated higher in alfalfa compared to fava bacteriocytes (ABAC vs. FBAC)
Differentially methylated between bacteriocytes and body cells (BAC vs. BODY)
Aphid bacteriocyte
Buchnera
phenyl-pyruvate
GOT2 2.6.1.1 100144899
Phe
PAH 1.14.16.1 10016697
Tyr
S-3-methyl-2-oxopentanoate
2-oxoisovalerate
4-methyl-2-oxopentanoate
BCAT 2.6.1.42 100167587
Ile
Val
Leu
Cys
CTH 4.4.1.1 100159197
L-Cystathionine
homocysteine
2.1.1.10 100168557
Met
Gln
Gln
APGLNT1 100159667
GS 6.3.1.2 100160139
NH3
Gln
GOGAT Cycle
Glu
GLTS 1.4.1.13 100158883
2-oxoglutarate
Glu

Figure 2-3.

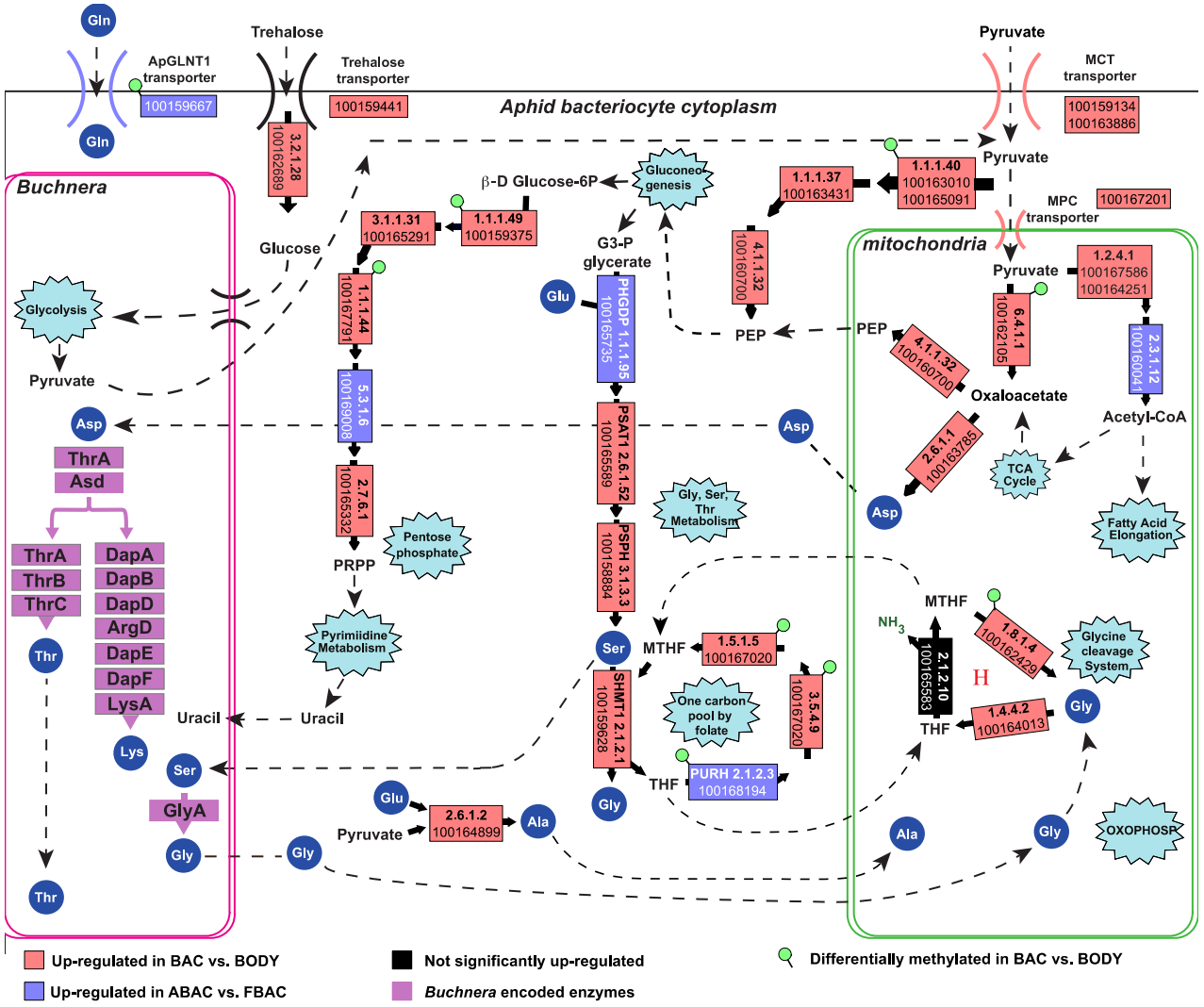
Gln
ApGLNT1 transporter
100159667
Gln
Trehalose
Trehalose transporter
100159441
Aphid bacteriocyte cytoplasm
Pyruvate
MCT transporter
100159134
100163886
3.2.1.28
100162689
Buchnera
β-D Glucose-6P
Gluconeo genesis
1.1.1.37
100163431
1.1.1.40
100163010
100165091
Pyruvate
MPC transporter
100167201
3.1.1.31
100165291
1.1.1.49
100159375
Glucose
G3-P glycerate
4.1.1.32
100160700
mitochondria
1.1.1.44
100167791
Glu
PHGDP 1.1.1.95
100165735
Pyruvate
1.2.4.1
100167586
100164251
Glycolysis
PEP
PEP
6.4.1.1
100162105
2.3.1.12
100160041
Pyruvate
4.1.1.32
100160700
Oxaloacetate
5.3.1.6
100169008
Asp
PSAT1 2.6.1.52
100165589
Asp
Acetyl-CoA
ThrA
Asd
2.6.1.1
100163785
TCA Cycle
2.7.6.1
100165332
Pentose phosphate
Gly, Ser, Thr Metabolism
Asp
Fatty Acid Elongation
ThrA
ThrB
ThrC
DapA
DapB
DapD
ArgD
DapE
DapF
LysA
PRPP
PSPH 3.1.3.3
100158884
MTHF
Thr
Pyrimiidine Metabolism
Ser
MTHF
1.5.1.5
100167020
NH3
1.8.1.4
100162429
Glycine cleavage System
Uracil
Uracil
SHMT1 2.1.2.1
100159628
One carbon pool by folate
3.5.4.9
100167029
2.1.2.10
100165583
H
Lys
Ser
THF
PURH 2.1.2.3
100168194
1.4.4.2
100164013
Gly
GlyA
Glu
2.6.1.2
100164899
Ala
Gly
Pyruvate
Gly
Ala
Gly
OXOPHOS
Thr
Up-regulated in BAC vs. BODY
Up-regulated in ABAC vs. FBAC
Not significantly up-regulated
Buchnera encoded enzymes
Differentially methylated in BAC vs. BODY

Figure 2-4.

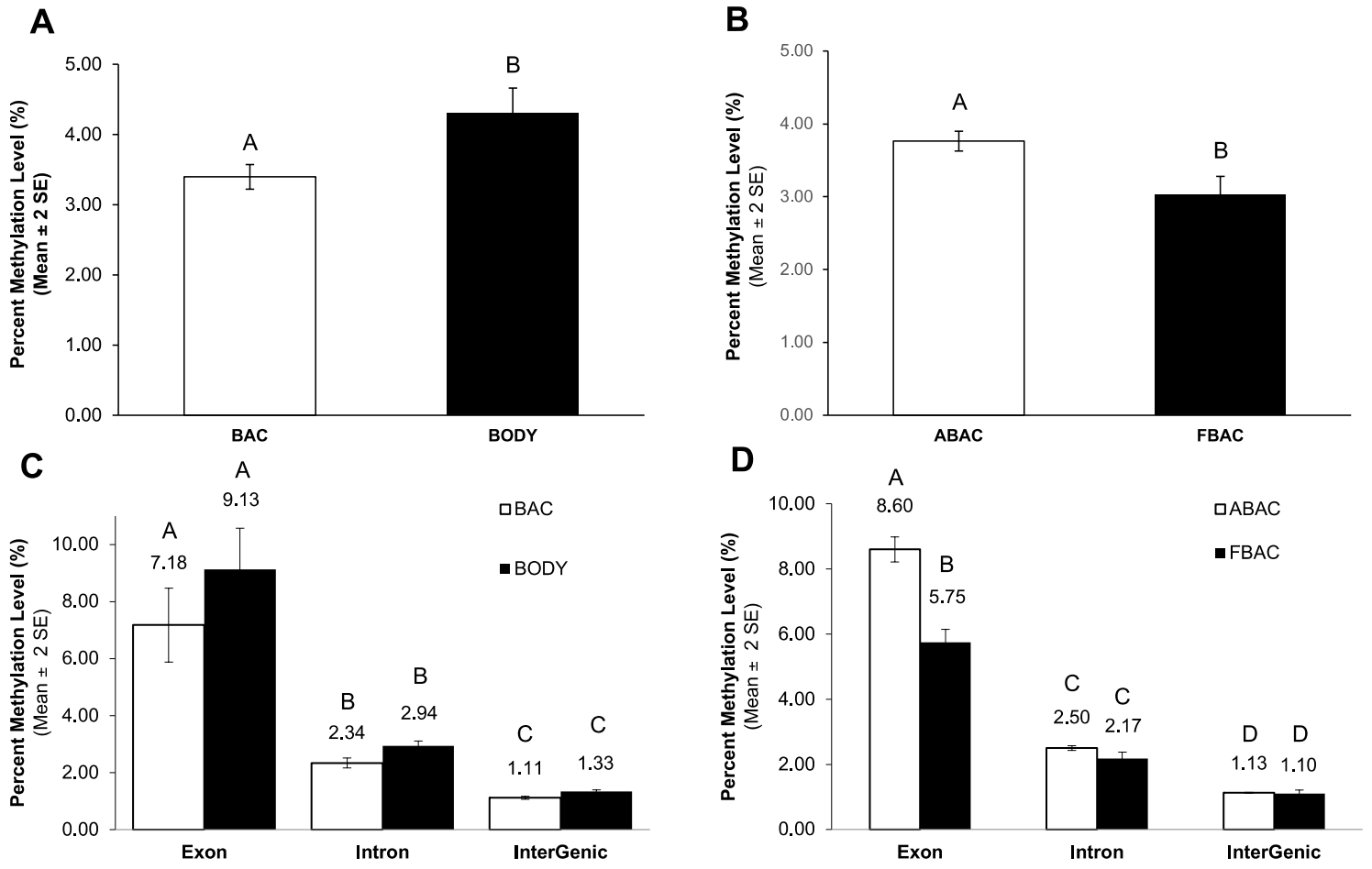
A
Percent Methylation Level (%)
(Mean ± 2 SE)
5.00
4.00
3.00
2.00
1.00
0.00
A
B
BAC
BODY
B
Percent Methylation Level (%)
(Mean ± 2 SE)
5.00
4.00
3.00
2.00
1.00
0.00
A
B
ABAC
FBAC
C
Percent Methylation Level (%)
(Mean ± 2 SE)
10.00
8.00
6.00
4.00
2.00
0.00
A
9.13
A
7.18
B
2.34
B
2.94
C
1.11
C
1.33
BAC
BODY
Exon
Intron
InterGenic
D
Percent Methylation Level (%)
(Mean ± 2 SE)
10.00
8.00
6.00
4.00
2.00
0.00
A
8.60
B
5.75
C
2.50
C
2.17
D
1.13
D
1.10
ABAC
FBAC
Exon
Intron
InterGenic

Figure 2-5.

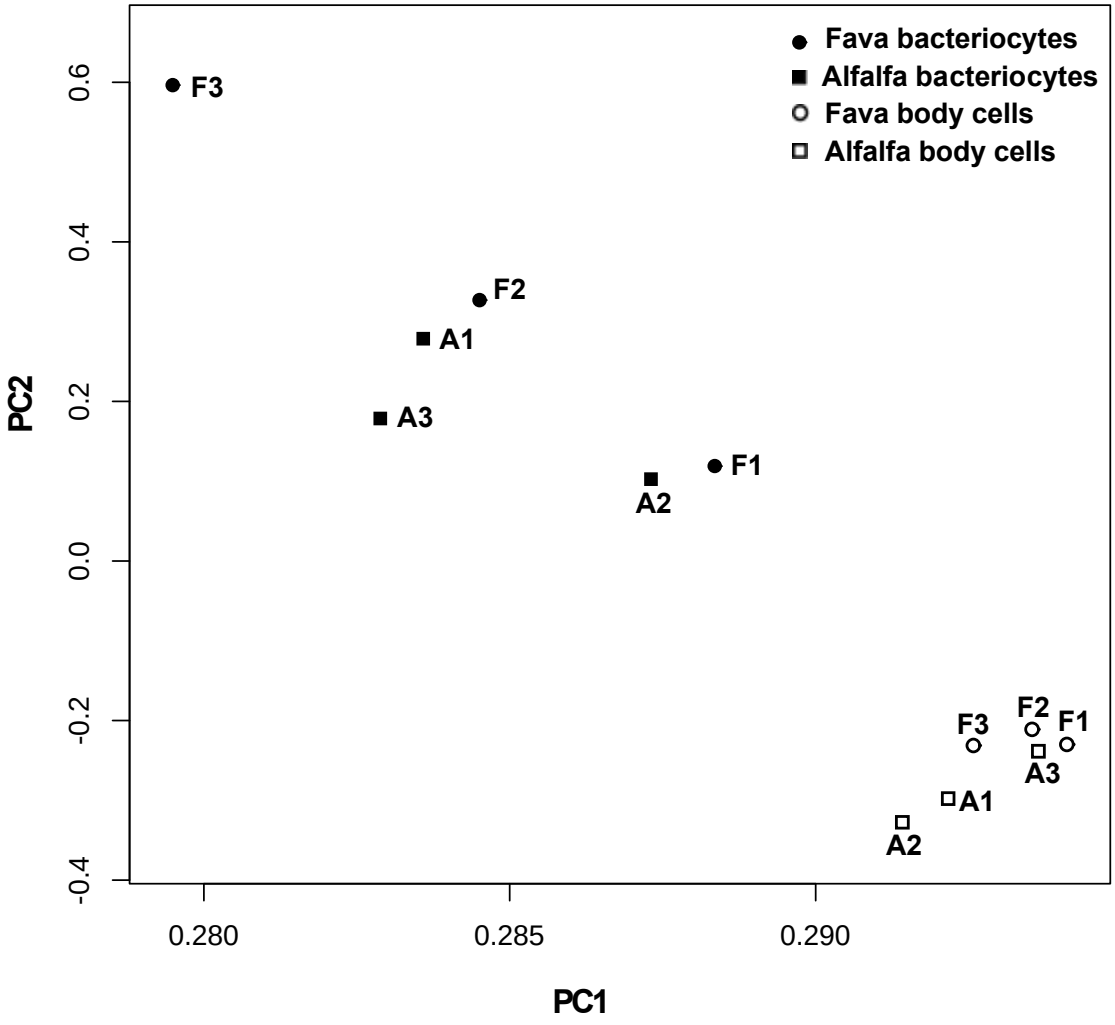
Fava bacteriocytes
Alfalfa bacteriocytes
Fava body cells
Alfalfa body cells
F3
F2
A1
A3
F1
A2
F3
F2
F1
A3
A1
A2
PC2
0.6
0.4
0.2
0.0
-0.2
-0.4
0.280
0.285
0.290
PC1

Table 2-1. Total RNA-Seq reads sequenced, quality trimmed, and successfully mapped as pairs for each bacteriocyte and body tissue sample in aphid host-plant treatments.

RNA-Seq samples	Total reads (76bps)	Total reads after trimming	Mapped reads (paired) aphid	Mapped reads (paired) aphid mitochondria	Mapped reads (paired) *Buchnera*
Fava sub-lines					
F1 Bacteriocytes	80,723,826	78,784,391	54,923,680	4,806,919	4,415,454
F1 Body Cells	79,079,310	77,647,276	70,404,510	2,201,097	0
F2 Bacteriocytes	81,633,998	80,354,255	62,761,468	5,706,771	4,044,574
F2 Body Cells	71,198,024	70,070,023	63,865,253	1,820,167	0
F3 Bacteriocytes	70,949,482	69,808,557	53,732,216	5,144,344	3,783,378
F3 Body Cells	65,093,760	63,966,507	58,063,394	1,797,792	0
Alfalfa sub-lines					
A1 Bacteriocytes	79,467,956	78,029,643	58,310,550	4,234,872	7,225,132
A1 Body Cells	74,620,738	73,329,565	65,953,928	2,525,267	0
A2 Bacteriocytes	59,473,084	58,527,569	45,256,013	3,340,620	4,239,528
A2 Body Cells	81,804,270	80,484,347	72,470,031	2,619,471	0
A3 Bacteriocytes	68,459,816	65,527,271	49,942,569	4,306,359	3,208,866
A3 Body Cells	68,934,700	66,654,812	59,286,048	2,129,278	0

F1, F2, and F3 denote 3 biological replicates of fava feeding aphids. A1, A2, and A3 denote 3 biological replicates of alfalfa feeding aphids.

Table 2-2. Number of significant[1], differentially expressed genes for each bacteriocyte comparison.

	ABAC vs. FBAC	BAC vs. BODY
Up-regulated	159	1,904
Down-regulated	103	4,211

ABAC and FBAC denote bacteriocytes of aphid lines feeding on alfalfa and fava, respectively. BAC and BODY denote bacteriocytes and body cells of both aphid lines feeding on alfalfa and fava, respectively. ABAC and FBAC have three biological replicates. BAC and BODY have 6 biological replicates. [1]FDR corrected p-value < 0.05, normalized read count values > 1.5x for BAC vs. BODY; FDR corrected p-value <0.1 for ABAC vs. FBAC.

Table 2-3a. Pathways significantly enriched in aphid bacteriocytes feeding on fava or alfalfa using Gene Set Enrichment Analysis (GSEA)

Description	Set size[a]	NES[b]	P_adj[c]
Bacteriocytes vs. Body cells			
Glycine, serine and threonine metabolism**	20	1.5	0.008
Glyoxylate and dicarboxylate metabolism**	18	1.49	0.017
Phenylalanine metabolism**	8	1.48	0.07
Pentose phosphate pathway**	17	1.46	0.044
Nicotinate and nicotinamide metaboslim**	10	1.43	0.027
Vitamin B6 metabolism**	3	1.4	0.037
Folate biosynthesis**	13	1.39	0.045
Biosynthesis of amino acids*	50	1.39	0.085
Phenylalanine, tyrosine and tryptophan biosynthesis*	3	1.36	0.074
Lipoic acid metabolism**	3	1.34	0.016
Hippo signaling pathway - fly*	43	-1.3	0.072
Notch signaling pathway**	19	-1.39	0.017
Other glycan degradation*	16	-1.44	0.087
Phototransduction - fly**	16	-1.46	0.002
Neuroactive ligand-receptor interaction**	27	-1.69	0

*p-value <0.1; **p-value <0.05; [a]Number of genes included in a gene set; [b]Normalized Enrichment Score; [c]Normalized p-value; Shaded rows indicate the pathways enriched in body cells.

Table 2-3b. Pathways significantly enriched in aphid bacteriocytes feeding on alfalfa relative to fava using Gene Set Enrichment Analysis (GSEA)

Description	Set size[a]	NES[b]	P_adj[c]
Alfalfa bacteriocytes vs. Fava bacteriocytes			
Synthesis and degradation of ketone bodies*	5	1.43	0.096
Vitamin B6 metabolism**	3	1.39	0
Aminoacyl-tRNA biosynthesis**	32	1.35	0
Pyruvate metabolism*	30	1.35	0.097
Jak-STAT signaling pathway*	18	1.29	0.092
Lipoic acid metabolism**	3	1.28	0
Butanoate metabolism**	15	1.27	0

*p-value <0.1; **p-value <0.05; [a]Number of genes included in a gene set; [b]Normalized Enrichment Score; [c]Normalized p-value.

Table 2-3c. Pathways significantly enriched in aphid bacteriocytes relative to body cells feeding on both host plants from differentially methylated genes between bacteriocytes relative to body cells using Gene Set Enrichment Analysis (GSEA)

Description	Set size[a]	NES[b]	P_adj[c]
Alfalfa bacteriocytes vs. Fava bacteriocytes			
Metabolic pathways **	47	1.68	0.006
Lysosome **	8	1.65	0
Protein processing in endoplasmic reticulum**	7	1.51	0.04
Selenocompound metabolism*	2	1.47	0.002
Tryptophan metabolism**	3	1.34	0.083

*p-value <0.1; **p-value <0.05; [a]Number of genes included in a gene set; [b]Normalized Enrichment Score; [c]Normalized p-value.

Table 2-4. Total whole genome bisulfite sequencing reads, high quality reads successfully mapped as pairs, and total read coverage for the whole genome, CpG sites, and genes for each bacteriocyte and body tissue sample in aphid host-plant treatments.

DNA samples	Total reads	Mapped reads (paired) *Buchnera*	Mapped reads (paired) aphid	Total read coverage aphid genome	Total read coverage CpGs (%)[a]	Total read coverage genes (%)[b]
Fava sub-lines						
F1 Bacteriocytes	101,895,398	59,637,111	14,278,056	1.2x	46.8	49.6
F1 Body Cells	45,538,395	0	26,129,311	4.8x	58.9	54.4
F2 Bacteriocytes	131,924,346	85,647,711	14,063,069	1.2x	47.5	46.1
F2 Body Cells	56,394,111	0	31,477,892	5.8x	62.9	62.9
F3 Bacteriocytes	61,926,293	39,367,593	6,830,107	0.9x	37.3	36.2
F3 Body Cells	44,062,440	0	24,627,528	4.6x	55.7	54.6
Alfalfa sub-lines						
A1 Bacteriocytes	107,886,414	73,506,570	11,090,840	0.9x	33.5	42.2
A1 Body Cells	28,669,685	0	17,578,080	3.2x	58.8	48.3
A2 Bacteriocytes	112,822,307	75,303,932	11,830,822	0.9x	48.6	43.1
A2 Body Cells	28,405,921	0	17,928,345	3.3x	53.0	50.6
A3 Bacteriocytes	81,042,088	54,282,855	7,982,616	0.7x	38.7	38.5
A3 Body Cells	88,555,344	0	22,527,414	3.1x	55.4	52.7

F1, F2, and F3 denote 3 biological replicates of aphid lines with fava bean treatment. A1, A2, and A3 denote 3 biological replicates of aphid lines with alfalfa treatment.

[a]Total read coverage CpGs are the percentage value of the CpG sites that are covered by the high-quality mapped reads with ≥10x coverage.

[b]Total read coverage genes are the percentage value of the aphid genes that are covered entirely by the high-quality mapped reads with ≥10x coverage.

Table 2-5a. Number of differentially methylated CpG sites and differentially expressed or spliced genes

Differentially Methylated CpG Sites[a]	Both differentially methylated and expressed genes[b]	Differentially spliced genes[c]	Both differentially methylated and spliced genes
Bacteriocytes vs. Body cells			
3,474	441	3,859	702
Fava bacteriocytes vs. Alfalfa bacteriocytes			
294	3	47	3

Table 2-5b. Number of differentially methylated CpG sites be genic regions

Differentially methylated CpG sites[a] - Total	Differentially methylated CpG sites - Exon	Differentially methylated CpG sites - Intron	Differentially methylated CpG sites - Intergenic
Bacteriocytes vs. Body cells			
3,474	1,625 (47%)	1,215 (35%)	630 (18%)
Fava bacteriocytes vs. Alfalfa bacteriocytes			
294	141 (48%)	89 (30%)	64 (22%)

Each group has 6 biological replicates from alfalfa and fava feeding aphids.

[a]Differentially methylated sites were determined with > 10% difference level, FDR-adjusted p-value < 0.01, with minimum of 10 read cutoff. [b]Differentially expressed genes were determined with FDR-adjusted p-value < 0.05, 1.5X cutoff based on normalized read counts. [c]Significance determined based on Jensen-Shannon divergences of splicing isoforms.

