Open Access Dissertations
Electronic Theses and Dissertations

2016-04-26

# The Role of Gene Co-option in the Evolution of Sap-Feeding Insect Endosymbioses (Hemiptera: Sternorrhyncha)

Rebecca Pate Duncan
*University of Miami,* reba.duncan@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

UNIVERSITY OF MIAMI


THE ROLE OF GENE CO-OPTION IN THE EVOLUTION OF SAP-FEEDING
INSECT ENDOSYMBIOSES (HEMIPTERA: STERNORRHYNCHA)



By

Rebecca Pate Duncan


A  DISSERTATION


Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy




Coral Gables, Florida

May 2016

UNIVERSITY OF MIAMI


A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy


THE ROLE OF GENE CO-OPTION IN THE EVOLUTION OF SAP-FEEDING
INSECT ENDOSYMBIOSES (HEMIPTERA: STERNORRHYNCHA)


Rebecca Pate Duncan


Approved:


_____
Alex C. C. Wilson, Ph.D.
Associate Professor of Biology

_____
Julia E. Dallman, Ph.D.
Assistant Professor of Biology


_____
Charles W. Luetje, Ph.D.
Professor of Molecular and Cellular
Pharmacology

_____
Guillermo Prado, Ph.D.
Dean of the Graduate School


_____
Angela E. Douglas, Ph.D.
Daljit S. and Elaine Sarkaria Professor
of Insect Physiology and Toxicology
Cornell University

DUNCAN, REBECCA PATE                                (Ph.D., Biology)

<u>The Role of Gene Co-option in the Evolution of</u>                (May 2016)
<u>Sap-Feeding Insect Endosymbioses</u>
<u>(Hemiptera: Sternorrhyncha)</u>

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Associate Professor Alexandra C.C. Wilson.
No. of pages in text. (169)

Symbiosis, the tightly co-evolved partnership between different species, fundamentally influences the evolutionary trajectory of the interacting species. However, with the help of emerging genomic resources, the mechanisms by which symbiotic partners are integrated into a functional organism are only just beginning to be revealed. As a system in which to investigate the mechanisms of host/symbiont integration, sap-feeding insects, which rely on intracellular symbionts for nutrient provisioning, benefit from growing genomic resources and a solid foundation resulting from years of research on the pea aphid, *Acyrthosiphon pisum*. At the core of the relationship between *A. pisum* and its endosymbiont *Buchnera aphidicola* is amino acid metabolism and exchange, and previous research suggests that amino acid exchange is mediated by amino acid transporters that were co-opted to operate in a symbiotic context. Inspired by recent discoveries in *A. pisum*, I use sap-feeding insects of the hemipteran suborder Sternorrhyncha to investigate the role of co-option of amino acid transporters in host/symbiont integration. In four chapters, I examine co-option with gene duplication and co-option in a single-copy amino acid transporter. Using microarray and RNA-seq data in the first three data chapters, I found support for both symbiotic and non-symbiotic traits driving the retention of duplicate amino acid transporters both at the level of the

suborder Sternorrhyncha and at the level of the family Aphididae. Further, co-option of duplicate amino acid transporters to the symbiotic interface involved expression evolution, and possibly functional evolution as well. In the fourth data chapter, I combined gene annotation, expression analysis, and functional characterization to investigate the evolutionary mechanism by which the single-copy *A. pisum* glutamine transporter ApGLNT1 was co-opted to be a key regulator of amino acid metabolism in *A, pisum*. Using electrophysiology, I functionally characterized ApGLNT1 orthologs from other sap-feeding and non sap-feeding insects. I discovered that ApGLNT1 is functionally conserved, indicating that ApGLNT1 evolved its role in symbiosis within its ancestral functional constraints. Together, these four studies indicate that co-option has played a role in metabolic integration of sternorrhynchan endosymbioses by recruiting amino acid transporters to operate in the novel context of symbiosis.

*To all of my teachers*
*Living and non-living*
*Young and old*
*Inner and outer*
*Human and non-human*
*Real and mythical*
*Past, present, and future*

# ACKNOWLEDGEMENTS

I would like to thank all of the current and past members of the Wilson lab that provided thoughtful discussion, feedback, help with experimental design and data analysis, as well as good company. I am also thankful to Luetje lab members who showed me the ropes of electrophysiology for Chapter 5 of this dissertation. In particular, I thank Dan Price, Fil Kushlan, Honglin Feng, Hsiao-Ling Lu, Devin Kepchia, Suhaila Rahman, Ben Sherman, and Ana Castro. I thank Doug Nguyen for assistance with data analysis in Chapter 4. I also am grateful for the generous support of colleagues who welcomed me to their labs to collect tissues and/or shared data sets that improved my comparative analyses. These colleagues include John McCutcheon, Cecilia Tambourindeguy, Kirsten Stelinski, Ellen Dotson, Don Gilbert, Pedro Lagerblad de Oliveira, Patrick Abbot, Nicole Gerardo, François Delmotte, and graduate students, postdocs, and/or staff in their labs. I thank Amy Roda and Scott Weihman for help collecting *P. citri* specimens in Miami, FL. I also thank all of my co-authors for their help in the published chapters of this dissertation, particularly Lubov Nathanson, Filip Husnik, JT Van Leuven, Honglin Feng, John McCutcheon, Don Gilbert, and Liliana Dávalos.

I thank my current committee members Angela Douglas, Julia Dallman, Chuck Luetje, as well as past committee members Bill Browne, Doug Crawford, and Dale Hedges for their thoughtful input on my research. I especially thank Chuck Luetje for opening his lab to me and supporting a collaborative Chapter 5 of this dissertation. And

of course, I am thankful for the guidance, support, and mentorship of my research advisor, Alex Wilson.

I thank my family for their love and support throughout graduate school: Susie and Gilbert Duncan, Michael Muench, Karl and Any Muench, and especially Arlo Muench, whose soul lights up my life more than anything in the world.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

**Chapter 1**

Introduction

How evolution generates novelty is one of the most fundamental questions in evolutionary biology. The tools to elucidate the genetic and genomic basis of novelty, however, have only recently been developed. Evolutionary novelty and its origins have captivated biologists since Darwin's time (Darwin 1859, Mayr 1963, Nitecki 1990, Pigliucci 2008, Shubin et al. 2009, Wagner & Lynch 2010, West-Eberhard 2008). Today, novelty often refers to morphological traits characterizing a particular species or higher taxonomic group, such as wing spots on butterflies (True & Carroll 2002), vertebrate eyes (Shubin et al. 2009), or the transition from fins to limbs (Spitz et al. 2001). Novelty, however, also evolves in other types of traits, like chemical defenses, organs, and behavior. These traits all have an underlying genetic basis, the discovery of which has been made possible by molecular genetic tools and high-throughput genome and transcriptome sequencing. For example, comparative sequence and expression data, often coupled with functional information, illuminate the evolutionary origins of snake venom cocktails (Hargreaves et al. 2014), a novel electrical communication organ in fish (Arnegard et al. 2010), and the shift in behavioral response to chemical signals found in herbivorous drosophilids that evolved from microbe-feeding ancestors (Goldman-Huertas et al. 2015). However, equally intriguing outstanding questions about the evolution of novelty apply to interactions between divergent organisms. For example, what mechanisms drive the evolution of herbivory (Goldman-Huertas et al. 2015, Russell et al. 2009) the shift in flowering plants from one pollinator to another (Schemske & Bradshaw 1999), and the origin and maintenance of intimate host/symbiont interactions (Duncan et

1

al. 2014, Gerardo et al. 2010, Lu et al. 2016, Macdonald et al. 2012, Price et al. 2011, 2014; Shigenobu & Stern 2013)?  Elucidating how novel interspecific relationships evolve is important for understanding large-scale biological systems and the evolution of biological diversity – interspecific interactions are critical to ecosystem function and the evolutionary trajectories of the interacting species.  Molecular genetic and genomic tools are also instrumental in uncovering the basis of many interspecific interactions, particularly those that must overcome metabolic (Hansen & Moran 2014, Salem et al. 2014, Tsai et al. 2013, Wilson & Duncan 2015), immunological (Foth et al. 2014, Gerardo et al. 2010, International Glossina Genome Initiative 2014, Tsai et al. 2013) or chemical challenges (Gloss et al. 2014, Tsai et al. 2013).  Similarly, these tools are necessary for understanding intimate interactions that take place at the cellular or subcellular level, such as interactions involving at least one unicellular organism (Nakabachi et al. 2014, Price et al. 2014, Russell et al. 2013).

**Symbiosis**

Symbiotic interactions are the most intimate interspecific interactions, often taking place between tightly co-evolving species.  Symbiosis is broadly defined as the interaction between two different organisms living in close physical association.  In symbiotic interactions, one species is called the host and the other is the symbiont. Superficially, symbiotic interactions can be grouped into different categories that fall along a continuum depending on whether they benefit or harm the host and symbiont. Interactions that harm the host but benefit the symbiont are called parasitism (e.g. tapeworms in vertebrate hosts), interactions that benefit the symbiont but neither benefit

nor harm the host are called commensalism (e.g. epiphytes on trees), and interactions that benefit both the host and symbiont are known as mutualism (e.g. angiosperms and pollinators). The intimacy and tight co-evolution of symbiotic interactions often results in co-evolutionary arms races between host and symbiont (Davies et al. 1989) or interdependence between host and symbiont (Wilson & Duncan 2015), making symbioses ideal for identifying genomic mechanisms underlying the evolution of novel interspecific interactions. Genome and transcriptome studies on hosts and symbionts are increasingly revealing genomic patterns that may be associated with the evolution of symbiosis, particularly parasitism (Foth et al. 2014, Kirkness et al. 2010, Tsai et al. 2013, Whiteman et al. 2012) and mutualism (Baumgarten et al. 2015, Ganot et al. 2011, Hansen & Moran 2011, Husnik et al. 2013, International Aphid Genomics Consortium 2010, Macdonald et al. 2012, Nakabachi et al. 2006, Shigenobu et al. 2000, Sloan et al. 2014, Young et al. 2011). These studies pave the way for functional studies to test hypotheses generated by high throughput sequencing data.

**Study System**

Particularly good models for studying the genomic basis of symbiosis are sap-feeding insects such as aphids, mealybugs, and psyllids. These insects, members of the insect order Hemiptera, have evolved intimately interdependent mutualisms with obligate bacterial symbionts that supplement their exclusive plant sap diet with vitamins and essential amino acids (McCutcheon & Dohlen 2011, Nakabachi et al. 2006, Shigenobu et al. 2000). In exchange, host insects provide their symbionts with a stable line of succession as well as non-essential amino acids that symbionts need for their own protein

synthesis and as precursors for synthesizing essential amino acids (Shigenobu & Wilson 2011). What makes these mutualisms so suitable for understanding the genomic basis of symbiosis is their extraordinary intimacy. Nearly all sap-feeding insects from the hemipteran suborders Sternorryncha and Auchenorrhyncha harbor their symbionts intracellularly in specialized organs called bacteriomes (Baumann 2005). Inside bacteriome cells (bacteriocytes), symbionts are enclosed within membrane-bound compartments. The compartmentalization of sap-feeding insect symbionts implies that interactions between host insects and their symbionts take place at cellular and sub-cellular levels, mediated by genes encoded in the host insect and symbiont genomes. In addition to anatomical intimacy between sap-feeding insects and their symbionts, their relationships are also extremely close because of long host/symbiont co-evolutionary history. Sap-feeding insects transmit their symbionts vertically from mother to offspring (Baumann 2005), and in some cases have done so for hundreds of millions of years (Moran et al. 1993) to the point that they have become so developmentally and metabolically integrated that neither partner can survive without the other. Integration between sap-feeding insects and their symbionts coupled with compartmentalization of symbionts within specialized insect cells is reflected in signatures that can be detected at the genomic level (Wilson & Duncan 2015).

**Host/symbiont integration in sap-feeding insects**

Since publication of the genome of the pea aphid, *Acyrthosiphon pisum* (International Aphid Genomics Consortium 2010), genomic and transcriptomic studies have rapidly advanced our understanding of host/symbiont developmental and metabolic

integration.  Most work on developmental integration has been conducted in *A. pisum*, and here I highlight some recent advances.  Using transcriptome sequencing and analysis combined with *in situ* hybridization, Shigenobu and Stern (2013) discovered several novel, bacteriocyte-specific genes encoding proteins with putative signal peptides.  Transcript localization of seven of these novel genes coincided with localization of the aphid symbiont *Buchnera aphidicola* from the initial infection of aphid blastoderm-stage embryos (stage 7) through to adulthood.  The signal peptides of the encoded proteins suggest that they are secreted either to the extracellular space surrounding bacteriocytes, or possibly to intracellular compartments housing *Buchnera* cells.  While the authors do not have direct evidence for the function of these novel secreted proteins, they speculate, on the basis of symbiont-targeting proteins in weevils and legumes, that the proteins mediate *A. pisum*/*Buchnera* interactions and may regulate *Buchnera* proliferation during development.  Three other recent studies have found differences in gene expression and protein localization between maternal and embryonic bacteriocytes, indicating that *Buchnera* function is integrated into the *A. pisum* developmental program.  Using immunolocalization, Nakabachi et al. (2014) investigated the cellular localization of the protein encoded by a bacterial gene that was transferred to the *A. pisum* genome, *RlpA4*.  Remarkably, RlpA4 localizes to *Buchnera* cells in maternal bacteriocytes, providing the first evidence for host/symbiont integration to the same degree of intimacy as organelles.  Significantly, RlpA4 is not found in embryonic bacteriocytes.  The function of RlpA4 is currently unknown.  Hansen and Degnan (2014) also found developmental differences in *Buchnera* gene expression.  While they found no transcriptional regulation between *Buchnera* from maternal and embryonic bacteriocytes, protein-level regulation was

evident. Some of the proteins that were differentially regulated were amino acid biosynthesis genes, suggesting that the nutritional role of *Buchnera* changes during development. More evidence for differences in nutrient provisioning through development is reported by Lu et al. (2016). In this study, the authors investigated the localization of a glutamine transporter, ApGLNT1 (see below), during *A. pisum* development. While ApGLNT1 localizes to the bacteriocyte plasma membrane in maternal bacteriocytes, supplying *Buchnera* with an important precursor for amino acid biosynthesis (Price et al. 2014), the transporter does not localize to the membrane of embryonic bacteriocytes. Instead, ApGLNT1 localizes to the maternal follicular epithelium surrounding the ovariole and embryonic epithelium surrounding individual embryos. The localization of ApGLNT1 around embryos suggests that it may play a role in supplying embryos with glutamine from maternal hemolymph. While these studies have filled some large gaps in our understanding in developmental integration between *A. pisum* and *Buchnera*, they also highlight areas where further research is needed.

Metabolic integration between sap-feeding insects and their symbionts is generally better understood than developmental integration. Genomic and transcriptomic data for *A. pisum*, the citrus mealybug *Planococcus citri*, the hackberry petiole gall psyllid *Pachypsylla venusta*, and the whitefly *Bemisia tabaci* predict that complete biosynthesis pathways for vitamins, some amino acids, and even peptidoglycan (in the *P. citri* symbiont *Moranella*) require combinations of eukaryotic and prokaryotic genes encoded in host and symbiont genomes (Husnik et al. 2013, Luan et al. 2015, McCutcheon & Dohlen 2011, Price & Wilson 2014, Sloan et al. 2014, Wilson et al. 2010). Experimental validation of host/symbiont metabolic collaboration has been demonstrated

for some amino acids in *A. pisum*. Investigating the prediction that the terminal steps of leucine and phenylalanine biosynthesis were carried out by host aphid enzymes, Russell et al. (2013) examined leucine or phenylalanine output from isolated *Buchnera* with and without *Buchnera*-free host cell fractions. They found that leucine and phenylalanine output significantly increased with the addition of host cell fractions. Further, the aphid enzymes carrying out those terminal reactions, Branched chain amino acid transaminase (EC 2.6.1.42) and Phenylalanine aminotransferase (EC 2.6.1.58), were enriched in bacteriocytes and their activity was detected through *in vitro* assays of host cell fractions. Taken together, these data indicate that host aphid enzymes are required to complete the *Buchnera* pathways for leucine and phenylalanine biosynthesis. Additionally, multiple studies have provided transcriptional evidence for host/symbiont metabolic collaboration, showing that transcripts of host genes predicted to be involved in collaborative biosynthesis are enriched in bacteriocytes (Hansen & Moran 2011, Husnik et al. 2013, Price & Wilson 2014, Sloan et al. 2014).

In addition to metabolic enzymes, evidence is emerging that host/symbiont metabolic integration also relies on transporter proteins. Membrane-bound transporter proteins are likely important players in the metabolic integration of sap-feeding insects with their symbionts because the interface between host insect tissues and the symbiont is made up of multiple membranes. In *A. pisum*, the host/symbiont interface is made up of three types of membranes: (1) The bacteriocyte plasma membrane, (2) the aphid-derived symbiosomal membrane that surrounds individual *Buchnera* cells, and (3) bacterial membranes consisting of *Buchnera* inner and outer membranes. Together, these membranes form a barrier to nutrient exchange between *Buchnera* and the aphid host

tissues outside of the bacteriome. While many types of transporters are likely important to these systems, I focus on amino acid transporters here because of the centrality of amino acid biosynthesis and exchange in sap-feeding insect symbioses. After the *A. pisum* genome was published (International Aphid Genomics Consortium 2010), Price et al. (2011) annotated amino acid transporters in the *A. pisum* genome belonging to two gene families: The amino acid polyamine organocation (APC) family (Transporter Classification (TC) #2.A.3) and the amino acid/auxin permease (AAAP) family (TC #2.A.18). They also examined expression profiles of APC and AAAP members. This study, the results of which are described below, inspired my dissertation by providing the first evidence that co-option of amino acid transporters played a key role in the metabolic integration of sap-feeding insects and their endosymbionts.

**Gene co-option and metabolic integration in sap-feeding insect endosymbioses**

There are many mechanisms by which natural selection can produce novel structures, genes, behaviors and relationships, but one that is particularly intriguing for sap-feeding insect symbiosis is gene co-option. Gene co-option (reviewed by True & Carroll 2002) happens when natural selection finds new functional roles for existing genes. Gene co-option provides a mechanism for the evolution of novel gene roles without creating separate genes *de novo*, and based on many well understood examples, it is clear that gene co-option has been an important mechanism driving evolutionary innovation across the tree of life. For example, the refractive function of vertebrate lens crystallins and the ice-binding function of teleost fish antifreeze proteins both evolved through co-option of genes with ancestral enzymatic roles (True & Carroll 2002).

Similarly, co-option of actins is thought be important in the evolution and diversification of cell types in both plants and metazoans, suggesting a role in the evolution of multicellularity (True & Carroll 2002). Given what previous studies have established about amino acid transporter gene families, amino acid transporter co-option may play an especially significant role in the evolution of host/symbiont interactions and metabolic integration in sap-feeding insects. In the following sections, I will explain different ways that gene co-option can take place and how each relates to amino acid transporters in the sap-feeding insect *A. pisum*.

*Gene co-option facilitated by gene duplication*

One way that gene co-option occurs is through gene duplication followed by functional divergence of daughter genes. Gene duplication is widely recognized as a mechanism driving the evolution of novel genes and other traits by producing raw material visible to natural selection. Following a gene duplication event, there are several possible outcomes, some of which lead to functional divergence of duplicates (paralogs). Short of paralog loss, the most common outcome (Lynch & Conery 2000), or functional conservancy for benefits of increased gene dosage, paralogs can diverge functionally in two general ways: 1) One paralog maintains the ancestral function while the other, released from purifying selection, is free to evolve a new function (neofunctionalization). 2) Alternatively, paralogs may partition the ancestral function(s) such that both paralogs are needed to maintain the ancestral functional role of the single, pre-duplication gene (subfunctionalization). Importantly, "functional" divergence includes both the evolution in the coding region, resulting in new or altered functional domains in the gene product,

as well as divergence in expression patterns across time and space, mediated by evolution in *cis*-regulatory domains controlling gene expression.

While neofunctionalization more obviously leads to gene co-option (by facilitating recruitment of a gene to a novel cell type or developmental stage and/or enabling a shift in e.g. enzyme activity or substrate specificity), subfunctionalization is equally capable of driving gene co-option. For example, a single multifunctional gene may experience evolutionary conflicts between its different functions, such that selection to optimize one functional role will compromise the other. In these cases, duplication can release these multifunctional genes from conflict, allowing selection to favor mutations optimizing conflicting functions in distinct paralogs (Marais & Rausher 2008). For example, ancient gene duplication events in mammalian hemoglobin led to several hemoglobin variants. These variants are expressed in different developmental stages and each is optimized to different oxygen requirements at different times during development (Wolpert et al. 2011), indicating that functional specialization involved evolution in both regulatory and coding regions of hemoglobin genes.

Gene duplication in amino acid transporters could have facilitated gene co-option in the evolution of symbiosis between sap-feeding insects and intracellular bacterial symbionts. In the pea aphid *A. pisum*, nutrient amino acid transporters in the APC and AAAP families underwent numerous taxon-specific gene duplications, resulting in more amino acid transporters in *A. pisum* than in non sap-feeding insects (Price et al. 2011). Among paralagous aphid transporters, functional divergence is supported by both gene expression analyses and computational tests for signatures of selection. Using quantitative reverse transcriptase PCR (qRT-PCR), Price *et al.* (2011) showed that some

amino acid transporter paralogs are highly expressed and enriched in *A. pisum* bacteriocytes while related paralogs are expressed in different tissues. These results suggest that after gene duplication, some paralogs were recruited to bacteriocytes as a result of selection to specialize in host/symbiont amino acid exchange. Further evidence for functional divergence and gene co-option was uncovered in a large, 10-paralog *slimfast* expansion, named for the related single-copy *Drosophila melanogaster* gene (ortholog), through computational tests for signatures of selection. These tests revealed an elevated rate of evolution along a branch leading to a loss of gut expression (dominant among *slimfast* paralogs) preceding the evolution of bacteriocyte expression, consistent with a release from purifying selection in the coding sequence (Price et al. 2011). Release from purifying selection could have resulted in a shift in transport activity either through neofunctionalization or subfunctionalization. Evolution in transport activity coinciding with the shift to bacteriocyte-specific expression is intriguing, suggesting a role for gene duplication in facilitating the co-option of ancestral amino acid transporters towards a novel, symbiotic role.

*Gene co-option in single copy genes*

Although gene duplication was once thought to be necessary for the evolution of novel gene functions (Ohno 1970), gene co-option can also occur in single-copy genes. Single-copy genes can be co-opted to perform new functions without compromising ancestral function either by evolving a new functional domain or by acquiring substitutions in the coding region that optimize a minor function in a promiscuous enzyme. For example, the enzyme encoded by the *Buchnera* gene *ilvC*, involved in the

conversion of pyruvate to 2-oxoisovalerate (involved in both valine and pantothenate (vitamin B5) biosynthesis), additionally catalyzes the conversion of 2-dehydropantoate to pantoate, the penultimate step in panthothenate biosynthesis. The latter step is typically carried out by the *panE* gene product, but *panE* is missing from the *Buchnera* genome. The dual function of *Buchnera ilvC* is supported by Price et al. (2014), who showed that *Buchnera ilvC* can rescue an *Escherichia coli* pantothenate auxotroph (*ilvC⁻/panE⁻*). In contrast, transformation with *E. coli ilvC* results in extremely limited growth. Thus, the ability to complement *panE* was presumably facilitated by substitutions in *Buchnera ilvC* that optimized the ancestral, limited *panE* activity without compromising *ilvC* activity.

Alternatively, single-copy genes can acquire novel *cis*-regulatory elements, enabling them to be re-deployed in new tissues or developmental contexts while the ancestral function of the coding region remains conserved. Gene co-option through changes to expression across time and space has been observed in classic developmental toolkit genes involved in the evolution of new plant and animal body plans. For instance, compound leaves in plants evolved from simple leaves by co-opting transcription factors like KNOX1 to be expressed not only in the apical meristem of the primary leaf, but also in the distal bulges of primary leaf primordia from which leaflets emerge, indicating the emergence of novel *cis*-regulatory regions regulating expression (True & Carroll 2002). The evolution of tetrapod limbs from fish fins is also associated with acquisition of novel enhancers regulating expression of the *HoxD* complex, involved in vertebrate limb development (Spitz et al. 2001).

Co-option in single copy amino acid transporters could also have played a role in the origin and maintenance of symbiosis in aphids and other sap-feeding insects. Indeed,

some of the most highly expressed amino acid transporters in the bacteriocytes of *A. pisum* are single-copy and tend to also be expressed in multiple tissues (Price et al. 2011). Functional characterization of the gene product of one single-copy amino acid transporter, *ApGLNT1*, confirms its importance in the *A. pisum*/*Buchnera* endosymbiosis by providing *A. pisum* with a mechanism to regulate *Buchnera* metabolism (Price et al. 2014, 2015). *ApGLNT1* localizes to the *A. pisum* bacteriocyte plasma membrane where it transports glutamine, an important precursor to amino acid metabolism in bacteriocytes by both *A. pisum* and *Buchnera*, from aphid hemolymph into bacteriocytes. Intriguingly, glutamine transport is inhibited by the presence of a metabolically downstream *Buchnera*-produced amino acid, arginine (Price et al. 2014). These data support an elegant model for how *A. pisum* can regulate bacteriocyte metabolic output through feedback inhibition by arginine in response to demand. Since glutamine is required as a precursor for both *A. pisum* and *Buchnera* mediated amino acid biosynthesis in bacteriocytes, increasing levels of arginine in the aphid hemolymph effectively shuts down amino acid biosynthesis in bacteriocytes. If the function of *ApGLNT1* is derived in *A. pisum* and other sap-feeding insects, then that would imply that evolution co-opted *ApGLNT1* towards a novel function in response to selection for a host-controlled regulatory mechanism in sap-feeding insect symbioses. Alternatively, if *ApGLNT1* function is conserved, then selection for its ancestral function in bacteriocytes could have driven this transporter to evolve expression in a novel, symbiosis related organ (the bacteriome).

**Research Aims**

In this dissertation, I examined the role of co-option of amino acid transporters in the evolution of symbiosis in sap-feeding insects like *A. pisum* with intracellular, bacterial symbionts. My overarching aims were to investigate co-option 1) involving gene duplication and 2) co-option in single-copy genes. I tested the role of amino acid transporter co-option involving gene duplication from three different perspectives in Chapters 2, 3 and 4 of this dissertation. In Chapter 2, I examined an alternative, sex-biased explanation for gene duplication in the *slimfast* expansion in *A. pisum* and the green peach aphid *Myzus persicae*. In Chapter 3, I used a comparative transcriptomic approach to test the importance of gene duplication in symbiosis across several different sap-feeding insect families. In Chapter 4, I also used a comparative transcriptomic approach to investigate the relative roles of symbiosis and other traits in maintaining amino acid transporter paralogs within the aphid family, Aphididae. In my final data chapter, Chapter 5, I examined the mechanism of gene co-option in recruiting the single-copy amino acid transporter *ApGLNT1* to bacteriocytes, where it plays a key role in host/symbiont metabolic integration. I completed Chapter 5 using a comparative phylogenetic approach combined with functional characterization, expression analysis, and genome annotation. Together, the four data chapters of this dissertation have advanced understanding of the role gene co-option plays in bringing insects and microbial symbionts together to form novel partnerships, changing the course of evolution for all organisms involved.

**Chapter 2**

Novel male-biased expression of paralogs in the aphid *slimfast* nutrient amino acid transporter expansion[1]

**Summary**

A major goal of molecular evolutionary biology is to understand the fate and consequences of duplicated genes. In this context, aphids are intriguing because the newly sequenced pea aphid genome harbors an extraordinary number of lineage-specific gene duplication relative to other insect genomes. Though many of their duplicated genes may be involved in their complex life cycle, duplications in nutrient amino acid transporters appear to be associated rather with their essential amino acid poor diet and the intracellular symbiosis aphids rely on to compensate for dietary deficits. Past work has shown that some duplicated amino acid transporters are highly expressed in the specialized cells housing the symbionts, including a paralog of an aphid-specific expansion homologous to the *Drosophila* gene *slimfast*. Previous data provide evidence that these bacteriocyte-expressed transporters mediate amino acid exchange between aphids and their symbionts. I report that some nutrient amino acid transporters show male-biased expression. Male-biased expression characterizes three paralogs in the aphid-specific *slimfast* expansion, and the male-biased expression is conserved across two aphid species for at least two paralogs. One of the male-biased paralogs has additionally experienced an accelerated rate of non-synonymous substitutions. This is the first study to document male-biased *slimfast* expression. My data suggest that the male-biased aphid *slimfast* paralogs diverged from their ancestral function to fill a functional

---

role in males.  Furthermore, my results provide evidence that members of the *slimfast* expansion are maintained in the aphid genome not only for the previously hypothesized role in mediating amino acid exchange between the symbiotic partners, but also for sex-specific roles.

**Background**

A fundamental goal of molecular evolutionary biology is to understand the evolutionary fate and biological consequences of duplicated genes.  Most duplicate genes fail to reach fixation (Lynch & Conery 2000), strongly suggesting that retained gene duplicates provide a selective advantage.  Recent genome sequencing projects have revealed that the water flea, *Daphnia pulex,* and the pea aphid, *Acyrthosiphon pisum,* are unique among arthropods in having exceptionally high levels of tandem, lineage-specific gene duplication and retention (Colbourne et al. 2011, Huerta-Cepas et al. 2010, International Aphid Genomics Consortium 2010, Jaubert-Possamai et al. 2010, Price et al. 2010, 2011; Rider et al. 2010, Shigenobu et al. 2010).  Apart from high gene duplication levels, both *Daphnia* and aphids, in contrast to other sequenced insects, are cyclical parthenogens with complex life cycles characterized by extensive polyphensim (Colbourne et al. 2011, International Aphid Genomics Consortium 2010).  The shared life cycle complexity and gene duplication levels of aphids and *Daphnia* suggest that polyphenism may be a major factor underlying the extensive gene duplications characterizing these two genomes.

Several aphid-specific gene family expansions have been suggested to be associated with the aphid life cycle and polyphenism (International Aphid Genomics Consortium

2010).  For example, several genes involved in early developmental signaling pathways have undergone aphid-specific duplications (Shigenobu et al. 2010), which may account for the differences observed in embryonic development between live-bearing, asexual females and egg-laying, sexual females (Miura et al. 2003, Shigenobu et al. 2010).  One major developmental difference between asexual and sexual females is the type of cell division that initiates oogenesis.  While sexual female oocytes undergo standard meiosis, asexual female oocytes undergo a modified form of meiosis before initiating development of a parthenogenetic embryo (Srinivasan et al. 2010).  The differences in cell division between asexual and sexual female oocytes might be accommodated by gene duplications found in cell cycle genes, some of which are differentially expressed in sexual and asexual female ovarioles (Srinivasan et al. 2010).  Additionally, since different morphs within the same clonal line are genetically identical, their morphological, biological and ecological differences must result from differences in gene expression (e.g. (Brisson et al. 2007)), which may be facilitated by duplicate, divergent chromatin remodeling genes (Rider et al. 2010).

In contrast to an hypothesized association with polyphenism, some aphid-specific gene duplications appear to be more connected with other aspects of aphid biology, such as their diet.  Aphids feed on plant phloem sap, a diet rich in sugar, which creates a high positive osmotic potential between the ingested phloem sap and the aphid hemolymph.  This osmotic challenge must be overcome in order for aphids to avoid desiccation (Karley et al. 2005, Price et al. 2010).  Aphids cope with their osmotically challenging diet both by voiding excess sugar as liquid waste called honeydew (Price et al. 2007), and by transporting sugars across the gut epithelium into their hemolymph via sugar

transporter proteins (Price et al. 2010). Interestingly, aphid sugar transporters have undergone aphid-specific gene duplications (Price et al. 2010), which may enable aphids to utilize a sugar rich diet.

Another group of diet-associated genes that have undergone aphid-specific duplications are the nutrient amino acid transporters (Price et al. 2011). Duplications in nutrient amino acid transporters may be explained by another feature of the aphid phloem sap diet. Apart from being rich in sugar, phloem sap is also deficient in essential amino acids. Essential amino acids cannot be synthesized *de novo* by metazoans, so their deficit in phloem is compensated for by the intracellular bacterial symbiont, *Buchnera aphidicola* (Munson et al. 1991, Shigenobu & Wilson 2011). *Buchnera* reside within membrane-bound compartments in specialized aphid cells called bacteriocytes, where they exchange non-essential amino acids from the aphid for the essential amino acids they synthesize (reviewed by Shigenobu and Wilson (2011)). Recently, Price et al. (Price et al. 2011) discovered that some nutrient amino acid transporters are highly expressed in bacteriocyte cells where *Buchnera* reside, suggesting that these amino acid transporters mediate amino acid exchange between aphids and *Buchnera.* Their finding that bacteriocyte-enriched amino acid transporters are often members of gene family expansions suggests that past duplication events enabled some duplicates to conserve their ancestral function while others diverged to fill a role in symbiosis.

Here I propose that in addition to a role in symbiosis, duplicate nutrient amino acid transporters have evolved sex-biased roles. In microarray data for the aphid species *Myzus persicae*, I unexpectedly discovered that two nutrient amino acid transporters have highly male-biased expression. These transporters, orthologous to the *Drosophila* gene

*slimfast*, were nested within an aphid-specific *slimfast* expansion among 10 *A. pisum*

paralogs.  Male-biased expression for these two paralogs is conserved in *A. pisum* and

detailed expression analysis in *A. pisum* revealed one additional male-biased *slimfast*

paralog and one asexual female-biased *slimfast* paralog.  Together, these data suggest that,

in addition to the symbiosis-based functions proposed by Price *et al*., (Price et al. 2011),

the aphid-specific *slimfast* expansion is maintained for sex-biased functions.


**Methods**

**Design, execution, and statistical analysis of microarray experiments**

I investigated gene expression based on sex using a microarray experiment

designed by my Ph.D. advisor, Alex C. C. Wilson.  Agilent microarrays targeting 10,478

*M. persicae* unigenes (Ramsey et al. 2007) and ESTs were used to quantify mRNA

expression.  The microarray measured gene expression between four treatments and two

aphid lines (W109 and W115), both collected in 2008 on tobacco in Windsor, CT by

Alex Wilson and her colleague.  The four treatments were males and three different

female reproductive morphs: sexual females, asexual females at short day, and asexual

females at long day.  Asexual females at long day were collected from standard asexual

cultures (14:10 Light:Dark, 20 ºC).  Sexual morphs were induced by changing the growth

temperature and light/dark cycle (10:14 L:D, 16 ºC) to simulate seasonal differences.

When sexual morphs appeared, males, sexual females and asexual females at short day

were collected.

Total RNA was extracted from ten flash frozen aphids per aphid line per treatment

using an RNeasy Mini Kit (Qiagen, Valencia, CA).  RNA quantity and quality was

assessed using an RNA LabChip on the Agilent 2100 Bioanalyzer.  One microgram of total RNA was labeled using an Amino Allyl MessageAmp aRNA Amplification Kit (Ambion) and Cy3- and Cy5-NHS ethers (Amersham).  Following labeling, the concentration of each sample was determined spectrophotometrically and equal amounts of labeled amplified RNA were competitively hybridized to the arrays.  Factorial design was used for this microarray experiment.  Four competitive hybridizations using the aphid line W109 (first biological replicate) were made as following: male → asexual female (long day) → sexual female → asexual female (short day) → male.  The direction of the arrow indicates Cy3 → Cy5 sample labeling.  An additional four competitive hybridizations were made with the reverse labeling (Cy5 → Cy3) using the aphid line W115 (second biological replicate).

Following hybridization and washing, microarrays were scanned at 5 μm resolution using a GenePix 4000B scanner (Molecular Devices).  The resulting images were analyzed using GenePix Pro 6.1 (Molecular Devices) and subject to quality control using Acuity 4.0 (Molecular Devices).   My colleague, Lubov Nathanson at the Hussman Institute for Human Genomics at University of Miami Miller School of Medicine, analyzed the microarray data.  Data  were analyzed using Linear Models for Microarray Data (LIMMA) implemented in R (Smyth 2005).  Data normalization was executed first within each array using loess normalization, then normalization between arrays was executed using the LIMMA quantile algorithm.  Differential expression and false discovery rates (FDR) were assessed using a linear model and empirical Bayes moderated F statistics (Smyth 2004, Smyth & Speed 2003).  All primary microarray data are accessible through the National Center for Biotechnology Information (NCBI) Gene

Expression Omnibus (GEO) database (http://www.ncbi.nih.gov/geo) under the GEO

series accession number GSE31024.

**Annotating sex-enriched genes**

I annotated sex-enriched genes in the microarray in Blast2GO (v.12.7.0) (Conesa

et al. 2005). Blast2GO implements a BLAST search and maps Gene Ontology (GO)

annotations to the query sequences based on the best BLAST hits. Next, the annotation

step evaluates candidate GO terms based on their hit similarity and annotation evidence

codes (e.g. experimentally validated, computationally validated, etc.) to assign the most

specific GO annotations possible to the original query sequences. The Blast2GO

platform can also be used to perform an InterProScan to annotate sequences based on

protein domain and motif signatures. GO terms corresponding to the InterPro annotation

are then merged with the GO annotations assigned by the annotation step.

