Genome Expansion and Differential Expression of Amino Acid Transporters at the Aphid/Buchnera Symbiotic Interface

Daniel R. G. Price,†1 Rebecca P. Duncan,†1 Shuji Shigenobu,2,3 and Alex C. C. Wilson*1

1Department of Biology, University of Miami, Coral Gables, Florida
2National Institute for Basic Biology Core Research Facilities, National Institute for Basic Biology, Myodaiji, Okazaki, Japan
3PRESTO, Japanese Science and Technology Agency, Kawaguchi, Saitama, Japan
†These authors contributed equally to the manuscript.
*Corresponding author: E-mail: acwilson@bio.miami.edu.
Associate editor: Jennifer Wernegreen

Abstract

In insects, some of the most ecologically important symbioses are nutritional symbioses that provide hosts with novel traits and thereby facilitate exploitation of otherwise inaccessible niches. One such symbiosis is the ancient obligate intracellular symbiosis of aphids with the γ-proteobacteria, Buchnera aphidicola. Although the nutritional basis of the aphid/Buchnera symbiosis is well understood, the processes and structures that mediate the intimate interactions of symbiotic partners remain uncharacterized. Here, using a de novo approach, we characterize the complement of 40 amino acid polyamine organocation (APC) superfamily member amino acid transporters (AATs) encoded in the genome of the pea aphid, Acyrthosiphon pisum. We find that the A. pisum APC superfamily is characterized by extensive gene duplications such that A. pisum has more APC superfamily transporters than other fully sequenced insects, including a ten paralog aphid-specific expansion of the APC transporter slimfast. Detailed expression analysis of 17 transporters selected on the basis of their phylogenetic relationship to five AATs identified in an earlier bacteriocyte expressed sequence tag study distinguished a subset of eight transporters that have been recruited for amino acid transport in bacteriocyte cells at the symbiotic interface. These eight transporters include transporters that are highly expressed and/or highly enriched in bacteriocytes and intriguingly, the four AATs that show bacteriocyte-enriched expression are all members of gene family expansions, whereas three of the four that are highly expressed but not enriched in bacteriocytes retain one-to-one orthology with transporters in other genomes. Finally, analysis of evolutionary rates within the large A. pisum slimfast expansion demonstrated increased rates of molecular evolution coinciding with two major shifts in expression: 1) a loss of gut expression and possibly a gain of bacteriocyte expression and 2) loss of expression in all surveyed tissues in asexual females. Taken together, our characterization of nutrient AATs at the aphid/Buchnera symbiotic interface provides the first examination of the processes and structures operating at the interface of an obligate intracellular insect nutritional symbiosis, offering unique insight into the types of genomic change that likely facilitated evolutionary maintenance of the symbiosis.

Key words: AAAP, APC, bacteriocyte, gene duplication, slimfast, symbiosis.

Introduction

Metazoans cannot synthesize nine essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) that are all required for normal growth and development and usually supplemented as a dietary component. A tenth amino acid, arginine, may also be considered essential, as demand for this amino acid is often not met by de novo biosynthesis and must also be supplemented as a dietary component. Insects belonging to the order Hemiptera, in particular, aphids, whiteflies, and planthoppers, face a unique nutritional challenge as they are specialist plant phloem-sap feeders and thus restricted to a diet with limiting amounts of essential amino acids (Sandström and Pettersson 1994; Karley et al. 2002). Despite these nutritional limitations, insect-phloem feeders are able to rapidly establish large populations on suboptimal diets. This in part can be attributed to nutritional symbiotic associations with bacteria that are able to metabolically compensate for the nutrient poor diet of the host. In nature, the nutritional association of insects with bacterial endosymbionts is widespread and is a large contributing factor to the success of many insects in exploiting otherwise inaccessible environmental niches (Wu et al. 2006; International Aphid Genomics Consortium 2010; Kirkness et al. 2010). The most extensively studied of the insect/bacterial nutritional symbioses is that of the pea aphid, Acyrthosiphon pisum and its γ-proteobacterial endosymbiont, Buchnera aphidicola (Moran et al. 2003). Experimental evidence, and more recently whole-genome sequence analysis of A. pisum and Buchnera, reveals an intimate metabolic collaboration between partners (Wilson et al. 2010), which fully meets the amino acid requirements of the holosymbiont (reviewed in Shigenobu and Wilson 2011).

Core to all Hemipteran insect nutritional symbioses is the provisioning of essential amino acids to the host by...
a bacterial symbiont, and in this regard, *Buchnera* is no exception (Moran et al. 2008; Shigenobu and Wilson 2011). Large populations of *Buchnera* are densely packed and housed in specialized aphid bacteriocyte cells. Adult female aphids contain between 60 and 80 bacteriocytes that are grouped to form a defined structure called the bacteriome, which is located in the aphid haemoceol (McLean and Houk 1973; Baumann et al. 1995). Within bacteriocytes, individual *Buchnera* cells, that number in the thousands (Mira and Moran 2002), are partitioned from the bacteriocyte cytoplasm by an aphid-derived membrane called the symbiosomal membrane (Baumann et al. 1995). This compartmentalization of *Buchnera* implies that supply of nonessential amino acids to *Buchnera* and return of essential amino acids to aphids is dependent on transporters that function at three membrane barriers: 1) the bacteriocyte plasma membrane, 2) the symbiosomal membrane, and 3) *Buchnera*’s Gram-negative envelope (Shigenobu and Wilson 2011). Collectively we refer to these three membrane barriers as the aphid/*Buchnera* symbiotic interface; an interface that includes two, likely differentiated, aphid membrane barriers: the bacteriocyte plasma membrane and the symbiosomal membrane.