Table 2-6. KEGG Pathways of the genes that are differentially methylated and differentially spliced between bacteriocyte and body samples.

Description	Set size[a]
Metabolic pathways	40
RNA transport	14
Spliceosome	12
Protein processing in endoplasmic reticulum	12
Wnt signaling pathway	12
mTOR signaling pathway	10
Endocytosis	9
Hippo signaling pathway - fly	9
Ubiquitin mediated proteolysis	8
Ribosome biogenesis in eukaryotes	7
mRNA surveillance pathway	7
RNA degradation	7
Glycerophospholipid metabolism	6
Biosynthesis of amino acids	6
Lysosome	6
Jak-STAT signaling pathway	6

[a]Number of genes that belong to the pathway

Chapter 3: Conserved and Lineage Specific Patterns of DNA Methylation in the Soybean Aphid (*Aphis glycines*) (Hemiptera: Aphididae)

Abstract

DNA methylation is associated with the modification of gene expression levels and therefore can impact an organism's phenotype without a change in the DNA sequence. While not all insects possess functional CpG DNA methylation systems recent studies have demonstrated that DNA methylation is linked to the regulation of adaptive traits in a diversity of insect orders. The soybean aphid, *Aphis glycines* Matsumura, is a major agricultural pest that displays a diversity of resistance phenotypes on soybean. The genomic basis for these phenotypes is largely unknown. To determine if *A. glycines* possesses a functional DNA methylation system, and if so, the location and patterns of DNA methylation throughout the genome, we performed whole genome bisulfite sequencing on the head tissue of *A. glycines*. Our results reveal that *A. glycines* possesses a functional CpG DNA methylation system and that *de novo* and aphid-specific genes have significantly lower levels of CpG DNA methylation compared to ancestral house-keeping genes. We also identified one cytochrome P450 gene and four effector genes that display CpG methylation in the exon regions of these genes. These five genes are aphid-specific in aphids and are prime gene candidates involved in insect-plant interactions, especially at the epigenetic level.

Introduction

In the most diverse group of animals (the insects) studies reveal that environmental cues may trigger cell reprogramming by DNA methylation, resulting in the regulation of adaptive traits. For example, DNA methylation and differential gene expression patterns have been associated with caste determination in honey bees, *Apis melifera* Linnaeus, and several species of ants (Foret *et al.,* 2012; Bonasio *et al.,* 2012), host plant use and pesticide resistance in aphids (Field and Blackman, 2003; Gong *et al.,* 2012; Kim *et al.,* 2018), and the manipulation of insect hosts by parasitoid wasps (Kumar and Kim, 2017). The impact of these epigenetic effects on insect ecology and evolution may be widespread as the presence of functional CpG methylation has been both predicted computationally and confirmed empirically from species belonging to a diversity of insect orders (Provataris *et al.,* 2018). Nevertheless, our basic understanding of the mechanisms underlying insect DNA methylation and their effect on gene function remains largely elusive. Several hypotheses have been proposed (reviewed in Glastad *et al.,* 2011). One strategy that is currently being used to gain a firmer understanding of insect DNA methylation and genome function, analogous to the Human epigenome project (Bernstein *et al.,* 2010), is to map the location of CpG DNA methylation for a diversity of insect tissue types and genomes. Ultimately, the accumulation of this type of epigenomic data will help the research community further understand the evolution and function of DNA methylation in insect genomes and beyond.

Insects that produce offspring from unfertilized egg cells (parthenogenic) are fascinating subjects to characterize on an epigenomic level because identical genotypes can produce diverse phenotypes. For example, parthenogenic aphids that are genetically identical are known to display a myriad of phenotypic morphs in response to particular environmental conditions (Ragsdale *et al.*, 2011; Brisson *et al.*, 2016). Several studies that have associated specific aphid phenotypes with distinct DNA methylation profiles include color morphs (Dombrovsky *et al.,* 2009) and host plant use in pea aphids, *Acyrthosiphon pisum* Harris (Kim *et al*., 2018), insecticide resistance (Field and Blackman, 2003) and sex morphs in peach potato aphids, *Myzus persicae* Sulzer (Mathers *et al*., 2018), and host plant virulence in Russian wheat aphids, *Diuraphis noxia* Kurdjumov (Gong *et al.,* 2012). These epigenetic associated phenotypes can have a major influence on aphid ecology and evolution and therefore having a greater understanding of aphid DNA methylation patterns is warranted.

Recently, the genome sequence of the *A. glycines* has become publicly available (Wenger *et al.,* 2017). This aphid species is of particular interest to the scientific community and agriculture because it is associated with phenotypic morphs (Voegtlin et al., 2004) and different virulence response to soybean, *Glycine max* (L.) Merr., genotypes with resistance genes (*Rag* = resistance to *A. glycines*) (Hill *et al.,* 2012). Moreover, this species is responsible for significant yield losses of soybean, and extensive pesticide use (Ragsdale *et al*., 2011). Accordingly, in this study we take advantage of the genomic resource and conduct whole genome bisulfite sequencing on *A. glycines*. Specifically, we determine if DNA methylation occurs in *A. glycines*, and if so compare and contrast DNA

methylation levels in both ancestral and *de novo* genes. We also determine the presence and location of DNA methylation within aphid genes that are involved in plant virulence to highlight specific gene candidates for future study.

Materials and Methods

Aphid culture, DNA extraction, and bisulfite sequencing

For this experiment, apterous vivipara adults of *A. glycines* biotype 2 were used. This biotype was described based on its distinct resistance against (*Rag* = resistance to *A. glycines*) 1 gene. A colony of *A. glycines* biotype 2 was isolated from Ohio (Kim *et al.*, 2008) and is maintained at the USDA-ARS laboratory located at the University of Illinois at Urbana-Champaign. Biotype 2 was reared on LD10-5903a (*Rag1*) in square BugDorm-44545 insect cages that are: 47.5x47.5x47.5 cm (MegaView Science Co., Taichung, Taiwan) in isolated plant tissue-culture chambers (Percival, TC-2) set at 23°C constant temperature, and a photoperiod of 16 h per day.

Genomic DNA was isolated from pooled aphid head tissues from approximately 50 vivipara adults using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) following manufacture guidelines. After DNA was extracted, each sample was quantified using the Qubit dsDNA HS Kit (Invitrogen, Carlsbad, CA). Up to 500 ng of DNA was used for the bisulfite conversion process with the EZ DNA Methylation Lightning™ Kit (Zymo Research, Irvine, CA). As an internal control, 0.5 ng of λ phage DNA (New England BioLabs, Ipswich, MA) was spiked into the aphid DNA samples to calculate the rate of false positives from the sodium bisulfite conversion treatments. The bisulfite library was

constructed using the TruSeq DNA Methylation Kit (Illumina, San Diego, CA) with the starting material of 50 ng of bisulfite-treated DNA following the manufacturer's guidelines. Strand-specific paired-end sequencing was conducted on HiSeq 4000 (Illumina, San Diego, CA) using a TruSeq SBS sequencing kit (Illumina) at University of Illinois at Urbana-Champaign's Roy J. Carver Biotechnology Center. Sequenced DNA reads were then deposited into the Sequence Read Archive at the National Center for Biotechnology Information (NCBI) under accession no. PRJNA509918.

Methylation data analysis

Whole genome bisulfite sequencing data were processed and analyzed similar to Kim *et al*. (2018). Briefly, raw read data from whole genome bisulfite sequencing were trimmed to remove Illumina index sequences using Trimmomatic (Bolger *et al.*, 2014). Methylation read data were aligned to the *A. glycine*s's genome (Wenger *et al.*, 2017) with Bowtie2 and Bismark as suggested in Krueger and Andrews (2011). Genome-wide methylation levels of cytosines were calculated using Bismark v0.19.1 (Krueger and Andrews, 2011). Site-specific CpG methylation data were calculated using the methylKit package in R (Akalin *et al.*, 2012). Only the CpG sites with the minimum coverage of 10 reads or more were considered in calculating the methylation levels based on the λ phage internal control (Kim *et al*., 2018).

Site-specific methylation levels of each CpG site were determined as the percentage of methylated cytosines out of all the reads mapped to the cytosine nucleotide position. Only CpG sites with a site-specific methylation level ≥ 50% were considered as "methylated". Methylation levels of each gene were calculated by averaging the site-

specific methylation level of all CpG sites within the genic region. The *A. pisum* genome annotation v2.0 (International Aphid Genomics Consortium, 2010) was used to identify orthologs and aphid-specific genes within the *A. glycines* genome using reciprocal best BLASTP hits (Camacho *et al.,* 2009) with the cut-off of 35% sequence identity and the e-value of 10^{-5}. Similarly, the *A. glycines* genome was compared against the *D. melanogaster* genome r6.22 (Gramates *et al.,* 2018) using BLAST and the human genome GRCh38 (Genome Reference Consortium) to identify orthologs within the *A. glycines* genome. Statistical comparison of the average methylation level between species-specific genes (i.e. *de novo* genes, genes unique in the soybean aphid genome) and genes with *A. pisum* homologs was performed with Fisher's exact test with $p < 0.05$. Genes were ranked into one-tenth intervals (deciles) according to gene-specific methylation levels. Functional analyses were conducted using the *A. glycines* Gene Ontology Annotations (Wegner *et al.,* 2017) available from Aphidbase (Legeai *et al.,* 2010). The top ten most abundant Gene Ontology Annotation categories, i.e. categories that contain the largest number of annotated genes from the *A. glycines* genome, were used for this analysis.

To identify DNA methyltransferase (DNMT) genes in the *A. glycines* genome, genomic sequences of *A. glycines* were obtained from the Bioinformatics Platform for Agroecosystem Arthropods (BIPAA) (Wegner *et al.,* 2017) and compared against the *DNMT* nucleotide sequences from the *A. pisum* genome v2.0 from GenBank (International Aphid Genomics Consortium, 2010). *Aphis glycines* orthologs for *DNMTs* were identified from reciprocal best BLAST hits (Camacho *et al.,* 2009) with an E-value

cut-off of 10^{-25}. Nucleotide coding sequences of DNMT insect homologs were obtained for *A. pisum*, *M. persicae*, *D. noxia*, and *A. mellifera* from existing annotations on GenBank. Obtained sequences were aligned using MAFFT 7.402 (Katoh and Standley, 2013). The aligned sequences were then used to generate a maximum likelihood phylogenetic tree using MEGA7 v7.0.26 (Kumar *et al.*, 2016) with 1,000 bootstrap iterations and Generalized Time Reversible (GTR) substitution model. The phylogenetic tree has been rooted with at the midpoint.

To understand the long-term evolutionary patterns of CpG methylation in the *A. glycines* genome, the CpG Observed/Expected ratio were calculated as (CpG frequency)/ (C frequency * G frequency) as in Elango *et al.* (2009) for each gene using a customized Perl script. The distributions of ratios were plotted using R v3.4.4 (R Core Team, 2018). The bimodality of the plots was tested with the diptest package v0.75 (Maechler, 2016).

Results and Discussion

Genome wide patterns of CpG methylation

To investigate the presence of functional CpG methylation in *A. glycines*, the whole genome bisulfite sequencing for the aphid head region was performed to characterize head specific methylation patterns that may be associated with host-plant interactions (e.g. salivary glands and taste receptors). A total of 22,975,863 reads were sequenced and aligned against the *A. glycines* genome with the mapping efficiency of 51.0%, which is similar to previous DNA methylation studies (Kim *et al.*, 2018).

Approximately 83% of *A. glycines* genes had a minimum of 10 reads per CpG site and were used for methylation analyses in this study.

In eukaryotes, DNA methylation primarily occurs on a cytosine residue in the dinucleotide context of cytosine followed by a guanine nucleotide (CpGs) compared to CHG and CHH (H = A, C, or T) (Feng *et al.*, 2010). Accordingly, insects with functional DNA methylation display higher genome-wide methylation levels for CpG dinucleotide sites compared to CHG and CHH nucleotide sites (Lyko and Maleszka, 2011). For *A. glycines*, it was found that the genome-wide percentage of methylated cytosine in CpG context was 4.1%, whereas the CHG and CHH context were 1.9% and 1.2%, respectively. These results are similar to average CpG, CHG, and CHH methylation levels observed previously in other insect species such as *A. mellifera* (Lyko *et al.* 2010) and *A. pisum* (Walsh *et al.*, 2010).

Identification of DNA methyltransferases

DNA methylation requires the enzymatic addition of a methyl group to DNA nucleotide bases by DNA methyltransferases (DNMTs) (Klose and Bird, 2006). In *A. pisum*, all three types of DNMT genes (*DNMT1*, *DNMT2*, and *DNMT3*) have been identified (Walsh *et al.* 2010). *DNMT2* sequences were included in this analysis due to its shared domain and sequence homology to *DNMT1* and *DNMT3* even though its function is not strictly involved in DNA methylation (Rai *et al.*, 2007). It was also found that *A. glycines* shares the same three *DNMT* homologs with *A. pisum* (Fig. 3-1). All the homologs share more than 79% nucleotide sequence similarity when compared to *A. pisum DNMTs*. Interestingly, two *A. glycines* genes, AG002259 and AG002260 were

identified as DNMT1 homologs, suggesting a recent gene duplication event. Therefore, it is likely that CpG methylation is conserved within the aphid family (Fig. 3-1); however, duplication of DNMT1 could lead to neofunctionalization, sub-functionalization, or the formation of a pseudogene for one of the paralogs as observed in other duplicated genes in another insect species (Kelleher and Markow, 2009).

Genome wide patterns of Observed/Expected CpG ratios

Methylated cytosine can be deaminated into uracil, which is then replaced with thymidine through DNA repair mechanisms resulting in C to T transitions (Coulondre *et al.*, 1978). As a consequence, through the process of mutation CpG dinucleotides are reduced over time across the genome. Such depletion can be measured by the normalized CpG content: the ratio of the observed number of CpG dinucleotide sites to the expected number of sites within a gene based on frequencies of cytosines and guanines (CpG O/E ratio) (Elango *et al.*, 2009). Therefore, the CpG O/E ratio for all genes were calculated within the *A. glycines* genome. The genome-wide distribution of the CpG O/E ratios shows a significant bimodality (Hartigans' dip test; p-value = 0.00254) (Fig. 3-2). A bimodal distribution of CpG O/E ratios indicates that there are two distinct groups of genes in terms of DNA methylation. Genes with high CpG O/E ratios are predicted to historically have low CpG methylation levels, whereas those with low CpG O/E ratios are predicted to historically have high levels of CpG methylation. Such bimodality can be found in other insect species with functional CpG methylation systems such as *A. melifera* and *A. pisum* (Glastad *et al.*, 2011; Walsh *et al.*, 2010). This is in contrast to the unimodal distributions observed in organisms without functional CpG methylation such

as fruit fly, *Drosophila melanogaster* Meigen, and red flour beetle, *Tribolium castaneum* Herbst (Elango *et al.*, 2009). Additionally, I ranked the genes according to their average methylation levels from the whole genome bisulfite data generated in this study. As expected, highly methylated genes generally showed lower CpG O/E ratios compared to genes with low CpG methylation levels (Fig. 3-3). These results indicate that these highly methylated genes have lost many CpG sites through deamination, however high CpG methylation levels of these genes have been maintained through evolutionary history.

Methylation levels of lineage specific genes

In insects, highly methylated genes generally are housekeeping genes that are constitutively expressed in all tissue types (Hunt *et al.*, 2013). Such ubiquitously expressed genes are often conserved at the amino acid and nucleotide sequence level (Elango *et al.*, 2009). It has been suggested that these housekeeping genes, many of which are transcribed at low to moderate levels, tend to show higher methylation levels compared to genes that are either highly expressed and/or newly emerged (Glastad *et al.*, 2014). It has also been hypothesized that gene body methylation, which is commonly observed in invertebrate organisms may suppress transcriptional noise (Huh *et al*., 2013). Collectively these observations suggest that the methylation level of genes is associated with the age of the gene in the genome (e.g. ancestral compared to newly emerged) in addition to the mRNA expression level. To investigate the CpG methylation levels within the *A. glycines* genome, we calculated the methylation levels of genes unique to the *A. glycines* genome (*de novo*) in addition to those conserved in other genomes that range in evolutionary distance from *A. glycines* such as: the *A. pisum* genome v2.0 (International

Aphid Genomics Consortium, 2010), the *D. melanogaster* genome r6.22 (Gramates *et al.*, 2017), and the *H. sapiens* genome GRCh38 (Genome Reference Consortium). From this analysis, 826 *A. glycines* genes have no homologs while 4,921 genes share homologs with only *A. pisum*. Furthermore, 1,266 *A. glycines* genes share homologs with both *A. pisum* and *D. melanogaster*, whereas 6,962 *A. glycines* genes share homologs in *A. pisum, D. melanogaster,* and *H. sapiens* (Fig. 3-4A). Within the genome of *A. glycines*, species- and aphid-specific genes show significantly lower methylation levels compared to genes that share homologs with *D. melanogaster* and humans (Fig. 3-4B; Tukey's HSD test, $p<0.05$). Specifically, the average methylation levels of the unique genes in the *A. glycines* genome and of the genes that share homology with *A. pisum* were 10.6% and 10.9%, respectively. In comparison, *A. glycines* genes that share homologs in *D. melanogaster* and *H. sapiens* have an average methylation level of 14.9%, whereas the genes that have homologs in all 4 species (*A. glycines, A. pisum, D. melanogaster,* and *H. sapiens*) have an average methylation level of 37.5% (Figure 3-4B). Our results are consistent with Provataris and colleagues (2018) who revealed that DNA methylation levels are significantly higher for the genes that are conserved.

Methylation level and gene functions

To investigate whether genes with different levels of CpG methylation are associated with specific functions we compared and ranked % methylation levels (in one-tenth intervals) to Gene Ontology Annotation Information (Fig. 3-5). There was a trend for genes that belong to ATP binding and nucleic acid binding groups to possess higher DNA methylation levels, whereas genes involved in transcription regulation possess

lower DNA methylation levels. However, these trends were not statistically significant indicating that there is not a strong association between Gene Ontology Annotations and average CpG methylation levels for *A. glycines* genes (Kruskal-Wallis chi-squared = 3.486, df = 8, p-value = 0.9003).

Cytochrome P450 genes are a group of enzymes that have diverse functions and are closely associated with an insect's survival and adaptation (Scott 1999). These genes are ubiquitously expressed in all insect tissue types including brain and salivary glands (Scott *et al.,* 1998). In *A. pisum*, one of the insect-specific cytochrome P450 genes, *CYP4G51*, is known to be correlated with hydrocarbon synthesis, which is directly linked to the insect's tolerance to desiccation (Chen *et al.,* 2016). In *M. persicae*, over-expression of a single P450 gene, *CYP6CY3*, results in resistance to neonicotinoid pesticides (Puinean *et al.,* 2010). *Aphis glycines* are host plant specialists on soybeans and demonstrate various mechanisms of adaptation to overcome host plant resistance (Hill *et al.,* 2012). One type of mechanism that has been suggested to overcome host plant resistance is the utilization of cytochrome P450 detoxification genes (Bansal *et al.,* 2014). Based on the data from Wenger *et al.* (2017), 68 *A. glycines* cytochrome P450 genes have been identified. Out of these 68 genes 14 showed differential expression when aphids feed on *Rag1* resistant compared to susceptible soybean plants (Bansal *et al.,* 2014; Table 3-1). We calculated the average methylation levels of these cytochrome P450 gene candidates from soybean aphid heads. The average methylation level of cytochrome P450 genes was 7.0%, which was significantly lower than the average CpG methylation level of all genes (24.6%; ± 34.86 STDEV). Four genes showed methylation levels that were

much higher than the rest of the cytochrome P450 genes and include: AG015040 (17.8%), AG000379 (73.2%), AG012883 (92.2%), and AG005349 (94.4%) (Table 3-1). All four genes have orthologs in *A. pisum*, *D. melanogaster* and *H. sapiens*, which coincide with our previous finding that highly methylated genes are often maintained throughout evolutionary history. Additionally, two of the 14 differentially expressed genes from Bansal et al (2014), AG005349 and AG005414, had 14 and two methylated CpGs in their exon regions, respectively.

Another important group of genes that is involved in insect-plant interactions are insect effector genes. Effector genes encode proteins that can modulate the plant defense responses in insect herbivores (Yates and Michel, 2018). Identification of the targets and/or mechanisms of effector proteins may reveal how insect herbivores such as aphids display host resistance. In *A. glycines*, a total of 94 putative effector genes that are secreted and not anchored in cell membranes have been suggested to be important in host plant resistance (Wenger *et al.*, 2017). Of such genes 87 have CpG sites. The average methylation levels of those 87 genes were 13.2% (±25.32 STDEV), indicating that these genes are generally hypomethylated compared to other genes in *A. glycines* genome. Only 11 genes had methylation levels ≥ 50% (Table 3-2). Of those 11 genes, two genes (AG006612 and AG007896) have orthologs within the *A. pisum* and *D. melanogaster*, while the remaining 9 genes share homology with *A. pisum*, *D. melanogaster*, and *H. sapiens* genes (Table 3-2). Although the methylation levels in this study were obtained from the head body region of *A. glycines*, head tissue is composed of a variety of different tissue types including the aphid salivary glands. Therefore, it is possible that

methylation levels for specific effector genes may be higher or lower for individual tissues types within the head region of the aphid body.

In addition to average methylation levels of genes, site-specific CpG methylation within genic regions may also influence gene expression in the form of gene activation and/or alternative splicing in insects (Field and Blackman, 2003; Foret *et al.*, 2012). For example, in *M. persicae*, overexpression of the E4 esterase gene confers the aphid resistant to insecticides (Field and Blackman, 2003). This study showed that the loss of CpG methylation within this gene silences gene expression, which results in the loss of insecticide resistance (Field and Blackman, 2003). We identified the methylated CpG sites for cytochrome P450 genes and effector genes (as detailed above) and found that 92.7% and 93.5% of the methylated CpG sites are confined in the exon regions, respectively (Table 3-1, 3-2). Only five aphid-specific genes for both effector and cytochrome P450 gene lists displayed CpG methylation: One cytochrome P450 gene (AG017396; Table 3-1) and four effector genes (AG001590, AG009361, AG010599, AG014673; Table 3-2). Specifically, all five genes had one methylated CpG site in their exon regions (Table 3-1, 3-2). These newly emerged and methylated genes are important candidates to investigate further for future insect-plant interaction studies.

Figure Legend:

Figure 3-1. Maximum likelihood analysis of the nucleotide sequences of the DNA methyltransferases (*DNMT*s). *DNMTs* from the pea aphid (*Acyrthosiphon pisum*; Ap), the green peach aphid (*Muzus persicae*; Mper), the Russian wheat aphid (*Diruaphis noxia*; Dnox), the soybean aphid (*Aphis glycines*; Agly; blue), and the honeybee (*Apis mellifera*; Amel) were aligned with MAFFT. Maximum likelihood phylogenetic tree constructions were performed based on the Jones *et al.* (1992) with a bootstrap parameter of 1000 using MEGA7 (Kumar *et al.*, 2016).

Figure 3-2. Distribution of the CpG ratios (CpG expected / CpG observed) in the coding sequence of *Aphis glycines*. The CpG frequencies for each gene were calculated from available genome sequence data. The expected CpG frequencies for each coding sequence were calculated based on its GC content. The middle, red vertical line depicts 1.0.

Figure 3-3. Mean CpG O/E of *Aphis glycines* coding sequences grouped by methylation level. The genes were binned into 10 groups by their average methylation levels in deciles. Functional categories represent the 10 most abundant Gene Ontology Annotation categories that contain annotated genes from the *A. glycines* genome. The observed CpG frequencies for each gene were calculated from available genome sequence data. The expected CpG frequencies for each coding sequence were calculated based on GC content of *A. glycines'*s genome.

