AMINO ACID METABOLISM IN MATERNAL BACTERIOCYTES OF THE PEA APHID, *ACYRTHOSIPHON PISUM*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Calum William Russell

August 2013

AMINO ACID METABOLISM IN MATERNAL BACTERIOCYTES OF THE PEA APHID, *ACYRTHOSIPHON PISUM*

Calum William Russell, Ph.D.

Cornell University 2013

The pea aphid contains symbiotic bacteria, *Buchnera aphidicola*, in specialized cells called bacteriocytes. These bacteria have been implicated with essential amino acid provisioning to the aphid. The host relies on their symbiotic bacteria for the 10 essential amino acids (EAAs) yet *Buchnera* lacks key genes in the biosynthesis of five of the ten essential amino acids (isoleucine, leucine, valine, phenylalanine, and methionine). I tested the hypothesis, derived from genome annotation, that the missing *Buchnera* reactions are mediated by host enzymes. To lay the metabolic framework for the bacteriocytes, we performed large-scale, high accuracy tandem mass spectrometry (nanoLC-LTQ-Orbtrap) to identify aphid and *Buchnera* proteins in the whole aphid body, purified bacteriocytes, isolated *Buchnera* cells and the residual bacteriocyte fraction (host fraction, HF). This study, in addition to being the first proteomic study of a bacteriocyte symbiosis, showed the enrichment (relative to the whole body) of candidate enzymes that could mediate the missing *Buchnera* reactions in isoleucine, leucine, valine, and phenylalanine synthesis in the bacteriocytes. I found an enrichment for a branched chain aminotransferase (BCAT) and an aspartate aminotransferase (GOT2) that are believed to mediate the terminal reaction of the branched chain amino acid (BCAA – isoleucine, leucine, valine) and phenylalanine synthesis, respectively. I also found enrichment for cystathionine-γ-lyase, which is believed to synthesize the carbon skeleton for isoleucine.

To validate the candidate host enzymes identified from proteomics, the bacteriocytes were separated into a *Buchnera*-fraction and a *Buchnera*-free host cell fraction (HF). Addition of HF to isolated *Buchnera* preparations significantly increased the production of leucine and

phenylalanine; and recombinant enzymes mediating the final reactions in branched chain amino acid and phenylalanine synthesis rescued the production of these EAAs by *Buchnera* preparations without HF. The likely precursors for the missing proximal reactions in isoleucine and methionine synthesis were identified, and differed from predictions based on genome annotations: the carbon skeleton for isoleucine synthesis was stimulated by homoserine and production of the homocysteine precursor of methionine was driven by cystathionine. The coevolution of shared metabolic pathways in this symbiosis can be attributed to host compensation for genomic deterioration in the symbiont, involving changes in host gene expression networks to recruit specific enzymes to the host cell.

BIOGRAPHICAL SKETCH

Calum was born on November 5, 1984 in Tübingen, Germany to David and Maureen Russell. He has lived in Tübingen, New York City, St. Louis, and Ithaca growing up. Rather than belonging to a family of nomads, his family moved as much as it did to allow for his father to eventually become a distinguished professor at Cornell University. Calum owes much of his love of insects to his grandmother, who encouraged him to his mother's dismay. He owes his misdirected belief that graduate school was a great idea to his father. He attended the State University of New York at Geneseo from 2002 until 2006 as a double-major of English Literature and Biology. This institution is incidentally where he learned to talk about himself in the third person and where he took his first entomology courses. There will be no crossover of these skills until many years later where he finds himself writing a biographical sketch for his thesis in an entomology department. Calum was a fairly nondescript, average student during his tenure at SUNY Geneseo. He was not Magna or Summa cum laude of anything (except Star Trek Deep Space Nine trivia), but he obtained high marks in his biology courses and tutored his friends in Organic Chemistry and upper level biology courses (for free). Calum spent the summer before his graduating year working in Michael Kotlikoff's lab, which introduced him to animal physiology, molecular biology, and basic experimental etiquette. The knowledge of how to think about designing experiments correctly and to have the moral and intellectual fortitude to fess up to negative results, and occasional botched experimental designs (despite thinking), has served him well throughout his graduate studies.

Calum began his graduate student career at Cornell University in the fall of 2007 and plans on graduating in the summer of 2013. During his almost 6 years here, he has gained a deeper appreciation for how animals interact with their microbes through graduate study first

with Ann Hajek and second with Angela Douglas. During his tenure at Cornell, Calum has learned and refined many metabolic techniques to study the metabolic interplay between aphids and their resident bacterial symbionts. If he could give one piece of advice to students seeking to study animals and their resident bacteria, it would be to work on a bacterium that could be cultured.

For my parents

ACKNOWLEDGEMENTS

I want to thank my major advisor (Angela Douglas) and committee members (Jeff Scott and Georg Jander) for their mentorship and for allowing me the opportunity to pursue a Ph.D in Entomology at Cornell University. I also want to thank the many lab mates I have had over the years for primarily their friendship and secondarily from preventing me from going bat-crap crazy. In no particular order, I want to thank Adam Dobson (Dobby), Dave Sannino (Baby 'D'), Peter Newell (the Pete-ster), John Chaston (John Chaston), Sophie Bouvaine (the Boss), Stephanie Westmiller (the bigger Boss), Sara Hermann (the Hermann-ator), and Emma Ridley (the 'ex'). Without their support and relatively calm demeanors, my PhD would have been a much more difficult experience. I want to thank all the friends I have made throughout my PhD (outside of the lab), namely Virginia Howick, Sarah Jandricic, Mark Jandricic, Erin Morris, Frank Rinkevich, and Erik Smith. Thank you all for showing me that there is a life outside of the lab and for giving me the support and encouragement I needed to get over life's hurdles.

I want to thank the NSF and the Palmer fellowship for support during my PhD, as well as the TA lines offered by the department.

And finally, I want to thank my parents. Without them, none of my thesis would be possible. Seriously, none of it.

TABLE OF CONTENTS

Biographical sketch iii
Dedication iv
Acknowledgements v
Table of Contents vii
List of Figures x
List of Tables xi

1 Literature Review

1.1 Introduction 1
1.2 Endosymbiosis vs. Organelle 1
1.3 Arthropod Intracellular Symbioses 2
1.4 Xylem and Phloem Feeder Adaptations 6
1.5 Gene loss within EAA pathways 7
1.6 Mapping of *Buchnera* and aphid metabolism 11
1.7 Aphid amino acid metabolism 12
1.8 Research goals 17

2 Quantitative proteomics of maternal bacteriocytes of the pea aphid, *Acyrthosiphon pisum*

2.1 Introduction 18
2.2 Materials and Methods 21
2.2.1 Experimental material and processing for proteomics . 21
2.2.2 Proteome analysis by Nano lc-LTQ-Orbitrap . . . 24
2.2.3 Processing of the MS data, database searches and upload into PRIDE 24
2.2.4 Selection of the best gene models, quantification and creation of protein groups with a high percentage of shared matched spectra 26
2.2.5 Functional classification of quantified proteins . . 27
2.2.6 Hierarchical clustering analysis 28
2.2.7 Transmission electron microscopy 28
2.3 Results 29
2.3.1 Biological properties of the fractions used for proteome analysis 29
2.3.2 Proteomics workflow for identification and quantification of aphid and *Buchnera* proteins 29
2.3.3 The *Buchnera* proteome 37
2.3.4 The aphid proteome 41
2.3.5 Cluster analysis of protein partitioning between aphid and *Buchnera* compartments 41
2.3.6 Aphid proteins enriched in bacteriocyte samples . . 43
2.3.7 Aphid enzymes in amino acid metabolism enriched in bacteriocytes 44