Sex-enriched *M. persicae* sequences were input as the query sequences and

compared to the NCBI protein sequence database using a TBLASTN search, keeping up

to 50 top hits at an e-value of 0.001 or less. Annotation was performed in two rounds to

maximize both the quality and number of annotations. First, the annotation step was run

with default evidence code weights, except that the evidence code IEA (Inferred from

Electronic Annotation) was weighted 0. For the second round, unannotated sequences

were passed through the annotation step using default evidence code weights. After the

annotation step, I ran the InterProScan and merged the InterPro results with the previous

annotations. Annotations were then augmented using Annex, a database of manually

reviewed relationships between the three GO categories (Myhre et al. 2006). Finally, the

GO-slim function was used to summarize the annotation results.  A fisher exact test was implemented in Blast2GO to examine overrepresented GO terms in male-enriched and female-enriched genes.  Significance was assessed by a FDR of 0.05 or less.

**Identification and annotation of male-enriched amino acid transporters**

Male-enriched amino acid transporters were annotated based on GO terms and InterProScan IDs from a list of sex-biased genes found in the microarray.  Annotations were manually validated by searching for protein domains in Pfam by hidden markov models (http://pfam.sanger.ac.uk/).  Two male-biased amino acid transporters were assigned orthology to genes in the pea aphid, *A. pisum*, by reciprocal BLAST and phylogenetic analysis.  Briefly, the *M. persicae* sequences were queried against the *A. pisum* non-redundant (nr) protein database on NCBI using a BLASTX search.  Query sequences were contigs assembled from *M. persicae* ESTs (Ramsey et al. 2007) (available at http://www.aphidbase.com/aphidbase/downloads).  The top *A. pisum* BLAST hit for each query was subsequently subjected to a local TBLASTN search against the entire set of *M. persicae* unigenes in the microarray.

For the phylogenetic analysis, male-biased *M. persicae* contigs were translated to protein and aligned to a profile of *A. pisum* homologs in SeaView (v.4) (Gouy et al. 2010) using a Clustal W (Thompson et al. 1994) plug-in.  The profile, from Price et al., (2011), consisted of all *A. pisum* members of the *slimfast* gene family, *slimfast* sequences from select other insects, and two outgroup sequences closely related to the *slimfast* gene family (Price et al. 2011).  The new alignment was then used to build a phylogeny in MrBayes (v.3.1.2) (Huelsenbeck & Ronquist 2001) using two simultaneous runs, each

with four chains. Chains were run for two million generations, after which the standard deviation of split frequencies between the runs was less than 0.01. Convergence was assessed in Tracer (v.1.5) (Rambaut et al. 2014). Briefly, Tracer plots the value of the log-likelihood and model parameters against the number of generations, visually displaying parameter convergence between the two runs and enabling us to determine that MrBayes sufficiently sampled the parameter space. A consensus tree was constructed after discarding generations making up the burn-in, as determined in Tracer.

**Analyzing sex-biased gene expression of *slimfast* homologs in *A. pisum***

Sex-dependent expression for seven out of ten *A. pisum slimfast* homologs was quantified by performing quantitative reverse transcriptase PCR (qRT-PCR) on three different *A. pisum* clonal lineages that included two lineages with intermediate lifecycles, LSR1 (Caillaud et al. 2002) and 7A (Moran et al. 2009) and one holocyclic lineage, 9-2-1 (Russell & Moran 2006) (see Wilson et al. (2003) for a description of the morphs produced by different aphid life cycle classes). The three paralogs for which I did not quantify expression shared so much sequence similarity and untranslated regions were so AT rich (~80%) that I was unable to design paralog-specific primers. For the two intermediate lineages (LSR1 and 7A), qRT-PCR was performed on the same aphid morphs used in the microarray, asexual females at long day, asexual females at short day, sexual females, and males. For the holocyclic lineage (9-2-1), which completely switches to the production of sexual morphs at short day, qRT-PCR was performed on asexual females at long day, sexual females, and males. Males and female morphs were collected as described above.

Total RNA was extracted from at least 6 whole, wingless adults of each morph using an RNeasy mini kit (Qiagen), and a DNase digest was performed to eliminate genomic DNA. cDNA was synthesized from 450 ng of total RNA using qScript cDNA SuperMix (Quanta Biosciences) and QPCR was performed using PerfeCTa® SYBR® Green FastMix® (Quanta Biosciences). Gene expression was compared between morphs using $2^{-DDC_T}$ methodology (Livak & Schmittgen 2001), with expression between morphs normalized to the expression of housekeeping gene glyceraldehyde-3-dehydrogenase (*GAPDH*, ACYPI009769 (Price et al. 2010)). Experiments were each performed in triplicate and included no template controls. Prior to running experiments, I verified that cDNA pools lacked genomic DNA by running no reverse transcription controls alongside no template controls and positive controls. QPCR reactions were performed on a Mastercycler® ep realplex[4] real-time PCR system (Eppendorf) using a program that began with 95º C for 5 minutes, followed by 40 cycles at 95º C for 15 sec, 52º C, and 60º C for 20 sec. Amplification profiles and melt-curves were analyzed in Mastercycler® ep realplex software (v.1.5) (Eppendorf). For information on primer sequences and efficiencies, see Price et al. (2011). Gene expression data were collectively normalized, allowing comparison of expression profiles across paralogs and aphid lines. Normalization was performed by converting $\Delta C_T$ values to z-scores as follows:

$$z = -10 \times \left( \frac{\Delta C_T - \overline{\Delta C_T}}{\sigma_{\Delta C_T}} \right)$$

The normalized expression for each gene from the three lines was compiled into a heat map where yellow indicates z > 0 and blue indicates z < 0.

**Molecular evolution analyses**

Rates of molecular evolution along phylogenetic tree branches were estimated for

aphid genes in the aphid specific *slimfast* expansion.  The analysis was conducted using

the same phylogeneny for which the methods were described above.  The amino acid

alignment used to reconstruct the tree was converted back to a codon alignment in

SeaView, and the codon alignment was used to estimate the ratio of non-synonymous

(dN) to synonymous (dS) substitution rates along branches, or dN/dS (also denoted $\omega$).

The $\omega$ ratio reflects the strength and type of selection in protein coding sequences.

Excess *dS* ($\omega < 1$) indicates that selection favors substitutions that conserve the amino

acid sequence (purifying selection), while equal dS and dN ($\omega = 1$) illustrates relaxed

selective constraints (neutral selection).  Excess dN ($\omega > 1$) indicates that selection favors

substitutions that change the amino acid sequence (positive selection).

The $\omega$ ratio was estimated for different branches in the aphid *slimfast* expansion by

maximum likelihood in the PAML package (Yang 2007).  Analyses implemented the

branch models (Bielawski 2003, Yang 1998) with one or more $\omega$ categories for branches

and one w across sites (model = 0 or 2, NSsites = 0).  The parameters $\omega$ and $\kappa$

(transition/transversion rate) were both estimated, starting from initial values of 0.2 and 2

respectively, and codon frequency was measured using a 3x4 codon table.  I tested for

elevated $\omega$ along specific branches using various models.  Briefly, I tested the null model

assuming equal $\omega$ across branches and several nested models allowing for different $\omega$

ratios for particular groups of branches and compared the log likelihoods under these

models with a likelihood ratio test (Yang 1998).  If the likelihood ratio test showed a

statistically significant improvement in log likelihood from one model to the next, then

the model with the higher log likelihood was supported.  Models I tested were inspired by the expression profiles of aphid paralogs in the expansion and I outline them in the results section.

## Results

### Microarray and automated annotation results

I mined a microarray comparing expression between male, sexual female, and asexual female *Myzus persicae* aphids for genes with differential expression between males and at least one female morph (i.e. genes that were not sex-neutral).  Comparisons were made with two types of asexual females: asexual females from perpetually parthenogenetic cultures requiring higher temperatures and long day lengths (long day asexual females) and asexual females collected from sexual cultures, induced by lower temperatures and shortened day lengths (short day asexual females).  Comparing males and sexual females to both long and short day asexual females enabled me to control for day length effects in gene expression (since sexual morphs are only present under short day conditions).

Genes with differential expression between males and at least one female reproductive morph are referred to as "male-enriched" or "female-enriched".  These sex-enriched genes were identified by fold change (4-fold or greater upregulation) and significance (FDR < 0.05).  Of the 10,478 *M. persicae* unigenes represented on the microarray, 768 (7%) were male-enriched and 725 (7%) were female-enriched (Supplemental Table 2.1; a list of all sex-enriched genes with descriptions, annotations and fold changes can be found in Additional files 2 and 3 from Duncan et al. (2011)).  Of

the male-enriched genes, 126 (16%) were upregulated in males relative to all three female morphs and of the female-enriched genes, 82 (11%) were enriched in all three female morphs relative to males. I refer to genes that are differentially expressed between males and all three female reproductive morphs as "male-biased" or "female-biased". One gene was enriched in both sexes depending upon the female morph: contig 2728 (see Additional file 4 (Duncan et al. 2011) for corresponding EST accession numbers) was upregulated in males relative to sexual females and upregulated in asexual females relative to males (Supplemental Table 2.1, and Additional files 2, and 3 (Duncan et al. 2011)).

Mapping and annotation results obtained from Blast2GO are summarized in Supplemental Table 2.1. The homology search returned significant BLAST hits for 501 (65%) male-enriched and 554 (76%) female-enriched genes. Gene Ontology (GO) terms were found for 371 (48%) male-enriched and 414 (57%) female-enriched genes. Round one annotation, which excluded electronic annotations, annotated 202 (26%) male-enriched and 291 (40%) female-enriched genes. Round two annotation, which included electronic annotations, annotated an additional 63 (8%) male-enriched and 51 (7%) female-enriched genes. InterProScan (Zdobnov & Apweiler 2001) added annotations to genes annotated in the previous steps and additionally annotated 42 (5%) male-enriched and 31 (4%) female-enriched genes that were previously unannotated. Annotations were assigned to a total of 307 (40%) male-enriched and 373 (51%) female-enriched genes (Supplemental Table 2.1).

Overrepresented GO terms are presented in Additional file 5 (Duncan et al. 2011). In total, 75 GO terms were significantly overrepresented in sex-enriched genes (FDR <

0.05).  Of these GO terms, 16 (21%) were overrepresented in males and 59 (79%) in females.  Female-enriched genes were represented by several GO classes such as nucleotide binding, nitrogen metabolism, cell cycle, cytoskeleton organization and organelles.  In contrast, male-enriched genes were dominated by GO terms related to transporter and channel activity (Additional file 5 (Duncan et al. 2011)).

**Annotation of sex-enriched nutrient amino acid transporters in *M. persicae***

From the sex-enriched genes identified in the microarray, I identified five nutrient amino acid transporters upregulated 4.2-fold to 20.1-fold in males (FDR < 0.05; Table 2.1 and Additional file 3 (Duncan et al. 2011)).  Four of the five male-enriched transporters were annotated in Blast2GO based on GO terms associated with significant BLAST hits (E ≤ 0.001) and/or InterProScan IDs (Table 2.1, Additional file 3 (Duncan et al. 2011)).  The fifth transporter had no BLAST hits associated with GO terms, but its one InterProScan ID was annotated as belonging to transmembrane amino acid transporters (Table 2.1).  After translating unigene nucleotide sequences to amino acid sequences based on the largest open reading frame, four of the five transporters significantly matched Pfam domains *AA_permease* (PF000324) or *Aa_trans* (PF01490) (Table 2.1).

Two transporters (contigs 1492 and 8321; See Additional file 4 (Duncan et al. 2011) for the GenBank accession numbers) were upregulated 5.7 to 20.1-fold in males relative to both sexual and asexual females (Table 2.1, Additional file 3 (Duncan et al. 2011)).  I further annotated these two transporters based on their strongly male-biased

expression.  Using BLASTX and reciprocal TBLASTN searches, *M. persicae* contig

1492 (Table 2.1, Additional file 4 (Duncan et al. 2011)) and the *A. pisum* amino acid

permease ACYPI005156 were reciprocal BLAST best hits.  These two genes also paired

with unambiguous support in a gene phylogeny for insect orthologs to the *Drosophilia*

gene *slimfast*, a cationic amino acid permease (Colombani et al. 2003) of which *A. pisum*

has a ten paralog aphid-specific expansion (Price et al. 2011) (Figure 2.1).  A BLASTX

search for the other *M. persicae* contig, 8321 (Table 2.1 and Additional file 4 (Duncan et

al. 2011)), against the *A. pisum* nr protein database returned the *A. pisum* gene

ACYPI003240 as the best hit, but the reciprocal TBLASTN search failed to return contig

8321 as the best hit for ACYPI003240.  Phylogenetic analysis paired contig 8321 and

ACYPI003240 with unambiguous support in the *slimfast* gene phylogeny (Figure 2.1).

Based on the homology of the aphid expansion to *slimfast*, I named the 10 *A. pisum*

paralogs *ACYPIslif1-10* (Table 2.2).

Intrigued by having identified male-enriched nutrient amino acid transporters, I

sought to discover if more sex-enriched amino acid transporters were represented in the

microarray that my annotation pipeline had failed to annotate.  TBLASTX and BLASTN

searches queried all 40 *A. pisum* nutrient amino acid transporters (as identified by Price et

al. (2011)) against the *M. persicae* unigenes in the microarray.  Apart from the 5 male-

enriched transporters annotated above, I uncovered 9 additional contigs with significant

($E \leq 0.001$) similarity to *A. pisum* nutrient amino acid transporters.  Of these 9 contigs, 5

were significantly (FDR<0.05) male-enriched (two of which were male-biased) and 3

were female-enriched.  Fold change for most of these contigs was less than four,

explaining why I did not identify them in my initial analysis.  In total, the *M. persicae*

microarray targeted 14 contigs that significantly matched nutrient amino acid transporter sequences, of which 10 were male-enriched or male-biased and 3 were female-enriched.

**Sex-biased expression in the aphid *slimfast* expansion**

In my survey for more amino acid transporters in the microarray, I found no additional *slimfast* paralogs.  I thus shifted the focus of this study to *A. pisum* so as to leverage its sequenced genome (International Aphid Genomics Consortium 2010) to analyze the evolution of male-biased aphid *slimfast*.  Expression was quantified by qRT-PCR between males and the three female morphs across three biological replicates. Expression was quantified for all *slimfast* paralogs but three, for which sequence similarity and AT richness in the untranslated regions prevented the design of paralog-specific primers (Price et al. 2011).

Expression results are presented as a heat map aligned against the *slimfast* gene phylogeny in Figure 2.2.  Paralogs *ACYPIslif5, 6,* and *10* were male-biased, with *ACYPIslif5* and *6* showing the highest male-biased expression (Average fold change ± SD = 5.0 ±2.2) and *ACYPIslif10* showing less upregulation in male expression (Average fold change ± SD = 2.0 ± 0.5).  Expression patterns for male-biased genes were consistent across all biological replicates (Figure 2.2).  Additional patterns that were consistent across the three biological replicates were (1) enriched expression for *ACYPIslif8* in asexual females at long day and (2) consistent low expression across all morphs for *ACYPIslif9*.

**Molecular evolution of male-biased *slimfast* paralogs**

The expression profiles of each *slimfast* paralog inspired the molecular evolutionary models I used to test if branches leading to male-biased paralogs experienced accelerated rates of evolution.  Four codon-based branch models (Bielawski 2003, Yang 1998) were constructed to test for differences in the ratio of non-synonymous to synonymous substitution rates along specific branches (dN/dS = $\omega$).  The null model (one ratio) assumed equal $\omega$ for all branches.  The first alternative model (two ratios) assumed different $\omega$ ratios for branches within and outside the aphid-specific *slimfast* expansion. The second alternative model (three ratios) assumed an additional, collective $\omega$ ratio for the branches leading to the three male-biased paralogs, and the third alternative model (five ratios) assumed a distinct $\omega$ for each branch leading to a male-biased paralog or clade (Figure 2.3).

Results for the models and likelihood ratio tests are given in Table 2.3.  The two ratios model had a significantly higher likelihood than the one ratio model and predicted a higher $\omega$ for branches within the aphid *slimfast* expansion, consistent with previous work (Price et al. 2011).  Therefore, the remaining two models were compared to the two ratios model.  The three ratios model was not significantly better than the two ratios model, but a significant improvement in likelihood was observed with the five ratios model.  Using the five ratios model, branches leading to male-biased clades (*ACYPIslif6*, contig 1492 and *ACYPIslif5*, contig 8321) had lower $\omega$ values than the average branch within the expansion (0.17-0.18 vs. 0.22).  In contrast, the branch leading to *ACYPIslif10* had a greater $\omega$ value than other expansion branches (0.61 vs. 0.22).

**Discussion**

The nutrient amino acid transporter *slimfast* plays an important role in nutritionally dependent processes, such as coordinated growth in both males and females and oogenesis in females. By acting as a nutrient sensor, *slimfast* stimulates processes like nutrient import, metabolism, and translation in response to nutrient availability via the Target of Rapamycin (TOR) signaling pathway (Hundal & Taylor 2009, Wullschleger et al. 2006). TOR signaling activation relies especially on *slimfast* expression in the insect fat body, evidenced by the fact that in *Drosophila melanogaster,* downregulating *slimfast* in fat body cells disrupts overall growth (Colombani et al. 2003). In mosquitoes, *slimfast* expression in the fat body also activates vitellogenesis at (Attardo et al. 2006), a key process in oogenesis. While oogenesis is reproductive in nature, it is fundamentally nutrient-driven because female reproductive processes, unlike male reproductive processes, are energetically costly and depend directly on nutrient availability (Wheeler 1996, Wigglesworth 1960). Although a male-specific role for *slimfast* has never been documented in any insect, I present evidence that some paralogs in the aphid-specific *slimfast* expansion have evolved a male functional role.

**Conserved male-biased expression of aphid-specific *slimfast* paralogs**

I identified five nutrient amino acid transporters with at least 4-fold enriched expression in males relative to at least one female aphid morph. Although the software Blast2GO failed to annotate all five of the contigs I identified from the microarray (Table 2.1 and Additional file 3 (Duncan et al. 2011)), InterProScan IDs and Pfam searches

provided strong support that all five contigs are amino acid transporters and that four of the five contain either the *Aa_trans* or *AA_permease* transmembrane domains (Table 2.1). These domains characterize two amino acid transporter families, the amino acid/polyamine/organocation (APC) family (TC #2.A.3), and the amino acid/auxin permease (AAAP) family (TC #2.A.18) (Castagna et al. 1997, Saier 2000, Saier et al. 2009, 2006). These amino acid transporter families contain all but a handful (Boudko et al. 2005) of the known nutrient amino acid transporters. Though one contig (8321) did not significantly match any Pfam domain (Table 2.1), its strongly supported phylogenetic position within the *A. pisum* expansion of *slimfast* (Figure 2.1), a member of the APC family, indicates that contig 8321 is also a member of the APC family. Of the five male-enriched nutrient amino acid transporters (Table 2.1 and Additional file 3 (Duncan et al. 2011)), contigs 1492 and 8321 (see Additional file 4 (Duncan et al. 2011) for GenBank accessions) are male-biased, being upregulated in males relative to all three female reproductive morphs.

Shifting focus to *A. pisum* supported a conserved male role for *slimfast.* The phylogenetic placement of contigs 1492 and 8321 indicates that they are orthologous to *A. pisum slimfast* paralogs *ACYPIslif6* and 5 respectively (Figure 2.1). Male-biased expression in these paralogs (Figure 2.2) suggests that their male-biased expression is conserved minimally in the Macrosiphini tribe, of which *A. pisum* and *M. persicae* are members. Further, qRT-PCR expression analysis revealed one more male-biased paralog, *ACYPIslif10* (Figure 2.2). Male expression in *ACYPIslif10* was low compared to the expression observed for *ACYPIslif5* and *6* (Figure 2.2), which may indicate that it is expressed only in a particular tissue and its true expression level is confounded by

quantifying expression in whole insects. Because *slimfast* is not known to have a male-specific role, the presence of male-biased *slimfast* paralogs in aphids is puzzling. If aphid *slimfast* paralogs retain a function in activating TOR signaling and nutrient-dependent processes, then the presence of male-biased paralogs could indicate that male and female aphids must overcome different nutritional challenges.

**The evolutionary origin of male-biased *slimfast* expression in aphids**

Given the phylogenetic distribution of male-biased expression (Figure 2.2), I am unable to conclusively reconstruct when male-biased expression evolved within the *slimfast* expansion. The current data set can explain the distribution of male-biased expression by two alternatives. First, male-biased expression could be derived, which could have happened in one of two ways. Either male-biased expression evolved independently for each paralog, or it evolved twice independently and was lost in the lineage leading to *ACYPIslif4*. A second, equally parsimonious, explanation is that male-biased expression is the ancestral state and was lost three times. Derived and ancestral male-biased expression in the expansion are thus equally plausible given the data. Distinguishing between these two explanations would be facilitated by data for the three paralogs lacking expression information (Figure 2.2).

Despite the lack of resolution in my results for when male-biased expression evolved in the *slimfast* expansion, support for derived male-biased expression in aphids can be gleaned from the literature. In a study examining sex-biased genes in seven *Drosophila* species, *slimfast* was significantly female-biased in five species and not

significantly enriched in either sex for the other two species (see Supplementary Tables from (Zhang et al. 2007)). Equal expression in females and males may reflect the ancestral state of the aphid *slimfast* expansion since *Drosophila* genomes have only one *slimfast* copy (http://phylomedb.org), and the essential role of *slimfast* in growth (Colombani et al. 2003) strongly suggests that *Drosophila slimfast* is under strong purifying selection. My molecular evolution analyses support strong purifying selection for *Drosophila* and other insect *slimfast*, evidenced by the significantly lower ω found for branches outside the aphid *slimfast* expansion (Table 2.3). Further, the known female and sex-neutral roles (Attardo et al. 2006, Colombani et al. 2003) but lack of documented male role suggests that male-biased expression (and corresponding male function) is derived among aphids.

**Evolution of a male-biased functional role in aphid *slimfast***

Male-biased genes reflect traits that increase male fitness, such as male-male competition, sexually selected characters and secondary sexual characters (Ellegren & Parsch 2007). The male roles of three aphid *slimfast* paralogs can be hypothesized from current and previous results. Although I did not examine tissue-level expression in this study, previous work by Price et al. (2011) quantified relative expression levels of *slimfast* paralogs in asexual female head, gut and bacteriocytes. While tissue-level expression patterns may not be identical in both sexes, these data provide a framework within which to formulate some of the possible testable hypotheses about the function of male-biased paralogs. In this context, all three male-biased paralogs show highly enriched expression in asexual female gut (Price et al. 2011), consistent with *slimfast*

expression in *Drosophila* (Colombani et al. 2003) and *Tribolium* (Supporting Table 3 from Morris et al. (2009)).

The digestive tract interfaces between an animal and its diet, playing a critical role in uptaking dietary nutrients. Dietary nutrient availability is central to the aphid/*Buchnera* nutritional symbiosis because deficient dietary amino acids must be synthesized by *Buchnera*. While *Buchnera*-containing bacteriocytes are abundant in females, males contain relatively few (Douglas 1989) and in extreme cases, males completely lack bacteriocytes (Braendle et al. 2003, Ishikawa 1992). This sex-based difference in a major nutrient provisioning cell type suggests that males and females differ in their ability to synthesize amino acids deficient in their diet. Differential amino acid biosynthesis could be compensated for by upregulating amino acid transporters in the male gut, enabling greater uptake of certain amino acids from phloem sap.

In contrast, other previously expression patterns suggest an alternative possibility for male function. In addition to being enriched in gut, two of the male-biased paralogs (*ACYPIslif6* and *ACYPIslif10*; Figure 2.2) were also enriched in the asexual female head (Price et al. 2011). These head-expressed paralogs could thus function in any of the head tissues included in the analysis, such as brain, eyes, mouthparts, antennae, or salivary glands, suggesting a role in sensory functions, such as locating a mate.

The possibility remains that the male-biased paralogs have evolved a different expression pattern (and function) from asexual females. One possible role for these male-biased paralogs that I cannot predict from female expression profiles is a role in male reproduction, such as spermatogenesis. Although *slimfast* has never been implicated in spermatogenesis, related mammalian transporters (*SLC7A1* and *SLC7A2*)

deliver L-Arginine to rat seminiferous tubule cells (Cerec et al. 2007), where

spermatogeneis begins. Thus, divergence of some *slimfast* paralogs to fill a male

reproductive role is plausible. In light of the molecular evolution results, paralog

*ACYPIslif10* is the best candidate for having a role in spermatogenesis since its terminal

branch has experienced an extremely accelerated rate of non-synonymous substitutions

(Table 2.3). Accelerated evolutionary rates are commonly associated with male-biased

genes, especially genes involved in sperm competition (Ellegren & Parsch 2007). While

the accelerated ω I observed is consistent with both positive and neutral/relaxed selection,

both types of selection can lead to functional divergence (Lynch 2007, Zhang 2003).

Additional studies can tease apart the various hypotheses I have presented for the role

played by male-biased *slimfast* paralogs in aphid biology.

**Insights on the maintenance of duplicate amino acid transporters in aphids**

By pointing towards a derived evolutionary origin of male-biased function in

*slimfast*, my results provide insights into a fundamental question of how the aphid

genome retains the *slimfast* expansion and other nutrient amino acid transporter

duplications (see also Price et al. (2011)). Given that most gene duplications fail to reach

fixation (Lynch & Conery 2000), it is intriguing that the aphid genome possesses more

nutrient amino acid transporters than other sequenced insects (Price et al. 2011). The

presence of these duplicate amino acid transporters indicates that there must be a

selective advantage to their maintenance. As mentioned above, some duplicate nutrient

amino acid transporters (including some *slimfast* paralogs) are highly enriched in

bacteriocytes, leading to the hypothesis that these transporters mediate nutrient exchange

across the *A. pisum/Buchnera* symbiotic interface (Price et al. 2011). A role in mediating endosymbiotic interactions strongly suggests that duplicate nutrient amino acid transporters and the *slimfast* expansion are maintained in the genome because they diverged to fill a novel role in symbiosis. The current study supports a different, but not mutually exclusive, hypothesis that the *slimfast* expansion (and possibly other nutrient amino acid transporter duplicates) is maintained because some paralogs diverged to fill novel, sex-specific roles. Future studies will be able to test the relative significance of symbiosis and sex in maintaining amino acid transporter gene duplications by examining the genetic architecture and expression of nutrient amino acid transporters in other phloem-feeding insects with different, less complex life cycles.

**Conclusions**

This study is the first to report evidence for a male function of the nutrient amino acid transporter *slimfast*. Using a microarray and qRT-PCR, I identified three male-biased *slimfast* paralogs in an aphid-specific gene family expansion, and two of those paralogs had conserved expression profiles across two aphid species. By integrating different sources of knowledge on the function and expression of *slimfast* and related nutrient amino acid transporters, I propose competing hypotheses for the function of male-biased *slimfast* in male-dependent nutritional, head, and reproductive functions. This study highlights the necessity of examining different types of expression data for functionally annotating genes in novel genomes. While previous work showed evidence that the aphid *slimfast* expansion is maintained in the *A. pisum* genome because some paralogs evolved a novel role in mediating symbiotic interactions (Price et al. 2011), this

study extends the hypothesis for the maintenance of the *slimfast* expansion to include paralog divergence for unexpected sex-dependent functions.  The integration of different expression profiles to assist in functionally characterizing genes is particularly important for novel genes in all organisms, especially gene duplications, which have the potential to evolve divergent functions from homologous single-copy genes in other organisms.
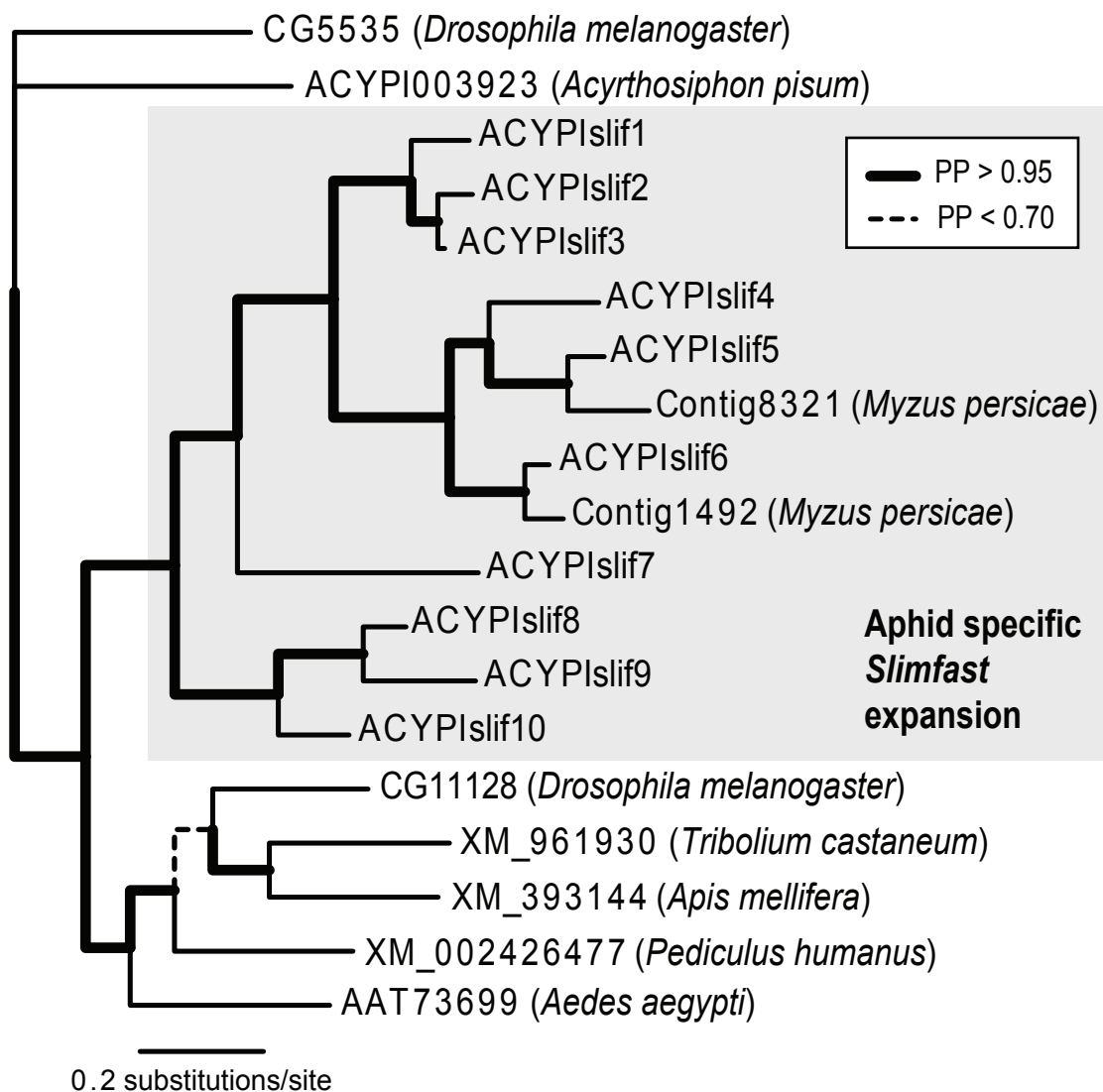
Figure 2.1. Bayesian phylogeny of the slimfast gene family. The aphid specific *slimfast* expansion is shown in the gray box. Posterior probability support for relationships is represented by branch width and style, as indicated in the key. Outgroup sequences were shown by Price et al. (2011) to belong to a closely related clade of the *slimfast* gene family. Gene IDs beginning with "ACYPI" belong to *A. pisum* and gene IDs beginning with "contig" belong to *M. persicae.*
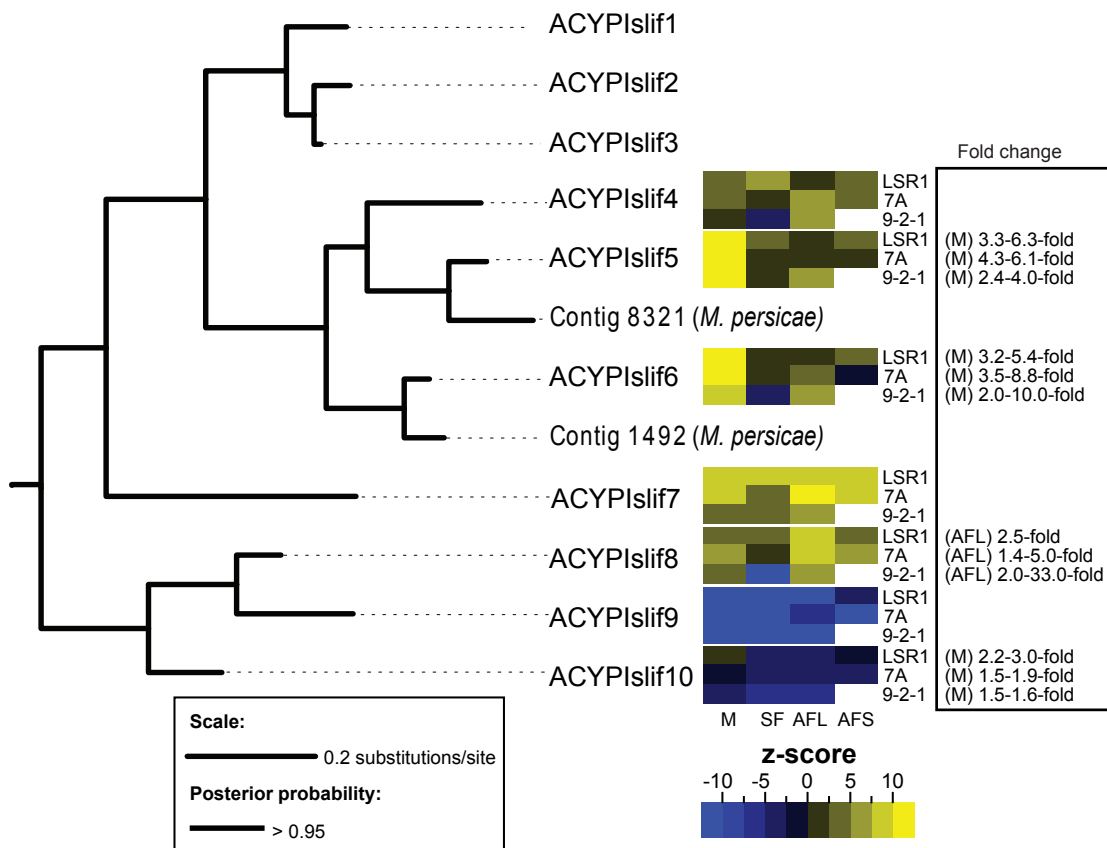
Figure 2.2. Quantitative gene expression analysis of *A. pisum slimfast* paralogs. Gene expression profiles are shown for three *A. pisum* lineages (LSR1, 7A, 9-2-1). Transcript levels for each *slimfast* paralog were quantified by qRT-PCR on cDNA for whole adult asexual females at long day (AFL) and short day (AFS), sexual females (SF), and males (M). The relative abundance of each paralog was normalized to the housekeeping gene *GAPDH* (ACYPI009769) and expression levels were standardized by converting them to z-scores and compiled into a heat map (for details, see methods). Expression was not quantified for *ACYPIslif1-3* because sequence similarity and AT richness in the untranslated regions precluded my ability to design paralog-specific qRT-PCR primers. Aphid line 9-2-1 does not produce asexual females at short day so I did not measure AFS expression for 9-2-1 (see methods for details). Yellow: z > 0; Blue: z < 0.
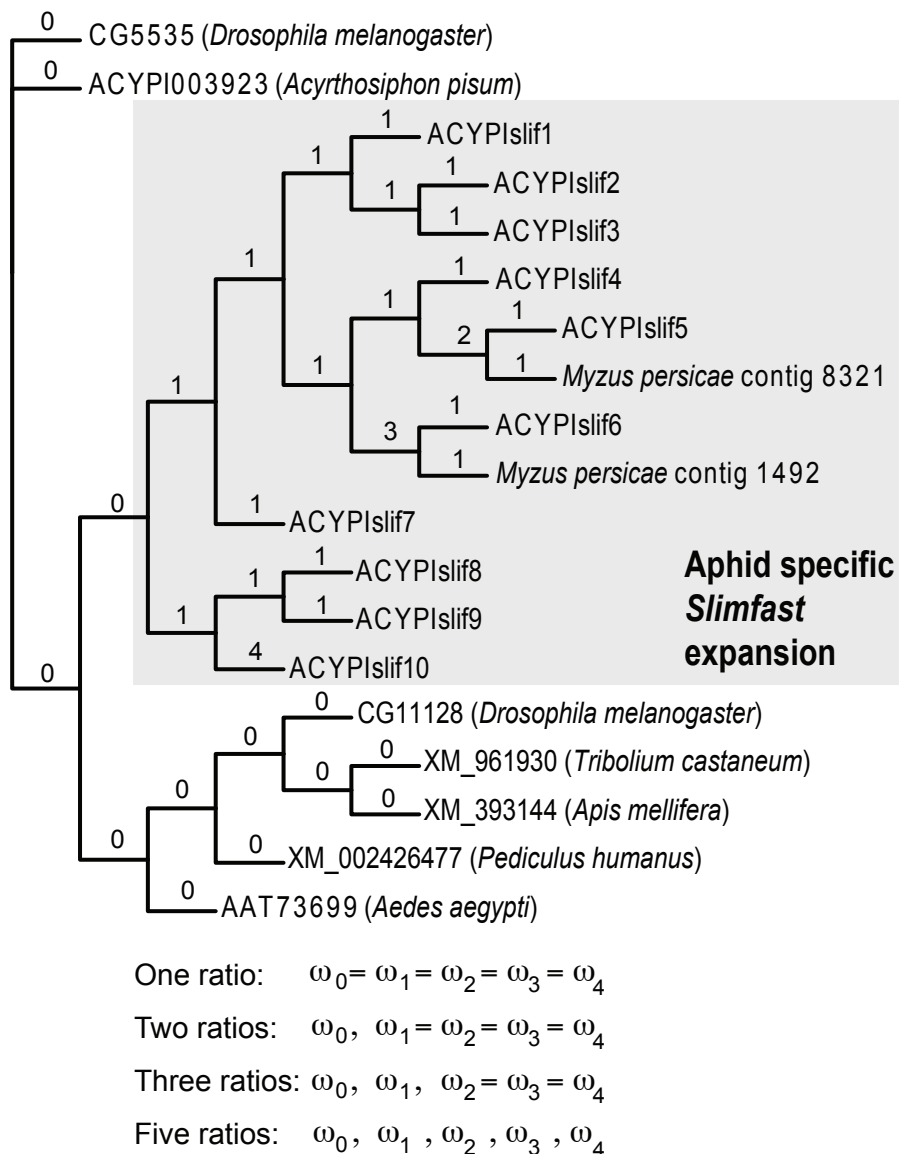
Figure 2.3. Molecular evolution models. Strategy used to test for accelerated rates of evolution along branches leading to male-baised aphid *slimfast* paralogs. Tree topology and branch lengths are identical to the phylogeny in Figure 2.1. Each branch was assigned a ω category as indicated. The models tested are outlined below the phylogeny. The null model (one ratio) assumed equal ω across all branches. The first alternative model (two ratios) assumed one ω ratio for branches outside the aphid-specific *slimfast* expansion and a different ω ratio for branches within the expansion. The second alternative model (three ratios) assumed an additional, collective ω ratio for the branches leading to the three male-biased paralogs, and third alternative model (five ratios) assumed a distinct ω for each branch leading to a male-biased paralog or clade.

