THESIS

RNA SEQUENCING IDENTIFIES GENES PUTATIVELY INVOLVED AT THE APHID-*BUCHNERA* SYMBIOTIC INTERFACE

Submitted by

Shailesh Raj Acharya

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2020

Master's Committee:

Advisor: Vamsi. J. Nalam

Jan Leach
Daniel Sloan

ABSTRACT

RNA SEQUENCING IDENTIFIES GENES PUTATIVELY INVOLVED AT THE APHID-*BUCHNERA* SYMBIOTIC INTERFACE

In aphids, the supply of essential amino acids depends on an ancient nutritional symbiotic association with the gamma-proteobacterium, *Buchnera aphidicola*. The endosymbiont converts abundant non-essential amino acids into essential amino acids that are supplied to the aphid. The long-term goal of the proposed work is to exploit the biochemical interdependence that exists between soybean aphid (*Aphis glycines*) and its primary endosymbiont to develop effective resistance in soybean (*Glycine max*). Little is known of the *A. glycines* and soybean amino acid transporters (AATs) that facilitate this exchange. The soybean aphid is the most important arthropod pest on soybean in North America and aphid outbreaks in major soybean growing regions of the country in the past has resulted in yield losses of up to 40%. In the current study, we used RNA-seq to identify amino acid transporters involved in the exchange of amino acids between the aphid and its endosymbiont. A total of 2121 genes were differentially expressed between the aphid and bacteriocytes with 516 genes showing up-regulation, while 1605 genes were down-regulated in the bacteriocytes. Analysis of GO terms revealed enrichment in membrane and transport associated processes. Our RNA-seq analysis of differentially expressed genes showed that one putative amino acid transporter: 72-RA, is up-regulated in the bacteriocytes. This work represents a first step towards understanding aphid dependency on its endosymbiotic bacteria and target them as a means of a novel aphid control strategy.

ACKNOWLEDGMENTS

Firstly, I would like to thank Dr. Vamsi J Nalam for his professional support as a mentor who provided me guidance throughout my master's research. His help in plant pathology and entomology has made me an enthusiast in the world of agriculture and pests. He has provided me with immense support in the current project, a formal start to my graduate career. During my ebb and flow throughout this project, his encouragement has made me an independent thinker and an interpretative problem solver. My research funded by USDA due to Vamsi's throughput supported my career as a master's student.

I would like to thank my committee members, Dr. Jan Leach and Dr. Daniel Sloan, for their valuable inputs in my research. I am grateful for their generosity shown during the global pandemic, which temporarily halted my research. I genuinely appreciate their ethics and professionalism.

I would like to thank our lab managers Elizabeth Blotevogel and Tessa Albrecht, for their valuable guidance. I wholeheartedly appreciate the positive work attitude I inherited from them. I would also like to thank my lab mate Jen Rushton for her helping hand in my experiments. My cohort Patrick Woods helped me learn bioinformatics and statistics skills. I am genuinely grateful to have acquired invaluable skills as a biologist from him.

Finally, I would like to thank all the professionals from our Agricultural Biology department who supported me in all forms throughout my master's research. I would like to appreciate our department head, Amy Charkowski, who has shown an immense supportive both personally and professionally.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

Agricultural crop production is a direct measure of food availability. The global demand for food production is estimated to double by 2050 (Hunter et al., 2017). However, agricultural production demands constant attention to mitigate challenges from pests, pathogens, and the environment. The increasing use of insecticides and pesticides is creating an environmental hazard. For decades, genetic control against pests has been deployed in addition to traditional pesticide use. However, pests and pathogens have developed resistance to genetic control and commonly used agrochemicals over time. Thus, suitable integrated pest management (IPM) strategies to control pests and pathogens is required. The current research attempts to identify ways to exploit the biochemical interdependence between soybean aphids and their endosymbiotic bacteria. The work serves as a foundation for a biologically based novel pest management strategy of controlling soybean aphids.

Aphids

Aphids (Hemiptera: *Aphididae*) are some of the most important pests that cause economic loss in agriculture. Aphids belong to the sub-order *Sternorrhyncha* which also includes other economically important pests like mealybugs and whiteflies. Aphids undergo parthenogenesis to produce asexual females and could either be apterous (wingless) or alate (winged) forms. However, during late summer, both male and female winged forms are produced, migrating to a secondary host plant for mating. The *Aphididae* family includes more than 5000 species worldwide, and about 450 of these species are known to be economically important agricultural pests (Emden & Harrington, 2007). About 100 species fall under the subfamily *Aphidinae,* which are some of the most commonly found aphid pests in the U.S (Emden & Harrington, 2007).

Aphids feed on phloem sap siphoning away nutrients essential for plant growth. Aphids utilize piercing/sucking mouthparts to feed exclusively on sugar-rich, nitrogen (N)-poor plant phloem sap. Aphid feeding causes withering in plants, and in the event of heavy infestation, death. Aphids increase their damage potential as they serve as vectors of several economically important viruses. More than 200 aphid species act as vectors for 50% of all insect-vectored viruses (Ng & Perry, 2004). The viruses vectored belong primarily to the *Potyviridae* family (Ng & Perry, 2004). Leaf yellowing and wrinkling in aphid fed plants are mainly due to the viruses transmitted by aphids. Yellow leaves have increased free amino acids that are beneficial to aphids (Hohn, 2007). Aphids secrete sugar-rich honeydew onto leaves due to feeding on a high-sugar diet, which serves as an excellent food source for sooty molds. The growth of the non-plant pathogenic sooty mold on leaves impairs the plant's ability for photosynthesis. It also creates an environment ideal for the growth and development of other fungal diseases (Sorensen, 2009).

The economic importance of aphids in agriculture can be attributed to their rapid asexual reproduction and voracious feeding, the transmission of viruses, and the promotion of fungal pathogens (Hohn, 2007; Ng & Perry, 2004; Sorensen, 2009). The rapid evolution of resistant aphid biotypes possesses an additional problem in mitigating agricultural loss

Aphid -*Buchnera* Symbioses

In nature, symbiosis plays a crucial role in animals' optimal survival in diverse ecological niches (Wilson & Duncan, 2015). Animals living in a balanced symbiosis benefit each other, and balanced symbiosis between a symbiont and its host is ubiquitous.

Symbionts living inside the host body are endosymbionts, and these can be either primary (obligate) or facultative. Primary or obligate endosymbionts are present inside the insect host that feeds on unbalanced diets and fulfills their dietary needs (Heyworth & Ferrari, 2015). They are

vertically transferred from the mother to offspring. Facultative endosymbionts are transmitted both horizontally and vertically and provides a selective advantage to the host (Guo et al., 2017; Herren et al., 2013).

Aphids and specifically pea aphids (*Acyrthosiphon pisum*) are a well-studied model organism for insect-symbiont interaction. In aphids, the primary endosymbiont is *Buchnera aphidicola*, a gammaproteobacterium. At least eight other facultative endosymbionts have been identified in pea aphids (Heyworth & Ferrari, 2015; Russell et al., 2013). Most of these are facultative endosymbionts that provide the pea aphid protection from predators and affect a wide range of life-history and ecologically important traits (Heyworth & Ferrari, 2015; Wernegreen, 2012). In soybean aphids (*Aphis glycines*), the primary endosymbiont is also *Buchnera aphidicola.* In addition, soybean aphids also house *Arsenophonus* as facultative endosymbionts that provide a general benefit to aphid growth and reproduction (Wille & Hartman, 2009; Wulff & White, 2015).

Characteristic features of Aphid -*Buchnera* endosymbiotic interface

Aphids rely on phloem sap from plants as their sole source of nutrients. They use their piercing and sucking mouthparts to feed on a nutritionally unbalanced phloem sap rich in sugar but low in nitrogen. Besides, the phloem sap consists of a higher proportion of non-essential amino acids (NEAAs) as compared to essential amino acids (EAAs) - i.e., amino acids that aphids cannot synthesize on their own. Aphids rely entirely on their endosymbionts to convert NEAAs obtained from phloem sap into essential amino acids. In return the aphid compensates the endosymbiont with essential carbohydrates and lipids (Chong & Moran, 2018; Moran et al., 2005).

Much of our knowledge of the relationship between the aphid and *Buchnera* comes from studies on pea aphid (Feng et al., 2019; Price et al., 2011; Price et al., 2014a; Wilson et al., 2010). *Buchnera* resides in a U-shaped organ called the bacteriome within the aphid hemocoel. The

bacteriome is a bilobed organelle that harbors about 90-110 bacteriocyte cells (Figure 1) (Wilson, 2020).

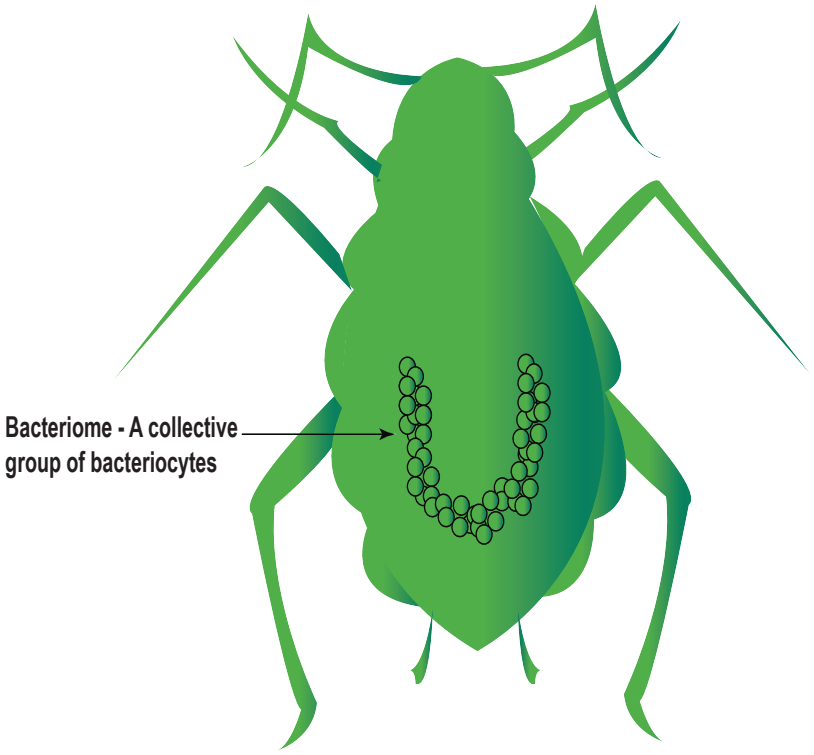


Figure 1. Location of bacteriocytes in an aphid.

The bacteriome is a U-shaped organ located at the posterior part of the aphid gut comprising 90-110 individual bacteriocyte cells.

Each bacteriocyte cells contains insect-derived sheath cells that connect large round cells within which thousands of *Buchnera* reside (Houk & Griffiths, 1980; Wilson., 2020). The number of bacteriocytes housing *Buchnera* vary depending upon aphid developmental stage. Bacteriocyte size and number are largest during aphid nymphal development but gradually decrease as aphids mature (Simonet et al., 2016). *Buchnera* is vertically transmitted from maternal aphids.

The location of the endosymbiont is not static and varies during an aphid's development. Initially, *Buchnera* is localized to the syncytium region and the hemolymph of mother aphids (Koga et al., 2012; Wilson, 2020), but their localization changes to the bacteriocyte cells after transmission to the progeny is complete.

The innermost membrane is *Buchnera*-derived and is bi-layered with an inner cell membrane and an outer cell-wall. The outermost aphid-derived membrane, the middle symbiosomal membrane, and the innermost *Buchnera* cell wall together form the Aphid-*Buchnera* symbiotic interface (Figure 2).

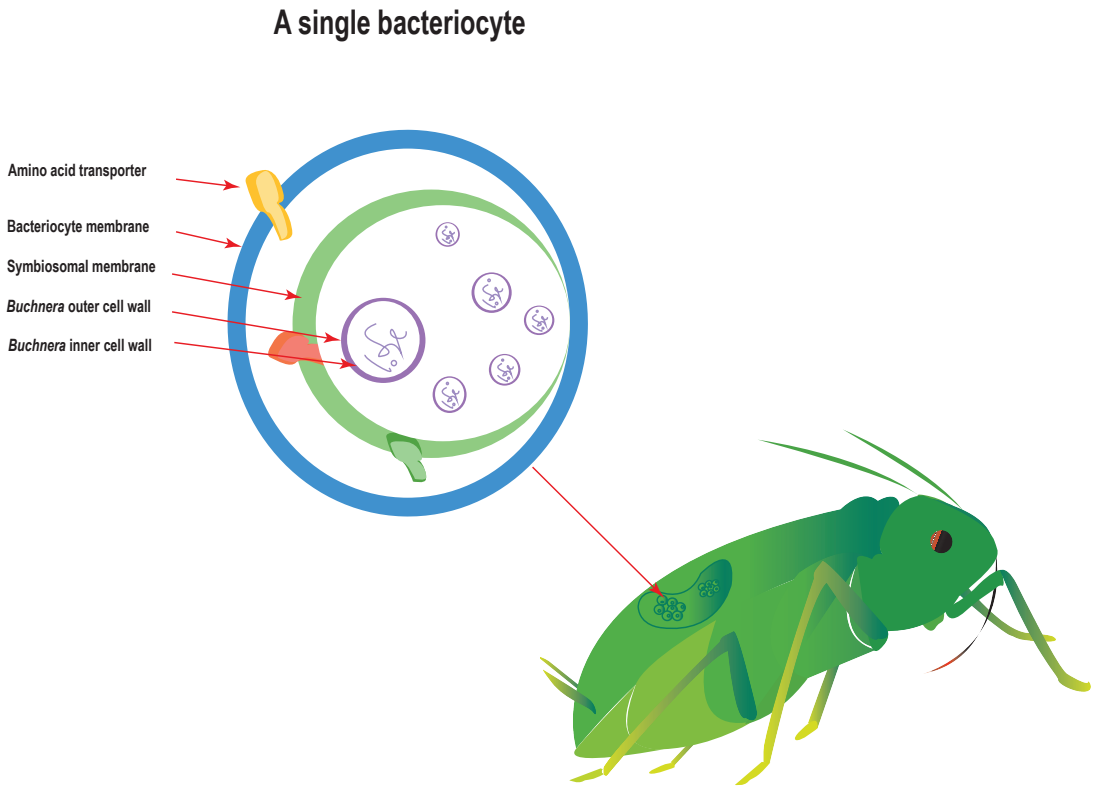


Figure 2. Aphid-*Buchnera* interface.

Each bacteriocyte contains an outer bacteriocyte membrane, a middle symbiosomal membrane, both of which are aphid derived. The symbiosomal membrane partitions aphid and *Buchnera aphidicola*

The endosymbiotic relationship between aphid and *Buchnera* has co-existed for over 160 million years (Douglas, 1998). This co-existence has resulted in a significant reduction in *Buchnera*'s genome (Moran et al., 2009; Wilson et al., 2010). The genetic changes in *Buchnera* include lack of genetic recombination, gradual loss of genes for DNA repair, codon bias due to high AT-rich regions and protein instability (Chong et al., 2019; Wilson, 2020). Concomitantly, there have been a loss of genes involved in multiple biosynthetic pathways in aphids including those involved in amino acid synthesis. The host-symbiont co-evolution thus complements each other for survival (Chong & Moran, 2018). Additionally, the aphid's ability to retain *Buchnera* as an endosymbiont is possibly due to the lack of pathogenicity genes in *Buchnera* and the absence of immune response genes in aphids (Douglas, 1998; Douglas et al., 2011).

The aphid's need for essential amino acids is a consequence of feeding on sugar-rich nitrogen-poor phloem predominated by NEAAs. Phloem is nutritionally limited and as aphids are unable to synthesize all ten essential amino acids (arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) de-novo, they must rely on *Buchnera* for their synthesis. The aphid compensates for the absence of NEAAs (asparagine, aspartate, glutamine, glutamate, proline, serine, and tyrosine) biosynthetic genes in *Buchnera*. However, host-endosymbiont dependency is not absolute. While some biosynthesis pathways are absent in either host or endosymbiont, genes involved in biosynthesis and regulation of the pathways are shared. For example, the synthesis of five essential amino acids (Phe, Ile, Leu, Val, and Met) requires genes from both the aphid and *Buchnera*. In pea aphid, two of the five amino acid synthesizing enzymes are retained in bacteriocytes, but the intermediary pathways are present in the *Buchnera* cytoplasm (Smith & Moran, 2020).

Sugar Exchange between Aphids and Buchnera

In the phloem, sugar is usually present at a higher concentration than free amino acids (Price et al., 2007). Sucrose is the principal sugar transported in the phloem. In aphid gut epithelium, sucrose is hydrolyzed by α-glucosidase/trans-glucosidase activity into glucose oligomers and monomers. The monomers of glucose enter the aphid hemolymph and are transported to fat bodies (Price et al., 2007). In animals, sugar transport across membranes is performed by two major protein transporters that belong to the sodium/solute symporter family (SSF; Pfam PF00474) and the major facilitator superfamily (MFS; Pfam CL0015) (Mueckler & Thorens, 2013; Price et al., 2010). The MFS superfamily consists of integral membrane transporters consisting of 12 transmembrane helices spanning the plasma membrane. Sugar transport via the MFS transporter can occur either along the concentration gradient, as uniporters, or against the concentration gradient as symporters or antiporters (Price et al., 2010). Although much is known about the MFS superfamily of transporters in animals (Mueckler & Thorens, 2013; Price & Gatehouse, 2014), our knowledge of these transporters in hemipterans is lacking. In the rice brown planthopper (*Nilaparvata lugens*), NIHT1, a low-affinity glucose transporter with sequence similarity to a mammalian glucose transporter, was identified that facilitates glucose transport from the gut into the hemolymph (Price et al., 2007).

Trehalose, a disaccharide of glucose, is the major sugar found in the hemolymph of aphids. In aphids, fat bodies or adipocyte cells surround gut tissues and are bathed by the hemolymph. Trehalose is formed by the immediate conversion of hydrolyzed glucose monomers in the fat bodies and released into the hemolymph (Moriwaki et al., 2003). Since aphids feed on a diet rich in sugar, this generates a positive osmotic pressure in the aphid gut that can result in dehydration (Kikawada et al., 2007; Moriwaki et al., 2003). Among the several mechanisms to counter

dehydration, aphids rapidly convert glucose to trehalose. Trehalose is also the major sugar that is utilized by *Buchnera*. In the bacteriocytes, trehalose is hydrolyzed by trehalase into glucose monomers. The higher expression of trehalase and trehalose transporters within bacteriocyte cells signifies the importance of trehalose (Moriwaki et al., 2003). In the bacteriocytes and *Buchnera,* trehalose is the metabolic precursor for glycolysis, oxidative phosphorylation, amino acid biosynthesis, and several other biosynthetic pathways, including those occurring in the bacteriocyte mitochondria. A facilitated high capacity trehalose transporter, Trehalose transporter 1 (TRET1), in sleeping chironomid (*Polypodium vanderplanki*) has been shown to be present inside the fat body (Kikawada et al., 2007). TRET1-like facilitated trehalose transporters have also been reported in different aphid genus (Chen et al., 2019; Li et al., 2019; Mathers, 2020; Richards et al., 2010; Voronova et al., 2020).

A recent study in pea aphids has reported two sugar transporters: *A. pisum* sugar transporter 3 (Ap_ST3) and *A. pisum* sugar transporter 4 (Ap_ST4) (Price & Gatehouse, 2014). These transporters share sequence similarity with mammalian GLUT transporters and are globally expressed in aphid tissue with high enrichment in gut tissues (Price & Gatehouse, 2014). As these transporters are globally expressed across aphid tissue, they may be involved in the transport of glucose, fructose, and other sugar monomers from the hemolymph to bacteriocytes.

Amino acid exchange between Aphids and *Buchnera*

The bi-directional transport of amino acids between *Buchnera* and aphids is carried about by transporters of the amino acid polyamine choline superfamily (APC). The APC superfamily consists of two subfamilies: Amino acid polyamine choline (APC) family and Amino acid/auxin permease (AAAP) family (Douglas, 2006; Price et al., 2011; Duncan et al., 2014). In pea aphids, 40 amino acid transporters have been identified, of which 18 belong to the APC family, and 22

belong to the AAAP family. Transporters from the APC family contain 12-14 transmembrane domains (Jack et al., 2000; Price et al., 2011) and can be uniporters, symporters, and antiporters. Transporters from the AAAP family have only 11 transmembrane domains (Price et al., 2011; Young et al., 1999) and are mainly symporters. Besides, they can transport multiple amino acids. The APC superfamily of transporters has been extensively studied in the pea aphid.

Research suggests that the large number of APC transporters in aphids is a consequence of extensive gene duplication, an evolutionary process that generates genes with new functions. The increased expression of amino acid transporters within the bacteriocyte cells indicates the importance of amino acid exchange between the aphid and its endosymbiont (Price et al., 2011). *Buchnera* retains all the essential amino acid synthesis genes absent in the aphid, and there is a need to transport these amino acids from *Buchnera* to the aphid. Conversely, the precursor NEAAs must be transported from the aphid to *Buchnera* for essential amino acid synthesis. Thus, the increased expression of genes encoding amino acid transporters is necessary.

ApGLNT1 and ApNEAAT1: two transporters at the *A. pisum-Buchnera* interface

Buchnera obtains non-essential amino acid (NEAA) precursors from the aphid. These precursors are transported across two distinct aphid-derived membranes, the bacteriocyte, and the symbiotic membrane. The precursors are then transported across *Buchnera's* outer and inner cell wall into its cytoplasm, following which essential amino acids (EAAs) are synthesized from the non-essential amino acid precursors. These are then transported out to the bacteriocyte cytoplasm (Figure 3). This bi-directional transport of EAAs and NEAAs through the compartmentalized membrane layers requires different amino acid transporters present at each membrane. In the pea aphid, two amino acid transporters have been identified that play a role in the bi-directional transport of NEAAs and EAAs. Both the transporters are members of the APC superfamily:

ApGLNT1 – a glutamine transporter and ApNEAAT1 – small dipolar amino acid transporter (Figure 3) (Feng et al., 2019; Price et al., 2014b) .

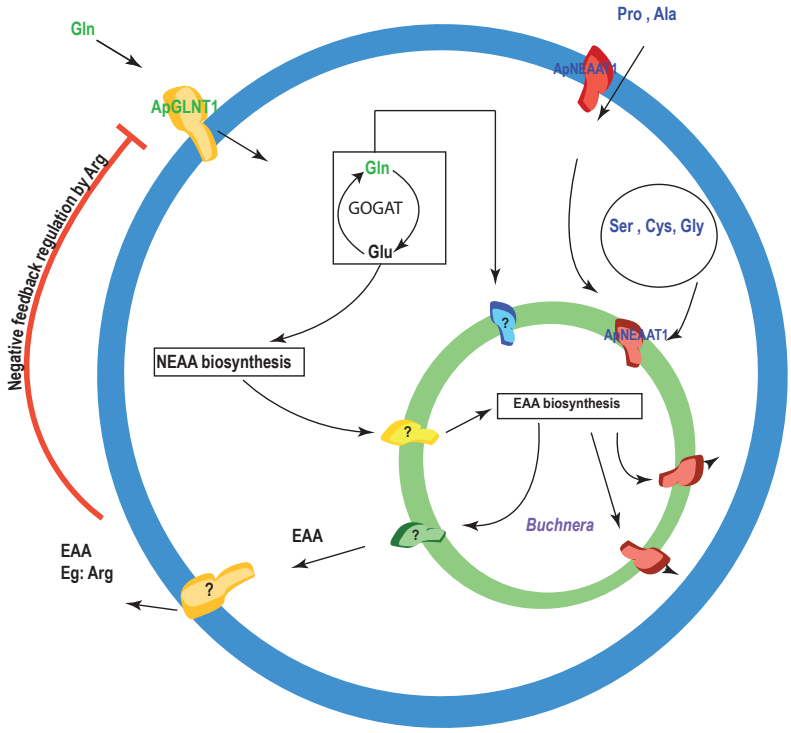


Figure 3. Amino acid transport pathways across the aphid-*Buchnera* symbiotic interface.

Glutamine is transported across the bacteriocyte membrane (blue) via a glutamine transporter, ApGLNT1. Glutamine is either converted into glutamate in the bacteriocyte cytoplasm to synthesize other NEAAs or is transported into *Buchnera* (across the symbiosomal membrane, green), serving as a precursor for the synthesis of EAAs. Small non-polar amino acids are transported from hemolymph to *Buchnera* via a NEAA transporter ApNEAAT1 present at the bacteriocyte membrane and the symbiosomal membrane.

Glutamine is an abundant NEAA present in the hemolymph of aphids. It acts as a precursor for the synthesis of several NEAAs and EAAs amino acids. Five out of seven NEAAs - aspartate, glutamine, glutamate, serine, and tyrosine - are synthesized in bacteriocyte cytoplasm from glutamine. In *Buchnera*, glutamine also acts as a precursor for the biosynthesis of EAAs - arginine,

histidine, and tryptophan. Since aphid lack genes to synthesize these EAAs, glutamine must be delivered into the *Buchnera* cytoplasm for their synthesis. ApGLNT1, located at the bacteriocyte membrane, facilitates glutamine transport from aphid hemolymph to the bacteriocyte cytoplasm. Glutamine in the bacteriocyte cytoplasm gets converted into glutamate via the GS-GOGAT cycle (Hansen & Moran, 2011). Glutamate and glutamine are either converted into EAAs and NEAAs acids or are transported into the *Buchnera* cytoplasm. The transport of glutamine and five NEAAs from the bacteriocyte cytoplasm to *Buchnera* occurs via uncharacterized transporters. The transporters at the symbiosomal membrane are also uncharacterized.