2.3.8 Aphid enzymes in gluconeogenesis enriched in bacteriocytes 51
2.3.9 Aphid enzymes in purine metabolism enriched in bacteriocytes 52
2.3.10 Aphid transporters enriched in bacteriocytes . . . 53
2.3.11 Aphid proteins depleted in bacteriocytes . . . 55
2.4 Discussion 55
2.4.1 Absence of selective protein transfer between aphid and *Buchnera* 55
2.4.2 Metabolite exchange between bacteriocytes and *Buchnera* . 57
2.4.3 The bacteriocyte as a habitat for symbiotic bacteria . . 61

3 Shared amino acid pathways in the pea aphid-*Buchnera* symbiosis
3.1 Introduction 65
3.2 Materials and Methods 67
3.2.1 Aphid rearing 67
3.2.2 Amino acid release and analysis 67
3.2.3 Immunoblots 69
3.2.4 Generation of recombinant enzymes 70
3.2.5 Statistical analyses 71
3.3 Results 72
3.3.1 Impact of substrate on release of isoleucine, leucine, phenylalanine, and valine 72
3.3.2 Metabolic source of 2-oxobutanoate, the precursor of isoleucine 72
3.3.3 Terminal transamination reactions in synthesis of leucine and phenylalanine 75
3.3.4 Metabolic source of homocysteine, the precursor of methionine 79
3.4 Discussion 82

4 Response of the maternal aphid bacteriocyte to dietary essential amino acids
4.1 Introduction 86
4.2 Materials and Methods 87
4.2.1 Aphid rearing 87
4.2.2 Diet composition and tissue extraction . . . 88
4.2.3 Proteomic analysis by nanoLC-LTQ-Orbitrap . . 88
4.2.4 Amino acid production with varying substrate concentration 88
4.2.5 UPLC analysis 89
4.2.6 Statistical analyses 89
4.3 Results 90
4.3.1 Pathways responsive to varying substrate concentration . 90
4.3.2 Amino acid composition with dietary deletions . . 92
4.3.3 Proteomic analysis of bacteriocytes 94
4.4 Discussion 98

5 Research summary and future directions
5.1 Research summary 101
5.2 Future directions 102

A Nitrogen excretion and assimilation in pea aphid
A.1 Introduction 103
A.2 Materials and Methods 103
A.2.1 Aphid rearing 103
A.2.2 Diet and honeydew extraction 103
A.2.3 Supplementation of aphid diet with nitrogen . . . 103
A.2.4 Glutamine synthetase activity 104
A.2.5 UPLC analysis 105
A.3 Results 105
A.3.1 Assimilation efficiency of amino acids . . . 106
A.3.2 Ammonia assimilation and glutamine synthetase activity . 108
A.3.3 Impacts of varying amino acid concentration on Hemolymph and Honeydew 113
A.4. Discussion 117

B Branched chain aminotransferase and phenylalanine aminotransferase activity in bacteriocytes
B.1 Introduction 118
B.2 Materials and Methods 118
B.2.1 Aphid rearing 118
B.2.2 Branched chain and phenylalanine aminotransferase assays 118
B.2.3 Immunoblots 120
B.3 Results 121
B.3.1 Transaminase activity 121
B.3.2 BCAT localization 122
B.4 Discussion 125

Bibliography 126

LIST OF FIGURES

1.1 Hypothesized nitrogen recycling in the bacteriocytes of pea aphids with their symbiotic bacteria *Buchnera aphidicola* 13
1.2 Predicted biosynthetic pathways for certain essential amino acids based upon gene content and metabolic experiments 15
2.1 The five fractions used for the proteome analysis 23
2.2 Experimental and bioinformatics workflow of the proteome analysis . 31
2.3 The proteome of the pea aphid and bacteriocyte fractions . . . 33
2.4 Proteome analysis of the *Buchnera* protein DdlB 40
2.5 Hierarchical cluster analysis of protein distribution across the five fractions 42
2.6 Reactions and transport of metabolites across the aphid-*Buchnera* interface, deduced from the proteome of bacteriocytes 47
2.7 Abundance of individual *Buchnera* enzymes involved in the synthesis of arginine within the purified *Buchnera* cells 49
3.1 Amino acid release dependence on substrates provided in medium . . 73
3.2 Isoleucine release by preparations of *Buchnera* with HF . . . 74
3.3 Leucine and Phenylalanine release by isolated *Buchnera* preparations . 77
3.4 Enzymatic activities in the HF 78
3.5 Methionine release by isolated *Buchnera* preparations . . . 81
4.1 Methionine and Isoleucine release from *Buchnera* 91
4.2 Workflow for proteomic experiments. 96
4.3 Comparison of bacteriocytes from diet- and plant-reared aphids . . 97
A.1 Assimilation efficiency of amino acids in the pea aphid . . . 107
A.2 The relative growth rate of aphids reared on diets with different concentrations of NH_4Cl 109
A.3 Glutamine synthetase activity of tissue and aphids reared on control or 10 mM NH_4Cl diets 112
A.4 Concentration of arginine in hemolymph and honeydew . . . 115
B.1 Activity assays for terminal transamination reactions . . . 123
B.2 General localization of BCAT in the bacteriocyte 124

LIST OF TABLES

1.1 Subset of intracellular nutritional symbionts associated with arthropods and nutrients provisioned to host based upon inferences from genomes . . . 3
1.2 Missing genes in essential amino acid biosynthetic pathways in Sternorrhynchan symbionts 10
2.1 Pearson's linear correlation analysis for the NadjSPC values across the three biological replicates for each of the five sample types 36
4.1 Amino acid composition of bacteriocytes from aphids reared on complete and isoleucine deletion diets 93
4.2 Amino acid composition of bacteriocytes from aphids reared on complete and methionine deletion diets 94
A.1 Amino acid concentration in tissues from aphids reared on control diet or 10 mM NH_4Cl diet 110
A.2 Amino acid concentration of honeydew from aphids reared on control or 10 mM NH_4Cl diet 111
A.3 Correlations between amino acid concentration of the diet and amino acid concentrations in the honeydew and hemolymph 116

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

The following chapter will provide a review of the pertinent literature for understanding the premise behind my thesis. The chapter will deal primarily with the metabolic interactions between the pea aphid and their intracellular symbiotic bacteria *Buchnera aphidicola*. Although my research has focused on essential amino acid biosynthesis in the aphid-*Buchnera* symbiosis, a review of arthropod symbionts and phloem-feeder adaptations follow.