Figure 3-4. Methylation levels of lineage-specific and orthologous *Aphis glycines* genes. A. Numbers of *Aphis glycines* (soybean aphid) orthologs that are shared between

Acyrthosiphon pisum (pea aphid), *Drosophila Melanogaster* (fruit fly), and *Homo sapiens* (human) are represented as numbers within the Venn diagram. The number located outside the three circles represents the number of unique genes found only within *A. glycines's* genome. Orthologous genes were identified using reciprocal BLASTP best hits against *Ac. pisum*, *D. melanogaster*, and *H. sapiens* genomes with a threshold of ≥ 35% sequence identity and a 10e-5 e-value. B. Average methylation levels of species-specific genes and genes that share homology to *Ac. pisum*, *D. melanogaster*, and *H. sapiens*. The numbers located underneath bars represent the number of genes that belong to each group. Different letters above bars indicate significant differences between gene groups (Fisher's exact test $p < 0.05$). Error bars ±2 SE.

Figure 3-5. Functional analysis of genes according to their methylation levels. The genes were divided into 10 groups by their average methylation levels. Gene ontology (GO) annotations were used to classify genes into functional groups. The top ten GO functional group categories representing the 10 most abundant *Aphis glycines* gene categories are shown above.

Figure 3-1.

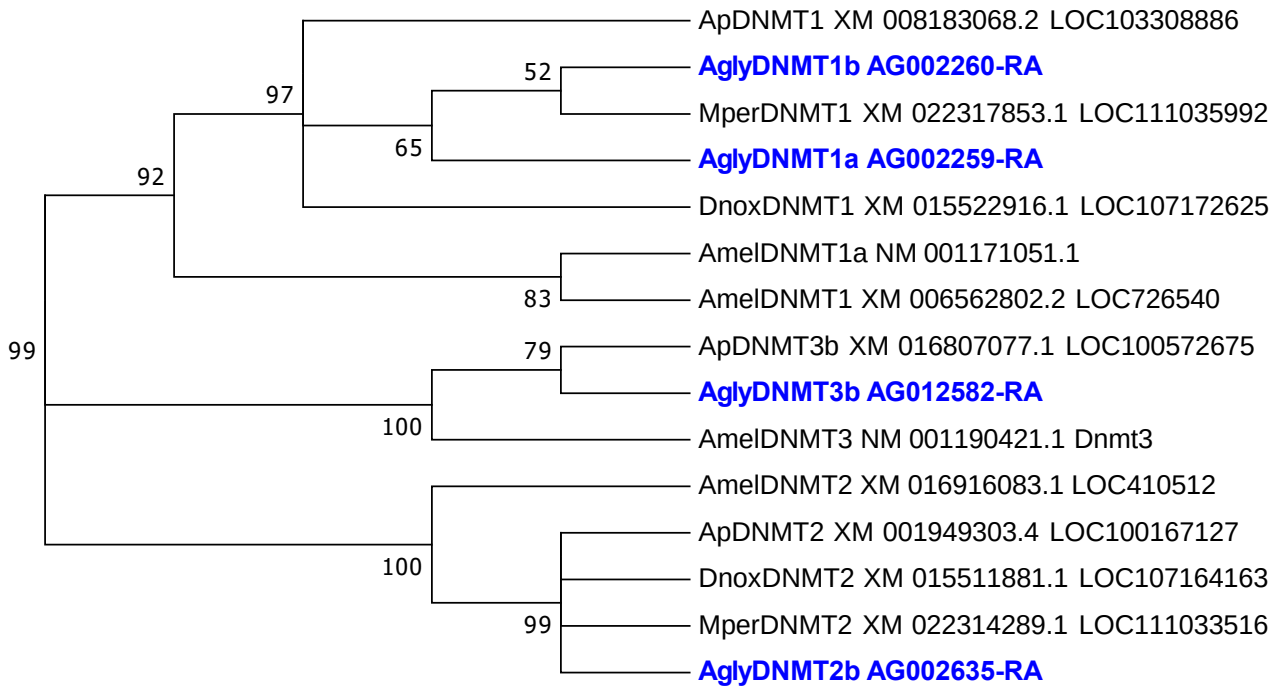
ApDNMT1 XM 008183068.2 LOC103308886
AglyDNMT1b AG002260-RA
MperDNMT1 XM 022317853.1 LOC111035992
AglyDNMT1a AG002259-RA
DnoxDNMT1 XM 015522916.1 LOC107172625
AmelDNMT1a NM 001171051.1
AmelDNMT1 XM 006562802.2 LOC726540
ApDNMT3b XM 016807077.1 LOC100572675
AglyDNMT3b AG012582-RA
AmelDNMT3 NM 001190421.1 Dnmt3
AmelDNMT2 XM 016916083.1 LOC410512
ApDNMT2 XM 001949303.4 LOC100167127
DnoxDNMT2 XM 015511881.1 LOC107164163
MperDNMT2 XM 022314289.1 LOC111033516
AglyDNMT2b AG002635-RA
52
97
65
92
83
99
79
100
100
99

Figure 3-2.

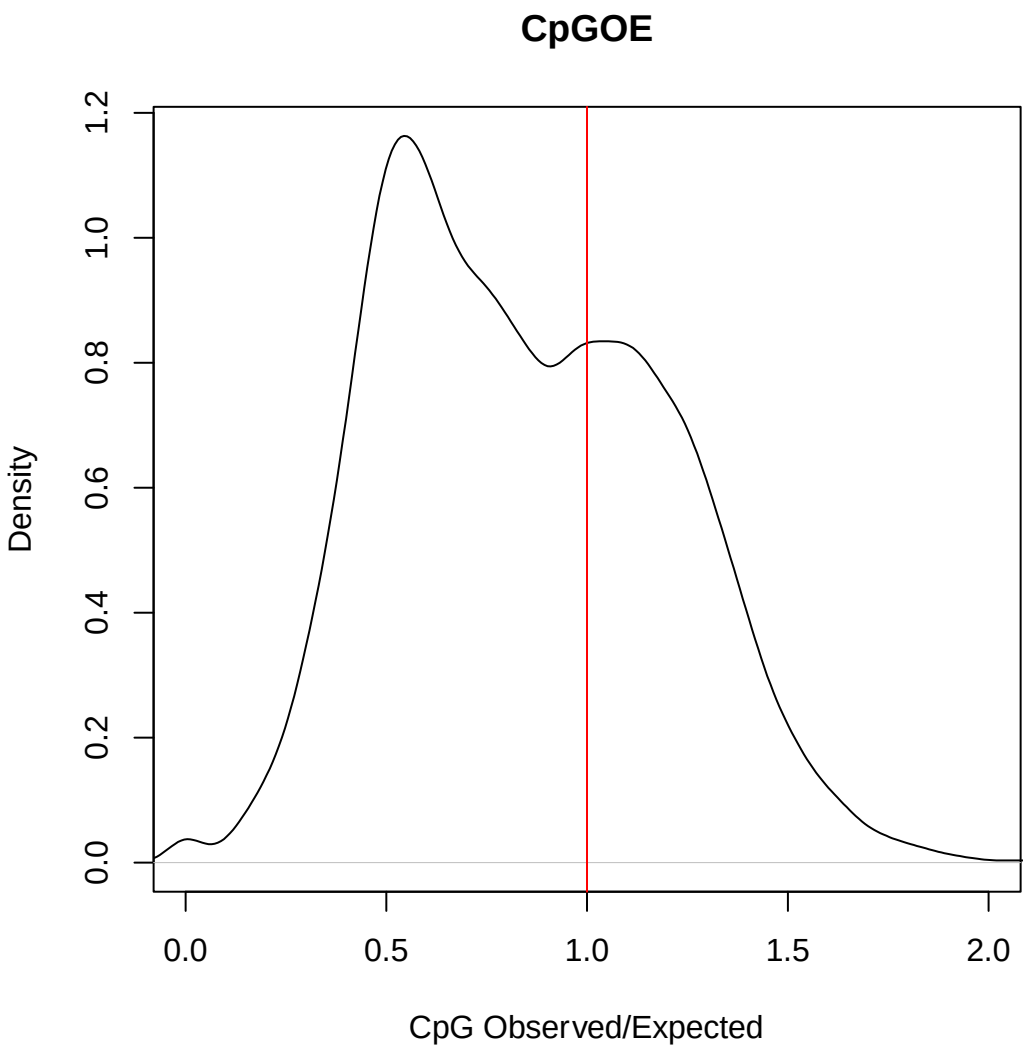
CpGOE
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
CpG Observed/Expected

Figure 3-3.

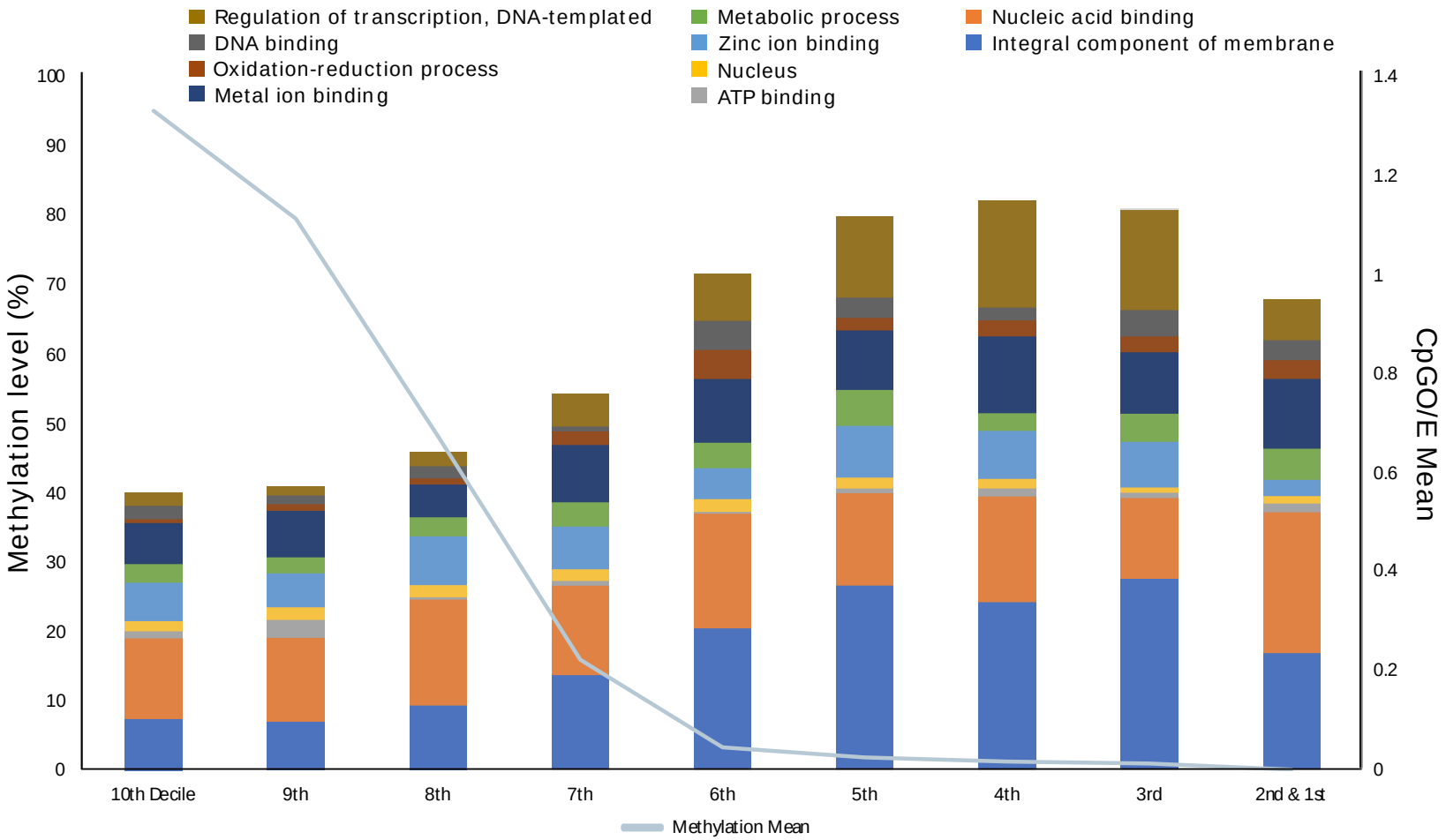
Regulation of transcription, DNA-templated
DNA binding
Oxidation-reduction process
Metal ion binding
Metabolic process
Zinc ion binding
Nucleus
ATP binding
Nucleic acid binding
Integral component of membrane
Methylation level (%)
CpGO/E Mean
100
90
80
70
60
50
40
30
20
10
0
1.4
1.2
1
0.8
0.6
0.4
0.2
0
10th Decile
9th
8th
7th
6th
5th
4th
3rd
2nd & 1st
Methylation Mean

Figure 3-4.

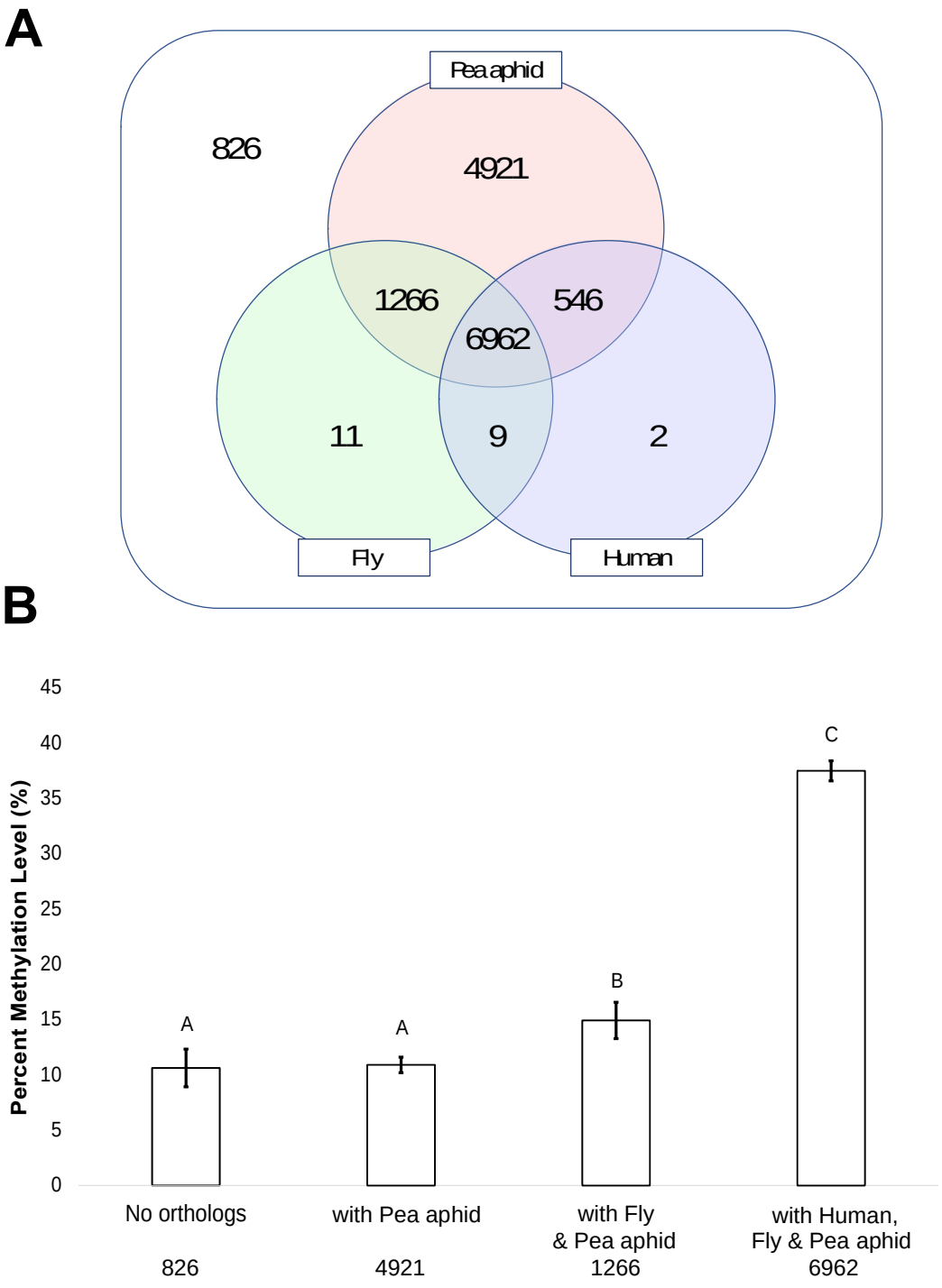
A
Pea aphid
826
4921
1266
546
6962
11
9
2
Fly
Human
B
Percent Methylation Level (%)
45
40
35
30
25
20
15
10
5
0
A
A
B
C
No orthologs
826
with Pea aphid
4921
with Fly
& Pea aphid
1266
with Human,
Fly & Pea aphid
6962

Figure 3-5.

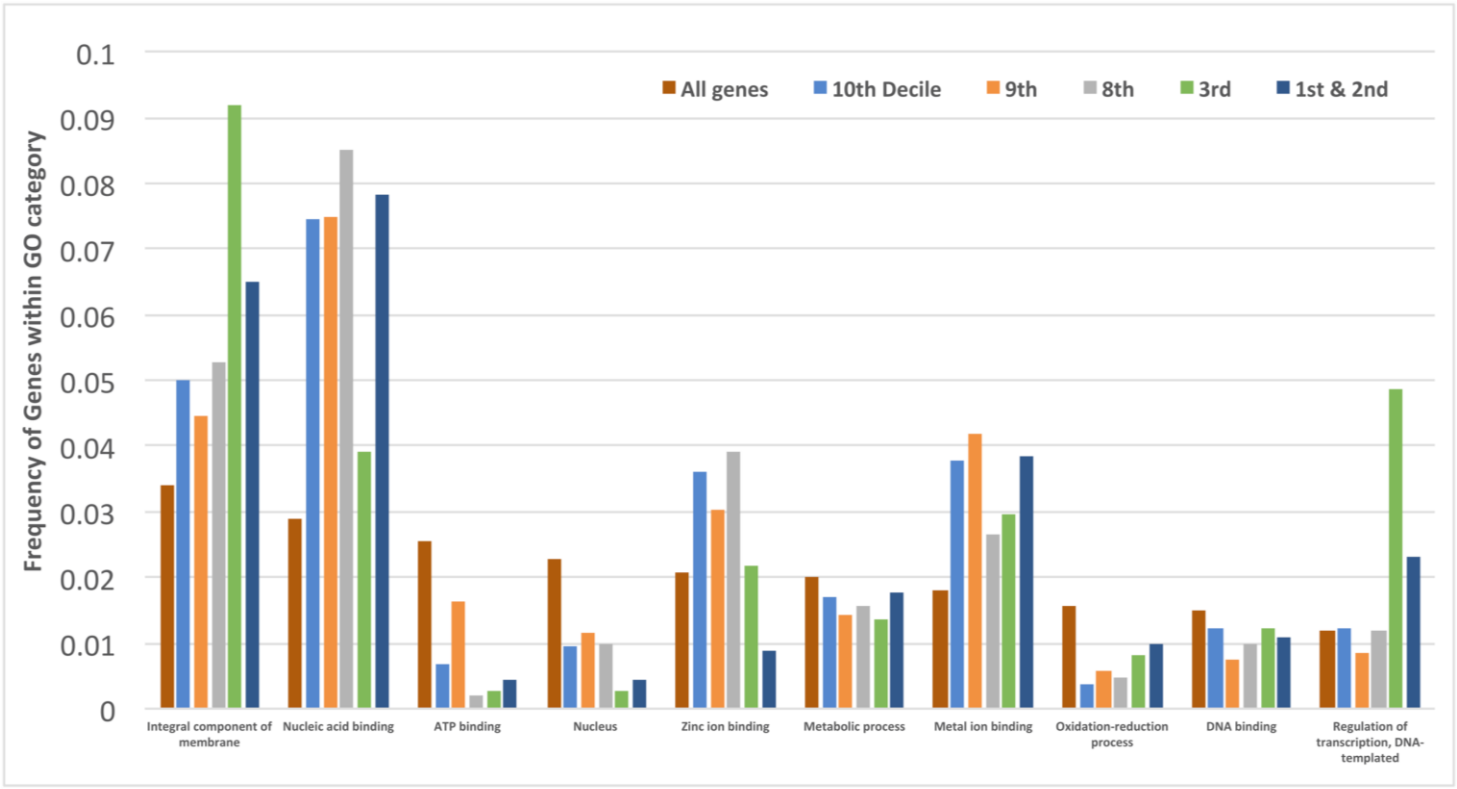
All genes
10th Decile
9th
8th
3rd
1st & 2nd
Frequency of Genes within GO category
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Integral component of membrane
Nucleic acid binding
ATP binding
Nucleus
Zinc ion binding
Metabolic process
Metal ion binding
Oxidation-reduction process
DNA binding
Regulation of transcription, DNA-templated

Table 3-1. Average CpG methylation levels of *A. glycines* cytochrome P450 genes and their corresponding orthologs in *A. pisum*, *D. melanogaster*, and *H. sapiens*.