The precursor NEAA glutamine is synthesized by *Buchnera* into EAA arginine. When produced in excess, arginine competes for the glutamine binding of ApGLNT1 (Price et al., 2014b). This substrate feedback inhibition by arginine reduces the transport of the precursor glutamine into bacteriocytes. The regulation of the transport of smaller dipolar NEAAs amino acids by ApNEAAT1 is concentration-dependent (Feng et al., 2019). The transport of NEAAs: proline, alanine, glycine, serine, and cysteine across bacteriocyte and symbiosomal membrane bidirectionally occur across their concentration gradient. For example, the synthesis of methionine, an EAA, requires the coordinated influx of serine into *Buchnera* and the efflux of cysteine intermediate into bacteriocyte (Feng et al., 2019). The bi-directional nature of ApNEAAT1 ensures the proper transport of these two dipolar amino acids required for methionine synthesis (Feng et al., 2019; Hansen & Moran, 2011; Smith & Moran, 2020)

The pea aphid has served as a model organism within the *Aphididae* family. Resources are abundant, including sequence data and genomic resources for pea. The soybean aphid shares 80% putative gene homologs with the pea aphid (1145 out of 1453 genes tested in soybean aphid) (Liu et al., 2012; Wenger et al., 2017). This homology is higher for sugar transporters (94.7%) and

highest for protein transporters (95.7%) (Liu et al., 2012). The high homology implies that a similar amino acid transport mechanism also exists at the soybean aphid/*Buchnera* symbiotic interface. The identity of such amino acid transporters at the soybean aphid/*Buchnera* interface remains unknown.

Objective

In the United States, soybean aphid (*Aphis glycines* Matsumura) is an important aphid pest on soybean. The aphid causes significant yield loss in soybean through direct feeding damage and as vectors of several economically important viruses. Besides, soybean aphids have also caused an increase in production costs due to the increasing need for scouting and insecticides, both for foliar and seed treatments.

Soybean (*Glycine max*) is a major cash crop grown in over 40 million hectares across the United States, generating about $39 billion in annual revenue (USDA, 2019). Thus, any disruption in production will have a significant impact on the economy. Soybean aphid (*Aphis glycines* Matsumura) is a significant arthropod pest of soybean that has resulted in yield loss of up to 40% in the past (USDA, 2019). The damage potential of soybean aphid has increased insecticide application 130-fold in less than ten years (USDA, 2019). Soybean aphids are also vectors of economically important viruses, including the Soybean mosaic virus and Alfalfa mosaic virus (Hill & Whitham, 2014). Despite deploying resistant soybean cultivars with individual genes for aphid resistance, soybean aphids have evolved biotypes to overcome host plant resistance. Thus, unique strategies to enhance host plant resistance are required to alternative classical breeding and current management practices.

This research aimed to identify novel mechanisms that disrupt the transport and biosynthesis of essential amino acids in soybean aphid. The objective of this research is to identify soybean aphid amino acid transporters recruited to the aphid/*Buchnera* symbiotic interface.

CHAPTER 2 - MATERIALS AND METHODS

This experiment is focused on identifying differentially expressed genes present at the soybean aphid/*Buchnera* interface and identifying which of these genes are involved in amino acid transport. We measured the expression of genes present at the aphid/*Buchnera* interface by performing RNA seq analysis on the whole body from aphids and bacteriocyte cells. Findings from the experiment would help to elucidate the complicated nutritional endosymbiotic relationship between soybean aphid and *Buchnera*.

Soybean aphid colony maintenance

Adult soybean aphids (*A. glycines*) were obtained from soybean fields at the Pinney Purdue Agricultural Center (PPAC), Wanatah, Indiana. They were reared in soybean variety Asgrow® AG3334 (Monsanto, St. Louis, MO). This variety was grown on Mastermix® 830 soilless media (Mastermix, Quakertown, PA) supplied with Osmocote 274850 Smart-Release Plant Food Plus fertilizer in a gallon pot (15.24 cm diameter x 14.61 cm height). Soybean plants with insect colonies were confined in BioQuip® rearing cage (1450N Rearing Cage, Aluminum, (45.7 x 45.7 x 76.2 cm) maintained at 16:8 (Light: Dark) photoperiod at 460 µmol/m2/sec photosynthetically active radiation (PAR) at 25˚C.

RNA extraction

Soybean aphids were collected from the reared colony for RNA extraction. RNA was extracted from two different treatment groups: Whole-body and Bacteriocytes. The whole-body treatment group comprised 30 adult aphids pooled together in an Eppendorf tube, which was considered a replicate. RNA was then extracted using Direct-zol RNA Miniprep Plus® (Zymo, Irvine, CA). The whole-body treatment group consisted of three replicates. For bacteriocytes

collection, aphids were dissected using 1X Phosphate Buffer Saline (PBS) solution under a dissecting microscope at 4X magnification. Bacteriocytes enriched in hemolymph were then pipetted into an Eppendorf tube containing 250ul of RNA later solution. This consisted of one replicate. Three such independent replicates of bacteriocytes enriched in hemolymph were collected. RNA from hemolymph enriched bacteriocytes were isolated using TRIzol™ LS Reagent (ThermoFisher Scientific, Waltham, MA) and Direct-zol RNA Miniprep Plus® kit (Zymo, Irvine, CA). A total of six samples, three replicates each for whole-body and bacteriocytes treatment, were obtained. Extracted RNAs' were quantified using Quant-iT™ RNA Assay Kit (ThermoFisher Scientific, Waltham, MA) and were immediately stored at -80˚C before shipping it to Novogene (Sacramento, CA) for sequencing.

RNA sequencing

All six samples, three from bacteriocytes and whole-body extractions, were subjected to total RNA Quality Control (QC) before library preparation for Illumina Sequencing. Nanodrop reading ($OD_{260/280}$) was used to test the RNA purity. Agarose gel (1%) electrophoresis was used to test for RNA degradation and any DNA/protein contamination. RNA integrity and quantification were done using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). One μg RNA per sample was used as the input material to prepare samples for sequencing. RNA sequencing libraries were created using the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA). Index codes were added to attribute sequences to each sample. Poly-T oligo-attached magnetic beads were used to purify mRNA from the total RNA samples. The mRNA fragmentation was done using divalent cations under elevated temperature using 5X NEBNext® First Strand Synthesis Reaction Buffer. Random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-) were used for first-strand cDNA synthesis. The second strand

cDNA was then synthesized using DNA Polymerase I and RNase H. Any additional dT overhangs were converted into blunt end cDNA with the exonuclease/polymerase activity. The NEBNext Adaptor was then ligated into the adenylated 3' cDNA ends to prepare for hybridization. The cDNA library fragments were then purified for 150-200 bps length fragments using the AMPure XP system (Beckman Coulter, Beverly, USA). PCR was performed using Phusion High-Fidelity DNA polymerase, universal PCR primers, and Index(X) primer followed by treatment of size selected adapter-ligated fragments with 3ul USER enzyme (NEB, USA) at 37^0C for 15 mins and 95^0C for 5mins. The PCR products were then purified, and quality assessment of PCR products was done using the Agilent Bioanalyzer 2100 system. An outline of our RNA sequencing analysis workflow is presented in Figure 4

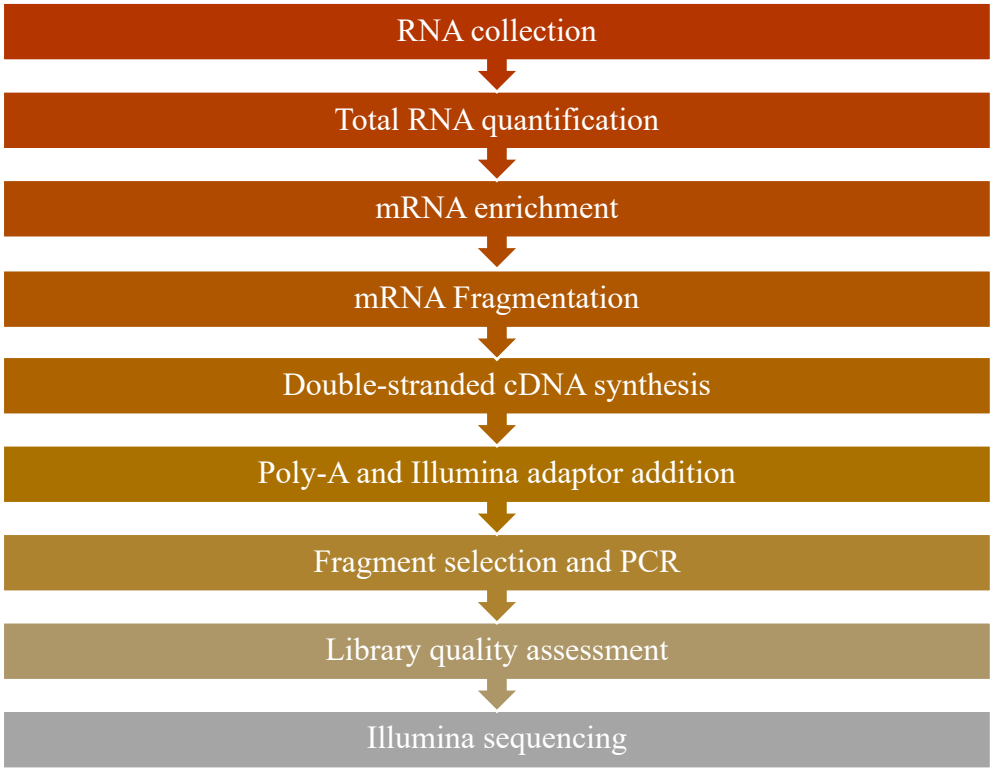


Figure 4: **Outline of RNA sequencing workflow.**
From sample collection to sample preparation for RNA sequencing on Illumina platform.

Filtering and quantification of gene expression level

Clean reads were obtained after trimming raw reads to remove Illumina adapter sequences, poly-N sequences, and low quality reads (Phred-score < 20), using fastp (Chen et al., 2018). The whole-genome sequence (*A. glycines* Biotype 1 v6) obtained from the Aphid base (BIPAA, 2017) was indexed and aligned to the filtered paired-end read using HISAT2 v2.1.0 (Kim et al., 2015). Read counts for those reads uniquely mapped to each treatment were measured using HTSeq v0.6.1 in union mode (Anders et al., 2015). Gene expression level was normalized using FPKM (Fragments Per Kilobase of reads per million base pair sequence) that accounts the effect of gene length and sequencing depth for the read counts mapped to a gene (Trapnell et al., 2010).

Differential gene expression analysis

Normalized read counts were used to perform differential gene expression analysis using DESeq v1.10.1 in R which uses negative binomial distribution model (Love et al., 2014). Log_2 fold change was calculated to determine the relative gene expression levels between bacteriocytes and whole-body RNA. Fold changes greater than one were selected for further analysis. The p-value obtained from differential gene expression was adjusted (p_{adj}) using the Benjamini and Hochberg's method to trim any false positives. Genes with adjusted p_{adj} less than 0.05 were considered to be differentially expressed.

GO Enrichment analysis

Gene Ontology (GO) enrichment analysis predicts molecular processes, biological functions, and cellular components for differentially expressed genes. GO enrichment analysis of differentially expressed genes was done using GOSeq v2.12 package in R which uses Wallenius non-central hypergeometric distribution model (Young et al., 2010) . Gene length bias was adjusted to give a corrected p-value since the adjustment solves the issue of likelihood of longer

gene having more GO categories (Mi et al., 2012). Corrected p-values <0.05 were significantly enriched GO terms. The number of differentially expressed genes with GO annotation was determined.

KEGG pathway analysis

KEGG pathway analysis gives a molecular network diagram showing the relationship between differentially expressed genes and the pathways they are involved in (Kanehisa & Goto, 2000). KEGG pathway analysis was done to access enriched metabolic pathways related to differentially expressed genes using KOBAS v3.0 (http://kobas.cbi.pku.edu.cn/) (Xie et al., 2011). Rich factor, Q-values, and gene counts enriched in KEGG pathway analysis were determined for DEGs. An outline of our RNA sequencing data analysis workflow is presented in Figure 5.

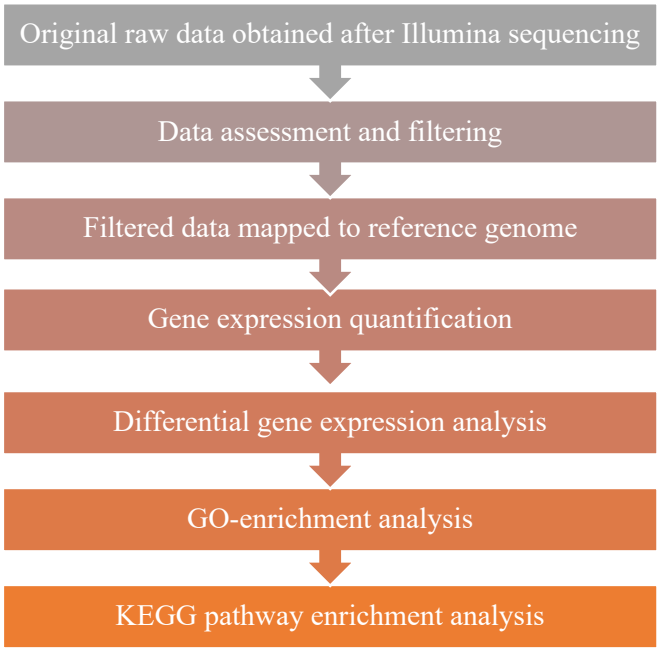


Figure 5. Outline of data analysis workflow.
Data analysis for Illumina reads obtained after RNA sequencing.

DEGs in soybean aphid bacteriocytes compared to DEGs in pea aphid bacteriocytes.

Top 100 upregulated DEGs in bacteriocytes with log_2fold change greater than two were chosen, and blastx search against pea aphid was done manually using NCBI BLASTx server. The top hits from this BLAST search were then manually searched for DEGs in pea aphid bacteriocytes based on the previous study (Smith & Moran, 2020). Additionally, BLASTx (NCBI, 2008) was performed using the fasta sequence for upregulated DEGs against the pea aphid protein database (GCF_005508785.1) obtained from Aphidbase (BIPAA, 2019).

Validation of DEGs

Before validating DEGs using qPCR, HMMER (v 2.3.2) search (Eddy, 2011) was performed to identify putative aphid amino acid transporters. The protein sequence (OGS6.0_20180125_proteins) for *A. glycines* biotype 1 was obtained from Aphidbase (BIPAA, 2017). OGS6.0_20180125_proteins was used as query to identify genes that contains amino acid transporters within the Pfam domains: AA_permease(PF00324.21), Aa_trans(PF01490.18) and APC_basic (PF05956.11) (El-Gebali et al., 2019) . The top 25 candidate transporters based on this search were chosen regardless of their expression profile in the RNAseq experiment. qPCR primers were designed for all 25 genes (Supplementary table 1) using Primer-BLAST (Ye et al., 2012). qPCR was conducted for all 25 genes using conditions from Supplementary figure 1 on CFX connect™ Real-Time System machine (BioRad®, Hercules, CA). The relative gene expression calculation for these 25 genes was based on the $2^{-\partial\partial C_T}$ method (Livak & Schmittgen, 2001). Gene expression level between bacteriocytes (BC) samples versus samples without bacteriocytes (WC) was calculated. The calculation was normalized with the whole-body sample (WB). RPS9 (40S ribosomal protein S9) served as a housekeeping gene for qPCR analysis. Any

putative amino acid transporter present in DEG analysis was cross-validated using the qPCR as described above.

Prediction of Transmembrane helices

TMHMM, a program that predicts transmembrane helices in proteins (DTU Bioinformatics, 2017), was used to predict transmembrane domains in genes from validated DEGs. FASTA sequence was supplied to the server. Generated transmembrane helices were based on Hidden Markov Model (Krogh et al., 2001; Sonnhammer et al., 1998).

CHAPTER 3 – RESULTS

Quality Control

Of the six cDNA libraries, three each for whole-body (WB) and bacteriocyte (BC), one of the three bacteriocyte-enriched (S1_BC) samples did not pass quality control (QC). Thus, the data analysis involves three whole-body (WB) samples than two bacteriocyte-enriched (BC) samples. For each sample sequenced, an average of 39.8 million clean paired-end reads was generated after raw reads filtering. Sequencing reads ranged from 27.5 million to 44.9 million in number (Table 1). The GC content of the samples is homogenous, suggesting that there were no sequencing errors. However, the GC content of the transcriptome was ~10% higher than the GC content of the assembled *A. glycines* genome (Wenger et al. 2020).

Table 1. Data quality control summary.

Sample	Raw reads	Clean reads	Raw bases	Clean bases	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
S1_WB	45526820	44731348	6.8G	6.7G	0.02	98.68	95.53	38.62
S2_WB	45668472	44930738	6.9G	6.7G	0.02	98.72	95.59	38.71
S3_WB	45205456	44390658	6.8G	6.7G	0.02	98.7	95.55	38.8
S2_BC	28308478	27589806	4.2G	4.1G	0.02	98.59	95.45	39.67
S3_BC	38303222	37460834	5.7G	5.6G	0.02	98.62	95.6	39.17

Raw reads represent original sequencing read counts. Clean reads were obtained after filtering raw reads. Raw bases and clean bases were calculated by multiplying read lengths (150 bps) with raw read or clean-read numbers respectively and saved in G unit. Error rate represents the average sequencing error rate and was calculated as $Q_{phred} = -10\log_{10}e$. Q20 and Q30 are the percentages of bases whose correct base recognition rates are greater than 99% and 99.9% in total bases, respectively. GC content is the percentages of G and C in total bases.

For the whole body sample, an average of 95% of the total reads was mapped to the genome, whereas for the bacteriocyte-enriched samples, an average of 77% mapped to the reference genome (Table 2).

Table 2. Overview of mapping status for reads mapped to *A. glycines* reference genome

Sample name	S1_WB	S2_WB	S3_WB	S2_BC	S3_BC
Total reads	44731348	44930738	44390658	27589806	37460834
Total mapped	42549339 (95.12%)	42786175 (95.23%)	42441936 (95.61%)	20362192 (73.8%)	30178197 (80.56%)
Multiple mapped	2774394 (6.2%)	4449547 (9.9%)	2088448 (4.7%)	2546211 (9.23%)	2339082 (6.24%)
Uniquely mapped	39774945 (88.92%)	38336628 (85.32%)	40353488 (90.91%)	17815981 (64.57%)	27839115 (74.32%)
Reads map to '+'	19882296 (44.45%)	19161617 (42.65%)	20167413 (45.43%)	8903308 (32.27%)	13914827 (37.15%)
Reads map to '-'	19892649 (44.47%)	19175011 (42.68%)	20186075 (45.47%)	8912673 (32.3%)	13924288 (37.17%)
Non-splice reads	21148865 (47.28%)	18613963 (41.43%)	20804953 (46.87%)	8454180 (30.64%)	14249847 (38.04%)
Splice reads	18626080 (41.64%)	19722665 (43.9%)	19548535 (44.04%)	9361801 (33.93%)	13589268 (36.28%)

The total reads represent clean filtered reads. Total mapped represents clean reads that were mapped to *A. glycines* reference genome (v6.0). Multiple mapped and uniquely mapped reads are the number of reads that were mapped to multiple sites and read uniquely mapped to this reference genome, respectively. Reads map to'+' and reads map to '-' are the number of reads mapped to the positive strand (+) and negative-strand (-), respectively. Splice reads are the number of reads either mapped to two exons or read that was segmented when mapped to this reference genome.

Of the 95% total reads from the whole-body samples, 88% of the reads were uniquely mapped, and only 69% of the reads were uniquely mapped in the bacteriocyte-enriched samples. Overall, on average 90% of the total mapped reads for both treatments were mapped to the exonic region of the genome (Figure 6).

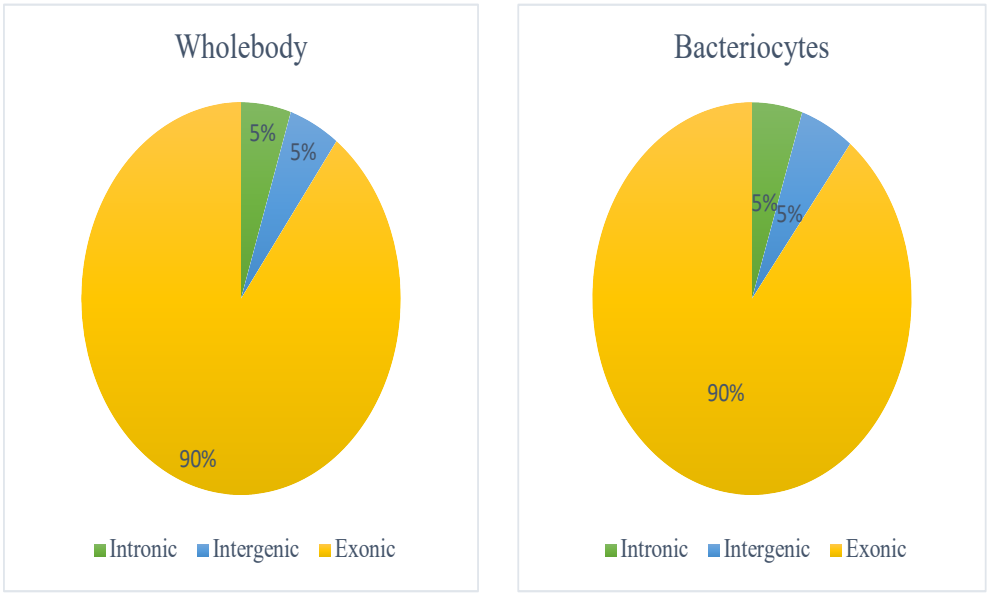


Figure 6: Reads mapped to the genomic regions of *A. glycines* reference genome.

The total reads obtained after mapping to the *A. glycines* reference genome shows that at least 90% of the reads were mapped to exons.

Bacteriocyte enrichment

Bacteriocyte samples were collected by dissecting soybean aphids at 4X under a stereomicroscope. For this purpose, adult apterous aphids of a similar age and size were chosen from the colony. After excision of the aphids' hind leg, the small circular cells (or bacteriocytes) that oozed from the leg (Figure 7) were collected. Bacteriocytes from 30 samples were pooled to form one replicate. To validate if our extraction protocol resulted in an enrichment of bacteriocytes, we aligned the clean reads to the reference genome of *Buchnera aphidicola* (Cassone et al., 2015). An average of 21.2% of clean reads for the whole-body sample mapped to the *Buchnera* genome, while an average of 31.9% of clean reads for the bacteriocyte samples mapped to the *Buchnera* genome (Table 3). Our results suggest that the dissection protocol that we used resulted in an enrichment of bacteriocytes.

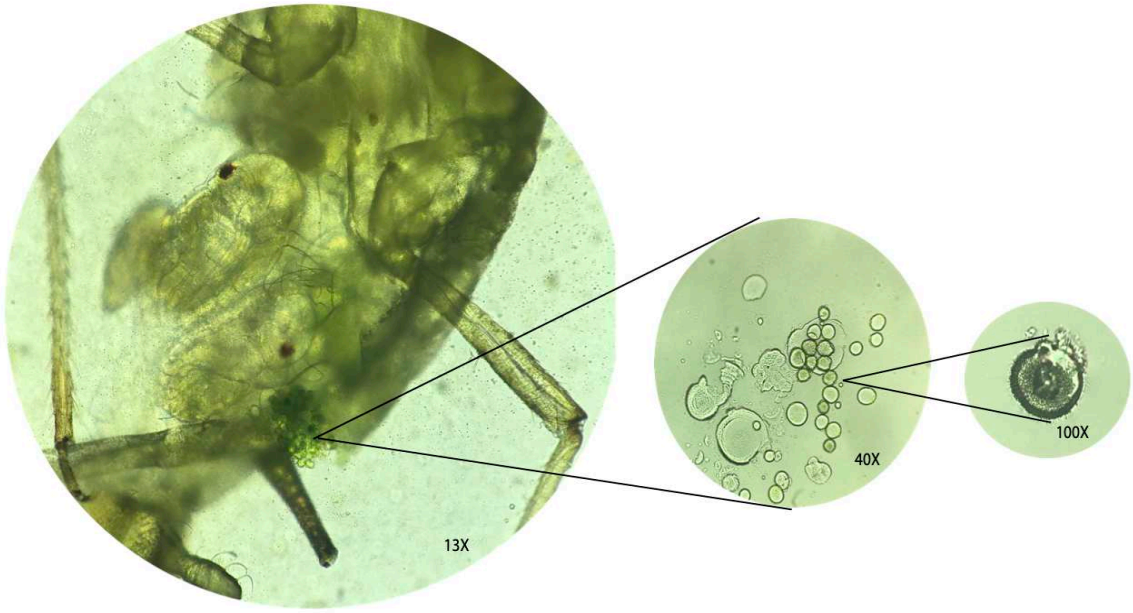


Figure 7: Bacteriocytes under the microscope

Aphids dissected under a stereo dissecting microscope at 4X magnification enlarged to 13X reveals clump of bacteriocytes. The magnified bacteriocyte clump is enlarged under 40X magnification. At 100X magnification, a single bacteriocyte visible with a distinct central nucleus surrounded by numerous *Buchnera.*

Table 3: Overview of mapping status for reads mapped to *B. aphidicola* reference genome

Sample name	S1_WB	S2_WB	S3_WB	S2_BC	S3_BC
Total reads	44731348	44930738	44390658	27589806	37460834
Total mapped	88753(19.8%)	113213(25.2%)	83041(18.7%)	896148(32.5%)	1175915(31.4%)
Uniquely mapped	43060(0.1%)	54171(0.12%)	40443(0.09%)	412237(1.5%)	548977(1.4%)

The total reads represent clean filtered reads. Total mapped represents clean reads mapped to the reference genome of *Buchnera aphidicola*. Total mapped and uniquely mapped reads are the total number of mapped reads and the reads uniquely mapped to this reference genome.

Differentially expressed genes in bacteriocytes

The uniquely mapped reads were selected for differential gene expression analysis followed by Log_{10} (FPKM+1) gene normalization for read length and sequencing depth (Supplementary figure 2). The repeatability of experimental replicate after Log_{10} (FPKM+1) normalization was measured using the Pearson-correlation coefficient. The whole-body replicates showed a high correlation ($R^2 \geq 0.93$) (Figure 6) among each other, while the bacteriocyte replicates had a weaker correlation ($R^2 \geq 0.72$) (Figure 7).