1.2 Endosymbiosis versus Organelles

Beneficial associations with microbes are commonplace in the animal kingdom. An ancient mitochondria-eukarya intracellular symbiosis gave the ancestor of all eukaryotes access to aerobic respiration (Margulis 1970). This beneficial arrangement, with the host gaining novel traits and the microorganisms gaining a stable environment, allowed for the massive radiation of the eukaryotes (Keeling et al 2005, Koonin 2010). Intracellular symbioses are not only ancient and ubiquitous to all extant eukaryotic cells, but have evolved independently in almost all lineages of eukaryotic life other than vertebrate animals (just one example of an intracellular symbiosis in vertebrates is known, algal cells in the North American spotted salamander – Kerney et al 2011). These secondary intracellular acquisitions of microbial partners after the evolution of multicellularity in eukaryotes have many features that are similar to mitochondria. Similar to mitochondria, these partners bring novel metabolic pathways and products to their host and have undergone massive genome erosion due to relaxed selective pressure (Moran 1996). Unlike mitochondria, these microbes have not transferred genes to the host genome that

are later transcribed, translated, and transferred back into the symbiont (IAGC 2010, Poliakov et al 2011), and these microbes are localized in specialized cells and are not found in every animal cell of the organism (Buchner 1965).

1.3 Arthropod Intracellular Nutritional Symbioses

The incorporation of bacterial pathways has provided a fitness advantage to animals that cannot meet their nutritional needs from their diet (Douglas 2009). In the phylum Arthropoda, these nutritional symbioses are widespread amongst groups that feed exclusively on blood, plant sap (phloem and xylem), and detritus. The intracellular symbioses occur in multiple orders in arthropoda including Coleoptera, Diptera, Hymenoptera, Hemiptera, Anoplura, Mallophaga, and Dictyoptera with acquisitions occurring independently in multiple lineages (Heddi et al 1999, Akman et al 2002, Gil et al 2003, Kirkness et al 2010, Sabree et al 2009, Shigenobu et al 2000). As can be expected, the symbioses are not uniform in the retention of metabolic capabilities, but tend to fulfill the nutritional requirements of the host not met by the diet (Table 1.1). In the case of phloem and xylem feeders, the microbial symbiont(s) retains genes coding for enzymes in essential amino acid biosynthesis (Shigenobu et al 2000, McCutcheon and Moran 2010, Santos-Garcia et al 2012, Nakabachi et al 2006, Sabree et al 2012, McCutcheon and von Dohlen 2011). As has been documented, the phloem is low in essential amino acids and the nitrogen is predominately in the form of non-essential amino acids which can include asparagine and glutamine (Gunduz and Douglas 2009). For insects that are obligate blood-feeders through their lifecycle (e.g. tsetse flies, lice), the symbionts have retained genes responsible for the biosynthesis of B vitamins (Akman et al 2002, Kirkness et al 2010), which have been shown to be low abundance in vertebrate blood (Edwards et al 1957). The detritivores typically have

symbionts that provision the host with essential amino acids, but a few have been implicated with recycling urea and ammonia (*Blattabacterium*) to usable nitrogenous compounds (Sabree et al 2009).

Table 1.1: Subset of intracellular nutritional symbionts associated with arthropods and nutrients provisioned to host based upon inferences from genomes

Host	Symbiont	Nutrient(s) provided	Reference
Coleoptera			
Sitophilus oryzae	*Sitophilus oryzae* primary symbiont (SOPE)	Pantothenic acid & riboflavin	Heddi et al 1999
Diptera			
Glossina brevipalpis	*Wigglesworthia glossinidia*	B-vitamins	Akman et al 2002
Hymenoptera			
Blochmannia floridanus	*Camponotus floridanus*	Essential amino acids	Gil et al 2003
Anoplura			
Riesia pediculicola	*Pediculus humanus humanus* L.	B-vitamins	Kirkness et al 2010
Dictyoptera			
Blattabacterium sp.	*Periplaneta americana*	Essential amino acids	Sabree et al 2009
Hemiptera			
Buchnera aphidicola	*Acyrthosiphon pisum*	Essential amino acids	Shigenobu et al 2000

The nutritional provisioning of the host by intracellular symbionts of phloem and blood-feeding insects has been tested through the elimination of the bacteria and rescuing host health by dietary supplementation (Puchta 1956, Nogge 1976, Mittler 1971), and by sequencing the

genomes of the symbionts (Shigenobu et al 2000, McCutcheon and Moran 2010, Santos-Garcia et al 2012, Nakabachi et al 2006, Sabree et al 2012, McCutcheon and von Dohlen 2011). Tsetse flies and lice that have their symbionts removed fail to thrive, but can be rescued by supplementing their blood-diet with B-vitamins (Puchta 1956, Nogge 1976), and the green peach aphid (*Myzus persicae*) has been shown to be partially rescued by supplementing their diet with essential amino acids (Mittler 1971). The genomes of *Wigglesworthia glossinidia* (symbiont of tsetse flies) and *Riesia pediculicola* (symbiont of lice) have retained the biosynthetic pathways for biotin, thiazole, lipoic acid, FAD, folate, pantothenate, thiamine, pyridoxine, and nicotinamide (B-vitamins – Akman et al 2002, Kirkness et al 2010) and the genomes of all sequenced *Buchnera aphidicola* have retained the biosynthetic pathways for 9-10 essential amino acids with the exceptions of *Buchnera* from *Baizongia pistacea* (pistachio aphids) and *Cinara cedri* (cedar aphids) (Shigenobu et al 2000, Tamas et al 2002, van Ham et al 2003, Perez-Brocal et al 2006, Lamelas et al 2011). *Buchnera* from the pistachio aphids have lost the proximal reactions for ornithine synthesis (required for arginine synthesis) yet have retained the distal reactions after ornithine synthesis indicating that the ornithine may be provisioned by the aphid as opposed to synthesized by *Buchnera* (van Ham et al 2003). *Buchnera* from cedar aphids have retained TrpEG, responsible for making anthranilate yet have lost TrpDCBA which converts anthranilate to tryptophan, as well as the capacity to make the non-essential amino acid cysteine (which is retained in all *Buchnera* investigated to date except for in *Schizaphis graminum*) (Perez-Brocal et al 2006). To make cysteine and tryptophan, it is theorized that the cedar aphids rely on metabolic contributions from another symbiont, *Serratia symbiotica*, which has genes coding for TrpDCBA as well as the cysteine biosynthetic pathway (Perez-Brocal et al 2006).

The incorporation of these microbial symbionts into the metabolism of the animal has shown some predictable, convergent patterns. Though the microbes of most lineages evolved over 100 mya (Moran et al, 1993) the genomes of these bacteria have eroded to less than 1 Mb and include the smallest bacterial genome sequenced to date at 138,927 base pairs (*Tremblaya princeps*) (McCutcheon and Moran 2010, McCutcheon and von Dohlen 2011). This is more than four times smaller than the smallest genome of any free-living bacterium, *Mycoplasma genitalium*, with a genome size of 580,076 base pairs (Bak et al 1969). These endosymbionts typically have a high AT bias, though this is by no means the absolute rule as *Hodgkinia cicadicola* and *Tremblaya princeps* have been shown to have high GC contents (McCutcheon and Moran 2010). Though gene losses follow reasonably predictable patterns (i.e. gene losses in particular gene families), the genes that are retained are not the same for each symbiont. The gene families that are reduced, but contain a random functional subset, are involved with DNA repair, cell envelope/peptidoglycan production, and replication and recombination pathways. As well as patterns for losses, there are genes that seem to be preferentially retained. These include genes that code for proteins involved with protein folding and stability: chaperones and chaperonins (McCutcheon and Moran 2011). In *Buchnera*, the GroES-GroEL chaperonin complex has been retained, of which GroEL has been shown to protect enzymatic activity of certain *Buchnera* enzymes (anthranilate synthase – Huang et al, 2008).