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DmelOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG005349*	uncharacterized	94.4	ACYPI010073	alphaTub84B-PA	NM_080386.3	14	14(14)
AG012883	cytochrome p450 cyp12a2-like	92.2	ACYPI003070	Cyp49a1-PE	XM_024452838.1	17	17(17)
AG000379	cytochrome p450 mitochondrial	73.2	ACYPI000973	sad-PA	XM_005260304.5	2	2(2)
AG015040	cytochrome p450	17.8	ACYPI56606	Cyp6a13-PA	NM_001202855.2	37	8(5)
AG017396	uncharacterized protein loc100574879	5.1	ACYPI37690	N/A	N/A	17	1(1)
AG015634	probable cytochrome p450 305a1	4.8	ACYPI009807	Cyp305a1-PA	NM_024514.4	12	1(1)
AG003118	probable cytochrome p450 6a13	3.3	ACYPI56631	Cyp6a13-PA	NM_001202855.2	5	0(0)
AG001124	cytochrome p450 4c1-like	2.9	ACYPI008643	Cyp4g15-PC	NM_207352.3	27	0(0)
AG005414*	probable cytochrome p450 mitochondrial	2.7	ACYPI005596	Cyp301a1-PA	NM_000785.3	67	2(2)
AG004557	probable cytochrome p450 305a1	2.5	ACYPI009807	Cyp303a1-PB	NM_183075.2	35	0(0)
AG014755	cytochrome p450 6k1-like	2.5	ACYPI004952	Cyp6a13-PA	NM_000765.4	61	0(0)
AG009373*	cytochrome p450 6bq11	2.3	ACYPI47738	Cyp6a14-PE	NM_001202855.2	18	0(0)
AG005036	probable cytochrome p450 6a14	2.2	ACYPI003740	Cyp6a19-PA	NM_000765.4	91	1(1)
AG003107*	probable cytochrome p450 6a13	2.1	ACYPI56631	Cyp6a13-PA	NM_057095.2	37	0(0)
AG000557	cytochrome p450 mitochondrial	1.9	ACYPI085958	dib-PA	NM_000785.3	76	0(0)
AG010227	cytochrome p450 4c1-like	1.9	ACYPI001913	Cyp4c3-PA	NM_207352.3	169	0(0)
AG002855*	probable cytochrome p450 6a13	1.7	ACYPI004952	Cyp6a13-PA	NM_001202855.2	41	0(0)
AG001122*	cytochrome p450 4c1-like	1.6	ACYPI008643	Cyp4g15-PC	XM_005262935.4	14	0(0)
AG004984	cytochrome p450 4c1-like	1.6	ACYPI002550	Cyp4aa1-PB	NM_207352.3	29	0(0)
AG000820	probable cytochrome p450 6a13	1.5	ACYPI008473	Cyp6a13-PA	N/A	12	0(0)
AG001156	cytochrome p450 4g15	1.5	ACYPI005113	Cyp4g15-PC	NM_207352.3	46	0(0)
AG007638	probable cytochrome p450 6a13	1.5	ACYPI008746	Cyp6a13-PA	NM_017460.5	13	0(0)
AG016855	cytochrome p450 4c1	1.5	ACYPI000608	Cyp312a1-PB	NM_207352.3	18	0(0)

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DmelOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG010228	cytochrome p450 4c1-like	1.4	ACYPI003803	Cyp4c3-PA	NM_207352.3	20	0(0)
AG000658*	uncharacterized	1.3	ACYPI000990	Cyp6a14-PE	NM_000765.4	210	0(0)
AG000965	probable cytochrome p450 6a13	1.3	ACYPI004952	Cyp6a13-PA	NM_001202855.2	56	0(0)
AG010565	cytochrome p450 4c1-like	1.3	ACYPI069651	Cyp4c3-PA	NM_207352.3	190	0(0)
AG011621*	probable cytochrome p450 6a13	1.3	ACYPI002699	Cyp6a13-PA	NM_057095.2	80	0(0)
AG012997	uncharacterized	1.3	ACYPI066175	N/A	N/A	44	0(0)
AG001390	probable cytochrome p450 305a1	1.2	ACYPI009602	Cyp305a1-PA	NM_000775.3	26	0(0)
AG003949*	probable cytochrome p450 6a13	1.1	ACYPI004295	Cyp6a2-PA	NM_000765.4	69	0(0)
AG005793	probable cytochrome p450 6a13	1.1	ACYPI006882	Cyp6a13-PA	NM_000765.4	92	0(0)
AG014886	retrovirus-related pol polyprotein from transposon	1.1	ACYPI25751	N/A	N/A	101	0(0)
AG015545*	probable cytochrome p450 6a13	1.1	ACYPI002836	Cyp6a2-PA	NM_001202855.2	440	0(0)
AG004983	cytochrome p450 4c1-like	1	ACYPI005978	Cyp4g15-PC	NM_207352.3	78	0(0)
AG014876	cytochrome p450 307a1	1	ACYPI001519	spo-PC	NM_000761.4	218	0(0)
AG004772	probable cytochrome p450 6a13	0.9	ACYPI004295	Cyp6a2-PA	NM_057095.2	64	0(0)
AG005943*	cytochrome p450 18a1	0.8	ACYPI004722	Cyp18a1-PB	NM_001025161.2	114	0(0)
AG005942	cytochrome p450 306a1	0.7	ACYPI006623	phm-PB	NM_000775.3	40	0(0)
AG017397	cytochrome p450 9e2-like	0.5	ACYPI47651	Cyp6a13-PA	N/A	49	0(0)
AG003517	ecdysone 20-monooxygenase isoform x2	0.3	ACYPI008228	shd-PD	XM_017027692.2	53	0(0)
AG000766	cytochrome p450 307a1-like	0.2	ACYPI000716	spo-PC	XM_005262717.2	77	0(0)
AG000559	uncharacterized	0	ACYPI070222	N/A	N/A	5	0(0)
AG000966	probable cytochrome p450 6a13	0	ACYPI004952	Cyp6a13-PA	XM_017012569.1	23	0(0)
AG003108	probable cytochrome p450 6a13	0	ACYPI000761	Cyp6a18-PB	N/A	7	0(0)
AG009673	probable cytochrome p450 303a1	0	ACYPI003371	Cyp303a1-PB	NM_000775.3	4	0(0)
AG010226*	cytochrome p450 4v2	0	ACYPI007609	Cyp4c3-PA	NM_207352.3	19	0(0)
AG011892	uncharacterized	0	ACYPI44663	N/A	N/A	1	0(0)
AG013706	cytochrome p450 4c1-like	0	ACYPI008643	Cyp4g15-PC	XM_005262935.4	11	0(0)

[a]Average percent methylation levels of each gene. [b]Ortholog found in *A. pisum* genome. [c]Ortholog found in *D. melanogaster* genome. N/A indicates no homologs were found. [d]Ortholog found in *H. sapiens* genome. N/A indicates no homologs were found. [d]Number of CpG sites displaying a site-specific methylation level of ≥ 50% (i.e. ≥ 50% of the cytosines mapped to a given CpG site were methylated); Number of methylated CpG sites ≥ 50% found in exons are represented within the parenthesis. [*]Genes that showed differential expression when aphids feed on Rag1 resistant compared to susceptible soybean plants (Bansal et al., 2014).

Table 3-2. Average CpG methylation levels of *A. glycines* effector genes and their corresponding orthologs in *A. pisum*, *D. melanogaster*, and *H. sapiens*.

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DemlOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG004899	stromal cell-derived factor 2	100	ACYPI007065	CG11999-PA	NM_006923.3	3	3(3)
AG006652	mitochondrial import inner membrane translocase subunit tim14	85.58201	ACYPI007001	CG7394-PC	NR_033721.1	6	5(5)
AG007788	neutral alpha-glucosidase ab isoform x1	83.49916	ACYPI009457	GCS2alpha-PC	XM_017017412.1	19	17(17)
AG011783	aael014548- partial	76.00317	ACYPI003960	Jafrac2-PC	NM_006406.1	12	10(10)
AG015985	ferritin subunit	74.48142	ACYPI009403	Fer1HCH-PD	NM_031894.2	10	8(8)
AG001735	hypoxia up-regulated protein 1 isoform x1	71.38703	ACYPI008996	CG2918-PB	XM_017017097.1	22	17(17)
AG006612	uncharacterized	70.159004	ACYPI001938	Tango1-PD	N/A	24	19(19)
AG004337	carboxypeptidase e-like	69.79875	ACYPI001238	svr-PN	NM_001873.3	32	26(15)
AG002467	mesencephalic astrocyte-derived neurotrophic factor homolog	65.404144	ACYPI008001	Manf-PA	NM_006010.5	9	6(6)
AG010733	coiled-coil domain-containing protein 47	61.06704	ACYPI006626	CG17593-PB	NM_020198.2	19	11(11)
AG007896	lachesin	61.06526	ACYPI008756	Lac-PA	N/A	33	21(19)
AG014950	uncharacterized	46.623287	ACYPI083041	Cont-PA	NM_020872.2	50	24(24)
AG010755	tm2 domain-containing protein cg10795	41.77778	ACYPI007262	CG10795-PA	NM_032027.2	5	1(1)
AG003194	endoplasmic reticulum resident protein 29	34.367424	ACYPI000995	wbl-PB	NM_006817.3	20	7(7)
AG004644	dnaj homolog subfamily b member 11	33.805088	ACYPI001453	shv-PA	NM_016306.5	14	5(4)
AG002109	protein disulfide-isomerase a6	19.885279	ACYPI008926	CaBP1-PB	NM_001282705.1	92	18(18)
AG011655	endoplasmin	18.421053	ACYPI009915	Gp93-PA	NM_003299.2	38	7(7)
AG011557	uncharacterized	18.353746	ACYPI008002	stau-PB	NM_001164380.1	47	9(8)
AG002322	protein disulfide-isomerase a3	17.652348	ACYPI005594	ERp60-PA	NM_005313.4	26	4(4)
AG010706	uncharacterized	9.749253	ACYPI006124	P58IPK-PA	NM_006260.4	61	5(4)
AG005025	calreticulin	9.1317835	ACYPI002622	Calr-PB	NM_004343.3	61	5(5)

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DemlOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG002737	uncharacterized	3.3641975	ACYPI004737	N/A	N/A	36	0(0)
AG005369	glucose dehydrogenase	2.840963	ACYPI000986	Gld-PA	NM_018397.4	35	0(0)
AG001590	uncharacterized protein loc100302326 precursor	2.7389903	ACYPI43360	N/A	N/A	19	1(1)
AG005212	uncharacterized	2.4908876	ACYPI000364	N/A	N/A	25	0(0)
AG006367	thrombin-like enzyme cerastocytin isoform x1	2.125192	ACYPI38795	CG33160-PA	NM_182559.2	25	0(0)
AG011982	circumsporozoite protein	2.123077	ACYPI004904	N/A	XR_001739283.1	25	0(0)
AG000320	odorant-binding protein partial	2.0986123	ACYPI006147	N/A	N/A	50	0(0)
AG014474	acetylcholinesterase	1.8826135	ACYPI009886	Ace-PB	NM_000055.3	43	0(0)
AG008799	uncharacterized	1.8518518	ACYPI002746	N/A	N/A	6	0(0)
AG009499	protein takeout	1.8170044	ACYPI001245	CG2016-PD	N/A	94	0(0)
AG001014	uncharacterized	1.7857143	ACYPI081664	N/A	N/A	8	0(0)
AG017236	uncharacterized	1.7486802	ACYPI008224	N/A	N/A	26	0(0)
AG007882	uncharacterized	1.7212257	ACYPI005750	N/A	N/A	25	0(0)
AG007198	uncharacterized protein loc100166059	1.7165748	ACYPI006963	CG32032-PA	N/A	99	0(0)
AG003095	gelsolin	1.712963	ACYPI008158	Gel-PJ	NM_001353053.1	27	0(0)
AG006928	microsomal triglyceride transfer protein large subunit	1.6715358	ACYPI008792	Mtp-PA	NM_001300785.1	121	0(0)
AG001237	cuticle protein isoform b-like	1.604111	ACYPI45293	N/A	N/A	47	0(0)
AG003862	flocculation protein flo11-like	1.5334085	ACYPI001019	Mur89F-PC	N/A	149	1(0)
AG000920	peritrophin a	1.531035	ACYPI009786	Peritrophin-A-PB	N/A	163	1(1)
AG009502	uncharacterized	1.4638093	ACYPI000473	N/A	N/A	19	0(0)
AG005907	immediate early response 3-interacting protein 1	1.461039	ACYPI31539	CG32069-PA	NM_016097.4	22	0(0)
AG006088	cuticle protein 21	1.4324446	ACYPI000889	Cpr64Ad-PB	N/A	115	0(0)
AG009351	uncharacterized	1.4137253	ACYPI55148	N/A	N/A	213	1(1)
AG007466	trehalase-like	1.4112477	ACYPI002298	Treh-PF	NM_007180.2	155	0(0)
AG010599	serta domain-containing protein 2	1.3916502	ACYPI21883	N/A	N/A	300	1(1)
AG015933	superoxide dismutase	1.358411	ACYPI003921	Sod3-PF	NM_000454.4	145	0(0)

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DemlOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG006077	myosin light chain smooth muscle	1.3572104	ACYPI007076	Strn-Mlck-PV	XM_024453098.1	153	0(0)
AG012058	maltase 2-like	1.3560208	ACYPI000002	Mal-A4-PA	NM_000341.3	65	0(0)
AG000918	chondroitin proteoglycan-2-like	1.3514293	ACYPI001579	Gasp-PA	N/A	387	1(1)
AG000769	uncharacterized	1.2963581	ACYPI082701	N/A	N/A	104	0(0)
AG006624	uncharacterized	1.2947729	ACYPI000490	N/A	N/A	95	0(0)
AG001587	cathepsin b	1.2820513	ACYPI000003	CtsB1-PC	XM_006716245.3	36	0(0)
AG014673	uncharacterized	1.2630662	ACYPI006100	N/A	N/A	264	1(1)
AG007889	uncharacterized protein loc100160727 isoform x2	1.1953505	ACYPI002002	N/A	N/A	89	0(0)
AG007758	uncharacterized	1.1792098	ACYPI004654	frm-PH	NM_000090.3	66	0(0)
AG008704	uncharacterized protein loc100158783 isoform x1	1.1460885	ACYPI000227	verm-PG	NM_198455.2	91	0(0)
AG010712	uncharacterized	1.099839	ACYPI006788	CG5867-PA	N/A	24	0(0)
AG007148	venom serine carboxypeptidase	1.0826603	ACYPI001008	CG4572-PD	XM_011515437.1	285	0(0)
AG005703	uncharacterized	1.0620115	ACYPI005411	CG9572-PA	NM_000435.2	188	0(0)
AG007505	pancreatic triacylglycerol lipase	1.0464479	ACYPI009369	CG13282-PB	XM_011512530.3	64	0(0)
AG004044	sarcalumenin isoform x2	0.98466504	ACYPI001446	CG9297-PE	XM_017023527.1	59	0(0)
AG004432	uncharacterized	0.9794919	ACYPI082430	N/A	N/A	33	0(0)
AG004515	uncharacterized	0.86296266	ACYPI007532	N/A	N/A	93	0(0)
AG012784	uncharacterized protein loc100167701	0.8392866	ACYPI008471	N/A	N/A	155	0(0)
AG008787	glucose dehydrogenase	0.7518797	ACYPI000288	CG9512-PB	NM_018397.4	14	0(0)
AG001859	cuticle protein 6	0.7330171	ACYPI001278	Cpr92F-PA	N/A	281	0(0)
AG001408	repetitive proline-rich cell wall protein 2-like	0.71220076	ACYPI005320	CG30101-PA	XM_017015401.1	96	0(0)
AG014020	rr1 cuticle protein 2	0.6449493	ACYPI001775	Cpr49Aa-PB	N/A	27	0(0)
AG010207	uncharacterized	0.60262316	ACYPI35369	N/A	N/A	31	0(0)
AG007860	maltase 2-like	0.5963944	ACYPI002659	Mal-B2-PD	NM_000341.3	86	0(0)
AG001995	apolipophorins	0.59339374	ACYPI000422	apolpp-PD	N/A	50	0(0)
AG011103	protein yellow	0.570175	ACYPI000479	yellow-e-PA	N/A	94	0(0)

ID	Description	MethLvl[a]	ApisumOrtholog[b]	DemlOrtholog[c]	HumanOrtholog[d]	NumCpG[e]	MethCpG[f]
AG001228	pancreatic triacylglycerol lipase-like	0	ACYPI001479	CG7367-PC	NM_001011709.2	21	0(0)
AG001589	uncharacterized	0	ACYPI009919	N/A	N/A	5	0(0)
AG008486	apolipoprotein d-like	0	ACYPI003166	N/A	N/A	13	0(0)
AG011899	cuticle protein 7-like	0	ACYPI004074	CG34461-PB	N/A	8	0(0)
AG011902	cuticle protein 7-like	0	ACYPI086044	Cpr66Cb-PA	N/A	7	0(0)
AG016490	uncharacterized protein loc100164214 precursor	0	ACYPI005249	N/A	N/A	2	0(0)
AG017235	uncharacterized	0	ACYPI006346	N/A	N/A	1	0(0)
AG017842	uncharacterized	0	ACYPI24280	N/A	N/A	2	0(0)

[a]Average percent methylation levels of each gene. [b]Ortholog found in *A. pisum* genome. [c]Ortholog found in *D. melanogaster* genome. N/A indicates no homologs were found. [d]Ortholog found in *H. sapiens* genome. N/A indicates no homologs were found. [d]Number of CpG sites displaying a site-specific methylation level of ≥ 50% (i.e. ≥ 50% of the cytosines mapped to a given CpG site were methylated); Number of methylated CpG sites ≥ 50% found in exons are represented within the parenthesis.

Chapter 4: Conserved and lineage specific patterns of gene expression profiles within bacteriocytes of the green peach aphid (*Myzus persicae*)

Abstract

Most plant-sap feeding insects have obligate symbiotic relationships with maternally transmitted bacteria. The green peach aphid, *Myzus persicae*, requires its nutritional endosymbiont, *Buchnera aphidicola*, for the production of essential amino acids that the insect cannot produce by itself. Such endosymbionts are harbored in specialized insect cells called bacteriocytes. Using RNA-sequencing, I performed a cell-specific transcriptome analysis on the bacteriocytes of *M. persicae* and identified differentially expressed genes between bacteriocytes and other body cells not infected with the endosymbiont, *Buchnera*. My results reveal that the aphid genes involved in essential amino acid production are up-regulated in bacteriocytes when compared to body cells. I then compared these differentially expressed genes to the set of differentially expressed genes previously identified from the bacteriocytes of the model aphid species, *Acyrthosiphon pisum,* which resides in the same aphid tribe, *Aphidini,* as *M. persicae*. Based on my results, both *M. persicae* and *A. pisum* bacteriocytes share very similar expression profiles, however, lineage-specific differences in some key enzymes involved in the symbiosis include asparaginase and pseudouridine kinase.

Introduction

Many plant-sap feeding insects harbor maternally transmitted, obligate bacterial endosymbionts (Gil *et al.*, 2004; Wernegreen, 2002). The insect host provides a stable environment for the bacterial symbiont to live. In exchange the bacteria aid by supplementing the host's nutrient deficient diet, taking metabolomic precursors from the host and converting them into essential amino acids, vitamins, and co-factors needed for proper insect growth, development, and reproduction (Douglas, 1998). While many symbiotic bacteria that are not obligate for insect survival are located in the host's hemolymph or gut, some insect hosts, including most sap-feeders, have evolved specialized cells called bacteriocytes that house endosymbiotic bacteria within the body cavity of an insect host (Gill *et al.*, 2004). These cells are thought to have originally evolved from special adipocytes and extraembryonic follicular nuclei, although this still remains unclear (Braendle *et al*., 2003; Nakabachi *et al*., 2014). Bacteriocyte-associated endosymbiosis has been described in a number of insect families, and bacteriocytes have been found within fat body tissue, inside the epithelium of the midgut, in the hemolymph, or clustered together in an organ-like structure called bacteriome (Moran and Telang, 1998; Baumann, 2005; Feldhaar and Gross, 2009). While bacteriocytes have been described in a number of important insect super families, little is known about how conserved gene regulation of this specialized cell is with its endosymbiont for the biosynthesis of essential amino acids and vitamins.

One of the best studied models of insect-endosymbiont interactions is the pea aphid, *Acyrthosiphon pisum*, and its obligate endosymbiont bacterium, *Buchnera*

aphidicola. This integrated host-symbiont metabolism has been well characterized biochemically and genetically, identifying which enzymes are expressed by the aphid and by its symbiont (Hansen and Moran, 2011; Poliakov *et al.*, 2011; Nakabachi *et al*., 2014; Hansen and Moran, 2014). Within bacteriocytes the aphid metabolism is integrated with *Buchnera* for the production of essential amino acids. This complementary relationship is hypothesized to be regulated by the aphid host through transporters, the GS/GOGAT cycle, and genes that complement *Buchnera*'s essential amino acid pathways (Hansen and Moran, 2011; Poliakov *et al.,* 2011; Price *et al*., 2014). Moreover, a study by Kim *et al*. (2018) suggests that CpG methylation may play a role in the differential regulation of key symbiosis genes in the bacteriocytes when aphids feed on different host plant diets. For example, the aphid glutamine transporter (*ApGLNT1*), which incorporates glutamine into bacteriocytes is both differentially methylated and expressed between bacteriocytes and body cells of *A. pisum*. This transporter is inhibited by extracellular arginine, which is produced by *Buchnera*, and is also thought to be post-transcriptionally regulated (Price *et al*., 2015). Nevertheless, what regulatory mechanisms are conserved and lineage specific among different aphid species is still unclear.

Similar to *A. pisum*, the green peach aphid, *Myzus persicae*, belongs to the *Aphidini* tribe within the subfamily Aphididae: Aphidinae (Nováková *et al.*, 2013). While both aphid species are pests and have similar lifecycles their host plant range varies dramatically (Mathers *et al*., 2017). For example, *A. pisum* is acknowledged as a specialist of *Fabaceae* host plants, where sympatric aphid populations are divided into a number of different biotypes that specialize on specific *Fabaceae* species (Frantz *et al.,*

2005). In contrast, *M. persicae* differs from *A. pisum* in that it is a true generalist herbivore and can feed on 40 different plant families including many economically and agriculturally important crop species (Blackman and Eastop, 2000). Furthermore, rapid transcriptional plasticity of multigene families has allowed *M. persicae* clones to colonize up to 100 species of host plants without genetic specialization (Mathers *et al.*, 2017). Given this transcriptional plasticity in response to diverse host plant diets, it is of interest to compare the regulation of bacteriocytes between a closely related aphid specialist and generalist, *A. pisum* and *M. persicae,* respectively.

In this study, I performed cell-specific transcriptome analyses to identify genes that are differentially expressed between the bacteriocytes and symbiont-free cells obtained from body tissues of the green peach aphid, *M. persicae*. I then compared the set of differentially expressed genes in the bacteriocytes of *M. persicae* to the set of genes previously known to be differentially expressed in the bacteriocytes of *A. pisum* (Hansen and Moran 2011; Kim *et al.*, 2018). Here I identify conserved and lineage specific mechanisms of aphid host regulation in bacteriocytes when aphids feed on the same host plant species, *Vicia fava*.

Materials and Methods

Aphid rearing, RNA extraction, and RNA sequencing

A genetically homogenous strain of *Myzus persicae* (Sulzer) from Medina-Ortega and Walker (2015) was divided into three sub-lines. The three sub-lines were reared on the same developmental stage of fava bean (*Vicia fava*; three weeks after germination) for

over 20 generations. Approximately 200 aphids that were 8 days old were dissected from each sub-line to co-collect both bacteriocytes and other body cells without *Buchnera*. Pooled RNA of bacteriocytes and body cells were extracted using the *Quick*-RNA Microprep kit (Zymo Research, Irvine, CA, USA). Extracted RNA samples were treated with DNase I and purified with RNA Clean & Concentrator kit (Zymo Research). Illumina library preparation and sequencing were conducted by the University of California, Davis Genome Center. Libraries were sequenced as paired-end 150-mers using Illumina HiSeq4000. Reads for all RNA-Seq samples were submitted to the Sequence Read Archive of the National Center for Biotechnology Information (NCBI).

RNA-Seq data analysis

Sequenced RNA reads were quality-checked with FASTQC v.0.11.8 (Andrews, 2010), and adapters and low-quality reads were trimmed using Trimmomatic v.0.36 (Bolger *et al*., 2014). The trimmed reads were aligned using HISAT2 v.2.1.0 (Kim *et al*., 2015) against the *M. persicae* Clone O whole genome assembly and annotation v.1.1 (Gauthier *et al.*, 2007), which was retrieved from AphidBase (Legeai *et al.*, 2010). The mapped reads for each gene and transcript were quantified as raw read counts using StringTie v.1.3.5 (Pertea *et al*., 2015). Differential expression of transcripts between bacteriocytes and body cells was determined using edgeR v.3.26.0 (Robinson *et al.*, 2010). Statistical significance for differentially expressed genes was determined if FDR adjusted p-values were ≤ 0.05 with at least 1.5-fold read count. Gene Set Enrichment Analysis (GSEA) (Subramanial *et al*., 2005) was used to determine which pathways were up- or down-regulated at the normalized $p < 0.01$, as described in Kim *et al*. (2018). The

A. pisum genome annotation v.2.0 (International Aphid Genomics Consortium, 2010) was used to identify homologs within the *M. persicae* genome using reciprocal best BLAST hits (Camacho *et al*., 2009) with the cut-off criteria ≥ 35% sequence identity and the e-value ≥ 10^{-5}. Lineage-specific genes of *M. persicae* were determined by having no hits, that meet the latter thresholds.

Results

A total of 62.5% and 82.9% of the RNAseq reads mapped to the *M. persicae* Clone O genome for bacteriocytes and body cells, respectively (Table 4-1). The remainder of the reads mapped primarily to the endosymbiont *Buchnera* (Table 4-1). A total of 1,578 genes were up-regulated in *M. persicae* bacteriocytes compared to body cells, while 2,938 genes were up-regulated in body cells compared to bacteriocytes (fold change ≥ 1.5x; normalized $p < 0.05$). To characterize the putative function of genes that are differentially regulated in bacteriocytes compared to body cells, I used Gene Set Enrichment Analysis (GSEA) (Subramania *et al*., 2005) using Gene Ontology (GO) terms. In bacteriocytes, 75 GO groups were up-regulated compared to body cells (Table 4-2A). The top ten up-regulated GO groups in bacteriocytes in descending order based on the GSEA normalized enrichment score were structural constituent of ribosome, ribosome, nucleosome, rRNA binding, protein heterodimerization activity, nucleosome assembly, chromosome, ribonucleoprotein complex, RNA-directed DNA polymerase activity, and RNA-dependent DNA biosynthetic process (Table 4-2A). I also identified 68 GO groups that were down-regulated in bacteriocytes compared to body cells (Table 4-

2B). Among such GO groups, the top ten groups in descending order based on the GSEA normalized enrichment score were structural constituent of cuticle, structural molecule activity, synapse, chitin binding, postsynaptic membrane, chitin metabolic process, extracellular ligand-gated ion channel activity, carbohydrate derivative binding, oxidoreductase activity, and G protein-coupled receptor signaling pathway. These results suggest that both transcription and translation of genes are much more active in bacteriocytes relative to body cells.