The molecular identities of aphid amino acid transporters (AATs) remain largely unknown but are anticipated to share common features with AATs of other species. Eukaryotic nutritional AATs, with few known exceptions, belong to the amino acid polyamine/organocation (APC) superfamily and are members of the amino acid/polyamine/organocation (APC) family (TC # 2.A.3) or the eukaryotic-specific amino acid/auxin permease (AAAP) family (TC # 2.A.18) (Castagna et al. 1997; Saier 2000b; Saier et al. 2006; Saier et al. 2009). Transports belonging to the APC family catalyze the uniport, symport, and antiport of a broad range of substrates across lipid membranes (Saier 2000a), whereas transporters belonging to the divergent AAAP family (Jack et al. 2000) have varying substrate specificities, catalyzing the symport of auxin (indole-3-acetic acid), single amino acids, or multiple amino acids across lipid membranes (Young et al. 1999; Saier 2000a). Transports from both families are integral membrane proteins and share a common topology, consisting of hydrophobic transmembrane (TM) domains that are linked by short hydrophilic loop regions that alternate from extracellular to intracellular positions. Typically, most APC transporters contain 12–14 TM domains (Jack et al. 2000), whereas AAAP transporters are shorter, usually with 11 TM domains (Young et al. 1999). Both transporter families function as secondary transporters, which couple substrate transport with an electrochemical gradient across the cell membrane. The driving force for transport is usually established in the cell by primary active transport processes that use ATP to establish a proton motive force (PMF) or sodium-ion motive force (SMF) across the cell membrane (Wolfersberger 1994).

Transcript profiling of isolated *A. pisum* bacteriocytes identified, on the basis of sequence homology to known insect transporters, two putative AATs: 1) aphid cationic amino acid transporter 1 (CAT1, ACYPI000584), represented by one expressed sequence tag (EST) and 2) CAT2 (ACYPI005118) enriched in bacteriocytes approximately 93.3-fold compared with whole insect expression levels and represented by 27 ESTs (Nakabachi et al. 2005). Publication of the complete genome sequence of *A. pisum* positions us to extend this initial study of Nakabachi et al. (2005), to identification of the complete set of aphid nutritional AATs, and to begin dissection of their organization and function. Thus, the goal of the present study was to identify and annotate the complement of nutritional AATs in *A. pisum* and distinguish the subset of transporters that have been recruited for amino acid transport at the symbiotic interface. In doing this, we found that *A. pisum* has a larger set of nutrient AATs than any other insect with a completely sequenced genome. Although the largest *A. pisum* expansion was identified in the APC family, AATs from both families have experienced aphid-specific duplication events that include examples of apparent specialization following gene duplication. Finally, analysis of evolutionary rates within the large APC family expansion found increased rates of molecular evolution along two lineages where major shifts occurred in gene expression, one lineage where gut expression was lost and bacteriocyte expression may have been gained and a second lineage where a global expression was lost in asexual females.

**Materials and Methods**

De novo Identification of Amino Acid Transporters *Acyrthosiphon pisum* Official Gene Set 1.0 (OGS 1.0) comprises 10,239 NCBI RefSeq gene models and 24,355 ab initio GLEAN models predicted from genome assembly Acyr 1.0 (ABLFO0000000.1, International Aphid Genomics Consortium 2010). OGS 1.0 and the 209,931 *A.pisum* ESTs available in NCBI’s dbEST September 2010 were used as a basis for the bioinformatic analyses described here. First, the OGS 1.0 protein data set was searched for proteins containing conserved sequence motifs for APC (T.C # 2.A.3) transporters (Jack et al. 2000; Saier 2000a; Saier et al. 2009) and AAAP (T.C # 2.A.18) transporters (Young et al. 1999; Saier 2000a; Saier et al. 2009) as defined by Pfam HMM (hidden Markov model) profile AA_permease, PF00324 and Aa_trans, PF01490, respectively (available at http://pfam.sanger.ac.uk/, Finn et al. 2008) using hmmsearch program, which is a part of the HMMER package (version 2.3.2) (Eddy 1998). Following de novo HMMER identification, the aphid phylome, which is a complete phylogenetic analysis of the aphid genome (http://www.phylomedb.org, Huerta-Cepas et al. 2010), was searched for any additional homologs. Finally, to ensure that identification of AATs in the *A. pisum* genome was as complete as possible, all known *Drosophila melanogaster* AATs were used in BlastP and TBLastN searches against Acyr 1.0 predicted proteins and genomic scaffolds, respectively. Following computational identification, we gathered all available EST and full-length transcript support for our gene models using BlastN searches against
all *A. pisum* NCBI dbEST data sets and the full-length transcript data set of Shigenobu et al. (2010). The complete list of 40 AATs identified using the above pipeline together with supporting transcriptomic data are listed in supplementary table S1, Supplementary Material online. Following de novo identification, all 40 putative AATs were mapped to *A. pisum* genomic scaffolds from assembly Acyr 2.0 (ABL00000000, available at NCBI in September 2010).

To facilitate comparative analyses and investigation of the molecular evolution of *A. pisum* AAT duplications, we searched the predicted protein sequences from the whole genomes of 16 taxa that included multiple insects as well as representative noninsect arthropods, vertebrates, a nematode, a plant, and yeast for APCs and AAAPs using HMMER 2.3.2 as described above for *A. pisum* Acyr 1.0. The 16 taxa included: *Anopheles gambiae*, *Apis mellifera*, *Arabidopsis thaliana*, *Bombyx mori*, *Caenorhabditis elegans*, *Danio rerio*, *Daphnia pulex*, *D. melanogaster*, *Gallus gallus*, *Homo sapiens*, *Ixodes scapularis*, *Mus musculus*, *Nasonia vitripennis*, *Pediculus humanus*, *Saccharomyces cerevisiae*, and *Tribolium castaneum* (supplementary table S3, Supplementary Material online).