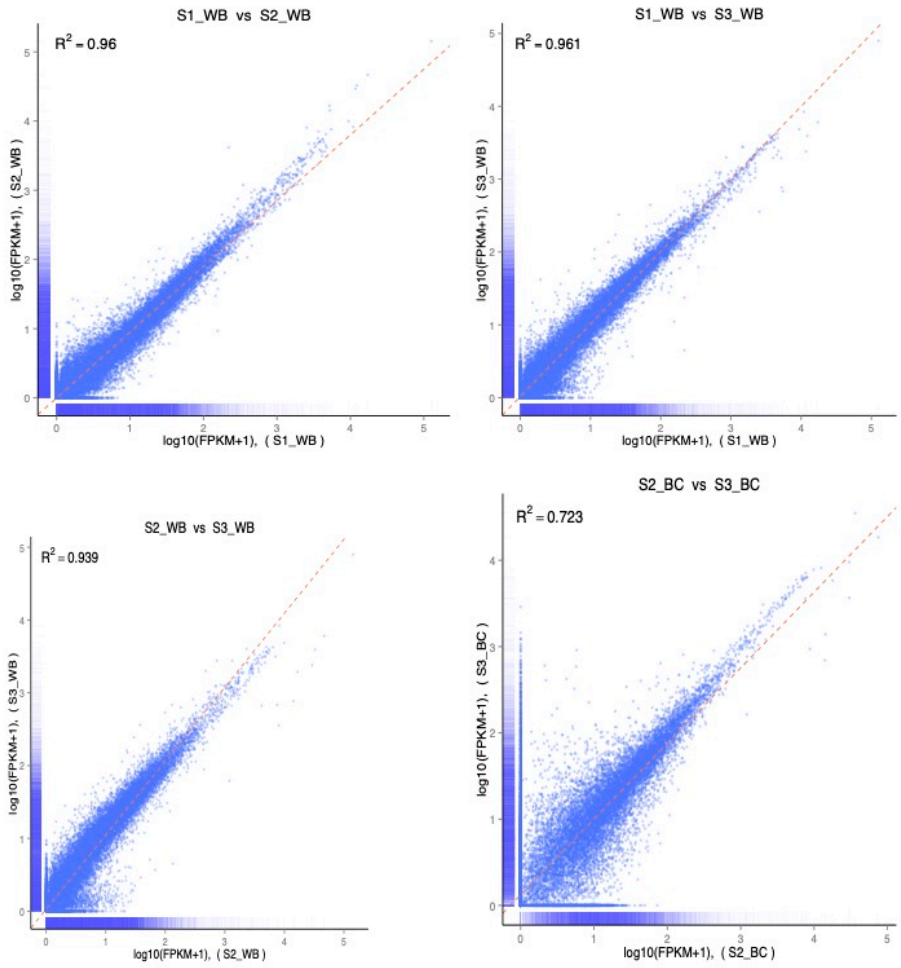


Figure 6: Scatter plot of normalized read counts(log_{10}(FPKM+1) between each replicate.

The plot demonstrates the correlation coefficient between the samples generated. R^2 is the square of the Pearson coefficient. Each axis represents normalized read count between replicates. (S1_W, S2_WB, and S3_WB are whole-body replicates. S2_BC and S3_BC are bacteriocyte replicates)

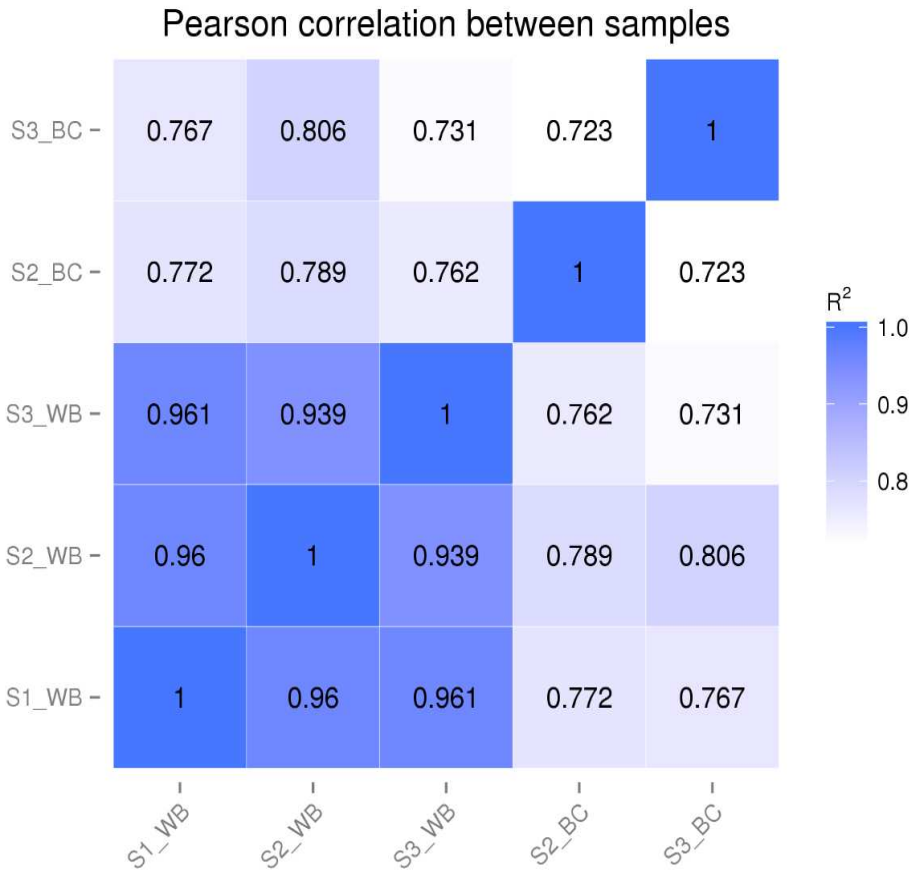


Figure 7: Pearson's correlation plot for normalized read counts (log_{10}(FPKM+1)). S1_WB, S2_WB, and S3_WB represent replicates for whole-body (WB) samples. S2_BC and S3_BC represent replicates for bacteriocyte (BC) samples.

Differential expression analysis between the whole body and bacteriocyte-enriched samples revealed that 2121 genes were differentially expressed. Of these, 1605 genes were down-regulated, and 516 genes were up-regulated in bacteriocyte-enriched samples (Figure 8; Supplementary table 2) as compared to the whole body. Among the up-regulated genes, 49% or 251 genes do not have a functional annotation or homology to proteins in the UniProt database. Thus, we chose a lenient log fold change cutoff of values greater than 1 to incorporate unannotated proteins that might have a significant role in our data analysis.

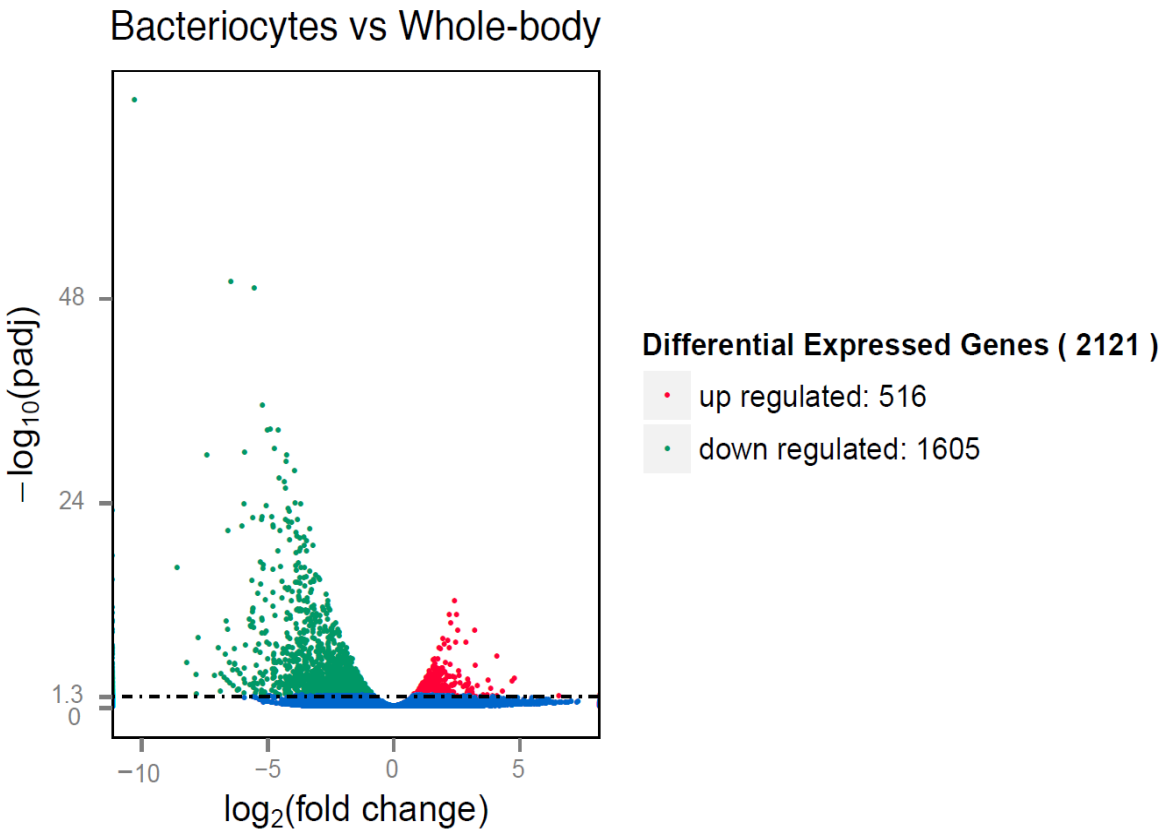


Figure 8: Volcano plot of differentially expressed genes.

Genes that are differentially expressed among bacteriocyte-enriched versus whole-body samples. The red dots show up-regulated (positive fold-change) genes in bacteriocytes and down-regulated in the whole body. The green dots show down-regulated (negative fold change) genes in bacteriocyte-enriched and up-regulated in the whole-body. The y-axis represents p-adjusted $\log_{10}$values of differentially expressed genes in five different replicates (Two from bacteriocyte-enriched and three from whole-body). The x-axis represents fold change in the expression of these genes among the two treatments. A positive fold change means a gene is significantly expressed to higher levels in bacteriocytes. Blue dots show genes that are not differentially expressed.

DEGs in soybean aphid bacteriocytes compared to DEGs in pea aphid bacteriocytes.

Similarities and differences between soybean and pea aphids were determined by comparing the top 100 up-regulated DEGs in bacteriocytes between the two species. The DEGs up-regulated in pea aphid bacteriocytes was obtained from (Smith & Moran, 2020). Of the 100 up-regulated DEGs, only four genes were common for both species. Additionally, five upregulated DEGs from our dataset were significantly downregulated in pea aphid bacteriocytes (Table 4).

Table 4: BLAST search of upregulated DEGs with upregulated DEGs in the pea aphid

Gene ID (*A.glycines*)	log2foldchange	Function	Gene ID (*A.pisum*)	Log2foldchange
AG6002106-RA	2.9	histone H3	LOC100166831	Negative
AG6035293-RA	2.8	putative nuclease HARBI1 transcript variant X1	LOC100573919	Negative
AG6028424-RA	2.2	neprilysin-2	LOC100159184	Negative
AG6032474-RA	2.2	uncharacterized protein	LOC100573825	Positive
AG6023685-RA	2.1	DNA mismatch repair protein Msh6	LOC100166946	Negative
AG6025782-RA	2.1	uncharacterized protein	LOC103307748	Positive
AG6028337-RA	2.0	lachesin like	LOC100168017	Negative
AG6009492-RA	2.0	uncharacterized protein	LOC100169409	Positive
AG6023289-RA	2.0	phenoloxidase 2	LOC100163393	Positive

The first and second columns show the gene identifier for our upregulated DEGs dataset for soybean aphid and its log2fold change value, respectively. The function column shows the BLAST result from the functional annotation of the pea aphid. The fourth and fifth columns show gene identifiers for pea aphid and the nature of upregulated DEGs in the previous study (Smith & Moran, 2020).

Furthermore, we also ran a manual BLASTp to identify a homolog of ApGLNT1, the glutamine transporter in the pea aphid. Our homology search revealed a putative amino acid transporter gene: AG6017791-RA. However, AG6017791-RA did not show significant differential expression in our RNAseq data.

GO analysis for DEGs in bacteriocytes

The Gene Ontology (GO) enrichment analysis revealed 2553 GO terms assigned for each domain (cellular component, biological processes, and molecular functions). GO analysis included 2307 GO terms for down-regulated DEGs. However, only 72 of these GO enriched terms were significant ($p < 0.05$) (Supplementary table 3). Among the significant GO terms for down-regulated genes, more than 500 genes were associated with the membrane (GO:0005575, domain: cellular component), 100 genes were associated with transport activity(GO:0022892, domain:

molecular function), and more than 150 genes were associated with localization for GO biological processes (GO:1902578) (**Figure 9**, **Supplementary table 3**).

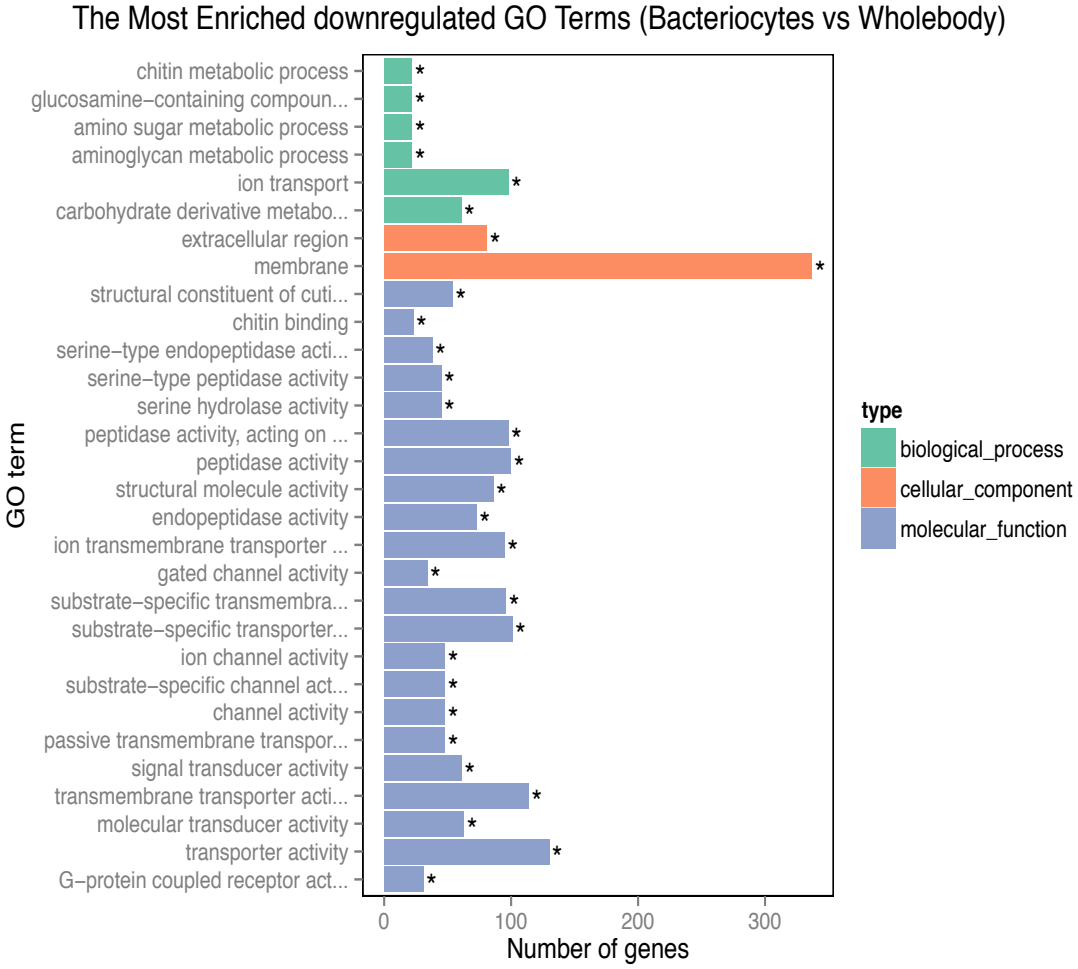


Figure 9. The most enriched GO terms for down-regulated DEGs.

Different colors are used for distinct biological processes, cellular components, and molecular function. GO terms marked by asterisks "*" shows genes within the GO terms with padj < 0.05 adjusted after multiple testing following Wallenius approximation model.

GO enrichment analysis for up-regulated DEGs revealed 1262 GO terms. More than 200 genes were associated with the binding activity (GO:0003674, domain: molecular function), more than 170 genes for GO biological processes (GO:0008150), and more than 100 genes for cellular component (GO:0005575). However, none of these GO terms for up-regulated genes were significant ($p < 0.05$) (Figure 10).

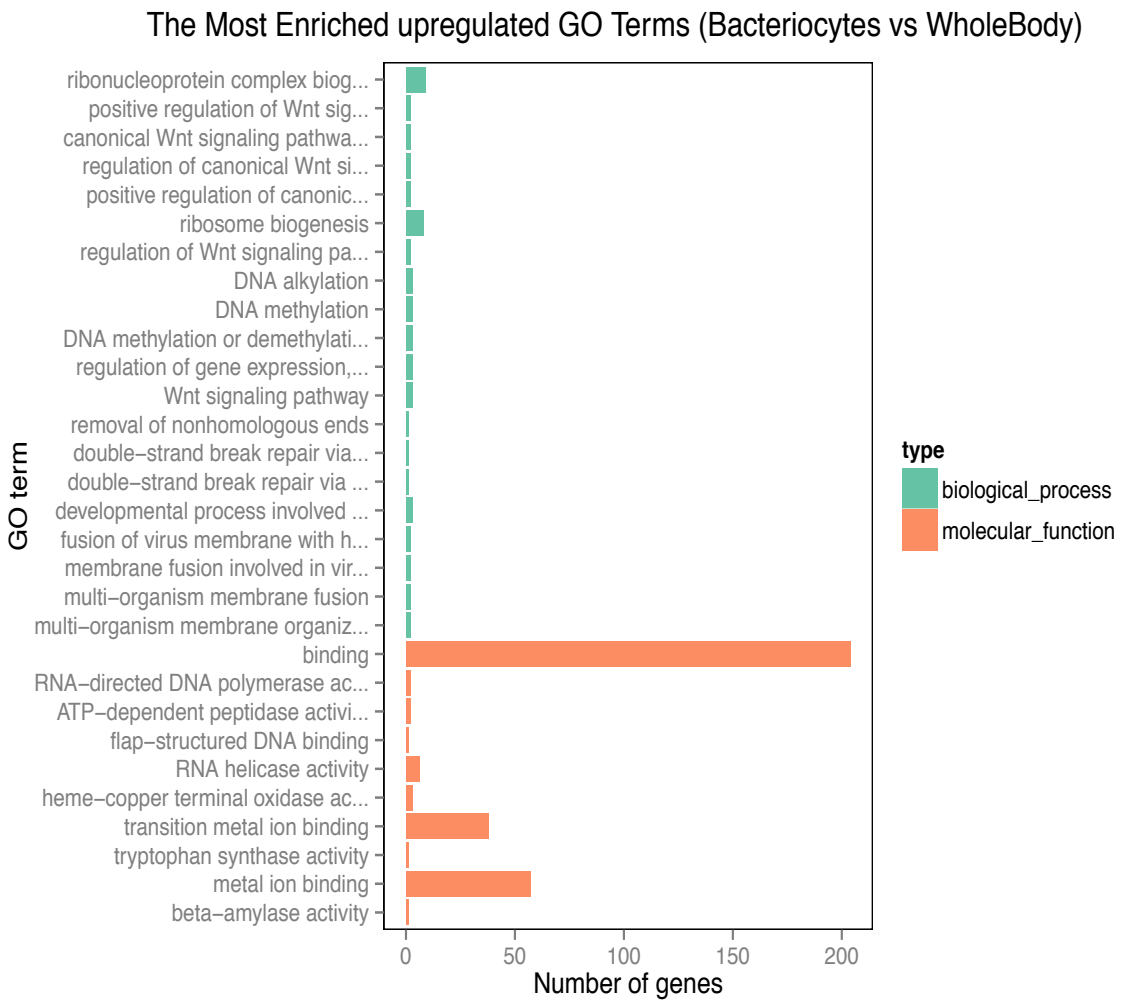


Figure 10: The most enriched GO terms for up-regulated DEGs.

The y-axis is GO terms enriched, and the x-axis is the number of differentially expressed genes. Different colors are used for distinct biological processes, cellular components, and molecular functions.

One of the primary goals of the RNA-seq was to identify genes involved in transport at the aphid-*Buchnera* symbiotic interface. We found that 236 GO terms were assigned to differentially expressed genes involved in the transport of ions, sugar, amino acids, and other substrates. Eight GO terms related to transport were significant ($p<0.05$) (Supplementary table 4). Sixty-eight GO terms were associated with transport among up-regulated DEGs in bacteriocytes. However, none of them were significant ($p <0.05$).

KEGG pathway enrichment analysis for DEGs

Pathway enrichment analysis for DEGs from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database showed 108 different KEGG pathway descriptors. KEGG pathway analysis for down-regulated DEGs had 92 KEGG descriptors. Only three out of 92 KEGG descriptors for down-regulated DEGs were significant ($p < 0.05$). Among the three KEGG pathway descriptors; 96 genes were associated with metabolic pathways, 15 with pathways involving neuroactive ligand-receptor interaction, and ten with phototransduction activity. KEGG pathway analysis for up-regulated DEGs had 55 KEGG descriptors. Two out of 55 KEGG descriptors for up-regulated DEGs were significant ($p <0.05$). Among these two KEGG descriptors, seven genes were associated with the ribosome biogenesis pathway, and four genes were associated with the pathway for biosynthesis of unsaturated fatty acids (Table 5, Figure 11). KEGG pathway analysis for up-regulated DEGs had 55 KEGG descriptors. Two out of 55 KEGG descriptors for up-regulated DEGs were significant ($p <0.05$). Among these two KEGG descriptors, seven genes were associated with the ribosome biogenesis pathway, and four genes were associated with the pathway for biosynthesis of unsaturated fatty acids (Table 6, Figure 12).

Table 5: KEGG pathways for differentially downregulated genes with $p < 0.05$

#Term	Pathway ID	Input number	P-value	Corrected p-value
Metabolic pathways	api01100	96	1.30E-03	0.040836368
Neuroactive ligand-receptor interaction	api04080	15	8.52E-06	0.000745975
Phototransduction	api04745	10	1.59E-05	0.000745975

Term: The description of KEGG pathways.
Input number: Number of DEGs with pathway annotation.
P-value: P-value in the hypergeometric test.
Corrected P-value: pathway with Corrected P-value < 0.05 are significantly enriched in DEGs.

Table 6: KEGG pathways for differentially upregulated genes with p < 0.05

#Term	Pathway ID	Input number	P-value	Corrected p-value
Ribosome biogenesis in eukaryotes	api03008	7	1.70E-03	0.04664208
Biosynthesis of unsaturated fatty acids	api01040	4	5.73E-04	0.03149484

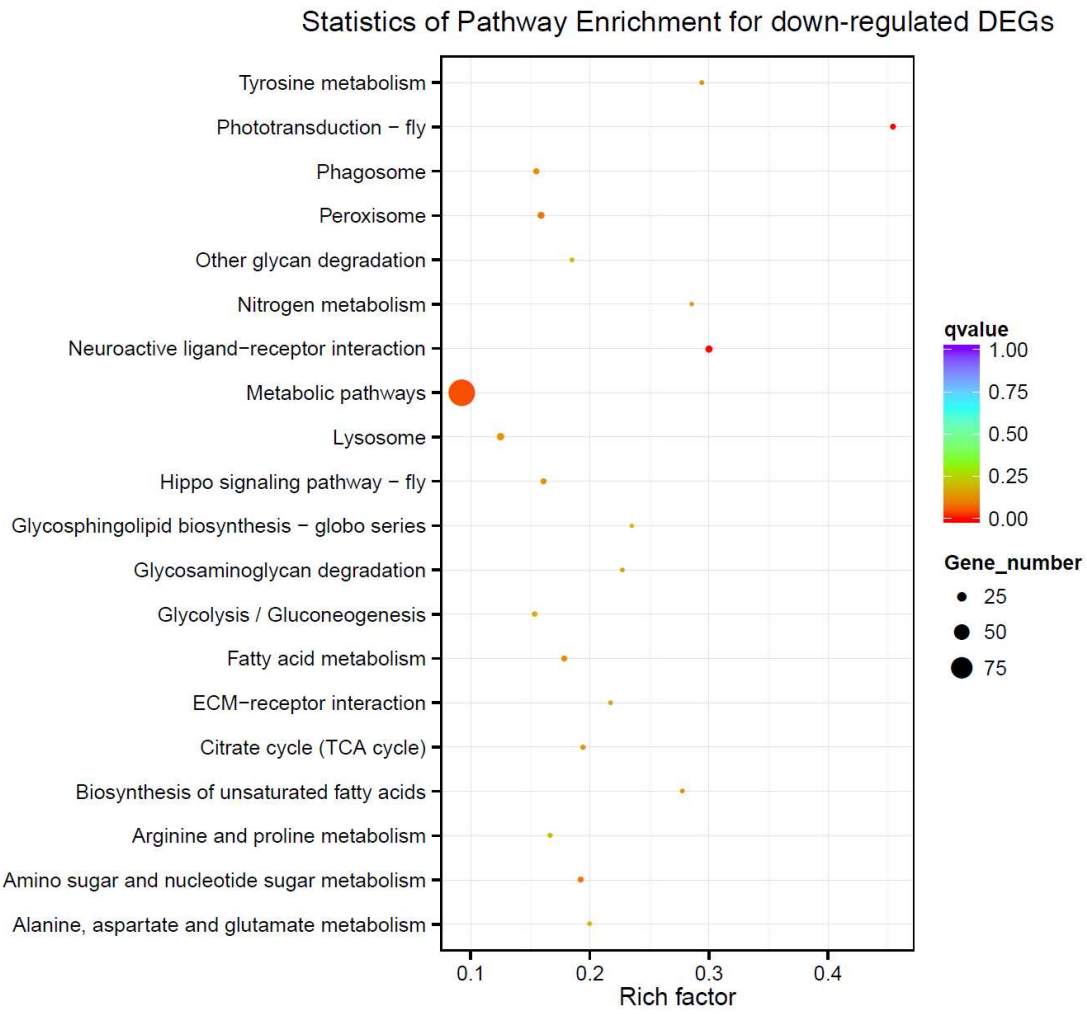


Figure 11: KEGG pathway analysis for differentially down-regulated genes.