Table 2.1. *M. persicae* male-enriched nutrient amino acid transporters identified in microarray and annotations

| Contig ID | GO terms | InterProScan IDs | Pfam domain | Fold change (Compared to)[1] |
|---|---|---|---|---|
| 1492 | GO:0006629: lipid metabolic process GO:0005215: transporter activity GO:0007165: signal transduction GO:0005886: plasma membrane GO:0006810: transport GO:0040007: growth GO:0006811: ion transport | IPR002293: Amino acid/polyamine transporter IPR004841: Amino acid permease domain IPR015606: Cationic amino acid transporter | AA_permease (PF00324) | 9.5-15.0 (AFL) 6.9-10.5 (AFS) 13.4-20.1 (SF) |
| 3389 | GO:0006810: transport | IPR013057: Amino acid transporter, transmembrane | Aa_trans (PF01490) | 4.2 (AFS) |
| 4712 | Not annotated in Blast2GO | IPR013057: Amino acid transporter, transmembrane | Aa_trans (PF01490) | 5.5 (SF) |
| 4891 | GO:0005623: Cell | IPR013057: Amino acid transporter, transmembrane | Aa_trans (PF01490) | 9.3 (SF) |
| 8321 | GO:0009605: response to external stimulus GO:0006950: response to stress GO:0006629: lipid metabolic process GO:0006810: transport GO:0050789: regulation of biol. process GO:0005634: nucleus GO:0005623: Cell GO:0005215: transporter activity GO:0040007: growth GO:0006139: nucleobase, nucleoside, and nucleic acid metabolic nucleotide proc. GO:0009058: biosynthetic process GO:0005886: plasma membrane GO:0006811: ion transport | IPR002293: Amino acid/polyamine transporter IPR015606: Cationic amino acid transporter | No significant Pfam hits | 5.7 (AFL) 7.2 (AFS) 7.6 (SF) |

[1] AFL: Asexual female at long day; AFS: Asexual female at short day; SF: Sexual female

Table 2.2. *A. pisum slimfast* gene identifications

| *A. pisum slimfast* name | *A. pisum* gene ID[1] |
|---|---|
| *ACYPIslif1* | ACYPI000537 |
| *ACYPIslif2* | ACYPI000584 |
| *ACYPIslif3* | ACYPI005472 |
| *ACYPIslif4* | ACYPI003523 |
| *ACYPIslif5* | ACYPI003240 |
| *ACYPIslif6* | ACYPI005156 |
| *ACYPIslif7* | ACYPI005118 |
| *ACYPIslif8* | ACYPI008904 |
| *ACYPIslif9* | ACYPI008323 |
| *ACYPIslif10* | ACYPI002633 |

[1] Gene IDs are recognized in GenBank

Table 2.3. Molecular evolution results

| Model | $\omega$ value | -lnL | Null[1] | P |
|---|---|---|---|---|
| One ratio | $\omega 0 = 0.16415$ | 26427.39 | | --- |
| | $\omega 1 = \omega 0$ | | | |
| | $\omega 2 = \omega 0$ | | | |
| | $\omega 3 = \omega 0$ | | | |
| | $\omega 4 = \omega 0$ | | | |
| Two ratios | $\omega 0 = 0.00452$ | 26300.94 | One ratio | ** |
| | $\omega 1 = 0.22750$ | | | |
| | $\omega 2 = \omega 1$ | | | |
| | $\omega 3 = \omega 1$ | | | |
| | $\omega 4 = \omega 1$ | | | |
| Three ratios | $\omega 0 = 0.00452$ | 26300.92 | Two ratios | NS |
| | $\omega 1 = 0.22624$ | | | |
| | $\omega 2 = 0.23358$ | | | |
| | $\omega 3 = \omega 2$ | | | |
| | $\omega 4 = \omega 2$ | | | |
| Five ratios | $\omega 0 = 0.00451$ | 26295.18 | Two ratios | ** |
| | $\omega 1 = 0.22356$ | | | |
| | $\omega 2 = 0.18305$ | | | |
| | $\omega 3 = 0.17040$ | | | |
| | $\omega 4 = 0.60752$ | | | |

[1] Null model against which the given model was tested using a likelihood ratio test.
NS = Not Significant
** indicates that $P < 0.05$

Supplemental Table 2.1. Summary of BLAST2GO results

| | Number of genes | | |
|---|---|---|---|
| | Male | Female | Total |
| Total found in array | 768 | 725 | 1492* |
| Significant BLAST hits | 501 | 554 | 1054* |
| Mapping | | | |
|     GO terms found | 371 | 414 | 784* |
| New annotations | | | |
|     1st Annotation | 202 | 291 | 492* |
|     2nd Annotation | 63 | 51 | 114 |
|     InterProScan | 42 | 31 | 73 |
|     Total annotated | 307 | 373 | 679* |

* One gene (Contig 2728) was enriched in males relative to asexual females and also enriched in sexual females relative to males

**Chapter 3**

Dynamic recruitment of amino acid transporters to the insect-symbiont interface[2]

**Summary**

Symbiosis is well known to influence bacterial symbiont genome evolution, and has recently been shown to shape eukaryotic host genomes. Intriguing patterns of host genome evolution, including remarkable numbers of gene duplications, have been observed in the pea aphid, a sap-feeding insect that relies on a bacterial endosymbiont for amino acid provisioning. Gene duplication has been proposed to be important for the evolution of symbiosis based on aphid-specific gene duplication in amino acid transporters, with some paralogs highly expressed in the cells housing symbionts (bacteriocytes). Here, I use a comparative approach to test the role of gene duplication in enabling recruitment of amino acid transporters to bacteriocytes. Using genomic and transcriptomic data, I annotate amino acid transporter from sap-feeding and non sap-feeding insects and find that, like aphids, amino acid transporter gene families have undergone independent large-scale gene duplications in three out of four additional sap-feeding insects. RNA-seq differential expression data indicate that, like aphids, the sap-feeding citrus mealybug possesses several lineage-specific bacteriocyte-enriched paralogs. Further, differential expression data combined with quantitative PCR support independent evolution of bacteriocyte enrichment in sap-feeding insect amino acid transporters. Although these data indicate that gene duplication is not necessary to initiate host/symbiont amino acid exchange, they support a role for gene duplication in enabling amino acid transporters to mediate novel host/symbiont interactions broadly in

---

[2] This chapter was published as: Duncan RP, Husnik F, Van Leuven JT, Gilbert DG, Dávalos LM, McCutcheon JP, Wilson ACC. 2014. Dynamic recruitment of amino acid transporters to the insect-symbiont interface. *Molecular Ecology*. 23(6): 1608-23.

the sap-feeding suborder Sternorrhyncha. In combination with recent studies on other symbiotic systems, gene duplication is emerging as a general pattern in host genome evolution.

**Background**

Interspecific interactions fundamentally impact the evolutionary trajectory of species and have long been known to influence characteristics such as morphology (Schemske & Bradshaw 1999), color patterns (Sandoval 1994), community structure (Kennedy 2010), and even behavior (Eberhard 2000). Furthermore, interactions between species shape an organism's genome in ways that are only just beginning to be appreciated. Not only do species interactions influence the genes and pathways directly involved in those interactions, but overall genome content, organization, expression, size and even base composition are influenced by interspecific interactions. The most intriguing examples of how genome evolution is shaped by interspecific interactions are found in obligate, endosymbiotic mutualists. For example, bacterial nutritional endosymbionts have undergone drastic genome reduction and gene loss in response to evolving an obligate endosymbiotic lifestyle (Bennett & Moran 2013, McCutcheon & Dohlen 2011, McCutcheon & Moran 2007, 2012; McCutcheon et al. 2009, Nakabachi et al. 2006, Nikoh et al. 2011, Sabree et al. 2009, 2012a; Shigenobu et al. 2000, Sloan & Moran 2012). Historically, symbiont genomes have received more attention than the genomes of their hosts, but as deep sequencing becomes cheaper and assembly technology advances, host genomes are providing insight into how symbiosis shapes genomes in eukaryotic hosts (Husnik et al. 2013, International Aphid Genomics

Consortium 2010, Kirkness et al. 2010, Nygaard et al. 2011, Young et al. 2011).  Four

interesting and novel (given current sampling) features of host genomes include: (1)

metabolic complementarity with symbionts in essential nutrient biosynthesis (Hansen &

Moran 2011, Husnik et al. 2013, McCutcheon & Dohlen 2011, Nygaard et al. 2011,

Shigenobu et al. 2000, Wilson et al. 2010), (2) loss or modulation of immune pathways

(Gerardo et al. 2010, Kim et al. 2011b, Ratzka et al. 2013), (3) maintenance and

expression of functional genes acquired horizontally from bacteria other than the obligate

symbiont (Husnik et al. 2013, Nikoh & Nakabachi 2009, Nikoh et al. 2010), and (4)

duplication of genes with functions that may facilitate symbiosis (Ganot et al. 2011, Price

et al. 2011, Shigenobu & Stern 2013, Young et al. 2011).  Although these features

suggest a role for symbiosis in shaping host genomes, some genomic attributes of

eukaryotic hosts come from isolated examples and a role for symbiosis in their

evolutionary origin remains untested.  One way to test the role of symbiosis in shaping

host genome evolution is by evaluating specific genomic traits within an evolutionary

framework.

An evolutionary framework is especially powerful in evaluating the extensive

gene duplication and differential expression of amino acid transporters observed in the

genome of the pea aphid, *Acyrthosiphon pisum*.  This evolutionary pattern may be

influenced by the relationship between *A. pisum* and its obligate bacterial endosymbiont,

*Buchnera aphidicola*.  *A. pisum*, a member of the insect order Hemiptera, feeds on plant

phloem sap, a diet deficient in key nutrients such as essential amino acids (Douglas 1993,

2006; Sandstrom & Pettersson 1994, Wilkinson & Douglas 2003).  Essential amino

acids – i.e. amino acids that animals are unable to synthesize *de novo* – are provided to

aphids by *Buchnera* in exchange for non-essential amino acids (Shigenobu et al. 2000).

Supply of non-essential amino acids to *Buchnera* and distribution of essential amino

acids from *Buchnera* to host tissues is mediated by amino acid transport across three key

membrane barriers that I collectively refer to as the symbiotic interface: (1) The plasma

membrane of the specialized aphid cells that house *Buchnera* (bacteriocytes), (2) the

host-derived symbiosomal membrane surrounding individual *Buchnera* cells, and (3) the

bacterial inner and outer membranes of individual *Buchnera* (Shigenobu & Wilson 2011).

Analyses of transcripts (Hansen & Moran 2011, Macdonald et al. 2012, Price et al. 2011)

and proteins (Poliakov et al. 2011) enriched in aphid bacteriocytes suggest that amino

acid flux at the aphid/*Buchnera* symbiotic interface is mediated by several aphid amino

acid tranpsorter from two gene families: the amino acid polyamine organocation (APC)

family (Transporter Classification (TC) #2.A.3), and the amino acid/auxin permease

(AAAP) family (TC #2.A.18) (Castagna et al. 1997, Saier 2000, Saier et al. 2009, 2006).

These two amino acid transporter families play important nutritional roles in insects

(Attardo et al. 2006, Colombani et al. 2003, Dubrovsky et al. 2002, Evans et al. 2009,

Goberdhan et al. 2005, Jin et al. 2003, Martin et al. 2000).  Some aphid amino acid

transporters enriched in bacteriocytes are paralogs derived from within an aphid-specific

gene expansion.  The membership of bacteriocyte-enriched amino acid transporters to an

aphid-specific expansion intrigues us because gene duplication can be a critical source of

raw genetic material for evolutionary innovation.  While gene duplication is random,

duplicates can be maintained in a genome for many reasons, including the evolution of

novel functions and/or the spaciotemporal partitioning of ancestral function across

paralogs (reviewed in Kondrashov 2012). Finding amino acid transporter gene duplicates with enriched expression in bacteriocytes is consistent with the hypothesis that gene duplication plays an important, possibly adaptive, role in recruiting amino acid transporters to the symbiotic interface of aphids and other sap-feeders (Price et al. 2011). This hypothesis predicts that other sap-feeders with obligate bacterial endosymbionts also maintain duplicated amino acid transporters with similar patterns of bacteriocyte enrichment.

Most sap-feeding insects are hemipterans and thus, testing the role of gene duplication in recruiting amino acid transporters to the symbiotic interface can be facilitated with genomic data from sap-feeding and non sap-feeding hemipteran taxa. Despite difficulties resolving higher-level hemipteran relationships (Campbell et al. 1995, Cryan & Urban 2011, Dohlen & Moran 1995, Grimaldi & Engel 2005, Song et al. 2012), current understanding of hemipteran suborders can facilitate selection of appropriate taxa to evaluate whether symbiosis influences amino acid transporter evolution. Ideal taxon sampling will span the three major hemipteran suborders of Sternorrhyncha, Auchenorrhyncha, and Heteroptera (see Figure 1). Sternorrhyncha, the suborder that includes aphids, will enable determination of whether the AAT duplications we discovered in the pea aphid (Price et al. 2011) pre- or postdate diversification of the Sternorrhyncha. Draft genomes and transcriptomes are available for four sternorrhynchan lineages including the pea aphid *A. pisum* (International Aphid Genomics Consortium 2010), the whitefly *Bemisia tabaci* (Wang et al. 2010), the potato psyllid *Bactericera cockerelli* (Nachappa et al. 2012), and the citrus mealybug *Planococcus citri* (Husnik et al. 2013). Auchenorrhyncha, a suborder that independently

evolved sap-feeding (Zrzavy 1990, 1992), will provide tests of independence (Weber & Agrawal 2012) in amino acid transporter evolutionary patterns. Here, I use transcriptome generated by a colleague for an auchenorrhynchan, the cicada *Diceroprocta semicincta*. Lastly, Heteroptera comprises mostly non sap-feeders, and inclusion of this suborder will provide a test of whether gene duplication in amino acid transporters is influenced by a general aspect of hemipteran biology unrelated to diet. A transcriptome is available for a blood-feeding heteropteran, the kissing bug *Rhodnius prolixus* (Ribeiro et al. 2014).

In this study, I use comparative transcriptomics and gene expression analyses to test the role of gene duplication in recruiting amino acid transporters to the sap-feeder symbiotic interface by pinpointing the relative timing of gene duplication in hemipteran amino acid transporters and quantifying expression of amino acid trasnporters in bacteriocytes. Importantly, the sap-feeding taxa I sampled have comparable symbiotic interfaces to the aphid/*Buchnera* system: one or more obligate, bacterial symbionts residing within host-derived membrane-bound compartments inside bacteriocytes (Table 3.1). Remarkably, I find that numerous gene duplications took place independently in sap-feeders of the suborder Sternorrhyncha. Consistent with observations of aphid amino acid transporters (Price et al. 2011), I find that citrus mealybug paralogous amino acid transporters are also differentially expressed at the symbiotic interface, with some paralogs enriched in bacteriocytes. Together, these data indicate that gene duplication has broadly played a role in recruiting amino acid transporters to operate at the symbiotic interface of sternorrhynchans.

**Methods**

**Insect collection and cultivation**

Adult female cicadas (*Diceroprocta semicincta*) were collected in Tucson, AZ by John McCutcheon and preserved in RNAlater (Ambion). Citrus mealybugs (*Planococcus citri*) were collected from coleus plants in the Utah State University Greenhouse in Logan, Utah (Dohlen et al. 2001, McCutcheon & Dohlen 2011) by Carol von Dohlen and raised on coleus plants in the laboratory at 25º C. Pea aphids (*Acyrthosiphon pisum*) from the genome line LSR1 (Caillaud et al. 2002) were raised on fava plants at 20º C. Both insect colonies were maintained under a photoperiodicity of 16:8 (L:D).

**Transcriptome sequencing and assembly**

Cicada transcriptomes were generated by JT Van Leuven. Total RNA was purified from either (1) bacteriocytes or (2) a combination of head, legs, and wing muscles (hereafter referred to as "insect") dissected from RNAlater preserved adult female cicadas according to manufacturer's protocols (MoBio PowerBiofilm RNA Isolation Kit). RNA was sent to Hudson Alpha Institute for Biotechnology for barcoded library preparation and Illumina HiSeq sequencing. Paired-end 100 nt reads were filtered to a minimum quality of 20 over 95% of the read, and 5 nt were trimmed from the 5' end. Insect (42,688,895 read pairs) and bacteriocyte (53,510,432 read pairs) reads were assembled into separate insect and bacteriocyte transcriptomes in Trinity (January 25, 2012 release) (Haas et al. 2013) using kmer_length=25 and min_contig_length=48.

To generate the mealybug whole body transcriptome, I filtered paired end, 100 nt reads sequenced from a mixed population of adult and penultimate instar females (Husnik

et al. 2013) to a minimum quality of 30 over 95% of the read. The resulting 58,812,530

read pairs were assembled with two different assembly packages. First, reads were

assembled in velvet (v.1.2) (Zerbino & Birney 2008) and oases (v.0.2) (Schulz et al.

2012) using variable k-mer lengths (between 33 and 63 nt) and resulting assemblies were

merged into one master assembly. Second, reads were assembled in Trinity using default

parameters (kmer_length = 25).

The whitefly (*Bemisia tabaci*) transcriptome was re-assembled by Don Gilbert

using 170,884,234 RNA-seq read pairs from a mixed population of adult males and

females (NCBI BioProject PRJNA89143). Reads were assembled with RNA assemblers

velvet/oases (v.1.2.03/v.0.2.06) (2012.02), SOAPdenovo-Trans (v.2011.12.22), and

Trinity (March 17, 2012 release), using multiple options. EvidentialGene (Gilbert 2013)

tr2aacds pipeline software was used to process the many resulting assemblies by coding

sequences, translate to proteins, score gene evidence and classify/reduce to a biologically

informative transcriptome of primary and alternate transcripts. The gene set is available

at http://arthropods.eugenes.org/EvidentialGene/arthropods/whitefly/whitefly1eg6/.


### *De novo* identification of hemipteran amino acid transporters

Amino acid transporters were identified using HMMER (v.3.0) (Eddy 2009) from

transcriptomes of cicada, mealybug, whitefly, the potato psyllid *Bactericera cockerelli*

(Nachappa et al. 2012), and the kissing bug *Rhodnius prolixus* (NCBI BioProject

PRJNA191820; mixed developmental stages and sexes). Briefly, using a stand-alone

PERL script underlying the open reading frame (ORF) prediction webserver hosted by

the Proteomics/Genomics Research Group at Youngstown State University

(http://bioinformatics.ysu.edu/tools/OrfPredictor.html), transcripts were translated into all six reading frames.  As described previously (Price et al. 2011), translated transcripts were searched for conserved functional domains associated with the APC (TC # 2.A.3) and AAAP (TC # 2.A.18) families of amino acid transporters (Castagna et al. 1997, Saier 2000, Saier et al. 2009, 2006) in HMMER (v.3) (Eddy 2009, Finn et al. 2011). Transcripts significantly matching APC or AAAP domains (E ≤ 0.001) were verified by BLASTX searches against the NCBI refseq database, and retained for further analyses if they showed significant (E ≤ 0.001) similarity to the APC or AAAP sequences from the fruit fly *Drosophila melanogaster* and/or *A. pisum*.

Alleles and splice variants were collapsed into a conservative set of representative transcripts for each insect by one of the following two methods depending on availability of genome sequence data: (1) Draft genome assemblies are available for mealybugs (Husnik et al. 2013) and kissing bugs (unpublished; hosted at vectorbase.org and NCBI), so we validated loci by mapping transcripts to genomic scaffolds by BLASTN searches. Of the transcripts mapping to the same region of a particular scaffold(s), the transcript encoding the longest protein was kept to represent the gene locus. In a few cases, 2-3 partial transcripts were merged into a single locus for phylogenetic analyses (Tables S1-S4 in supporting file 1 in Duncan et al. (2014)). In all cases, partial transcripts mapped side by side to genomic scaffolds on the same strand. Additionally in all cases, the partial transcript mapping upstream in the genome aligned to the 5' end of other, full-length AAT loci and the downstream partial transcript aligned to the 3' end. (2) In contrast, whiteflies, psyllids, and cicadas lack draft genome assemblies. In these insects, transcripts that have been diverging for a short period of time were collapsed into

representative loci.  Time of transcript divergence was determined by estimating the

pairwise rate of synonymous substitutions (dS) by the Goldman and Yang method

(Goldman & Yang 1994), a common proxy for relative age of homologous gene pairs

(e.g. paralogs within a species or orthologs between species) (Lynch & Conery 2000).  I

collapsed all transcripts with a dS of less than 0.25, keeping the longest sequence to

represent the locus (supporting file 2 from Duncan et al. (2014)).  This cutoff dS (0.25) is

the average dS between orthologs of two aphid species (*A. pisum* and *Myzus persicae*)

that diverged between 32 and 53 million years ago (International Aphid Genomics

Consortium 2010, Kim et al. 2011a).  When closely related transcripts for a particular

taxon were partial and non-overlapping or had a very short region of overlap (50 bp or

less), we removed the shortest of the pair to ensure conservative estimates of locus

number.  To confirm the accuracy of using pairwise dS to collapse related transcripts into

loci, we performed the same analysis on related aphid paralogs in the APC gene family

(supporting file 2 from Duncan et al. (2014)), all of which map to unique regions of the

aphid genome (Price et al. 2011).  We found three aphid-specific paralogs with pairwise

dS measurments below 0.25 (supporting file 2 from Duncan et al. (2014)), indicating that

our approach to estimate locus number may collapse true paralogs that duplicated

relatively recently. Thus, importantly, our estimation of locus number is conservative.


**Phylogenetic analyses**

Gene phylogenies for the APC and AAAP amino acid transporter families were

estimated using sequences from citrus mealybug, potato psyllid, whitefly, cicada, and

kissing bug as well as previously annotated AATs (Price et al. 2011) from the pea aphid,

the human body louse (*Pediculus humanus*), the fruit fly (*D. melanogaster*), a tick (*Ixodes scapularis*), and humans (*Homo sapiens*). Outgroup sequences were aphid and/or fruit fly genes closely related to APC and AAAP gene families and members of the same transporter superfamily (Price et al. 2011). Full length protein sequences were aligned in MAFFT (Katoh et al. 2002) using default parameters and resulting alignments were trimmed in trimAl (v.1.2) (Capella-Gutiérrez et al. 2009) using a gap threshold of 25%.

Phylogenies were estimated using Maximum Likelihood (ML) and Bayesian methods. ML phylogenies were estimated in RAxML (v.7.2.8) (Ott et al. 2007, Stamatakis 2006) using the protein evolution model LG+G (the best-fit model as determined by ProtTest (v.2.4) using the Akaike Information Criterion (Abascal et al. 2005)) and the fast bootstrap option. The number of bootstrap replicates for each analysis was chosen by the bootstrap convergence criterion "autofc" implemented in RAxML. Bayesian phylogenies were reconstructed in MrBayes (v.3.1.2) (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003) using two runs with 4 chains per run. The LG protein substitution matrix is not available in MrBayes, so phylogenies were inferred using WAG+G. Analyses were allowed to run until the standard deviation of split frequencies between runs dropped below 0.05. Convergence of estimated parameters was confirmed in Tracer (v.1.5) (Rambaut et al. 2014) and of topology in AWTY (Nylander et al. 2008), assuming a burn-in of 10% of generations. The criteria supported convergence, so the first 10% of generations were discarded and phylogenies sampled in the remaining generations were used to estimate a 50% majority-rule consensus tree.

The AAAP family contained a large amount of sequence divergence, preventing convergence of the Markov Chain Monte Carlo (MCMC) in Bayesian phylogenetic

analyses. Therefore, we estimated the phylogeny of a reduced set of AAAP genes corresponding to a monophyletic clade supported in a preliminary maximum likelihood analysis (Figure S1 in supporting file 1 from Duncan et al. (2014)).

**Gene conversion analyses**

Lineage-specific AAT expansions were assessed for the possibility of gene conversion using the program GENECONV (Sawyer 1989).  Codon alignments were produced by the Clustal W plugin of SeaView (Gouy et al. 2010), and run in GENECONV using three different mismatch penalties, g0, g1 and g2.  Applying different mismatch penalties to the analysis facilitates the identification of recent gene conversion and ancient gene conversion that may be partially masked by the accumulation of different substitutions between paralogs.

**Expression analysis by quantitative reverse transcriptase PCR**

Expression profiles of select AATs were measured by quantitative reverse transcriptase PCR (qRT-PCR) in whole bodies and bacteriocytes of adult female LSR1 pea aphids, a mixture of adult and penultimate female citrus mealybugs and adult female potato psyllids (from the same colonies used for the potato psyllid transcriptome (Nachappa et al. 2012)).  Bacteriocytes were dissected from 100 female aphids, mealybugs, or psyllids in 0.9% RNase free NaCl, and immediately stabilized by placing in TRI Reagent (Ambion).  Total RNA was extracted from dissected bacteriocytes and whole female bodies of each insect using the TRI Reagent procedure (Ambion), treated with DNase I in solution and cleaned up using the RNeasy Mini Kit (Qiagen).  First

strand cDNA was synthesized from 500 ng of RNA from each tissue, using qScript

cDNA Supermix (Quanta Biosciences) and following the manufacturer's protocol.

qRT-PCR assays were performed as previously described (Price et al. 2011) using

one biological replicate and three technical replicates for each gene/tissue.  Primers were

subject to BLASTN searches against genomic and/or transcriptomic data sets using a

word length of 7, an expect threshold of 1000, and without the low-complexity filter.  In

all cases, only the target sequence was returned as a hit for each pair of forward and

reverse primers.  To confirm that primers amplified only one locus, we analyzed melt

curves from our qRT-PCR results.  With the exception of one gene, which was discarded

from analysis, all melt curves showed one clear peak, indicating a single product.  No

template controls and no reverse transcriptase controls (controlling for RNA

contaminated with gDNA) were run in parallel with unknown samples.  Identifiers,

sequences, amplification efficiency and optimization details for primers used in qRT-

PCR assays are listed in table S6 (supporting file 1 from Duncan et al. (2014)).

Expression for target genes within a particular insect was compared between whole insect

and bacteriocytes using $2^{-\Delta\Delta CT}$ methodology (Livak & Schmittgen 2001) with expression

normalized to either glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) in aphids or

the 60S Ribosomal Protein L7 (*RPL7*) in mealybugs and psyllids.  Expression data within

each insect were collectively normalized by converting $\Delta C_T$ to z-scores as follows:

$$z = -10 \times \left( \frac{\Delta C_T - \overline{\Delta C_T}}{\sigma_{\Delta C_T}} \right)$$

Normalized expression values were compiled into a heat map where $z > 0$ (high

expression) is represented as yellow and $z < 0$ (low expression) is represented as blue.

**Differential expression quantification**

Filip Husnik quantified global differential expression between mealybug insect

and bacteriocyte tissues using the whole body transcriptome generated in this study and

previously a published bacteriocyte transcriptome (Husnik et al. 2013). Differential

expression analyses were conducted with the PERL script pipeline implemented in

Trinity. Briefly, raw RNA-seq reads were mapped to transcripts using Bowtie (v.0.12.7)

(Langmead et al. 2009) and mapped reads were counted by RSEM (v.1.1.18) (Li &

Dewey 2011). Data were normalized by TMM (trimmed mean of M values) and

transcripts significantly differentially expressed between whole body and bacteriocytes

were identified using the Bioconductor package edgeR (v.2.10) (Robinson et al. 2010).

Digital expression values of differentially expressed transcripts are presented in

supporting file 3 (Duncan et al. 2014) as "fragments per kilobase of exon per million

fragments mapped" (FPKM). Differential expression was not quantified for cicada

bacteriome vs. insect tissues because cicadas lacked gene duplications.

**Results and Discussion**

**Nutrient amino acid transporter families are expanded in the Sternorrhyncha**

Consistent with our pea aphid work (Price et al. 2011), all sternorrhynchan

hemipterans we sampled (Table 3.1, Figure 3.1) possessed expanded amino acid

transporter families relative to non sap-feeding insects (kissing bug, human body louse,

and the fruit fly) (Table 3.2).  In particular, citrus mealybugs, potato psyllids, and

whiteflies possessed 36-38 amino acid transporter loci across both gene families;

relatively large amino acid transporter numbers compared with the 20 loci in the non sap-

feeding hemipteran annotated here (kissing bug; Table 3.2) and 22-28 amino acid

transporter loci in other insects annotated by Price et al. (2011) (the fruit fly *D.*

*melanogaster*, the body louse *P. humanus*, the honey bee *Apis mellifera*, the flour beetle

*Tribolium castaneum*, the silkworm moth *Bombyx mori*, the wasp *Nasonia vitripennis,*

and the mosquito *Anopheles gambiae*).  In contrast, I identified only 26 amino acid

transporter loci in cicada, a sap-feeder belonging to the hemipteran suborder

Auchenorrhyncha (Table 3.1, Figure 3.1).


**Amino acid transporter expansions in sap-feeding insects result from both ancient
and recent gene duplication events**

To clarify the evolutionary mechanism and timing of events that led to amino acid

transporters expanding in sap-feeding insects, I estimated phylogenies for the APC and

AAAP amino acid transporter families Figures 3.2A and 3.3A). The phylogenies revealed

that gene duplications occurred on two time scales.  First, two ancient gene duplication

events (one in each gene family) predate hemipteran diversification (marked by pale

green boxes in Figures 3.2A and 3.3A).  Second, consistent with previous observations in

aphids (Price et al. 2011), multiple, more recent, gene duplications occurred

independently in sternorrhynchan taxa following their divergence from a common

ancestor (marked by gray boxes in Figures 3.2A and 3.3A).  In contrast, analyses failed to

support any Auchenorrhyncha-specific gene duplications in either amino acid transporter

family.  That said, in four instances (marked in Figures 3.2A and 3.3A with asterisks) I

found phylogenetic support for close relationships between 2-3 cicada (auchenorrhynchan) loci and one kissing bug (heteropteran) locus. Two scenarios could explain these close relationships. First, amino acid transporter duplication could have taken place independently in the linage leading to cicadas and no gene duplication took place in kissing bugs, but sequence similarity among orthologs prevents resolution of cicada-specific clades. Second, assuming species tree B (Fig 3.1B), gene duplications took place in the common ancestor of cicadas and kissing bugs but paralogs were only retained in cicada. Of the two scenarios, the second is the least parsimonious, requiring that cicadas retain their paralogs and that kissing bugs lose all but one paralog in three independent instances.

Price et al. (2011), previously found that pea aphid amino acid transporter paralogs were tandemly arrayed in the genome. Although new amino acid transporters in this study were annotated from transcriptome data, a draft genome assembly for the citrus mealybug (Husnik et al. 2013) enabled me to preliminarily assess paralog arrangement in that genome. In the mealybug AAAP expansion (Figure 3A), three pairs of paralogs map to different regions of the same scaffold within ~4 kbp or less of each other (Figure 3.4, Table S2 in supporting file 1 from Duncan et al. (2014)), indicating that these paralogs are tandemly arrayed in the mealybug genome. These tandemly arrayed paralogs thus resulted from localized gene duplication (as opposed to whole genome duplication). No other mealybug amino acid transporter loci shared a scaffold (Tables S1-S2 in supporting file 1 from Duncan et al. (2014)), which could at least in part be due to the poor quality of the assembly (Husnik et al. 2013). In the kissing bug genome, several transcripts mapped to the same scaffold (Tables S3-S4 in supporting file 1 from Duncan et al. (2014)), but

were usually separated by large genomic regions between 19 kbp and 1.2 Mbp. The only

exception was that two loci were separated by 5.7 kbp (Tables S3-S4 in supporting file 1

from Duncan et al. (2014)). Despite the short distance between those two loci, the

phylogeny (Figure 3.3A) indicates that they did not result from a recent gene duplication

event in the lineage leading to kissing bugs.

**Amino acid transporter evolution within the Sternorrhyncha**

One unexpected result of this work is finding that amino acid transporters have

undergone gene family expansions independently in each of the sternorrhynchans I

sampled (aphids, mealybugs, psyllids and whiteflies). Consistent molecular and

morphological phylogenetic support for the monophyly of Sternorrhyncha (Campbell et

al. 1995, Cryan & Urban 2011, Dohlen & Moran 1995, Grimaldi & Engel 2005, Hennig

1981, Song et al. 2012) indicates that aphids, mealybugs, whiteflies, and psyllids

inherited sap-feeding from their common ancestor. I thus assumed that the common

ancestor also had an amino acid-provisioning endosymbiont that was later replaced in

three, or perhaps all four, lineages we sampled (explaining why each lineage has a

different symbiont, Table 3.1). Importantly, this common ancestor required that amino

acid transporters mediate host/symbiont amino acid exchange. Retention of

independently duplicated paralogs in sternorrhynchans could be explained in four ways.

First, my understanding that symbiosis predates sternorrhynchan diversification could be

wrong, and each lineage independently evolved symbiosis and comparable symbiotic

interfaces. Second, the importance of different transporters could depend on the

symbiont lineage. Third, some amino acid transporter gene duplications in these taxa

could appear to be more recent than they truly are if tandem arrays of paralogs have undergone concerted evolution through gene conversion or non-homologous crossing over after the major sternorrhynchan lineages (aphids, mealybugs and other scale insects, whiteflies and psyllids) began diversifying (e.g. Colbourne et al. 2011). I found evidence of gene conversion only among a few paralogs in aphids and whiteflies (Table S5 in supporting file 1 from Duncan et al. (2014)), indicating that AAT paralogs are largely evolving independently of one another. However, I cannot rule out the possibility that gene conversion played a more important role in paralog diversification at some time in the past. Fourth, consistent with my previous discovery of male-biased amino acid transporter paralogs in aphids, presented in Chapter 2 of this dissertation and published as Duncan et al. (2011), many paralogs are probably retained in sternorrhynchan genomes for lineage-specific roles not related to symbiosis. Most aphid and mealybug amino acid transporer paralogs are not enriched in bacteriocytes, supporting a role for non-symbiotic factors in driving the maintenance of amino acid transporter paralogs in these insects. Given the important role that amino acid transporters play broadly in animals it is not surprising that paralog maintenance in sternorrhynchan genomes is not only driven by symbiosis. For example, nutrient amino acid transporters also mediate amino acid uptake from the gut into hemolymph (insect blood) (Colombani et al. 2003, Morris et al. 2009, Price et al. 2011). Further, some nutrient amino acid transporters play a role in nutrient sensing (Attardo et al. 2006, Colombani et al. 2003). Lastly, some amino acid transporters transport neurotransmitters, likely explaining their expression in aphid heads (Price et al. 2011). Accepting their many roles, it is not surprising that some insects without intracellular, amino acid provisioning symbionts maintain lineage-specific amino

acid transporter duplications (e.g. Fig, 3.3 and Price *et al.* 2011). However, that sternorrhynchan sap-feeding insects maintain more amino acid transporter duplications in their genomes than other, non sap-feeding insects is compelling and suggests that gene duplication has facilitated amino acid tranpsorter recruitment to bacteriocytes - at least in the Sternorrhyncha. Nevertheless, the absence of duplicates in cicada indicates that gene duplication is not a prerequisite for initiation of host/symbiont amino acid exchange in sap-feeding insects; an interpretation consistent with the fact that some single-copy amino acid trasnporters also operate at the symbiotic interface in aphids (Price et al. 2011) and mealybugs (Figure 3.3C).