To investigate the integrated aphid metabolism with its endosymbiont *Buchnera* within bacteriocytes, I examined the gene expression profiles of key metabolic genes and transporters that Kim *et al*. (2018) identified as important in amino acid biosynthesis in *A. pisum*. All 37 amino acid biosynthesis genes and 76 transporter genes from *A. pisum* had homologs present in the *M. persicae* genome. Out of 37 aphid amino acid biosynthesis genes that have been characterized previously in *A. pisum* bacteriocytes, 24 genes were up-regulated in the bacteriocytes of *M. persicae* relative to body cells, and three genes were down-regulated in bacteriocytes compared to body cells (Table 4-3). Moreover, out of 76 active transporter genes that have been characterized previously in *A. pisum* bacteriocytes, 57 genes were up-regulated in bacteriocytes relative to body cells, whereas five transporter genes were down-regulated (Table 4-4). Furthermore, similar to *A. pisum* both Glutamine synthetase (*GS*) and Glutamate synthase (*GLTS*), both of which are key genes in GS/GOGAT cycle and recycle ammonia into glutamine and glutamate (Hansen and Moran, 2011), are significantly up-regulated in bacteriocytes of *M. persicae* relative to body cells (Figure 4-1; Table 4-3).

To determine which metabolic pathways were up and down-regulated in *M. persicae* bacteriocytes in comparison to *A. pisum*, I identified KEGG pathways for each *M. persicae* gene. There are 22 KEGG pathways that are up-regulated in the *M. persicae* bacteriocytes relative to body cells, and four KEGG pathways that are down-regulated (Table 4-5). Among the 22 up-regulated pathways, seven pathways were commonly up-regulated in bacteriocytes of both *A. pisum* and *M. persicae* compared to body cells (Table 4-5A). These pathways are ribosome (API03010), pyrimidine metabolism (API00240), DNA replication (API03030), RNA polymerase (API03020), oxidative phosphorylation (API00190), homologous recombination (API03440), and purine metabolism (API00230).

Using a reciprocal best-blast hit approach, I identified 5,428 lineage specific genes in the *M. persicae* genome. Out of 5,428 lineage specific genes, 2,667 genes were expressed in at least two out of three biological replicates. Of those, 50 and 31 genes are significantly up-regulated and down-regulated respectively in bacteriocytes relative to body cells (Table 4-6). A total of 46% and 39% of these genes that are up- and down-regulated, respectively, in bacteriocytes compared to body cells have unknown function.

Discussion

Results from my study reveal that bacteriocytes of *M. persicae* have both conserved and lineage specific regulation when feeding on the same host plant species as *A. pisum*. Similar to *A. pisum,* bacteriocyte pathways that are related to gene transcription and translation are up-regulated much higher in *M. persicae* bacteriocytes compared to body

cells (Table 4-2A). Also, similar to *A. pisum* both genes involved in the GS/GOGAT cycle are significantly up-regulated in *M. persicae* bacteriocytes relative to body cells. The GS/GOGAT cycle in *A. pisum* bacteriocytes is hypothesized to be important for this nitrogen limited symbiosis, because the enzymes *GS* and *GOGAT* recycle waste ammonia into amino donors for the production of essential amino acids (Hansen and Moran, 2011). Moreover, similar to *A. pisum*, all of *M. persicae*'s genes that complement *Buchnera*'s essential amino acid biosynthesis pathways are up-regulated in bacteriocytes compared to body cells except for aspartate transaminase (E.C.2.6.1.1; 000034170), which enables interconversion between glutamate and aspartate as well as the terminal step of phenylalanine biosynthesis (Figure 4-1) (Hansen and Moran, 2011).

As observed in Price *et al.* (2014) as well as in Kim *et al*. (2018) in *A. pisum*, *ApGLNT1*, which imports glutamine into bacteriocytes, is significantly up-regulated in the bacteriocytes relative to body cells (Figure 4-1). This glutamine transporter has been previously suggested to be the main transporter responsible for glutamine uptake into bacteriocytes when arginine is limited (Price *et al.*, 2010). Such imported glutamine may be utilized in the GS/GOGAT cycle and is converted into glutamate, which is then used as an amino donor for the production of other essential amino acids.

Another interesting result from my data is the up-regulation of the gene involved in the uracil salvage pathway. Since *Buchnera* is unable to produce its own uracil, it depends on the host for the production of uracil. Contrary to the down-regulation of pseudouridine kinase (LOC100165128) in the bacteriocytes of the pea aphid feeding on fava plants relative to body cells (Kim *et al*., 2018), the *M. persicae* homolog for this

enzyme (000118100) is significantly up-regulated in the bacteriocytes compared to body cells. Such up-regulation of pseudouridine kinase was only observed in bacteriocytes of alfalfa feeding pea aphids (Kim *et al*., 2018). This result suggests that compared to *A. pisum*, *M. persicae* bacteriocytes requires more uracil for either *Buchnera* and/or the aphid when fed on fava.

Compared to the *Buchnera* strain found in the pea aphid, *Buchnera* of *M. persicae* possesses 21 additional genes including asparaginase, which converts asparagine and ammonia to aspartate (Jiang *et al*., 2013). The *M. persicae* genome has two copies of asparaginase (000052870 and 000113450), both of which are up-regulated in bacteriocytes compared to body cells less than two-fold (Figure 4-1; Table 4-3). This is much lower than previously observed for *A. pisum* on the same host plant where asparaginase (LOC100164179) is up-regulated over 23 times higher in bacteriocytes compared to body cells (Kim *et al*., 2018). In turn, instead of the aphid host, asparaginase encoded within *Buchnera* from *M. persicae* may play a key role in ammonia recycling and aspartate production in bacteriocytes in contrast to the pea aphid -*Buchnera* symbioses. Interestingly, *Buchnera* of the bluegreen aphid, *Acyrthosiphon kondoi*, is the only known *Buchnera* strain that encodes asparaginase (Jiang *et al*., 2013). A previous study showed that three dominant non-essential amino acids in fava bean phloem sap are asparagine, glutamine, and serine. Asparagine takes up more than 45% of the total amino acid concentration from fava bean phloem sap (Sandström and Pettersson, 1994). However, a labeled isotope study in *A. pisum* on artificial diet suggests that such non-essential amino acids transported into bacteriocytes most likely are synthesized *de novo*

by aphid-encoded enzymes (Haribal and Jander, 2015). However, it is unknown if this phenomenon occurs in *M. persicae* as well.

In summary, my findings indicate that *M. persicae* bacteriocytes have very similar gene expression profiles to the *A. pisum* bacteriocytes. However, lineage specific gene expression patterns in this generalist aphid species includes asparaginase and peudouridine kinase. Future studies are needed to investigate how such tissue-specific gene expression patterns are regulated and if DNA methylation is involved as suggested in Kim *et al*. (2018) in *A. pisum*.

Figure 4-1. Differentially expressed *M. persicae* amino acid metabolism genes between bacteriocytes and body cells.

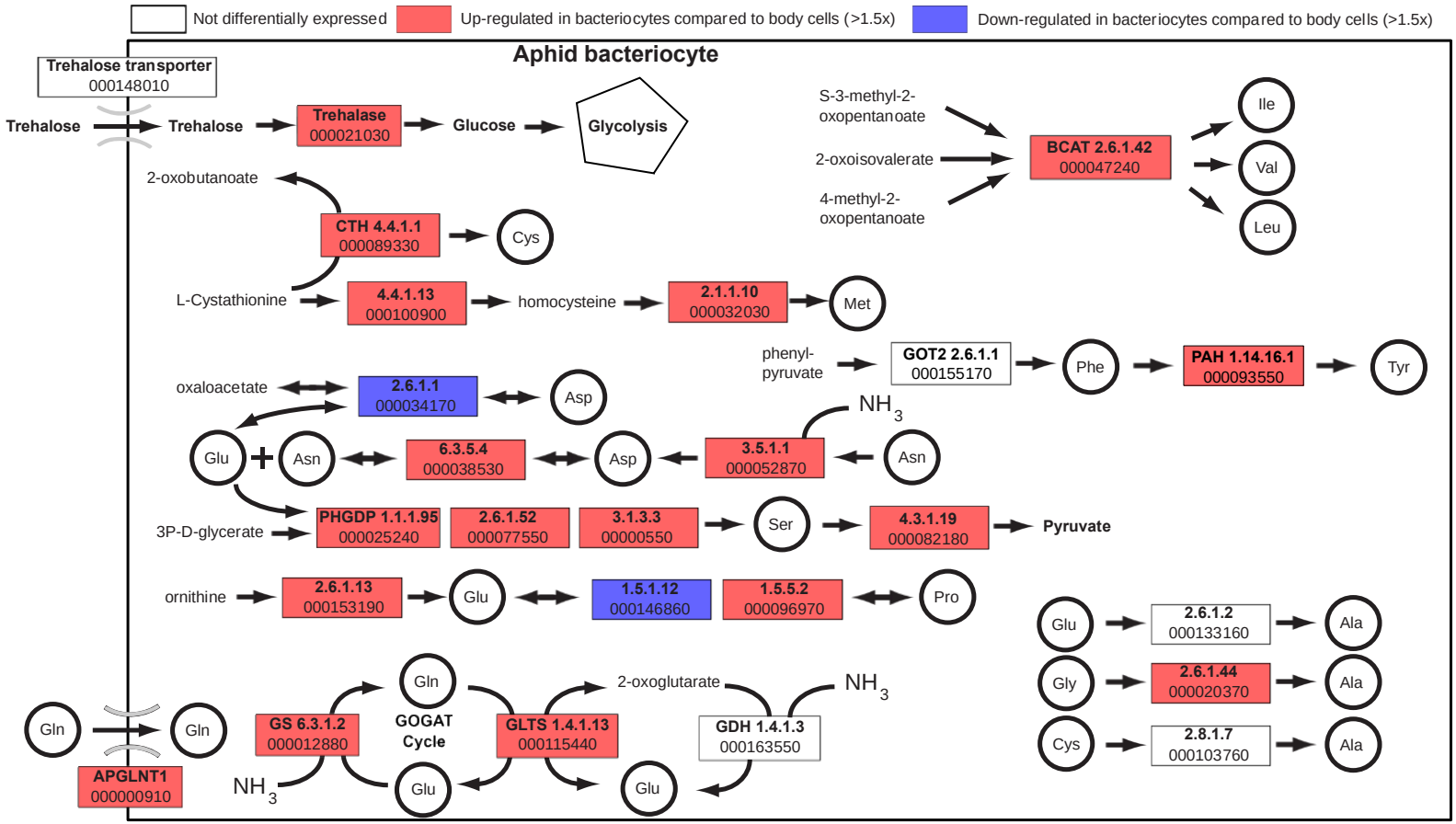

Table 4-1. RNA-Seq results and overall mapping rates

RNA-Seq samples	Total reads	Paired reads	Mapped reads (paired)	Unpaired reads	Mapped reads (unpaired)	Overall alignment rate	Alignment rate to *Buchnera*
MP1 Bacteriocytes	57,975,178	41,976,455	27,013,852	15,998,723	10,530,759	68.04%	25.19%
MP1 Body Cells	61,267,690	43,403,664	35,200,554	17,864,026	15,117,611	86.16%	1.62%
MP2 Bacteriocytes	59,028,122	39,482,824	18,235,342	19,545,298	8,773,269	48.39%	44.13%
MP2 Body Cells	55,216,125	39,790,326	26,354,498	15,425,799	11,704,280	74.09%	1.05%
MP3 Bacteriocytes	44,507,411	30,036,619	20,143,940	14,470,792	10,292,357	71.06%	20.24%
MP3 Body Cells	55,353,883	38,599,434	32,055,709	16,754,449	14,784,680	88.37%	1.21%

MP1, MP2, and MP3 denotes three sub-lines of *Myzus persicae* reared on fava bean plants.

Table 4-2A. Top 10 up-regulated Gene Ontology (GO) groups in *M. persicae* bacteriocytes compared to body cells based on normalized enrichment score from Gene Set Enrichment Analysis (GSEA).

Name	Description	SIZE[a]	NES[b]	NOM p-val[c]
GO:0003735	Structural constituent of ribosome	137	2.0879283	0.0
GO:0005840	Ribosome	178	1.902873	0.0
GO:0000786	Nucleosome	62	1.8507292	0.0
GO:0019843	rRNA binding	16	1.8498222	0.0
GO:0046982	Protein heterodimerization activity	73	1.8072857	0.0
GO:0006334	Nucleosome assembly	38	1.7873666	0.0
GO:0005694	Chromosome	153	1.7867577	0.0
GO:0030529	Ribonucleoprotein complex	241	1.7622097	0.0
GO:0003964	RNA-directed DNA polymerase activity	154	1.7596092	0.0
GO:0006278	RNA-dependent DNA biosynthetic process	154	1.7515945	0.0

Table 4-2B. Top 10 down-regulated Gene Ontology (GO) groups in *M. persicae* bacteriocytes compared to body cells based on normalized enrichment score from Gene Set Enrichment Analysis (GSEA).

Name	Description	SIZE[a]	NES[b]	NOM p-val[c]
GO:0042302	Structural constituent of cuticle	93	-3.5752213	0.0
GO:0005198	Structural molecule activity	158	-3.5505188	0.0
GO:0045202	Synapse	40	-2.8712983	0.0
GO:0008061	Chitin binding	40	-2.7951887	0.0
GO:0045211	Postsynaptic membrane	28	-2.7364106	0.0
GO:0006030	Chitin metabolic process	40	-2.7179766	0.0
GO:0005230	Extracellular ligand-gated ion channel activity	27	-2.6015377	0.0
GO:0097367	Carbohydrate derivative binding	50	-2.5442722	0.0
GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	42	-2.4865189	0.0
GO:0007186	G protein-coupled receptor signaling pathway	157	-2.450538	0.0

[a]Number of genes included in a gene set based on GO; [b]Normalized enrichment score; [c]Normalized p-value based on the size of the group.

Table 4-3. Expression levels of essential and non-essential amino acid aphid genes in bacteriocytes of *M. persicae* and *A. pisum*.

Essential Amino Acid Complementary Genes

E.C. Number	Name	Pathway[a]	Gene ID[b]	*A. pisum* homolog[c]	p-value[d]	Fold Change[e]	p-value_DK[f]	FC_DK[g]
4.3.1.19	Threonine ammonia-lyase	Ile	MYZPE13164_0_v1.0_000082180	LOC100165866	**0.0002**	**4.0644**	0.2453	0.3584
2.6.1.42	BCAT	Ile, Val, Leu	MYZPE13164_0_v1.0_000047240	LOC100167587	**0.0000**	**4.8460**	**0.0000**	**4.7930**
2.6.1.1	Aspartate transaminase, Got2	Phe, Tyr	MYZPE13164_0_v1.0_000155170	Got2	0.0248	1.4811	**0.0000**	**31.508**
2.6.1.1	Aspartate transaminase	Phe, Tyr	MYZPE13164_0_v1.0_000034170	LOC100161812	**0.0009**	**0.2792**	1.0000	1.0000
2.6.1.1	Aspartate transaminase	Phe, Tyr	MYZPE13164_0_v1.0_000168850	LOC100165255	0.0526	1.4528	0.8987	1.1104
1.14.16.1	Phenylalanine 4-monooxygenase, PAH	Tyr	MYZPE13164_0_v1.0_000093550	LOC100166971	**0.0000**	**31.9886**	**0.0000**	**8.5976**
4.4.1.1	Cystathionine gamma-lyase; CTH	Met	MYZPE13164_0_v1.0_000089330	LOC100159197	**0.0000**	**50.0204**	0.0475	0.6862
4.4.1.8	Putative cystathionine gamma-lyase 2	Met	MYZPE13164_0_v1.0_000009200	LOC100168016	**0.0161**	**2.0799**	0.5602	1.2336
4.4.1.8	Cystathionase-like	Met	MYZPE13164_0_v1.0_000089320	LOC100159560	0.0312	1.8305	**0.0000**	**197.03**
2.1.1.10	Homocysteine S-methyltransferase	Met	MYZPE13164_0_v1.0_000032030	LOC100168557	**0.0000**	**8.3957**	**0.0000**	**28.387**
2.1.1.10	Homocysteine S-methyltransferase	Met	MYZPE13164_0_v1.0_000049310	LOC100159972	0.4773	1.2332	0.6365	0.6066

Non-essential Amino Acid Aphid Genes

E.C. Number	Name	Pathway[a]	Gene ID[b]	*A. pisum* homolog[c]	p-value[d]	Fold Change[e]	p-value_DK[f]	FC_DK[g]
1.4.1.13	Glts	Glu	MYZPE13164_0_v1.0_000115440	LOC100158883	**0.0000**	**4.3816**	**0.0022**	**3.0617**
2.6.1.13	Ornithine aminotrasferase	Glu	MYZPE13164_0_v1.0_000153190	LOC100168809	**0.0000**	**9.9060**	**0.0000**	**8.6364**
1.5.1.2	Pyrroline-5-carboxylate reductase	Glu	MYZPE13164_0_v1.0_000153590	LOC100161005	**0.0004**	**2.3270**	0.0665	0.0799
1.5.1.12	1-Pyrroline-5-carboxylate dehydrogenase	Glu	MYZPE13164_0_v1.0_000146860	LOC100165747	0.0160	0.6298	0.9255	0.9527
1.4.1.3	GDH [NAD(P)+]	Glu	MYZPE13164_0_v1.0_000163550	LOC100169613	0.0887	0.6957	0.1765	0.5322
1.1.1.95	Phosphoglycerate dehydrogenase; Phgdp	Ser	MYZPE13164_0_v1.0_000025240	Phgdp	**0.0000**	**3.4298**	**0.0000**	**7.8181**
1.1.1.95	Phosphoglycerate dehydrogenase	Ser	MYZPE13164_0_v1.0_000053730	LOC100160394	0.0013	1.8308	0.3566	1.7868
2.6.1.52	Phosphoserine transaminase	Ser	MYZPE13164_0_v1.0_000077550	LOC100163589	**0.0000**	**3.8905**	**0.0000**	**22.5669**
2.6.1.52	Phosphoserine transaminase	Ser	MYZPE13164_0_v1.0_000077770	LOC100163589	**0.0000**	**3.6612**	**0.0000**	**22.5669**
3.1.3.3	Phosphoserine phosphatase	Ser	MYZPE13164_0_v1.0_000005550	LOC100158884	**0.0000**	**4.3984**	**0.0000**	**13.5146**
2.6.1.1	Got2	Asp	MYZPE13164_0_v1.0_000155170	Got2	0.0248	1.4811	**0.0000**	**31.5086**
2.6.1.1	Aspartate transaminase	Asp	MYZPE13164_0_v1.0_000034170	LOC100165255	**0.0009**	**0.2792**	0.8987	1.1104
6.3.5.4	Asparagine synthase (glytamine-hydrolyzing)	Asp	MYZPE13164_0_v1.0_000038530	LOC100160265	**0.0253**	**2.3235**	0.0537	0.0692
6.3.5.4	Asparagine synthase (glytamine-hydrolyzing)	Asp	MYZPE13164_0_v1.0_000015370	LOC100161762	0.0188	1.5543	0.1598	3.5330
6.3.1.2	GS	Gln	MYZPE13164_0_v1.0_000012880	Gs2	**0.0000**	**4.0477**	0.1068	1.4728
6.3.1.2	Glutamate-ammonia ligase	Gln	MYZPE13164_0_v1.0_000117760	LOC100165282	0.1746	2.1112	1.0000	1.0000
3.5.1.1	Asparaginase	Asn	MYZPE13164_0_v1.0_000052870	LOC100164179	0.0231	1.8097	**0.0000**	**23.8200**
3.5.1.1	Asparaginase	Asn	MYZPE13164_0_v1.0_000113450	LOC100158730	0.0024	1.8314	1.0000	1.0000
3.5.1.1	N-terminal asparagine amidohydrolase	Asn	MYZPE13164_0_v1.0_000046110	LOC100160095	0.0036	1.7285	1.0000	1.0000
2.6.1.2	Alanine aminotransferase 2	Ala	MYZPE13164_0_v1.0_000133160	LOC100164899	0.8937	0.9744	0.1335	1.8115
2.6.1.44	Alanine-glyoxylate transaminase	Ala	MYZPE13164_0_v1.0_000020370	LOC100163196	**0.0000**	**4.5962**	0.4334	0.4728
2.6.1.44	Alanine-glyoxylate transaminase	Ala	MYZPE13164_0_v1.0_000121380	LOC100165269	0.0875	1.4569	0.1801	3.6455

Non-essential Amino Acid Aphid Genes

E.C. Number	Name	Pathway[a]	Gene ID[b]	*A. pisum* homolog[c]	p-value[d]	Fold Change[e]	p-value_DK[f]	FC_DK[g]
4.1.2.5	Threonine aldolase	Gly	MYZPE13164_0_v1.0_000067680	LOC100161178	0.0060	1.8059	0.1961	2.2066
4.1.2.5	Threonine aldolase	Gly	MYZPE13164_0_v1.0_000084550	LOC100159142	0.0701	0.6235	0.3317	0.3039
1.4.4.2	Glycine dehydrogenase (decarboxylating)	Gly	MYZPE13164_0_v1.0_000149810	LOC100164013	**0.0000**	**15.7687**	**0.0000**	**5.5008**
1.8.1.4	Dihydrolipoyl dehydrogenase	Gly	MYZPE13164_0_v1.0_000017420	LOC100162429	0.5618	1.1148	**0.0067**	**2.8343**

[a]Amino acid enzymes that participate in from KEGG. [b]Gene ID assigned by *M. persicae* annotation v.1.0. [c]Pea aphid homolog. [d]Normalized p-value of differentially expressed genes between bacteriocytes and body cells. [e]Fold change of raw read counts of bacteriocytes over body cells. [f]Normalized p-values and [g]fold change of expression values of bacteriocytes over body cells in pea aphids feeding on fava plants (Kim *et al.*, 2018). Red and blue highlights indicate up-regulated and down-regulated genes in bacteriocytes compared to body cells, respectively ($p < 0.05$, at least 1.5x fold change). Bold ones indicate up- or down-regulated genes in bacteriocytes compared to body cells with at least 2x fold change.

Table 4-4. Differentially expressed active transporters including amino acid transporters in the bacteriocyte relative to the body in *M. persicae* and in *A. pisum*.