**Acyrthosiphon pisum** Amino Acid Transporter Gene Annotation

Gene models with partial sequence coverage were validated by DNA sequencing, resulting in minor changes to the predicted genes. Briefly, full-length coding sequences were amplified from whole asexual female *A. pisum* cDNA and checked by Sanger sequencing using standard ABI Big Dye terminator v3.1 reactions. Reaction products were analyzed on 3130xl genetic analyzer (ABI), and sequence data assembled into contiguous sequences using Sequencher (version 4.9). All sequence data have been deposited in GenBank with accession numbers given in supplementary table S1, Supplementary Material online. Sequence differences usually resulted from either silent nucleotide polymorphisms or incorrect prediction of intron/exon boundaries in regions of the gene without EST support. Gene models have been amended using APOLLO gene annotation tools available at AphidBase (http://www.aphidbase.com, cited 2011 Jun 2; Legeai et al. 2010).

**Molecular Phylogenetic Analysis**

Amino acid sequence similarity between APC and AAAP family members is very low, and therefore, phylogenetic reconstruction of each family was performed in separate analyses. Phylogenetic reconstruction included transporter family sequence homologs identified by HMMER searches (as described above) from *H. sapiens*, *I. scapularis* and the following insect taxa *A. mellifera*, *T. castaneum*, *D. melanogaster*, and *P. humanus*, which were specifically chosen to include diverse insect taxa, representing Hymenoptera, Coleoptera, Diptera, and Phthiraptera. Functionally characterized *Aedes aegypti* transporters (APCs: AeaLAT (Jin et al. 2003), AaSlif (Attardo et al. 2006), and AaiCAT2 (Attardo et al. 2006) and AAAP: AeaPAT1 (Evans et al. 2009)) were also included in the analysis. Outgroups for the APC family included two *A. pisum* transporters from a clade of predicted Na-K-Cl transporters (supplementary table S1, Supplementary Material online), which we resolved as sister to the APC family in preliminary analyses, whereas outgroup transporters for the AAAP family belonged to the solute-sodium symporter family (T.C #2.A.21), which is a member of the Amino Acid/Polyamine/Organocation (APC) superfamily. Prior to alignment, the highly variable N and C terminal residues of the transporters were trimmed from sequences. Following trimming, the remaining sequences were aligned using MAFFT v.6.818b (Katoh et al. 2005), and the alignments checked by eye. Phylogenies for both families were reconstructed using maximum likelihood (ML) and Bayesian inference. ML analyses were conducted in GARLI (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006) using a grid computing strategy (Cummings and Huskamp 2005) through a grid service developed by Bazinet et al. (2007); available at http://www.molecularevolution.org/. GARLI analyses were conducted using default parameters and the amino acid substitution model WAG+I+I-F, as identified in ProtTest v. 1.3 (Abascal et al. 2005). Bayesian analyses were conducted using MrBayes v.3.1.2 (Huelsenbeck and Ronquist 2001) on the TeraGrid supercomputer through the CIPRES web portal v. 3.0 (http://www.phylo.org/sub_sections/portal/). We applied MrBayes default parameters and priors and ran chains until the standard deviation of split frequencies between runs converged by falling below 0.01. Finally, in order to check for appropriate sampling during parameter optimization across trees, we visually inspected trace files in Tracer v.1.4 (written by Andrew Rambaut and Alexei Drummond, available at http://tree.bio.ed.ac.uk/software/tracer/). Tracer was also used to determine burn-in values to construct the majority rule Bayesian consensus tree.

**Expression Analysis**

Real-time quantitative polymerase chain reaction (QPCR) was performed on *A. pisum* cDNA from bacteriocytes, gut and head (including salivary glands) tissues, and whole parthenogenetic adult females of three *A. pisum* lineages that included LSR1 (Caillaud et al. 2002, the parental line of the inbred line used in the Aphid Genomics Consortium 2010), 5A (Sandstrom et al. 2002), and 9-2-1 (Russell and Moran 2006). Tissues were dissected from 100 parthenogenetic females in lineages that included LSR1 (Caillaud et al. 2002, the parental line of the inbred line used in the *A. pisum* genome project; International Aphid Genomics Consortium 2010), 5A (Sandstrom and Moran 2001), and 9-2-1 (Russell and Moran 2006). Tissues were dissected from 100 parthenogenetic females in nuclease-free 0.9% saline and immediately lysed in TRI Reagent (Ambion). Total RNA was isolated and further purified by RNA cleanup with an RNeasy mini kit (Qiagen), which included a DNaseI digest. Whole insect total RNA was isolated from ten wingless adult females using an RNeasy mini kit and on-column RNase-free DNase treatment. First-strand cDNAs were synthesized from 500 ng total RNA from whole *A. pisum*, and head and gut tissues and 100 ng of total RNA from bacteriocyte tissues, using qScript cDNA SuperMix (Quanta Biosciences) in standard 20 μl reactions.