The y-axis shows the pathway name, and the x-axis shows the Rich factor. Dot size represents the number of different genes. q-value represent pathways with the corrected p-value with a significant level of 0.05.

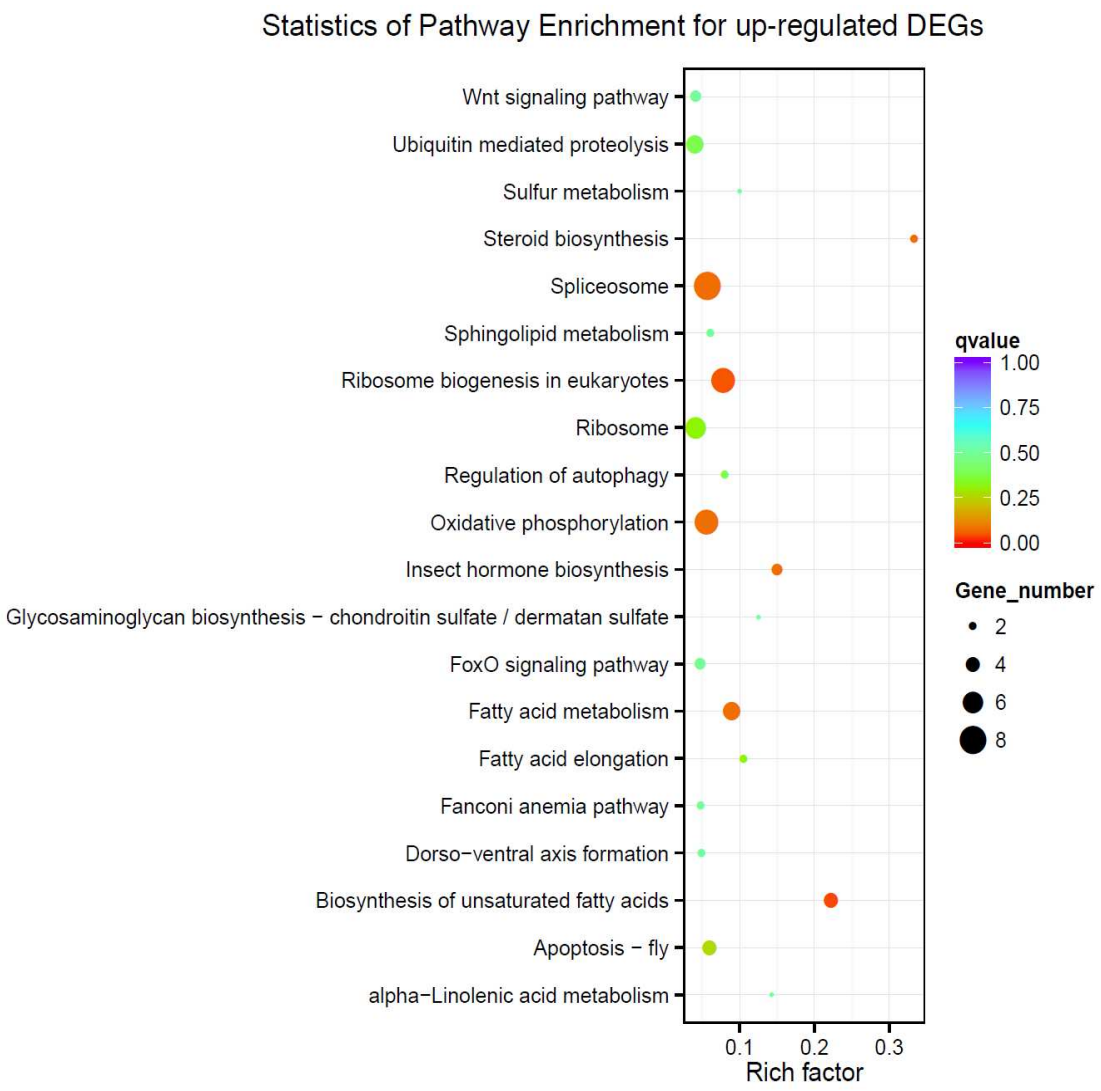


Figure 12: KEGG pathway analysis for differentially up-regulated genes.

The y-axis shows the pathway name, and the x-axis shows the Rich factor. Dot size represents the number of different genes. q-value represents pathways with the corrected p-value with a significant level of 0.05.

Differential expression analysis identifies genes involved in transport activity

Differential expression analysis showed 59 transcripts involved in the transport of ions, sugar, amino acids, and other substrates. Of these, the majority (56 transcripts) were down-regulated in the bacteriocytes, and only three transcripts were up-regulated (Supplementary table 5, Table 7).

Table 7: Transport related differentially up-regulated genes

Gene_id	log2 Fold Change	pval	padj	Uniprot_Clus	Function
AG6031606-RA	4.69	0.00	8.00E-04	A9ZSY3	Facilitated trehalose transporter
AG6020433-RA	1.21	0.00	6.00E-03	Q28J44	Zinc transporter
AG6005572-RA	1.16	0.00	5.00E-03	Q09143	High affinity cationic amino acid transporter 1

Gene_id: The gene identifier for differentially expressed genes
log2FoldChange: log2(BC/WB)
pvalue(pval): The statistical p-value
qvalue(padj): p-value after normalization. The smaller the q-value is, the more significant the difference is.
Uniprot_clus: Uniprot ID belonging to the eukaryotic cluster group.

Based on the functional annotation available for soybean aphids, our differential expression analysis showed only one gene with a putative function in amino acid transport. Based on the gene identifier from the soybean aphid biotype 1 genome, this amino acid transporter has a gene ID: AG6005572-RA (GeneID-72). GeneID-72 is a 280 aa long, high-affinity cationic amino acid transporter. The genome sequence of biotype 4 soybean aphid is also available (Mathers, 2020), and BLASTp reveals a paralog for GeneID-72, AG006001-PA (GeneID-01). Although geneID-01 is listed as a high-affinity cationic amino acid transporter, the gene model in the biotype 4 genome shows a protein consisting of 597 aa residues making it twice the length of GeneID-72.

HMMER search identifies putative amino acid transporters

Previous studies with pea aphids indicate that at least ten amino acid transporters are up-regulated in bacteriocytes (Price et al., 2014). However, in our analysis, we were only able to identify one amino acid transporter, which could, in part, be due to a large number of unannotated transcripts. To identify the complete set of soybean amino acid transporters that function at the aphid-*Buchnera* symbiotic interface, we used HMMER to search the soybean aphid genome for proteins containing the Pfam domains; PF00324 (AA permease) and PF01490 (AA transporter

protein) (Price et al., 2011). We identified 25 *A. glycines* genes with sequence homology with AA permeases (Supplementary table 6) and 22 genes with homology with AA transporter proteins (Supplementary table 7). However, none of these identified amino acid transporters showed differential expression in our RNA-seq data.

To ensure that this was indeed the case, we determined the relative gene expression of these genes in a fresh set of samples. An additional sample type was included in this analysis, consisting of the whole aphid from which the bacteriocytes had been removed (WB-BC = WC). qPCR analysis of the 25 putative amino acid transporter shows that only two genes: AG6031286-RA (GeneID-86) and AG6005572-RA(GeneID-72), have significantly higher expression in bacteriocytes (Figure 13; Supplementary table 8). While GeneID-72 was present as an up-regulated gene in our RNAseq DEGs analysis, GeneID-86 is not. Additionally, AG6017791-RA, the homolog to ApGLNT1 in the pea aphid, did not show bacteriocyte enrichment in our qPCR analysis (Supplementary table 8).

Transmembrane domain prediction for putative amino acid transporters

Transmembrane helices prediction using TMHMM server 2.0 identified transmembrane helices based on supplied amino acid sequences. Six transmembrane helices were present in the AG6005572-RA (Figure 14a, Supplementary table 9), while 13 transmembrane helices were present in AG006001-PA (Figure 14b, Supplementary table 10) spanning the plasma membrane. Eleven transmembrane helices were present in AG6031286-RA, spanning the plasma membrane (Figure 14c, Supplementary table 11).

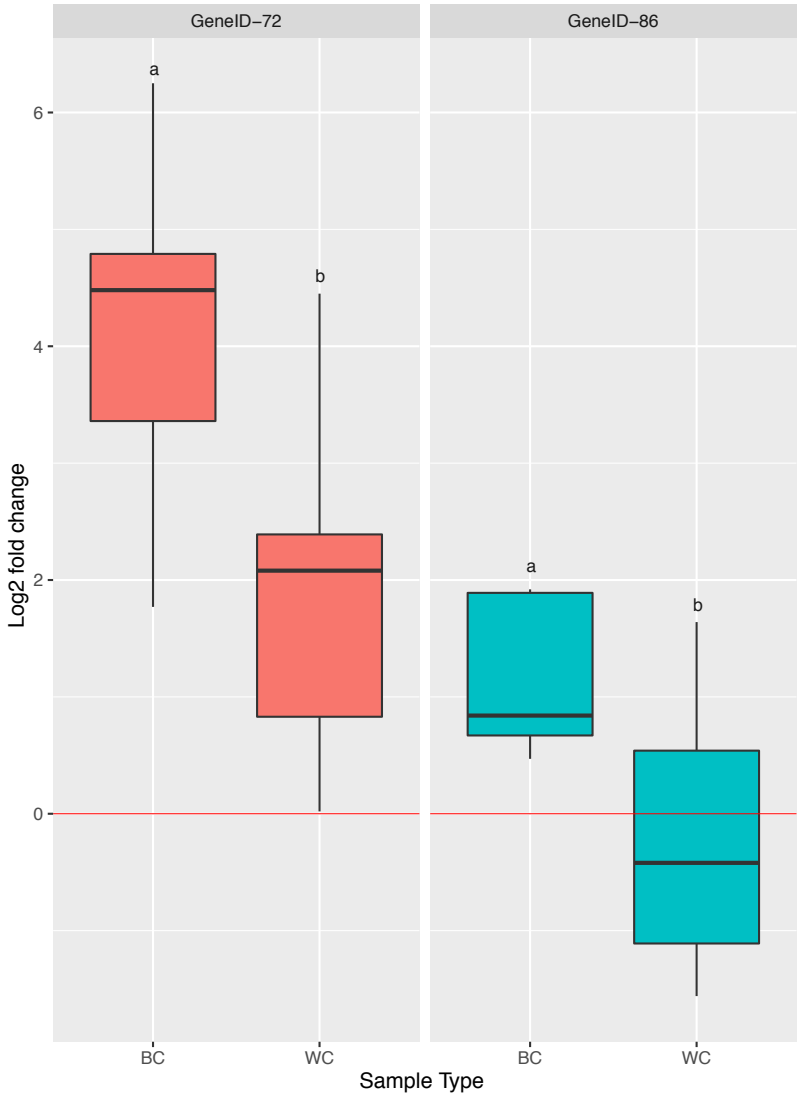


Figure 13: Validation of differentially expressed genes (DEGs) from quantitative PCR (qPCR).
GeneID-72, a high-affinity cationic amino-acid transporter obtained from RNAseq data, was enriched in hemolymph enriched bacteriocytes. GeneID-86 shows bacteriocyte enrichment among the top 25 amino-acid transporters chosen from HMMER search on the Pfam domain. GeneID-86 was not differentially expressed in our RNAseq data.
BC: Hemolymph enriched bacteriocyte cells
WC: Aphid with hemolymph enriched bacteriocyte cells removed
WB: BC and WC were normalized against aphid whole body (WB) aphid.

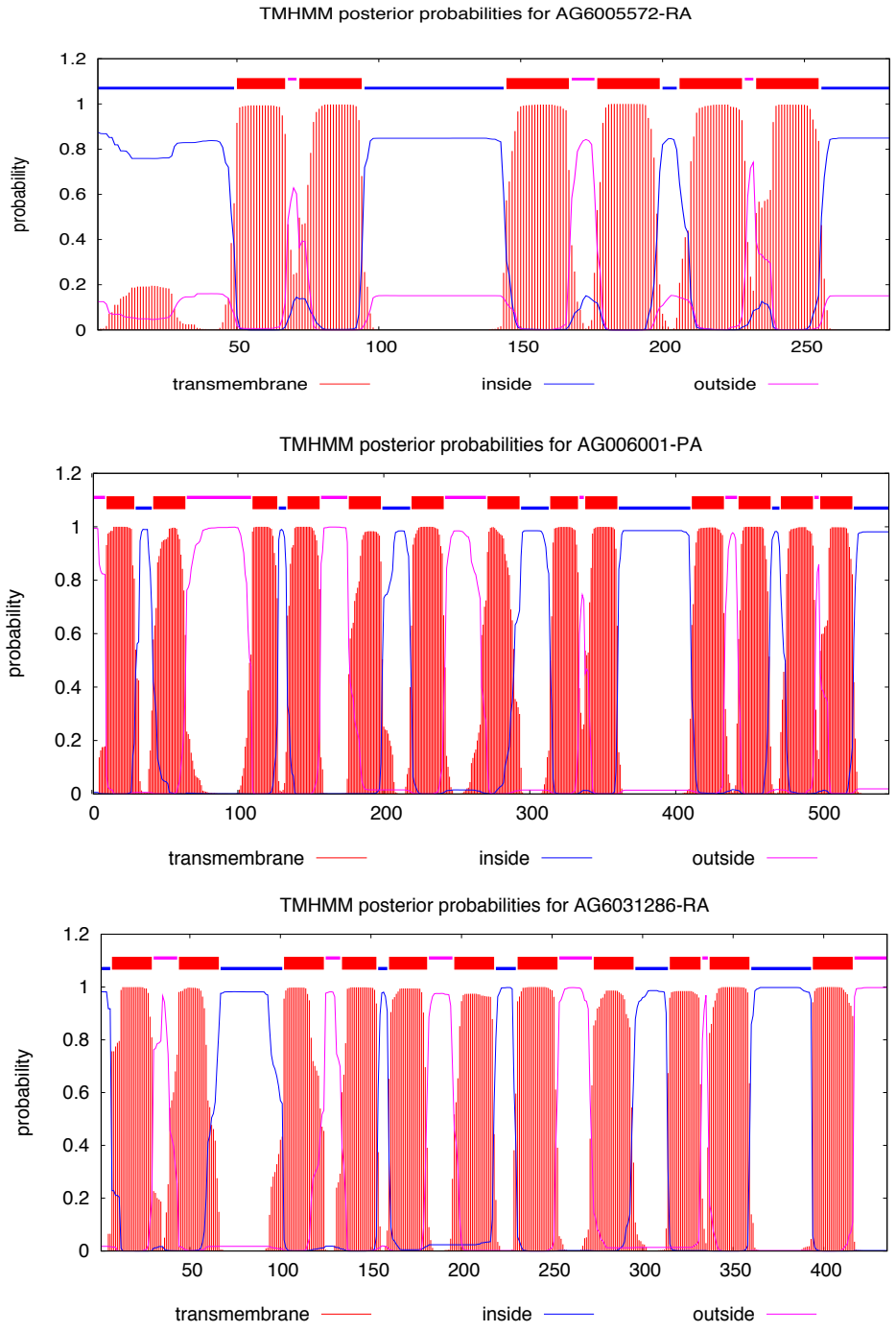


Figure 14: **Prediction model of transmembrane helices in putative amino acid transporter.** The prediction model of transmembrane helices for a. AG6005572-RA shows six helices b. AG006001-PA shows thirteen helices and c. AG6031286-RA shows eleven helices spanning the plasma membrane. The red vertical bar shows the transmembrane domain, while the horizontal red box shows the gene associated with each helix. The pink and blue zigzag lines show outside and inside of the cell membrane, respectively. The red vertical bar spans across these membrane layers.

CHAPTER 4 – DISCUSSION

One of the most important factors to consider in our study is the enrichment of bacteriocytes in our samples. The protocol for the bacteriocyte enrichment we followed has been used previously for pea aphids (Feng et al., 2019; Price et al., 2011; Price et al., 2014b; Smith & Moran, 2020). Microscopy using compound microscopes indicates that the dissection protocol we followed allowed for the collection of *Buchnera* -enriched bacteriocytes. Besides, despite eukaryotic library preparation being used for sequencing our experimental samples, we aligned the RNA sequence reads to the reference genome of soybean aphid's *Buchnera aphidicola*. The significance of this result is the successful isolation of bacteriocyte samples containing *Buchnera.*

Furthermore, our dataset's bacteriocyte enrichment was validated by comparing differentially expressed genes to pea aphid bacteriocytes from previous studies (Duncan et al., 2014; Smith & Moran, 2020). The homology search for up-regulated DEGs in soybean and pea aphid bacteriocytes revealed four genes in common to both aphids. However, three out of four upregulated genes in bacteriocytes were uncharacterized protein. One of the genes, AG6023289-RA, shows homology to LOC100163393 in pea aphids is annotated as a phenoloxidase enzyme. Phenol oxidases are abundant in insect hemolymph and act as an essential part of insects' innate immune against entomopathogens (González & Córdoba, 2012). The enrichment of this enzyme in both aphids' bacteriocytes indicates a vital role for *Buchnera* in the synthesis and activation of phenoloxidases, possibly by supplying precursor amino acids for enzyme synthesis. However, the enrichment of this enzyme could solely be due to hemolymph presence in our bacteriocyte samples.

The bacteriocyte cell contains several metabolic pathways that reflect the expression level's differences compared to the aphid body. We found an amino acid transporter AG6005572-RA that is enriched in *A. glycines* bacteriocytes. Our result was based on the genome annotation of Biotype 1 *A. glycines*. The genome annotation of *A. glycines* Biotype 4 is also recently available (Mathers, 2020). Based on the newer annotation of Biotype 4 *A. glycines*, AG6005572-RA had a homolog: AG006001-PA. In-fact, AG6005572-RA was a partial sequence of AG006001-PA. The partial annotation explains smaller amino acid residue in AG6005572-RA, possibly due to poor functional annotation of Biotype 1 *A. glycines*. Protein homology search with the model aphid *A. pisum* shows that AG6005572-RA has sequence similarity with a cationic amino acid transporter ACYPI003923. AG6005572-RA was also similar to *Drosophila melanogaster* Slimfast CG11128-PC (Hoskins et al., 2015). This cationic amino acid transporter belongs to the Amino acid/Auxin permease (AAAP) family with 14 transmembrane domains (Duncan et al., 2014). Based on our transmembrane domain search, AG006001 had 13 transmembrane domains. AG6005572 had only six transmembrane domains, which also provide additional support for partially annotated AG6005572-RA. The A. *pisum amino* acid transporter ACYPI003923, however, showed differential gene expression in late embryo stage (Rabatel et al., 2013). Bacteriocyte enrichment for ACYPI003923 was absent in *A. pisum*.

Many differentially up-regulated genes in our dataset were either annotated as hypothetical protein or did not have a functional annotation at all. While the new, improved genome annotation is available for Biotype 4 *A. glycines*, an improved functional annotation for Biotype 1 *A.glycines* is also necessary. A well-built, functional annotation could provide insights into what the hypothetical proteins in our dataset encode. Our dataset's top 10 candidate genes with the highest log fold change activity had no functional annotation except for a trehalose transporter.

Our differential expression data revealed a sugar transporter: AG6031606-RA, which is a putative transporter of trehalose. Several trehalose transporters and trehalase metabolism genes were enriched in bacteriocytes in the pea aphid (Hansen & Moran, 2011; Smith & Moran, 2020). Trehalose is considered a significant glucose source in bacteriocytes. The revelation of several trehalose transporters in pea aphids with sequence homology with AG6031606-RA shows that trehalose transport is vital in bacteriocyte cells. Thus, the trehalose transporter: AG6031606-RA from our dataset is possibly involved in transport from hemolymph to bacteriocytes. An experimental validation would be required to ascertain the transporter activity of AG6031606-RA. Our plan is to characterize the plasma membrane transport proteins using *Xenopus* oocyte standard transport assays based on previous studies (Price et al., 2014a; Yao et al., 2000).

Conclusion

In summary, our research identifies a candidate amino acid transporter AG6005572-RA. The localization of this transporter could either be on the outer bacteriocyte membrane or the symbiosomal membrane. Thus, experimental validation is required to answer the localization, which can be performed using immunolocalization techniques. The functional characterization of AG6005572-RA is also needed to identify which amino acids are transported across the plasma membrane within the aphid-*Buchnera* interface. The validated experimental implementation of localization and functional characterization of two amino acid transporters in pea aphid was performed in Xenopus oocyte (Feng et al., 2019; Price et al., 2014b). It remains to be seen whether a similarly experimental process can be applied to our candidate amino acid transporter. The location and function of AG6005572-RA within the compartmentalized bacteriocytes will be our subsequent research.

Future Directions

Soybean aphid is one of the major pest problems of soybean in the United States' Midwest region. Soybean aphid has combated several IPM strategies that include cultural, chemical, and genetic control strategies over time. The need for biological control is necessary—our research aimed at identifying amino acid transporters at the soybean aphid-*Buchnera* symbiotic interface. The dependency of soybean aphid on *Buchnera* for essential amino acids provides a compelling reason to study the interaction. The primary interaction between soybean aphid and *Buchnera* is to transport amino acids from the aphid body to *Buchnera* and vice versa. The transport occurs between two distinct membranes localized within bacteriocytes.

The current study involved the isolation of bacteriocytes from aphids, followed by RNA sequencing. We identified a single amino acid transporter that was enriched in bacteriocytes. However, multiple such amino acid transporters are enriched in pea aphid bacteriocyte. The enrichment of a single amino acid transporter in soybean aphid than pea aphid could be due to multiple reasons, including poor functional annotation of biotype 1 soybean aphid used in our current study, insufficient and improper bacteriocyte collection method or data analysis artifacts. We collected bacteriocytes from soybean aphids by collecting hemolymph oozing out from the hind leg. This method was chosen for two reasons: Bacteriocytes are suspended in hemolymph at the posterior region, forming a U-shaped organ. The collection of hemolymphs from this location would also mean collecting bacteriocytes. The bacteriocytes collection method in other research that involved pea aphid was done by dissection. The pea aphid was dissected to separate body organs from bacteriocytes. Isolated bacteriocytes were collected for sequencing. Since soybean aphid is much smaller than pea aphid, a precise dissection is difficult.

For this reason, enough bacteriocytes might not have been collected in our experiment. As a result, significant enrichment in bacteriocyte genes was not prominent in our RNA sequenced data. Thus, a modification to our sample collection method in the future might provide additional gene expression data. The improvement to the current bacteriocyte collection method could be done by using hemocytometer to count the number of bacteriocyte cells before extracting RNA from these bacteriocytes.

The current RNA sequencing of bacteriocytes and aphid body samples were done on a pool of samples collected. Pooling might lead to sample variation. A high-resolution sequencing approach would be a more rigorous analysis. Single-cell RNA sequencing of isolated bacteriocytes may provide a precise measure of genes enriched in bacteriocytes. However, this method would also require would require proper dissection and isolation of individual bacteriocytes.

Apart from sample collection and sequencing methods, our research's limitation was poor genome annotation of soybean aphid. Soybean aphids are classified based on which soybean resistant genes they can overcome. The soybean aphid in our experiment was a biotype 1 variety. Due to the poor functional annotation of biotype 1 soybean aphids, many differentially expressed genes in our dataset did not have any functional annotation. Additionally, many genes that were differentially expressed did not have a complete sequence.

We collected bacteriocytes from soybean aphids by collecting hemolymph oozing out from the hind leg after dissection, a similar method applied to pea aphid bacteriocyte isolation. Isolated bacteriocytes were collected for sequencing. Since soybean aphid is much smaller than pea aphid, a precise dissection, however, is difficult. For this reason, enough bacteriocytes might not have been collected in our experiment. As a result, significant enrichment in bacteriocyte genes was not prominent in our RNA sequenced data. Thus, a modification to our sample collection method in

the future might provide additional gene expression data. The improvement of the current bacteriocyte collection method could be done using a hemocytometer to count the bacteriocyte cells' number before extracting RNA from these bacteriocytes.

Consequently, functionally annotated biotype 1 soybean aphid might have a partial gene sequence. These partial gene sets could also just be a section of an intron. The de-novo sequencing method could annotate the un-annotated or poorly annotated genes in our future RNA sequencing data analysis. The de-novo sequence annotation could provide stringent information on protein functions.

REFERENCES

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

BIPAA. (2017). *Aphis glycines Ag_bt1 | Aphis glycines | BIPAA*. https://bipaa.genouest.org/sp/aphis_glycines/organism/Aphis/glycines-Ag_bt1

BIPAA. (2019). *BioInformatics Platform for Agroecosystem Arthropods*.