Even though general patterns exist for gene loss and retention, this evolution of losses is not due to phylogenetic relatedness. The primary endosymbionts of aphids, psyllids, tsetse flies, and carpenter ants are all γ-proteobacteria, whereas the Auchenorrhynchans (sharpshooters, spittlebugs, cicadas) have one primary symbiont that is a Flavobacterium with one other primary symbiont that is either an α-proteobacterium, β-proteobacterium, or a γ-protobacterium

(Shigenobu et al 2000, Nakabachi et al 2006, Akman et al 2002, Gil et al 2003, McCutcheon and Moran 2010). Like the Auchenorrhynchans, mealybugs (Sternorrhynchan insects) have two primary symbionts: *Tremblaya princeps* (β-proteobacteria) and *Moranella endobia* (γ-proteobacteria (McCutcheon and von Dohler 2011). Though there is no way of definitively knowing how these symbionts were acquired by the host, I can reasonably infer that these acquisitions were independent of one another due to their distinct phylogenetic identity.

1.4 Xylem and Phloem Feeder Adaptations

The xylem and phloem feeders live on a diet insufficient in essential amino acids and at the extremes of osmotic stresses. Phloem feeders deal with a high osmotic difference between the diet and the host tissues due to the sucrose content in the plant sap, whereas xylem feeders deal with very dilute sap (Gunduz and Douglas 2009, Hocking 1980, Brodbeck et al 1993). Phloem feeders have evolved multiple adaptions to deal with the high concentration of sucrose and low concentration of essential amino acids. The high expression of sucrase, coupled with sugar transporters, in the gut allow the sucrose to be degraded to fructose and glucose (with glucose subsequently made into oligosaccharides) and transported into the aphid to lower the osmotic potential (Karley et al 2005, Cristofoletti et al 2003). The amino acids in the diet are taken up at a very high rate, with an assimilation efficiency approaching 1.0 (Prosser et al 1992). This high efficiency may be due to the expansion and specialization of the amino acid transporter gene repertoire in aphids (Price et al 2011). The xylem feeders face a very different set of problems. The xylem fluid, similar to phloem sap, has a low concentration of essential amino acids (predominately made up of asparagine and glutamine in both cases), but unlike phloem sap, xylem fluid is also low in carbon sources (glucose, sucrose and fructose) (Brodbeck et al 1993).

Xylem feeders have overcome the dilute nature of their diet with multiple adaptations: shunting fluid from anterior to posterior gut via the filter chamber to prevent water flooding into the hemocoel, having very high feeding rates (100-300 times dry weight per day), having a high assimilation efficiency of all carbohydrates and amino acids (approaching 1.0), and synchronizing feeding rates to fluctuations in xylem chemistry (Brodbeck et al 1993).

1.5 Gene loss within EAA pathways

The Sternorrhynchan insects (aphids, whiteflies, scales, mealybugs, and psyllids) all contain primary intracellular symbionts that have been implicated in essential amino acid biosynthesis based upon the sequenced genome of the bacterial endosymbiont (Shigenobu et al 2000, Santos-Garcia et al 2012, Sabree et al 2012, McCutcheon and von Dohler 2011, Nakabachi et al 2006). Similar patterns in genome erosion occur between these phylogenetically distinct bacteria (γ-proteobacteria, β-proteobacteria, and flavobacteria). Though the massive initial genome erosion is similar between these bacteria, a small subset of genes coding for enzymes in several essential amino acid biosynthetic pathways have been lost by these bacteria. In *Buchnera aphidicola* (aphids), *Carsonella rudii* (psyllids), *Portiera aleyrodidarum* (whiteflies), *Uzinura diaspidicola* (scales), and *Tremblaya princeps/Moranella endobia* (mealybugs), the only gene retained in the methionine biosynthetic pathways is *metE*, which converts homocysteine to methionine. Genes that code for more proximal enzymes in the pathway, *metB* and *metC* (converts cysteine to cystathionine, and cystathionine to homocysteine respectively) have been lost in all the primary symbionts of Sternorrhynchan sequenced to date. This indicates that the homocysteine is either of host or dietary origin. Similar to the methionine pathway, genes have been lost for the production of the branched chain amino acids (BCAAs) and phenylalanine. The

BCAAs encompass isoleucine, leucine, and valine; of which the terminal reaction, IlvE (a branched chain aminotransferase – BCAT), and a proximal reaction for isoleucine production, IlvA, have been lost in *Buchnera aphidicola* (Shigenobu et al 2000). In other Sternorrhynchan symbionts, both these genes have been lost (*Tremblaya/Moranella, Uzinura diaspidicola*), ilvA has solely been lost (*Carsonella rudii)*, or ilvE has solely been lost (*Portiera aleyrodidarum*) (Shigenobu et al 2000, Santos-Garcia et al 2012, Sabree et al 2012, McCutcheon and von Dohler 2011, Nakabachi et al 2006). The *ilvE* gene codes for an enzyme that transaminates the oxo-acids of leucine, isoleucine, and valine using glutamate as an amino donor. The *ilvA* gene codes for an enzyme that degrades threonine to 2-oxobutanoate, which forms the carbon skeleton of isoleucine. The only other widespread loss in genes that are responsible for amino acid biosynthesis is the loss of *tyrB* in phenylalanine synthesis. *TyrB* has been lost in all sequenced Sternorrhynchan symbionts to date and is responsible for the terminal reaction in the biosynthetic pathway which converts phenylpyruvate to phenylalanine (Shigenobu et al 2000, Santos-Garcia et al 2012, Sabree et al 2012, McCutcheon and von Dohler 2011, Nakabachi et al 2006). These convergent gene losses in pathways that are essential for producing amino acids, which the host requires, is despite the independent acquisition of these symbionts and independent evolution of a nutritional symbiosis. A potential explanation, involving the use of host enzymes to mediate the mission reactions catalyzed by the missing bacterial enzymes, is explored in Wilson et al (2010) for aphids that may shed light on this pattern of convergent evolution.