AAT gene expression was compared in different tissues using 2^ΔΔCT methodology (Livak and Schmittgen 2001), with transporter expression across tissues normalized to...
glyceraldehyde-3-phosphate dehydrogenase (GAPDH, ACYPI009769) and reported in each tissue relative to whole insect expression. Primers were designed using Primer3Plus software for real-time PCR (Untergasser et al. 2007), and specificity checked using Primer-BLAST (available at NCBI) against the A. pism Refseq gene models and Acyr 1.0. For comparative gene expression analysis, all QPCR primers had amplification efficiencies >90% and <105%. All primer sequences and amplification efficiencies are given in supplementary table S2, Supplementary Material online. Each experiment was performed in triplicate and included no template controls and no reverse transcription controls with each run. Ten microtiter reactions comprised 1 × Perfecta SYBR Green FastMix (Quanta Biosciences), 200 nM of forward and reverse primers, and cDNA derived from 8 ng total RNA (for whole insect, head and gut tissues) or cDNA derived from 1.6 ng total RNA for bacteriocytes. PCR reactions were performed on a Mastercycler ep realplex4 real-time PCR system (Eppendorf); thermal cycling conditions were 95 °C 5 min, followed by 40 cycles at 95 °C for 15 s, 52 °C for 15 s, and 60 °C for 20 s. Analysis of amplification profiles and melt curves was performed using Mastercycler ep realplex software (version 1.5). Finally, gene expression data for both APC and AAAP transporters were collectively normalized, allowing comparison of expression profiles across both families. The normalized expression for each transporter gene from LSR1 was compiled into the heat map presented in figure 4 (z-score × 10, [z = {each value—average}/standard deviation]). Yellow: z > 0, blue z < 0, and black z = 0.

Molecular Evolution Analysis
Rates of molecular evolution were estimated for a subset of A. pism APC transporters that form an A. pism-specific slimfast expansion. The analysis was conducted using the ten A. pism paralogs that form the APC family expansion (ACYPI003523, ACYPI005156, ACYPI003240, ACYPI005337, ACYPI005846, ACYPI005472, ACYPI005118, ACYPI008904, ACYPI003232, and ACYPI002633), insect slimfast orthologs from A. aegypti (AaSlif, AA773699), D. melanogaster (slimfast, CG11128), T. castaneum (XP_967023), A. mellifera (XP_393144), and P. humanus (PHUM258130), and outgroup sequences from the sister clade of the expansion (A. pism, ACYPI003923 and D. melanogaster, CG5535).

The amino acid alignment of these 17 sequences was generated in MAFFT (Katoh et al. 2005) and used to reconstruct the phylogeny of the genes in MrBayes v.3.1.2. Bayesian trace plots were analyzed, and the burn-in value determined as described above and the resulting majority rule consensus tree used as input for the molecular evolution analyses.

Full-length nucleotide coding sequences (aligned in MAFFT) were used to estimate the difference in nonsynonymous (dN) and synonymous (dS) substitution rates, measured by the ratio dN/dS (also denoted ω). The strength and type of selection in protein coding sequences are reflected by ω such that excess synonymous substitutions (ω < 1) indicate that selection favors substitutions that conserve the amino acid sequence (purifying selection), whereas equal synonymous and nonsynonymous substitution rates (ω = 1) illustrate relaxed selective constraints to retain the amino acid sequence (neutral selection) and excess nonsynonymous substitutions (ω > 1) indicate that selection favors substitutions that change the amino acid sequence (positive selection).

We analyzed branch-based and site-based models to detect lineages under accelerated rates of evolution and to detect specific sites under positive selection. For both model types, ω was calculated in the aphid slimfast expansion by ML in the PAML package (Yang 2007). Models implemented in PAML do not accommodate gene conversion or recombination (Anisimova et al. 2003; Casola and Hahn 2009), so gene conversion among aphid paralogs was assessed in GENECONV (Sawyer 1989; available at http://www.math.wust.edu/~sawyer/GENECONV). GENECONV showed evidence of recent gene conversion between two paralogs (ACYPI005846 and ACYPI005472; P < 0.05, Bonferroni corrected), which we take into account when interpreting our results (see Results and Discussion).

Branch models of molecular evolution (Yang 1998) were implemented with one or more ω categories for branches and one ω across sites (model = 0 or 2, NSites = 0). ω and κ (transition/transversion rate) were both estimated, starting from initial values of 0.2 and 2, respectively, and codon frequency was measured using a 3 × 4 codon table. We tested for elevated ω along specific branches using various models assuming variable selective pressures among branches. Briefly, we tested the null model assuming equal ω across branches and several nested models allowing for different ω ratios for particular branches and compared the log likelihoods under these models using 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 we tested were inspired by the expression profiles of aphid paralogs in the expansion, and we outline them in the results section.

Site-based models of molecular evolution (Nielsen and Yang 1998; Yang et al. 2000) analyzed selection at sites across the slimfast gene family, including other insect orthologs and excluding outgroup sequences not in the slimfast family. To determine candidate sites that had experienced positive selection, we implemented three pairs of models and compared them using a likelihood ratio test. The first model pair tested for variable ω across sites by comparing the null model M0 that assumed a homogeneous ω across sites, with model M3 that allowed ω to vary among three categories. The second and third model pairs, M1a/M2a (Wong et al. 2004; Yang et al. 2005) and M7/M8, tested for positively selected sites. M1a (nearly neutral) constrains ω at all sites to be either one (neutral selection) or less than one (purifying selection), whereas model M2a (positive selection) allows for a third ω category that is greater than one (positive selection). Model M7 allows ω at all sites to fall into one of ten rate categories along a beta distribution (between 0 and 1). Model M8 (positive selection) maintains
the beta distribution and allows for an additional \( \omega \) category greater than one. A significantly greater likelihood resulting from M2a and M8 is evidence that sites are under positive selection.