**Amino acid transporter recruitment to the symbiotic interface is dynamic**

I measured expression of paralogs resulting from both ancient and recent gene duplication events because both could play a role in recruiting amino acid transporters to the symbiotic interface. Although most examples of gene duplication giving rise to novelty involve gene duplication evolving concurrently with or after the origin of new traits, there are some examples of gene duplication pre-dating the evolution of novelty (Arnegard et al. 2010, Ben Trevaskis et al. 1997). Expression patterns in both the anciently and recently duplicated amino acid transporters (Figure 3.2 and 3.3) indicate that amino acid trasnporters were recruited independently to the bacteriocytes of different sap-feeding insect lineages. In the ancient duplications predating hemipteran diversification, qRT-PCR results for aphids, mealybugs, and psyllids indicate that bacteriocyte enrichment is derived in the psyllid transporter *Bcoc-APC12* (Figure 3.2B and Figure S2 from Duncan et al. (2014)) and the aphid transporter *ACYPI000536*

(Figure 3.3B and Figure S2 in supporting file 1 from Duncan et al. (2014)) is derived.

This finding is consistent with sap being a derived diet within Hemiptera (Cobben 1979, Zrzavy 1990, 1992), requiring that amino acid transporters be independently recruited to the symbiotic interface after hemipteran suborders (and these orthologs) diverged from their common ancestor. Biological replication within each sternorrhynchan lineage (psyllids, mealybugs and aphids) would provide finer resolution of the extent to which expression is or is not conserved in these taxa. However, lack of within species biological replication does not compromise our finding that expression of orthologous amino acid transporters is not conserved.

Similarly with respect to the recent taxon-specific gene duplications, qRT-PCR from this study (Figure 3.2, 3.3, supporting file 1 from Duncan et al. (2014)) and Price *et al.* (2011) together with RNA-seq differential expression data (Hansen & Moran 2011, Macdonald et al. 2012, Price et al. 2011) support independent amino acid transporter recruitment to the symbiotic interface in pea aphids and citrus mealybugs (Figure 3.2C and 3.3C). Notably, bacteriocyte expression in aphid amino acid transporters is remarkably consistent across qRT-PCR and RNA-seq studies that together include data from four different pea aphid lineages at different developmental stages (Duncan et al. 2014, Hansen & Moran 2011, Macdonald et al. 2012, Price et al. 2011). Similar to what was previously reported for pea aphids (Hansen & Moran 2011, Price et al. 2011), six mealybug-specific paralogs have enriched bacteriocyte expression (Figure 3.3C, supporting file 3 from Duncan et al. (2014)). As reported previously by Price et al. (2011), expression profiles among aphid APC paralogs (Figure 3.2C) are most parsimoniously explained by bacteriocyte enrichment evolving after (and potentially

being enabled by) gene duplication; an argument based on the fact that bacteriocytes are a novel, derived tissue and most aphid APC paralogs, like their insect orthologs, are highly expressed in gut. In contrast, the distribution of bacteriocyte-enrichment among mealybug AAAP paralogs (Figure 3.3C) lacks a clear most parsimonious explanation. Bacteriocyte enrichment/expression could be derived or ancestral, consistent with either neofunctionalization or subfunctionalization of paralogs. Furthermore, some paralogs may be functionally redundant and are maintained for dosage reasons or are differentially expressed across time and space. Indeed, some aphid amino acid transporter paralogs are enriched in head and gut tissues, (Price et al. 2011), and others have male-biased expression (Duncan et al. 2011). Distinguishing between these explanations will be facilitated with functional data for paralogs of this expansion and their orthologs in other insects. However, that multiple paralogs show bacteriocyte enrichment together with substantial sequence divergence among paralogs (indicated by long branches) strongly suggests that at least some mealybug paralogs have evolved novel functional roles. These results are thus consistent with the hypothesis that gene duplication played a role in recruiting mealybug amino acid transporters to the symbiotic interface, enabling them to carry out novel, symbiotic functions.

Interestingly, expression patterns indicate that aphids and mealybugs use different amino acid transporters at their symbiotic interface. For example, bacteriocyte enriched amino acid trasnporters are not orthologous between aphids and mealybugs (Figures 3.2 and 3.3). Recruitment of different amino acid transporters in aphids and mealybugs could reflect differences in nutritional demand between these insects, or could simply result from chance. Alternatively, amino acid transporters could be functionally dynamic,

with similar environmental pressures experienced by aphids and mealybugs resulting in distinct amino acid transporter loci converging upon common functional roles.

**Differential amino acid transporter expansion among sap-feeding hemipterans is consistent with co-evolutionary patterns of host/symbiont metabolic collaboration**

Despite evidence that gene duplication has facilitated recruitment of amino acid transporters to the symbiotic interface in the Sternorrhyncha, cicadas demonstrate that gene duplication is not necessary to initiate novel sap-feeder/symbiont amino acid exchange. Cicadas did not experience expansions in their amino acid transporters, a pattern that may relate to a dietary difference between cicadas and sternorrhynchan sap-feeders. While sternorrhynchans feed on plant phloem sap (Gullan et al. 2003), the source of sap for cicadas is the plant xylem (White & Strehl 1978), a more dilute source of nitrogen than phloem (Redak et al. 2004). It is unclear how amino acid concentration *per se* could influence host insect amino acid transporter evolution and recruitment to the symbiotic interface. However, differences in individual amino acid content could potentially influence the nutritional demands of different sap-feeding insects and thus the evolutionary trajectory of amino acid transporters operating at the symbiotic interface. However, recent sequencing of the symbiont genomes of a phloem-feeding auchenorrhynchan suggests that differences in amino acid transporter copy number between sternorrhynchans and auchenorrhynchans are not driven by diet.

Bennett and Moran (2013) recently sequenced *Sulcia muelleri* and *Nasuia deltocephalinicola,* the obligate symbionts of the phloem feeding auchenorrhynchan *Macrosteles quadrilineatus*. Their work highlights an important genomic difference between the obligate symbioses of sternorrhynchans and auchenorrhynchans. Obligate

symbionts of both sternorrhynchans and auchenorrhynchans play a major role in providing their hosts with essential amino acids. However, while symbionts of both phloem-feeding and xylem-feeding auchenorrhynchans retain relatively autonomous metabolic pathways (Bennett & Moran 2013, McCutcheon & Moran 2007, McCutcheon et al. 2009, Wu et al. 2006), sternorrhynchan symbionts lack some genes for crucial metabolic steps; metabolic steps that the host has been demonstrated to complement (Russell et al. 2013). For example, sternorrhynchan symbionts typically lack genes necessary to complete the terminal steps in branch-chain amino acid and phenylalanine biosynthesis as well as the step required to synthesize homocysteine for methionine biosynthesis (Husnik et al. 2013, McCutcheon & Dohlen 2011, Nakabachi et al. 2006, Sabree et al. 2012b, Shigenobu et al. 2000, Sloan & Moran 2012). These missing steps are carried out by host insect enzymes (McCutcheon & Dohlen 2011, Shigenobu & Wilson 2011, Wilson et al. 2010). This within metabolic pathway host/symbiont collaboration likely necessitates host/symbiont exchange of intermediate metabolites, a step that is not required in auchenorrhynchans that possess metabolically autonomous symbionts (McCutcheon & Moran 2007, McCutcheon et al. 2009, Wu et al. 2006). Therefore, gene duplication could have enabled, through neofunctionalization of paralogs, the evolution of novel transporters capable of transporting intermediate metabolites in amino acid biosynthesis pathways, facilitating pathway partitioning between hosts and symbionts. Once functional data are available for these transporters, this hypothesis can be tested. Thus, current evidence suggests that differences in amino acid transporter copy number between sternorrhynchans and auchenorrhynchans are driven by differences in the extent of host/symbiont metabolic independence.

**Gene duplication and the evolution of novel, symbiotic interactions**

The generation of genomic resources for non-model organisms, including the partners of symbiotic systems, make it possible to understand how intimate symbiotic relationships have influenced genome evolution in both symbionts (McCutcheon & Dohlen 2011, McCutcheon & Moran 2007, 2012; McCutcheon et al. 2009, Nakabachi et al. 2006, Nikoh et al. 2011, Sabree et al. 2009, 2012a; Shigenobu et al. 2000) and hosts (Husnik et al. 2013, International Aphid Genomics Consortium 2010, Kirkness et al. 2010, Nygaard et al. 2011, Young et al. 2011). The pea aphid/*Buchnera* symbiosis was the first symbiotic system to have both host and symbiont genomes sequenced, providing the first insights into how host genomes are shaped by symbiosis. Here, I provide evidence that one of those insights applies more broadly to sternorrhynchan sap-feeding insects: gene duplication plays a role in recruiting amino acid transporters to operate at the host/symbiont interface. Further, recent studies in other, very divergent, symbiotic systems also invoke gene duplication in the evolution of genes with symbiotic functions. For example, in legumes, an ancient whole genome duplication event in the ancestor of the major papilionoid subfamily was followed by some paralogs evolving enriched expression in symbiotic root nodules. This pattern correlates with the evolution of many important Nod factor signaling components that are critical for legume/*Rhizobium* recognition and the initiation of nodulation in this subfamily (Young et al. 2011). Additionally, gene duplication may have facilitated the origin of leghemoglobin, a special hemoglobin protein that legumes use to remove oxygen from symbiotic root nodules, facilitating symbiotic nitrogen fixation (Anderson et al. 1996, Ben Trevaskis et al. 1997). Similarly, in an anenome/dinoflagellate symbiosis, cnidarian-specific paralogs gave rise

to three genes proposed to function in symbiosis. All three of these cnidarian-specific paralogs are both enriched in individuals hosting symbionts (as opposed to aposymbiotic individuals lacking symbionts) and preferentially expressed in the gastroderm, where symbionts are housed (Ganot et al. 2011). Together with the results presented here, these plant and cnidarian studies suggest that gene duplication facilitates recruitment of non-symbiotic genes to play a role in symbiosis broadly across symbiotic systems. The independent evolution in diverse symbiotic systems of gene duplication followed by expression in tissues that host symbionts, however intriguing, does not in itself provide insight into the potential adaptive significance of gene duplication in the evolution of symbiosis-related genes. The crucial next step to deciphering the role of gene duplication in the evolution of symbiotic interactions will be functional characterization within a phylogenetic framework, which will reveal if paralogs preferentially expressed at the host/symbiont interface have also evolved novel symbiotic functions.

**A.**



**B.**
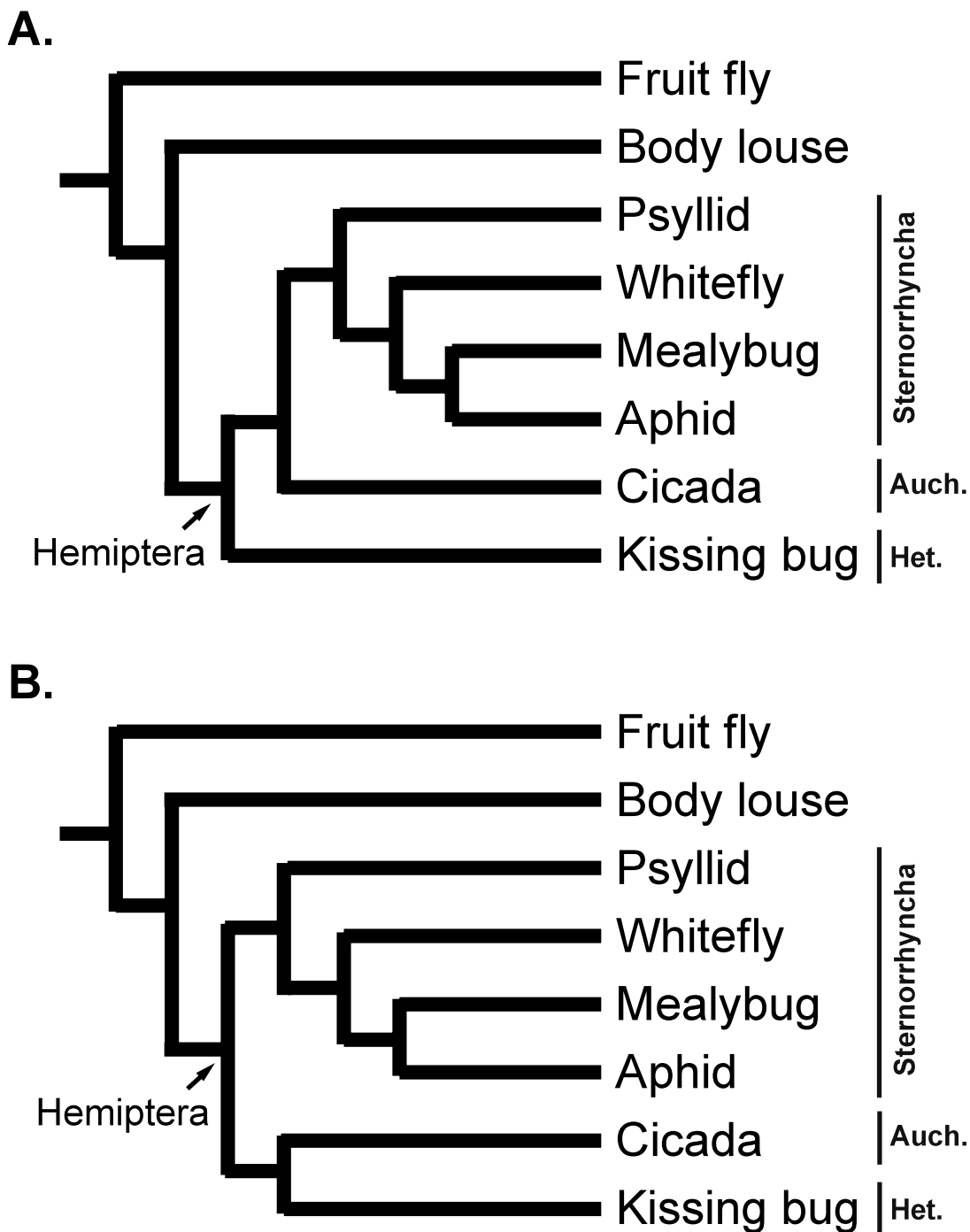


Figure 3.1. Alternative hypotheses for phylogenetic relationships among sampled hemipterans. (A) Sternorrhyncha + cicada sister to kissing bug (consistent with (Hennig 1981, Song et al. 2012)). (B) Sternorrhyncha sister to cicada + kissing bug (consistent with (Campbell et al. 1995, Dohlen & Moran 1995, Grimaldi & Engel 2005, Zrzavy 1992). Suborders are indicated to the right of taxon names: Sternorrhyncha (aphids, mealybugs, whiteflies and psyllids), Auchenorrhyncha (cicadas), and Heteroptera (kissing bugs).

Figure 3.2. APC (TC # 2.A.3) phylogeny and bacteriocyte expression. (A) Bayesian gene phylogeny for amino acid transporters (AATs) in the APC family. Hemipteran-specific gene duplications and taxon-specific expansions are highlighted with green or gray boxes, respectively. Asterisks denote possible cicada-specific paralogs. Branches are color-coded based on taxon and clade support ≥ 50% (posterior probability and ML bootstrap support) is indicated on branches/nodes as described in the key. (B) qRT-PCR expression data generated in this study for Hemiptera-specific gene duplication are presented both as a heat map for whole insect ("WI") and bacteriocyte ("B") and as differential expression ("DE") between bacteriocyte and whole insect. Heat map expression data are normalized across all tissues and genes within each insect, but not across insects. (C) Differential expression between whole insect and bacteriocyte is indicated for aphid and mealybug genes in boxes to the right of gene IDs, as indicated in the key. RNA-seq differential expression data for aphids are from Hansen & Moran (2011) and Macdonald *et al.* (2012). qRT-PCR data not generated here are from Price *et al.* (2011). Expression is marked as ambiguous ("A") if different transcripts or data sets show inconsistent relative bacteriocyte expression.

**A.**



ACYPI001649
ACYPI007138

Outgroups:
— Human
— Tick
— Body Louse
— Fruit fly

Hemipterans:
— Pea aphid
— Citrus mealybug
— Potato psyllid
— Whitefly
— Cicada
— Kissing bug

Posterior probability:
— ≥ 0.95
— 0.80-0.95
-- < 0.80

Bootstrap support:
● ≥ 95%
● 80-95%
○ < 80%

SLC7A1
SLC7A2
SLC7A3
SLC7A4
ACYPI005006
Bcoc-APC6
CG13248
PHUM169950
Btab-APC6
Pcit-APC5
ISCW001425
ISCW007095
SLC7A14
ACYPI009985
Pcit-APC4
ISCW018487
Btab-APC5
Bcoc-APC7
PHUM499080
CG12531

ACYPI003523
ACYPI003240
ACYPI005156
ACYPI000537
ACYPI000584
ACYPI005472
ACYPI005118
ACYPI008904
ACYPI008323
ACYPI002633

Aphid-specific expansion

Pcit-APC1
Pcit-APC2
Pcit-APC3
Rpro-APC1
Bcoc-APC1
Bcoc-APC2
Bcoc-APC3
Bcoc-APC4

Mealybug-specific expansion

Psyllid-specific expansion

Dsem-APC1
Btab-APC2
Btab-APC1
Rpro-APC3

ACYPI003923
Bcoc-APC5
Btab-APC3
Btab-APC4
Dsem-APC2 ✱
Dsem-APC3 ✱
CG7255
CG5535
PHUM169840
CG11128
PHUM258130
Pcit-APC6
Rpro-APC2
PHUM169830

SLC7A8
ACYPI003130
ACYPI009467
Bcoc-APC9
Btab-APC9
Pcit-APC8
Dsem-APC5
CG6070
CG3297
PHUM136660
Pcit-APC9
Bcoc-APC10
Dsem-APC6
Btab-APC10
ACYPI005720
CG1607
Rpro-APC4
Dsem-APC4
Bcoc-APC8
Btab-APC7
Pcit-APC7
CG12317
PHUM462600
Bcoc-APC11
Btab-APC8
Dsem-APC7
Rpro-APC7

SLC7A11
ACYPI006356
CG9413
Rpro-APC5
PHUM212700
Bcoc-APC12
Dsem-APC8
Btab-APC11
Pcit-APC10
ACYPI001378
Btab-APC12
Bcoc-APC13
Dsem-APC10 ✱
Rpro-APC6
Dsem-APC9 ✱
ISCW005964

Hemipteran-specific duplication (Fig. 2B)

0.6 substitutions/site

**B.**

WI B DE

ACYPI006356
Pcit-APC11
Bcoc-APC12
PHUM212700
CG9413
ACYPI001378
Bcoc-APC13

**qRT-PCR Expression**

Low ▮▮▮▮▮▮▮ High

Differential Expression (fill):

▮ > 2-fold upregulated in bacteriocyte
▯ No difference
▮ > 2-fold downregulated in bacteriocyte
▮ Data not available
Ⓐ Ambiguous expression

Differential Expression source (outline):

▮ This study only
▯ This study and previous study
⬚ Previous study only

**C.**

RNA-seq qPCR

ACYPI005006
PcitAPC5

ACYPI009985
Pcit-APC4

ACYPI003523
ACYPI003240
ACYPI005156
ACYPI000537
ACYPI000584
ACYPI005472
ACYPI005118
ACYPI008904
ACYPI008323
ACYPI002633
Pcit-APC1
Pcit-APC2
Pcit-APC3

ACYPI003923

Pcit-APC6

ACYPI003130
ACYPI009467
Pcit-APC8

Pcit-APC9

ACYPI005720
Pcit-APC7

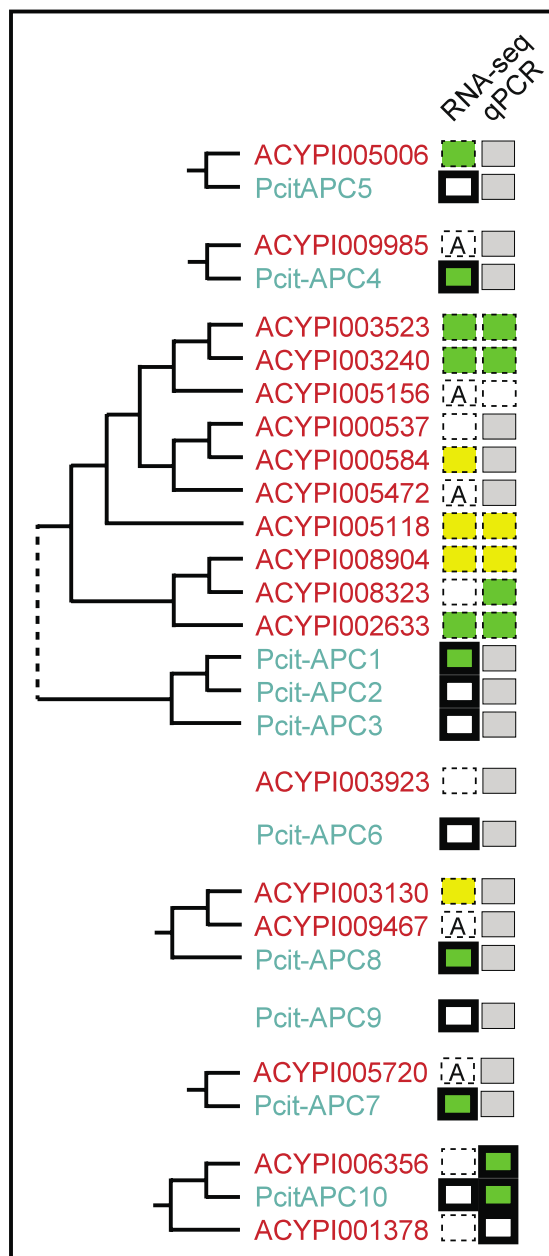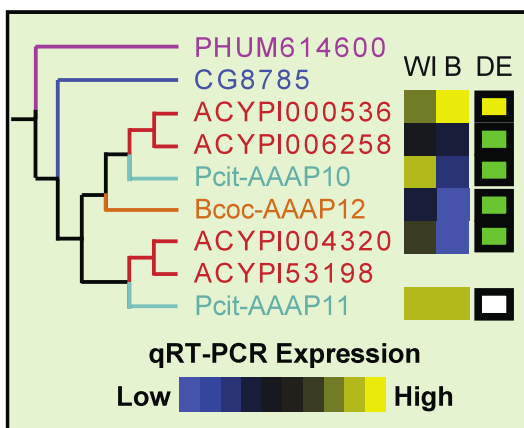ACYPI006356
PcitAPC10
ACYPI001378

Figure 3.3. Partial AAAP (TC # 2.A.18) phylogeny and bacteriocyte expression. (A) Bayesian gene phylogeny for amino acid transporters (AATs) in the AAAP family. Hemipteran-specific gene duplications and taxon-specific expansions are highlighted with green or gray boxes, respectively. Asterisks denote possible cicada-specific paralogs. Branches are color-coded based on taxon and clade support $\geq 50\%$ (posterior probability and ML bootstrap support) is indicated on branches/nodes as described in the key. (B) qRT-PCR expression data generated in this study for Hemiptera-specific gene duplication are presented both as a heat map for whole insect ("WI") and bacteriocyte ("B") and as differential expression ("DE") between bacteriocyte and whole insect. Heat map expression data are normalized across all tissues and genes within each insect, but not across insects. (C) Differential expression between whole insect and bacteriocyte is indicated for aphid and mealybug genes in boxes to the right of gene IDs, as indicated in the key. RNA-seq differential expression data for aphids are from Hansen & Moran (2011) and Macdonald *et al.* (2012). qRT-PCR data not generated here are from Price *et al.* (2011). Expression is marked as ambiguous ("A") if different transcripts or data sets show inconsistent relative bacteriocyte expression.
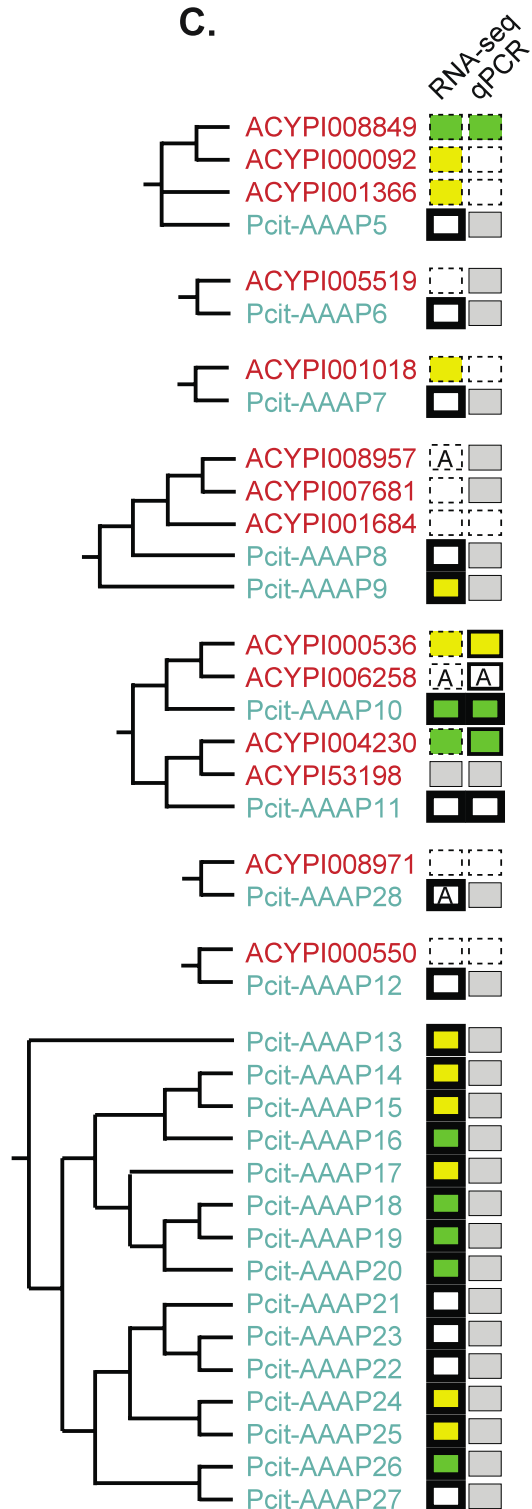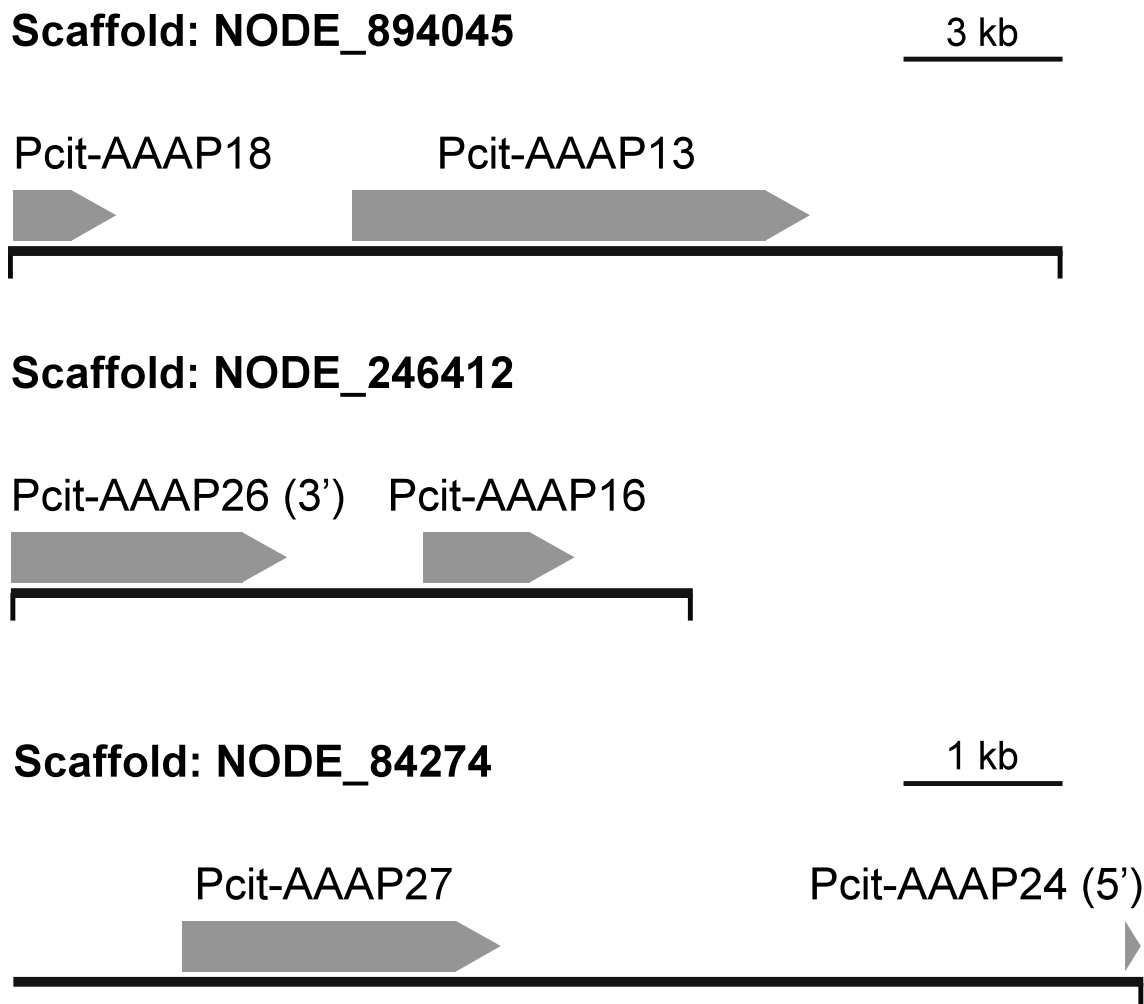
Figure 3.4. Paralogs in mealybug-specific AAAP expansion are tandemly arrayed in the genome. Schematic illustrating the arrangement of mealybug AAAP paralogs along genomic scaffolds. Gray arrows depict the position and 5'-3' direction of representative transcripts (including introns) along four mealybug genomic scaffolds. Each row represents a different scaffold. The top two scaffolds are depicted at the same scale (upper scale bar) and the bottom scaffold is depicted at a different scale (bottom scale bar).

Table 3.1. Hemipteran taxa and associated symbionts

| Taxon | Diet | Obligate symbiont(s) | Symbiont classification | Symbiont localization | Symbiosomal membrane |
|---|---|---|---|---|---|
| **Sternorrhyncha** | | | | | |
| Pea aphid | Phloem sap | *Buchnera aphidicola*[1] | γ proteobacteria | Bacteriocytes | Yes |
| Citrus mealybug | Phloem sap | *Tremblaya princeps*[2] | β proteobacteria | Bacteriocytes | Yes |
| | | *Moranella endobia*[3] | γ proteobacteria | Nested within *Tremblaya*[4] | Not applicable |
| Whitefly | Phloem sap | *Portiera aleyrodidarum*[5] | γ proteobacteria | Bacteriocytes | Yes |
| Potato psyllid | Phloem sap | *Carsonella ruddii*[6] | γ proteobacteria | Bacteriocytes | Yes |
| **Auchenorrhyncha** | | | | | |
| Cicada | Xylem sap | *Sulcia muelleri*[7] | Bacteroidetes | Bacteriocytes | Yes |
| | | *Hodgkinia cicadicola*[8] | α proteobacteria | Bacteriocytes | |
| **Heteroptera** | | | | | |
| Kissing bug | Vertebrate blood | *Rhodococcus rhodnii*[9] | Actinobacteria | Gut lumen | No |

References: [1] (Munson et al. 1991); [2] (Thao et al. 2002); [3] (McCutcheon & Dohlen 2011); [4] (Dohlen et al. 2001); [5] (Thao & Baumann 2004); [6] (Thao et al. 2000), [7] (Moran et al. 2005); [8] (McCutcheon et al. 2009); [9] (Goodfellow & Alderson 1977)

Table 3.2. Amino acid transporters in sampled insects

| | APC Loci | AAAP Loci | Total |
|---|---|---|---|
| Pea aphid | $18^1$ | $22^1$ | 40 |
| Citrus mealybug | $10^2$ | $28^2$ | 38 |
| Whitefly | $12^3$ | $24^3$ | 36 |
| Potato psyllid | $13^3$ | $25^3$ | 38 |
| Cicada | $10^3$ | $16^3$ | 26 |
| Kissing bug | $7^2$ | $13^2$ | 20 |
| Human body louse | $8^1$ | $13^1$ | 21 |
| Fruit fly | $10^1$ | $17^1$ | 27 |

[1] Distinct loci confirmed by mapping transcripts to genomic scaffolds (Price et al. 2011)
[2] Distinct loci confirmed by mapping transcripts to genomic scaffolds (this study)
[3] Estimated number of loci based on the rate of synonymous substitutions (dS) between paralogs being greater than 0.25

**Chapter 4**

Gene family expansions in aphids maintained by endosymbiotic and non-symbiotic traits[3]

**Summary**

Facilitating the evolution of new gene functions, gene duplication is a major mechanism driving evolutionary innovation. Gene family expansions relevant to host/symbiont interactions are increasingly being discovered in eukaryotes that host endosymbiotic microbes. Such discoveries entice speculation that gene duplication facilitates the evolution of novel, endosymbiotic relationships. Here, using a comparative transcriptomic approach combined with differential gene expression analysis, I investigate the importance of endosymbiosis in retention of amino acid transporter paralogs in aphid genomes. To pinpoint the timing of amino acid transporter duplications I inferred gene phylogenies for five aphid species and three outgroups. I found that while some duplications arose in the aphid common ancestor concurrent with endosymbiont acquisition, others predate aphid divergence from related insects without intracellular symbionts, and still others appeared during aphid diversification. Interestingly, several aphid-specific paralogs have conserved enriched expression in bacteriocytes, the insect cells that host primary symbionts. Conserved bacteriocyte enrichment suggests that the transporters were recruited to the aphid/endosymbiont interface in the aphid common ancestor, consistent with a role for gene duplication in facilitating the evolution of endosymbiosis in aphids. In contrast, the temporal variability of amino acid transporter duplication indicate that endosymbiosis is not the only trait driving selection for retention

---

of amino acid paralogs in sap-feeding insects. This study cautions against simplistic interpretations of the role of gene family expansion in the evolution of novel host/symbiont interactions by further highlighting that multiple complex factors maintain gene family paralogs in the genomes of eukaryotes that host endosymbiotic microbes.

**Background**

Gene family expansion is a key player in the evolution of innovation (Arnegard et al. 2010, Deng et al. 2010, Kondrashov 2012, Ohno 1970, Voordeckers et al. 2012), but elucidating the factors maintaining paralogs in a genome can be tricky (Innan & Kondrashov 2010). The vast majority of paralogs that arise in a population are lost through stochastic birth and death processes (Lynch & Conery 2000), presenting an evolutionary conundrum when ancient paralogs are found. A complete understanding of paralog retention requires detailed knowledge of a number of traits, such as paralog expression and function, and the ecological and molecular mechanisms driving paralog diversification. While expression, function, ecology, and molecular traits facilitate explaining the maintenance of anciently acquired paralogs, obtaining data about all these traits is often intractable. In these cases, a combination of natural history and phylogeny of the taxa of interest can provide a framework for proposing the ecological and biological factors that underlie paralog retention.

One trait that may drive paralog maintenance is the evolution of endosymbiosis between multicellular eukaryotes and microorganisms. In endosymbiosis, microbial symbionts reside intracellularly within hosts, commonly providing a nutritional benefit. New gene paralogs may facilitate the establishment and maintenance of a symbiotic

partnership through expression diversification (e.g. from a non-symbiotic tissue to a symbiotic tissue) or possibly through functional diversification (e.g. towards a function specialized for endosymbiosis). Indeed, genomic studies suggest a role for gene duplication in endosymbiosis – these studies have found lineage-specific duplication in genes functionally relevant to symbiotic interactions across systems as divergent as the legume/*Rhizobia* endosymbiosis (Young et al. 2011), the endosymbiosis between corals or anenomes and *Symbiodinium* (Baumgarten et al. 2015, Shinzato et al. 2011), and the endosymbiosis between the pea aphid *Acyrthosiphon pisum* and its endosymbiont *Buchnera aphidicola* (Huerta-Cepas et al. 2010, Price et al. 2011). *A. pisum*, a sap-feeding insect, has undergone expansions in over 2000 gene families (Huerta-Cepas et al. 2010, International Aphid Genomics Consortium 2010), some of which may be mechanistically important for its relationship with *Buchnera*. Particularly intriguing in *A. pisum* are lineage-specific expansions in amino acid transporter genes (Dahan et al. 2015, Duncan et al. 2014, Price et al. 2011, Wilson & Duncan 2015) – genes whose membrane-bound protein products are crucial for nutritional exchange between *A. pisum* and *Buchnera* (Price et al. 2014, 2015). Indeed, amino acid transporters are over-represented among *A. pisum* gene families that underwent large expansions resulting in more than 10 paralogs (Huerta-Cepas et al. 2010). Similarly, I recently discovered that several sap-feeding insects (also with endosymbionts) experienced lineage-specific expansions in amino acid transporters (Duncan et al. 2014). Interestingly, independent expansion of amino acid transporters in multiple sap-feeding insect lineages is a pattern that parallels other independently evolved signatures of host/endosymbiont genome co-evolution (Wilson & Duncan 2015).