The pea aphid genome contains genes that code for enzymes with equivalent activity to those lost by the bacterial symbionts. The prevailing hypothesis before the genome was sequenced was that the bacterial symbiont must contain biosynthetic enzymes with broad activities and substrates (Shigenobu et al 2000). After the genome was sequenced, and the

metabolic potential of the aphid was mapped, several genes were implicated with essential amino acid biosynthesis (Wilson et al 2010). Though animals cannot synthesize essential amino acids de novo, they do contain enzymes that can mediate the terminal reactions for branched chain amino acid synthesis (branched chain aminotransferase – BCAT) and phenylalanine synthesis (aspartate aminotransferase – GOT2). These enzymes are commonplace in animals and were found in the pea aphid genome. The aphid BCAT, which catalyzes a reversible reaction, was predicted to mediate the missing terminal step (IlvE) in BCAA synthesis. The proximal reaction generating 2-oxobutanoate (IlvA) was predicted to be mediated by threonine dehydratase, which has an equivalent activity (threonine -> 2-oxobutanoate). The aphid aspartate aminotransferase, GOT2, was predicted to mediate the terminal reaction for phenylalanine synthesis. All of these reactions have been shown to occur with homologues of these aphid enzymes and fall well within the annotated activity of these enzymes (Lain-Guelbenzu et al 1990, Hall et al 1993). While aphid candidates for mediating the missing reactions in BCAA and phenylalanine biosynthesis were readily available, no animal enzyme has ever been annotated or discovered that would fill in for the missing MetB and MetC genes. Animals cannot contribute to methionine synthesis. Wilson et al (2010) posits that the two terminal steps in the transulfuration pathway which normally degrades homocysteine to cysteine via cystathionine beta synthase (homocysteine -> cystathionine) and cystathionine gamma lyase (cystathionine -> cysteine) are running in reverse to produce the homocysteine required by the symbiont to produce methionine. While there has never been any indication that these enzymes can run in reverse, there are no other enzymes that have been annotated to work on cystathionine in the pea aphid.

The convergent evolution between Sternorrhynchan symbionts to lose these genes (Table 1.2) is due to the relaxed selective pressure of containing multiple enzymes within the host-

symbiont complex that can perform the same reactions. This relaxed selective pressure could cause the loss of the bacterial genes with no associated fitness costs. But to date, no experimental evidence exists that validates the shared metabolic pathways in any of the Sternorrhynchan symbionts and their hosts, only genomic and in certain instances transcriptomic evidence exists to support these hypotheses (Wilson et al 2010, Hansen and Moran 2011).

Table 1.2: Missing genes in essential amino acid biosynthetic pathways in Sternorrhynchan symbionts.

Symbiont	*ilvA*	*ilvE*	*metB*	*metC*	*tyrB*
Buchnera aphidicola	-	-	-	-	-
Carsonella ruddii	-	+	-	-	-
Portiera aleyrodidarum	+	-	-	-	-
Tremblaya/ Moranella	-	-	-	-	-
Uzinura diaspidicola	-	-	-	-	-

Pathways complete in these symbionts except for above genes (+ = present, - = absent from the genome). (Shigenobu et al 2000, Santos-Garcia et al 2012, Sabree et al 2012, McCutcheon and von Dohler 2011, Nakabachi et al 2006)

1.6 Mapping of *Buchnera* and aphid metabolism

The eroded gene content of *Buchnera* has allowed for an opportunity to map and model the metabolic inputs and outputs of this bacterium. As *Buchnera* APS contains only 620 genes, 617 of which have known orthologues in *Escherichia coli* (Shigenobu et al 2000), the metabolic potential of the bacterium could easily be mapped to *E. coli* maps and models (Edwards and Palsson 2000). There are three keystone studies designed to map and understand the metabolic inputs and outputs of either *Buchnera* or the aphid-*Buchnera* symbiosis: Thomas et al 2009, Macdonald et al 2011 and Macdonald et al 2012. The initial study mapped the *Buchnera* genes and created a metabolic network for *Buchnera*, which uncovered several striking features. Of all the reactions in the network, 35% of reactions contribute to essential amino acid synthesis, with the penultimate carbon sources as glucose and mannitol and the penultimate sources of nitrogen as glutamine, glutamate, and aspartate. One of the greatest mysteries of the symbiosis is whether and how does the *Buchnera* respond to host requirement for essential amino acids. Flux analyses show that the output can vary in linear proportion to the inputs into the system, perhaps illustrating that the host can meet their own nutritional requirements by differential provisioning of carbon and nitrogen to the symbiont (Thomas et al 2009). A subsequent study demonstrated that different lines of aphids had differential growth rates when certain essential amino acids were deleted from the diet. Certain lines had requirements for essential amino acids (line 1-arg and met, line 2-none, line 3-arg, lys, thr, trp, line 4-his and met) that were predicted to be the result of genome erosion in the *Buchnera* preventing synthesis of these amino acids. Oddly enough, these *Buchnera* had no mutations that would account for these requirements of the aphid for essential amino acids. These amino acid requirements are only for essential amino acids in the pathways that are not theorized to be shared between host and symbiont (terminal reactions

only), hinting that the *Buchnera* may not be releasing, as opposed to synthesizing at lower rates, these solely synthesized amino acids (selfish). The metabolic network constructed is extremely rigid and has a theorized low range of fluxes with 62% of the metabolic reactions unable to alter their flux, with reactions in pyrmidine metablism showing the greatest ranges of predicted fluxes. All these lines of *in silico* evidence present a model of a bacterium that cannot modulate its own metabolism (Macdonald et al 2011).

The final model incorporated the hypothesis that the aphid mediates the terminal reactions of the branched chain amino acids and phenylalanine. The concept of nitrogen recycling has been invoked in many symbioses, including the pea aphid-*Buchnera* symbiosis (Hansen and Moran 2011). In previous studies, the glutamine synthetase/glutamine oxoglutarate aminotransferase (aka glutamate synthase) (GS/GOGAT) is believed to incorporate free ammonia into glutamate and glutamine, which is utilized by *Buchnera* to make essential amino acids. But these predictions did not take into account the glutamate produced by the GS/GOGAT would also be used to the transamination reactions predicted to occur outside of the bacteria. The *Buchnera* would provide the carbon in the form of the oxo-acids of BCAAs and phenylpyruvate and the host would utilize the glutamate made by the GS/GOGAT cycle as an amino donor for the terminal transamination. And finally, the free ammonia is predicted to largely be generated by the bacteriocytes which do not service as a sink for free ammonia to be upgraded to essential amino acids (Macdonald et al 2012) (Figure 1.1).