Results and Discussion

The A. pism Genome Encodes a Large Expanded Family of Nutrient Amino Acid Transporters

Strikingly, compared with other insects with fully sequenced genomes, the A. pism genome contains extensive gene family duplications (International Aphid Genomics Consortium 2010). Phylogenomic analysis of A. pism identified 2,459 gene families that possess A. pism lineage-specific duplications (International Aphid Genomics Consortium 2010), most of which are small to moderate in size (<10 paralogs), but 9% of which represent large A. pism-specific expansions containing >10 closely related duplicates (Huerta-Cepas et al. 2010). To date, detailed analyses have been made of large gene family expansions in A. pism odorant and gustatory receptors (Smadja et al. 2009), sugar transporters (Price et al. 2010), and proteases (Rispe et al. 2008). We now report in detail on the large expansion of AATs belonging to the APC transporter superfamily.

The A. pism genome contains 40 putative APC superfamily AATs that include 18 APC transporters and 22 putative AAAP transporters (supplementary table S1, Supplementary Material online), positioning A. pism as the most APC superfamily AAT-rich insect genome of all those included in our de novo analysis (supplementary table S3, Supplementary Material online). Further, gene expression data in the form of ~88,000 ESTs from the cotton aphid, Aphis gossypii, demonstrate that the APC and AAAP expansions predate divergence of the Macrosiphini and Aphidini at 50–70 Ma (von Dohlen et al. 2006) and is therefore likely common to members of the Aphidinae. Specifically, orthologs for 16 of the 18 A. pism APC transporters are present in the A. gossypii EST data set with only ACYP1005720 and ACYP1005118 not represented in A. gossypii ESTs. Similarly, 20 of the 22 A. pism AAAP transporters had EST support in A. gossypii with only ACYP144297 and ACYP1004320 not represented in A. gossypii ESTs. Thus, although both APC and AAAP families are characterized by aphid-specific expansions, there are marked differences in the size and patterning of the expansions between these two families. The most prominent of these differences being that the APC transporter family expansion is dominated by a single large expansion of 10 transporters and one pair of duplicates (fig. 1 and supplementary figure S1a, Supplementary Material online), whereas in contrast, small expansions that typically contain three to four aphid-specific paralogs characterize the aphid AAAP transporter family expansion (fig. 2 and supplementary figure S1b, Supplementary Material online).

Members of the large APC family expansion, which current evidence indicates is unique to aphids, are orthologous to slimfast, an insect-specific transporter represented by single orthologs in D. melanogaster, A. aegypti, P. humanus, T. castaneum, and A. mellifera. The slimfast gene is one of just a small handful of functionally characterized insect nutrient AATs. Functional characterization of D. melanogaster slimfast (CG11128, Colombani et al. 2003) demonstrated that it transports arginine, whereas transport specificity of A. aegypti slimfast remains unknown (AaSlif, Attardo et al. 2006). Both slimfast orthologs are commonly expressed in fatbody, the insect analog of the vertebrate liver, where it acts as a nutrient sensor (Colombani et al. 2003; Attardo et al. 2006). Specifically, experimental downregulation of D. melanogaster slimfast expression in fatbody tissues results in flies phenotypically similar to those raised under poor nutrient conditions (Colombani et al. 2003). Fatbody-expressed A. aegypti slimfast plays a role in nutritional signaling during mosquito reproduction such that transport of particular cationic amino acids by the fatbody is important for the initiation of vitellogenesis. Downregulation of A. aegypti slimfast together with another cationic amino acid transporter, iCAT2, strongly decreases the expression of vitellogenesis related genes (Attardo et al. 2006).

In remarkable contrast to the large expansion of slimfast in A. pism, and the otherwise rarity of gene duplication affecting other APC transporters during aphid evolution, the AAAP transporters have experienced multiple independent gene duplication events. Further, comparison of figures 1 and 2 demonstrates that although gene family expansions in other taxa are also rare in the APC transporter family, they are relatively common across all taxa including humans and Drosophila in the AAAP transporter family (fig. 2). That said, A. pism has the greatest number of small independent AAAP expansions compared with all other insects in our analysis, which accounts for the fact that the A. pism genome encodes for more AAAP transporters than any other insect genome included in our analysis (fig. 2 and supplementary table S3, Supplementary Material online).

Phylogenetic reconstruction of A. pism AAAP transporters with related sequences from other taxa reveals a conspicuous arthropod expanded clade (labeled “Arthropod expanded clade” in figures) that shows evidence of several ancient gene duplications in the common ancestor of D. melanogaster and A. aegypti (ISCW018959 and ISCW018960) with all other AAAP transporters found outside the arthropod expanded clade more frequently cluster with single gene representatives from each taxon and tend to have one-to-one human orthologs. A. pism, Drosophila, and Tribolium transporters in the arthropod expanded clade have undergone independent species-specific expansions with each small expansion typically containing three to four closely related transporters. We designate the three A. pism expansions (1), (2), and (3) (fig. 2). In contrast, arthropod gene duplications outside of the arthropod expanded clade are only observed for A. pism transporters (ACYP126212 and ACYP144297) and I. scapularis transporters (ISCW018960 and ISCW018959) with all other AAAP transporters (with the exception of some human transporters) present as singletons (fig. 2).
**FIG. 1.** Phylogenetic analysis of *Acyrthosiphon pisum* APC transporters. Bayesian phylogenetic tree of APC transporters (T.C # 2.A.3) from hemimetabolous insects *A. pisum* (red) and *Pediculus humanus* (purple) and other taxa including *Tribolium castaneum* (light blue), *Apis mellifera* (olive), *Drosophila melanogaster* (blue), *Ixodes scapularis* (green), and *Homo sapiens* (black). Additional AATs from *Aedes aegypti* (orange), with known transport function were included in the analysis. All sequences are available from NCBI database using accession numbers indicated in the phylogenetic tree. For each branch posterior probabilities are shown by line thickness, as indicated in the key. The scale bar indicates an evolutionary distance of 0.3 amino acid substitutions per position in the sequence. *Acyrthosiphon pisum* Na-K-Cl cotransporters ACYPI001649 and ACYPI007138 were used to root the tree. All transporters with a known transport function are shaded with a gray box. *Acyrthosiphon pisum* transporters located on genomic scaffold 434, accession GL350054, are boxed in red.
Duplicated Amino Acid Transporter Genes Are Clustered in the *A. pisum* Genome