Transporter type	Gene ID[a]	*A. pisum* homolog[b]	p-value[c]	FC[d]	p_DK[e]	FC_DK[f]
Sugar transporters; similar to AGAP007667-PA	MYZPE13164_0_v1.0_000165530	LOC100161811	**0.0000**	**7.7312**	**0.0000**	**6.6154**
Sugar transporters; similar to AGAP007484-PA	MYZPE13164_0_v1.0_000083640	LOC100161021	**0.0000**	**34.3043**	**0.0000**	**83.2334**
Sugar transporters; similar to AGAP007483-PA	MYZPE13164_0_v1.0_000095960	LOC100160251	**0.0000**	**9.3837**	**0.0184**	**16.4564**
Sugar transporters; similar to AGAP004457-PA	MYZPE13164_0_v1.0_000066560	LOC100163256	**0.0000**	**2.9733**	**0.0000**	**18.5932**
Sugar transporter; similar to CG10960 CG10960-PB	MYZPE13164_0_v1.0_000048890	LOC100159728	**0.0000**	**4.3424**	**0.0000**	**8.4799**
Sugar transporter; similar to AGAP007484-PA	MYZPE13164_0_v1.0_000095970	LOC100169115	**0.0000**	**10.2615**	**0.0000**	**12.8193**
Sugar transporter; similar to AGAP001236-PA	MYZPE13164_0_v1.0_000165580	LOC100167659	0.0000	0.1987	0.1139	0.2263
Sugar transporter	MYZPE13164_0_v1.0_000112690	LOC100169458	**0.0000**	**39.0764**	**0.0001**	**16.4742**
Sugar transporter	MYZPE13164_0_v1.0_000090690	LOC100166904	**0.0000**	**5.8009**	**0.0000**	**6.3131**
Sugar transporter	MYZPE13164_0_v1.0_000047970	LOC100168622	**0.0000**	**8.4687**	**0.0112**	**4.1694**
Sugar transporter	MYZPE13164_0_v1.0_000165650	LOC100163428	**0.0000**	**13.4306**	**0.0028**	**3.1511**
Sugar transporter	MYZPE13164_0_v1.0_000061890	LOC100162563	0.4034	1.1648	**0.0001**	**46.3554**
Protein transporter; synaptic vesicle glycoprotein 2; similar to CG14691 CG14691-PA	MYZPE13164_0_v1.0_000046810	LOC100163807	**0.0000**	**22.6514**	**0.0000**	**73.6993**
Protein transporter; Peptide-acetyl-CoA transporter; similar to SD08430p	MYZPE13164_0_v1.0_000172010	LOC100162627	**0.0000**	**4.8975**	0.1300	2.2597
Protein transporter; Peptide-acetyl-CoA transporter; similar to SD08430p	MYZPE13164_0_v1.0_000076090	LOC100167419	**0.0000**	**37.2666**	**0.0005**	**41.0958**
Protein transporter; Peptide-acetyl-CoA transporter; similar to conserved hypothetical protein	MYZPE13164_0_v1.0_000134340	LOC100163344	**0.0000**	**6.9779**	**0.0043**	**28.0812**
Protein transporter; Peptide-acetyl-CoA transporter; similar to CG9706 CG9706-PA	MYZPE13164_0_v1.0_000134360	LOC100159262	**0.0000**	**23.7743**	**0.0002**	**42.3865**
Protein transporter; Peptide-acetyl-CoA transporter; similar to CG9706 CG9706-PA	MYZPE13164_0_v1.0_000134350	LOC100159922	**0.0332**	**2.0254**	0.1043	7.8382

Transporter type	Gene ID[a]	*A. pisum* homolog[b]	p-value[c]	FC[d]	p_DK[e]	FC_DK[f]
Protein transporter; Peptide-acetyl-CoA transporter; similar to CG9706 CG9706-PA	MYZPE13164_0_v1.0_000134330	LOC100165404	**0.0000**	**2.3692**	**0.0020**	**5.9923**
Protein transporter; Peptide-acetyl-CoA transporter; similar to acetyl-coenzyme A transporter	MYZPE13164_0_v1.0_000052240	LOC100169490	**0.0000**	**31.6533**	**0.0001**	**45.9975**
Organic acid transporters;monocarboxylic acid transporters ;similar to AGAP007743-PA	MYZPE13164_0_v1.0_000139420	Slc16a10	**0.0000**	**5.4426**	0.0559	2.3417
Organic acid transporter; sodium-dependent inorganic; similar to CG15094 CG15094-PA	MYZPE13164_0_v1.0_000080830	LOC100164256	**0.0020**	**2.2100**	**0.0000**	**0.1209**
Organic acid transporter; sodium-dependent inorganic; similar to AGAP007732-PA	MYZPE13164_0_v1.0_000001520	LOC100168152	**0.0000**	**36.2201**	**0.0000**	**167.3351**
Organic acid transporter; sodium-dependent inorganic; similar to AGAP007732-PA	MYZPE13164_0_v1.0_000041540	LOC100162864	**0.0000**	**3.7945**	0.0045	1.7041
Organic acid transporter; similar to sodium/solute symporter	MYZPE13164_0_v1.0_000034680	LOC100169257	**0.0000**	**7.0234**	**0.0089**	**7.0787**
Organic acid transporter; similar to organic cation transporter, partial	MYZPE13164_0_v1.0_000107160	LOC100167082	**0.0000**	**17.1978**	**0.0000**	**34.2830**
Organic acid transporter	MYZPE13164_0_v1.0_000059130	LOC100166357	**0.0000**	**44.9876**	**0.0125**	**2.9984**
Organic acid transporter	MYZPE13164_0_v1.0_000028930	LOC100164485	**0.0000**	**6.9744**	**0.0000**	**8.1582**
Organic acid transporter	MYZPE13164_0_v1.0_000107170	LOC100169129	**0.0000**	**9.2166**	0.6338	1.1738
amino acid transporter; cationic AA transmembrane	MYZPE13164_0_v1.0_000044170	LOC100168178	**0.0000**	**38.2664**	**0.0000**	**16.2945**
amino acid transporter; AA transmembrane transporter	MYZPE13164_0_v1.0_000043830	LOC100161944	**0.0000**	**6.8370**	0.0214	1.7009
amino acid transporter, cationic AA transmembrane	MYZPE13164_0_v1.0_000044160	LOC100159188	0.1314	0.3507	**0.0042**	**25.7679**
amino acid transporter, cationic AA transmembrane	MYZPE13164_0_v1.0_000035830	LOC100164077	**0.0000**	**2.8221**	**0.0001**	**4.5045**
<ApGLNT1> similar to proton-coupled amino acid transporter 1	MYZPE13164_0_v1.0_000000910	LOC100159667	**0.0000**	**5.6637**	0.7900	1.0965
active transporters;sodium/nucleoside cotransporter; similar to AGAP005989-PA, partial	MYZPE13164_0_v1.0_000007570	LOC100162704	**0.0000**	**4.7534**	0.2016	3.5968
active transporters; vacuolar H+-ATPase; hypothetical protein	MYZPE13164_0_v1.0_000058710	Vha16	0.0645	1.4825	**0.0002**	**2.9876**
active transporters; vacuolar H[+] ATPase 14kD subunit [Acyrthosiphon pisum]	MYZPE13164_0_v1.0_000040680	Vha14	0.0683	1.3909	**0.0001**	**7.1259**
active transporters; vacuolar ATPase subunit D	MYZPE13164_0_v1.0_000065770	Vha36	0.8274	0.9600	0.0658	2.0846
active transporters; vacuolar ATP synthase subunit G-like protein	MYZPE13164_0_v1.0_000196550	LOC100144888	0.0831	1.3964	**0.0005**	**40.1428**

Transporter type	Gene ID[a]	*A. pisum* homolog[b]	p-value[c]	FC[d]	p_DK[e]	FC_DK[f]
active transporters; sulfate transporter; similar to AGAP002331-PA	MYZPE13164_0_v1.0_000155760	LOC100162682	**0.0040**	**2.4044**	0.8832	0.9298
active transporters; stunted-like isoform B	MYZPE13164_0_v1.0_000141780	LOC100144904	**0.0000**	**2.1590**	**0.0017**	**2.3455**
active transporters; solute carrier family 25 member 35, hypothetical protein	MYZPE13164_0_v1.0_000116610	LOC100167176	**0.0001**	**2.7887**	**0.0000**	**44.4436**
active transporters; similar to vacuolar ATP synthase	MYZPE13164_0_v1.0_000009650	VhaAC39	0.0586	1.4358	0.8361	1.1419
active transporters; similar to sulfate transporter, similar to AGAP002331-PA	MYZPE13164_0_v1.0_000084570	LOC100165717	**0.0000**	**6.1334**	0.0660	3.8842
active transporters; similar to sulfate transporter	MYZPE13164_0_v1.0_000135130	LOC100159835	**0.0000**	**5.9559**	**0.0002**	**12.2931**
active transporters; similar to sulfate transporter	MYZPE13164_0_v1.0_000048810	LOC100164052	**0.0001**	**0.4118**	**0.0007**	**4.3376**
active transporters; similar to sodium/solute symporter	MYZPE13164_0_v1.0_000174730	LOC100159703	**0.0000**	**7.3192**	0.2311	2.6418
active transporters; similar to potassium/chloride symporter, putative, partial	MYZPE13164_0_v1.0_000044120	LOC100159106	**0.0000**	**6.0973**	**0.0001**	**5.6970**
active transporters; similar to pH-sensitive chloride channel	MYZPE13164_0_v1.0_000150760	LOC100162616	0.3903	1.2164	0.2979	2.3377
active transporters; similar to oligopeptide transporter	MYZPE13164_0_v1.0_000137150	LOC100164171	0.0010	0.5089	0.3867	2.2111
active transporters; similar to inwardly rectifying k+ channel	MYZPE13164_0_v1.0_000086020	LOC100160517	**0.0000**	**17.4107**	**0.0040**	**3.5752**
active transporters; similar to GA14898-PA	MYZPE13164_0_v1.0_000157230	LOC100163045	**0.0000**	**2.5120**	**0.0213**	**2.8623**
active transporters; similar to CG9990 CG9990-PA	MYZPE13164_0_v1.0_000017500	LOC100166151	0.3495	1.3856	0.3829	1.7245
active transporters; similar to CG6293 CG6293-PA	MYZPE13164_0_v1.0_000152690	LOC100158812	**0.0000**	**6.9194**	0.3355	0.4465
active transporters; similar to CG3876 CG3876-PA	MYZPE13164_0_v1.0_000045130	LOC100161800	**0.0000**	**4.7186**	**0.0000**	**76.4717**
active transporters; similar to CG32580 CG32580-PA	MYZPE13164_0_v1.0_000052470	LOC100160550	**0.0000**	**14.0119**	**0.0000**	**23.7027**
active transporters; similar to ATPase	MYZPE13164_0_v1.0_000183750	LOC100161355	0.0452	1.4959	**0.0009**	**2.8158**
active transporters; similar to aquaporin	MYZPE13164_0_v1.0_000080910	LOC100168499	**0.0000**	**2.2698**	**0.0000**	**11.2122**
active transporters; similar to AGAP011839-PA	MYZPE13164_0_v1.0_000147660	LOC100166886	**0.0000**	**2.8655**	0.0004	1.9375
active transporters; similar to AGAP007119-PA	MYZPE13164_0_v1.0_000190590	LOC100162446	0.1384	1.3020	0.6808	1.3904
active transporters; similar to AGAP002331-PA	MYZPE13164_0_v1.0_000155740	LOC100162816	**0.0000**	**7.5102**	0.0660	3.8842
active transporters; similar to AGAP000958-PA	MYZPE13164_0_v1.0_000166220	aralar1	0.0181	1.5023	**0.0126**	**4.3683**

Transporter type	Gene ID[a]	*A. pisum* homolog[b]	p-value[c]	FC[d]	p_DK[e]	FC_DK[f]
active transporters; similar to 2-oxoglutarate carrier protein	MYZPE13164_0_v1.0_000178330	LOC100159664	**0.0050**	**2.0174**	**0.0008**	**1.8578**
active transporters; putative vacuolar ATP synthase subunit E; hypothetical protein	MYZPE13164_0_v1.0_000157240	LOC100165126	0.0455	1.4650	**0.0001**	**2.7850**
active transporters; putative vacuolar ATP synthase proteolipid subunit; hypothetical protein	MYZPE13164_0_v1.0_000008540	LOC100165919	0.0359	1.4766	**0.0237**	**3.0997**
active transporters; major facilitator superfamily domain-containing protein 5; similar to predicted protein	MYZPE13164_0_v1.0_000149210	LOC100165727	0.6642	1.1228	**0.0000**	**5.9622**
active transporters; hypothetical protein	MYZPE13164_0_v1.0_000113140	LOC100163312	0.0145	1.5417	**0.0109**	**5.6807**
active transporters; hypothetical protein	MYZPE13164_0_v1.0_000144390	LOC100164284	**0.0000**	**6.9388**	**0.0000**	**77.5527**
active transporters; equilibrative nucleoside transporter; similar to AGAP003892-PA	MYZPE13164_0_v1.0_000173910	LOC100165130	0.0107	1.8397	0.0798	2.9653
active transporters; abc transporter; similar to CG17646 CG17646-PA	MYZPE13164_0_v1.0_000091600	LOC100164565	**0.0000**	**17.2801**	**0.0000**	**7.8252**
ABCG (White) subfamily; similar to abc transporter	MYZPE13164_0_v1.0_000005200	LOC100166685	**0.0000**	**3.0198**	0.1381	3.4804
active transporters; similar to sodium/solute symporter	MYZPE13164_0_v1.0_000174730	LOC100159703	**0.0000**	**7.3192**	0.2311	2.6418
active transporters; similar to potassium/chloride symporter, putative, partial	MYZPE13164_0_v1.0_000044120	LOC100159106	**0.0000**	**6.0973**	**0.0001**	**5.6970**
active transporters; similar to pH-sensitive chloride channel	MYZPE13164_0_v1.0_000150760	LOC100162616	0.3903	1.2164	0.2979	2.3377
active transporters; similar to oligopeptide transporter	MYZPE13164_0_v1.0_000137150	LOC100164171	0.0010	0.5089	0.3867	2.2111
active transporters; similar to inwardly rectifying k+ channel	MYZPE13164_0_v1.0_000086020	LOC100160517	**0.0000**	**17.4107**	**0.0040**	**3.5752**
active transporters; similar to GA14898-PA	MYZPE13164_0_v1.0_000157230	LOC100163045	**0.0000**	**2.5120**	**0.0213**	**2.8623**
active transporters; similar to CG9990 CG9990-PA	MYZPE13164_0_v1.0_000017500	LOC100166151	0.3495	1.3856	0.3829	1.7245
active transporters; similar to CG6293 CG6293-PA	MYZPE13164_0_v1.0_000152690	LOC100158812	**0.0000**	**6.9194**	0.3355	0.4465
active transporters; similar to CG3876 CG3876-PA	MYZPE13164_0_v1.0_000045130	LOC100161800	**0.0000**	**4.7186**	**0.0000**	**76.4717**
active transporters; similar to CG32580 CG32580-PA	MYZPE13164_0_v1.0_000052470	LOC100160550	**0.0000**	**14.0119**	**0.0000**	**23.7027**
active transporters; similar to ATPase	MYZPE13164_0_v1.0_000183750	LOC100161355	0.0452	1.4959	**0.0009**	**2.8158**
active transporters; similar to aquaporin	MYZPE13164_0_v1.0_000080910	LOC100168499	**0.0000**	**2.2698**	**0.0000**	**11.2122**
active transporters; similar to AGAP011839-PA	MYZPE13164_0_v1.0_000147660	LOC100166886	**0.0000**	**2.8655**	0.0004	1.9375
active transporters; similar to AGAP007119-PA	MYZPE13164_0_v1.0_000190590	LOC100162446	0.1384	1.3020	0.6808	1.3904

Transporter type	Gene ID[a]	*A. pisum* homolog[b]	p-value[c]	FC[d]	p_DK[e]	FC_DK[f]
active transporters; similar to AGAP002331-PA	MYZPE13164_0_v1.0_000155740	LOC100162816	**0.0000**	**7.5102**	0.0660	3.8842
active transporters; similar to AGAP000958-PA	MYZPE13164_0_v1.0_000166220	aralar1	0.0181	1.5023	**0.0126**	**4.3683**
active transporters; similar to 2-oxoglutarate carrier protein	MYZPE13164_0_v1.0_000178330	LOC100159664	**0.0050**	**2.0174**	**0.0008**	**1.8578**
active transporters; putative vacuolar ATP synthase subunit E; hypothetical protein	MYZPE13164_0_v1.0_000157240	LOC100165126	0.0455	1.4650	**0.0001**	**2.7850**
active transporters; putative vacuolar ATP synthase proteolipid subunit; hypothetical protein	MYZPE13164_0_v1.0_000008540	LOC100165919	0.0359	1.4766	**0.0237**	**3.0997**
active transporters; major facilitator superfamily domain-containing protein 5; similar to predicted protein	MYZPE13164_0_v1.0_000149210	LOC100165727	0.6642	1.1228	**0.0000**	**5.9622**
active transporters; hypothetical protein	MYZPE13164_0_v1.0_000113140	LOC100163312	0.0145	1.5417	**0.0109**	**5.6807**
active transporters; hypothetical protein	MYZPE13164_0_v1.0_000144390	LOC100164284	**0.0000**	**6.9388**	**0.0000**	**77.5527**
active transporters; equilibrative nucleoside transporter; similar to AGAP003892-PA	MYZPE13164_0_v1.0_000173910	LOC100165130	0.0107	1.8397	0.0798	2.9653
active transporters; abc transporter; similar to CG17646 CG17646-PA	MYZPE13164_0_v1.0_000091600	LOC100164565	**0.0000**	**17.2801**	**0.0000**	**7.8252**
ABCG (White) subfamily; similar to abc transporter	MYZPE13164_0_v1.0_000005200	LOC100166685	**0.0000**	**3.0198**	0.1381	3.4804
ABCC (CFTR/MRP) subfamily; similar to Multidrug-Resistance like Protein 1 CG6214-PK	MYZPE13164_0_v1.0_000032080	LOC100165098	**0.0000**	**0.2505**	1.0000	1.0000
ABCC (CFTR/MRP) subfamily; similar to ATP-dependent bile acid permease	MYZPE13164_0_v1.0_000072860	LOC100168114	**0.0000**	**7.3012**	0.1680	3.4363
ABCC (CFTR/MRP) subfamily; similar to ATP-dependent bile acid permease	MYZPE13164_0_v1.0_000072850	LOC100166328	0.2279	1.2914	**0.0107**	**22.7033**
ABCC (CFTR/MRP) subfamily; similar to ATP-dependent bile acid permease	MYZPE13164_0_v1.0_000160860	LOC100161113	**0.0000**	**2.2869**	0.4834	0.5919
ABCC (CFTR/MRP) subfamily; similar to ATP-dependent bile acid permease	MYZPE13164_0_v1.0_000118840	LOC100165493	**0.0034**	**2.4912**	0.2334	2.4326

[a]Gene ID assigned by *M. persicae* annotation v.1.0. [b]Pea aphid homolog. [c]Normalized p-value of differentially expressed genes between bacteriocytes and body cells. [d]Fold change of raw read counts of bacteriocytes over body cells. [e]Normalized p-values and [f]fold change of expression values of bacteriocytes over body cells in pea aphids feeding on fava plants (Kim *et al.*, 2018). Red and blue highlights indicate up-regulated and down-regulated genes in bacteriocytes compared to body cells, respectively ($p < 0.05$, at least 1.5x fold change). Bold ones indicate up- or down-regulated genes in bacteriocytes compared to body cells with at least 2x fold change.

Table 4-5A. Up-regulated *A. pisum* KEGG pathways in bacteriocytes of *M. persicae* compared to body cells.

NAME	Description	SIZE[a]	NES[b]	NOM p-val[c]
API03010	**Ribosome**	**111**	**2.4374619**	**0**
API03040	Spliceosome	110	2.0748997	0
API00240	**Pyrimidine metabolism**	**65**	**1.7872552**	**0**
API03410	Base excision repair	22	1.7258344	0.00364964
API03013	RNA transport	110	1.6823648	0
API03008	Ribosome biogenesis in eukaryotes	56	1.6723169	0
API03015	mRNA surveillance pathway	65	1.6566565	0.00109051
API03030	**DNA replication**	**32**	**1.6332701**	**0.00233918**
API03020	**RNA polymerase**	**22**	**1.580519**	**0.00855746**
API03460	Fanconi anemia pathway	29	1.5775697	0.00843374
API04120	Ubiquitin mediated proteolysis	74	1.5657086	0.00322928
API00190	**Oxidative phosphorylation**	**86**	**1.5507746**	**0**
API04340	Hedgehog signaling pathway -fly	21	1.5330284	0.01392405
API03022	Basal transcription factors	30	1.523735	0.0157385
API03440	**Homologous recombination**	**22**	**1.5132399**	**0.03198032**
API03018	RNA degradation	56	1.5017705	0.00883978
API04320	Dorso-ventral axis formation	21	1.4957622	0.03041363
API03420	Nucleotide excision repair	32	1.4809945	0.02450408
API04310	Wnt signaling pathway	55	1.4779882	0.0189099
API00230	**Purine metabolism**	**83**	**1.4740024**	**0.01055966**
API00010	Glycolysis/Glucogenesis	30	1.4362799	0.03452381
API03050	Proteasome	35	1.3978068	0.04262673

Table 4-5B. Down-regulated *A. pisum* KEGG pathways in bacteriocytes of *M. persicae* compared to body cells

NAME	Description	SIZE[a]	NES[b]	NOM p-val[c]
API04080	Neuroactive ligand-receptor interaction	27	-1.6515603	0.01015228
API00430	Taurine and hypotaurine metabolism	9	-1.5705205	0.04347826
API04745	Phototransduction - fly	15	-1.5555499	0.04347826
API00510	N-Glycan biosynthesis	28	-1.4940711	0.04605263

[a]Number of genes included in a gene set based on *A. pisum* KEGG pathways;

[b]Normalized enrichment score; [c]Normalized p-value based on the size of the group.

Bolded pathways are the pathways that are also up-regulated in *A. pisum* bacteriocytes compared to body cells from Kim *et al.* (2018).

Table 4-6A. *Myzus persicae* lineage-specific genes up-regulated in bacteriocytes compared to body cells.