The mechanistic importance of gene duplication in amino acid transporters for endosymbiosis in these insects is further supported by the observation that some lineage-specific paralogs in *A. pisum* and the citrus mealybug *Planococcus citri* are enriched in bacteriocytes (Duncan et al. 2014, Price et al. 2011), the specialized insect cells where symbionts reside.  Bacteriocyte enrichment of some paralogs implies paralog evolution to operate in a symbiotic context because bacteriocytes represent the interface between the insect host and the endosymbiont.  This host/endosymbiont interface is made up of three membrane barriers separating host tissues from endosymbionts: (1) The plasma membrane surrounding the bacteriocyte, (2) the insect-derived membrane surrounding individual symbiont cells (symbiosomal membrane), and (3) the inner and outer bacterial membranes of each symbiont cell.  While the symbiosomal membrane is the most immediate interface between host (bacteriocyte cytoplasm) and endosymbiont, the bacteriocyte plasma membrane is also an important part of the host/endosymbiont interface because of the role it plays in regulating metabolic output of the symbiont (Price et al. 2014).  In addition to independent recruitment of amino acid transporter paralogs to the host/symbiont interface of both *A. pisum* and *P. citri*, tests for signatures of selection in the pea aphid found, in one expansion, an elevated rate of evolution in the transition from high gut expression to enriched bacteriocyte expression.  The elevated rate of evolution suggests functional evolution corresponding to a shift towards symbiotic expression (Price et al. 2011).  Lastly, expansions in sap-feeding insects are significantly associated with increased gene duplication rates and decreased gene loss rates (Dahan et al. 2015), supporting an adaptive explanation for the retention of duplicate amino acid

transporters – an explanation that may relate to shared traits among these insects, such as endosymbiosis with nutrient-provisioning bacteria.

Despite the evidence supporting endosymbiosis as a factor influencing the retention of amino acid transporter paralogs in sap-feeding insects, transcriptomic and expression data indicate that other biological factors are also at play. For example, in aphids, some paralogs show biased expression in males (Duncan et al. 2011), where symbionts are less abundant than in females (Douglas 1989). In fact, accelerated rates of evolution also correlate with the evolution of male-biased expression in one male-biased paralog, suggesting a derived sex-biased function. Further, despite bacteriocyte enrichment in some citrus mealybug and pea aphid paralogs, most paralogs are not enriched in bacteriocytes (Duncan et al. 2014, Price et al. 2011), suggesting that other features of these insects influence paralog retention. Lastly, while lineage-specific duplications in amino acid transporters occur in four sap-feeding insects – pea aphids (Price et al. 2011), citrus mealybugs, potato psyllids, and whiteflies (Duncan et al. 2014)– I found no evidence for duplication in the sap-feeding cicada (Duncan et al. 2014), indicating that gene duplication in amino acid transporters is not necessary for the evolution of endosymbiosis between sap-feeding insects and bacteria. These data do not rule out the possibility that gene duplication has played an important role in the evolution and maintenance of endosymbiosis in some sap-feeding insects. Even so, evidence that other factors influence paralog retention makes it unclear if endosymbiosis plays a primary or secondary role in selection for the maintenance of amino acid transporter paralogs.

Here, using a comparative transcriptomic approach, I leverage the phylogeny and natural history of aphids and their close relative, the grape phylloxera *Daktulosphaira vitifoliae* (see Table 4.1 and Figure 4.1 for taxonomic classifications and relationships among taxa in this study), to investigate life history traits underlying the retention of amino acid transporter paralogs. Aphids and phylloxera belong to different families of the hemipteran group Aphidomorpha – Aphididae and Phylloxeridae, respectively. The common ancestor of Aphididae established an endosymbiotic relationship with the bacterium *Buchnera aphidicola* around 160-280 MYA (Moran et al. 1993), and since that initial infection, *Buchnera* has been vertically inherited by nearly all extant aphids. In contrast, Phylloxeridae, including *D. vitifoliae*, lacks *Buchnera* or other intracellular symbionts (Medina et al. 2011, Vorwerk et al. 2007). If endosymbiosis between aphids and *Buchnera* initially drove paralog retention in amino acid transporters, I expect to find that duplication took place, at least initially, in the aphid common ancestor. If, however, duplication predates the aphid/phylloxera split or postdates aphid diversification, then the initial trait influencing paralog maintenance is more likely not endosymbiosis with *Buchnera*, but another trait that evolved concurrently with gene duplication.

**Methods**

**Taxon sampling, specimen collection, identification, and vouchering**

I sampled amino acid transporters from several aphid genera representing 3 subfamilies and multiple tribes across the Aphididae (Table 4.1, Figure 4.1). Aphid species included *Acyrthosiphon pisum*, *Myzus persicae*, *Aphis nerii*, *Tamalia coweni*, and *Pemphigus obesinymphae*. I consider these taxa as representative of aphids because they

include tribes at a range of positions across the aphid phylogeny, including members of

Pemphigini, which are usually supported as sister to the rest of aphids (Nováková et al.

2013). Additionally, I sampled amino acid transporters from an outgroup of aphids, the

grape phylloxera *Daktulosphaira vitifoliae*, and two Aphidoidea outgroups (*Pediculus*

*humanus* and *Drosophila melanogaster*; Figure 4.1).

M. persicae* and *A. pisum* data were generated from established isofemale

laboratory lines. *M. persicae* RNA-seq and differential expression data, generated in this

study by Honglin Feng, came from laboratory clones G006, G002, and BTI Red (also

known as USDA) (Ramsey et al. 2007). Data for *A. pisum*, generated in previously

published studies using RNA-seq and qRT-PCR (Duncan et al. 2014, Hansen & Moran

2011, Macdonald et al. 2012, Price et al. 2011), came from laboratory clones LSR1

(Caillaud et al. 2002), 9-2-1 (Russell & Moran 2006), 5A (Sandström & Moran 2001)

and CWR09/18 (Macdonald et al. 2012). *A. nerii* were collected by myself and

colleagues from *Asclepias spp.* in Miami, Florida (Miami-Dade county), Atlanta, Georgia

(Dekalb county), and Minnesota. The Atlanta and Miami populations are maintained as

isofemale clones in the laboratory of Patrick Abbot at Vanderbilt University. *A. nerii*

were identified based on host plant and distinctive morphology. *T. coweni* were collected

by Miller et al. (2015) from various sites and host plants in Arizona, Nevada and

California. *P. obesinymphae* were collected in the vicinity of Nashville, Tennessee

(Davidson county) on *Populus deltoides* subsp. *deltoides* Pantrick Abbot, who identified

them by based on distinctive gall morphology. *D. vitifoliae* were collected at the

vineyards of Château Couhins in Bordeaux, France on *Vitis vinifera cv. Cabernet franc*

by colleagues at the French National Institute for Agricultural Research (INRA), who

also identified them.  An isofemale clone of *D. vitifoliae* (INRA-Pcf7) is maintained at

INRA.  Voucher specimens for *T. coweni* are deposited in the Smithsonian Institution

Department of Entomology, the Canadian National Collection of Insects, and collections

at Washington State University and California State University, Chico (Miller et al. 2015).

Additionally, I annotated transcripts corresponding to *cytochrome c cxidase subunit 1*

(*CO1*) for *A. nerii*, *T. coweni*, *P. obesinymphae*, and *D. vitifoliae* to serve as identity

vouchers.  Briefly, I used a protein sequence for *A. pisum CO1* (Genbank ID

YP_002323931.1) as a query in local TBLASTN searches against transcriptomes for the

other aphids and *D. vitifoliae*.  Top hits (all with E = 0.0) were used as queries in

reciprocal BLASTX searches against the NCBI refseq protein database to confirm

homology with *A. pisum CO1*.  Reciprocal BLAST searches returned one *CO1* sequence

for each species except *A. nerii*, which had two *CO1* sequences.  Sequences are provided

in Supplemental file 1 from Duncan et al. (2016).


**Transcriptome sequencing and assembly**

Transcriptomes were sequenced for *A. nerii* and *M. persicae*.  Total RNA was

extracted by colleagues Stephanie Chiang and Honglin Feng from whole adult, asexual

female *A. nerii* bodies and a combination of whole bodies, bacteriocyte, and gut for adult,

asexual female *M. persicae*.  Total RNA was sent to the Hussman Institute for Human

Genomics (University of Miami Miller School of Medicine) for library preparation and

paired end sequencing on the Illumina HiSeq platform.  Raw RNA-seq reads for *A. nerii*

and *M. persicae* were deposited in the NCBI Sequence Read Archive (SRA) under

BioProject PRJNA296778.  Raw reads for *T. coweni* (a combination of paired end and

single end reads) and *P. obesinymphae* (single end reads) were provided by Patrick Abbot, and are available on GenBank in the SRA (*T. coweni* accession numbers: SRX1305377, SRX1305445, SRX1305282, SRX1304838 (Miller et al. 2015); *P. obeiynymphae* reads are stored under BioProject PRJNA301746). Reference transcriptomes were assembled for *T. coweni*, *P. obesinymphae*, *A. nerii* and the *M. persicae* G006 clone. Reads for these taxa were filtered to a minimum quality score of 30 over 95% of the read, resulting in a combination of paired end and single end reads for *A. nerii* and G006. All reads from each taxon kept after the filtering process were assembled into a single reference transcriptome for each species in Trinity (7/17/14 release) (Haas et al. 2013) using the Blacklight system at the Pittsburgh Supercomputing Center. A fully assembled transcriptome for *D. vitifoliae* was provided by colleagues at INRA, for which raw sequencing reads are accessible via the NCBI SRA (BioProject PRJNA294954). The other three taxa in my data set – *A. pisum*, *P. humanus,* and *D. melanogaster* – have fully sequenced genomes, from which I used amino acid transporter sequences that were annotated in a previous study (Price et al. 2011).

**Amino acid transporter annotation**

With the help of University of Miami undergraduate Doug Nguyen and my colleague Honglin Feng, I used transcriptomes to annotate amino acid transporters in the Amino Acid-Polyamine-Organocation (APC) (TC # 2.A.3) and Amino Acid-Auxin-Permease (AAAP) (TC # 2.A.18) families from *T. coweni, P. obesinymphae*, *M. persicae*, *A. nerii*, and *D. vitifoliaeas* as previously described (Duncan et al. 2014, Price et al. 2011). Briefly, I used a stand-alone PERL script underlying the ORF prediction webserver

hosted at http://bioinformatics.ysu.edu/tools/OrfPredictor.html, where transcripts were

translated into the six reading frames. The translated transcripts were searched for

functional domains that significantly matched (E ≤ 0.001) known APC, and AAAP

families in HMMER v3.0 (Eddy 2009, Finn et al. 2011). HMMER hits were verified

through BLAST searches to the NCBI refseq protein database. Transcripts with

significant similarity (E ≤ 0.001) to APC or AAAP sequences from *D. melanogaster* or *A.*

*pisum* were selected for further computational processing.

Transcriptomes generated many unique but similar transcripts identified as APC

or AAAP members through HMMER and BLAST. I collapsed amino acid transporter

transcripts into conservative sets of representative loci for each taxon using methods I

developed previously in Chapter 3 of this dissertation and published as Duncan et al.

(2014). For *M. persicae*, which has a draft genome sequence, I collapsed all transcripts

that mapped to the same location in the genome. For remaining taxa, I followed a series

of steps. First, I collapsed all transcripts with the same Trinity component number into

the longest representative transcript. Next, I followed the methods previously described

(Duncan et al. 2014). Briefly, I estimated the pairwise rate of synonymous substitutions

(Ks) among transcripts that clustered together in preliminary phylogenetic analyses using

the Goldman Yang method (Goldman & Yang 1994) in KaKs_Calculator (v.1.2) (Zhang

et al. 2006). I collapsed all transcripts with a pairwise Ks value < 0.25, keeping the

longest sequence to represent the locus. If two similar transcripts met the cutoff Ks of ≥

0.25, but overlapped less than fifty base pairs, I collapsed the shorter transcript into the

longer transcript to validate a conservative estimation of locus number. I chose the

threshold Ks value (0.25) because I previously found that it slightly underestimates the

number of true amino acid transporter paralogs in *A. pisum,* collapsing only three very recently duplicated APC paralogs (Duncan et al. 2014). Thus, this threshold is appropriate for conservative estimation of amino acid transporter paralogs in related species.

**Differential expression analysis**

The differential expression of *Myzus persicase* amino acid transporters between bacteriocyte and whole body tissues was quantified by Honglin Feng using the transcriptome data from this study. *M. persicae* clones G006, G002, and BTIRed, were treated as replicates. Differential expression analysis was conducted with the RSEM package (v.1.2.22) (Li & Dewey 2011) and edgeR (v.3.10.2) from the Bioconductor package (Robinson et al. 2010). Briefly, the processed forward RNA-seq reads for all three *M. persicae* clones were mapped to reference transcripts in a strand-specific manner using bowtie2 (v.2.2.4) (Langmead & Salzberg 2012) and mapped reads were counted using PERL script rsem-calculate-expression.pl from the RSEM package. The counts from RSEM were scaled to the whole transcriptomes and normalized by relative log expression (RLE). The significantly differentially expressed amino acid transporters between bacteriocyte and whole insect were identified using edgeR negative binomial models. Four amino acid transporter sequences comprised several truncated, partial transcripts that supported full-length gene models in the *M. persicae* genome (*Mper-APC09*, *Mper-APC12*, *Mper-AAAP06*, and *Mper-AAAP20*; the *M. persicae* draft genome assembly is available at aphidbase.com). In these four cases, the differential expression

analysis was performed by mapping raw RNA-seq reads to the gene models instead of the transcripts.

**Phylogenetic analysis**

Transcript sequences were translated to protein using SeaView (v.4.5.2) (Gouy et al. 2010), and protein sequences were aligned using MAFFT (v.7.158b) (Katoh et al. 2002) using default parameters. Alignments were trimmed in TRIMAL (v.1.4) (Capella-Gutiérrez et al. 2009) using a gap threshold of 25%. ProtTest (v.3.4) (Abascal et al. 2005) determined the best-fit model of protein evolution to be either LG+G (APC family) or LG+I+G (AAAP) based on the Akaike Information Criterion. Maximum-likelihood (ML) phylogenies were inferred for the APC and AAAP families in RAxML (v.8.0.26) (Random Axelerated Maximum Likelihood) (Ott et al. 2007, Stamatakis 2006) using the best-fit model of protein evolution and the fast bootstrap option. Bootstrap replicate number was chosen by the bootstrap convergence criterion 'autofc'.

I further inferred Bayesian phylogenies in MrBayes (v.3.2) (Ronquist & Huelsenbeck 2003) using WAG+G (APC family) or WAG+I+G (AAAP), as MrBayes does not implement the LG amino acid substitution model. Two independent runs, each with four chains, were run for one to five million generations, until the standard deviation of split frequencies between runs converged to below 0.01. Appropriate parameter sampling and convergence were determined by visually inspecting trace files in Tracer (v.1.6) (Rambaut et al. 2014). Tracer was also used to determine burn-in values of each data set (10% of generations), which I discarded when constructing Bayesian consensus trees. In the figures presented here, I mapped ML bootstrap support onto Bayesian

consensus trees using SumTrees v3.0 from the DendroPy package (Sukumaran & Holder 2010).

Although I annotated all amino acid transporters in the AAAP family, because of large divergence in that family (Duncan et al. 2014), I inferred the relationships among a reduced set of sequences corresponding to the "arthropod expanded clade". The arthropod expanded clade consists of arthropod orthologs to the mammalian SLC36 family of proton-coupled amino acid transporters (Price et al. 2011, Thwaites & Anderson 2011). Two human SLC36 sequences (*SLC36A1*, *SLC36A2*), previously shown to belong to the sister clade of the arthropod expanded clade (Price et al. 2011), were used as outgroups. Outgroup sequences for the APC family are members of the sister clade of Na-K-Cl transporters (*ACYPI001649*, *ACYPI007138*) (Price et al. 2011). Untrimmed transcript sequences translated into protein as well as trimmed Bayesian and ML alignments of the APC family and reduced AAAP family ("arthropod expanded clade") are provided as Supplemental files 2-5 from Duncan et al. (2016).

**Results**

**The *slimfast* expansion predates aphid-phylloxera divergence**

Previously, Price et al. (2011) identified an aphid-specific expansion in the APC transporter *slimfast* (named for the *Drosophila melanogaster* ortholog, CG11128 (Colombani et al. 2003)). Here, I find APC members of different Aphidomorpha (Table 4.1, Figure 4.1) interleaved within four subclades of the *slimfast* expansion. These data imply that *slimfast* expanded into minimally four paralogs before the divergence of aphids and the grape phylloxera, *Daktulosphaira vitifoliae* (Figure 4.2).

**Amino acid transporters duplicated at multiple time scales in aphid evolution**

*APC family*

Following the aphid/phylloxera split, additional gene duplication events were inferred in the *slimfast* expansion at four different taxonomic levels within aphids: (1) predating aphid diversification (Figure 4.2, gray boxes labeled A), (2) predating diversification of the subfamily Aphidinae (Figure 4.2, gray box labeled B), (3) predating diversification of the tribe Macrosiphini (Figure 4.2, gray boxes labeled C), as well as (4) at least one duplication event specific to *Acyrthosiphon pisum* (Figure 4.2, gray box labeled D).  Additionally, the phylogenetic analysis supports one *D. vitifoliae*-specific gene duplication following aphid/phylloxera divergence (Figure 4.2, gray box labeled D1).  Species varied in the number of *slimfast* paralogs they encoded, ranging from four in *Tamalia coweni* to 10 in *A. pisum*.  Variation in number of *slimfast* paralogs resulted at least in part from differences in additional gene duplications during aphid diversification, but may also indicate gene losses in some species or under-sampling.

*AAAP family*

In contrast to the APC *slimfast* expansion, no AAAP duplications predated the phylloxera/aphid divergence.  However, mirroring other duplication patterns I found in the APC *slimfast* expansion, duplication in the AAAP family happened at different time scales during aphid evolution (Figure 4.3).  I found strong support for several duplications predating aphid diversification (figure 4.3, gray boxes labeled A), as well as support for duplications specific to the tribe Macrosiphini (Figure 4.3, gray boxes labeled C).  Asterisks beside gene IDs in Figures 4.2 and 4.3 mark genes that may be products of

species-specific gene duplication. However, given our tree topology, I am unable to confidently infer that these sequences result from species-specific duplication.

**Bacteriocyte expression is dynamic among aphids**

Differential expression analyses identified 25 significantly differentially expressed amino acid transporters in the *M. persicae* transcriptome (based on a 2-fold difference in either direction and P ≤ 0.05, FDR ≤ 0.05; supplementary table 1 from Duncan et al. (2016)). Differentially expressed amino acid transporters included 11 bacteriocyte-enriched transcripts and 14 bacteriocyte-depleted transcripts. Of these differentially expressed amino acid transporters, ten bacteriocyte-enriched and twelve bacteriocyte-depleted transcripts were members of the APC family or the reduced AAAP family that appear in phylogenetic analyses (Figure 4.4).

Differential expression results for amino acid transporters in *M. persicae* were qualitatively compared with previously published expression results for orthologous amino acid transporters in *A. pisum* (Duncan et al. 2014, Hansen & Moran 2011, Macdonald et al. 2012, Price et al. 2011). Bacteriocyte expression was not necessarily consistent between *A. pisum* and *M. persicae*, evident by three patterns displayed by orthologous pairs: (1) Both have the same relative bacteriocyte expression (enriched, depleted or not significantly different from whole insect), (2) one species has enriched or depleted bacteriocyte expression while its ortholog is not significantly differentially expressed between whole insect and bacteriocyte, or (3) one species has bacteriocyte-enriched expression while the other has bacteriocyte-depleted expression (Figure 4.4). In the APC family, three pairs of *A. pisum*/*M. persicae* orthologs showed conserved

bacteriocyte enrichment: *ACYPI008904/Mper-APC08*, *ACYPI005118/Mper-APC10*, and

*ACYPI003130/Mper-APC04*.  Notably, the first two pairs are members of the *slimfast*

expansion (Figures 4.2 and 4.4).  In the AAAP family, two orthologous pairs show

conserved bacteriocyte enrichment: *ACYPI000536/Mper-AAAP11* and

*ACYPI001366/Mper-AAAP03*, both of which are members of aphid-specific expansions

(Figures 4.3 and 4.4).


**Discussion**

Recent studies on symbiotic organisms support a role for gene duplication in the

evolution of endosymbiosis (Baumgarten et al. 2015, Dahan et al. 2015, Duncan et al.

2014, Price et al. 2011, Shinzato et al. 2011, Young et al. 2011).  However, most studies

do not address the possibility of a gene duplication/endosymbiosis connection in more

than one species and, as I have found in sap-feeding insects, non-symbiotic traits may

influence the retention of duplicate genes (Duncan et al. 2011, 2014).  To gain insight

into the role of endosymbiosis in retaining aphid-specific amino acid transporter paralogs,

I used a comparative transcriptomic approach to pinpoint the timing of amino acid

transporter duplications in the Aphidomorpha (Table 4.1, Figure 4.1).  My results support

a complex and dynamic evolutionary history of amino acid transporters in these insects –

a history that was likely shaped by multiple biological and ecological factors.  In support

of a role for gene duplication in the evolution of endosymbiosis, I inferred several

duplication events in the aphid common ancestor, corresponding to the acquisition of the

aphid endosymbiont, *Buchnera* (Figures 4.1-4.3).  However, I also inferred duplication

events both earlier and later than the evolution of endosymbiosis in aphids.  Duplication

at these earlier and later time scales implies that gene duplication and retention has also been driven by factors other than endosymbiosis.

**The *slimfast* expansion was not driven by the evolution of endosymbiosis**

I posit that the aphid/phylloxera *slimfast* expansion, at least initially, was not driven by endosymbiosis. Phylloxera lack an endosymbiont, and assuming that the shared ancestor of aphids and phylloxera also lacked an endosymbiont, the expansion predated the evolution of endosymbiosis. However, the relationship of the aphid/phylloxera lineage, derived from within Sternorrhyncha, suggests that endosymbiosis originated in the common ancestor of Sternorrhyncha and was secondarily lost in phylloxera. Even if endosymbiosis originated in the sternorrhynchan common ancestor, the *slimfast* expansion likely was not driven primarily by primary endosymbiosis because the expansion postdates aphid divergence from other major sternorrhynchan lineages (Duncan et al. 2014). Furthermore, despite their shared ancestry, sternorrhynchan lineages may have evolved endosymbiosis independently, supported by ongoing discoveries of convergent patterns of host/symbiont genome coevolution (reviewed by Wilson and Duncan (2015)). Given the timing of the *slimfast* expansion, its origin is more likely influenced by a trait shared by aphids and phylloxera, such as their complex life cycle that involves both sexual and asexual reproduction (Blackman & Eastop 2000, Forneck & Huber 2009). Indeed, I previously reported on male-biased and asexual female-biased *slimfast* paralogs in aphids in Chapter 2 of this dissertation (and published as Duncan et al. (2011)), supporting the notion that *slimfast* paralogs were retained as a result of selection for divergence to fulfill sex-specific roles.

**Dynamic evolution of amino acid transporters in aphids**

*APC family*

Despite evidence that the *slimfast* expansion was driven by a shared trait between aphids and phylloxera, retention of *slimfast* paralogs in aphids was most likely influenced by more complex and dynamic factors and processes. For example, gene duplication within the *slimfast* expansion continued after aphids and phylloxera diverged (Figure 4.2), implying that (1) additional selective pressures, perhaps shifting from ancestral selective pressures, influenced the retention of additional *slimfast* paralogs as they emerged, (2) additional *slimfast* paralogs emerged by chance and were retained through non-adaptive processes (e.g. through the classic model of subfunctionalization known as Duplication, Degeneration, Complementation (Force et al. 1999)), or (3) aphids are particularly prone to gene duplication. Notably, these three possibilities are not mutually exclusive and could all be operating.

Another important aspect of paralog evolution in the *slimfast* expansion is highlighted by two pairs of *A. pisum*/*M. persicae* paralogs that have conserved bacteriocyte enrichment: *ACYPI008904*/*Mper-APC08* and *ACYPI005118*/*Mper-APC10*. A change in expression in these two orthologous pairs from the typical gut expression of *slimfast* (Price et al. 2011) towards bacteriocyte enrichment implies that these *slimfast* paralogs were recruited to the aphid/*Buchnera* symbiotic interface and retained for a role in endosymbiosis. In addition to being recruited to bacteriocytes, *ACYPI008904* and *Mper-APC08* may have diverged functionally from other members of their subclade (Figure 4.2). Subclade 3 of the *slimfast* expansion experienced gene duplication in the common ancestor of Macrosiphini, resulting in multiple orthologous paralogs for *A.*

*pisum* and *M. persicae* while the other three aphid species each have only one gene. Gene duplication in subclade 3 provides an opportunity for functional divergence among paralogs while still fulfilling their ancestral function in two ways: (1) by evolving novel function and/or expression (neofunctionalization) or (2) by partitioning, and possibly optimizing or specializing, ancestral functions in sister paralogs (subfunctionalization). Indeed, supporting divergence following gene duplication in subclade 3, while *ACYPI008904* and *Mper-APC08* are both highly enriched in bacteriocytes, the closely related paralogs *ACYPI002633* and *Mper-APC09* are both bacteriocyte-depleted (Price et al. 2011) (Figure 4.3). In fact, previous work in *A. pisum* revealed that *ACYPI008904* and *ACYPI002633* have very different expression profiles – while *ACYPI008904* is enriched in bacteriocytes (Price et al. 2011) as well as asexual adult females (Duncan et al. 2011), *ACYPI002633* is enriched in both gut (Price et al. 2011) and adult male *A. pisum* (Duncan et al. 2011). Additionally, the branches leading to both *ACYPI002633* and the clade containing *ACYPI008904* and *ACYPI008323* experienced accelerated rates of evolution (Duncan et al. 2011, Price et al. 2011), supporting the possibility that gene duplication was followed by both expression and functional divergence.

Without expression data for the other three aphid species, I cannot infer if the different expression profiles of the *A. pisum* and *M. persicae* APC paralogs result from neofunctionalization or subfunctionalization, both of which have implications for the role of endosymbiosis in the evolution of amino acid transporters. If bacteriocyte expression is novel in *ACYPI008904* and *Mper-APC08*, then amino acid transporter recruitment to bacteriocytes is an ongoing process in aphid evolution. In contrast, if single-copy aphid orthologs are also expressed in bacteriocytes, then we could infer that *Buchnera* infection

coincided with recruitment of the ancestral, single-copy gene in subclade 3 to aphid bacteriocytes.  In light of expression data for *M. persicae* (supplemental file 6 from Duncan et al. (2016)) and *A. pisum* (Hansen & Moran 2011, Macdonald et al. 2012, Price et al. 2011), I predict that the ancestral aphid gene of subclade 3 operated at the symbiotic interface, coincident with *Buchnera* infection.  After all, the sister clade, subclade 4, contains the only additional *slimfast* paralogs with conserved bacteriocyte enrichment between *A. pisum* (*ACYPI005118*) and *M. persicae* (*Mper-APC10*).  Thus, parsimony would predict that aphid *slimfast* members in both subclades 3 and 4 inherited bacteriocyte expression from their common ancestor – or, since subclades 3 and 4 appeared before the aphid/phylloxera split, their ancestral gene may have been expressed in the cells that gave rise developmentally to aphid bacteriocytes after the aphids and phylloxera diverged from a common ancestor (Wilson & Duncan 2015).

A scenario in which bacteriocyte expression evolved in single-copy aphid orthologs would imply that the ancestral aphid gene of subclade 3 operated in endosymbiosis in addition to other, previously defined functions in place prior to aphid/phylloxera divergence.  Importantly, there is a precedent for genes with broad expression to play important roles in endosymbiosis.  Indeed, the *A. pisum* AAAP member *ACYPI001018* is both globally highly expressed and also a key regulator of *Buchnera* metabolic output (Price et al. 2014, 2015) – a role that is very possibly conserved in *M. persicae,* given the conserved high bacteriocyte expression of the ortholog *Mper-AAAP05*.

*AAAP family*

Amino acid transporters in the AAAP family, like the *slimfast* paralogs, expanded at different time scales during aphid evolution.  AAAP members resulting from duplication in the aphid common ancestor, including *ACYPI000536/Mper-AAAP11* and *ACYPI001366/Mper-AAAP03*, are prime candidates for facilitating transporter divergence towards a role in endosymbiosis.  The fact that both these pairs of orthologs have conserved bacteriocyte enrichment while related paralogs have different expression profiles (Figure 4.4, Supplemental file 6 from Duncan et al. (2016)) suggests that gene duplications in the aphid common ancestor were followed by functional divergence to symbiotic and non-symbiotic roles.  The phylogeny in figure 4.3 also supports three gene duplications predating the diversification of the tribe Macrosiphini.  Retention of these sets of orthologs is unlikely to have been driven by acquisition of the primary endosymbiont *Buchnera*, given that they lack conservation of differential expression between *A. pisum* and *M. persicae.*

**Gene duplication and endosymbiosis in aphids**

The factors driving gene duplication in the amino acid transporters of Aphidomorpha are complex – a pattern that may also influence our understanding of gene duplications in other host genomes and their role in endosymbiosis.  My data continue to point towards the importance of non-symbiotic traits in driving selection for paralog retention.  At the same time, my data support a role for endosymbiosis in maintaining duplicated amino acid transporters in aphids.  Importantly, I find that the most parsimonious explanation for conserved bacteriocyte-enrichment in orthologous pairs is

that amino acid transporters were recruited to bacteriocytes in the aphid common ancestor, coinciding with acquisition of the primary, obligate endosymbiont, *Buchnera*.

Given the repeated support found for both symbiotic and non-symbiotic roles for gene duplication in sap-feeding insects (Dahan et al. 2015, Duncan et al. 2011, 2014; Price et al. 2011), I caution against drawing strong conclusions about the role endosymbiosis plays in gene duplications found in the genomes of other symbiotic systems. An understanding of the natural history of a focal taxon can indeed help with making predictions about the evolutionary significance of interesting genomic patterns like gene duplication. However, by using a comparative approach, I have found that multiple complex factors maintain paralogs in a group of genes that are functionally critical to host/symbiont interactions. Moving forward, I advocate using a comparative framework with as much information as possible about the expression and function of duplicated genes. Differential expression analysis of symbiotic and non-symbiotic tissues using RNA-seq or qPCR provides a valuable layer of information that contributes to our ability to infer the role of duplicated genes and the factors contributing to their retention in a genome. These genomic and transcriptomic approaches pave the way for the next phase in understanding why genomes evolved their particular architecture through application of functional genomic approaches.
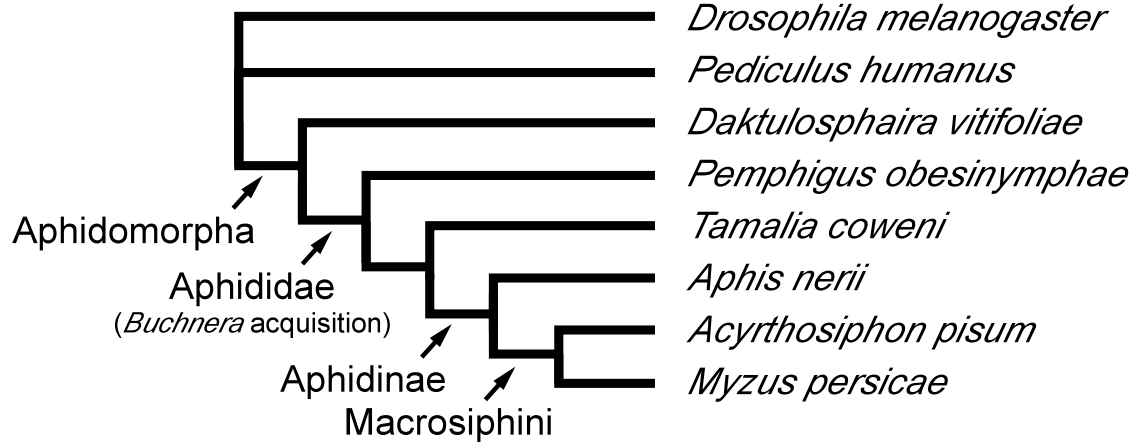
Figure 4.1. Phylogenetic relationships among sampled taxa. Acquisition of the aphid endosymbiont *Buchnera* and relevant higher taxonomic classifications are mapped. Tree structure is based on phylogenetic analyses reported by Novakova et al. (2013), Misof et al. (2014), and Dahan et al. (2015).