Cassone, B. J., Wenger, J. A., & Michel, A. P. (2015). Whole Genome Sequence of the Soybean Aphid Endosymbiont Buchnera aphidicola and Genetic Differentiation among Biotype-Specific Strains . *Journal of Genomics*, *3*, 85–94. https://doi.org/10.7150/jgen.12975

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (Rhopalosiphum maidis Fitch). *GigaScience*, *8*(4). https://doi.org/10.1093/gigascience/giz033

Chong, R. A., & Moran, N. A. (2018). Evolutionary loss and replacement of Buchnera, the obligate endosymbiont of aphids. *ISME Journal*, *12*(3), 898–908. https://doi.org/10.1038/s41396-017-0024-6

Chong, R. A., Park, H., & Moran, N. A. (2019). Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz082

Douglas, A. E. (1998). Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria Buchnera. In *Annual Review of Entomology* (Vol. 43, pp. 17–37). https://doi.org/10.1146/annurev.ento.43.1.17

Douglas, A. E. (2006). Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*, *57*(4), 747–754. https://doi.org/10.1093/jxb/erj067

Douglas, A. E., Bouvaine, S., & Russell, R. R. (2011). How the insect immune system interacts with an obligate symbiotic bacterium. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1704), 333–338. https://doi.org/10.1098/rspb.2010.1563

DTU Bioinformatics. (2017). *TMHMM Server, v. 2.0*. http://www.cbs.dtu.dk/services/TMHMM/

Duncan, R. P., Husnik, F., Van Leuven, J. T., Gilbert, D. G., Dávalos, L. M., McCutcheon, J. P., & Wilson, A. C. C. (2014). Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Molecular Ecology*, *23*(6), 1608–1623. https://doi.org/10.1111/mec.12627

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10). https://doi.org/10.1371/journal.pcbi.1002195

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Feng, H., Edwards, N., Anderson, C. M. H., Althaus, M., Duncan, R. P., Hsu, Y.-C., Luetje, C. W., Price, D. R. G., Wilson, A. C. C., & Thwaites, D. T. (2019). Trading amino acids at the aphid– Buchnera symbiotic interface . *Proceedings of the National Academy of Sciences*, *116*(32), 16003–16011. https://doi.org/10.1073/pnas.1906223116

González-Santoyo, I., & Córdoba-Aguilar, A. (2012). Phenoloxidase: A key component of the insect immune system. In *Entomologia Experimentalis et Applicata* (Vol. 142, Issue 1, pp. 1–16). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1570-7458.2011.01187.x

Guo, J., Hatt, S., He, K., Chen, J., Francis, F., & Wang, Z. (2017). Nine facultative endosymbionts in aphids. A review. In *Journal of Asia-Pacific Entomology* (Vol. 20, Issue 3, pp. 794–801). https://doi.org/10.1016/j.aspen.2017.03.025

Hansen, A. K., & Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2849–2854. https://doi.org/10.1073/pnas.1013465108

Herren, J. K., Paredes, J. C., Schüpfer, F., & Lemaitre, B. (2013). Vertical transmission of a Drosophila endosymbiont via cooption of the yolk transport and internalization machinery. *MBio*, *4*(2). https://doi.org/10.1128/mBio.00532-12

Heyworth, E. R., & Ferrari, J. (2015). A facultative endosymbiont in aphids can provide diverse ecological benefits. *Journal of Evolutionary Biology*, *28*(10), 1753–1760. https://doi.org/10.1111/jeb.12705

Hill, J. H., & Whitham, S. A. (2014). Control of Virus Diseases in Soybeans. In *Advances in Virus Research* (Vol. 90, pp. 355–390). Academic Press Inc. https://doi.org/10.1016/B978-0-12-801246-8.00007-X

Hohn, T. (2007). Plant virus transmission from the insect point of view. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 104, Issue 46, pp. 17905–17906). https://doi.org/10.1073/pnas.0709178104

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., Pfeiffer, B. D., George, R. A., Svirskas, R., Krzywinski, M., Schein, J., Accardo, M. C., Damia, E., Messina, G., Méndez-Lago, M., De Pablos, B., Demakova, O. V., Andreyeva, E. N., … Celniker, S. E. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research*, *25*(3), 445–458. https://doi.org/10.1101/gr.185579.114

Houk, E. J., & Griffiths, G. W. (1980). Intracellular symbiotes of the Homoptera. *Annual Review of Entomology*, *25*(1), 161–187.

Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., & Mortensen, D. A. (2017). Agriculture in 2050: Recalibrating targets for sustainable intensification. In *BioScience* (Vol. 67, Issue 4, pp. 386–391). Oxford University Press. https://doi.org/10.1093/biosci/bix010

Jack, D. L., Paulsen, I. T., & Saier, J. (2000). The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology*. https://doi.org/10.1099/00221287-146-8-1797

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. In *Nucleic Acids Research* (Vol. 28, Issue 1, pp. 27–30). https://doi.org/10.1093/nar/28.1.27

Kikawada, T., Saito, A., Kanamori, Y., Nakahara, Y., Iwata, K. I., Tanaka, D., Watanabe, M., & Okuda, T. (2007). Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(28), 11585–11590. https://doi.org/10.1073/pnas.0702538104

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360.

https://doi.org/10.1038/nmeth.3317
Koga, R., Meng, X. Y., Tsuchida, T., & Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(20). https://doi.org/10.1073/pnas.1119212109
Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315
Li, Y., Park, H., Smith, T. E., Moran, N. A., & Singh, N. (2019). Gene Family Evolution in the Pea Aphid Based on Chromosome-Level Genome Assembly. *Molecular Biology and Evolution*, *36*(10), 2143–2156. https://doi.org/10.1093/molbev/msz138
Liu, S., Chougule, N. P., Vijayendran, D., & Bonning, B. C. (2012). Deep Sequencing of the Transcriptomes of Soybean Aphid and Associated Endosymbionts. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045161
Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262
Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. In *Genome Biology* (Vol. 15, Issue 12). https://doi.org/10.1186/s13059-014-0550-8
Mathers, T. C. (2020). Improved genome assembly and annotation of the soybean aphid (aphis glycines matsumura). *G3: Genes, Genomes, Genetics*, *10*(3), 899–906. https://doi.org/10.1534/g3.119.400954
Mi, G., Di, Y., Emerson, S., Cumbie, J. S., & Chang, J. H. (2012). Length Bias Correction in Gene Ontology Enrichment Analysis Using Logistic Regression. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046128
Moran, N. A., Dunbar, H. E., & Wilcox, J. L. (2005). Regulation of transcription in a reduced bacterial genome: Nutrient-provisioning genes of the obligate symbiont Buchnera aphidicola. *Journal of Bacteriology*, *187*(12), 4229–4237. https://doi.org/10.1128/JB.187.12.4229-4237.2005
Moran, N. A., McLaughlin, H. J., & Sorek, R. (2009). The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*. https://doi.org/10.1126/science.1167140
Moriwaki, N., Matsushita, K., Nishina, M., & Kono, Y. (2003). High concentrations of trehalose in aphid hemolymph. *Applied Entomology and Zoology*. https://doi.org/10.1303/aez.2003.241
Mueckler, M., & Thorens, B. (2013). The SLC2 (GLUT) family of membrane transporters. In *Molecular Aspects of Medicine*. https://doi.org/10.1016/j.mam.2012.07.001
NCBI. (2008). *Quick start - BLAST® Command Line Applications User Manual - NCBI Bookshelf*. https://www.ncbi.nlm.nih.gov/books/NBK279680/
Ng, J. C. K., & Perry, K. L. (2004). Transmission of plant viruses by aphid vectors. In *Molecular Plant Pathology* (Vol. 5, Issue 5, pp. 505–511). https://doi.org/10.1111/J.1364-3703.2004.00240.X
Price, D. R. G., Duncan, R. P., Shigenobu, S., & Wilson, A. C. C. (2011). Genome Expansion and Differential Expression of Amino Acid Transporters at the Aphid/Buchnera Symbiotic Interface. *Molecular Biology and Evolution*, *28*(11), 3113–3126.

https://doi.org/10.1093/molbev/msr140
Price, D. R. G., Tibbles, K., Shigenobu, S., Smertenko, A., Russell, C. W., Douglas, A. E., Fitches, E., Gatehouse, A. M. R., & Gatehouse, J. A. (2010). Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional characterization. *Insect Molecular Biology*, *19*(SUPPL. 2), 97–112. https://doi.org/10.1111/j.1365-2583.2009.00918.x
Price, D. R.G., Karley, A. J., Ashford, D. A., Isaacs, H. V., Pownall, M. E., Wilkinson, H. S., Gatehouse, J. A., & Douglas, A. E. (2007). Molecular characterisation of a candidate gut sucrase in the pea aphid, Acyrthosiphon pisum. *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.ibmb.2006.12.005
Price, Daniel R.G., Feng, H., Baker, J. D., Bavan, S., Luetje, C. W., & Wilson, A. C. C. (2014a). Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 320–325. https://doi.org/10.1073/pnas.1306068111
Price, Daniel R.G., Feng, H., Baker, J. D., Bavan, S., Luetje, C. W., & Wilson, A. C. C. (2014b). Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 320–325. https://doi.org/10.1073/pnas.1306068111
Price, Daniel R.G., & Gatehouse, J. A. (2014). Genome-wide annotation and functional identification of aphid GLUT-like sugar transporters. *BMC Genomics*, *15*(1), 647. https://doi.org/10.1186/1471-2164-15-647
Price, Daniel R.G., Wilkinson, H. S., & Gatehouse, J. A. (2007). Functional expression and characterisation of a gut facilitative glucose transporter, NlHT1, from the phloem-feeding insect Nilaparvata lugens (rice brown planthopper). *Insect Biochemistry and Molecular Biology*. https://doi.org/10.1016/j.ibmb.2007.07.001
Rabatel, A., Febvay, G., Gaget, K., Duport, G., Baa-Puyoulet, P., Sapountzis, P., Bendridi, N., Rey, M., Rahbé, Y., Charles, H., Calevro, F., & Colella, S. (2013). Tyrosine pathway regulation is host-mediated in the pea aphid symbiosis during late embryonic and early larval development. *BMC Genomics*, *14*(1), 235. https://doi.org/10.1186/1471-2164-14-235
Richards, S., Gibbs, R. A., Gerardo, N. M., Moran, N., Nakabachi, A., Stern, D., Tagu, D., Wilson, A. C. C., Muzny, D., Kovar, C., Cree, A., Chacko, J., Chandrabose, M. N., Dao, M. D., Dinh, H. H., Gabisi, R. A., Hines, S., Hume, J., Jhangian, S. N., … Hunter, W. (2010). Genome sequence of the pea aphid Acyrthosiphon pisum. *PLoS Biology*, *8*(2). https://doi.org/10.1371/journal.pbio.1000313
Russell, J. A., Weldon, S., Smith, A. H., Kim, K. L., Hu, Y., Łukasik, P., Doll, S., Anastopoulos, I., Novin, M., & Oliver, K. M. (2013). Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, *22*(7), 2045–2059. https://doi.org/10.1111/mec.12211
Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas, J., Heddi, A., & Calevro, F. (2016). Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Scientific Reports*. https://doi.org/10.1038/srep19967
Smith, T. E., & Moran, N. A. (2020). Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and Buchnera. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 2113–2121. https://doi.org/10.1073/pnas.1916748117

Sonnhammer, E. L., von Heijne, G., & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, *6*, 175–182.

Sorensen, J. T. (2009). Aphids. In *Encyclopedia of Insects* (pp. 27–31). Elsevier Inc. https://doi.org/10.1016/B978-0-12-374144-8.00008-4

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. https://doi.org/10.1038/nbt.1621

USDA. (2019). *USDA Crop Production 2019 Summary*.

Van Emden, H. F., & Harrington, R. (2007). Aphids as crop pests. In *Aphids as Crop Pests*. Cabi. https://doi.org/10.2135/cropsci2008.02.0001br

Voronova, N. V., Levykina, S., Warner, D., Shulinski, R., Bandarenka, Y., & Zhorov, D. (2020). Characteristic and variability of five complete aphid mitochondrial genomes: Aphis fabae mordvilkoi, Aphis craccivora, Myzus persicae, Therioaphis tenera and Appendiseta robiniae (Hemiptera; Sternorrhyncha; Aphididae). *International Journal of Biological Macromolecules*, *149*, 187–206. https://doi.org/10.1016/j.ijbiomac.2019.12.276

Wenger, J. A., Cassone, B. J., Legeai, F., Johnston, J. S., Bansal, R., Yates, A. D., Coates, B. S., Pavinato, V. A. C., & Michel, A. (2017). Whole genome sequence of the soybean aphid, Aphis glycines. *Insect Biochemistry and Molecular Biology*, *102917*. https://doi.org/10.1016/j.ibmb.2017.01.005

Wernegreen, J. J. (2012). Endosymbiosis. In *Current Biology* (Vol. 22, Issue 14, pp. R555–R561). Cell Press. https://doi.org/10.1016/j.cub.2012.06.010

Wille, B. D., & Hartman, G. L. (2009). Two species of symbiotic bacteria present in the soybean aphid (Hemiptera: Aphididae). *Environmental Entomology*, *38*(1), 110–115. https://doi.org/10.1603/022.038.0113

Wilson, A. C.C., Ashton, P. D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P. F., MacDonald, S. J., Schwartz, J. F., Thomas, G. H., & Douglas, A. E. (2010). Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. In *Insect Molecular Biology* (Vol. 19, Issue SUPPL. 2, pp. 249–258). https://doi.org/10.1111/j.1365-2583.2009.00942.x

Wilson, Alex C.C. (2020). Regulation of an insect symbiosis. In *Advances in Insect Physiology* (Vol. 58, pp. 207–232). Academic Press Inc. https://doi.org/10.1016/bs.aiip.2020.04.004

Wilson, Alex C.C., & Duncan, R. P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(33), 10255–10261. https://doi.org/10.1073/pnas.1423305112

Wulff, J. A., & White, J. A. (2015). The endosymbiont arsenophonus provides a general benefit to soybean aphid (Hemiptera: Aphididae) regardless of host plant resistance (Rag). *Environmental Entomology*, *44*(3), 574–581. https://doi.org/10.1093/ee/nvv031

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y., & Wei, L. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, *39*(SUPPL. 2). https://doi.org/10.1093/nar/gkr483

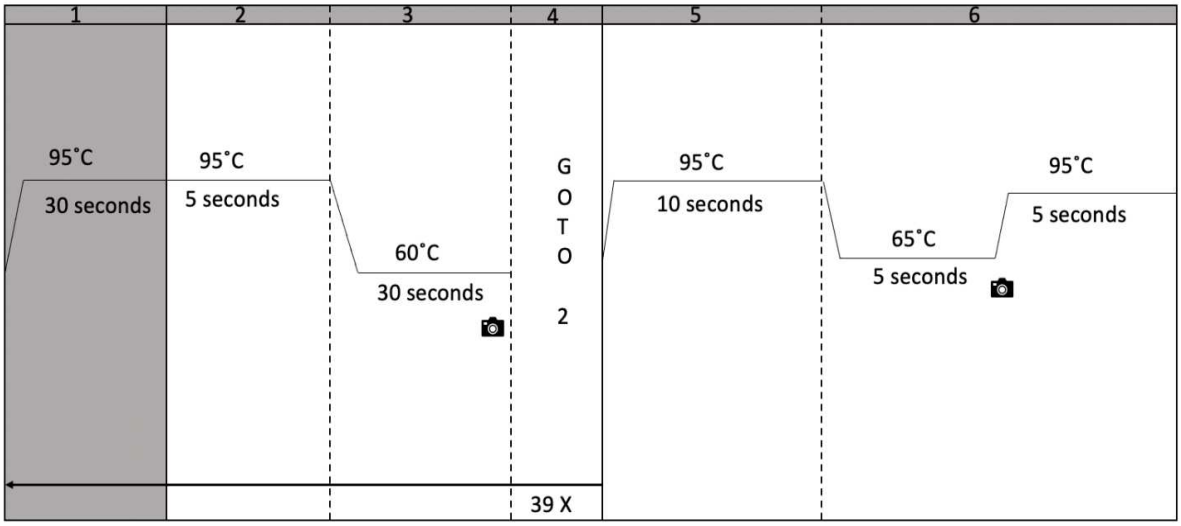
Yao, S. Y. M., Cass, C. E., & Young, J. D. (2000). The Xenopus oocyte expression system for the cDNA cloning and characterization of plasma membrane transport proteins. In *Membrane Transport* (pp. 47–78).

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Young, G. B., Jack, D. L., Smith, D. W., & Saier, M. H. (1999). The amino acid/auxin: Proton symport permease family. *Biochimica et Biophysica Acta - Biomembranes*, *1415*(2), 306–322. https://doi.org/10.1016/S0005-2736(98)00196-5

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2). https://doi.org/10.1186/gb-2010-11-2-r14

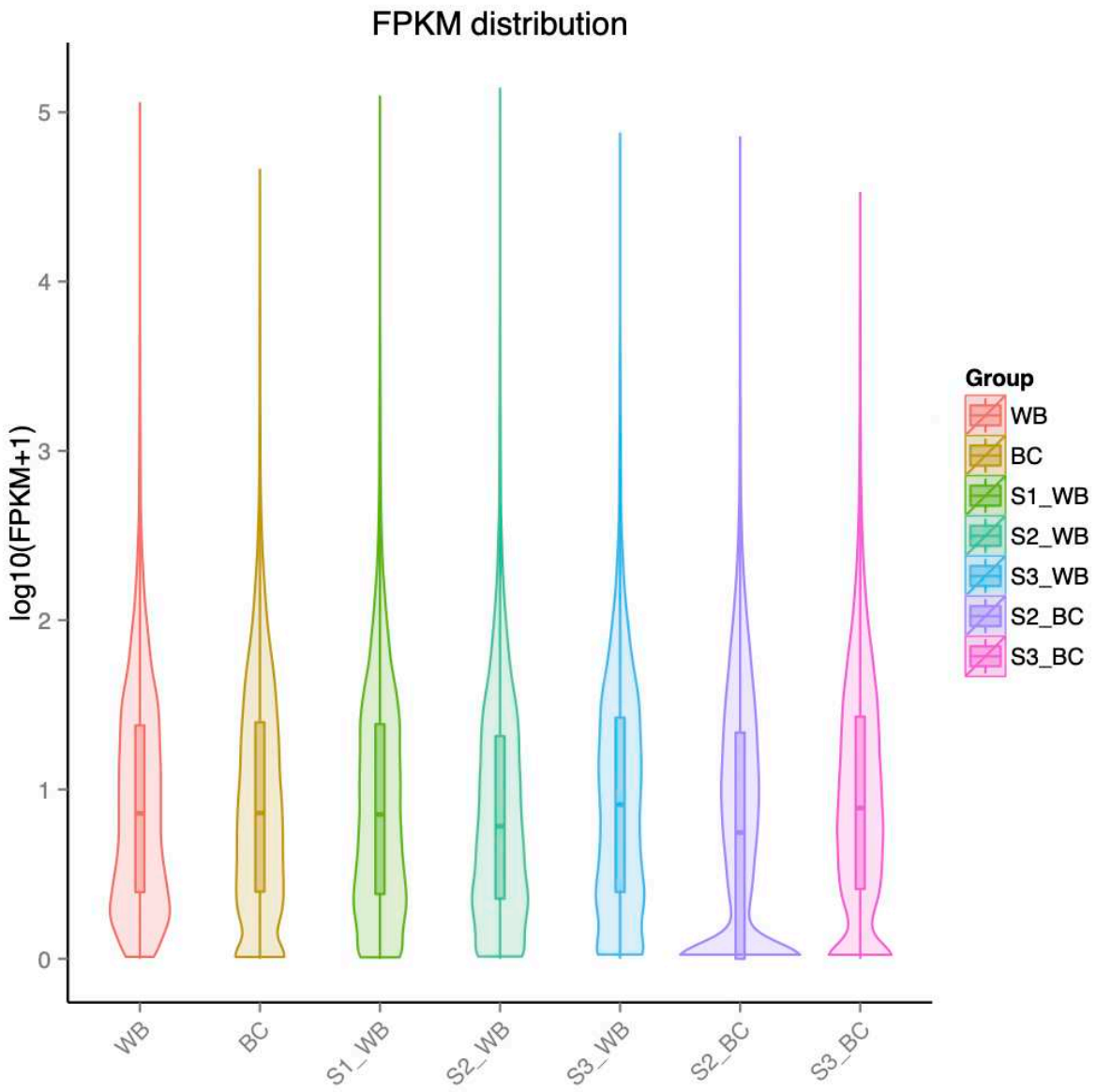
APPENDICES



Supplementary figure 1 : **qPCR cycling condition adapted from BioRad® software.**

1. Initial denaturation 2. Denaturation 3. Primer Annealing

1. 95 ˚C for 0:30 2. 95˚C for 0:05 3. 60˚C for 0:30 + Plate Read 4. GOTO 2, 39 more times 5. 95˚C for 0:10 6. Melt Curve 65˚C to 95˚C, increment 0.5 ˚C, for 0:05 + Plate Reads END



Supplementary figure 2 : Violin plot showing the normalization of read counts after applying Log_{10}(FPKM+1) normalization.

The y-axis shows the log-transformed normalized read-counts, and the x-axis shows the treatments (WB and BC) with their replicates (S1_WB to S3_WB and S2_BC and S3_BC).

Supplementary table 1. Primers for qPCR. Gene-specific primers for 25 putative amino acid transporters were designed using Primer-Blast.

Primer	Seq 5'->3'	Amplicon Size
AG6007894 F	CGGCGACCGGACCATATTAG	144
AG6007894 R	ATCACGTGTTCGGTCAGCTT	
AG6012777 F	TGCTCGGTTCAGAACTTCTAGT	96
AG6012777 R	ACAAGTGCACAGTAAGAGGACC	
AG6035555 F	TGGGAAGTACATGCCGAAATA	102
AG6035555 R	TATTTCATCGCTTTTCACGGC	
AG6034404 F	TGTGAGGTAACTTTTGATGCT	117
AG6034404 R	AGAGGTTAACTTGGAATTAATTGTT	
AG6005572 F	ACCTTATTGTGCAAGCAGCC	115
AG6005572 R	AGTGCATCGGCCTTAGCTATC	
AG6015419 F	AACGCCACCTAAATTGTATTAAAGA	213
AG6015419 R	TAGTACACAGGCGAAAACCTT	
AG6019085 F	TTTTCTTTTGGCAATTGCATGG	222
AG6019805 R	ACCCAACAATGCAGTAGTTACCA	
AG6026401 F	GGACGTTTCTGGTTCAACGG	136
AG6026401 R	GTCGTTACGGTTGTGATCGT	
AG6017614 F	GAGAATGACACTGCGCACAC	80
AG6017614 R	AGCCATCAAGACTTGCATAAGTT	
AG6005841 F	TGATATTCTACGCGGTGCCC	114
AG6005841 R	TGGCCAGCAGCCGGA	
AG6005568 F	GCTTTCAGAGCGCTTTACGA	125
AG6005568 R	AGACGCACCGAGTCTTTGAG	
AG6008892 F	ACCTCAAGAGATATTTCAAGAGCAA	96
AG6008892 R	GACGACTTGGTTGGAGAAACC	
AG6020364 F	GACCATCCTCATATGTCCCCAC	77
AG6020364 R	TGCCGTATCATTGGCAGCTA	

AG6008884 F	AGACGTGGATAAGATTTGGAGTG	225
AG600884 R	ACGCAATAGCCATCATCTACTAA	
AG6013324 F	ATTGTTTGGCACTGTTCGGC	144
AG6013324 R	CAGTTGGGAGGTATGTGTCGT	
AG6008879 F	GTTTGTACTTCTTATAGTTTGCACG	124
AG6008879 R	TGGTTGGAAGGTTGTTCCGT	
AG6023366 F	GACATCGTGGGCGGATATGT	202
AG6023366 R	ACGATGTGTACGTATTGTGGGT	
AG6027265 F	TCAGGTTTACGTTTATCGCCA	137
AG6027265 R	CCTGACGAAATTCTCGGGTT	
AG6031286 F	TGTTCCTTTTGCCGCCATTG	101
AG6031286 R	CGCATTTGGAAACGTGCGAT	
AG6029680 F	TGACTGTCGCTTATGTATCGAA	110
AG6029680 R	TGACGTTTTTGGTGCCATCT	
AG6011975 F	AAGTCACTGCCACAGTTTGT	137
AG6011975 R	GTCTGTCACCAGCTATTATATACCG	
AG6012001 F	TTCCTGGCGCTGATATTCCC	98
AG6012001 R	ATCACCAATCGCCAGCTGAA	
AG6029795 F	CTCTATGGCTGGCTTCCTCG	78
AG6029795 R	TCGTTGGGAAGATTGAGCGT	
AG6000219 F	CCTCTGCCTTGTGTAGCGTT	112
AG6000219 R	TGTTCAAAAATGATAACTGTGTGCT	
AG6008387 F	TGAAGATGAATGTATGGACTGGA	133
AG6008387 R	TAACCGTTGTTCCTGCAAGC	

Supplementary table 2: Differentially up-regulated genes in bacteriocytes

Gene_id	Readcount_BC	Readcount_WB	log2 Fold Change	pval	padj
AG6000092-RA	654.21	227.29	1.53	0.00	0.00
AG6000123-RA	1951.26	954.24	1.03	0.00	0.01
AG6000141-RA	205.04	64.54	1.67	0.00	0.03
AG6000157-RA	242.98	64.01	1.92	0.00	0.00
AG6000180-RA	449.94	126.00	1.84	0.00	0.00
AG6000335-RA	1072.87	472.93	1.18	0.00	0.01
AG6000364-RA	195.35	70.36	1.47	0.00	0.03
AG6000483-RA	429.82	168.68	1.35	0.00	0.02
AG6000496-RA	219.17	64.60	1.76	0.00	0.01
AG6000497-RA	1007.98	434.95	1.21	0.00	0.01
AG6000518-RA	1515.98	744.99	1.03	0.00	0.02
AG6000542-RA	857.09	429.35	1.00	0.00	0.05
AG6000651-RA	2773.71	1125.70	1.30	0.00	0.01
AG6000656-RA	67.21	11.65	2.53	0.00	0.04
AG6000755-RA	102.94	21.27	2.28	0.00	0.01
AG6000965-RA	2738.86	931.37	1.56	0.00	0.00
AG6001114-RA	479.28	184.54	1.38	0.00	0.01
AG6001320-RA	783.74	289.75	1.44	0.00	0.00
AG6001367-RA	430.26	170.22	1.34	0.00	0.01
AG6001383-RA	406.55	184.19	1.14	0.00	0.03
AG6001398-RA	2798.91	1495.07	0.90	0.00	0.02
AG6001438-RA	115.27	28.68	2.01	0.00	0.03
AG6001453-RA	1364.64	532.89	1.36	0.00	0.00
AG6001493-RA	6579.10	3677.71	0.84	0.00	0.04
AG6001513-RA	431.20	189.08	1.19	0.00	0.05
AG6001662-RA	1171.96	494.28	1.25	0.00	0.00
AG6001789-RA	677.71	339.25	1.00	0.00	0.05
AG6001790-RA	906.54	429.55	1.08	0.00	0.03
AG6001861-RA	1248.73	400.58	1.64	0.00	0.00
AG6001895-RA	909.20	465.76	0.97	0.00	0.05
AG6001898-RA	1938.01	578.35	1.74	0.00	0.00
AG6002039-RA	1759.07	899.33	0.97	0.00	0.02
AG6002042-RA	334.33	112.07	1.58	0.00	0.01
AG6002106-RA	50.09	6.56	2.93	0.00	0.03
AG6002129-RA	462.71	179.04	1.37	0.00	0.00
AG6002171-RA	542.44	249.54	1.12	0.00	0.03