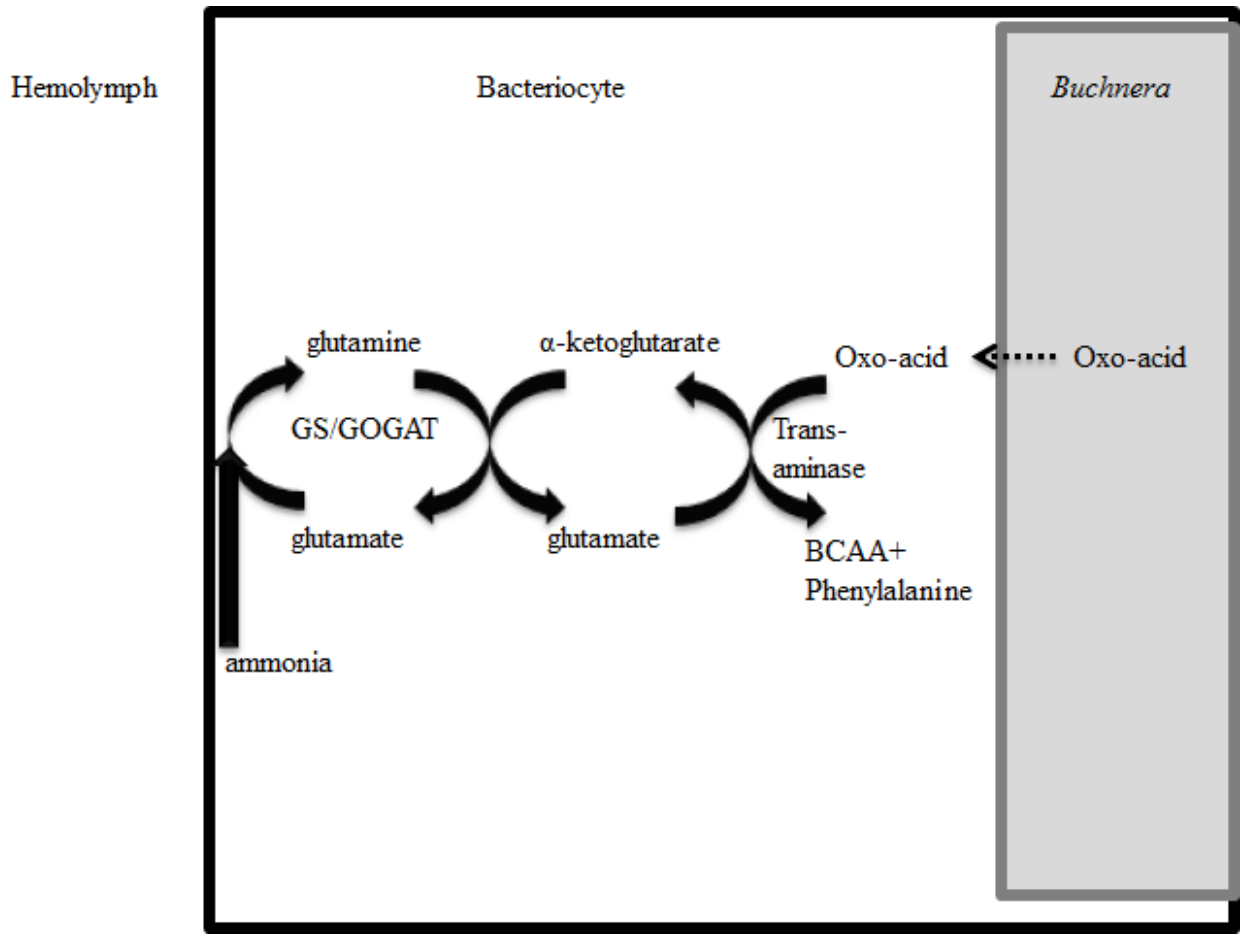


Figure 1.1: Hypothesized nitrogen recycling in the bacteriocyte of pea aphids with their symbiotic bacteria *Buchnera aphidicola* (modified from Macdonald et al 2012). BCAA=branched chain amino acid.

1.7 Aphid amino acid metabolism

Many metabolic studies have been undertaken in the aphid-*Buchnera* symbiosis. These studies have occurred at the organismal level (aphid), at the bacteriocyte level, and at the *Buchnera* level. Though many of these studies occurred before the genomes of *Buchnera* and the pea aphid were released, the findings directly relate to generating hypotheses about shared metabolism between aphids and their resident bacteria. Aphids reared on diets supplemented with ^{14}C labeled amino acids followed some expected patterns, and some unexpected, of radioactivity incorporated into amino acids. Labeled aspartate was incorporated into threonine,

isoleucine, and lysine, which is to be expected as the carbon of aspartate is incorporated into these 3 essential amino acids. Threonine, which is expected to provide the carbon skeleton for isoleucine, is not incorporated into isoleucine. As aspartate, a precursor of threonine, is incorporated into isoleucine and threonine (yet not threonine into isoleucine) several possible metabolic explanations arise (Febvay et al 1995). It is possible that the threonine (from dietary aspartate) produced by the *Buchnera* is the sole source for the carbon skeleton for isoleucine synthesis, which would exclude the dietary threonine. Another possibility is that the carbon from aspartate is entering into the isoleucine biosynthetic pathway via another route that has previously not been annotated or explored. Another oddity is the lack of incorporation of the carbon from glutamate into arginine, as the arginine biosynthetic pathway (from glutamate) is intact in *Buchnera aphidicola* APS. Furthering the oddity of arginine, ^{14}C proline was incorporated into arginine. One possible explanation for these findings is that dietary amino acids are not directly utilized for the biosynthesis of essential amino acids by *Buchnera,* which is an explanation supported by the de novo synthesis of certain essential amino acids from sucrose (valine, isoleucine, leucine, phenylalanine, and lysine) (Febvay et al 1999). Another explanation is that proline is somehow incorporated into arginine (Figure 1.2)

Other studies have focused on particular biosynthetic pathways in *Buchnera*: tryptophan and methionine synthesis. *Myzus persicae,* the green peach aphid, fed labeled inorganic sulfate led to an incorporation of the sulfur into cysteine and methionine, indicating that sulfate is being incorporated into the sulfur amino acids (Douglas 1988). In the pea aphid, *Buchnera* can make cysteine from serine and inorganic sulfate but lacks the genes coding for enzymes that convert cysteine to homocysteine, the immediate precursor of methionine. Either the *Buchnera* of *Myzus persicae* contain genes that are missing in the pea aphid *Buchnera*, or the aphid is mediating