Figure 4.2. Bayesian phylogeny of amino acid transporters in the APC family. Branches are color coded by taxon, as indicated in the key. Node support is shown both as branch weight (Bayesian posterior probability) and circles on nodes (Maximum Likelihood bootstrap support). The *slimfast* expansion is marked in the upper right, along with the four subclades resulting from duplication events that predate aphid/phylloxera divergence. Duplication events occurring at different time scales in aphid evolution gave rise to lineage-specific clades highlighted by gray boxes. Letters in gray boxes refer to (A) clades resulting from duplication events predating Aphididae divergence, (B) clades resulting from duplication events predating Aphidinae divergence, (C) clades resulting from duplication events predating Macrosiphini divergence, and (D) species-specific duplication events.
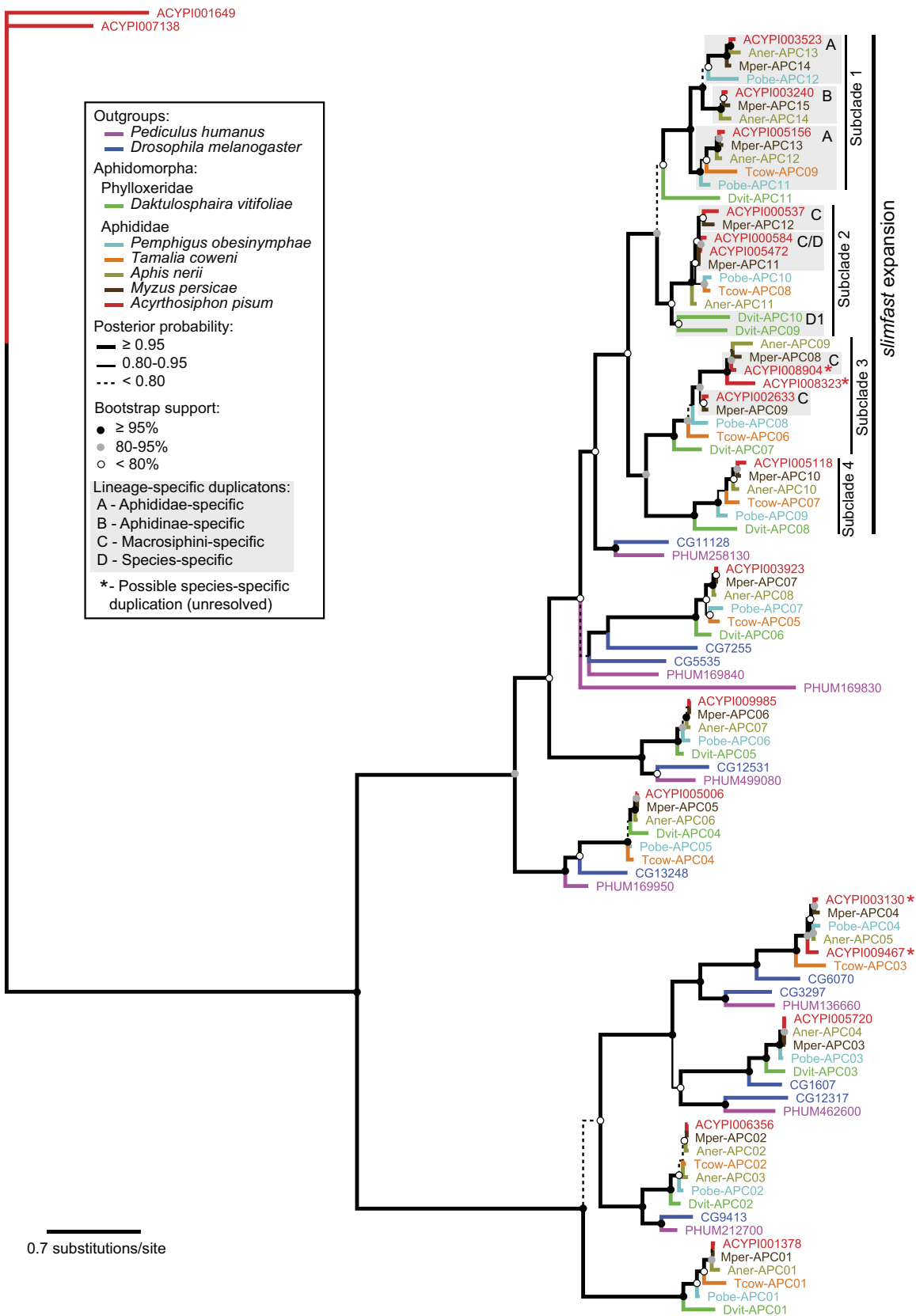
**APC family**



0.7 substitutions/site

Figure 4.3. Bayesian phylogeny of amino acid transporters in the AAAP family. Branches are color coded by taxon, as indicated in the key. Node support is shown both as branch weight (Bayesian posterior probability) and circles on nodes (Maximum Likelihood bootstrap support). Duplication events occurring at different time scales in aphid evolution gave rise to lineage-specific clades that are highlighted by gray boxes. Letters in gray boxes refer to (A) clades resulting from duplication events predating Aphididae divergence, and (C) clades resulting from duplication events predating Macrosiphini divergence. Outgroup sequences (*SLC36A1* and *SLC36A2*) are from humans.
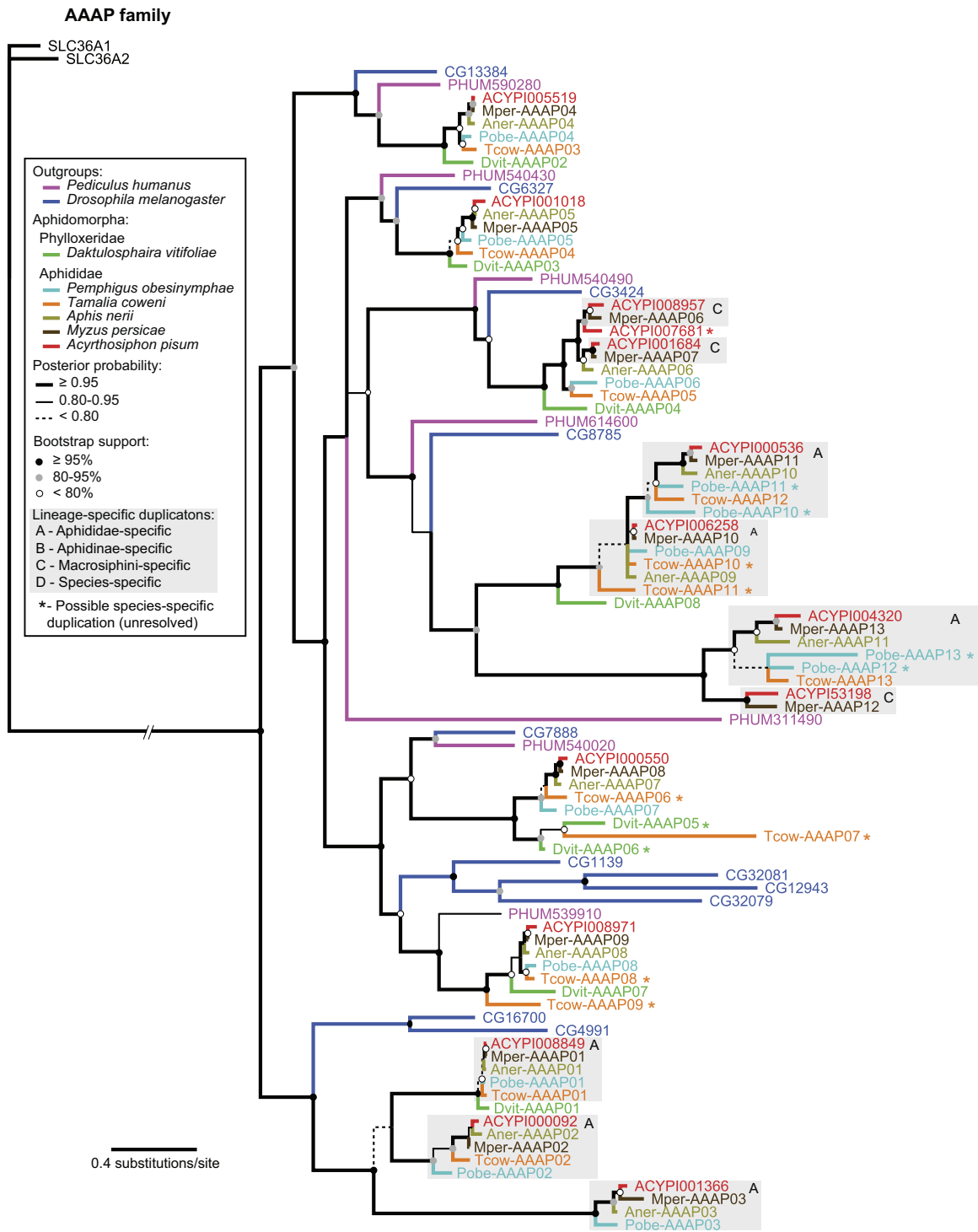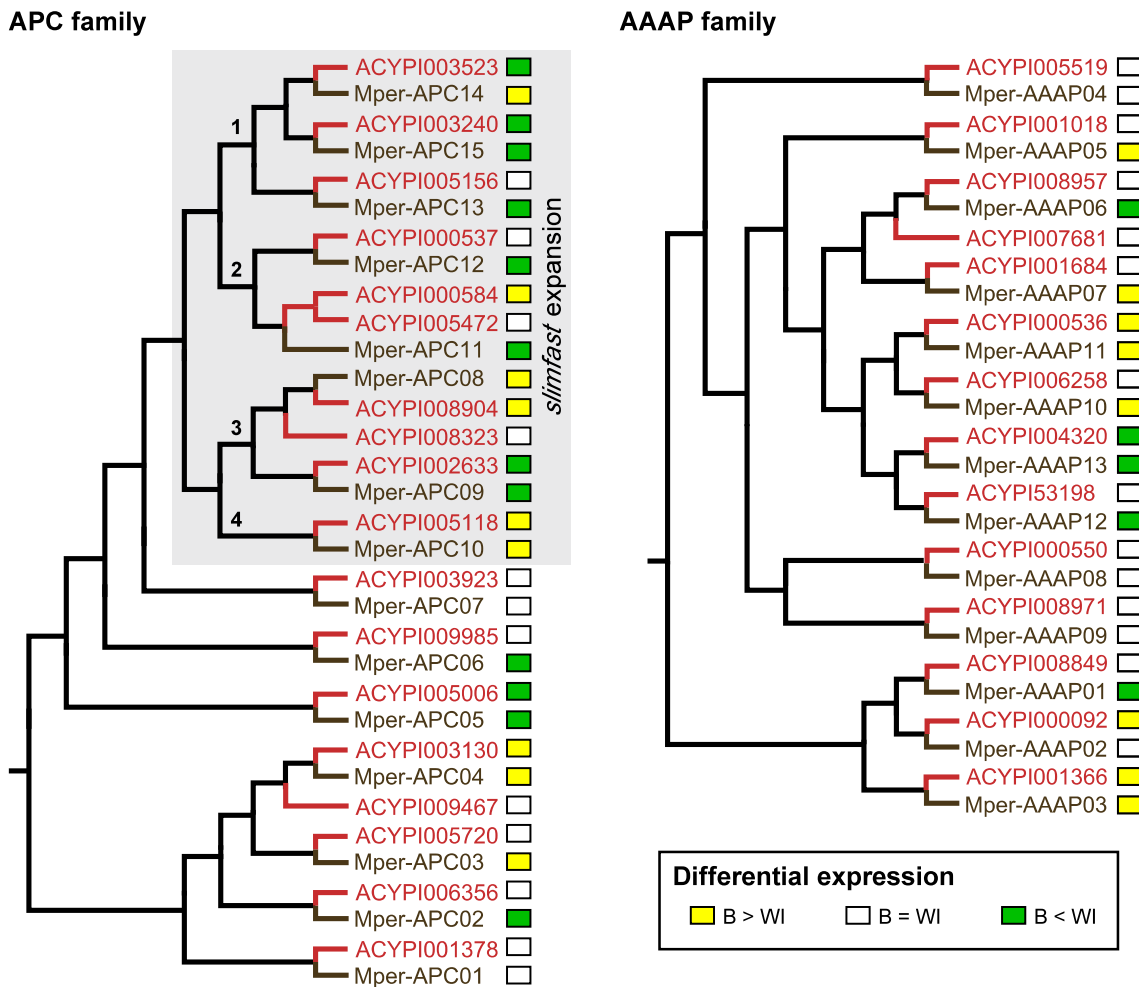
Figure 4.4. Differential expression of amino acid transporters in *A. pisum* and *M. persicae*. Phylogenies depict relationships between *A. pisum* and *M. persicae* amino acid transporters in the APC and AAAP families (based on Figures 4.2 and 4.3). The *slimfast* expansion in the APC family is highlighted with a gray box, and numbers are mapped onto nodes to indicate the four *slimfast* subclades resulting from gene duplications predating aphid/phylloxera divergence. Twofold or greater differential expression between bacteriocyte (B) and whole insect (WI) is mapped onto trees. Expression data for *M. persicae* amino acid transporters are reported in Supplemental file 6 from Duncan et al. (2016). Differential expression for *A. pisum* transporters is based on consistent differential expression in the same direction found across four studies: Hansen & Moran (2011), Price et al. (2011), Macdonald et al. (2012), and Duncan et al. (2014).

Table 4.1. Taxon sampling within Aphidomorpha

| **Family** |
| --- |
|   **Subfamily** |
|     **Tribe** |
|       ***Genus species*** |
| Phylloxeridae |
|       *Daktulosphaira vitifoliae* |
| Aphididae |
|   Eriosomatinae |
|     Pemphigini |
|       *Pemphigus obesinymphae* |
|   Tamaliinae |
|       *Tamalia coweni* |
|   Aphidinae |
|     Aphidini |
|       *Aphis nerii* |
|     Macrosiphini |
|       *Myzus persicae* |
|       *Acyrthosiphon pisum* |

**Chapter 5**

Co-option of a conserved amino acid transporter enables metabolic integration between sap-feeding insects and their endosymbionts[4]

**Summary**

Sap-feeding insects like the pea aphid, *Acyrthosiphon pisum*, depend on endosymbionts for amino acids. Recent work suggests that a key regulator of amino acid supply by the endosymbiont *Buchnera* is the aphid-encoded glutamine transporter ApGLNT1. By regulating supply of precursor glutamine to the *Buchnera* host cells (bacteriocytes) through feedback inhibition by *Buchnera*-produced arginine, ApGLNT1 plays an important role in *A. pisum*/*Buchnera* metabolic integration. Global ApGLNT1 expression in *A. pisum* and its presence in genomes of non-symbiotic insects implies that it was co-opted to function in symbiosis. Understanding the mechanism of co-option will help answer the long-standing question of how metabolic integration evolves in endosymbiotic relationships. Using a combination of comparative gene annotation, gene expression analysis, and functional characterization of ApGLNT1 orthologs in other insects, I examined the role of gene co-option in metabolic integration in sap-feeding insect endosymbioses. Annotation of arginine biosynthesis genes in sap-feeding insect transcriptomes revealed that sap-feeding insects vary in their requirement for symbiotic provisioning of arginine – a necessary requirement for ApGLNT1-mediated metabolic integration. Gene expression analyses of ApGLNT1 orthologs also showed variation in employment of ApGLNT1 orthologs in bacteriocytes. Functional characterization of ApGLNT1 orthologs in sap-feeding and non sap-feeding insects revealed that ApGLNT1 and its orthologs are functionally conserved, implying that it was co-opted to a symbiotic

---

[4] This chapter will be submitted for publication with authors RP Duncan, CW Luetje, and ACC Wilson.

role not through functional evolution but through recruitment to be expressed in bacteriocytes. These data indicate that evolving a tightly integrated endosymbiotic relationship is possible within the functional constraints of the ancestral host genome.

**Background**

Bacterial endosymbionts live intracellularly within eukaryotic hosts and fundamentally shape the evolutionary trajectory of their hosts. By giving their hosts access to their genomic repertoire, bacterial endosymbionts underlie the evolution of complex phenotypes (Moran 2007), enabling hosts to thrive in novel environments. Endosymbiosis has evolved countless times, especially in insects (Douglas 1989). Persistent interactions between hosts and symbionts, some of which have co-evolved for hundreds of millions of years (Moran et al. 1993), result in the partners becoming integrated to the extent that they are no longer separate organisms, but rather indispensible parts of a whole (McFall-Ngai et al. 2013). Mechanisms of host/endosymbiont developmental and metabolic integration are being illuminated with the help of genomic tools (Hansen & Moran 2011, Lu et al. 2016, Luan et al. 2015, Macdonald et al. 2012, McCutcheon & Dohlen 2011, Price & Wilson 2014, Price et al. 2014, Shigenobu & Stern 2013, Sloan et al. 2014, Wilson et al. 2010), but the evolutionary origin of host/endosymbiont integration remains largely a mystery.

Recent work on sap-feeding insects in the Hemipteran suborder Sternorrhyncha (e.g. aphids, mealybugs, and psyllids) demonstrates that they are especially good candidates to investigate the evolutionary origin of host/endosymbiont integration, particularly of metabolic integration. Sternorrhynchans and their endosymbionts have

evolved extremely intimate and interdependent relationships.  The insect hosts critically

depend on their endosymbionts for provisioning of vitamins and essential amino acids.

Endosymbionts reside in membrane-bound compartments in specialized insect organs

called bacteriomes.  Inside bacteriome cells (bacteriocytes), symbionts produce essential

amino acids from metabolic precursors provided by their insect hosts.  Endosymbionts

are equally dependent on their hosts, as nutritional provisioning requires complementary

host and endosymbiont gene sets in biosynthesis of vitamins, amino acids, and even

phospholipids (Hansen & Moran 2011, Husnik et al. 2013, Macdonald et al. 2012,

McCutcheon & Dohlen 2011, Price & Wilson 2014, Shigenobu et al. 2000, Sloan et al.

2014, Wilson et al. 2010, Wilson & Duncan 2015).  Symbiont localization in membrane-

bound compartments presents a barrier to nutrient exchange – a barrier that can be

crossed by host encoded, membrane-bound transporters, such as amino acid transporters.

Studies of the pea aphid *Acyrthosiphon pisum* and its endosymbiont *Buchnera aphidicola*

suggest that amino acid transporters play an important role in host/symbiont metabolic

integration in sap-feeding insects.  For example, the amino acid transporter ApGLNT1 is

thought to be a key player in *A. pisum*/*Buchnera* metabolic integration.  Based on

functional characterization of ApGLNT1, Price et al. (2014) proposed a model in which

ApGLNT1 regulates *Buchnera* and *A. pisum* amino acid metabolism in bacteriocytes,

providing the aphid host with a mechanism to control amino acid supply in response to

demand.  Robust knowledge of amino acid biosynthesis in sternorrhynchans coupled with

evidence that ApGLNT1 regulates amino acid metabolism in *A. pisum* provides a

framework within which to examine how these insects and their endosymbionts evolved

metabolic integration with respect to amino acid biosynthesis and exchange.

Assessing possible mechanisms for the evolution of metabolic integration in sternorrhynchans requires detailed understanding of how ApGLNT1 is hypothesized to regulate amino acid biosynthesis in both aphids and *Buchnera* in bacteriocytes (Price et al. 2014). ApGLNT1 belongs to the amino acid/auxin permease (AAAP) familiy of transporters (Transporter Classification #2.A.18), falling within a large arthropod-specific expansion (Price et al. 2011). Localizing to the bacteriocyte plasma membrane, ApGLNT1 transports the non-essential amino acid glutamine into bacteriocytes, supplying the cells with an essential precursor for the biosynthesis of the remaining 19 protein amino acids (Price et al. 2014). Importantly, glutamine import into bacteriocytes is inhibited by a *Buchnera*-produced amino acid – arginine (Price et al. 2014). Arginine is typically produced by metazoans through the urea cycle, but aphids have lost four of the five urea cycle genes from their genome, and with them the ability to make their own arginine (Wilson et al. 2010). Although arginine is present in phloem sap, its concentration is relatively low and falls short of what is required to support aphid growth (Gündüz & Douglas 2009). Therefore, aphids rely on *Buchnera* for arginine supply. Glutamine is a substrate for arginine biosynthesis, but most *Buchnera*-produced amino acids require other non-essential amino acids as nitrogen donors (Shigenobu & Wilson 2011). Glutamine, however, indirectly acts as a precursor to all amino acid biosynthesis in bacteriocytes through its role in producing the other three substrates of *Buchnera*-produced amino acids (glutamate, aspartate, and serine) – the enzyme glutamate synthase (EC 1.4.1.13) uses glutamine (nitrogen donor) and 2-oxoglutarate (carbon backbone) to produce glutamate. In turn, glutamate is the nitrogen for the biosynthesis of aspartate and serine. Importantly, genes catalyzing the conversion of glutamine to glutamate and

glutamate to aspartate and serine are enriched in aphid bacteriocytes (Hansen & Moran 2011).  Therefore, inhibiting glutamine transport into bacteriocytes effectively shuts down all amino acid biosynthesis in bacteriocytes (Price et al. 2014).  Feedback inhibition of ApGLNT1 by arginine therefore provides *A. pisum* with a mechanism to control the metabolic output of *Buchnera* in response to demand (Price et al. 2014). Remarkably, despite the simplicity of this regulatory mechanism, ApGLNT1 is not specialized for bacteriocytes.  Expression information and phylogenetic analyses show that the symbiotic function of ApGLNT1 is derived, not ancestral.  Apart from being highly expressed in *A. pisum* bacteriocytes, ApGLNT1 is expressed in *A. pisum* head, gut, ovariole, and developing embryos, indicating that it plays non-symbiotic roles in other tissues (Lu et al. 2016, Price et al. 2011).  In addition, ApGLNT1 shares one-to-one orthology with genes in other insect genomes, including non sap-feeding insects without bacteriocyte-associated symbionts (Price et al. 2011), implying that its ancestral role was non-symbiotic and it was co-opted towards a symbiotic function in *A. pisum*.  Whether similarly elegant mechanisms of host/symbiont integration have evolved in other systems remains to be determined.

Understanding the functional evolution of ApGLNT1 will illuminate the mechanism by which it was co-opted to facilitate metabolic integration of *A. pisum* and *Buchnera*.  Occurring via several mechanisms that involve either duplicated or single-copy genes, gene co-option (True & Carroll 2002) results when evolution finds new functions for previously existing genes.  Co-option can result from gene duplication because redundant copies of a gene (paralogs) can diverge functionally, providing an opportunity for a new function to evolve or for an ancestral, secondary function to

become optimized in one of the copies.  Additionally, paralogs can undergo evolution in

*cis*-regulatory regions, leading to expression in new tissues (Lynch 2007).  In single-copy

genes, co-option can take place through evolution in the coding sequence in a way that

leaves the ancestral function intact, and/or expression evolution through changes in *cis*-

regulatory elements.  ApGLNT1 is encoded by a single-copy gene in *A. pisum* and

several other insects, including other sternorrhynchans and outgroup taxa (Duncan et al.

2014, Price et al. 2011).  Thus, the mechanism of co-option minimally took place in its

*cis*-regulatory region in response to selection for ApGLNT1 function in a derived cell

type (bacteriocytes).  However, ApGLNT1 could also have undergone evolution in its

coding region, resulting in functional evolution in response to selection for an optimized,

symbiotic function.

Here, I use sap-feeding insects as a model to address the role of gene co-option in

host/endosymbiont metabolic integration using a combination of gene annotation, gene

expression analysis, and functional characterization in a phylogenetic framework.  In this

study, I aim to answer two questions.  First, are ApGLNT1 orthologs likely to play a role

in regulating host/endosymbiont metabolism in other sternorrhynchans through the same

mechanism proposed for *A. pisum*?  If so, then I predict that other sternorrhynchans (1)

have lost the urea cycle, leaving them dependent on their endosymbionts for arginine

supply, (2) express their ApGLNT1 ortholog in their bacteriocytes, and (3) their

ApGLNT1 orthologs are functionally equivalent to *A. pisum* ApGLNT1.  Secondly, I aim

to investigate if ApGLNT1 evolved functionally from its ancestral role to fulfill its role in

host/symbiont metabolic integration.  By testing if ApGLNT1 orthologs play a regulatory

role in other sternorrhynchans, I will improve our understanding of the generality of

feedback inhibition in the establishment of host/symbiont metabolic integration. By examining the functional evolution of ApGLNT1 across sternorrhyncans and outgroups, I will be able to determine if the regulatory role of ApGLNT1 was preceded by evolution of a bacteriocyte-specific function, or recruitment of a conserved function to a derived, endosymbiotic context.

## Methods

### Taxon sampling

To examine both the functional significance of ApGLNT1 and its orthologs in Sternorrhyncha and its functional evolution in insects, I focused on three sternorrhynchan insects representing three of the suborder's four major groups (psyllids, scales, and aphids) and two outgroup insects from two different insect superorders. Sternorrhynchans included in this study are the pea aphid *Acyrthosiphon pisum*, the citrus mealybug *Planococcus citri*, and the Asian citrus psyllid *Diaphorina citri*. All three of these sap-feeding insects possess amino acid provisioning bacterial symbionts that are all localized within membrane-bound compartments in host insect bacteriocyte cells. As outgroups, I selected the human body louse *Pediculus humanus* (superorder Paraneoptera, order Phthiraptera), and the flour beetle *Tribolium castaneum* (superorder Endopterygota, order Coleoptera). All samples came from laboratory colonies subject to previous studies with the exception of *P. citri*. In addition to using samples from a colony of *P. citri* maintained in a greenhouse at Utah State University in Logan, UT (also used in studies by von Dohlen et al. (2001), McCutcheon & von Dohlen (2011), Husnik et al. (2013), and Duncan et al. (Duncan et al. 2014)), I also included samples from a population of *P.*

*citri* that I collected from a cacao fruit at the United States Department of Agriculture

Subtropical Horticultural Research Station in Miami, FL. The Miami population is now

maintained as a laboratory colony in the lab of my Ph.D. advisor, Alex Wilson. Also

used were *A. pisum* lab lines 7A (Moran et al. 2009) and 9-2-1 (Russell & Moran 2006), a

colony of *D. citri* from the lab of Kirsten Pelz-Stelinski at the University of Florida Citrus

Research and Education Center, and the genome lines of *T. castaneum* and *P. humanus*.

Relationships between these taxa are depicted in Figure 5.1.


**Transcriptome sequencing and assembly**

Whole insect (female) and bacteriome RNA-seq data were generated for the new

*P. citri* population from Miami, FL, *D. citri*, and *A. pisum*. Total RNA was extracted for

each replicate from tissues preserved in TRI Reagent (Ambion) according to the

manufacturer guidelines from 10-20 adult (*A. pisum* and *D. citri*) or fourth instar (*P. citri*)

females and bacteriomes dissected from 100 adult (*A. pisum* and *D. citri*) or fourth instar

(*P. citri*) females and stored in trizol. Resulting RNA was DNAseI treated before

sending to the Hussman Institute for Human Genomics at the University of Miami Miller

School of Medicine for library preparation and paired-end, strand-specific sequencing on

the Illumina HiSeq 2500 platform. Raw RNA-seq reads were deposited in the NCBI

Sequence Read Archive under BioProject PRJNA315109.

Reference transcriptomes were assembled for each insect species in Trinity

(v.2.1.1) (Haas et al. 2013) by combining reads for each species, including previously

sequenced *P. citri* RNA-seq paired-end data reported by Husnik et al. (2013) and Duncan

et al. (2014). Raw reads were filtered to a minimum quality score of 30 over 95% of the

read and remaining single-end reads were discarded, resulting in filtered, paired-end reads. Reference transcriptomes were assembled on the Pegasus supercomputer at the University of Miami Center for Computational Science. Transcriptomes for *D. citri* and *A. pisum* were assembled using the strand-specific parameter in order to benefit from the information provided by strand-specificity. The reference assembly for *P. citri*, however, was assembled in non strand-specific mode, as it contained reads that were not strand-specific from a previous data set (Duncan et al. 2014, Husnik et al. 2013).

**Annotation**

*Urea cycle*

Urea cycle genes were annotated in *P. citri* and *D. citri* through a reciprocal BLAST approach. Urea cycle protein sequences from *Drosophila melanogaster* or, when absent from *D. melanogaster,* from *P. humanus* (Table 1) were used as query sequences for TBLASTN searches against *P. citri* and *D. citri* transcriptome assemblies generated here (described above). Top hits from TBLASTN searches included multiple transcripts with the same Trinity gene ID number, indicating that they were variants of the same locus. Therefore, the top transcript for each unique Trinity gene ID found by TBLASTN searches was selected as a representative for a reciprocal blast (BLASTX) search against NCBI refseq proteins. The transcripts subjected to reciprocal BLASTX searches were also mapped to the genome scaffolds of their host species to confirm their presence in the genome. Transcripts identified from insect transcriptomes that significantly matched bacterial proteins in BLASTX searches were used as query sequences in BLASTN searches against the insect draft genome to confirm presence on insect genomic scaffolds.

Genomic scaffolds containing bacterial genes were further examined for the presence of

insect genes to confirm that scaffolds are part of the insect genome and not from bacterial

contaminants or symbionts.  The *P. citri* draft genome was reported by Husnik et al.

(2013), and the *D. citri* draft genome is available on NCBI (BioProject PRJNA251515)

and through the i5K BLAST server hosted at the National Agricultural Library website:

https://i5k.nal.usda.gov/webapp/blast/.  I compared annotation results to the urea cycle

genes found for *P. humanus* and *T. castaneum*, stored in the Kyoto Encyclopedia of

Genes and Genomes (KEGG).

*ApGLNT1 orthologs*

ApGLNT1 orthologs were identified from *P. citri* and *D. citri* reference

transcriptomes using a reciprocal BLAST approach.  The protein sequence for ApGLNT1

(isoform NP_001233190.1) was used as a query sequence in TBLASTN searches against

*P. citri* and *D. citri* transcriptomes.  Top hits with the same Trinity gene ID were then

collapsed into a single locus and the top hit was used as a query for BLASTX searches

against the NCBI refseq proteins to confirm orthology to ApGLNT1.  The resulting *P.

citri* ortholog was confirmed to be identical in nucleotide sequence to the ApGLNT1

ortholog supported by phylogenetic analysis in Chapter 3 of this dissertation (also

published (Duncan et al. 2014)).  The same process was done for ApGLNT1 in the *A.

pisum* reference transcriptome to enable comparison of bacteriocyte expression.

**Bacteriocyte expression analysis**

Bacteriocyte expression was quantified for ApGLNT1 and its orthologs in *P. citri* and *D. citri* using the RNA-seq reads that made up the reference assemblies generated here. The two *A. pisum* clones (7A and 9-2-1) and *P. citri* populations (Miami, FL and Logan, UT) were treated as biological replicates for those two species. Transcript abundance was estimated using the PERL script pipeline implemented in Trinity. Briefly, filtered reads from bacteriocyte libraries were mapped to ApGLNT1 or orthologs in the appropriate reference assembly using Bowtie (v.0.12.1) (Langmead et al. 2009) and read counts were estimated in RSEM (v.1.2.25) (Li & Dewey 2011). Transcript abundance estimates were used to generate <u>F</u>ragments <u>P</u>er <u>K</u>ilobase of transcript per <u>M</u>illion reads mapped (FPKM) as a measure of gene expression for ApGLNT1 and each ortholog.

**Cloning ApGLNT1 orthologs**

Expression constructs were generated for ApGLNT1 (isoform NP_001233190.1) and its orthologs in *P. citri, D. citri, P. humanus*, and *T. castaneum*. Orthologs from *P. humanus* and *T. castaneum* were identified through phylogenetic analysis in previous work (Price et al. 2011). Full-length coding sequences were PCR amplified from whole insect cDNA with Phusion High Fidelity DNA polymerase (Finnzymes). Primers (Supplemental Table 5.1) contained a 5' Kozak initiation sequence and a restriction site (BGlII or BamHI) and a 3' XbaI restriction site. PCR amplification was conducted on a Mastercycler ep thermal cycler (Eppendorf) using an initial denaturing step at 98º C for 30 s followed by 30 cycles of 98º C for 10 s, 60 º C for 30 s, and 72º C for 1 min, and ending with a step at 72º C for 5 min. PCR products were initially blunt end cloned into

pJET1.2 for sequence verification by Sanger sequencing. Next, vectors containing the correct sequences were double digested with the appropriate restriction enzymes and ligated into BamHI and XbaI sites in the expression vector pCS2+. After generating expression constructs, amino acid transporter coding sequences were once again verified by Sanger sequencing.

**ApGLNT1 expression in *Xenopus* oocytes**

*Xenopus laevis* oocytes were injected with copy RNA (cRNA) generated for ApGLNT1 and each ortholog using the SP6 mMESSAGE mMACHINE kit (Ambion). Experimental oocytes were injected with DEPC treated water containing either 23 ng (ApGLNT1 and orthologs from *P. citri*, *P. humanus*, and *T. castaneum*) or 46 ng (*D. citri* ortholog) of cRNA. Control oocytes for each run were injected with the same volume of DEPC treated water, according to the ortholog being examined. Oocytes were incubated at 17º C in Barth's solution (in mM: 88 NaCl, 1 KCl, 2.4 NaHCO, 0.3 CaNO, 0.41 CaCl, 0.82 MgSO, 15 HEPES (pH 7.6) and 100 µg/l amikacin, 50 µg/l ciprofloxacin, and 50 µg/l tetracycline) for at least 2 days before performing electrophysiology recordings.

**Two-electrode voltage clamp**

Two-electrode voltage clamp (TEVC) was used to functionally characterize ApGLNT1 and its orthologs in *P. citri*, *D. citri*, *P. humanus*, and *T. castaneum*. Experiments were based on previously demonstrated transport function of ApGLNT1 (Price et al. 2014), in which amino acid transport is proton-dependent (Price et al. 2015). Briefly, I examined glutamine concentration-response, pH dependence, substrate

specificity, arginine inhibition of glutamine transport, and response to arginine. Current

deflections (responses) generated by amino acids were recorded on an OpusXpress

6000A automated parallel electrophysiology robot (Molecular Devices). Micropipettes

filled with 3 M KCl had resistances of 0.2-1.9 MW. Oocyte membrane potentials were

held at -70 mV and continuously perfused with ND96 (in mM: 96 NaCl, 2 KCl, 1 CaCl, 1

MgCl, 5 HEPES, pH 7.5). Amino acid solutions were prepared the same day as

recordings by adding individual amino acids to ND96 at specific concentrations (see

below and in figures). Amino acid solutions were applied to oocytes for 60 s at a flow

rate of 1.0 ml/min followed by a washout period of 180 s with ND96 at a flow rate of 4.6

ml/min. Water-injected oocytes were run alongside oocytes expressing transporters as a

negative control for all experiments. Responses were measured using Clampfit 9.1

software (Molecular Devices) by comparing the mean baseline current to the mean

current at 30 s following the initiation of the application. Data were further analyzed in

Prism 5.0c software (GraphPad) based on the design of each experiment.


*Glutamine concentration-response*

A series of concentrations of glutamine were applied to oocytes expressing

ApGLNT1 or one of its orthologs. Raw responses to each glutamine concentration were

normalized to the maximal response and then averaged to generate concentration-

response curves. The concentration of glutamine yielding half the maximal response

($EC_{50}$) was determined through nonlinear regression by fitting concentration-response

data to the equation $I = I_{max}/(1+(EC_{50}/X)^n)$, where I is the current response to a given

concentration (X) of glutamine, $I_{max}$ is maximal response, and n is the apparent Hill coefficient.

*pH dependence*

Glutamine (5 mM) was prepared in ND96 buffered with HEPES (pH 7.5, 8.5) or MES (pH 5.5, 6.5). A 30 s application of ND96 of each pH was immediately followed by a 60 s co-application of glutamine solution of the same pH. Glutamine applications were then followed by a 180 s washout with ND96 (pH 7.5). Raw responses to glutamine at each pH were normalized to the response to glutamine at pH 5.5 and averaged. Statistical significance of response differences to different pH values was assessed with a one-way ANOVA followed by Tukey's multiple comparisons test.

*Amino acid screen*

Substrate specificity was analyzed qualitatively through amino acid screens. Amino acid screens were performed in two runs, one for nonessential amino acids and one for essential amino acids with glutamine as a positive control. Cysteine was not included in the nonessential amino acid screen because it induced a current deflection in water-injected control oocytes comparable to the current deflection observed in oocytes expressing ApGLNT1 or one of its orthologs in other insects. Amino acid solutions were applied at 2 mM to oocytes expressing all orthologs but that of *D. citri*, which has low sensitivity to glutamine. Instead, I applied amino acid solutions at 10 mM for all amino acids but tyrosine, which was applied at 2.5 mM because it is insoluble at higher concentrations. For the nonessential amino acid screen of oocytes expressing the *D. citri*

ortholog, aspartate and and glutamate were not included because at the higher 10 mM

concentration used for *D. citri* amino acid screens, they also induced a current deflection

in both transporter-expressing and water-injected oocytes.


*Arginine inhibition*

Arginine inhibition of glutamine transport was examined by alternating

applications of $EC_{50}$ glutamine and $EC_{50}$ glutamine plus arginine at a series of

concentrations (0.2 mM, 0.4 mM, 1 mM, 2 mM, 5 mM).  Raw responses to each

glutamine + arginine application were normalized to the preceding glutamine application

and then averaged for each arginine concentration.  Averaged responses to each arginine

concentration were used to generate arginine inhibition curves for ApGLNT1 and each

ortholog.  The half maximal inhibitory concentration ($IC_{50}$) of arginine was calculated

through nonlinear regression by fitting concentration-inhibition data to the equation

$I=I_{max}/(1+(X/IC_{50})^n)$, where $I_{max}$ is the maximal response in the absence of inhibitor ($EC_{50}$

glutamine in the absence of arginine) and n is the apparent Hill coefficient.


*Arginine response*

Arginine response was assessed by applying a normalizer (EC25 glutamine)

followed by a series of applications of arginine (1 mM, 3 mM, 5 mM, and 10 mM).  Raw

responses were qualitatively compared to the response of the oocyte to the normalizer.

**Results**

**The urea cycle was lost in Sternorrhyncha**

Similar to the pea aphid *Acyrthosiphon pisum* (Wilson et al. 2010), urea cycle gene annotation in the citrus mealybug *Planococcus citri* and the Asian citrus psyllid *Diaphorina citri* revealed that four of the five urea cycle genes of eukaryotic origin (EC 2.1.3.3, EC 3.5.3.1, EC 4.3.2.1, and EC 6.3.4.5) have been lost from the genome of both insects (Table 1 and Figure 5.2). Both *P. citri* and *D. citri*, along with *A. pisum* (Wilson et al. 2010), retain the gene encoding nitric oxide synthase (EC 1.14.13.39; Table 5.1 and Figure 5.2), which catalyzes the reversible conversion between arginine and citruline. Despite the loss of the eukaryotic urea cycle gene arginosuccinate lyase (EC 4.3.2.1), which catalyzes the conversion of L-arginosuccinate to arginine, *D. citri* encodes two copies of arginosuccinate lyase of bacterial origin (also known as *argH*; Table 5.2).

In contrast to *P. citri* and *D. citri*, the urea cycle is more intact in the outgroups *Pediculus humanus* and *Tribolium castaneum* (Figure 5.2). On the basis of KEGG annotations, the *P. humanus* genome encodes arginase (EC 3.5.3.1), ornithine carbamoyltransferase (EC 2.1.3.3) and nitric oxide synthase (EC 1.14.13.39) while the *T. castaneum* genome encodes all urea cycle genes except for ornithine carbamoyltransferase (EC 2.1.3.3) (Figure 5.2).


**ApGLNT1 and its *P. citri* ortholog are expressed in bacteriocytes**

In reference transcriptomes for each of the three sternorrhynchans (*A. pisum*, *P. citri*, and *D. citri*), one locus was found to represent ApGLNT1 or its ortholog (see table 5.3 for Trinity gene IDs). Bacteriocyte expression was then analyzed using the transcript

abundance estimation pipeline in Trinity. Results of the transcript abundance quantification are shown in table 5.3. Expression was measured in $\underline{F}$ragments $\underline{P}$er $\underline{K}$ilobase of transcript per $\underline{M}$illion reads mapped (FPKM) of representative transcripts for ApGLNT1 and its orthologs in *P. citri* and *D. citri*. The *A. pisum* transcript representing ApGLNT1 had a FPKM of 103.33 for the 7A clone and 99.35 for the 9-2-1 clone, confirming that ApGLNT1 is expressed in *A. pisum* bacteriocytes (Price et al. 2011, 2014). The *P. citri* ortholog is also expressed in bacteriocytes, with an FPKM of 314.87 for the Logan, UT population and 278.37 for the Miami, FL population. In contrast, expression analyses failed to support bacteriocyte expression for the *D. citri* ortholog of ApGLNT1, with a FPKM of 0 (Table 5.3).