AG6002220-RA	1699.09	728.87	1.22	0.00	0.00
AG6002231-RA	393.13	162.13	1.28	0.00	0.03
AG6002347-RA	1842.96	924.93	0.99	0.00	0.02
AG6002352-RA	292.22	103.20	1.50	0.00	0.01
AG6002363-RA	463.33	226.13	1.03	0.00	0.05
AG6002492-RA	328.08	137.88	1.25	0.00	0.03
AG6002669-RA	859.99	296.95	1.53	0.00	0.00
AG6002683-RA	2598.27	1146.21	1.18	0.00	0.00
AG6002705-RA	322.52	98.22	1.72	0.00	0.01
AG6002911-RA	392.72	177.93	1.14	0.00	0.05
AG6002916-RA	457.67	128.79	1.83	0.00	0.04
AG6003049-RA	52.11	3.62	3.85	0.00	0.01
AG6003065-RA	1654.95	794.40	1.06	0.00	0.02
AG6003178-RA	738.33	288.89	1.35	0.00	0.00
AG6003212-RA	486.39	224.68	1.11	0.00	0.05
AG6003472-RA	1054.05	398.44	1.40	0.00	0.00
AG6003559-RA	45768.15	23578.80	0.96	0.00	0.01
AG6003705-RA	879.66	427.44	1.04	0.00	0.02
AG6003752-RA	918.55	441.79	1.06	0.00	0.03
AG6003955-RA	1036.07	531.39	0.96	0.00	0.04
AG6004054-RA	658.46	260.65	1.34	0.00	0.00
AG6004104-RA	368.67	144.94	1.35	0.00	0.03
AG6004112-RA	1525.64	707.58	1.11	0.00	0.01
AG6004204-RA	974.40	487.13	1.00	0.00	0.03
AG6004234-RA	225.97	69.07	1.71	0.00	0.01
AG6004293-RA	634.35	163.93	1.95	0.00	0.00
AG6004486-RA	1525.33	262.74	2.54	0.00	0.00
AG6004499-RA	289.77	62.26	2.22	0.00	0.00
AG6004510-RA	1039.36	313.57	1.73	0.00	0.00
AG6004551-RA	157.80	40.58	1.96	0.00	0.02
AG6004571-RA	241.92	40.74	2.57	0.00	0.00
AG6004575-RA	387.72	113.99	1.77	0.00	0.00
AG6004601-RA	932.69	197.07	2.24	0.00	0.00
AG6004623-RA	457.90	213.31	1.10	0.00	0.05
AG6004705-RA	5751.49	3050.42	0.91	0.00	0.02
AG6004781-RA	386.21	107.31	1.85	0.00	0.05
AG6004850-RA	387.07	145.66	1.41	0.00	0.01
AG6004853-RA	726.31	358.38	1.02	0.00	0.03
AG6004974-RA	365.80	136.28	1.42	0.00	0.02

AG6004999-RA	6330.20	2706.05	1.23	0.00	0.00
AG6005118-RA	879.42	429.21	1.03	0.00	0.04
AG6005141-RA	131.33	32.37	2.02	0.00	0.03
AG6005154-RA	436.00	130.62	1.74	0.00	0.00
AG6005209-RA	677.59	311.27	1.12	0.00	0.03
AG6005219-RA	693.48	269.61	1.36	0.00	0.00
AG6005392-RA	887.50	433.25	1.03	0.00	0.03
AG6005571-RA	66.36	8.76	2.92	0.00	0.01
AG6005572-RA	1635.60	733.24	1.16	0.00	0.01
AG6005618-RA	3870.28	1488.73	1.38	0.00	0.00
AG6005650-RA	920.90	280.01	1.72	0.00	0.00
AG6005697-RA	1080.09	528.20	1.03	0.00	0.02
AG6005746-RA	2890.58	1532.38	0.92	0.00	0.02
AG6005788-RA	73.80	9.61	2.94	0.00	0.01
AG6005790-RA	1382.90	534.51	1.37	0.00	0.00
AG6005810-RA	642.42	247.82	1.37	0.00	0.00
AG6005813-RA	1308.40	506.31	1.37	0.00	0.01
AG6005816-RA	521.72	94.87	2.46	0.00	0.00
AG6005817-RA	341.90	139.42	1.29	0.00	0.02
AG6005825-RA	999.72	394.72	1.34	0.00	0.00
AG6005830-RA	1878.58	557.91	1.75	0.00	0.00
AG6005832-RA	326.67	119.09	1.46	0.00	0.01
AG6005991-RA	290.41	96.09	1.60	0.00	0.01
AG6005993-RA	65.18	7.82	3.06	0.00	0.02
AG6006014-RA	905.71	392.63	1.21	0.00	0.00
AG6006256-RA	629.10	275.16	1.19	0.00	0.02
AG6006415-RA	245.21	49.38	2.31	0.00	0.02
AG6006504-RA	331.58	131.74	1.33	0.00	0.02
AG6006519-RA	1411.95	536.34	1.40	0.00	0.01
AG6006544-RA	623.15	261.10	1.26	0.00	0.01
AG6006654-RA	179.40	62.74	1.52	0.00	0.04
AG6006924-RA	373.68	113.73	1.72	0.00	0.00
AG6006925-RA	157.38	9.26	4.09	0.00	0.00
AG6007157-RA	825.92	315.51	1.39	0.00	0.02
AG6007310-RA	5300.45	1856.12	1.51	0.00	0.02
AG6007333-RA	217.36	78.56	1.47	0.00	0.04
AG6007342-RA	776.59	232.91	1.74	0.00	0.00
AG6007362-RA	263.72	89.92	1.55	0.00	0.02
AG6007517-RA	14562.43	4706.56	1.63	0.00	0.00

AG6007536-RA	242.24	72.41	1.74	0.00	0.01
AG6007815-RA	158.65	37.24	2.09	0.00	0.01
AG6007945-RA	1052.76	466.14	1.18	0.00	0.01
AG6007954-RA	879.15	476.89	0.88	0.00	0.04
AG6007970-RA	2967.83	1230.55	1.27	0.00	0.00
AG6008083-RA	87.94	6.74	3.71	0.00	0.03
AG6008087-RA	450.23	178.40	1.34	0.00	0.02
AG6008160-RA	1268.76	690.02	0.88	0.00	0.05
AG6008180-RA	1097.43	541.88	1.02	0.00	0.03
AG6008253-RA	1276.33	654.23	0.96	0.00	0.02
AG6008305-RA	494.78	185.63	1.41	0.00	0.04
AG6008415-RA	929.91	289.64	1.68	0.00	0.01
AG6008436-RA	124.74	37.61	1.73	0.00	0.05
AG6008499-RA	244.92	62.04	1.98	0.00	0.00
AG6008511-RA	193.96	69.36	1.48	0.00	0.05
AG6008535-RA	587.09	268.46	1.13	0.00	0.04
AG6008583-RA	1237.57	553.60	1.16	0.00	0.01
AG6008820-RA	311.47	111.14	1.49	0.00	0.01
AG6008846-RA	617.94	280.38	1.14	0.00	0.02
AG6008905-RA	284.03	101.11	1.49	0.00	0.02
AG6008953-RA	308.91	100.38	1.62	0.00	0.02
AG6009009-RA	65.50	2.37	4.79	0.00	0.00
AG6009076-RA	627.94	179.10	1.81	0.00	0.00
AG6009077-RA	2705.32	1492.79	0.86	0.00	0.03
AG6009187-RA	366.02	135.63	1.43	0.00	0.01
AG6009193-RA	279.45	96.02	1.54	0.00	0.02
AG6009417-RA	333.27	124.57	1.42	0.00	0.01
AG6009492-RA	8846.11	2172.66	2.03	0.00	0.01
AG6009686-RA	142.93	46.61	1.62	0.00	0.04
AG6009723-RA	769.41	237.33	1.70	0.00	0.00
AG6009736-RA	558.63	253.89	1.14	0.00	0.01
AG6009779-RA	2131.37	1100.39	0.95	0.00	0.03
AG6009827-RA	1097.19	554.61	0.98	0.00	0.05
AG6009923-RA	3237.91	1038.68	1.64	0.00	0.00
AG6009951-RA	551.04	161.88	1.77	0.00	0.00
AG6010016-RA	606.90	292.42	1.05	0.00	0.03
AG6010033-RA	2918.71	1554.23	0.91	0.00	0.03
AG6010059-RA	1065.20	332.18	1.68	0.00	0.00
AG6010314-RA	251.18	83.83	1.58	0.00	0.01

AG6010315-RA	144.84	31.99	2.18	0.00	0.01
AG6010358-RA	164.61	47.72	1.79	0.00	0.03
AG6010513-RA	8053.30	3625.62	1.15	0.00	0.01
AG6010514-RA	448.96	176.40	1.35	0.00	0.01
AG6010528-RA	247.35	90.79	1.45	0.00	0.03
AG6010631-RA	8064.28	2802.11	1.53	0.00	0.00
AG6010678-RA	1512.20	587.62	1.36	0.00	0.00
AG6010698-RA	3263.19	1105.93	1.56	0.00	0.00
AG6010742-RA	1440.90	700.87	1.04	0.00	0.02
AG6010963-RA	1036.37	525.31	0.98	0.00	0.04
AG6011045-RA	163.67	58.21	1.49	0.00	0.03
AG6011051-RA	132.61	43.40	1.61	0.00	0.05
AG6011075-RA	204.49	51.68	1.98	0.00	0.01
AG6011078-RA	914.88	345.34	1.41	0.00	0.00
AG6011081-RA	139.51	31.82	2.13	0.00	0.01
AG6011117-RA	637.29	250.93	1.34	0.00	0.01
AG6011199-RA	1015.53	480.39	1.08	0.00	0.02
AG6011222-RA	588.80	238.26	1.31	0.00	0.01
AG6011248-RA	567.63	238.76	1.25	0.00	0.02
AG6011281-RA	444.72	172.94	1.36	0.00	0.01
AG6011306-RA	812.25	393.85	1.04	0.00	0.02
AG6011359-RA	80.82	8.15	3.31	0.00	0.00
AG6011460-RA	218.22	68.73	1.67	0.00	0.01
AG6011500-RA	565.09	245.83	1.20	0.00	0.02
AG6011524-RA	6199.37	2533.47	1.29	0.00	0.00
AG6011751-RA	1111.27	587.57	0.92	0.00	0.04
AG6011913-RA	500.97	212.54	1.24	0.00	0.02
AG6011968-RA	3777.25	1426.82	1.40	0.00	0.00
AG6012119-RA	3851.67	2075.80	0.89	0.00	0.03
AG6012120-RA	870.93	304.91	1.51	0.00	0.00
AG6012174-RA	603.72	295.46	1.03	0.00	0.02
AG6012221-RA	627.62	168.41	1.90	0.00	0.00
AG6012309-RA	191.94	43.65	2.14	0.00	0.02
AG6012318-RA	567.55	272.53	1.06	0.00	0.03
AG6012462-RA	1779.09	804.66	1.14	0.00	0.00
AG6012463-RA	341.12	84.76	2.01	0.00	0.03
AG6012504-RA	893.21	178.71	2.32	0.00	0.01
AG6012565-RA	226.94	67.04	1.76	0.00	0.02
AG6012631-RA	662.50	293.31	1.18	0.00	0.02

AG6012651-RA	892.32	368.90	1.27	0.00	0.00
AG6012676-RA	841.36	282.61	1.57	0.00	0.00
AG6012754-RA	95.39	13.27	2.85	0.00	0.00
AG6012798-RA	198.86	60.81	1.71	0.00	0.01
AG6012900-RA	906.97	464.42	0.97	0.00	0.05
AG6012960-RA	944.29	444.41	1.09	0.00	0.03
AG6013166-RA	385.91	132.85	1.54	0.00	0.00
AG6013195-RA	95.70	21.09	2.18	0.00	0.04
AG6013403-RA	1797.27	905.87	0.99	0.00	0.02
AG6013455-RA	2689.70	1065.64	1.34	0.00	0.03
AG6013592-RA	2274.42	1268.24	0.84	0.00	0.05
AG6013654-RA	111.65	24.94	2.16	0.00	0.03
AG6013717-RA	2385.28	1190.43	1.00	0.00	0.02
AG6013767-RA	1356.80	562.23	1.27	0.00	0.00
AG6013781-RA	1042.96	319.76	1.71	0.00	0.00
AG6013882-RA	1404.35	655.23	1.10	0.00	0.01
AG6014003-RA	1362.02	709.27	0.94	0.00	0.03
AG6014047-RA	219.17	72.76	1.59	0.00	0.02
AG6014201-RA	1363.26	654.13	1.06	0.00	0.02
AG6014267-RA	1180.60	609.46	0.95	0.00	0.03
AG6014305-RA	1168.87	392.31	1.58	0.00	0.00
AG6014469-RA	490.22	147.16	1.74	0.00	0.00
AG6014475-RA	1004.71	436.51	1.20	0.00	0.00
AG6014523-RA	618.26	283.69	1.12	0.00	0.03
AG6014636-RA	760.24	344.50	1.14	0.00	0.01
AG6014720-RA	249.38	103.96	1.26	0.00	0.04
AG6014746-RA	965.69	443.18	1.12	0.00	0.01
AG6014777-RA	881.99	406.51	1.12	0.00	0.01
AG6014779-RA	5009.81	2295.97	1.13	0.00	0.00
AG6014911-RA	39.45	3.39	3.54	0.00	0.03
AG6014922-RA	63.48	8.49	2.90	0.00	0.03
AG6014933-RA	591.24	154.12	1.94	0.00	0.00
AG6015009-RA	440.88	193.25	1.19	0.00	0.03
AG6015257-RA	771.50	275.81	1.48	0.00	0.00
AG6015277-RA	1060.53	474.65	1.16	0.00	0.01
AG6015310-RA	1619.79	739.44	1.13	0.00	0.00
AG6015317-RA	1622.99	683.44	1.25	0.00	0.00
AG6015518-RA	3891.05	2034.94	0.94	0.00	0.03
AG6015687-RA	780.31	369.58	1.08	0.00	0.04

AG6015697-RA	474.06	200.96	1.24	0.00	0.03
AG6015823-RA	868.81	327.53	1.41	0.00	0.00
AG6015958-RA	323.81	113.04	1.52	0.00	0.01
AG6015985-RA	767.67	374.31	1.04	0.00	0.03
AG6015999-RA	446.01	177.99	1.33	0.00	0.01
AG6016012-RA	664.60	208.32	1.67	0.00	0.01
AG6016659-RA	497.34	227.49	1.13	0.00	0.05
AG6016966-RA	312.32	68.26	2.19	0.00	0.00
AG6017004-RA	838.72	380.49	1.14	0.00	0.01
AG6017010-RA	288.52	76.41	1.92	0.00	0.01
AG6017064-RA	381.98	156.30	1.29	0.00	0.03
AG6017113-RA	314.88	105.92	1.57	0.00	0.01
AG6017407-RA	293.62	92.53	1.67	0.00	0.03
AG6017421-RA	894.53	425.80	1.07	0.00	0.03
AG6017534-RA	790.76	337.56	1.23	0.00	0.00
AG6017536-RA	1152.44	633.18	0.86	0.00	0.04
AG6017604-RA	2152.26	926.71	1.22	0.00	0.02
AG6017631-RA	1029.79	356.16	1.53	0.00	0.00
AG6017638-RA	203.74	67.31	1.60	0.00	0.03
AG6017690-RA	111.55	24.41	2.19	0.00	0.01
AG6017701-RA	89.75	19.21	2.22	0.00	0.04
AG6017711-RA	158.23	30.17	2.39	0.00	0.00
AG6017713-RA	63.38	12.27	2.37	0.00	0.05
AG6017930-RA	2174.97	1105.24	0.98	0.00	0.02
AG6017961-RA	387.40	153.71	1.33	0.00	0.02
AG6017998-RA	177.70	47.93	1.89	0.00	0.01
AG6018074-RA	523.82	145.81	1.85	0.00	0.00
AG6018225-RA	202.25	66.58	1.60	0.00	0.03
AG6018295-RA	309.02	71.03	2.12	0.00	0.00
AG6018476-RA	2879.75	1109.01	1.38	0.00	0.00
AG6018481-RA	2033.84	923.95	1.14	0.00	0.01
AG6018729-RA	412.49	142.32	1.54	0.00	0.00
AG6018839-RA	1311.04	390.13	1.75	0.00	0.00
AG6019022-RA	622.92	188.68	1.72	0.00	0.02
AG6019039-RA	212.57	71.38	1.57	0.00	0.02
AG6019127-RA	869.79	339.71	1.36	0.00	0.00
AG6019347-RA	1813.39	654.38	1.47	0.00	0.00
AG6019368-RA	2951.07	1101.02	1.42	0.00	0.00
AG6019420-RA	2308.40	768.94	1.59	0.00	0.00

AG6019469-RA	598.79	143.76	2.06	0.00	0.00
AG6019493-RA	345.17	100.96	1.77	0.00	0.00
AG6019497-RA	388.34	107.49	1.85	0.00	0.00
AG6019539-RA	188.54	36.43	2.37	0.00	0.00
AG6019615-RA	459.09	223.47	1.04	0.00	0.05
AG6019638-RA	64.87	7.54	3.10	0.00	0.01
AG6019686-RA	2414.26	870.53	1.47	0.00	0.00
AG6019729-RA	344.12	110.45	1.64	0.00	0.00
AG6019823-RA	281.81	120.46	1.23	0.00	0.05
AG6019915-RA	537.32	197.25	1.45	0.00	0.04
AG6020032-RA	940.07	440.41	1.09	0.00	0.01
AG6020047-RA	835.09	368.84	1.18	0.00	0.01
AG6020260-RA	997.89	440.09	1.18	0.00	0.01
AG6020270-RA	1340.95	712.98	0.91	0.00	0.05
AG6020274-RA	4902.38	1997.87	1.30	0.00	0.00
AG6020294-RA	475.86	116.50	2.03	0.00	0.00
AG6020301-RA	4599.83	1489.80	1.63	0.00	0.00
AG6020433-RA	807.77	350.39	1.21	0.00	0.01
AG6020446-RA	568.83	227.97	1.32	0.00	0.00
AG6020861-RA	579.14	284.28	1.03	0.00	0.04
AG6020894-RA	2901.99	1267.77	1.19	0.00	0.00
AG6021107-RA	752.60	393.44	0.94	0.00	0.03
AG6021119-RA	392.41	183.55	1.10	0.00	0.03
AG6021195-RA	2948.71	1515.40	0.96	0.00	0.02
AG6021559-RA	504.48	218.02	1.21	0.00	0.02
AG6021593-RA	1810.60	801.96	1.17	0.00	0.00
AG6021598-RA	2599.55	876.18	1.57	0.00	0.00
AG6021628-RA	501.41	209.91	1.26	0.00	0.01
AG6021698-RA	493.21	147.29	1.74	0.00	0.00
AG6021751-RA	474.47	118.98	2.00	0.00	0.00
AG6021889-RA	157.92	42.49	1.89	0.00	0.01
AG6021974-RA	612.61	113.79	2.43	0.00	0.00
AG6022174-RA	265.11	79.34	1.74	0.00	0.01
AG6022189-RA	638.68	210.19	1.60	0.00	0.00
AG6022207-RA	999.05	422.04	1.24	0.00	0.01
AG6022266-RA	766.60	262.65	1.55	0.00	0.00
AG6022281-RA	524.58	245.45	1.10	0.00	0.03
AG6022320-RA	3127.47	1529.76	1.03	0.00	0.01
AG6022369-RA	591.57	260.47	1.18	0.00	0.02

AG6022389-RA	2394.02	906.19	1.40	0.00	0.00
AG6022458-RA	2429.96	893.18	1.44	0.00	0.00
AG6022542-RA	601.99	248.30	1.28	0.00	0.01
AG6022549-RA	967.48	409.89	1.24	0.00	0.00
AG6022596-RA	89.01	19.57	2.19	0.00	0.03
AG6022606-RA	217.35	61.64	1.82	0.00	0.01
AG6022778-RA	537.87	167.41	1.68	0.00	0.00
AG6022820-RA	3064.84	1132.88	1.44	0.00	0.00
AG6022868-RA	552.25	202.80	1.45	0.00	0.03
AG6022910-RA	593.06	232.40	1.35	0.00	0.01
AG6022919-RA	286.06	109.88	1.38	0.00	0.02
AG6022926-RA	670.39	345.91	0.95	0.00	0.04
AG6022969-RA	455.78	125.32	1.86	0.00	0.00
AG6023042-RA	587.01	284.56	1.04	0.00	0.03
AG6023137-RA	1169.06	366.01	1.68	0.00	0.03
AG6023269-RA	133.78	32.78	2.03	0.00	0.01
AG6023289-RA	1557.56	387.82	2.01	0.00	0.00
AG6023339-RA	712.79	350.25	1.03	0.00	0.05
AG6023550-RA	1065.54	565.14	0.91	0.00	0.05
AG6023640-RA	3058.14	1622.00	0.91	0.00	0.04
AG6023649-RA	1126.12	574.13	0.97	0.00	0.05
AG6023650-RA	150.79	35.79	2.07	0.00	0.01
AG6023671-RA	4764.43	1037.29	2.20	0.00	0.00
AG6023680-RA	2221.23	464.97	2.26	0.00	0.00
AG6023681-RA	294.02	114.46	1.36	0.00	0.05
AG6023685-RA	170.57	39.22	2.12	0.00	0.00
AG6023810-RA	851.81	410.84	1.05	0.00	0.01
AG6023822-RA	1972.49	956.27	1.04	0.00	0.01
AG6023885-RA	218.42	62.15	1.81	0.00	0.01
AG6023886-RA	177.28	41.59	2.09	0.00	0.01
AG6023913-RA	102.08	10.65	3.26	0.00	0.05
AG6024018-RA	209.80	70.77	1.57	0.00	0.04
AG6024028-RA	1022.79	459.19	1.16	0.00	0.01
AG6024179-RA	2611.63	431.66	2.60	0.00	0.00
AG6024222-RA	503.43	194.78	1.37	0.00	0.00
AG6024238-RA	142.50	44.81	1.67	0.00	0.05
AG6024264-RA	521.93	252.07	1.05	0.00	0.04
AG6024306-RA	279.78	81.08	1.79	0.00	0.00
AG6024312-RA	624.34	191.80	1.70	0.00	0.00

AG6024444-RA	771.37	354.75	1.12	0.00	0.05
AG6024687-RA	227.58	30.39	2.90	0.00	0.00
AG6024812-RA	8632.68	3980.92	1.12	0.00	0.00
AG6024844-RA	182.27	65.12	1.49	0.00	0.05
AG6024956-RA	2267.10	1140.15	0.99	0.00	0.01
AG6025062-RA	592.66	263.67	1.17	0.00	0.02
AG6025073-RA	1679.38	639.82	1.39	0.00	0.00
AG6025260-RA	8300.89	4255.36	0.96	0.00	0.01
AG6025438-RA	2357.63	1157.18	1.03	0.00	0.01
AG6025694-RA	21.80	0.23	6.54	0.00	0.05
AG6025718-RA	509.47	182.46	1.48	0.00	0.00
AG6025732-RA	3699.33	1643.94	1.17	0.00	0.00
AG6025780-RA	454.37	130.62	1.80	0.00	0.02
AG6025782-RA	174.08	41.22	2.08	0.00	0.00
AG6025885-RA	835.07	327.93	1.35	0.00	0.03
AG6026103-RA	1148.93	408.88	1.49	0.00	0.00
AG6026141-RA	915.57	295.37	1.63	0.00	0.00
AG6026248-RA	125.59	27.54	2.19	0.00	0.01
AG6026653-RA	1348.57	656.51	1.04	0.00	0.05
AG6026741-RA	629.96	311.23	1.02	0.00	0.04
AG6026765-RA	569.35	238.56	1.26	0.00	0.01
AG6026828-RA	329.77	141.33	1.22	0.00	0.03
AG6026836-RA	860.82	340.09	1.34	0.00	0.00
AG6027013-RA	806.29	385.53	1.06	0.00	0.02
AG6027072-RA	109.32	15.17	2.85	0.00	0.00
AG6027089-RA	32.86	1.93	4.09	0.00	0.04
AG6027112-RA	467.69	173.59	1.43	0.00	0.00
AG6027136-RA	349.64	68.90	2.34	0.00	0.00
AG6027267-RA	485.44	225.01	1.11	0.00	0.04
AG6027279-RA	352.08	145.34	1.28	0.00	0.04
AG6027343-RA	407.08	124.21	1.71	0.00	0.00
AG6027367-RA	899.21	435.67	1.05	0.00	0.03
AG6027392-RA	5242.83	2488.46	1.08	0.00	0.00
AG6027491-RA	1040.61	395.55	1.40	0.00	0.04
AG6027519-RA	2171.73	1074.09	1.02	0.00	0.01
AG6027528-RA	642.41	291.65	1.14	0.00	0.01
AG6027619-RA	1915.63	1007.72	0.93	0.00	0.03
AG6027623-RA	1443.66	693.11	1.06	0.00	0.01
AG6027692-RA	307.00	118.93	1.37	0.00	0.04

AG6027754-RA	47.01	6.28	2.90	0.00	0.05
AG6027829-RA	327.62	129.66	1.34	0.00	0.04
AG6028004-RA	309.46	97.81	1.66	0.00	0.00
AG6028072-RA	3570.19	962.55	1.89	0.00	0.00
AG6028322-RA	202.68	58.62	1.79	0.00	0.02
AG6028337-RA	236.07	57.68	2.03	0.00	0.00
AG6028404-RA	1157.64	481.72	1.26	0.00	0.00
AG6028424-RA	91.46	20.27	2.17	0.00	0.03
AG6028432-RA	1133.24	396.63	1.51	0.00	0.00
AG6028449-RA	52.43	6.21	3.08	0.00	0.03
AG6028455-RA	863.24	207.63	2.06	0.00	0.00
AG6028646-RA	53.49	6.11	3.13	0.00	0.03
AG6028712-RA	1444.33	567.03	1.35	0.00	0.00
AG6028740-RA	1185.17	619.19	0.94	0.00	0.04
AG6028767-RA	448.96	185.14	1.28	0.00	0.02
AG6028873-RA	808.17	329.91	1.29	0.00	0.01
AG6028941-RA	376.23	127.28	1.56	0.00	0.01
AG6028942-RA	630.69	282.33	1.16	0.00	0.02
AG6029001-RA	292.32	94.09	1.64	0.00	0.04
AG6029098-RA	755.67	163.54	2.21	0.00	0.00
AG6029120-RA	1628.36	785.59	1.05	0.00	0.02
AG6029138-RA	950.50	214.83	2.15	0.00	0.00
AG6029139-RA	1672.00	541.18	1.63	0.00	0.00
AG6029203-RA	1107.55	389.71	1.51	0.00	0.00
AG6029219-RA	668.01	247.59	1.43	0.00	0.00
AG6029229-RA	277.45	116.94	1.25	0.00	0.03
AG6029267-RA	918.66	344.28	1.42	0.00	0.00
AG6029285-RA	1042.99	440.19	1.24	0.00	0.00
AG6029346-RA	114.75	13.36	3.10	0.00	0.01
AG6029466-RA	994.34	271.48	1.87	0.00	0.00
AG6029489-RA	625.40	279.90	1.16	0.00	0.02
AG6029722-RA	2469.52	1180.99	1.06	0.00	0.01
AG6029772-RA	2407.21	1289.02	0.90	0.00	0.04
AG6029911-RA	116.76	30.47	1.94	0.00	0.03
AG6029980-RA	430.45	154.71	1.48	0.00	0.04
AG6029992-RA	3930.40	538.22	2.87	0.00	0.00
AG6030077-RA	1111.23	537.82	1.05	0.00	0.03
AG6030176-RA	259.37	87.92	1.56	0.00	0.01
AG6030288-RA	81.56	6.11	3.74	0.00	0.00