GeneID	Predicted Description	p_adj[a]	FC[b]
MYZPE13164_0_v1.0_000162770	Unknown protein	0.0000	85304.2026
MYZPE13164_0_v1.0_000171630	Unknown protein	0.0000	6528.3961
MYZPE13164_0_v1.0_000108570	Aconitate hydratase, mitochondrial	0.0000	1180.4620
MYZPE13164_0_v1.0_000108560	3-hydroxymethyl-3-methylglutaryl-CoA lyase, cytoplasmic	0.0000	533.4632
MYZPE13164_0_v1.0_000068350	Wd-repeat protein	0.0058	496.2745
MYZPE13164_0_v1.0_000170590	Unknown protein	0.0000	456.5992
MYZPE13164_0_v1.0_000179560	Unknown protein	0.0000	408.0337
MYZPE13164_0_v1.0_000159160	Unknown protein	0.0000	266.7278
MYZPE13164_0_v1.0_000052310	Unknown protein	0.0000	182.1008
MYZPE13164_0_v1.0_000037990	Retrovirus-related Pol polyprotein from	0.0000	167.3799
MYZPE13164_0_v1.0_000190910	Lysozyme 3	0.0000	161.3485
MYZPE13164_0_v1.0_000030840	ACYPI005168 protein	0.0000	156.2941
MYZPE13164_0_v1.0_000108550	Unknown protein	0.0000	151.3995
MYZPE13164_0_v1.0_000183580	ACYPI005979 protein	0.0000	139.9608
MYZPE13164_0_v1.0_000183570	ACYPI005979 protein	0.0000	135.0705
MYZPE13164_0_v1.0_000158740	Unknown protein	0.0000	112.9298
MYZPE13164_0_v1.0_000200120	Unknown protein	0.0000	111.5479
MYZPE13164_0_v1.0_000188000	Unknown protein	0.0013	85.1454
MYZPE13164_0_v1.0_000183550	ACYPI005979 protein	0.0000	84.0798
MYZPE13164_0_v1.0_000129200	Nuclease	0.0000	82.6140
MYZPE13164_0_v1.0_000138760	ATP-dependent RNA helicase	0.0008	66.5552
MYZPE13164_0_v1.0_000003380	Unknown protein	0.0000	53.8907
MYZPE13164_0_v1.0_000101650	Unknown protein	0.0000	41.4600
MYZPE13164_0_v1.0_000158570	Nucleic-acid-binding protein from transposon x-element	0.0000	32.2792
MYZPE13164_0_v1.0_000122020	Unknown protein	0.0000	28.1369
MYZPE13164_0_v1.0_000017490	Unknown protein	0.0001	24.4602
MYZPE13164_0_v1.0_000059470	ACYPI000160 protein	0.0000	21.3181
MYZPE13164_0_v1.0_000194800	Zinc finger MYM-type protein 1	0.0000	20.4793
MYZPE13164_0_v1.0_000161990	Nucleic-acid-binding protein from transposon X-element	0.0002	19.2572
MYZPE13164_0_v1.0_000069250	Unknown protein	0.0003	18.5442
MYZPE13164_0_v1.0_000103580	Tetraspanin	0.0001	18.0885
MYZPE13164_0_v1.0_000007030	Unknown protein	0.0000	14.9021
MYZPE13164_0_v1.0_000160130	ACYPI008213 protein	0.0008	12.0666

MYZPE13164_0_v1.0_000056380	Unknown protein	0.0015	10.6661
MYZPE13164_0_v1.0_000067100	Small ubiquitin-related modifier	0.0001	10.0829
MYZPE13164_0_v1.0_000174420	Histone H1-II-1, putative	0.0004	9.9229
MYZPE13164_0_v1.0_000089610	Unknown protein	0.0000	9.5480
MYZPE13164_0_v1.0_000001640	Unknown protein	0.0000	9.3760
MYZPE13164_0_v1.0_000115630	Unknown protein	0.0000	9.3498
MYZPE13164_0_v1.0_000177900	ACYPI005147 protein	0.0083	8.9127
MYZPE13164_0_v1.0_000083710	ACYPI007976 protein	0.0002	8.5055
MYZPE13164_0_v1.0_000125320	Non-capsid protein NS-1	0.0000	8.2548
MYZPE13164_0_v1.0_000023750	Unknown protein	0.0073	8.1590
MYZPE13164_0_v1.0_000061670	Histone H1	0.0003	7.9134
MYZPE13164_0_v1.0_000191640	Unknown protein	0.0011	5.4512
MYZPE13164_0_v1.0_000060750	Ankyrin repeat and KH domain-containing protein 1	0.0001	5.0667
MYZPE13164_0_v1.0_000176920	Unknown protein	0.0001	4.9596

[a]Normalized p-value of differentially expressed genes between bacteriocytes and body cells. [b]Fold change of raw read counts of bacteriocytes over body cells.

Table 4-6B. *Myzus persicae* lineage-specific genes down-regulated in bacteriocytes compared to body cells.

GeneID	Predicted description	p_adj[a]	FC[b]
MYZPE13164_0_v1.0_000079410	PI-PLC X domain-containing protein 1	0.0005	47521.0767
MYZPE13164_0_v1.0_000169210	Unknown protein	0.0058	5342.7911
MYZPE13164_0_v1.0_000080250	UPF0439 protein C9orf30	0.0006	4140.3384
MYZPE13164_0_v1.0_000173480	Calcitonin receptor-like protein	0.0001	4024.1810
MYZPE13164_0_v1.0_000040930	Down syndrome cell adhesion molecule-like protein	0.0009	3487.3344
MYZPE13164_0_v1.0_000082620	Unknown protein	0.0006	3222.0703
MYZPE13164_0_v1.0_000060420	Zinc finger mym-type protein 1-like protein	0.0012	2938.7560
MYZPE13164_0_v1.0_000133590	ZNF687 protein	0.0041	314.3002
MYZPE13164_0_v1.0_000125550	Unknown protein	0.0000	92.8848
MYZPE13164_0_v1.0_000125520	Unknown protein	0.0000	66.7573
MYZPE13164_0_v1.0_000193690	Unknown protein	0.0025	66.7229
MYZPE13164_0_v1.0_000125540	Unknown protein	0.0000	54.2762
MYZPE13164_0_v1.0_000132160	Unknown protein	0.0000	26.2662
MYZPE13164_0_v1.0_000097390	Unknown protein	0.0000	21.4967
MYZPE13164_0_v1.0_000133350	Unknown protein	0.0000	20.3913

MYZPE13164_0_v1.0_000125510	Unknown protein	0.0000	18.9336
MYZPE13164_0_v1.0_000072950	Unknown protein	0.0020	16.1258
MYZPE13164_0_v1.0_000164820	ACYPI005590 protein	0.0059	14.7063
MYZPE13164_0_v1.0_000196990	Unknown protein	0.0000	14.4186
MYZPE13164_0_v1.0_000158490	General transcription factor II-I repeat domain-containing protein 2B	0.0016	14.3023
MYZPE13164_0_v1.0_000192210	Multiple inositol polyphosphate phosphatase 1	0.0000	13.7270
MYZPE13164_0_v1.0_000163270	Vesicular acetylcholine transporter-like Protein	0.0001	13.5528
MYZPE13164_0_v1.0_000187730	Nuclease harbi1-like protein	0.0024	11.8762
MYZPE13164_0_v1.0_000131420	Transposase domain-containing protein	0.0014	11.5374
MYZPE13164_0_v1.0_000064860	Major facilitator superfamily domain-containing protein 6	0.0001	10.8506
MYZPE13164_0_v1.0_000006970	ACYPI003979 protein	0.0004	9.0265
MYZPE13164_0_v1.0_000001530	Proton-coupled folate transporter	0.0000	8.9133
MYZPE13164_0_v1.0_000044010	Cathepsin B	0.0032	7.8902
MYZPE13164_0_v1.0_000029870	Unknown protein	0.0004	6.7499
MYZPE13164_0_v1.0_000143570	Lipopolysaccharide-induced tumor necrosis factor-alpha factor like protein	0.0069	5.9525
MYZPE13164_0_v1.0_000156040	RNA-directed DNA polymerase from	0.0005	5.2029

[a]Normalized p-value of differentially expressed genes between bacteriocytes and body cells. [b]Fold change of raw read counts of bacteriocytes over body cells.

Chapter 5: Conclusion

Herbivore insects are required to overcome lots of obstacles to survive on their diets. Such obstacles include seasonal availability of the food resource, both chemical and physical plant defense systems, and highly imbalanced nutrients and amino acids from plant parts. Therefore, herbivore insects have evolved different strategies to survive and reproduce. One type of strategy is the symbiotic association with obligate, nutritional microbes, which requires complex crosstalk and regulation between the host and the symbionts depending on what the host feeds upon.

In this dissertation, three aphid species (*A. pisum*, *A. glycines*, and *M. persicae*) were explored to investigate the gene regulation of specialized cells that harbor nutritional symbionts, called bacteriocytes. In the second chapter, I used the model system for insect-microbe symbiosis between pea aphids and *Buchnera*, to investigate the association between tissue-specific and host plant-specific gene expression and CpG methylation profiles. Through this study, I found that DNA methylation may play a key role in the regulation of aphid genes involved in the aphid-*Buchnera* symbiosis. In the third chapter of my dissertation, I used another species of aphid, *A. glycines*, to explore the methylation levels of lineage-specific and evolutionarily conserved genes. Coinciding with the previous research on other insect species with functional CpG methylation, aphid-specific genes tend to have lower methylation levels throughout the genome of *A. glycines*. Furthermore, I identified several highly methylated genes involved in plant-insect interactions. In the last chapter, I compared the tissue-specific gene expression patterns of *M. persicae* bacteriocytes to *A. pisum* bacteriocytes, when they fed on the

same host plant species. Although overall gene expression patterns in regard to amino acid biosynthesis are highly similar, I identified some key genes that have species-specific expression profiles within bacteriocytes of *M. persicae.*

One of the reasons why insects are the most numerous and successful group of organisms on Earth is their ability to quickly adapt to diverse microhabitats. While *A. pisum* have a somewhat broad range of host plants in *Fabaceae* family (Chapter 2), *A. glycines* has a narrow range of host plants (Chapter 3), and *M. persicae* is a true generalist feeding on the host plants that belong to over 40 different families (Chapter 4). Investigating the differences and similarities of gene regulation between three aphid species with diverse host plant ranges is an important steppingstone to further study how insects can utilize nutritional symbionts to adapt and survive on different host plants.

Bibliography

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics International* 11, 36-42.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., *et al*. (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13, R87. doi: 10.1186/gb-2012-13-10-r87

Alvarado, S., Rajakumar, R., Abouheif, E., and Szyf, M. (2015). Epigenetic variation in the Egfr gene generates quantitative variation in a complex trait in ants. Nat. Comm. 6, 6513. doi: 10.1038/ncomms7513

Amarasinghe, H. E., Clayton, C. I., and Mallon, E. B. (2014). Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proc. Biol. Sci. 281*, 20132502. doi: 10.1098/rspb.2013.2502

Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M., and Melino, G. (2014). Serine and glycine metabolism in cancer. *Trends Biochem. Sci.* 39, 191-198. doi: 10.1016/j.tibs.2014.02.004

Anderson, O. S., Sant, K. E., and Dolinoy, D. C. (2012). Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *J. Nutr. Biochem.* 23, 853-859. doi: 10.1016/j.jnutbio.2012.03.003

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Augui, S., Nora, E. P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* 12, 429–442. doi: 10.1038/nrg2987

Balaton, B. P., Cotton, A. M., and Brown, C. J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol. Sex Differ.* 6, 1–11. doi: 10.1186/s13293-015-0053-7

Bansal, R., Mian, M. A., Mittapalli, O., Michel, A. P. (2014). RNA-Seq reveals a xenobiotic stress response in the soybean aphid, *Aphis glycines*, when fed aphid-resistant soybean. *BMC Genomics.* 15, 972. doi: 10.1186/1471-2164-15-972

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Re. Microbiol.* 59, 155-189. doi: 10.1146/annurev.micro.59.030804.121041

Bebek, G., Bennett, K. L., Funchain, P., Campbell, R., Seth, R., Scharpf, J., et al. (2012). Microbiomic subprofiles and MDR1 promoter methylation in head and neck squamous cell carcinoma. Hum. Mol. Gen. 21, 1557–1565. doi: 10.1093/hmg/ddr593

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., *et al*. (2010). The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 28, 1045-1048. doi: 10.1038/nbt1010-1045

Biergans, S. D., Jones, J. C., Treiber, N., Galizia, C. G., and Szyszka, P. (2012). DNA methylation mediates the discriminatory power of associative long-term memory in honeybees. *PLoS ONE* 7:e39349. doi: 10.1371/journal.pone.0039349

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321, 209–213. doi: 10.1038/321209a0

Blackman, R. L., and Eastop, V. F. (2000). *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120. doi: 10.1093/bioinformatics/btu170

Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Curr. Biol. 22, 1755–1764. doi: 10.1016/j.cub.2012.07.042

Bookout, A. L., Cummins, C. L., Magelsdorf, D. J., Pesola, J. M., and Kramer, M.F. (2006). High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol.* 1, 1-28. doi: 10.1002/0471142727.mb1508s73

Braendle, C., Miura, T., Bickel, R., Shingleton, A. W., Kambhampati, S., and Stern, D. L. (2003). Developmental origin and evolution of bacteriocytes in the aphid–Buchnera symbiosis. PLoS Biol. 1:e21. doi: 10.1371/journal.pbio.0000021

Brisson, J. A., Jaquiery, J., Legeai, F., Le Trionnaire, G., and Tagu. D. (2016). Genomics of phenotypic plasticity in aphids. In Czosnek H, Ghanim M, editors. Management of Insect Pests to Agriculture: Lessons Learned from Deciphering their Genome, *Transcriptome and Proteome*. (pp. 65-96). Springer International Publishing, Cham.

Brownlie, J. C., Cass, B. N., Riegler, M., Witsenburg, J. J., Iturbe-Ormaetxe, I., McGraw, E. A., et al. (2009). Evidence for metabolic provisioning by a common invertebrate endosymbiont, Wolbachia pipientis, during periods ofnutritional stress. PLoS Pathog. 5:e1000368. doi: 10.1371/journal.ppat.1000368

Brumin, M., Kontsedalov, S., and Ghanim, M. (2011). Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. Insect Sci. 18, 57–66. doi: 10.1111/j.1744-7917.2010.01396.x

Burgers, W. A., Blanchon, L., Pradhan, S., De Launoit, Y., Kouzarides, T., and Fuks, F. (2007). Viral oncoproteins target the DNA methyltransferases. Oncogene 26, 1650–1655. doi: 10.1038/sj.onc.1209950

Camacho, C., Coulouris, G., Avagyan, V., Ma, N, Papadopoulos, J., *et al.* (2009). BLAST+: architecture and applications. *BMC Bioinformatics*. 10, 421. doi: 10.1186/1471-2105-10-421

Chang, A., Scheer, M., Grote, A., Schomburg, I., and Schomburg, D. (2009). BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.* 37, D588-592. doi: 10.1093/nar/gkn820

Chen, N., Fan, Y. L., Bai, Y., Li, X. D., Zhang, Z. F., *et al.* (2016). Cytochrome P450 gene, CYP4G51, modulates hydrocarbon production in the pea aphid, *Acyrthosiphon pisum*. *Insect Biochem. Mol. Biol.* 76, 84-94. doi: 10.1016/j.ibmb.2016.07.006

Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*. 274, 775-780

Daub, M. E., and Ehrenshaft, M. (2000)The photoactivated *Cercospora* toxin cercosporin: contributions to plant disease and fundamental biology. *Annu. Rev. Phytopathol.* 38, 461-490. doi: 10.1146/annurev.phyto.38.1.461

Diers, A. R., Broniowska, K. A., Chang, C. F., and Hogg, N. (2012). Pyruvate fuels mitochondrial respiration and proliferation of breast cancer cells: effect of monocarboxylate transporter inhibition. *Biochem. J.* 444, 561-571. doi: 10.1042/BJ20120294

di Salvo, M. L., Contestabile, R., Paiardini, A., and Maras, B. (2013). Glycine consumption and mitochondrial serine hydroxymethyltransferase in cancer cells: the heme connection. *Med. Hypotheses* 80, 633-636. doi: 10.1016/j.mehy.2013.02.008

Dixon, A. F. G. (1977). Aphid ecology: life cycles, polymorphism, and population regulation. *Annual Review of Ecology and Systematics* 8, 329-353. doi: 10.1146/annurev.es.08.110177.001553

Dombrovsky, A., Arthaud, L., Ledger, T. N., Tares, S., and Robichon, A. (2009). Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction. Genome Res. 19, 2052–2063. doi: 10.1101/gr.091611.109

Dunbar, H. E., Wilson, A. C., Ferguson, N. R., and Moran, N. A. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. PLoS Biol. 5:e96. doi: 10.1371/journal.pbio.0050096

Durdevic, Z., Hanna, K., Gold, B., Pollex, T., Cherry, S., Lyko, F., et al. (2013). Efficient RNA virus control in Drosophila requires the RNA methyltransferase Dnmt2. EMBO Rep. 14, 269–275. doi: 10.1038/embor.2013.3

Dyall, S. D., Brown, M. T., and Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. Science 304, 253-257. doi: 10.1126/science.1094884

Elango, N., Hunt, B. G., Goodisman, M. A. D., and Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene explression in castes of the honey bee, *Apis melifera*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11206–11211. doi: 10.1073/pnas.0900301106

Fang, J. Y., Mikovits, J. A., Bagni, R., Petrow-Sadowski, C. L., and Ruscetti, F. W. (2001). Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases de novo methylation. J. Virol. 75, 9753–9761. doi: 10.1128/JVI.75.20.9753-9761.2001

Febvay G., Rahbé, Y.,Rynkiewicz, M., Guillaud, J., and Bonnot, G. (1999). Fate of dietary sucrose and neosynthesis of amino acids in the pea aphid, *Acyrthosiphon pisum*, reared on different diets. *J. Exp. Biol.* 202, 2639-2652.

Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97-109. doi: 10.1038/nrg3142

Feldhaar, H., and Gross, R. (2009). Insects as hosts for mutualistic bacteria. *Int. J. Med. Microbiol.* 299, 1-8. doi: 10.1016/j.ijmm.2008.05.010

Feng, S., Cokus, S. J., Zhang, X., Chen, P., Bostick, M. *et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA* 107, 8689-8694. doi: 10.1073/pnas.1002720107

Ferrari, J., Via, S., and Godfray, H. C. (2008). Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution* 62: 2508-2524. doi: 10.1111/j.1558-5646.2008.00468.x

Ferrari, J., West, J. A., Via, S., Godfray, H. C. (2012) Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. *Evolution* 66: 375-390. doi: 10.1111/j.1558-5646.2011.01436.x

Field, L. M. (2000). Methylation and expression of amplified esterase genes in the aphid Myzus persicae (Sulzer). Biochem. J. 349, 863–868. doi: 10.1042/bj3490863

Field, L. M., and Blackman, R. L. (2003). Insecticide resistance in the aphid *Myzus persicae* (Sulzer): chromosome location and epigenetic effects on esterase gene expression in clonal lineages. *Biol. J. Linn. Soc.* 79, 107–113. doi: 10.1046/j.1095-8312.2003.00178.x

Field, L. M., Crick, S. E., and Devonshire, A. L. (1996). Polymerase chain reaction-based identification of insecticide resistance genes and DNA methylation in the aphid Myzus persicae (Sulzer). Insect Mol. Biol. 5, 197–202. doi: 10.1111/j.1365-2583.1996.tb00054.x

Field, L. M., Devonshire, A. L., Ffrench-Constant, R. H., and Forde, B. G. (1989). Changes in DNA methylation are associated with loss of insecticide resistance in the peach-potato aphid Myzus persicae (Sulz.). FEBS Lett. 243, 323–327. doi: 10.1016/0014-5793(89)80154-1

Field, L. M., Lyko, F., Mandrioli, M., and Prantera, G. (2004). DNA methylation in insects. Insect Mol. Biol. 13, 109–115. doi: 10.1111/j.0962-1075.2004.00470.x

Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., et al. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. U.S.A. 109, 4968–4973. doi: 10.1073/pnas.1202392109

Frantz, A., Plantegenest, M., Mieuzet, L., and Simon, J. C. (2005). Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. *J. Evol. Biol.* 19, 392-401. doi: 10.1111/j.1420-9101.2005.01025.x

Gadjev, I. (2015). Nature and nuture: lamarck's legacy. Biol. J. Linn. Soc. 114, 242–247. doi: 10.1111/bij.12439

Gama-Sosa, M. A., Midgett, R. M., Slagel, V. A., Githens, S., Kuo, K. C., Gehrke, C. W., et al. (1983). Tissue-specific differences in DNA methylation in various mammals. Biochim. Biophys. Acta 740, 212–219. doi: 10.1016/0167-4781(83)90079-9

Gauthier, J., Legeai, F., Zasadzinski, A., Rispe, C., and Tagu, D. (2007). AphidBase: a database for aphid genomic resources. *Bioinformatics*. 23,783-784. doi: 10.1093/bioinformatics/bt682

Gavery, M. R., and Roberts, S. B. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (Crassostrea gigas). BMC Genomics 11:483. doi: 10.1186/1471-2164-11-483

Gil, R., Latorre, A., and Moya, A. (2004). Bacterial endosymbionts of insects: insights from comparative genomics. *Environ. Microbiol.* 6, 1109-1122. doi: 10.1111/j.1462-2920.2004.00691.x

Glastad, K. M., Hunt, B. G., and Goodisman, M. A. D. (2012). Evidence of a conserved functional role for DNA methylation in termites. *Insect Mol. Biol.* 22, 143–154. doi: 10.1111/imb.12010

Glastad, K. M., Hunt, B. G., and Goodisman, M. A. A. (2014). Evolutionary insights into DNA methylation in insects. *Curr. Opin. Insect Sci.* 1, 25-30. doi: 10.1016/j.cois.2014.04.001

Glastad, K. M., Hunt, B. G., Yi, S. V., and Goodisman, M. A. D. (2011). DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol Biol.* 20, 553-565. doi: 10.1111/j.1365-2583.2011.01092.x

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., et al. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311, 395–398. doi: 10.1126/science.1120976

Gong, L., Cui, F., Sheng, C., Lin, Z., Reeck, G., Xu, J., et al. (2012). Polymorphism and methylation of four genes expressed in salivary glands of Russian Wheat Aphid (Homoptera: Aphididae). J. Econ. Entomol. 105, 232–241. doi: 10.1603/EC11289

Gramates, L. S., Marygold, S. J., dos Santos, G., Urbano, J., Antonazzo, G., *et al*. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res.* 45, D663-D671. doi:10.1093/nar/gkw1016

Hansen, A. K., and Degnan, P. H. (2014) Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J.* 8, 2490-2502. doi: 10.1038/ismej.2014.121

Hansen, A. K., and Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. U.S.A. 108, 2849–2854. doi: 10.1073/pnas.1013465108

Hansen, A. K., and Moran, N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol. Ecol. 23, 1473–1496. doi: 10.1111/mec.12421

Haribal M., and Jander, G. (2015). Stable isotope studies reveal pathways for the incorporation of non-essential amino acids in *Acyrthosiphon pisum* (pea aphids). *J. Exp. Biol*. 218, 3797-3806. doi: 10.1242/jeb.129189

Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* 15, 1371–1373. doi: 10.1038/nn.3218

Hick, C. A., Field, L. M., and Devonshire, A. L. (1996). Changes in the methylation of amplified esterase DNA during loss and reselection of insecticide resistance in peach-potato aphids, Myzus persicae. Insect Biochem. Mol. Biol. 26, 41–47. doi: 10.1016/0965-1748(95)00059-3

Hill, C. B., Chirumamilla, A., and Hartman, G. L. (2012). Resistance and virulence in the soybean – *Aphis glycines* interaction. *Euphytica*. 186, 635-646. doi: 10.1007/s10681-012-0695-z

Hirayama, C., and Nakamura, M. (2001). Regulation of glutamine metabolism during the development of *Bombyx mori* larvae. *Biochim. Biophys. Acta.* 1572, 131-137. doi: 10.1016/S0304-4165(02)00207-6.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., et al. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. Science 298, 129–149. doi: 10.1126/science.1076181

Hosokawa, T., Ishii, Y., Nikoh, N., Fujie, M., Satoh, N., and Fukatsu, T. (2016). Obligate bacterial mutualists evolving from environmental bacteria in natural insect populations. Nat. Microbiol. 1:15011. doi: 10.1038/nmicrobiol.2015.11

Huh, I., Zeng, J., Park, T., Yi, S. V. (2013). DNA methylation and transcriptional noise. *Epigenetics Chromatin* 6, 9. doi: 10.1186/1756-89345-6-9

Humphreys, N. J., and Douglas, A. E. (1997). Partitioning of symbiotic bacteria between generations of an insect: a quantitative study of a *Buchnera sp.* in the pea aphid (*Acyrthosiphon pisum*) reared at different temperatures. *Appl. Environ. Microbiol.* 63, 3294-3296.