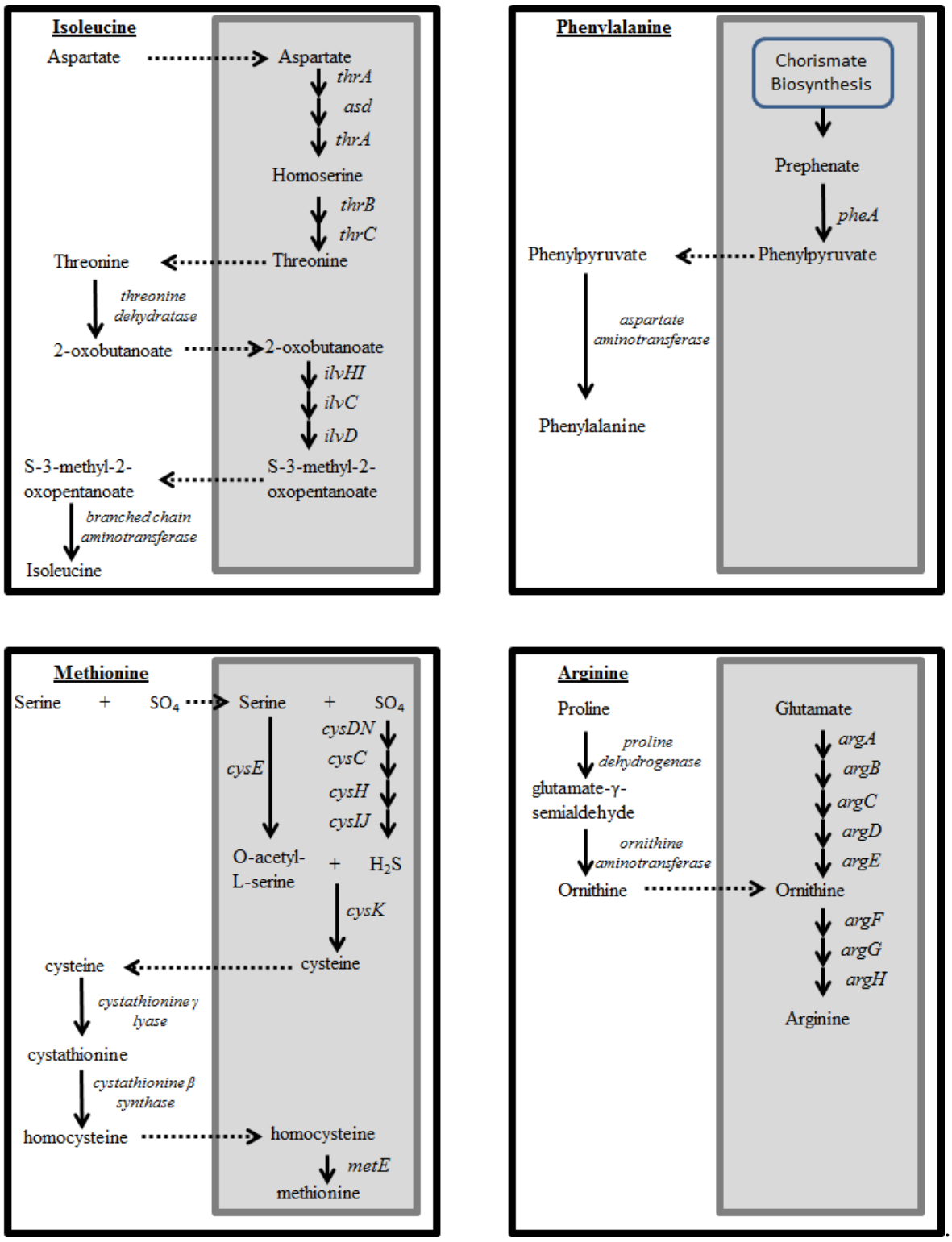


Figure 1.2: Predicted biosynthetic pathways for certain essential amino acids based upon gene content and metabolic experiments (shaded box is *Buchnera*, white box is the aphid; solid line is enzymatic step, dotted line is transportation step)

these missing reactions in a pathway I don't currently understand. Tryptophan synthesis has been inferred by the excretion of tryptophan into the honeydew from aphids feeding on tryptophan-free diet. Further evidence for tryptophan production by *Buchnera* is the presence of tryptophan synthetase activity in symbiotic aphids that is lost when the bacteria are disrupted by antibiotics (Douglas and Prosser 1992).

At the bacteriocyte level, there is some indication that ^{15}N glutamine is taken up and subsequently converted to ^{15}N glutamate inside the cells (Sasaki and Ishikawa 1995). This conversion is probably due to the aphid glutamate synthase activity, which can regenerate 2-oxoglutarate using glutamine as an amino donor. Unfortunately, the subsequent amino donations from glutamate in the cytosol of the bacteriocytes were not addressed. At the bacterial level, several labeling studies indicate that certain amino acids are taken up and utilized by *Buchnera*: glutamate and aspartate (Sasaki and Ishikawa 1995, Whitehead and Douglas 1993). The uptake of glutamate and aspartate is especially important, as these two non-essential amino acids have been predicted to be required for essential amino acid biosynthesis from metabolic models (Macdonald et al 2011). Crude *Buchnera* preps isolated from whole aphids (filtered through 5 µm pores) were shown to incorporate the ^{15}N from ^{15}N-glutamate into the branched chain amino acids (isoleucine, leucine, and valine), phenylalanine, alanine, aspartic acid, glutamine, and proline (Sasaki and Ishikawa 1995). The incorporation of the amino group into the conjugate oxo-acids of the branched chain amino acids and phenylalanine is expected to occur outside of *Buchnera* due to missing the IlvE and TyrB genes. There are multiple explanations for the transamination without containing the appropriate transaminases: aphid enzymes that may mediate these reactions associate with *Buchnera*, the preparations of *Buchnera* were crude and contained aphid enzymes (likely), or *Buchnera* contains enzymes that can mediate these missing

reactions that are not IlvE and TyrB. The incorporation of the ^{15}N from ^{15}N-glutamate into the non-essential amino acids (alanine, aspartic acid, glutamine, and proline) is expected as the amino group can be donated to pyruvate to make alanine (via the alanine aminotransferase,) donated to another glutamate to make glutamine (via the glutamine synthetase), or glutamate can be converted to proline (via the pyrroline-5-carboxylate dehydrogenase) by enzymes that are encoded in the aphid genome. This release/synthesis of non-essential amino acids that only the aphid can make is further evidence that these *Buchnera* preparations are contaminated with aphid enzymes. Continuing in being an odd essential amino acid, arginine does not seem to have incorporated ^{15}N from ^{15}N-glutamate. This is odd as the initial N comes from glutamate in the hypothesized biosynthetic pathway for arginine in *Buchnera*. This may be indicative that arginine is not synthesized from glutamate, but from another source of carbon and nitrogen.

The hypothesis of shared metabolic pathways is compatible with previous labeled studies. The transamination reactions of the branched chain amino acids and phenylalanine of crude *Buchnera* preparations support the belief that there are active transaminases in the aphid-*Buchnera* symbiosis, the incorporation of labeled sulfate into cysteine and methionine hints at the possibility of enzymes that can convert cysteine to homocysteine (the immediate precursor to methionine), and an inability of ^{15}N glutamate to be an amino donor for arginine synthesis (yet ^{14}C proline contributes carbon) hints at an alternative biosynthetic pathway for this essential amino acid not considered before (Figure 1.2)

1.8 Research Goals

This thesis will address two outstanding questions in the pea aphid-*Buchnera* symbiosis: how do *Buchnera* make essential amino acids while missing required enzymes and does the symbiosis respond to changes in dietary content of amino acids?

CHAPTER 2

QUANTITATIVE PROTEOMICS OF MATERNAL BACTERIOCYTES OF THE PEA APHID, *ACYRTHOSIPHON PISUM**

2.1 Introduction

Animals are metabolically impoverished. They are unable to synthesize 9 or 10 of the essential amino acids (EAAs) (Douglas 2010) and various cofactors required for the function of key metabolic enzymes (*e.g.* the B vitamins). Repeatedly over evolutionary time, animals have resolved metabolic limitations by forming intimate symbioses with bacteria possessing the metabolic capabilities that they lack (Douglas 2010, Moran 2007). Among the great diversity of animal-microbial symbioses with a metabolic basis, the bacteriocyte symbioses in insects are remarkable for their intimacy (Embley and Martin 2006, Douglas 1989). The symbiotic microorganisms are intracellular, located in just one cell type, known as the bacteriocyte, which sole function appears to be to maintain the symbiosis; Bacteriocytes are transmitted vertically from the mother insect to her offspring usually via the egg cytoplasm. The association is obligate for both partners. The insect hosts, which generally feed on nutritionally unbalanced diets, benefit from nutrients provided by the bacteria (Douglas 1989, Buchner 1965). For example, all insects feeding through the life cycle on plant sap (*e.g.* aphids, whitefly, and leafhoppers) and