AG6030348-RA	19.25	0.00	Inf	0.00	0.03
AG6030424-RA	490.97	178.37	1.46	0.00	0.01
AG6030505-RA	639.53	282.08	1.18	0.00	0.02
AG6030591-RA	209.70	64.43	1.70	0.00	0.02
AG6030632-RA	5005.02	2249.95	1.15	0.00	0.00
AG6030798-RA	313.27	111.25	1.49	0.00	0.01
AG6030800-RA	392.82	165.78	1.24	0.00	0.03
AG6030934-RA	868.83	289.73	1.58	0.00	0.00
AG6030935-RA	149.31	52.65	1.50	0.00	0.05
AG6030937-RA	83.06	19.95	2.06	0.00	0.04
AG6031389-RA	218.95	64.45	1.76	0.00	0.03
AG6031394-RA	870.30	422.22	1.04	0.00	0.02
AG6031531-RA	752.89	353.45	1.09	0.00	0.02
AG6031540-RA	263.08	88.28	1.58	0.00	0.02
AG6031542-RA	3330.02	960.67	1.79	0.00	0.00
AG6031545-RA	1858.03	565.92	1.72	0.00	0.00
AG6031550-RA	272.44	88.87	1.62	0.00	0.01
AG6031606-RA	54.88	2.12	4.69	0.00	0.00
AG6031694-RA	1688.26	925.15	0.87	0.00	0.05
AG6031882-RA	154.30	38.73	1.99	0.00	0.01
AG6031891-RA	155.04	28.72	2.43	0.00	0.00
AG6032129-RA	116.55	14.93	2.97	0.00	0.00
AG6032320-RA	103.58	23.14	2.16	0.00	0.03
AG6032400-RA	802.16	316.53	1.34	0.00	0.02
AG6032474-RA	242.04	54.44	2.15	0.00	0.00
AG6032530-RA	374.02	170.23	1.14	0.00	0.03
AG6032669-RA	47.21	5.10	3.21	0.00	0.05
AG6032675-RA	161.32	46.98	1.78	0.00	0.02
AG6032678-RA	2296.37	430.41	2.42	0.00	0.00
AG6032682-RA	299.57	108.30	1.47	0.00	0.01
AG6033362-RA	436.31	151.43	1.53	0.00	0.00
AG6033427-RA	52.00	6.67	2.96	0.00	0.04
AG6033704-RA	371.03	166.07	1.16	0.00	0.04
AG6033823-RA	2276.68	741.39	1.62	0.00	0.01
AG6034177-RA	1006.22	411.99	1.29	0.00	0.00
AG6034387-RA	365.17	160.24	1.19	0.00	0.04
AG6034504-RA	292.22	103.33	1.50	0.00	0.02
AG6034510-RA	246.92	87.04	1.50	0.00	0.03
AG6034534-RA	580.72	255.53	1.18	0.00	0.02

AG6034561-RA	541.05	209.22	1.37	0.00	0.01
AG6034601-RA	574.47	284.93	1.01	0.00	0.04
AG6034609-RA	459.72	144.43	1.67	0.00	0.00
AG6034611-RA	117.39	22.65	2.37	0.00	0.02
AG6035050-RA	708.23	231.80	1.61	0.00	0.00
AG6035293-RA	108.90	16.13	2.75	0.00	0.00
AG6035420-RA	1167.73	496.84	1.23	0.00	0.00
AG6035560-RA	184.92	37.68	2.29	0.00	0.00
AG6035634-RA	80.71	4.07	4.31	0.00	0.01
AG6035709-RA	525.53	233.47	1.17	0.00	0.03
AG6035827-RA	871.98	425.45	1.04	0.00	0.04
AG6035914-RA	222.36	68.17	1.71	0.00	0.01
AG6036215-RA	293.50	105.97	1.47	0.00	0.02
AG6036220-RA	693.21	289.46	1.26	0.00	0.03
AG6036334-RA	11278.76	6471.02	0.80	0.00	0.03
AG6036372-RA	242.89	60.14	2.01	0.00	0.00
AG6036806-RA	2081.05	1091.48	0.93	0.00	0.03
AG6037135-RA	54.77	8.71	2.65	0.00	0.05
AG6037152-RA	154.73	31.09	2.32	0.00	0.00
AG6037187-RA	259.78	95.57	1.44	0.00	0.04
AG6037217-RA	80.49	5.34	3.91	0.00	0.04
AG6038802-RA	469.26	115.98	2.02	0.00	0.03
AG6038902-RA	1324.16	235.34	2.49	0.00	0.00
AG6038916-RA	300.30	64.58	2.22	0.00	0.00
AG6038918-RA	121.98	15.97	2.93	0.00	0.00
AG6038931-RA	241.40	93.43	1.37	0.00	0.03
AG6039182-RA	891.91	284.24	1.65	0.00	0.00
AG6039240-RA	1516.67	608.35	1.32	0.00	0.00
AG6039263-RA	2823.07	729.95	1.95	0.00	0.00
AG6039281-RA	795.24	229.42	1.79	0.00	0.00
AG6039282-RA	372.41	40.20	3.21	0.00	0.00
AG6039302-RA	205.97	50.59	2.03	0.00	0.02
AG6039369-RA	448.12	195.86	1.19	0.00	0.04
AG6039432-RA	666.03	298.57	1.16	0.00	0.01
AG6039686-RA	193.64	53.61	1.85	0.00	0.01
AG6039695-RA	191.62	60.11	1.67	0.00	0.02
AG6039729-RA	144.19	33.96	2.09	0.00	0.01
AG6039730-RA	2685.06	330.55	3.02	0.00	0.01
AG6039731-RA	176.21	18.69	3.24	0.00	0.00

AG6039734-RA	28917.71	10989.99	1.40	0.00	0.02
AG6039735-RA	8260.14	1072.24	2.95	0.00	0.00
AG6039846-RA	114.95	26.73	2.10	0.00	0.02
AG6039928-RA	349.86	104.46	1.74	0.00	0.00
AG6040405-RA	1851.76	513.02	1.85	0.00	0.00
AG6040423-RA	1345.78	452.42	1.57	0.00	0.00
AG6040426-RA	199.28	55.66	1.84	0.00	0.01
AG6040590-RA	138.34	31.68	2.13	0.00	0.05
AG6040592-RA	6717.42	2748.16	1.29	0.00	0.01
AG6041833-RA	516.52	158.78	1.70	0.00	0.01
AG6041909-RA	289.66	92.83	1.64	0.00	0.03
AG6042133-RA	143.87	23.16	2.64	0.00	0.00

Gene_id: The gene identifier for differentially expressed genes
Readcount_BC: The read count values of bacteriocytes after normalization
Readcount_WB: The read count values of whole body after normalization
log2FoldChange: log2(BC/WB)
pvalue(pval): The statistical p-value
qvalue(padj): p-value after normalization. The smaller the q-value is, the more significant the difference is.

Supplementary table 3: GO enriched terms for down-regulated DEGs with $p < 0.05$

GO_accession	Description	Term type	Over-represented p-value	Corrected p-value	DEG_ item
GO:0042302	structural constituent of cuticle	molecular_function	5.60E-39	2.52E-35	54
GO:0008061	chitin binding	molecular_function	1.64E-14	3.69E-11	23
GO:0006030	chitin metabolic process	biological_process	6.44E-14	7.24E-11	22
GO:1901071	glucosamine-containing compound metabolic process	biological_process	6.44E-14	7.24E-11	22
GO:0006040	amino sugar metabolic process	biological_process	2.59E-13	2.33E-10	22
GO:0004252	serine-type endopeptidase activity	molecular_function	2.31E-11	1.73E-08	38
GO:0008236	serine-type peptidase activity	molecular_function	3.95E-10	2.22E-07	45
GO:0017171	serine hydrolase activity	molecular_function	3.95E-10	2.22E-07	45
GO:0006022	aminoglycan metabolic process	biological_process	1.99E-09	9.94E-07	22
GO:0005576	extracellular region	cellular_component	3.01E-09	1.35E-06	81
GO:0070011	peptidase activity, acting on L-amino acid peptides	molecular_function	5.81E-09	2.37E-06	98
GO:0008233	peptidase activity	molecular_function	2.06E-08	7.71E-06	100
GO:0005198	structural molecule activity	molecular_function	3.40E-08	1.18E-05	86
GO:0004175	endopeptidase activity	molecular_function	4.80E-08	1.54E-05	73
GO:0015075	ion transmembrane transporter activity	molecular_function	5.58E-08	1.67E-05	95
GO:0006811	ion transport	biological_process	8.26E-08	2.32E-05	98
GO:0022836	gated channel activity	molecular_function	1.83E-07	4.83E-05	34
GO:0022891	substrate-specific transmembrane transporter activity	molecular_function	1.96E-07	4.91E-05	96
GO:0022892	substrate-specific transporter activity	molecular_function	2.32E-07	5.34E-05	101
GO:0016020	membrane	cellular_component	2.38E-07	5.34E-05	337
GO:1901135	carbohydrate derivative metabolic process	biological_process	5.91E-07	1.15E-04	61
GO:0005216	ion channel activity	molecular_function	6.16E-07	1.15E-04	48

GO:0022838	substrate-specific channel activity	molecular_function	6.16E-07	1.15E-04	48
GO:0015267	channel activity	molecular_function	6.40E-07	1.15E-04	48
GO:0022803	passive transmembrane transporter activity	molecular_function	6.40E-07	1.15E-04	48
GO:0004871	signal transducer activity	molecular_function	2.68E-06	4.63E-04	61
GO:0022857	transmembrane transporter activity	molecular_function	4.96E-06	8.26E-04	114
GO:0060089	molecular transducer activity	molecular_function	5.54E-06	8.90E-04	63
GO:0005215	transporter activity	molecular_function	7.49E-06	1.16E-03	130
GO:0004930	G-protein coupled receptor activity	molecular_function	1.02E-05	1.53E-03	31
GO:0015276	ligand-gated ion channel activity	molecular_function	2.09E-05	2.93E-03	22
GO:0022834	ligand-gated channel activity	molecular_function	2.09E-05	2.93E-03	22
GO:0015672	monovalent inorganic cation transport	biological_process	2.83E-05	3.86E-03	43
GO:0006813	potassium ion transport	biological_process	3.05E-05	3.93E-03	20
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	molecular_function	3.06E-05	3.93E-03	31
GO:0015077	monovalent inorganic cation transmembrane transporter activity	molecular_function	5.38E-05	6.71E-03	42
GO:0007186	G-protein coupled receptor signaling pathway	biological_process	6.70E-05	8.14E-03	37
GO:0044765	single-organism transport	biological_process	8.23E-05	9.73E-03	159
GO:0008237	metallopeptidase activity	molecular_function	1.10E-04	1.26E-02	28
GO:0022890	inorganic cation transmembrane transporter activity	molecular_function	1.14E-04	1.29E-02	59
GO:1902578	single-organism localization	biological_process	1.23E-04	1.35E-02	160
GO:0008324	cation transmembrane transporter activity	molecular_function	1.43E-04	1.53E-02	63

GO:0038023	signaling receptor activity	molecular_function	1.92E-04	2.01E-02	44
GO:0005230	extracellular ligand-gated ion channel activity	molecular_function	2.14E-04	2.08E-02	15
GO:0034220	ion transmembrane transport	biological_process	2.22E-04	2.08E-02	31
GO:0004888	transmembrane signaling receptor activity	molecular_function	2.30E-04	2.08E-02	41
GO:0033993	response to lipid	biological_process	2.43E-04	2.08E-02	8
GO:0043401	steroid hormone mediated signaling pathway	biological_process	2.43E-04	2.08E-02	8
GO:0048545	response to steroid hormone	biological_process	2.43E-04	2.08E-02	8
GO:0071383	cellular response to steroid hormone stimulus	biological_process	2.43E-04	2.08E-02	8
GO:0071396	cellular response to lipid	biological_process	2.43E-04	2.08E-02	8
GO:0071407	cellular response to organic cyclic compound	biological_process	2.43E-04	2.08E-02	8
GO:0006836	neurotransmitter transport	biological_process	2.45E-04	2.08E-02	7
GO:0009755	hormone-mediated signaling pathway	biological_process	2.89E-04	2.36E-02	8
GO:0032870	cellular response to hormone stimulus	biological_process	2.89E-04	2.36E-02	8
GO:0005261	cation channel activity	molecular_function	2.94E-04	2.36E-02	25
GO:0006812	cation transport	biological_process	3.29E-04	2.57E-02	61
GO:0014070	response to organic cyclic compound	biological_process	3.33E-04	2.57E-02	8
GO:0071495	cellular response to endogenous stimulus	biological_process	3.41E-04	2.57E-02	8
GO:0004872	receptor activity	molecular_function	3.43E-04	2.57E-02	46
GO:0006629	lipid metabolic process	biological_process	3.93E-04	2.83E-02	56
GO:0005342	organic acid transmembrane transporter activity	molecular_function	3.96E-04	2.83E-02	10
GO:0046943	carboxylic acid transmembrane transporter activity	molecular_function	3.96E-04	2.83E-02	10
GO:0042886	amide transport	biological_process	4.50E-04	3.16E-02	3

GO:0006508	proteolysis	biological_process	4.79E-04	3.28E-02	78
GO:0009725	response to hormone	biological_process	4.81E-04	3.28E-02	8
GO:0022843	voltage-gated cation channel activity	molecular_function	5.69E-04	3.81E-02	9
GO:0008514	organic anion transmembrane transporter activity	molecular_function	5.81E-04	3.84E-02	10
GO:0015370	solute:sodium symporter activity	molecular_function	6.57E-04	4.28E-02	8
GO:0009719	response to endogenous stimulus	biological_process	7.27E-04	4.67E-02	8
GO:0015171	amino acid transmembrane transporter activity	molecular_function	7.38E-04	4.67E-02	8
GO:0005575	cellular_component	cellular_component	7.78E-04	4.85E-02	530

GO_accession: Gene Ontology entry.
Description: Detail description of Gene Ontology.
Term_type: GO types, including cellular component, biological process and molecular function.
Over_represented_pValue: P-value in hypergeometric test.
Corrected_pValue: Corrected P-value, GO with Corrected P-value < 0.05 are significantly enriched in DEGs.
DEG_item: Number of DEGs with GO annotation.

Supplementary table 4: Gene Ontology (GO-terms) for differentially expressed genes (DEGs) involved in all transport activity within the aphid bacteriocyte cell.

GO_accession	Description	Term type	Overrepresented p-value	Corrected_p-value	DEG_item
GO:0006811	ion transport	biological_process	0.000	0.002	111
GO:0015075	ion transmembrane transporter activity	molecular_function	0.000	0.007	104
GO:0006813	potassium ion transport	biological_process	0.000	0.007	23
GO:0022891	substrate-specific transmembrane transporter activity	molecular_function	0.000	0.008	107
GO:0022892	substrate-specific transporter activity	molecular_function	0.000	0.008	113
GO:0022803	passive transmembrane transporter activity	molecular_function	0.000	0.008	51
GO:0015672	monovalent inorganic cation transport	biological_process	0.000	0.037	47
GO:0015077	monovalent inorganic cation transmembrane transporter activity	molecular_function	0.000	0.040	47
GO:0005342	organic acid transmembrane transporter activity	molecular_function	0.001	0.062	11
GO:0046943	carboxylic acid transmembrane transporter activity	molecular_function	0.001	0.062	11
GO:0022857	transmembrane transporter activity	molecular_function	0.001	0.080	126
GO:0006836	neurotransmitter transport	biological_process	0.001	0.086	7
GO:0042886	amide transport	biological_process	0.001	0.086	3

GO:0008514	organic anion transmembrane transporter activity	molecular_function	0.001	0.086	11
GO:0005215	transporter activity	molecular_function	0.001	0.093	145
GO:0022890	inorganic cation transmembrane transporter activity	molecular_function	0.002	0.129	66
GO:0071705	nitrogen compound transport	biological_process	0.002	0.157	22
GO:0051050	positive regulation of transport	biological_process	0.003	0.189	4
GO:0008324	cation transmembrane transporter activity	molecular_function	0.003	0.191	70
GO:0015849	organic acid transport	biological_process	0.003	0.199	16
GO:0046942	carboxylic acid transport	biological_process	0.003	0.199	16
GO:0044765	single-organism transport	biological_process	0.003	0.199	183
GO:0015711	organic anion transport	biological_process	0.004	0.202	16
GO:0015171	amino acid transmembrane transporter activity	molecular_function	0.004	0.202	8
GO:0006812	cation transport	biological_process	0.004	0.202	68
GO:0034220	ion transmembrane transport	biological_process	0.005	0.202	32
GO:0006865	amino acid transport	biological_process	0.005	0.202	8
GO:0015291	secondary active transmembrane transporter activity	molecular_function	0.006	0.202	12
GO:0022898	regulation of transmembrane transporter activity	biological_process	0.006	0.202	3

GO:0032409	regulation of transporter activity	biological_process	0.006	0.202	3
GO:0032411	positive regulation of transporter activity	biological_process	0.006	0.202	3
GO:0032412	regulation of ion transmembrane transporter activity	biological_process	0.006	0.202	3
GO:0032414	positive regulation of ion transmembrane transporter activity	biological_process	0.006	0.202	3
GO:0034762	regulation of transmembrane transport	biological_process	0.006	0.202	3
GO:0034764	positive regulation of transmembrane transport	biological_process	0.006	0.202	3
GO:0034765	regulation of ion transmembrane transport	biological_process	0.006	0.202	3
GO:0034767	positive regulation of ion transmembrane transport	biological_process	0.006	0.202	3
GO:0043266	regulation of potassium ion transport	biological_process	0.006	0.202	3
GO:0043268	positive regulation of potassium ion transport	biological_process	0.006	0.202	3
GO:0043270	positive regulation of ion transport	biological_process	0.006	0.202	3
GO:1901016	regulation of potassium ion transmembrane transporter activity	biological_process	0.006	0.202	3
GO:1901018	positive regulation of potassium ion transmembrane	biological_process	0.006	0.202	3

	transporter activity				
GO:1901379	regulation of potassium ion transmembrane transport	biological_process	0.006	0.202	3
GO:1901381	positive regulation of potassium ion transmembrane transport	biological_process	0.006	0.202	3
GO:1904062	regulation of cation transmembrane transport	biological_process	0.006	0.202	3
GO:1904064	positive regulation of cation transmembrane transport	biological_process	0.006	0.202	3
GO:0006820	anion transport	biological_process	0.006	0.205	24
GO:0003333	amino acid transmembrane transport	biological_process	0.006	0.212	7
GO:0098656	anion transmembrane transport	biological_process	0.006	0.212	7
GO:1903825	organic acid transmembrane transport	biological_process	0.006	0.212	7
GO:0015079	potassium ion transmembrane transporter activity	molecular_function	0.006	0.213	13
GO:0015833	peptide transport	biological_process	0.008	0.261	2
GO:0030001	metal ion transport	biological_process	0.010	0.307	48
GO:0046873	metal ion transmembrane transporter activity	molecular_function	0.010	0.310	45
GO:0005326	neurotransmitter transporter activity	molecular_function	0.013	0.370	5
GO:0008504	monoamine transmembrane transporter activity	molecular_function	0.015	0.418	2

GO:0015222	serotonin transmembrane transporter activity	molecular_function	0.015	0.418	2
GO:0015085	calcium ion transmembrane transporter activity	molecular_function	0.018	0.468	10
GO:1901618	organic hydroxy compound transmembrane transporter activity	molecular_function	0.024	0.578	2
GO:0015081	sodium ion transmembrane transporter activity	molecular_function	0.025	0.580	14
GO:0008519	ammonium transmembrane transporter activity	molecular_function	0.027	0.612	3
GO:0005388	calcium-transporting ATPase activity	molecular_function	0.030	0.654	2
GO:0008509	anion transmembrane transporter activity	molecular_function	0.032	0.665	15
GO:0022804	active transmembrane transporter activity	molecular_function	0.033	0.687	22
GO:0098660	inorganic ion transmembrane transport	biological_process	0.035	0.707	25
GO:0098662	inorganic cation transmembrane transport	biological_process	0.035	0.707	25
GO:0098655	cation transmembrane transport	biological_process	0.036	0.722	25
GO:0005310	dicarboxylic acid transmembrane transporter activity	molecular_function	0.040	0.755	4
GO:0045261	proton-transporting ATP synthase complex,	cellular_component	0.044	0.810	3

	catalytic core F(1)				
GO:1902495	transmembrane transporter complex	cellular_component	0.050	0.893	5
GO:1990351	transporter complex	cellular_component	0.050	0.893	5
GO:0006835	dicarboxylic acid transport	biological_process	0.051	0.900	7
GO:0015078	hydrogen ion transmembrane transporter activity	molecular_function	0.052	0.911	21
GO:0090533	cation-transporting ATPase complex	cellular_component	0.056	0.911	2
GO:0016469	proton-transporting two-sector ATPase complex	cellular_component	0.056	0.911	17
GO:0051049	regulation of transport	biological_process	0.058	0.922	5
GO:0019829	cation-transporting ATPase activity	molecular_function	0.075	1.000	4
GO:1902600	hydrogen ion transmembrane transport	biological_process	0.079	1.000	17
GO:0098533	ATPase dependent transmembrane transport complex	cellular_component	0.081	1.000	2
GO:0055085	transmembrane transport	biological_process	0.081	1.000	86
GO:0042773	ATP synthesis coupled electron transport	biological_process	0.082	1.000	6
GO:0045259	proton-transporting ATP synthase complex	cellular_component	0.088	1.000	9
GO:0006810	transport	biological_process	0.088	1.000	219
GO:0071804	cellular potassium ion transport	biological_process	0.100	1.000	3

GO:0071805	potassium ion transmembrane transport	biological_process	0.100	1.000	3
GO:0006818	hydrogen transport	biological_process	0.106	1.000	17
GO:0015992	proton transport	biological_process	0.106	1.000	17
GO:0051180	vitamin transport	biological_process	0.118	1.000	2
GO:0010959	regulation of metal ion transport	biological_process	0.121	1.000	3
GO:0043269	regulation of ion transport	biological_process	0.121	1.000	3
GO:0015225	biotin transporter activity	molecular_function	0.126	1.000	1
GO:0015878	biotin transport	biological_process	0.126	1.000	1
GO:0015985	energy coupled proton transport, down electrochemical gradient	biological_process	0.140	1.000	9
GO:0015986	ATP synthesis coupled proton transport	biological_process	0.140	1.000	9
GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	cellular_component	0.140	1.000	6
GO:0033177	proton-transporting two-sector ATPase complex, proton-transporting domain	cellular_component	0.143	1.000	11
GO:0015858	nucleoside transport	biological_process	0.155	1.000	1
GO:0034257	nicotinamide riboside transmembrane transporter activity	molecular_function	0.155	1.000	1
GO:0034258	nicotinamide riboside transport	biological_process	0.155	1.000	1

GO:0015716	organic phosphonate transport	biological_process	0.157	1.000	1
GO:0042916	alkylphosphonate transport	biological_process	0.157	1.000	1
GO:0007034	vacuolar transport	biological_process	0.179	1.000	3
GO:0015889	cobalamin transport	biological_process	0.183	1.000	1
GO:0046933	proton-transporting ATP synthase activity, rotational mechanism	molecular_function	0.189	1.000	2
GO:0033227	dsRNA transport	biological_process	0.200	1.000	3
GO:0051032	nucleic acid transmembrane transporter activity	molecular_function	0.200	1.000	3
GO:0051033	RNA transmembrane transporter activity	molecular_function	0.200	1.000	3
GO:0022904	respiratory electron transport chain	biological_process	0.207	1.000	6
GO:0051588	regulation of neurotransmitter transport	biological_process	0.209	1.000	1
GO:0051589	negative regulation of neurotransmitter transport	biological_process	0.209	1.000	1
GO:0006816	calcium ion transport	biological_process	0.211	1.000	5
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	biological_process	0.220	1.000	8
GO:0015991	ATP hydrolysis coupled proton transport	biological_process	0.220	1.000	8
GO:0090662	ATP hydrolysis coupled transmembrane transport	biological_process	0.220	1.000	8

GO:0042775	mitochondrial ATP synthesis coupled electron transport	biological_process	0.221	1.000	3
GO:0070588	calcium ion transmembrane transport	biological_process	0.221	1.000	2
GO:0046961	proton-transporting ATPase activity, rotational mechanism	molecular_function	0.221	1.000	2
GO:0006892	post-Golgi vesicle-mediated transport	biological_process	0.229	1.000	1
GO:0006893	Golgi to plasma membrane transport	biological_process	0.229	1.000	1
GO:0032386	regulation of intracellular transport	biological_process	0.229	1.000	1
GO:0032388	positive regulation of intracellular transport	biological_process	0.229	1.000	1
GO:0033157	regulation of intracellular protein transport	biological_process	0.229	1.000	1
GO:0042996	regulation of Golgi to plasma membrane protein transport	biological_process	0.229	1.000	1
GO:0042998	positive regulation of Golgi to plasma membrane protein transport	biological_process	0.229	1.000	1
GO:0043001	Golgi to plasma membrane protein transport	biological_process	0.229	1.000	1
GO:0051222	positive regulation of protein transport	biological_process	0.229	1.000	1
GO:0090316	positive regulation of intracellular protein transport	biological_process	0.229	1.000	1
GO:1903649	regulation of cytoplasmic transport	biological_process	0.229	1.000	1