Hunt, B. G., Glastad, K. M., Yi, S. V., and Goodisman, M. A. (2013). The function of intragenic DNA methylation: Insights from insect epigenomes. *Integr. Comp. Biol.* 53, 319-328. doi: 10.1093/icb/ict003

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol.* 8, e1000313. doi: 10.1371/journal.pbio.1000313

Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., *et al.* (2012). Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* 336, 1040-1044. doi: 10.1126/science.1218595

Jankard, R., and Herman, J. P. (2008). Limbic regulation of hypohalamo-pituitary-adrenocortical function during acute and chronic stress. Ann. N.Y. Acad. Sci. 1148, 64–73. doi: 10.1196/annals.1410.012

Jeltsch, A. (2010). Phylogeny of methylomes. Science 328, 837–838. doi: 10.1126/science.1190738

Jiang, Z., Jones, D. H., Khuri, S., Tsinoremas, N. F., Wyss, T., *et al*. (2013). Comparative analysis of genome sequences from four strains of the *Buchnera aphidicola* Mp endosymbiont of the green peach aphid, *Myzus persicae*. *BMC Genomics*. 14, 917. doi: 10.1186/1471-2164-14-917

John, R. A. (1995). Pyridoxal phosphate-dependent enzymes. *Biochim. Biophys. Acta.* 1248, 81-96. doi: 10.1016/0167-4838(95)00025-P

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492. doi: 10.1038/nrg3230

Jung, J. K., Arora, P., Pagano, J. S., and Jang, K. L. (2007). Expression of DNA methyltransferase 1 is activated by hepatitis B virus X protein via a regulatory circuit involving the p16INK4a-cyclin D1-CDK 4/6-pRb-E2F1 pathway. Cancer Res. 67, 5771–5778. doi: 10.1158/0008-5472.CAN-07-0529

Kaltenpoth, M., Gottler, W., Herzner, G., and Strohm, E. (2005). Sytmbiotic vacteria protect wasp larvae from funcgal infestation. Curr. Biol. 15, 475–479. doi: 10.1016/j.cub.2004.12.084

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27-30. doi: 10.1093/nar/28.1.27

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772-780. doi: 10.1093/molbev/mst010

Kelleher, E. S., and Markow, T. A. (2009). Duplication, selection, and gene conversion in a *Drosophila majovensis* female reproductive protein family. *Genetics.* 181, 1451-1465. doi: 10.1534/genetics.108.099044

Keseler, I. M., Bonavides-Marinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., *et al*. (2009). EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic Acids Res.* 37, D464-470. doi: 10.1093/nar/gkn751

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. 12, 357-60. doi: 10.1038/nmeth.3317

Kim, D., Minhas, B. F., Li-Byarlay, H., and Hansen, A. K. (2018). Key transport and ammonia recycling genes involved in aphid symbiosis respond to host-plant specialization. *G3*. 8, 2433-2443. doi: 10.1534/g3.118.200297

Klose, R. J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci*. 31, 89-97. doi: 10.1016/j.tibs.2005.12.008

Koga, R., Meng, X., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Nat. Acad. Sci.* 109, E1230-E1237. doi: 10.1073/pnas.1119212109

Kono M., Koga, R, Shimada, M., and Fukatsu, T. (2008). Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. *Appl. Environ. Microbiol.* 74, 4175-4184. doi: 10.1128/AEM.00250-08

Krauss, V., Eisenhardt, C., and Unger, T. (2009). The genome of the stick insect *Medauroidea extradentata* is strongly methylated within genes and repetitive DNA. *PLoS ONE*. 4, e7223. doi: 10.1371/journal.pone.0007223

Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571-1572. doi: 10.1093/bioinformatics/btr167

Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. Science 319, 1827–1830. doi: 10.1126/science.1153069

Kumar, S., and Kim, Y. (2017). An endoparasitoid wasp influences host DNA methylation. *Sci. Rep*. 7, 43287. doi: 10.1038/srep43287

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870-1874. doi: 10.1093/molbev/msw054

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development 130, 5083–5090. doi: 10.1242/dev.00716

Labuschagne, C. F., Van Den Broek, N. J., Mackay, G. M., Vousden, K. H., and Maddocks, O. D. (2014). Serine, but not glycine, supports one-carbon metabolism and proliferation of cancer cells *Cell Rep.* 7, 1248-1258. doi: 10.1016/j.celrep.2014.04.045

Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., *et al.* (2010). AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Mol. Biol.* 19, 5-12. doi: 10.1111/j.1365-2583.2009.00930.x

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature. 366, 362–365. doi: 10.1038/366362a0

Li-Byarlay, H., Li, Y., Stround, H., Feng, S., Newman, T. C., Kaneda, M., et al. (2013). RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. Proc. Natl. Acad. Sci. U.S.A. 110, 12750–12755. doi: 10.1073/pnas.1310735110

Liadouze, I., Febvay, G., Guillaud, J., and Bonnot, G. (1995). Effect of diet on the free amino acid pools of symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum*. *J. Insect Physiol*. 41, 33-40. doi: 10.1016/0022-1910(94)00085-U

Locasale, J. W. (2013) Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat. Rev. Cancer* 13, 572-583. doi: 10.1038/nrc3557

Lockett, G. A., Helliwell, P., and Maleszka, R. (2010). Involvement of DNA methylation in memory processing in the honey bee. *Neuroreport* 21, 812–816. doi: 10.1097/WNR.0b013e32833ce5be

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16-26. doi: 10.1016/j.cell.2010.11.056

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., *et al.* (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLOS Biol.* 8, e1000506. doi: 10.1371/journal.pbio.1000506

Lyko, F., and Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. *Trends Genet.* 27, 127-131. doi: 10.1016/j.tig.2011.01.003

Machado-Assefh, C. R., Lopez-Isasmendi, G., Tjallingii, W. F., Jander, G., and Alvarez, A. E. (2015). Disrupting *Buchnera aphidicola*, the endosymbiotic bacteria of *Myzus*

persicae, delays host plant acceptance. *Arthropod-Plant Interactions*. doi: 10.1007/s11829-015-9394-8

Maddocks, O. D., Berkers, C. R., Mason, S. M., Zheng, L., Blyth, K. *et al.* (2013). Serine starvation induces stress and p53-dependent metabolic remodeling in cancer cells. *Nature* 24, 542-546. doi: 10.1038/nature11743

Maechler, M. (2016). diptest: Hartigan's dip test statistic for unimodality – corrected. R package version 0.75-7. https://CRAN.R-project.org/package-diptest

Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., *et al*. (2017). Rapid transcriptional plasticity of duplicated gene clusters enbales a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* 18, 27. doi: 10.1186/s13059-016-1145-3

Mathers, T. C., Mugford, S. T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., et al. (2018). Sex-specific changes in the aphid DNA methylation landscape. *bioRxiv*. 286302. doi: 10.1101/286302

Maunakea, A. K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. 23, 1256–1269. doi: 10.1038/cr.2013.110

Medina-Ortega, K. J., and Walker, G. P. (2015). Faba been forisomes can function in defence against generalist aphids. *Plant Cell Environ.* 38, 1167-1177. doi: 10.1111/pce.12470

Mielke, P. W., and Berry, K. J. (2003). Multivariate multiple regression analyses: A permutation method for linear models. *Physicol. Rep* 91, 3-9. doi: 10.2466/pr0.2002.91.1.3

Mira A., and Moran, N. A. (2002). Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* 44, 137-143. doi: 10.1007/s00248-002-0012-9

Mischke, M., and Plosch, T. (2013). More than just a gut instinct-the potential interplay between a baby's nutrition, its gut microbiome, and the epigenome. Am. J. Physiol. Regul. Integr. Comp. Physiol. 304, R1065–R1069. doi: 10.1152/ajpregu.00551.2012

Moczek, A. P., and Snell-Rood, E. C. (2008). The basis of bee-ing different: the role of gene silencing in plasticity. Evol. Dev. 10, 511–513. doi: 10.1111/j.1525-142X.2008.00264.x

Moran, N. A., and Telan, A. (1998). Bacteriocyte-associated symbionts of insects. *BioScience.* 48, 295-304. doi: 10.2307/1313356

Mukherjee, K., and Baudach, A. F. (2016). Epigenetic Control of Polyphenism in Aphids, pp. 89-99 in *Biology and Ecology of Aphids*, edited by A. Vilcinskas. CRC Press, Boca Raton, Florida.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., et al. (2005). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. Proc. Natl. Acad. Sci. U.S.A. 102, 5477–5482. doi: 10.1073/pnas.0409034102

Nakabachi, A., Ishida, K., Hongoh, Y., Ohkuma, M., and Miyagishima, S.-Y. (2014). Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Curr. Biol. 24, R640–R641. doi: 10.1016/j.cub.2014.06.038

Nakajima, T., Yamashita, S., Maekita, T., Niwa, T., Nakazawa, K., and Ushijima, T. (2009). The presence of a methylation fingerprint of Helicobacter pylori infection in human gastric mucosae. Int. J. Cancer 124, 905–910. doi: 10.1002/ijc.24018

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z. J., Ren, Q., et al. (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. Science 316, 1718–1723. doi: 10.1126/science.1138878

Niculescu, M. D., and Zeisel, S. H. (2002). Diet, methyl donors and DNA methylation: interactions between dietary folate, methionine and choline. *J. Nutr.* 132, 2333S-2335S. doi: 10.1093/jn/132.8.2333S

Nishikori, K., Morioka, K., Kubo, T., and Morioka, M. (2009) Age- and morph-dependent activation of the lysosomal system and *Buchnera* degradation in aphid endosymbiosis. *J. Insect Physiol.* 55, 351-357. doi: 10.1016/j.jinsphys.2009.01.001

Nováková, E., Hypša, V., Klein, J., Foottit, R. G., von Dohlen, C. D., *et al*. (2013). Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Mol. Phylogenet. Evol.* 68, 42-54. doi: 10.1016/j.ympev.2013.03.016

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., *et al*. (2017). vegan: Community ecology package. R package version 2.4-3. https://CRAN.R-project.org/pacakge=vegan

Oliver, K. M., Moran, N. A., and Hunter, M. S. (2005). Variation in resistance to parasitism in aphids is due to symbionts not host genotype. Proc. Natl. Acad. Sci. U.S.A. 102, 12795–12800. doi: 10.1073/pnas.0506131102

Ono, M., Swanson, J. J., Field, L. M., Devonshire, A. L., and Siegfried, B. D. (1999). Amplification and methylation of an esterase gene associated with insecticide-resistance in greenbugs, *Schizaphis graminum* (Rondani) (Homoptera: Aphididae). *Insect Biochem. Mol. Biol.* 29, 1065–1073. doi: 10.1016/S0965-1748(99)00082-X

Park, J., Peng, Z., Zeng, J., Elango, N., Park, T., Wheeler, D., et al. (2011). Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes. *Mol. Biol. Evol.* 28, 3345–3354. doi: 10.1093/molbev/msr168

Parkinson, J. F., Gobin, B., and Hughes, W. O. (2016). Heritability of symbiont density reveals distinct regulatory mechanisms in a tripartite symbiosis. *Ecol. Evol.* 6, 2053-2060. doi: 10.1002/ece3.2005

Pasquier, C., Clément, M., Dombrovsky, A., Penaud, S., Da Rocha, M., *et al.* (2014). Environmentally selected aphid variants in clonality context display differential patterns of methylation in the genome. *PloS One* 9, e115022. doi: 10.1371/journal.pone.0115022

Patalano, S., Hore, T. A., Reik, W., and Sumner, S. (2012). Shifting behavior: epigenetic reprogramming in eusocial insects. *Curr. Opin. Cell Biol.* 24, 367–373. doi: 10.1016/j.ceb.2012.02.005

Peccoud, J., de la Huerta, M., Bonhomme, J., Laurence, C., Outreman, Y., *et al.*, (2014). Widespread host-dependent hybrid unfitness in the pea aphid species complex. *Evolution* 68: 2983-2995. doi: 10.1111/evo.12478

Pegoraro, M., Bafna, A., Davies, N. J., Shuker, D. M., and Tauber, E. (2016). DNA methylation changes induced by long and short photoperiods in Nasonia. Genome Res. 26, 203–210. doi: 10.1101/gr.196204.115

Perroud, B., Lee, J., Valkova, N., Dhirapong, A., Lin, P., *et al.* (2006). Pathway analysis of kidney cancer using proteomics and metabolic profiling. *Mol. Cancer* 5, 64. doi: 10.1186/1476-4598-5-64

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protocols*. 11, 1650-1667. doi: 10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J.T., *et al*. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotech.* 33, 290-295. doi: 10.1038/nbt.3122

Peter, I. S., and Davidson, E. H. (2015). Genomic Control Process: Development and Evolution. New York, NY: Elsevier.

Poliakov, A., Russell, C. W., Ponnala, L., Hoops, H. J., Sun, Q., Douglas, A. E., et al. (2011). Large-scale label-free quantitative proteomics of the pea aphid-buchnera symbiosis. Mol. Cell. Proteomics 10:M110.007039. doi: 10.1074/mcp.m110.007039

Price, D. R., Feng, H., Baker, J. D., Bavan, S., Luetje, C. W., and Wilson, A. C. (2014). Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. Proc. Natl. Acad. Sci. U.S.A. 111, 320–325. doi: 10.1073/pnas.1306068111

Provataris, P., Meusemann, K., Niehuis, O., Grath, S., and Misof, B. (2018). Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. *Genome Biol. Evol.* 10, 1185-1197. doi: 10.1093/gbe/evy066

Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., *et al.* (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet.* 6, e1000999. doi: 10.1371/journal.pgen.1000999

Rai, K., Chidester, S., Zavala, C. V., Manos, E. J., James, S. R., *et al.* (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.* 21, 261-266. doi: 10.1101/gad.1472907

Ragsdale, D. W., Landis. D. A., Brodeur. J., Heimpel. G. E., and Desneux. N. (2011). Ecology and management of the soybean aphid in North America. *Annu. Rev Entomol.* 56:375-399.

Razin, A., and Cedar, H. (1994). DNA methylation and genomic imprinting. Cell. 77, 473–476. doi: 10.1016/0092-8674(94)90208-9

R Core Team. (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Regev, A., Lamb, M. J., and Jablonka, E. (1998). The rold of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol. Biol. Evol.* 15, 880–891. doi: 10.1093/oxfordjournals.molbev.a025992

Reik, W., Collick, A., Norris, M. L., Barton, S. C., and Surani, M. A. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* 328, 248–251. doi: 10.1038/328248a0

Robertson, K. D., Hayward, S. D., Ling, P. D., Samid, D., and Ambinder, R. F. (1995). Transcriptional activation of the Epstein-Barr virus latency C promoter after 5-azacytidine treatment: evidence that demethylation at a single CpG site is crucial. *Mol. Cell. Biol.* 15, 6150–6159. doi: 10.1128/MCB.15.11.6150

Robinson, K. L., Tohidi-Esfahani, D., Ponton, F., Simpson, S. J., Sowrd, G. A., and Lo, N. (2015). Alternative migratory locust phenotypes are associated with differences in the expression of genes encoding the methylation machinery. *Insect Mol. Biol.* 25, 105–115. doi: 10.1111/imb.12203

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26, 139-140. doi:10.1093/bioinformatics/btp616

Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L., and Almouzni, G. (2015). Epigenomics: Roadmap for regulation. *Nature* 518, 314-316. doi: 10.1038/518314a

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., et al. (2008). The genome of the model beetle and pest Tribolium castaneum. Nature 452, 949–955. doi: 10.1038/nature06784

Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L., and Almouzni, G. (2015). Epigenomics: roadmap for regulation. Nature 518, 314–316. doi: 10.1038/518314a

Salem, H., Kreutzer, E., Sudakaran, S., and Kaltenpoth, M. (2012). Actinobacteria as essential symbionts in firebugs and cotton strainers (Hemiptera, Pyrrhocoridae). Environ. Microbiol. 15, 1956–1968. doi: 10.1111/1462-2920.12001

Sandström J., and Pettersson, J. (1994). Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. *J. Insect Physiol.* 40, 947-955. doi: 10.1016/0022-1910(94)90133-3

Scaraffia, P. Y., Zhang, Q., Thorson, K., Wysocki, V. H., and Miesfeld, R. L. (2010). Differential ammonia metabolism in *Aedes aegypti* fat body and midgut tissues. *J. Insect Physiol.* 56, 1040-1049. doi: 10.1016/j.jinsphys.2010.02.016

Scarborough, C. L., Ferrari, J., and Godfray, H. C. J. (2005). Aphid protected from pathogen by endosymbiont. Science 310, 1781–1781. doi: 10.1126/science.1120180

Schwenk, R. W., Vogel, H., and Schürmann, A. (2013). Genetic and epigenetic control of metabolic health. Mol. Metab. 2, 337–347. doi: 10.1016/j.molmet.2013.09.002

Scott, J.G. (1999). Cytochromes P450 and insecticide resistance. *Insect Biochem. Mol. Biol*. 29, 757-777. doi: 10.1016/S0965-1748(99)00038-7

Scott, J. G., Liu, N., and Wen, Z. (1998). Insect cytochromes P450: diversity, insecticide resistance and tolerance to plant toxins. *Comp. Biochem. Physio.l C Pharmacol. Toxicol. Endocrinol*. 121, 147-155.

Scott, J. J., Oh, D. C., Yuceer, M. C., Klepzig, K. D., Clardy, J., and Currie, C. R. (2008). Bacterial protection of beetle-fungus mutualism. Science 322, 63. doi: 10.1126/science.1160423

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407, 81–86. doi: 10.1038/35024074

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B. et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79. doi: 10.1038/ nature10442

Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S. *et al.* (2016) Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep*. 6, 19967. doi: 10.1038/srep19967

Smith, C. R., Mutti, N. S., Jasper, W. C., Naidu, A., Smith, C. D., and Gadau, J. (2012). Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. *PLoS ONE* 7, e42433. doi: 10.1371/journal.pone.0042433

Snell-Rood, E. C., Troth, A., and Moczek, A. P. (2012). DNA methylation as a mechanism of nutritional plasticity: limited support from horned beetles. *J. Exp. Zool. Part B Mol. Dev. Evol.* 320, 22–34. doi: 10.1002/jez.b.22479

Stoll, S., Feldhaar, H., Fraunholz, M. J., and Gross, R. (2010). Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiol.* 10, 308. doi: 10.1186/1471-2180-10-308

Stouthamer, R., and Werren, J. H. (1993). Microbes associated with parthenogenesis in wasps of the genus Trichogramma. J. Invertebr. Pathol. 61, 6–9. doi: 10.1006/jipa.1993.1002

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., *et al*. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Nat. Acad. Sci. USA* 102, 15545-15550. doi: 10.1073/pnas.0506580102

Sudakaran, S., Kost, C., and Kaltenpoth, M. (2017). Symbiont acquisition and replancement as a source of ecological innovation. Trends Microbiol. 25, 375-390. doi: 10.1016/j.tim.2017.02.014

Suzuki, M. M., Kerr, A. R. W., De Sousa, D., and Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* 17, 625-631. doi: 10.1101/gr.6163007

Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. 9, 465–476. doi: 10.1038/nrg2341

Takahashi, K., Sugi, Y., Nakano, K., Tsuda, M., Kurihara, K., Hosono, A., et al. (2011). Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. J. Biol. Chem. 286, 35755–35762. doi: 10.1074/jbc.M111.271007

Tammen, S. A., Friso, S., and Choi, S. W. (2013). Epigenetics: The link between nature and nurture. *Mol. Aspects Med.* 34, 753-764. doi: 10.1016/j.mam.2012.07.018

Teixeira, L., Ferreira, A., and Ashburner, M. (2008). The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biol. 6, e2. doi: 10.1371/journal.pbio.1000002

Terrapon, N., Cai, L., Robertson, H., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636. doi: 10.1038/ncomms4636

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7, 562-578. doi: 10.1038/nprot.2012.016

Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029-1033. doi: 10.1126/science.1160809

Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., *et al.* (2014). Insects recycle endosymbionts when the benefit is over. *Curr. Biol.* 24, 2267-2273. doi: 10.1016/j.cub.2014.07.065

Voegtlin, D. J., Halbert, S. E., and Qiao, G. (2004). A Guide to Separating *Aphis glycines* Matsumura and Morphologically Similar Species that Share Its Hosts. *Ann. Entomol. Soc. Am.* 97, 227–232. doi: 10.1093/aesa/97.2.227

Vogel, K. J., and Moran, N. A. (2011). Sources of variation in dietary requirements in an obligate nutritional symbiosis. *Proc. R. Soc. Lond. B Biol. Sci.* **b** 115-121. doi: 10.1098/rspb.2010.1304

Vorburger, C., Sandrock, C., Gouskov, A., Castaneda, L. E., and Ferrari, J. (2009). Genotypic variation and the role of defensive endosymbionts in an all- parthenogenetic host-parasitoid interaction. Evolution 63, 1439–1450. doi: 10.1111/j.1558-5646.2009.00660.x

Walsh, T. K., Brisson, J. A., Robertson, H. M., Gordon, K., Jaubert-Possamai, S., Tagu, D., et al. (2010). A functional DNA methylation system in the pea aphid, Acyrthosiphone pisum. Insect Mol. Biol. 19, 215–228. doi: 10.1111/j.1365-2583.2009.00974.x

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., et al. (2004). Epigenetic programming by maternal behavior. Nat. Neurosci. 7, 847–854. doi: 10.1038/nn1276

Wedd, L., Kucharski, R., and Maleszka, R. (2016). Differentially methylated obligatory epialleles modulate context-dependent *LAM* gene expression in the honeybee *Apis mellifera*. *Epigenetics* 11, 1-10. doi: 10.1080/15592294.2015.110-7695

Wenger, J. A., Cassone, B. J., Legeai, F., Johnston, J. S., Bansal, R., *et al*. (2017). Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.* doi: 10.1016/j.ibmb.2017.01.005

Wernegreen, J. J. (2004). Endosymbiosis: lessons in conflict resolution. *PLoS Biol.* 2, e68. doi: 10.1371/journal.pbio.0020068

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* 327, 343–348. doi: 10.1126/science.1178028

Wilson, A. C., Ashton, P. D., Calevro, F., Charles, H., Colella, S., et al., (2010). Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. Insect Mol. Biol. 19, 249-258. doi: 10.1111/j.1365-2583.2009.00942.x

Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat. Rev. Microbiol. 4, 183–192. doi: 10.1038/nrmicro1350

Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., et al. (2010). Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nat. Biotechnol. 28, 516–520. doi: 10.1038/nbt.1626

Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L., and Reinberg, D. (2015). DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu. Rev. Entomol. 60, 435–452. doi: 10.1146/annurev-ento-010814-020803

Yates, A. D., and Michel, A. (2018). Mechanisms of aphid adaptation to host plant resistance. *Curr Opin Insect Sci.* 26, 41-49. doi: 10.1016/j.cois.2018.01.003

Ye, Y. H., Wollfit, M., Huttley, G. A., Rancès, E., Caragata, E. P., Popovici, J., et al. (2013). Infection with a virulent strain of Wolbachia disrupts genome wide-patterns of cytosine methylation in the mosquito Aedes aegypti. PLoS ONE 8:e66482. doi: 10.1371/journal.pone.0066482

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., et al. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev. 29, 1256–1270. doi: 10.1101/gad.257840.114

Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Scince* 328, 916-919. doi: 10.1126/science.1186366

Zhang, G., Hussain, M., O'Neill, S. L., and Asgari, S. (2013). Wolbachia uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in Aedes aegypti. Proc. Natl. Acad. Sci. U.S.A. 110, 10276–10281. doi: 10.1073/pnas.1303603110

Zhang, J., Xing, Y., Li, Y., Yin, C., Ge, C., and Li, F. (2015). DNA methyltransferases have an essential role in female fecundity in brown planthopper, Nilaparvata lugens. Biochem. Biophys. Res. Commun. 464, 83–88. doi: 10.1016/j.bbrc.2015.05.114

Zhang, M., Chen, J., Liang, S., Li, G., Wang, F., and Ahmad, I. (2015). Differentially methylated genomic fragments related with sexual dimorphism of rice pests, Sogatella furcifera. Insect Sci. 22, 731–738. doi: 10.1111/1744-7917.12179

Zhang, Q. L., Su, L. E., and Lu, Z. Q. (2015). Expression of DNA methylation and histone acetylation related genes in response to bacterial infection in the silkworm, *Bombyx mori*. *Acta Entomol. Sin.* 58, 941–949.

Zhou, X., Chen, J., Shike Liang, M., and Wang, F. (2013). Differential DNA methylation between two wing phenotypes adults of *Sogatella furcifera*. *Genesis* 51, 819–826. doi: 10.1002/dvg.22722

Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol. Mol. Biol. Rev. 68, 745-770. doi: 10.1128/MMBR.68.4.745-770.2004

Zwier, M. V., Verhulst, E. C., Zwahlen, R. D., Beukeboom, L. W., and van de Zande, L. (2012). DNA methylation plays a crucial role during early *Nasonia* development. *Insect Mol. Biol.* 21, 129–138. doi: 10.1111/j.1365-2583.2011.01121.x