Members of large *A. pisum* gene family expansions, those that contain >10 in-family paralogs, are frequently arranged as tandem duplicates in the genome (Rispe et al. 2008; Smadja et al. 2009; Price et al. 2010). Clusters of APC and AAAP gene duplicates follow this pattern. A total of seven members of the *A. pisum* APC family expansion...
Global Expression Analysis of *A. pismum* Amino Acid Transporters Within a Phylogenetic Framework Facilitates Initiation of Functional Assignment

The post-genome work we present here thoroughly catalogs a biologically important and understudied family of membrane-bound transporters. Although computational identification of putative transporters is relatively easy and accurate, precise prediction of preferred substrates remains more of a challenge both because of the diverse substrate specificities of APC and AAAP family members, and the paucity of functionally characterized insect AATs. By completing a comprehensive analysis of the tissue-specific expression patterns of APC superfamily members functioning at the *A. pismum/Buchnera* symbiotic interface, this work initiates elucidation of the metabolic integration of these symbiotic partners.

We analyzed, using QPCR, the expression of seven *A. pismum* APC transporters and ten *A. pismum* AAAP transporters relative to whole insect in three tissues (head, gut, and bacteriocyte) from three genetically discrete lines of *A. pismum* (fig. 4 and supplementary fig. S2, Supplementary Material online). Given our interest in the symbiotic interface, these 17 APC superfamily transporters were selected for expression analysis on the basis of their phylogenetic relationship to the set of five bacteriocyte expressed AATs identified by comparative analysis of tissue-specific EST libraries (Nakabachi et al. 2005). With very few exceptions, relative gene expression patterns across tissues and transporters were quantitatively and qualitatively similar in all three aphid lines (supplementary fig. S2, Supplementary Material online). The complete set of expression data for *A. pismum* lineage, LSR1, standardized across all 17 transporters is presented in figure 4. Importantly, figure 4 facilitates comparison of expression among transporters within LSR1 and for each transporter across tissues. Broadly, the following insights emerge from this analysis of the LSR1 data. The most highly expressed of the 17 AATs analyzed are the AAAP transporters ACYP1000550, ACYP1008971, and ACYP1001018 (fig. 4B). The one-to-one orthology and high apparently tissue non-specific expression of these three transporters lead us to hypothesize that they retain an evolutionarily conserved cellular housekeeping function. Second, members of the large *A. pismum* APC family expansion are predominantly gut expressed (fig. 4A). Third, both APC and AAAP families contain transporters that show strong differences in gene expression in gut versus bacteriocyte. Contrasting gut/bacteriocyte expression is observed most notably in the APCs ACYP1002633 (strongly upregulated in gut but weakly expressed in bacteriocyte) and ACYP1008904 (strongly upregulated in bacteriocyte but most weakly expressed in gut) (fig. 4A). Contrasting gut/bacteriocyte expression was also observed in AAAPs ACYP1006258, ACYP1004320, and ACYP1001366 (fig. 4B). Fourth, of the 17 AATs analyzed, only one APC, ACYP1008904 (fig. 4A) and three AAAPs, ACYP1000536, ACYP1006258, and ACYP1000092 (fig. 4B) were enriched >4-fold in LSR1 bacteriocyte tissue relative to whole insect. Fifth, in considering the LSR1 data as presented in figure 4 together with corresponding expression data from aphid lines 5A and 9-2-1 (supplementary fig. S2, Supplementary Material online), we generate a list of eight AATs that are operating at the aphid/Buchnera symbiotic interface (A list consistent with the next-generation RNA sequencing work of Hansen and Moran (2011) published during revision of this manuscript, see supplementary table S1, Supplementary Material online.). These eight transporters that are highly expressed and/or enriched in bacteriocyte tissue relative to whole insect include four transporters that show
bacteriocyte-biased expression and four that are highly expressed in bacteriocytes but are also expressed at high levels in other tissues as determined by their expression relative to whole insects. Of the bacteriocyte-biased genes, AAAP ACYPI000536 and APC ACYPI008904 are the two most highly expressed AATs in bacteriocyte tissue (table 1), whereas the remaining two, AAAP ACYPI000092 and APC ACYPI005118 are enriched 4-fold in bacteriocyte tissue from all three A. pisum lines but are not highly expressed (table 1). The remaining four AATs that are operating at the aphid/Buchnera symbiotic interface are all AAAPs and include three transporters, ACYPI001018, ACYPI000550, and ACYPI008971, that retain one-to-one orthology to transporters in other taxa and one transporter, ACYPI001366, that belongs to AAAP Expansion 3 (fig. 2).