**ApGLNT1 and its orthologs are functionally conserved**

Like ApGLNT1, orthologs in sternorrhynchans *P. citri* and *D. citri*, as well as in outgroups *P. humanus* and *T. castaneum*, transport glutamine with high specificity in a pH-dependent manner (Figures 5.3-5.5), and that glutamine transported is inhibited by the presence of arginine (Figures 5.6-5.7). Glutamine concentration-response curves (Figure 5.3) indicate that the concentration dependence of glutamine transport is in the same order of magnitude between ApGLNT1 and most orthologs examined here, with $EC_{50}$ concentrations between $1.5 \pm 0.2$ mM and $5.7 \pm 0.8$ mM (mean $\pm$ standard error). Orthologs from *T. castaneum* and *D. citri* were less sensitive to glutamine, with $EC_{50}$ values at $15.1 \pm 7.2$ mM and $35.2 \pm 18.5$ mM respectively, but high error precludes my ability to confidently conclude that these orthologs are significantly less sensitive to glutamine than the other orthologs. Glutamine transport increases with decreasing pH (Figure 5.4; one-way ANOVA followed by Tukey's multiple comparison test. *A. pisum*:

$F(3,20) = 160.5$, $p < 0.05$; *P. citri*: $F(3,12) = 601.8$, $p < 0.05$; *D. citri*: $F(3,24) = 232.6$, $p < 0.05$; *P. humanus*: $F(3,28)=45.4$, $p < 0.05$; *T. castaneum*: $F(3,44) = 673.7$, $p < 0.05$), consistent with proton-coupled transport described for ApGLNT1 (Price et al. 2015). Further, amino acid screens show that oocytes expressing ApGLNT1 or it orthologs respond only to glutamine among the protein amino acids at the concentration I tested (2 mM for ApGLNT1 and orthologs from *P. citri*, *P. humanus*, and *T. castaneum*; 10 mM for *D. citri* ortholog; Figure 5.5). Glutamine transport in ApGLNT1 orthologs is inhibited by the presence of arginine (Figure 5.6) at a comparable concentration range, with $IC_{50}$ values ranging from $2.5 \pm 0.2$ mM to $6.1 \pm 0.8$ mM (Figure 5.7).

Price et al. (2014) previously showed that while glutamine was the primary substrate of ApGLNT1, it also recognizes arginine as a substrate and transports it at very high concentrations compared with glutamine. I found that apart from ApGLNT1, only the *T. castaneum* ortholog responds to arginine at the concentrations I examined (1 mM, 3 mM, 5 mM, and 10 mM; Figure 5.8). Similarly to ApGLNT1, the *T. castaneum* ortholog responds to arginine only at high concentrations compared with glutamine (Figure 5.8).

**Discussion**

**Variation in host/symbiont metabolic integration across Sternorrhyncha**

Variation in host-encoded arginine biosynthesis genes suggests that *A. pisum*, *P. citri*, and *D. citri* are not metabolically integrated through the same mechanism. Urea cycle annotation revealed that *P. citri*, like *A. pisum*, is unable to synthesize arginine, indicating that it must receive arginine either through its diet and/or from its

endosymbionts (Figure 5.2, Table 5.1).  Indeed, *P. citri* endosymbionts *Candidatus*

Tremblaya princeps (Thao et al. 2002) and its nested partner (Dohlen et al. 2001)

*Candidatus* Moranella endobia (McCutcheon & Dohlen 2011) together encode the entire

pathway for arginine biosynthesis (McCutcheon & Dohlen 2011).  Exclusive symbiont

provisioning of arginine in *P. citri* is consistent with using ApGLNT1 to regulate amino

acid supply through feedback inhibition by arginine, as has been described in *A. pisum*

(Price et al. 2014).  In contrast, urea cycle annotation in *D. citri* revealed that it encodes

two copies of arginosuccinate lyase (EC 4.3.2.1), the gene capable of catalyzing the last

step of arginine biosynthesis (Figure 5.2, Table 5.1).  Both copies are of bacterial origin

(Table 5.2), consistent with other psyllid species (Sloan et al. 2014).  Sloan et al. (2014)

found that bacterial arginosuccinate lyase in the psyllid genome originated through an

ancient horizontal gene transfer event followed by a gene duplication event in the psyllid

genome.

Although the *D. citri* endosymbiont *Candidatus* Carsonella rudii DC encodes the

full arginine biosynthesis pathway (Nakabachi et al. 2013) and thus is able to provision

its host with arginine, having two copies of the gene encoding the terminal step could

give *D. citri* the ability to synthesize arginine outside of bacteriocytes, as long as the

metabolic precursor from Carsonella DC can be transported to different parts of the body.

In fact, genomic information for another psyllid species suggests that it is possible to

move the metabolic precursor of arginine minimally out of Carsonella DC into

bacteriocytes.  The hackberry petiole gall psyllid, *Pachypsylla venusta*, also encodes

bacterial versions of arginosuccinate lyase, but it endosymbiont does not encode the

terminal step of arginine biosynthesis (Sloan et al. 2014).  Thus, the host-encoded genes

are necessary to complete the arginine biosynthesis pathway in this system, which may occur in the bacteriocyte cytoplasm after exporting L-arginosuccinate from Carsonella PV. Further analysis of expression profiles of *D. citri*-encoded arginosuccinate lyase will help determine if *D. citri* relies exclusively on Carsonella DC for arginine provisioning, but the fact remains that, based on the genes encoded in the host genome, arginine biosynthesis is a possibility in *D. citri*, but not in *P. citri* or *A. pisum*. This difference in arginine biosynthesis capabilities calls into question whether ApGLNT1 plays equivalent roles in metabolic integration for these three sternorrhynchans.

Apart from differences in arginine biosynthesis, differences in expression of ApGLNT1 and its orthologs in *D. citri* and *P. citri* point towards variation in mechanisms of metabolic integration among these insects. ApGLNT1 is highly expressed in bacteriocytes (Price et al. 2011, 2014), which I confirmed through a bacteriocyte expression analysis (Table 5.3). Similarly, the *P. citri* ortholog of ApGLNT1 is also expressed in bacteriocytes, while the *D. citri* ortholog is not (Table 5.3). Expression of the *P. citri* ortholog in bacteriocytes is consistent with ApGLNT1-mediated regulation of amino acid supply by *P. citri* endosymbionts through the same mechanism found in *A. pisum* (Price et al. 2014), and can be confirmed through immunolocalization of the *P. citri* ortholog of ApGLNT1. The complete absence of bacteriocyte expression of the *D. citri* ApGLNT1 ortholog, on the other hand, implies that the *D. citri* ApGLNT1 ortholog is not employed in a symbiotic context. Thus, if ApGLNT1 functions in host/endosymbiont metabolic integration in *A. pisum* and *P. citri*, *D. citri* employs a different mechanism for metabolic integration with its endosymbiont.

Distinct metabolic integration mechanisms in *A. pisum* and *D. citri* are probably not explained by differences in the basic path of nitrogen metabolism in their bacteriocytes.  In *A. pisum*, glutamine is present at a high concentration in the hemolymph and is actively taken up by bacteriocytes (Sasaki & Ishikawa 1995), perhaps by ApGLNT1 (Price et al. 2014).  Inside bacteriocytes, glutamine has several important metabolic roles.  Not only is it an amino donor for purine synthesis, but it is also an amino donor for the *Buchnera*-produced amino acids arginine and histidine (Hansen & Moran 2011, Shigenobu & Wilson 2011).  Furthermore, glutamine feeds into bacteriocyte-mediated ammonia recycling through the enzymes glutamine synthase and glutamine oxoglutarate aminotransferase (GS/GOGAT), which recycles nitrogen from ammonia waste to be used in amino acid biosynthesis (Hansen & Moran 2011, Macdonald et al. 2012).  The GS/GOGAT cycle helps aphids manage ammonia production in bacteriocytes (Hansen & Moran 2011, Macdonald et al. 2012) and, importantly, produces glutamate, an important direct or indirect substrate for most aphid- and *Buchnera*-produced amino acids (Hansen & Moran 2011, Shigenobu & Wilson 2011).  Based on current knowledge, nitrogen metabolism is probably very similar in *D. citri*.  *D. citri* also needs glutamine and glutamate as substrates for amino acid biosynthesis (among other metabolic roles).  Further *D. citri* likely also employs the GS/GOGAT cycle to recycle waste ammonia into amino acids, on the basis that in another psyllid species, the hackberry petiole gall psyllid *Pachypsylla venusta*, the genes encoding the enzymes responsible for the GS/GOGAT cycle are enriched in bacteriocytes (Sloan et al. 2014).  Therefore, I propose that *D. citri* uses a different mechanism than *A. pisum* to

generate glutamine in its bacteriocytes, possibly by using a different glutamine transporter at its bacteriocyte plasma membrane.

**An ancestral functional role in nutrition for ApGLNT1**

Functional characterization of ApGLNT1 orthologs indicate that this amino acid transporter is functionally conserved across *P. citri*, *D. citri*, *P. humanus*, and *T. castaneum*. Functional conservation of ApGLNT1 and its orthologs across sap-feeding and non sap-feeding insects implies that it evolved its ancestral function in response to selection for a non-symbiotic role before the origin of endosymbiosis in Sternorrhyncha. Further supporting the notion that the ancestral functional role was nonsymbiotic (at least, with respect to bacteriocyte associated endosymbionts) is that this gene is highly expressed in gut tissues in adult female *A. pisum* (Price et al. 2011) and *T. castaneum* larvae (Morris et al. 2009). Gut expression of ApGLNT1 and its *T. castaneum* ortholog suggests that gut tissues are, at least in part, home to the ancestral functional role of this amino acid transporter, but it could play a role in other tissues. In fact, during development of parthenogenetic embryos in *A. pisum*, ApGLNT1 localizes to the follicular epithelium surrounding the ovariole and embryonic membrane surrounding individual embryos (Lu et al. 2016). Interestingly, known expression of ApGLNT1 and orthologs takes place at three important nutritional interfaces: (1) the gut epithelium is the interface between the diet and the insect hemolymph, (2) the bacteriocyte plasma membrane forms part of the interface between the endosymbiont and host tissues, and (3) the follicular epithelium surrounding developing embryos is an interface between embryonic tissues and maternal hemolymph, through which maternal nutrients are likely

provided.  Taken together, these data suggest that the ancestral functional role of ApGLNT1 and its orthologs is nutritional, and perhaps this amino acid transporter functions as a nutrient sensor, as proposed by Lu et al. (2016).

Interestingly, ApGLNT1 has four splice variants, which may be optimized to function in different contexts across time and space.  All of the differences in the splice forms are found in the N-terminal region of the protein, upstream of the transmembrane region where transported substrates bind (Gao et al. 2009, Shaffer et al. 2009).  Therefore, I do not expect substrate recognition to differ between the isoforms.  Differences in the N-terminal region of isoforms could affect substrate affinity, however, as has been shown in sugar transporters (Shulze et al. 2000).  Future work can determine the degree to which ApGLNT1 isoforms differ functionally, and whether they are optimized for different biological contexts.

**Metabolic integration through co-option of ancestral function**

The strong functional conservation of ApGLNT1 and its orthologs imply that its role in host/endosymbiont metabolic integration did not require evolution from its ancestral function towards a symbiotic function.  Instead, the symbiotic role for ApGLNT1 appears to involve evolution in its expression profile, possibly mediated by evolution in *cis*-regulatory regions.  Thus, metabolic integration between *A. pisum* and *Buchnera*, and possibly between *P. citri* and its endosymbionts, is made possible by co-opting the ApGLNT1 ancestral function to a novel, symbiotic context.  That co-option of ApGLNT1 does not require evolution at the functional level makes sense, given the support for a nutritional role of this amino acid transporter and the important nutritional

provisioning role of bacteriocytes.  Co-option of the ancestral ApGLNT1 function for host/symbiont metabolic integration also suggests the intriguing possibility that evolving a tightly integrated endosymbiotic relationship, at least for insects, does not require functional innovation, but rather is possible within the host's ancestral functional genomic constraints.
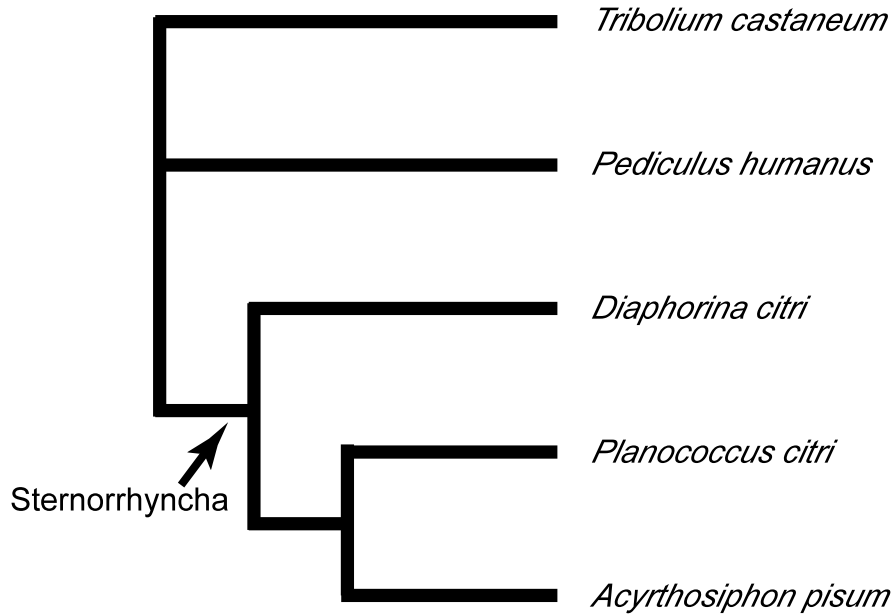
*Tribolium castaneum*

*Pediculus humanus*

*Diaphorina citri*

*Planococcus citri*

Sternorrhyncha

*Acyrthosiphon pisum*

Figure 5.1. Relationships between sampled taxa. ApGLNT1 and orthologs were examined from three sternorrhynchan sap-feeders (*A. pisum*, *P. citri*, and *D. citri*) and two non sap-feeding outgroups (*P. humanus* and *T. castaneum*). Relationships depicted here are based on studies by Cryan et al. (2011), Misof et al. (2014), Dahan et al. (2015).
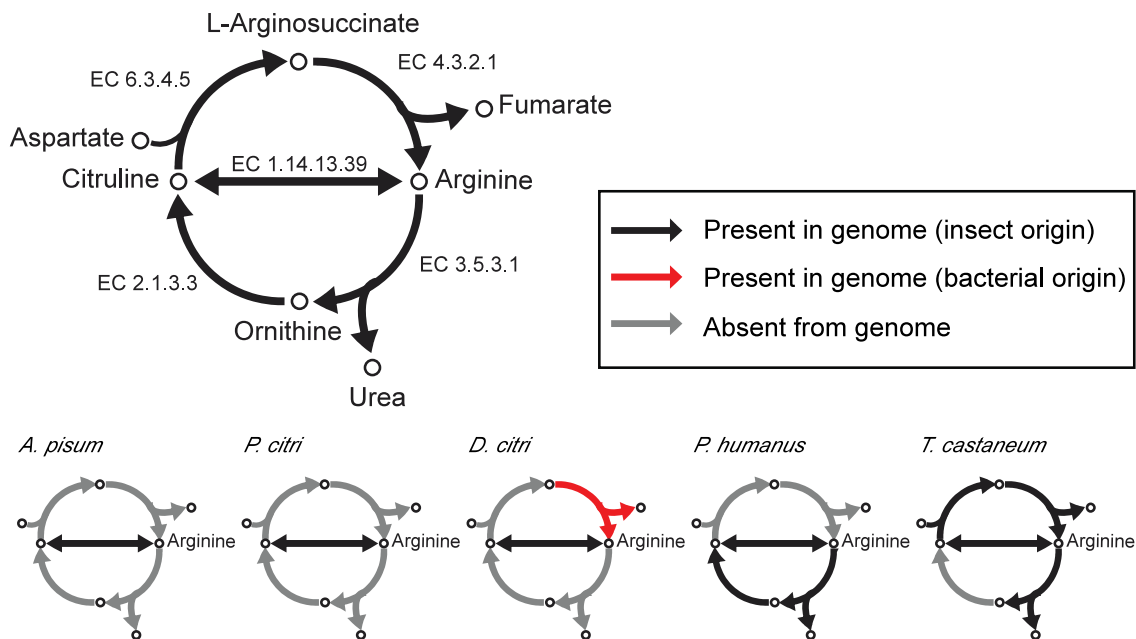
Figure 5.2. The urea cycle was lost in sternorrhynchans but not in non sap-feeding outgroups. Pictured above is a schematic of the complete urea cycle, where arrows represent enzymes catalyzing reactions in the cycle, and circles represent metabolites involved in the cycle. Pictured below are schematics of the urea cycle as encoded in the genomes of *P. citri*, *D. citri*, *P. humanus*, and *T. castaneum*, where black arrows represent genes of insect origin present in the genome, red arrows represent genes of bacterial origin present in the insect genome, and grey arrows represent genes that are absent from the genome. Four of the five urea cycle genes (EC 2.1.3.3, EC 3.5.3.2, EC 4.3.2.1, and EC 6.3.4.5) were lost in *P. citri* and *D. citri*, as has also been observed in *A. pisum* (Wilson et al. 2010). *D. citri* secondarily gained bacterial arginosuccinate lyase (EC 4.3.2.1) through horizontal gene transfer. The urea cycle is more complete in *P. humanus* (missing EC 4.3.2.1 and EC 6.3.4.5) and *T. castaneum* (missing EC 2.1.3.3 only).

Figure 5.3. ApGLNT1 and orthologs transport glutamine. On the left are representative current recordings showing responses to different concentrations (in mM) of glutamine in oocytes expressing ApGLNT1 or one of its orthologs (top trace is each set), or oocytes injected with water (bottom trace in each set). Length of black bars represent length of glutamine application to oocyte. On the right are glutamine concentration-response curves generated from data collected for ApGLNT1 or orthologs from each species, normalized to the maximum response. Each point on the curve represents mean ± standard error. Sample sizes for each trace range from 6 to 13 oocytes.

## A. pisum ApGLNT1



500 nA | 100 s

EC$_{50}$= 5.7 ± 0.8 mM

## P. citri



300 nA | 100 s

EC$_{50}$= 2.1 ± 0.2 mM

## D. citri



300 nA | 100 s

EC$_{50}$= 35.2 ± 18.5 mM

## P. humanus



500 nA | 100 s

EC$_{50}$= 1.5 ± 0.2 mM

## T. castaneum



1000 nA | 100 s

EC$_{50}$= 15.1 ± 7.2 mM

Figure 5.4. Glutamine transport is pH dependent.  On the left are representative current recordings showing responses to 5 mM glutamine solutions at pH 8.5, 7.5, 6.5, and 5.5 in oocytes expressing ApGLNT1 or one of its orthologs (top trace in each set), or oocytes injected with water (bottom trace in each set).  Arrowheads above traces indicate the application of ND96 at a given pH, and black bars represent co-applications of glutamine at the same pH.  Bar graphs on the right show mean response ± standard error for each pH as a percentage of the response to glutamine at pH 5.5.  Sample sizes for each trace range from 4 to 12 oocytes.

## *A. pisum* ApGLNT1
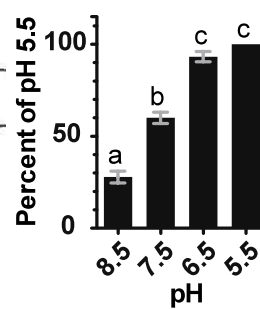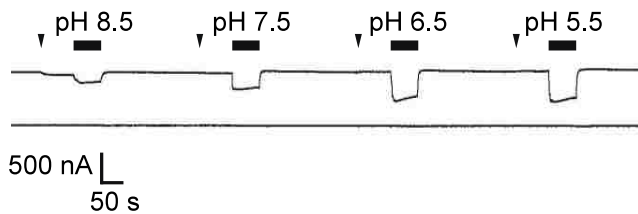


## *P. citri*
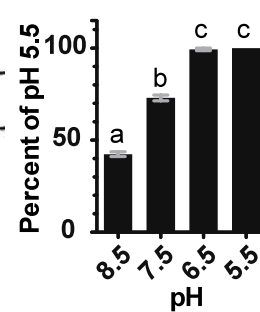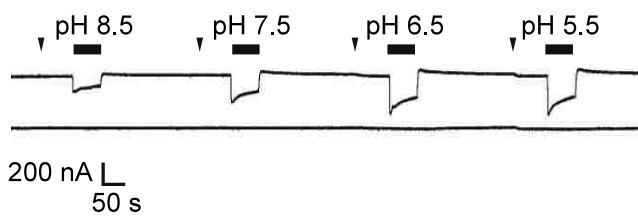


## *D. citri*
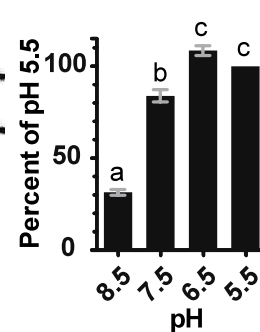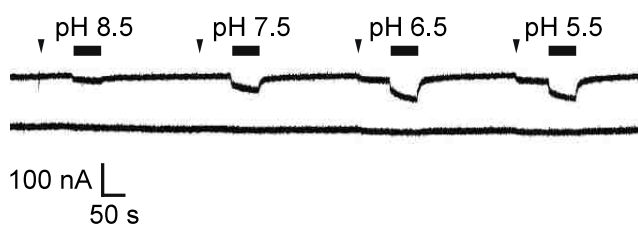


## *P. humanus*
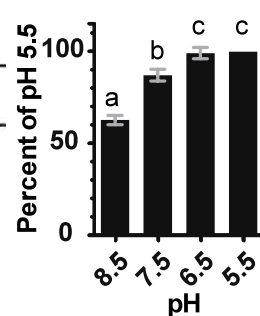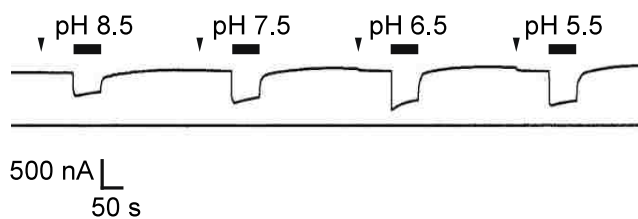


## *T. castaneum*
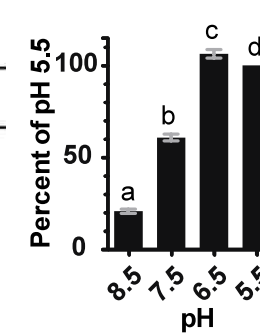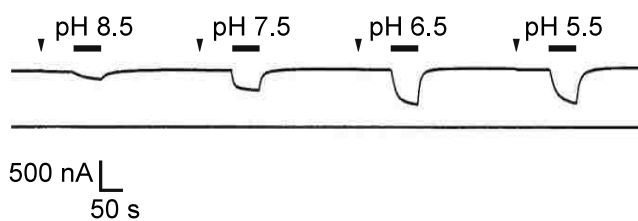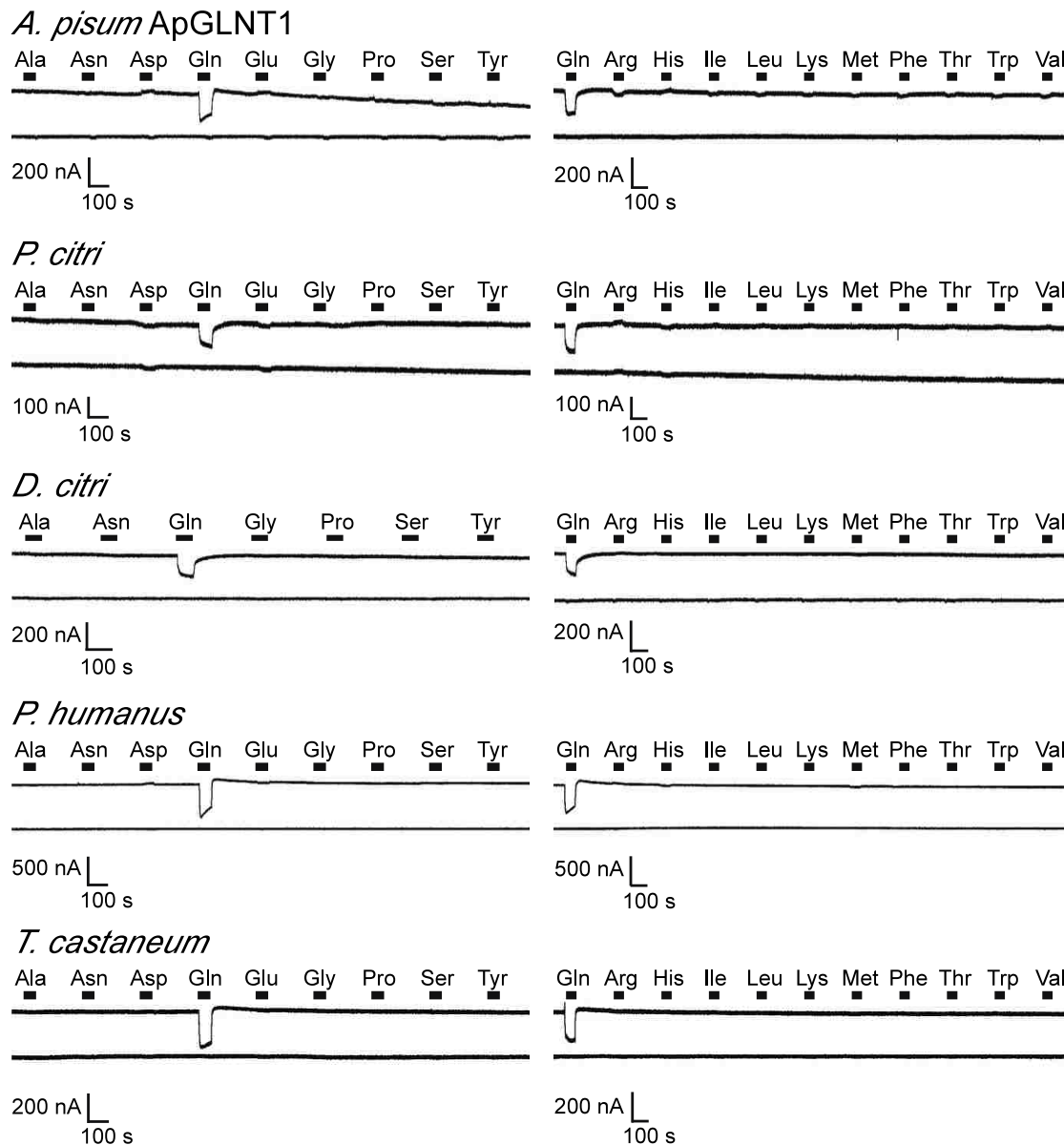
Figure 5.5. Glutamine is transported with high specificity. On the left are representative current recordings showing responses to nonessential amino acids in oocytes expressing ApGLNT1 or one of its orthologs (top trace), or oocytes injected with water (bottom trace). On the right are representative current recordings showing responses to essential amino acids and glutamine (positive control) in oocytes expressing ApGLNT1 or one of its orthologs (top trace), or oocytes injected with water (bottom trace). Amino acids were applied at 2 mM for oocytes expressing transporters from *A. pisum*, *P. citri*, *P. humanus*, and *T. castaneum*. Amino acids were applied at 10 mM for oocytes expressing the *D. citri* ortholog, with the exception of tyrosine. Tyrosine was applied at 2.5 mM because it is insoluble at higher concentrations. Amino acids omitted from screen induced comparable current deflections in both transporter expressing and water-injected oocytes. Sample sizes for each trace range from 4 to 12 oocytes.

## A. pisum ApGLNT1

EC50
Gln    0.2          0.4          1          2          4

200 nA
        200 s

0.2          0.4          1          2          4

200 nA
        100 s

## P. humanus

EC50
Gln    0.2          0.4          1          2          4

200 nA
        200 s

0.2          0.4          1          2          4

200 nA
        100 s

## P. citri

EC50
Gln    0.2          0.4          1          2          4

400 nA
        200 s

0.2          0.4          1          2          4

400 nA
        100 s

## T. castaneum

EC50
Gln    0.2          0.4          1          2          4

1000 nA
        200 s

0.2          0.4          1          2          4

1000 nA
        100 s

## D. citri

EC50
Gln    0.2          0.4          1          2          4

500 nA
        200 s

0.2          0.4          1          2          4

500 nA
        100 s
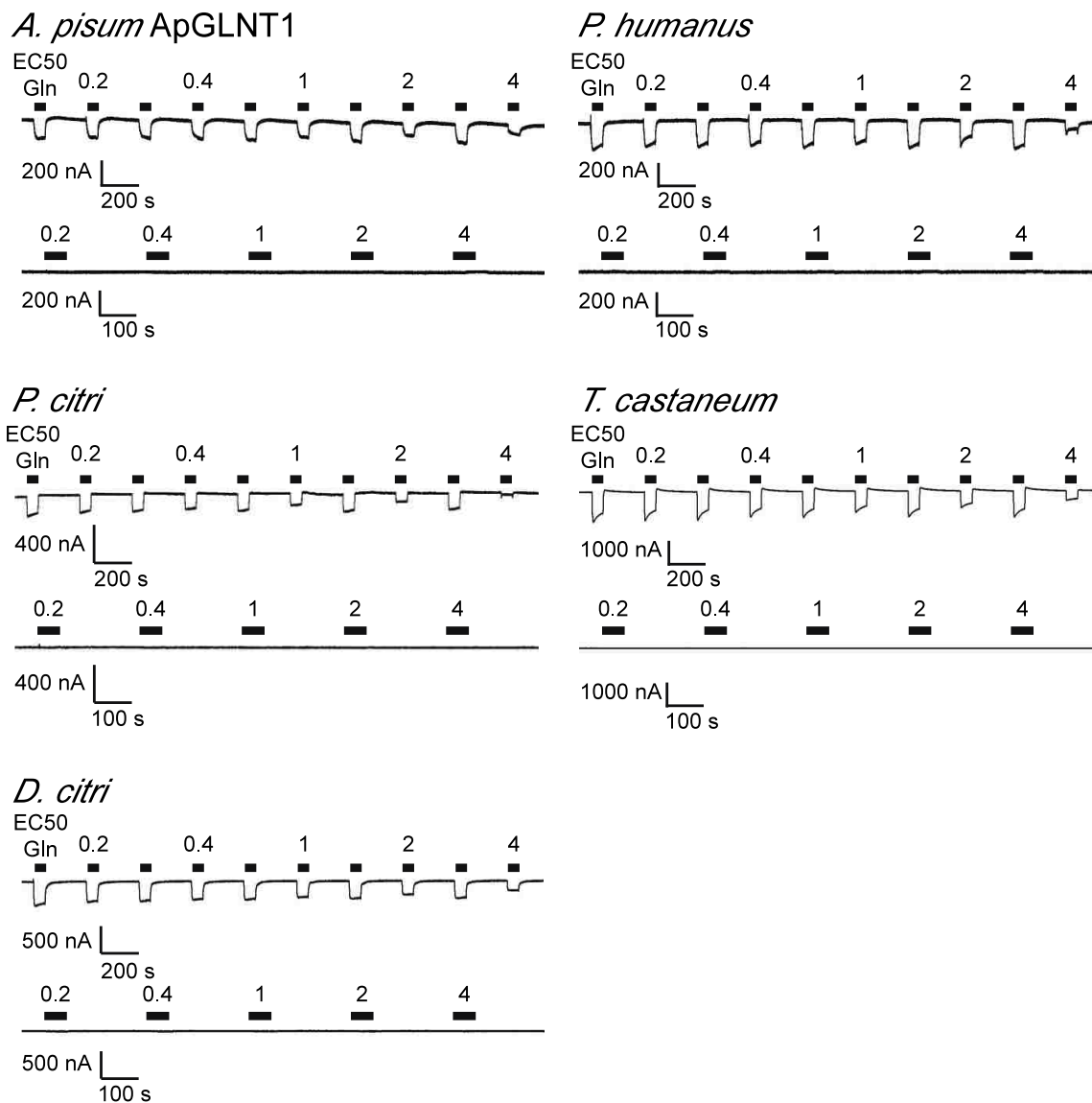
Figure 5.6. Arginine inhibits glutamine transport. Representative current recordings showing responses to alternating applications of $EC_{50}$ glutamine and $EC_{50}$ glutamine plus different concentrations of arginine (in mM) in oocytes expressing ApGLNT1 or one of its orthologs (top trace). Arginine alone at the same concentrations was applied to oocytes injected with water (bottom trace). Black bars represent the duration of each application. Sample sizes for each trace range from 5 to 12 oocytes.

**Arginine inhibition**



A. *pisum.* $IC_{50} = 5.1 \pm 0.9$ mM
P. *citri.* $IC_{50} = 2.5 \pm 0.2$ mM
D. *citri.* $IC_{50} = 6.1 \pm 0.8$ mM
P. *humanus.* $IC_{50} = 2.8 \pm 2.0$ mM
T. *castaneum.* $IC_{50} = 3.3 \pm 0.1$ mM
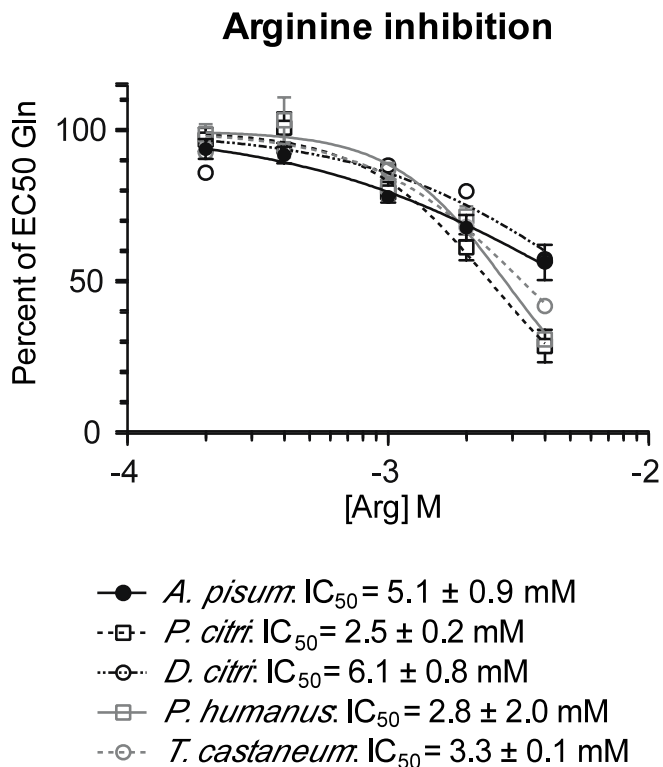
Figure 5.7. Arginine inhibition of glutamine transport is comparable in ApGLTN1 and orthologs. Inhibition curves for oocytes expressing ApGLNT1 or one of its orthologs, generated from oocyte responses to $EC_{50}$ glutamine plus a particular concentration of arginine (0.2 mM, 0.4 mM, 1 mM, 2 mM, or 4 mM) expressed as a percentage of the response to $EC_{50}$ glutamine alone. The $IC_{50}$ for each transporter is given in the key. Each point represents mean ± standard error. Sample sizes for each point range from 5 to 12 oocytes.
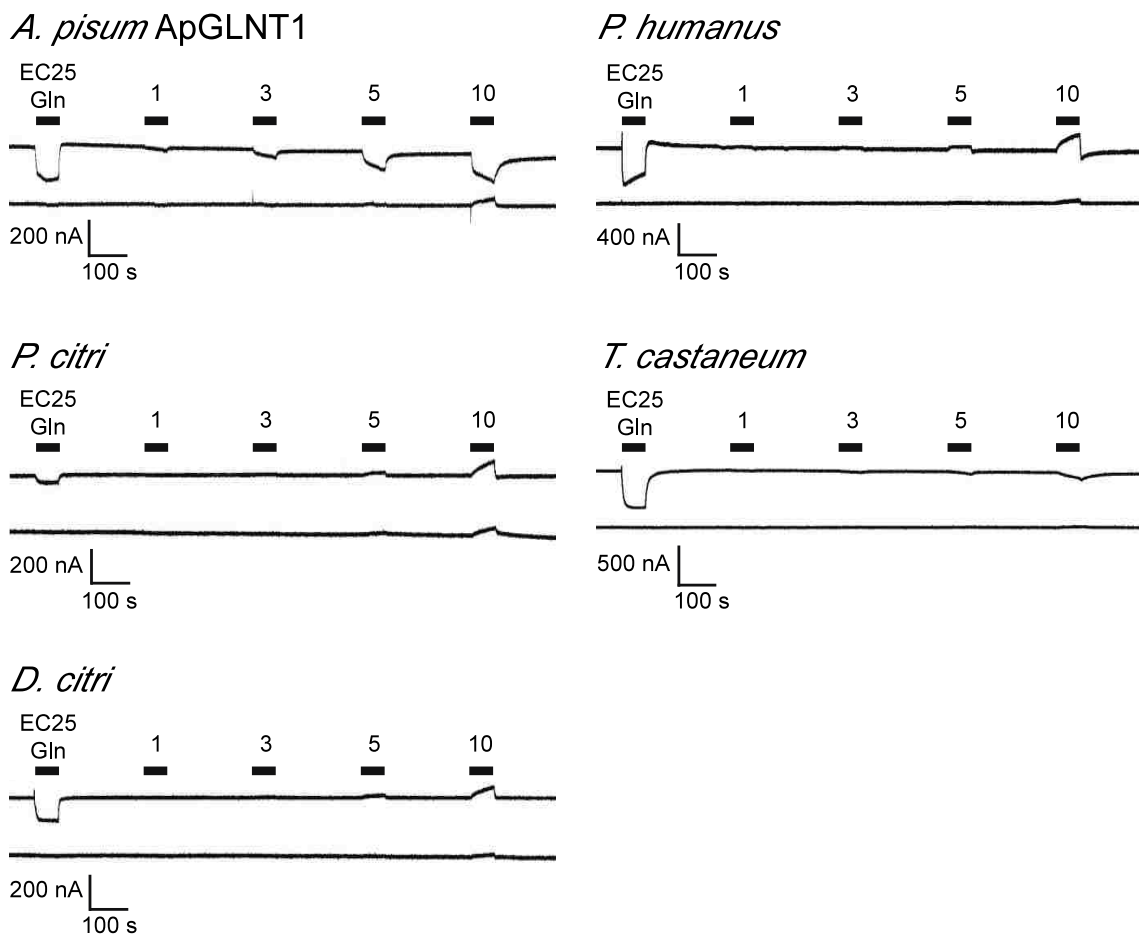
Figure 5.8. Arginine response by ApGLNT1 and orthologs. Representative current recordings showing responses a normalizer ($EC_{25}$ glutamine) to different concentrations (in mM) of arginine in oocytes expressing ApGLNT1 or one of its orthologs (top trace), or oocytes injected with water (bottom trace). Length of black bars represent length of amino acid application to oocyte. Sample sizes for each trace range from 4 to 17 oocytes.