GO:1903651	positive regulation of cytoplasmic transport	biological_process	0.229	1.000	1
GO:0072509	divalent inorganic cation transmembrane transporter activity	molecular_function	0.230	1.000	14
GO:0015932	nucleobase-containing compound transmembrane transporter activity	molecular_function	0.246	1.000	4
GO:0051181	cofactor transport	biological_process	0.255	1.000	3
GO:0022884	macromolecule transmembrane transporter activity	molecular_function	0.264	1.000	3
GO:0033176	proton-transporting V-type ATPase complex	cellular_component	0.266	1.000	6
GO:0060627	regulation of vesicle-mediated transport	biological_process	0.284	1.000	1
GO:0006120	mitochondrial electron transport, NADH to ubiquinone	biological_process	0.296	1.000	2
GO:0045263	proton-transporting ATP synthase complex, coupling factor F(o)	cellular_component	0.305	1.000	6
GO:0033179	proton-transporting V-type ATPase, V0 domain	cellular_component	0.307	1.000	4
GO:0006869	lipid transport	biological_process	0.313	1.000	7
GO:0022900	electron transport chain	biological_process	0.327	1.000	7
GO:0045156	electron transporter, transferring electrons within	molecular_function	0.338	1.000	1

	the cyclic electron transport pathway of photosynthesis activity				
GO:0015399	primary active transmembrane transporter activity	molecular_function	0.341	1.000	10
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	molecular_function	0.341	1.000	10
GO:0000275	mitochondrial proton-transporting ATP synthase complex, catalytic core F(1)	cellular_component	0.345	1.000	1
GO:0006814	sodium ion transport	biological_process	0.350	1.000	8
GO:0051183	vitamin transporter activity	molecular_function	0.352	1.000	1
GO:0015743	malate transport	biological_process	0.356	1.000	3
GO:0005313	L-glutamate transmembrane transporter activity	molecular_function	0.362	1.000	1
GO:0015172	acidic amino acid transmembrane transporter activity	molecular_function	0.362	1.000	1
GO:0015179	L-amino acid transmembrane transporter activity	molecular_function	0.362	1.000	1
GO:0015800	acidic amino acid transport	biological_process	0.362	1.000	1
GO:0015807	L-amino acid transport	biological_process	0.362	1.000	1
GO:0015813	L-glutamate transport	biological_process	0.362	1.000	1

GO:0071702	organic substance transport	biological_process	0.367	1.000	69
GO:0051223	regulation of protein transport	biological_process	0.368	1.000	1
GO:0015696	ammonium transport	biological_process	0.373	1.000	1
GO:0005319	lipid transporter activity	molecular_function	0.374	1.000	3
GO:0051051	negative regulation of transport	biological_process	0.383	1.000	1
GO:0015706	nitrate transport	biological_process	0.386	1.000	1
GO:0072348	sulfur compound transport	biological_process	0.388	1.000	3
GO:0015721	bile acid and bile salt transport	biological_process	0.398	1.000	1
GO:0043574	peroxisomal transport	biological_process	0.409	1.000	1
GO:0015740	C4-dicarboxylate transport	biological_process	0.411	1.000	3
GO:0015850	organic hydroxy compound transport	biological_process	0.414	1.000	1
GO:0006122	mitochondrial electron transport, ubiquinol to cytochrome c	biological_process	0.436	1.000	1
GO:0015099	nickel cation transmembrane transporter activity	molecular_function	0.441	1.000	1
GO:0015675	nickel cation transport	biological_process	0.441	1.000	1
GO:0035444	nickel cation transmembrane transport	biological_process	0.441	1.000	1
GO:1901264	carbohydrate derivative transport	biological_process	0.459	1.000	1
GO:0033180	proton-transporting V-type ATPase, V1 domain	cellular_component	0.464	1.000	2
GO:0015886	heme transport	biological_process	0.494	1.000	2

GO:1901678	iron coordination entity transport	biological_process	0.494	1.000	2
GO:0035725	sodium ion transmembrane transport	biological_process	0.511	1.000	2
GO:0008272	sulfate transport	biological_process	0.551	1.000	2
GO:0015116	sulfate transmembrane transporter activity	molecular_function	0.551	1.000	2
GO:0015718	monocarboxylic acid transport	biological_process	0.553	1.000	2
GO:0005315	inorganic phosphate transmembrane transporter activity	molecular_function	0.554	1.000	1
GO:0015931	nucleobase-containing compound transport	biological_process	0.560	1.000	5
GO:0070838	divalent metal ion transport	biological_process	0.577	1.000	9
GO:0012510	trans-Golgi network transport vesicle membrane	cellular_component	0.605	1.000	1
GO:0030140	trans-Golgi network transport vesicle	cellular_component	0.605	1.000	1
GO:1901677	phosphate transmembrane transporter activity	molecular_function	0.605	1.000	1
GO:0051184	cofactor transporter activity	molecular_function	0.610	1.000	4
GO:0006817	phosphate ion transport	biological_process	0.619	1.000	1
GO:0070070	proton-transporting V-type ATPase complex assembly	biological_process	0.620	1.000	1
GO:0070072	vacuolar proton-transporting V-type ATPase complex assembly	biological_process	0.620	1.000	1

GO:0005753	mitochondrial proton-transporting ATP synthase complex	cellular_component	0.621	1.000	4
GO:1901682	sulfur compound transmembrane transporter activity	molecular_function	0.633	1.000	2
GO:0072511	divalent inorganic cation transport	biological_process	0.650	1.000	9
GO:0015232	heme transporter activity	molecular_function	0.656	1.000	1
GO:0005337	nucleoside transmembrane transporter activity	molecular_function	0.658	1.000	1
GO:0016197	endosomal transport	biological_process	0.673	1.000	1
GO:0015698	inorganic anion transport	biological_process	0.691	1.000	4
GO:0070071	proton-transporting two-sector ATPase complex assembly	biological_process	0.709	1.000	1
GO:0050657	nucleic acid transport	biological_process	0.718	1.000	4
GO:0050658	RNA transport	biological_process	0.718	1.000	4
GO:0008565	protein transporter activity	molecular_function	0.731	1.000	3
GO:0000276	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	cellular_component	0.749	1.000	3
GO:0000221	vacuolar proton-transporting V-type ATPase, V1 domain	cellular_component	0.771	1.000	1
GO:0016471	vacuolar proton-transporting V-type ATPase complex	cellular_component	0.771	1.000	1

GO:0065002	intracellular protein transmembrane transport	biological_process	0.778	1.000	1
GO:1901505	carbohydrate derivative transporter activity	molecular_function	0.781	1.000	1
GO:0071806	protein transmembrane transport	biological_process	0.794	1.000	1
GO:0008521	acetyl-CoA transporter activity	molecular_function	0.816	1.000	2
GO:0051185	coenzyme transporter activity	molecular_function	0.816	1.000	2
GO:0017119	Golgi transport complex	cellular_component	0.840	1.000	1
GO:0008643	carbohydrate transport	biological_process	0.844	1.000	2
GO:0015093	ferrous iron transmembrane transporter activity	molecular_function	0.847	1.000	4
GO:0015684	ferrous iron transport	biological_process	0.847	1.000	4
GO:0015103	inorganic anion transmembrane transporter activity	molecular_function	0.851	1.000	2
GO:0006826	iron ion transport	biological_process	0.874	1.000	4
GO:0016192	vesicle-mediated transport	biological_process	0.880	1.000	23
GO:0005381	iron ion transmembrane transporter activity	molecular_function	0.882	1.000	4
GO:0030658	transport vesicle membrane	cellular_component	0.903	1.000	2
GO:0006913	nucleocytoplasmic transport	biological_process	0.907	1.000	4
GO:0051169	nuclear transport	biological_process	0.913	1.000	4
GO:0048193	Golgi vesicle transport	biological_process	0.921	1.000	6
GO:0046915	transition metal ion	molecular_function	0.922	1.000	5

	transmembrane transporter activity				
GO:0030133	transport vesicle	cellular_component	0.929	1.000	2
GO:1902582	single-organism intracellular transport	biological_process	0.935	1.000	20
GO:0006891	intra-Golgi vesicle-mediated transport	biological_process	0.941	1.000	1
GO:0010496	intercellular transport	biological_process	0.947	1.000	1
GO:0046739	transport of virus in multicellular host	biological_process	0.947	1.000	1
GO:0046740	transport of virus in host, cell to cell	biological_process	0.947	1.000	1
GO:1902586	multi-organism intercellular transport	biological_process	0.947	1.000	1
GO:0015031	protein transport	biological_process	0.948	1.000	35
GO:0000041	transition metal ion transport	biological_process	0.950	1.000	5
GO:0006886	intracellular protein transport	biological_process	0.952	1.000	26
GO:0012507	ER to Golgi transport vesicle membrane	cellular_component	0.952	1.000	1
GO:0030134	ER to Golgi transport vesicle	cellular_component	0.952	1.000	1
GO:0044766	multi-organism transport	biological_process	0.954	1.000	1
GO:0046794	transport of virus	biological_process	0.954	1.000	1
GO:0051028	mRNA transport	biological_process	0.966	1.000	1
GO:0006888	ER to Golgi vesicle-mediated transport	biological_process	0.973	1.000	1
GO:0046907	intracellular transport	biological_process	0.977	1.000	33
GO:0016482	cytoplasmic transport	biological_process	0.988	1.000	10

Supplementary table 5: DEGs related to transport activity

Gene_id	log2 FoldChange	p_{val}	p_{adj}	Uniprot_clus
AG6000014-RA	-1.15	0.00	2.00E-03	Q7PPA5
AG6000213-RA	-1.92	0.00	3.00E-06	Q9W056
AG6000301-RA	Inf	0.00	3.00E-02	Q80ZD3
AG6001926-RA	-1.97	0.00	1.00E-03	P43006
AG6002809-RA	-1.70	0.00	4.00E-03	Q55EH8
AG6002811-RA	-5.20	0.00	1.00E-04	Q8T674
AG6002818-RA	-2.73	0.00	1.00E-03	Q55EH8
AG6003753-RA	-1.62	0.00	7.00E-03	B0WC46
AG6004014-RA	-3.11	0.00	4.00E-04	Q7RTX9
AG6004214-RA	-2.33	0.00	4.00E-03	Q6DIV6
AG6004355-RA	-1.45	0.00	4.00E-04	Q8BLQ7
AG6004563-RA	-2.02	0.00	9.00E-03	P79110
AG6005572-RA	1.16	0.00	5.00E-03	Q09143
AG6007279-RA	-1.03	0.00	1.00E-02	F1Q4S1
AG6007887-RA	Inf	0.00	1.00E-03	O00400
AG6008864-RA	-1.19	0.00	1.00E-02	Q09143
AG6008882-RA	-4.33	0.00	3.00E-03	P30823
AG6009281-RA	-1.06	0.00	3.00E-02	Q8IRI6
AG6009836-RA	-3.54	0.00	2.00E-06	Q86WA9
AG6010039-RA	-3.49	0.00	5.00E-12	Q55EH8
AG6010975-RA	-2.34	0.00	2.00E-05	Q00910
AG6011595-RA	-2.04	0.00	3.00E-02	Q6DCX5
AG6012164-RA	-2.23	0.00	7.00E-10	Q8T674
AG6012170-RA	-1.70	0.00	3.00E-02	Q9V7S5
AG6012269-RA	-1.89	0.00	4.00E-05	Q9VCA2
AG6013193-RA	-1.14	0.00	1.00E-02	B0WC46
AG6014444-RA	-1.90	0.00	3.00E-02	Q05B21
AG6015062-RA	-2.12	0.00	1.00E-04	Q6DCX5
AG6016794-RA	-1.77	0.00	5.00E-04	Q16720
AG6017166-RA	-2.07	0.00	6.00E-08	Q4R3Y4
AG6017322-RA	-1.47	0.00	1.00E-03	D3Z5L6
AG6018036-RA	-1.23	0.00	4.00E-02	Q9V7S5

AG6019601-RA	-2.25	0.00	6.00E-03	Q9QXA6
AG6019810-RA	-1.01	0.00	3.00E-02	Q9C6W5
AG6020015-RA	-2.39	0.00	5.00E-02	O17444
AG6020433-RA	1.21	0.00	6.00E-03	Q28J44
AG6020983-RA	-2.56	0.00	2.00E-03	P51905
AG6021045-RA	-1.34	0.00	1.00E-02	A5LGM7
AG6021261-RA	-3.54	0.00	1.00E-07	P13607
AG6022731-RA	-2.44	0.00	4.00E-07	Q55EH8
AG6022769-RA	-3.29	0.00	4.00E-02	Q8BGK6
AG6022771-RA	-3.34	0.00	4.00E-03	Q9N1Q4
AG6022850-RA	-1.84	0.00	1.00E-06	P46411
AG6023736-RA	-2.75	0.00	1.00E-03	Q8N695
AG6023879-RA	-2.49	0.00	4.00E-03	Q1LZI2
AG6024349-RA	-4.30	0.00	2.00E-04	O00400
AG6024828-RA	-2.03	0.00	5.00E-02	Q7RTY0
AG6025111-RA	-3.89	0.00	2.00E-03	Q7K4Y6
AG6025503-RA	-2.21	0.00	1.00E-03	O60669
AG6025988-RA	-1.92	0.00	2.00E-02	O88575
AG6027273-RA	-4.73	0.00	1.00E-04	Q9W056
AG6027412-RA	-1.53	0.00	2.00E-04	P59889
AG6027799-RA	-5.24	0.00	2.00E-09	Q17766
AG6028216-RA	-1.88	0.00	1.00E-02	Q9H8M5
AG6028243-RA	-2.36	0.00	3.00E-08	Q55EH8
AG6028993-RA	-2.96	0.00	2.00E-09	Q55EH8
AG6029799-RA	-1.34	0.00	1.00E-02	Q9W056
AG6031606-RA	4.69	0.00	8.00E-04	A9ZSY3
AG6031948-RA	-2.04	0.00	1.00E-04	Q24048

Gene_id: The gene identifier for differentially expressed genes
log2FoldChange: log2(BC/WB)
pvalue(pval): The statistical p-value
qvalue(padj): p-value after normalization. The smaller the q-value is, the more significant the difference is.
Uniprot_clus: Uniprot ID belonging to eukaryotic cluster group.

Supplementary table 6: HMMER output for A. glycines amino-acid transporters belonging AA_permease family

HMM query: AA_permease								
Accession: PF00324.21								
Description: Amino acid permease protein					Target sequence database: OGS6.0_20180125_proteins			
E-value	**score**	**bias**	**E-value**	**score**	**bias**	**exp**	**N**	**Sequence**
3.7E-90	304	32.5	4.4E-90	303.8	32.5	1.1	1	AG6002163-PA
9.4E-83	279.6	35.8	1.1E-82	279.4	35.8	1.1	1	AG6015502-PA
4.9E-68	231.1	43.1	6.3E-68	230.7	43.1	1.1	1	AG6015990-PA
2.1E-63	215.8	27.7	2.4E-63	215.6	27.7	1	1	AG6015508-PA
5.3E-55	188.1	36.1	4.4E-37	129	4.1	2.1	2	AG6011675-PA
3.2E-36	126.2	29.7	3.2E-36	126.2	29.7	2.6	2	AG6008867-PA
1.2E-35	124.3	27.5	1.2E-35	124.3	27.5	2.5	2	AG6008879-PA
2.4E-35	123.3	24.7	2.4E-35	123.3	24.7	1.9	2	AG6008864-PA
7.3E-34	118.4	23.2	9.2E-34	118.1	23.2	1.1	1	AG6008893-PA
1.6E-31	110.7	35.5	5.3E-31	109	35.3	1.8	1	AG6008882-PA
6.7E-25	88.8	8.2	6.7E-25	88.8	8.2	2.4	2	AG6008884-PA
6.9E-25	88.8	33.2	9.3E-25	88.4	33.2	1.1	1	AG6015418-PA
9.6E-25	88.3	26.9	9.6E-25	88.3	26.9	2.5	2	AG6023366-PA
9.8E-24	85	29.6	2.2E-22	80.5	19.4	2.2	2	AG6022627-PA
2.3E-22	80.5	18.8	2.9E-22	80.2	18.8	1.1	1	AG6008890-PA
2.7E-22	80.2	35	4.7E-22	79.5	35	1.3	1	AG6020364-PA
4E-22	79.7	19.2	4.9E-22	79.4	19.2	1	1	AG6005568-PA
5.1E-20	72.7	26.5	6.8E-20	72.3	26.5	1.1	1	AG6019601-PA
1.4E-17	64.7	14.1	1.4E-17	64.7	14.1	2.6	3	AG6004355-PA
1.3E-12	48.4	11.8	1.5E-12	48.1	11.8	1	1	AG6022769-PA
9.7E-11	42.1	9.8	1.2E-10	41.9	9.8	1	1	AG6004359-PA
9.3E-10	38.9	12.6	1.3E-09	38.4	12.6	1.2	1	AG6022771-PA
9.8E-09	35.5	8.8	1.1E-08	35.4	8.8	1	1	AG6008885-PA
1.2E-08	35.3	11.3	1.4E-08	35	11.3	1.1	1	AG6022629-PA

Supplementary table 7: HMMER output for A. glycines amino-acid transporters belonging AA_trans family

HMM query: AA_trans								
Accession: PF01490.18								
Description: Transmembrane amino acid transporter protein					**Target sequence database: OGS6.0_20180125_proteins**			
E-value	score	bias	E-value	score	bias	exp	N	Sequence
1.60E-92	311.6	36.4	8.40E-53	180.8	13.6	3	2	AG6029795-PA
8.30E-87	292.8	19.6	1.00E-86	292.5	19.6	1	1	AG6004214-PA
1.80E-74	252.1	30.7	2.10E-74	251.9	30.7	1	1	AG6001806-PA
8.00E-74	250	29.2	9.90E-74	249.7	29.2	1	1	AG6017791-PA
6.80E-68	230.5	30	8.40E-68	230.2	30	1	1	AG6000213-PA
2.00E-67	228.9	38.5	2.30E-67	228.7	38.5	1	1	AG6034430-PA
6.30E-65	220.7	37.2	7.50E-65	220.4	37.2	1.1	1	AG6011975-PA
2.40E-64	218.8	35.1	3.10E-64	218.4	35.1	1	1	AG6013324-PA
4.60E-64	217.8	30.3	5.30E-64	217.7	30.3	1	1	AG6027273-PA
1.90E-63	215.8	8.8	1.90E-63	215.8	8.8	1.4	2	AG6029802-PA
4.10E-60	204.8	29.4	5.10E-60	204.5	29.4	1	1	AG6011997-PA
7.20E-48	164.5	25.8	8.70E-48	164.3	25.8	1.1	1	AG6000219-PA
1.10E-46	160.6	21.1	9.10E-37	128	14.6	2	2	AG6012003-PA
5.50E-45	155	35.6	7.40E-45	154.6	35.6	1	1	AG6029680-PA
1.80E-42	146.7	2.1	2.90E-31	109.8	0	2.2	2	AG6029799-PA
8.80E-39	134.6	17.9	1.30E-38	134	17.9	1.2	1	AG6031286-PA
2.30E-15	57.5	37.2	3.90E-10	40.2	27.5	2.3	2	AG6008514-PA
1.20E-14	55	13.6	2.90E-08	34.1	0.6	3.2	3	AG6027265-PA
3.00E-11	43.9	1.4	3.40E-11	43.7	1.4	1	1	AG6000217-PA
9.40E-11	42.3	20.6	5.00E-10	39.9	20.6	2	1	AG6008387-PA
9.70E-09	35.6	20.9	1.20E-08	35.4	20.9	1.1	1	AG6041405-PA
0.0041	17.1	0.8	0.0054	16.7	0.8	1.2	1	AG6012001-PA

Accession: The accession of the query profile
E-value: The expectation value (statistical significance) of the target
Score: The score (in bits) for the target/query correction
Bias: Biased composition correction. Higher bias score may represent false positives
Score on 5th column represents overall sequence score after bias composition correction
Bias on 6th column represents the bias composition that was applied to the bit score
Exp: is the expected number of domains according to HMMER's statistical mode

N, is the number of domains that HMMER's domain postprocessing and annotation pipeline finally decided to identify, annotate, and align in the target sequence.

Supplementary table 8: Log_2fold change from qPCR for candidate amino acid transporters identified from HMMER search.

Gene ID	Log2 fold change				P-value
	BC		WC		
	Avg	SE	Avg	SE	
AG6007894-RA	0.09	0.76	-1.81	0.51	0.05
AG6012777-RA	0.91	0.62	-6.19	1.36	0.00**
AG6035555-RA	1.21	0.30	-0.23	0.14	0.00**
AG6034404-RA	-2.06	1.23	-1.31	0.48	0.58
AG6005572-RA	4.09	0.46	1.96	0.47	0.01*
AG6017614-RA	1.93	0.62	1.26	0.52	0.45
AG6005841-RA	2.88	0.74	-0.52	0.31	0.00**
AG6005568-RA	-0.44	0.27	-0.28	0.38	0.73
AG6008892-RA	0.40	0.19	0.47	0.20	0.78
AG6020364-RA	-0.03	0.98	0.53	0.18	0.59
AG6008884-RA	0.02	0.29	0.23	0.40	0.68
AG6013324-RA	0.20	0.46	-0.03	0.16	0.64
AG6008879-RA	1.33	0.19	4.71	1.10	0.01*
AG6023366-RA	-0.57	0.34	0.51	0.17	0.01*
AG6027265-RA	N/A	N/A	N/A	N/A	N/A
AG6031286-RA	1.17	0.20	-0.29	0.36	0.00
AG6029680-RA	1.65	0.16	1.25	0.13	0.07
AG6011975-RA	-1.58	0.23	-0.23	0.21	0.00
AG6012001-RA	-1.32	0.19	-0.84	0.14	0.05
AG6029795-RA	-0.34	0.23	-0.57	0.13	0.41
AG6000219-RA	-0.97	0.25	-0.08	0.17	0.01
AG6008387-RA	N/A	N/A	N/A	N/A	N/A
AG6017791-RA	0.16	0.11	1.27	0.70	0.16

Validation of differentially expressed genes (DEGs) from quantitative PCR (qPCR). GeneID-72, a high-affinity cationic amino acid transporter obtained from RNAseq data, was enriched in hemolymph enriched bacteriocytes. GeneID-86 shows bacteriocyte enrichment among the top 25 amino acid transporters chosen from HMMER search on the Pfam domain. GeneID-86 was not differentially expressed in our RNAseq data.

Supplementary table 9: TMHMM v.2 prediction model of transmembrane helices for AG6005572-RA

AG6005572-RA		
Length: 280		
Number of predicted TMHs: 6		
Exp number of AAs in TMHs: 130.6		
Exp number, first 60 AAs: 16.2		
Total prob of N-in: 0.87471		
POSSIBLE N-terminal signal sequence		
inside	1	49
TM helix	50	67
outside	68	71
TM helix	72	94
inside	95	144
TM helix	145	167
outside	168	176
TM helix	177	199
inside	200	205
TM helix	206	228
outside	229	232
TM helix	233	255
inside	256	280

Length: The length of protein sequence
Number of predicted TMHs: The total number of predicted transmembrane helices
Exp number of AAs in TMHs: The expected number of amino acids in transmembrane helices. (Values for amino acids greater than 18 are most likely a transmembrane protein)
Exp number, first 60 AAs: The expected number of amino acids in transmembrane helices in the first 60 amino acids.
Total prob of N-in: The total probability that the N-term is on the cytoplasmic side of the membrane.

Supplementary table 10. TMHMM v.2 prediction model of transmembrane helices for AG006001-PA

AG006001-RA		
Length: 546		
Number of predicted TMHs: 13		
Exp number of AAs in TMHs: 280.6		
Exp number, first 60 AAs: 0.00581		
Total prob of N-in: 0.87471		
POSSIBLE N-term signal sequence		
inside	1	9
TM helix	10	29
outside	30	41
TM helix	42	64
inside	65	109
TM helix	110	127
outside	128	133
TM helix	134	156
inside	157	175
TM helix	176	198
outside	199	218
TM helix	219	241
inside	242	270
TM helix	271	293
inside	294	313
TM helix	314	333
outside	334	337
TM helix	338	360
inside	361	410
TM helix	411	433
outside	434	442
TM helix	443	465
inside	466	471
Tm helix	472	494
outside	495	498
TM helix	499	521
inside	522	546

Length: The length of protein sequence for AG006001-PA
Number of predicted TMHs: Total number of predicted transmembrane helices for AG006001-PA
Exp number of AAs in TMHs: The expected number of amino acids in transmembrane helices. Values for amino acids greater than 18 are most likely a transmembrane protein.
Exp number, first 60 AAs: The expected number of amino acids in transmembrane helices in the first 60 amino acids of AG006001-PA
Total prob of N-in: The total probability that the N-term is on the cytoplasmic side of the membrane.

Supplementary table 11: TMHMM v.2 prediction model of transmembrane helices for AG6031286-RA

AG6031286-RA		
Length: 435		
Number of predicted TMHs: 11		
Exp number of AAs in TMHs: 238.20745		
Exp number, first 60 AAs: 42.42069		
Total prob of N-in: 0.98221		
POSSIBLE N-term signal sequence		
inside	1	6
TMhelix	7	29
outside	30	43
TMhelix	44	66
inside	67	101
TMhelix	102	124
outside	125	133
TMhelix	134	153
inside	154	159
TMhelix	160	181
outside	182	195
TMhelix	196	218
inside	219	230
TMhelix	231	253
outside	254	272
TMhelix	273	295
inside	296	314
TMhelix	315	332
outside	333	336
TMhelix	337	359

inside	360	393
TMhelix	394	416
outside	417	435

Length: The length of protein sequence for AG6031286-RA
Number of predicted TMHs: Total number of predicted transmembrane helices for AG6031286-RA
Exp number of AAs in TMHs: The expected number of amino acids in transmembrane helices. Values for amino acids greater than 18 are most likely a transmembrane protein.
Exp number, first 60 AAs: The expected number of amino acids in transmembrane helices in the first 60 amino acids of AG6031286-RA
Total prob of N-in: The total probability that the N-term is on the cytoplasmic side of the membrane.