Given the paucity of characterized insect AATs and the generally poor relationship between AAT sequence and function, it is not possible to make accurate predictions of transport specificity for the eight aphid/Buchnera symbiotic interface AATs. However, it is possible on the basis of earlier experimental work to make predictions about the amino acids we might expect to see transported at the symbiotic interface (Shigenobu and Wilson 2011). Most importantly, based largely on the work of Sasaki and Ishikawa (1995) and Whitehead and Douglas (1993), we expect to find that the bacteriocyte plasma membrane and symbiosomal
membrane are differentiated with respect to the AATs they contain such that glutamine is transported across the bacteriocyte plasma membrane with high efficiency, while the symbiosomal membrane has a limited capacity to transport glutamine but an extremely high capacity to transport glutamate. Future work focused on functionally characterizing the transport specificity and subcellular localization (to bacteriocyte plasma membrane or symbiosomal membrane) of the eight aphid/Buchnera symbiotic interface AATs is the next step necessary for advancing our understanding of host/symbiont metabolic integration.

Molecular Evolutionary Insights into the Evolution of Altered Expression Profiles in the Aphid *slimfast* Expansion

As the best-studied insect nutrient AAT, we know that *slimfast* is highly expressed in the gut of *Drosophila* (Colombani et al. 2003), *Tribolium* (Supporting table 3 from Morris et al. 2009), and *A. pismum* (fig. 4, supplementary fig. S2, Supplementary Material online). Functionally, *slimfast* is well established as a nutrient sensor and necessary upstream activator of TOR signaling-mediated growth and oogenesis (Colombani et al. 2003; Attardo et al. 2006). On the basis of the functional importance of *slimfast*, the conservation of *slimfast* gut expression across the Diptera, Coleoptera, and Hemiptera and the fact that bacteriocytes are a derived cell type with no clear progenitor cell type (Brandele et al. 2003), we argue that bacteriocyte expression of *slimfast* paralogs ACYPI005118 and ACYPI008904 is derived.

Although expression divergence in duplicate transporters necessitates changes in upstream promoters regulating spatial and temporal expression, functional divergence in differentially expressed duplicate genes can be driven by either regulatory or coding region evolution (Force et al. 1999; Lynch and Force 2000; Papp et al. 2003; Hahn 2009). If coding region evolution plays a role in the functional divergence of bacteriocyte-expressed aphid *slimfast* paralogs, we expect to find evidence of accelerated rates of evolution along branches associated with a change in expression profile, indicating that they underwent either positive or relaxed selection. Positive selection would imply that functional divergence was driven by adaptive fixation of mutations in one duplicate conferring a new or specialized function (reviewed in Hahn 2009). In contrast, relaxed selection would imply that functional divergence was driven by relaxed selective constraints on some duplicates, resulting in degenerative mutations, for example, in residues important for gut but not bacteriocyte function (Force et al. 1999; Hahn 2009).

To test if accelerated rates of evolution correlate with the evolution of bacteriocyte expression in the *slimfast* expansion, we constructed four nested branch models of molecular evolution (Yang 1998) targeting branches where bacteriocyte expression was gained and/or gut expression was lost (branches ω3 and ω4 in fig. 5A). We were also curious about the paralog exhibiting a global reduction in expression (ACYPI008323; see fig. 4), which may be a pseudogene and thus may be experiencing relaxed selective constraints, so we also analyzed its terminal branch (*ω5* in fig. 5A).

Results for the branch models are reported in table 2. The significantly improved likelihood estimated for the two ratios model compared with the one ratio model (table 2) indicates that branches within the aphid *slimfast* expansion are evolving, on average, at a significantly higher rate than branches outside the expansion, providing a background rate against which to test for accelerated evolutionary rates along specific lineages in the *slimfast* expansion. The three ratios model (fig. 5B) found no significant increase in ω, on average, across the four branches with shifted expression relative to the rest of the expansion (table 2). The final six ratios model indicated that two of the four branches of interest have evolved at a significantly higher rate than the rest of the expansion. The first of these two is the branch predating the duplication event resulting in ACYPI008904 and ACYPI008323 (*ω3* in fig. 5A), and the second is the terminal branch (*ω5* in fig. 5A) leading up to the paralog that underwent a global reduction in expression, ACYPI008323 (fig. 4). Importantly, the rates of evolution along these two branches are likely to be underestimated because of possible gene conversion between *slimfast* paralogs ACYPI00584 and ACYPI005472.

### Table 1. Ranked Bacteriocyte Gene Expression Levels of APC and AAAP Transporters.

<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Tissue Expression, Relative to GAPDH (GAPDH = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WI</td>
</tr>
<tr>
<td>ACYPI000536a</td>
<td>0.02 ± 0.001</td>
</tr>
<tr>
<td>ACYPI008904b</td>
<td>0.01 ± 0.001</td>
</tr>
<tr>
<td>ACYPI001018a</td>
<td>0.07 ± 0.002</td>
</tr>
<tr>
<td>ACYPI000550a</td>
<td>0.15 ± 0.005</td>
</tr>
<tr>
<td>ACYPI008971a</td>
<td>0.09 ± 0.005</td>
</tr>
<tr>
<td>ACYPI001366a</td>
<td>0.03 ± 0.001</td>
</tr>
<tr>
<td>ACYPI000092a</td>
<td>0.01 ± 0.001</td>
</tr>
<tr>
<td>ACYPI001684a</td>
<td>0.03 ± 0.001</td>
</tr>
<tr>
<td>ACYPI005118b</td>
<td>0.01 ± 0.001</td>
</tr>
</tbody>
</table>