Table 5.1. Urea cycle annotation results for *P. citri* and *D. citri*

| Gene | EC number | Query species | Query species ID | Gene present in genome?[1] | |
| | | | | *P. citri* | *D. citri* |
|---|---|---|---|---|---|
| Ornithine carbamoyltransferase | EC 2.1.3.3 | *P. humanus*[2] | PHUM252380 | No | No |
| Arginase | EC 3.5.3.1 | *D. melanogaster* | CG18104 | No | No |
| Arginosuccinate lyase | EC 4.3.2.1 | *D. melanogaster* | CG9510 | No | Yes[3] |
| Arginosuccinate synthase | EC 6.3.4.5 | *D. melanogaster* | CG1315 | No | No |
| Nitric oxide synthase | EC 1.14.13.39 | *D. melanogaster* | CG6713 | Yes | Yes |

[1] Transcript maps to genomic scaffold.

[2] Query sequence from *P. humanus* was used because the ortholog is not present in the *D. melanogaster* genome

[3] Bacterial origin. Maps to scaffold containing insect genes. See Table 2 for more information

Table 5.2. Evidence that bacterial *argH* genes are encoded in the *D. citri* genome

| Description | *D. citri* scaffold | NCBI accession | *D. melanogaster* orthologs on scaffold[1] |
|---|---|---|---|
| Argininosuccinate lyase, *argH*-like 1 | Scaffold 14 | NW_007377453.1 | CG4797 CG8234 CG31006 CG12251 CG31121 |
| | Scaffold 1907 | NW_007379333.1 | None |
| Argininosuccinate lyase, *argH*-like 2 | Scaffold 21 | NW_007377459.1 | CG8205 CG1660 CG18012 CG1618 CG16885 CG18578 CG4933 |
| | Scaffold 145 | NW_007377582.1 | CG16886 CG43227 CG10640 CG15890 CG11820 |
| | Scaffold 79909 | NW_007456984.1 | None |

[1] *A. pisum* hits also found for Scaffolds 14, 21, and 145

Table 5.3. Bacteriocyte expression of ApGLNT1 and orthologs in the Sternorrhyncha

| Species | Trinity gene ID | Bacteriocyte FPKM 1[1] | Bacteriocyte FPKM2[2] |
|---------|-----------------|------------------------|------------------------|
| *A. pisum* | TRINITY_DN69409_c1_g2 | 103.33 | 99.35 |
| *P. citri* | TRINITY_DN32903_c0_g1 | 314.87 | 278.37 |
| *D. citri* | TRINITY_DN38561_c0_g1 | 0 | N/A |

[1] Biological replicate 1: 7A for *A. pisum*, Logan, UT population for *P. citri*

[2] Biological replicate 2: 9-2-1 for *A. pisum*, Miami, FL population for *P. citri*. Not applicable for *D. citri*.

Supplemental Table 5.1. Primers used for cloning ApGLNT1 and orthologs into pCS2+

| Species | Primer | RE site | Primer sequence (5' – 3') |
|---------|--------|---------|----------------------------|
| *A. pisum* | Fwd | BamHI | TT**GGATCC**<u>ATAATGG</u>CGCATCATTTCGC |
| | Rev | XbaI | TT**TCTAGA**TCACTGATTGAACGTTACGAT |
| *P. citri* | Fwd | BGlII | TT**AGATCT**<u>ATAATGG</u>CTGACATGCAAGAAAAT |
| | Rev | XbaI | AA**TCTAGA**TCACTGGGTTGCTGCTGC |
| *D. citri* | Fwd | BGlII | TT**AGATCT**<u>ATAATGG</u>CTCACCGGTCAAATA |
| | Rev | XbaI | AA**TCTAGA**TCAGTGATGTTCCTTATTAAACCA |
| *P. humanus* | Fwd | BGlII | TT**AGATCT**<u>ATAATGG</u>CGGAAAAAGTTGAAA |
| | Rev | XbaI | TT**TCTAGA**TCATTCGCTGTTAAATGTTTCC |
| *T. castaneum* | Fwd | BGlII | TT**AGATCT**<u>ATAATGG</u>CTCGAAAATCTGGAAC |
| | Rev | XbaI | TT**TCTAGA**TTATGAACTATTAGCTCTCTCAATGAT |

Restriction sites are in **boldface**

Kozak initiation sequences are <u>underlined</u>

**Chapter 6**

Conclusion

The question of how eukaryotic hosts and their obligate symbionts evolved to be

so tightly interconnected that neither partner can exist without the other is at the forefront

of the field of symbiosis. Increasingly, the view of symbiosis is shifting from a

mutualism between independent entities to one of interconnected parts of a whole

organism, or holobiont – a view that is supported by studies revealing how hosts and

symbionts are inextricably linked developmentally and metabolically (Wilson & Duncan

2015). Based on the research I present here, an important mechanism driving metabolic

integration in sap-feeding insects and their obligate, intracellular symbionts is co-option

of amino acid transporters. Amino acid transporter co-option mediates metabolic

integration through gene duplication and single-copy transporters. Amino acid

transporters are co-opted by evolving novel bacteriocyte expression and likely through

evolution in their coding sequences as well.

I dedicated three chapters of my dissertation to exploring the role of gene

duplication in host/symbiont amino acid exchange because the mechanisms underlying

paralog retention in sap-feeding insects are complex. My dissertation was inspired by

evidence that gene duplication is mechanistically important for nutrient exchange in the *A.

pisum*/*Buchnera* symbiosis (Price et al. 2011). However, my subsequent discovery that

paralogs in the *slimfast* expansion were maintained for sex-specific roles (Chapter 2)

indicates that paralog retention does not exclusively result from selection for mediating

host/symbiont interactions. Further supporting the notion that amino acid transporter

paralogs are maintained for non-symbiotic roles, most amino acid transporter paralogs in

149

*A. pisum* (Price et al. 2011) and *P. citri* (Chapter 3) are either not significantly differentially expressed or show depleted expression in bacteriocytes relative to whole insects.  Lastly, my inference that the origin of the *slimfast* expansion did not coincide with the acquisition of *Buchnera* in aphids (Chapter 4) implies that retention of *slimfast* paralogs was not initially driven by symbiosis.

Despite evidence that selection for non-symbiotic functions facilitates maintenance of amino acid transporter paralogs, my data also supported a role for symbiosis in paralog retention.  For example, in addition to aphids, members of the three other major sternorrhynchan lineages (psyllids, whiteflies and mealybugs) convergently evolved large amino acid transporter expansions (Chapter 3) – expansions that are significantly associated with increased gene duplication rates and decreased gene loss rates (Dahan et al. 2015).  These independent expansions in sap-feeding insect lineages coupled with enriched bacteriocyte expression in some *P. citri* and *A. pisum* paralogs strongly suggest that they resulted from selection for a trait that these insects have in common, such as symbiosis.  Interestingly, independent expansion of amino acid transporters in these sternorrhynchan lineages parallels other independently evolved genomic patterns that further cement their metabolic integration with their symbionts, such as host/symbiont metabolic collaboration and acquisition of non-symbiont bacterial genes through horizontal gene transfer to the insect genome (Wilson & Duncan 2015).  Further supporting symbiosis as a trait influencing amino acid transporter paralog retention, several aphid-specific gene duplications resulted in paralogs with conserved upregulation in bacteriocytes between *A. pisum* and *M. persicae* (Chapter 4).  Conservation of bacteriocyte expression in amino acid transporters that resulted from a

duplication event coinciding with *Buchnera* acquisition suggests that duplication was followed by recruitment to the aphid/*Buchnera* symbiotic interface. More expression data for other aphid species will confirm whether bacteriocyte enrichment in aphid-specific transporters is truly conserved and associated with the evolutionary origin of symbiosis with *Buchnera*. Additionally, while expression data provide a glimpse into the functional role of amino acid transporter paralogs, functional characterization is the next step towards full understanding of their symbiotic role. Functional characterization of related paralogs with different expression profiles will further illuminate whether amino acid transporter co-option towards symbiotic roles has involved functional evolution in addition to expression evolution.

In addition to mediating metabolic integration through gene duplication, co-option has also occurred in a single-copy amino acid transporter, ApGLNT1. Price et al. (Price et al. 2014) found that ApGLNT1 regulates amino acid supply in response to demand in the *A. pisum*/*Buchnera* system. By functionally characterizing ApGLNT1 orthologs in other sap-feeding and non sap-feeding insects (Chapter 5), I discovered that ApGLNT1 is functionally conserved. Conservation of ApGLNT1 function across the insects I sampled indicates that co-option of ApGLNT1 towards a symbiotic role in *A. pisum* did not occur in response to selection for a novel, symbiotic-specific function but rather in response to selection for the ancestral role to operate in bacteriocytes. These data indicate that host/symbiont metabolic integration can evolve within the functional constraints of host genes that existed before the symbiont entered the picture. Chapter 5 also sheds light on another role for constraint in the evolution of metabolic integration mediated by ApGLNT1. Through characterization of ApGLNT1 orthologs in the citrus mealybug *P.*

*citri* and the Asian citrus psyllid *D. citri*, I inferred that ApGLNT1 recruitment to bacteriocytes likely took place in the common ancestor of aphids and mealybugs, after its divergence from psyllids. The lack of ApGLNT1 expression in *D. citri* bacteriocytes indicates that it does not function in host/symbiont integration in the *D. citri* system, possibly because *D. citri* maintains the genetic capacity to produce the ApGLNT1 inhibitor, arginine. Thus, arginine biosynthesis in *D. citri*, unlike in *P. citri* and *A. pisum*, is not necessarily coupled with import of precursor glutamine into bacteriocytes by ApGLNT1. Based on these data, I posit that the ability of sap-feeding insects to use ApGLNT1 to regulate amino acid supply in response to demand is constrained by host genomic repertoire – a pattern that has been previously identified as a signature of host/symbiont integration in insect nutritional endosymbioses (Wilson & Duncan 2015).

Gene co-option has been shown to be important in the evolution of many biological innovations, including new gene functions (True & Carroll 2002), tissues (Jandzik et al. 2015), organs (Harlin-Cognato et al. 2006), and developmental circuits (True & Carroll 2002). Here, I have shown that gene co-option also plays a role in the evolution of intimate host/endosymbiont relationships. In sap-feeding insects, metabolic integration with their endosymbionts through co-option of amino acid transporters minimally involves changes in gene expression, probably mediated by evolution in *cis*-regulatory regions. Future work will elucidate whether co-option of amino acid transporters, particularly duplicated ones, also occurs through evolution of transporter function, such as substrate recognition and selectivity. Importantly, these insights were facilitated by genomic resources for these insects that were developed before and during my dissertation. Genome and transcriptome data sets are already rapidly accumulating in

other symbiotic systems, positioning them for studies on the mechanisms of host/symbiont integration similar to those that have already been conducted in sap-feeding insects. As the stories of other systems play out, and as research continues on sap-feeding insects, we will begin to see if gene co-option emerges as a general mechanism facilitating host/symbiont integration.

## References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21(9):2104–5

Anderson CR, Jensen EO, LLewellyn DJ, Dennis ES, Peacock WJ. 1996. A new hemoglobin gene from soybean: a role for hemoglobin in all plants. *Proc Natl Acad Sci USA*. 93(12):5682–87

Arnegard ME, Zwickl DJ, Lu Y, Zakon HH. 2010. Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proc Natl Acad Sci USA*. 107:22172–77

Attardo GM, Hansen IA, Shiao S-H, Raikhel AS. 2006. Identification of two cationic amino acid transporters required for nutritional signaling during mosquito reproduction. *J Exp Biol*. 209(16):3071–78

Baumann P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59:155–89

Baumgarten S, Simakov O, Esherick LY, Liew YJ, Lehnert EM, et al. 2015. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci USA*. 112(38):11893–98

Ben Trevaskis, Watts RA, Andersson CR, Llewellyn DJ, Hargrove MS, et al. 1997. Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. *Proc Natl Acad Sci USA*. 94:12230–34

Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol*. 5(9):1675–88

Bielawski J. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J Struct Funct Genomics*. 3:201-12

Blackman R, Eastop VF. 2000. *Aphids on the World's Crops*. John Wiley & Sons, LTD. 2nd ed.

Boudko D, Kohn A, Meleshkevitch E, Dasher M, Seron T, et al. 2005. Ancestry and progeny of nutrient amino acid transporters. *Proc Natl Acad Sci USA*. 102(5):1360–65

Braendle C, Miura T, Bickel RD, Shingleton A, Kambhampati S, Stern DL. 2003. Developmental origin and evolution of bacteriocytes in the aphid-Buchnera symbiosis. *PLoS Biol*. 1(1):70–76

Brisson JA, Davis GK, Stern DL. 2007. Common genome-wide patterns of transcript accumulation underlying the wing polyphenism and polymorphism in the pea aphid (*Acyrthosiphon pisum*). *Evol Dev*. 9(4):338–46

Caillaud M, Boutin M, Braendle C, Simon J-C. 2002. A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). *Heredity*. 89:346–52

Campbell BC, Steffen-Campbell JD, Sorensen JT, Gill RJ. 1995. Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst Entomol*. 20(3):175–94

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972–73

Castagna M, Shayakul C, Trotti D, Sacchi VF, Harvey WR, Hediger MA. 1997. Molecular characteristics of mammalian and insect amino acid transporters: implications for amino acid homeostasis. *J Exp Biol*. 200(Pt 2):269–86

Cerec V, Piquet-Pellorce C, Aly HAA, Touzalin A-M, Jegou B, Bauche F. 2007. Multiple pathways for cationic amino acid transport in rat seminiferous tubule cells. *Biol Reprod*. 76(2):241–49

Cobben RH. 1979. On the Original Feeding Habits of the Hemiptera (Insecta): a Reply to Merrill Sweet. *Ann Entomol Soc Am.* 72(6):711-15

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science*. 331(6017):555–61

Colombani J, Raisin S, Pantalacci S, Radimerski T, Montagne J, Leopold P. 2003. A nutrient sensor mechanism controls *Drosophila* growth. *Cell*. 114(6):739–49

Conesa A, Gotz S, Garcia-Gomez J, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21(18):3674–76

Cryan JR, Urban JM. 2011. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Syst Entomol*. 37(1):7–21

Dahan RA, Duncan RP, Wilson ACC, Dávalos LM. 2015. Amino acid transporter expansions associated with the evolution of obligate endosymbiosis in sap-feeding insects (Hemiptera: sternorrhyncha). *BMC Evol Biol*. 15(1):52

Darwin C. 1859. *The Origin of Species*. New York: Signet Classics

Davies NB, Bourke AF, de L Brooke M. 1989. Cuckoos and parasitic ants: Interspecific brood parasitism as an evolutionary arms race. *Trends Ecol Evol*. 4(9):274–78

Deng C, Cheng C-HC, Ye H, He X, Chen L. 2010. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Natl Acad Sci USA*. 107(50):21593–98

Dohlen von CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature*. 412(6845):433–36

Dohlen von CD, Moran NA. 1995. Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J Mol Evol*. 41(2):211–23

Douglas AE. 1989. Mycetocyte symbiosis in insects. *Biol Rev Camb Philos Soc*. 64(4):409–34

Douglas AE. 1993. The nutritional quality of phloem sap utilized by natural aphid populations. *Ecol Entomol*. 18(1):31–38

Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. *J. Exp. Bot.* 57(4):747–54

Dubrovsky EB, Dubrovskaya VA, Berger EM. 2002. Juvenile hormone signaling during oogenesis in *Drosophila melanogaster*. *Insect Biochem Mol Biol.* 32(11):1555–65

Duncan RP, Feng H, Nguyen DM, Wilson ACC. 2016. Gene family expansions in aphids maintained by endosymbiotic and non-symbiotic traits. *Genome Biol Evol.* 8(3):753-64

Duncan RP, Husnik F, Van Leuven JT, Gilbert DG, Dávalos LM, et al. 2014. Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Mol Ecol*. 23(6):1608–23

Duncan RP, Nathanson L, Wilson AC. 2011. Novel male-biased expression in paralogs of the aphid *slimfast* nutrient amino acid transporter expansion. *BMC Evol Biol*. 11:253

Eberhard WG. 2000. Spider manipulation by a wasp larva. *Nature*. 406(6793):255–56

Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 23(1):205–11

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 8(9):689–98

Evans AM, Aimanova KG, Gill SS. 2009. Characterization of a blood-meal-responsive proton-dependent amino acid transporter in the disease vector, *Aedes aegypti*. *J. Exp. Biol.* 212(Pt 20):3263–71

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39(Web Server issue):W29–W37

Force A, LYNCH M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 151(4):1531–45

Forneck A, Huber L. 2009. (A)sexual reproduction - a review of life cycles of grape phylloxera, *Daktulosphaira vitifoliae*. *Entomol Exp Appl*. 131(1):1–10

Foth BJ, Tsai IJ, Reid AJ, Bancroft AJ, Nichol S, et al. 2014. Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. *Nat Genet*. 46(7):693–700

Ganot P, Moya A, Magnone V, Allemand D, Furla P, Sabourault C. 2011. Adaptations to endosymbiosis in a cnidarian-dinoflagellate association: differential gene expression and specific gene duplications. *PLoS Genet.* 7(7):e1002187

Gao X, Lu F, Zhou L, Dang S, Sun L et al. 2009. Structure and mechanism of an amino acid antiporter. *Science*. 324:1565-68

Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, et al. 2010. Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome Biol*. 11(2):R21

Gloss AD, Vassão DG, Hailey AL, Nelson Dittrich AC, Schramm K, et al. 2014. Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol Biol Evol*. 31(9):2441–56

Goberdhan DCI, Meredith D, Boyd CAR, Wilson C. 2005. PAT-related amino acid transporters regulate growth via a novel mechanism that does not require bulk transport of amino acids. *Development*. 132(10):2365–75

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11(5):725–36

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. 2015. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA*. 112(10):3026–31

Goodfellow M, Alderson G. 1977. The actinomycete-genus Rhodococcus: a home for the "rhodochrous" complex. *J Gen Microbiol*. 100(1):99–122

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 27(2):221–24

Grimaldi D, Engel MS. 2005. *Evolution of the Insects*. Cambridge University Press, New York

Gullan PJ, Downie DA, Steffan SA. 2003. A new pest species of the mealybug genus *Ferrisia* Fullaway (Hemiptera: Pseudococcidae) from the United States. *Ann Entomol Soc Am*. 96(6):723–37

Gündüz EA, Douglas AE. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings: Biological Sciences*. 276(1658):987–91

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494–1512

Hansen AK, Degnan PH. 2014. Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J*. 8(12):2490–2502

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA*. 108(7):2849–54

Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol*. 23(6):1473–96

Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol*. 6(8):2088–95

Harlin-Cognato A, Hoffman EA, Jones AG. 2006. Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish. *Proc Natl Acad Sci USA*. 103(51):19407–12

Hennig W. 1981. *Insect Phylogeny*. John Wiley & Sons

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8):754–55

Huerta-Cepas J, Marcet-Houben M, Pignatelli M, Moya A, Gabaldon T. 2010. The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrthosiphon pisum* genes. *Insect Mol Biol*. 19:13–21

Hundal HS, Taylor PM. 2009. Amino acid transceptors: gate keepers of nutrient exchange and regulators of nutrient signaling. *Am J Physiol-Endoc M*. 296(4):E603–13

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. 153(7):1567–78

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 11(2):97–108

International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*. 8(2):e1000313

International Glossina Genome Initiative. 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science*. 344(6182):380–86

Ishikawa H. 1992. Soldier and male of an eusocial aphid *Colophina arma* lack endosymbiont: implications for physiological and evolutionary interaction between host and symbiont. *J Insect Physiol.* 38(12):1033-42

Jandzik D, Garnett AT, Square TA, Cattell MV, Yu J-K, Medeiros DM. 2015. Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature*. 518(7540):534–37

Jaubert-Possamai S, Rispe C, Tanguy S, Gordon K, Walsh T, et al. 2010. Expansion of the miRNA pathway in the hemipteran insect Acyrthosiphon pisum. *Mol Biol Evol*. 27(5):979–87

Jin X, Aimanova K, Ross LS, Gill SS. 2003. Identification, functional characterization and expression of a LAT type amino acid transporter from the mosquito *Aedes aegypti*. *Insect Biochem Mol Biol*. 33(8):815–27

Karley AJ, Ashford DA, Minto LM, Pritchard J, Douglas AE. 2005. The significance of gut sucrase activity for osmoregulation in the pea aphid, *Acyrthosiphon pisum*. *J Insect Physiol*. 51(12):1313–19

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–66

Kennedy P. 2010. Ectomycorrhizal fungi and interspecific competition: species interactions, community structure, coexistence mechanisms, and future research directions. *New Phytol*. 187(4):895–910

Kim H, Lee S, Jang Y. 2011a. Macroevolutionary patterns in the Aphidini aphids (Hemiptera: Aphididae): diversification, host association, and biogeographic origins. *Plos One*. 6(9):e24749

Kim JH, Min JS, Kang JS, Kwon DH, Yoon KS, et al. 2011b. Comparison of the humoral and cellular immune responses between body and head lice following bacterial challenge. *Insect Biochem Mol Biol*. 41(5):332–39

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA*. 107(27):12168–73

Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B*. 279(1749):5048–57

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–59

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3):R25

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12(1):323

Livak K, Schmittgen T. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 25(4):402–8

Lu H-L, Price DRG, Wikramanayake A, Chang C-C, Wilson ACC. 2016. Ontogenetic differences in localization of glutamine transporter ApGLNT1 in the pea aphid demonstrate that mechanisms of host/symbiont integration are not similar in the maternal versus embryonic bacteriome. *Evodevo*. 7(1):1

Luan J-B, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, et al. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol*. 7(9):2635–47

Lynch M. 2007. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates, Inc.

Lynch M, Conery J. 2000. The evolutionary fate and consequences of duplicate genes. *Science*. 290(5494):1151–55

Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proc R Soc B*. 279(1740):2965–73

Marais Des DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*. 454(7205):762–65

Martin JF, Hersperger E, Simcox A, Shearn A. 2000. *minidiscs* encodes a putative amino acid transporter subunit required non-autonomously for imaginal cell proliferation. *Mech Dev*. 92:155–67

Mayr E. 1963. *Animal Species and Evolution*. Cambridge: Harvard University Press

McCutcheon JP, Dohlen von CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol.* 21(16):1366–72

McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci USA*. 106(36):15394–99

McCutcheon JP, Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci USA*. 104(49):19392–97

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10(1):13–26

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Loso T, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA*. 110(9):3229–36

Medina RF, Nachappa P, Tamborindeguy C. 2011. Differences in bacterial diversity of host-associated populations of *Phylloxera notabilis Pergande* (Hemiptera: Phylloxeridae) in pecan and water hickory. *J Evol Biol*. 24(4):761–71

Miller DG, Lawson SP, Rinker DC, Estby H, Abbot P. 2015. The origin and genetic differentiation of the socially parasitic aphid *Tamalia inquilinus*. *Mol Ecol*. 24(22):5751–66

Misof B, Liu S, Meusemann K, Peters RS, Donath A, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 346(6210):763–67

Miura T, Braendle C, Shingleton A, Sisk G, Kambhampati S, Stern DL. 2003. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). *J Exp Zool B Mol Dev. Evol.* 295(1):59–81

Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA*. 104 Suppl 1:8627–33

Moran NA, McLaughlin HJ, Sorek R. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*. 323(5912):379–82

Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc B*. 253(1337):167–71

Moran NA, Tran P, Gerardo NM. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl. Environ. Microbiol.* 71(12):8802–10

Morris K, Lorenzen MD, Hiromasa Y, Tomich JM, Oppert C, et al. 2009. *Tribolium castaneum* larval gut transcriptome and proteome: A resource for the study of the coleopteran gut. *J. Proteome Res.* 8(8):3889–98

Munson MA, Baumann P, Kinsey MG. 1991. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated primary endosymbionts of aphids. *Int J Syst Bacteriol*. 41(4):566–68

Myhre S, Tveit H, Mollestad T, Laegreid A. 2006. Additional gene ontology structure for improved biological reasoning. *Bioinformatics*. 22(16):2020–27

Nachappa P, Levy J, Tamborindeguy C. 2012. Transcriptome analyses of *Bactericera cockerelli* adults in response to "*Candidatus* Liberibacter solanacearum" infection. *Mol Genet Genomics*. 287(10):803–17

Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S-Y. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr Biol*. 24(14):R640–41

Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, et al. 2013. Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol*. 23(15):1478–84

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, et al. 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science*. 314(5797):267

Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T. 2011. Reductive evolution of bacterial genome in insect gut environment. *Genome Biol Evol*. 3:702–14

Nikoh N, McCutcheon JP, Kudo T, Miyagishima S-Y, Moran NA, Nakabachi A. 2010. Bacterial genes in the aphid genome: absence of functional gene transfer from Buchnera to its host. *PLoS Genet*. 6(2):e1000827

Nikoh N, Nakabachi A. 2009. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol*. 7:12

Nitecki MH, ed. 1990. *Evolutionary Innovations*. Chicago: University of Chicago Press

Nováková E, Hypša V, Klein J, Foottit RG, Dohlen von CD, Moran NA. 2013. Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Mol Phylogenet Evol*. 68(1):42–54

Nygaard S, Zhang G, Schiøtt M, Li C, Wurm Y, et al. 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res*. 21(8):1339–48

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*. 24(4):581–83

Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag

Ott M, Zola J, Stamatakis A, Aluru S. 2007. Large-scale maximum likelihood-based phylogenetic analysis on the IBM BlueGene/L. *Sc '07*, p. 1. New York, New York, USA: ACM Press

Pigliucci M. 2008. What, if anything, is an evolutionary novelty? *Philos of Sci*. 75(5):887–98

Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, et al. 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Mol Cell Proteomics*. 10(6):M110.007039

Price DRG, Duncan RP, Shigenobu S, Wilson ACC. 2011. Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface. *Mol Biol Evol*. 28(11):3113–26

Price DRG, Feng H, Baker JD, Bavan S, Luetje CW, Wilson ACC. 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proc Natl Acad Sci USA*. 111(1):320–25

Price DRG, Tibbles K, Shigenobu S, Smertenko A, Russell CW, et al. 2010. Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional characterization. *Insect Mol Biol*. 19:97–112

Price DRG, Wilkinson HS, Gatehouse JA. 2007. Functional expression and characterisation of a gut facilitative glucose transporter, NlHT1, from the phloem-feeding insect *Nilaparvata lugens* (rice brown planthopper). *Insect Biochem Mol Biol*. 37(11):1138–48

Price DRG, Wilson ACC. 2014. A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol.* 12(1):110

Price DRG, Wilson ACC, Luetje CW. 2015. Proton-dependent glutamine uptake by aphid bacteriocyte amino acid transporter ApGLNT1. *Biochim Biophys Acta.* 1848(10 pt A):2085-91

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v.1.6. http://beast.bio.ed.ac.uk/Tracer

Ramsey JS, Wilson ACC, De Vos M, Sun Q, Tamborindeguy C, et al. 2007. Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics*. 8:423

Ratzka C, Gross R, Feldhaar H. 2013. Gene expression analysis of the endosymbiont-bearing midgut tissue during ontogeny of the carpenter ant *Camponotus floridanus*. *J Insect Physiol*. 59(6):611–23

Redak RA, Purcell AH, Lopes JRS, Blua MJ, Mizell RF, Andersen PC. 2004. The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. *Annu Rev Entomol*. 49:243–70

Ribeiro J, Genta FA, Sorgine M, Logullo R, Mesquita RD, et al. 2014. An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. *PLoS Negl Trop Dis*. 8(1):e2594

Rider SD Jr, Srinivasan DG, Hilgarth RS. 2010. Chromatin-remodelling proteins of the pea aphid, *Acyrthosiphon pisum* (Harris). *Insect Mol Biol*. 19:201–14

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26(1):139–40

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19(12):1572–74

Russell CW, Bouvaine S, Newell PD, Douglas AE. 2013. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl Environ Microbiol.* 79(19):6117–23

Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc R Soc B*. 273(1586):603–10

Russell JA, Moreau CS, Goldman-Huertas B, Fujiwara M, Lohman DJ, Pierce NE. 2009. Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc Natl Acad Sci USA*. 106(50):21236–41

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, et al. 2012a. Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl Environ Microbiol.* 78(1):204–10

Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. 2012b. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects. *Environ. Microbiol.*

Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci.* 106(46):19521–26

Saier MH. 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev.* 64(2):354–411

Saier MH, Yen MR, Noto K, Tamang DG, Elkan C. 2009. The Transporter Classification Database: recent advances. *Nucleic Acids Res.* 37(Database issue):D274–78

Saier MHJ, Tran CV, Barabote RD. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34(Database issue):D181–86

Salem H, Bauer E, Strauss AS, Vogel H, Marz M, Kaltenpoth M. 2014. Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proc R Soc B.* 281(1796):20141838

Sandoval CP. 1994. Differential visual predation on morphs of *Timema cristinae* (Phasmatodeae:Timemidae) and its consequences for host range. *Biol J Linn Soc.* 52(4):341–56

Sandstrom J, 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(11):947–55

Sandström JP, Moran NA. 2001. Amino acid budgets in three aphid species using the same host plant. *Physiol Entomol.* 26(3):202–11

Sasaki T, Ishikawa H. 1995. Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrthosiphon pisum*. *J Insect Physiol.* 41(1):41–46

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6(5):526–38

Schemske DW, Bradshaw HD. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc Natl Acad Sci USA.* 96(21):11910–15

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 28(8):1086–92

Schulze W, Weise A, Frommer WF, Ward, JM. 2000. Function of the cytosolic N-terminus of sucrose transporter AtSUT2 in substrate affinity. *FEBS Lett*. 485:189-94

Shaffer PL, Goehring A, Shankaranarayanan A, Gouaux E. 2009. Structure and mechanism of a Na$^+$-independent amino acid tranpsorter. *Science*. 325:1010-14

Shigenobu S, Bickel RD, Brisson JA, Butts T, Chang CC, et al. 2010. Comprehensive survey of developmental genes in the pea aphid, *Acyrthosiphon pisum*: frequent lineage-specific duplications and losses of developmental genes. *Insect Mol Biol*. 19:47–62

Shigenobu S, Stern DL. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc R Soc B*. 280(1750):20121952

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*. 407(6800):81–86

Shigenobu S, Wilson ACC. 2011. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cell Mol Life Sci*. 68(8):1297–1309

Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, et al. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature*. 476(7360):320–23

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. *Nature*. 457(7231):818–23

Sloan DB, Moran NA. 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology Letters*. 8(6):986–89

Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, et al. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol*. 31(4):857–71

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 3(1):Article3

Smyth GK. 2005. Limma linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, eds. R Gentleman, V Carey, W Huber, R Irrizary, S Duboit, pp. 397–420. New York: Springer

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. *Methods*. 31(4):265–73

Song N, Liang A-P, Bu C-P. 2012. A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. *Plos One*. 7(11):e48778

Spitz F, Gonzalez F, Peichel C, Vogt TF, Duboule D, Zákány J. 2001. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev*. 15(17):2209–14

Srinivasan DG, Fenton B, Jaubert-Possamai S, Jaouannet M. 2010. Analysis of meiosis and cell cycle genes of the facultatively asexual pea aphid, *Acyrthosiphon pisum* (Hemiptera: Aphididae). *Insect Mol Biol*. 19 Suppl 2:229–39

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21):2688–90

Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 26(12):1569–71

Thao ML, Baumann P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl Environ Microbiol*. 70(6):3401–6

Thao ML, Gullan PJ, Baumann P. 2002. Secondary (gamma-Proteobacteria) endosymbionts infect the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl Environ Microbiol*. 68(7):3190–97

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl Environ Microbiol*. 66(7):2898–2905

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22(22):4673–80

Thwaites DT, Anderson CMH. 2011. The SLC36 family of proton-coupled amino acid transporters and their potential role in drug transport. *Br J Pharmacol*. 164(7):1802–16

True JR, Carroll SB. 2002. Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol*. 18(1):53–80

Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 496(7443):57–63

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, et al. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol*. 10(12):e1001446

Vorwerk S, Martinez-Torres D, Forneck A. 2007. *Pantoea agglomerans*-associated bacteria in grape phylloxera (*Daktulosphaira vitifoliae*, Fitch). *Agric For Entomol*. 9(1):57–64

Wagner GP, Lynch VJ. 2010. Evolutionary novelties. *Curr Biol*. 20(2):R48–R52

Wang X-W, Luan J-B, Li J-M, Bao Y-Y, Zhang C-X, Liu S-S. 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*. 11:400

Weber MG, Agrawal AA. 2012. Phylogeny, ecology, and the coupling of comparative and experimental approaches. *Trends Ecol Evol*. 27(7):394–403

West-Eberhard MJ. 2008. Toward a modern revival of Darwin's theory of evolutionary novelty. *Philos of Sci*. 75(5):899–908

Wheeler D. 1996. The role of nourishment in oogenesis. *Annu Rev Entomol*. 41:407–31

White J, Strehl CE. 1978. Xylem feeding by periodical cicada nymphs on tree roots. *Ecol Entomol*. 3(4):323–27

Whiteman NK, Gloss AD, Sackton TB, Groen SC, Humphrey PT, et al. 2012. Genes involved in the evolution of herbivory by a leaf-mining, Drosophilid fly. *Genome Biol Evol*. 4(9):900–916

Wigglesworth V. 1960. Nutrition and reproduction in insects. *Proc Nutr Soc*. 19:18-23

Wilkinson TL, Douglas AE. 2003. Phloem amino acids and the host plant range of the polyphagous aphid, Aphis fabae. *Entomol Exp Appl*. 106(2):103–13

Wilson ACC, Ashton PD, 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 Suppl 2:249–58

Wilson ACC, Duncan RP. 2015. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc Natl Acad Sci USA*. 112(33):201423305–10261

Wilson ACC, Sunnucks P, Hales D. 2003. Heritable genetic variation and potential for adaptive evolution in asexual aphids (Aphidoidea). *Biol J Linn Soc*. 79(1):115–35

Wolpert L, Tickle C, Jessell T, Lawrence P, Meyerowitz E, et al. 2011. *Principles of Development*. Oxford: Oxford University Press. 4 ed.

Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, et al. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol*. 4(6):e188–092

Wullschleger S, Loewith R, Hall M. 2006. TOR signaling in growth and metabolism. *Cell*. 124(3):471–84

Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15(5):568–73

Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–91

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, et al. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*. 480(7378):520–24

Zdobnov EM, Apweiler R. 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 17(9):847–48

Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 18(5):821–29

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol*. 18(6):292–98

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature*. 450(7167):233–8

Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*. 4(4):259–63

Zrzavy J. 1990. Evolution of Hemiptera: An attempt at synthetic approach. *Proc Sixth Internat Symp Scale Insects Studies*. Cracow, August 6-12, 1990. Agric Univ Press: 19-22

Zrzavy J. 1992. Evolution of antennae and historical ecology of the hemipteran insects (Paraneoptera). *Acta Entomol Bohemoslov*. 89:77–86