**Note:** Expression levels in whole adult insect (WI), head, and gut tissues are also shown relatively to GAPDH, expression levels < 0.01 are not reported (x). Relative expression shown as mean ± standard deviation (n = 3). ACYPI001684 is not included in the eight transporters that are highly expressed and/or enriched in bacteriocytes because its expression is not consistent across all three aphid lines (supplementary fig. S2, Supplementary Material online).

a Ranked expression levels of AAAP transporters (T.C # 2.A.18) in *Acyrthosiphon pisum* bacteriocytes, shown relative to GAPDH (ACYPI009769) expression, GAPDH = 1.

b Ranked expression levels of APC transporters (T.C # 2.A.3) in *A. pismum* bacteriocytes, shown relative to GAPDH (ACYPI009769) expression, GAPDH = 1.
the number of models of molecular evolution tested, where model names indicate A. pisum phylogeny for each branch. In branch 1 (fig. 5) and loss of global expression in ACYPI008323. Future work can resolve the origin of bacteriocyte expression in ACYPI008323. However, positive selection may be difficult to detect in our data set, even with the more powerful site-based models. The power of site-based models to detect positive selection depends on the number of sequences, sequence length, and the optimal sequence divergence. Thus, the sequence divergence in our data set may not provide the optimal power to detect positive selection. Nevertheless, relaxed selective constraints can still lead to functional divergence and maintenance of duplicate genes by driving complementary degenerative mutations in duplicates so that each duplicate child must be maintained in order to fulfill the complete function(s) of the parent gene (Force et al. 1999).

With respect to ACYPI008323, the paralog exhibiting a global reduction in expression, our molecular evolution analyses are consistent with the possibility that this paralog has been pseudogenized. However, the codon sequence of this paralog lacks frame-shifting mutations and premature stop codons, strongly indicating that ACYPI008323 is functional. The nearly absent expression of ACYPI008323 in this study suggests that it lacks a role in asexual female aphids and instead functions in an unexamined aphid reproductive or developmental morph (e.g., males or sexual females). Future studies examining expression profiles of more aphid morphs and tissues together with protein localization can determine if ACYPI008323 encodes a functional gene product and if so, identify its functional role.
Gene Duplication and the Evolution of the Aphid/Buchnera Obligate Intracellular Symbiosis

The aphid/Buchnera symbiosis is the first obligate intracellular partnership for which both the host and symbiont have fully sequenced genomes (Shigenobu et al. 2000; International Aphid Genomics Consortium 2010). Although symbiont genome evolution has been extensively studied and is well understood to be characterized by extreme reduction in genome size, base composition bias generally toward A-T richness and rapid DNA sequence evolution (Moran et al. 2008), the picture of concomitant host genome evolution is just beginning to emerge. To date, the pea aphid genome project has provided three key insights into host genome evolution (International Aphid Genomics Consortium 2010). First, there are several instances of host genome gene loss that are implicated as host adaptations for maintenance of the symbiosis. These include, loss of genes central to the Immunodeficiency pathway (Gerardo et al. 2010), purine salvage (Ramsey et al. 2010), and the entire urea cycle (Wilson et al. 2010). Second, amino acid metabolism, the process at the core of the aphid/Buchnera symbiosis is characterized by extensive and unprecedented metabolic collaboration between partners at the level of genes within metabolic pathways (Wilson et al. 2010; Hansen and Moran 2011; Shigenobu and Wilson 2011). Third, the pea aphid has acquired multiple genes of bacterial origin that are highly and in some cases apparently exclusively expressed at the symbiotic interface (Nihoko et al. 2010). This study is the first to present evidence for a fourth insight: that gene duplication may be mechanistically important for host adaptation at the symbiotic interface. All four of the AATs that show bacteriocyte-biased gene expression are members of aphid-specific expansions (figs. 1 and 2), and three of the four have close paralogs with different expression profiles that include members highly expressed at the diet/gut interface (fig. 4). On the basis of this fourth insight and the fact that gene duplication is an important evolutionary force that facilitates the acquisition of novel traits (Ohno 1970; Hahn et al. 2007), we propose that gene duplication has been mechanistically important to the evolution of obligate intracellular symbioses. More specifically, we predict that bacteriome-biased genes will be members of taxon-specific gene family expansions.

An adaptive role for AAT gene duplications in the evolution of the aphid/Buchnera symbiosis predicts that the duplications were present in the aphid common ancestor, that bacteriocyte expression is derived, and that bacteriocyte expression predates aphid diversification from a common ancestor (Coddington 1988). At present, we cannot discount the possibility that the gene duplications and/or bacteriocyte expression arose more recently than the ~200 My old (Moran et al. 1993) symbiosis between aphids and Buchnera. Nevertheless, EST support for orthologous Acyrthosiphon gossypii slimfast indicates that the APC and AAAP expansions were present in the common ancestor of the Aphidinae, the family comprising the majority of aphid species. Future research in other bacteriome associated symbioses, including basal aphid lineages and other phloem-feeding insects, will facilitate testing our hypothesis and help to develop a clear understanding of how symbiont acquisition has shaped the evolution of host genome architecture.

Supplementary Material

Supplementary tables S1–S3 and figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the University of Miami startup funds to A.C.C.W. We thank University of Miami undergraduate, Ryan McMinds for technical assistance, Joe Bielawski for advice on the molecular evolution analyses, Nancy Moran for critical feedback, and three anonymous reviewers for their comments on the manuscript.

References


Zwickl D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion [PhD dissertation]. The University of Texas.