* Presented with minor modifications from the originally published article Poliakov A[**], Russell CW[**], Ponnala L, Hoops HJ, Sun Q, Douglas AE and van Wijk KJ. 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. Molecular and Cellular Proteomics 10: M10.007039. [**] = joint first authors. Tissue isolation and metabolic mapping was undertaken by Calum Russell.

vertebrate blood (*e.g.* tsetse fly, bedbugs, and sucking lice) have bacteriocyte symbionts that provide their insect host with nutrients in short dietary supply (EAAs for plant sap feeders and B vitamins for blood-feeders) (Wilcox et al 2003, Douglas 2006). The obligacy of the symbiosis for the bacterial partner can be attributed to the very small size of their genomes, generally <1 Mb and, in some taxa, of comparable size to mitochondria and plastids (Pais et al 2008). Relaxed selection on genes not required in the insect habitat, and genome erosion linked to the very small effective population size of the vertically transmitted bacteria, both contribute to gene loss of these symbionts (Embley and Martin 2006, McCutcheon 2010, Fares et al 2002). It has been suggested that some bacteriocyte symbionts are organelles (*e.g.* as defined by requiring import of host nuclear-encoded gene products, as in mitochondria and plastids (Moran 1996)) or on an evolutionary trajectory (*e.g.* as evidenced by extensive loss of bacterial gene content) to organelle status (Toft and Andersson 2010).

Symbiosis function, especially the nutritional role of the symbiotic microorganisms, is relatively well-studied from a whole-organism perspective for several bacteriocyte symbioses, but understanding of the molecular basis of these interactions is currently fragmentary. A coherent explanation for the sustained nutrient release from the symbionts and the coordinated growth/proliferation of the two partners, in terms of the metabolic pathways and immune effectors of the host cell and their interactions with the bacteria is lacking. The symbiosis between the pea aphid *Acyrthosiphon pisum* and the γ-proteobacterium *Buchnera aphidicola* is particularly amenable for molecular analysis because the genomes of both partners have been sequenced (IAGC 2010, Shigenobu et al 2000). Moreover, a wealth of experimental data supports the function of *Buchnera*, the upgrading of nonessential amino acids to essential amino acids (Douglas 1998).

The aim of this study was to address two unresolved key questions in aphid-*Buchnera* symbiosis. First, are any aphid proteins targeted specifically to the *Buchnera* cells? This question is of great evolutionary significance because the targeting of nuclear-encoded proteins is the hall-mark of a bacterial-derived organelle (Cavalier-Smith 2003, Martin 2010); the presence of such host proteins within the *Buchnera* cells would indicate that these bacterial cells are on their way to become organelles (or should be defined as organelles already). Second, which aphid proteins are enriched in bacteriocytes relative to the whole aphid body? This class of proteins is predicted to include aphid proteins that contribute directly to the symbiotic relationship. Of particular interest are proteins with known or predicted metabolic or immune function, contributing to nutrient flux among the partners and regulation of the *Buchnera* cells, respectively. In addition, "orphan" proteins (without obvious homologs in other organisms) would suggest specific adaptation of the aphid to its symbiotic relation with *Buchnera*. Finally, it is conceivable *Buchnera* cells secrete specific proteins into the bacteriocyte to contribute to host-symbiont signal exchange or as a nutritional source of EAAs.

To answer these questions, proteome analysis was carried out in which the proteome of whole aphids was compared with purified intact bacteriocytes, partially purified and highly purified *Buchnera* cells, and the host fraction of bacteriocytes. Mass-spectrometry-based label-free spectral counting was applied to quantify the distribution of proteins across these fractions, in combination with hierarchical cluster analysis. To date, the pea aphid and other aphid species have been the subject of various proteome analyses (An Nguyen et al 2007, Le Trionnaire et al 2009, Carolan et al 2009, Francis et al 2010, Wang et al 2010), but most did not address the symbiotic relationship. Only one published study was specific to the aphid symbiosis; proteins extracted from isolated *Buchnera* cells were analyzed by peptide mass fingerprinting confirming

the expression of 50 *Buchnera* proteins (Maezawa et al 2006). The current study identified more than 1900 aphid and 400 *Buchnera* proteins across these various aphid, bacteriocyte, and *Buchnera* fractions. The mass spectrometry-based identification was performed with a high resolution, high mass accuracy instrument, operating at its highest resolution (100,000), thus yielding a valuable data set that can be used to improve aphid genome annotation and provide evidence for predicted *Buchnera* protein models. This study provides not only excellent proteome coverage of both aphids and *Buchnera*, but also answers several important and fundamental questions regarding the symbiotic relationship.

2.2 Materials and Methods

2.2.1 Experimental Material and Processing for Proteomics

The pea aphid *Acyrthosiphon pisum* was collected from an alfalfa crop at Freeville farms, Freeville, NY in June 2009, and reared on preflowering *Vicia faba* cv. Windsor at 20 °C with 16 h light: 8 h dark (120 μE m^{-2} s^{-1}). Three independent replicate experiments were conducted on 7-day-old apterous fourth-instar larvae taken from cultures in three consecutive months between December 2009 and February 2010. The five samples per experiment (see Figure 2.1) comprised: whole aphid bodies (WB), bacteriocytes dissected from the maternal body cavity (BC), partially purified *Buchnera* cells isolated from dissected bacteriocytes (Bu-1), *Buchnera* cells purified by Percoll gradient centrifugation (Bu-2), and supernatant fraction of dissected bacteriocytes (BR). WB samples comprised five aphids flash-frozen in liquid nitrogen, homogenized frozen in a mortar and pestle, and resuspended in 2× SDS-PAGE loading buffer (125 mm Tris-HCl pH 6.8, 10% v/v β-mercaptoethanol, 20% v/v glycerol, 4% w/v SDS, few grains of bromphenol blue). For BC samples, the maternal bacteriocytes were dissected with fine pins from 60 aphids into 30 μl PBS (0.15 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, pH 7.4) final

volume, and mixed with 10 µl 4× SDS-PAGE loading buffer. For Bu-1, Bu-2, and BR samples, maternal bacteriocytes were dissected from 100, 700, and 300 aphids, respectively, into ice-cold extraction buffer (25 mM KCl, 10 mM $MgCl_2$, 35 mM Tris-HCl pH 7.5, 0.25 M sucrose), homogenized as above, and centrifuged at 600 × *g* for 5 min at 4 °C. For BR samples, the supernatant was collected, and Bu-1 samples comprised the pellet resuspended in loading buffer. To obtain Bu-2 samples, the pellet was resuspended in 0.2 ml 0.25 m sucrose, layered onto a step gradient of 45%, 27%, 18%, and 9% Percoll in 0.25 m sucrose. After centrifugation at 200 × *g* for 20 min, the band of *Buchnera* cells (confirmed by microscopical examination) at the interface between 27 and 18% Percoll was aspirated off, diluted fivefold in the extraction buffer, centrifuged at 700 × *g* for 5 min, and resuspended in SDS-PAGE loading buffer. All samples in loading buffer were incubated at 90 °C for 5 min prior to separation by SDS-PAGE.

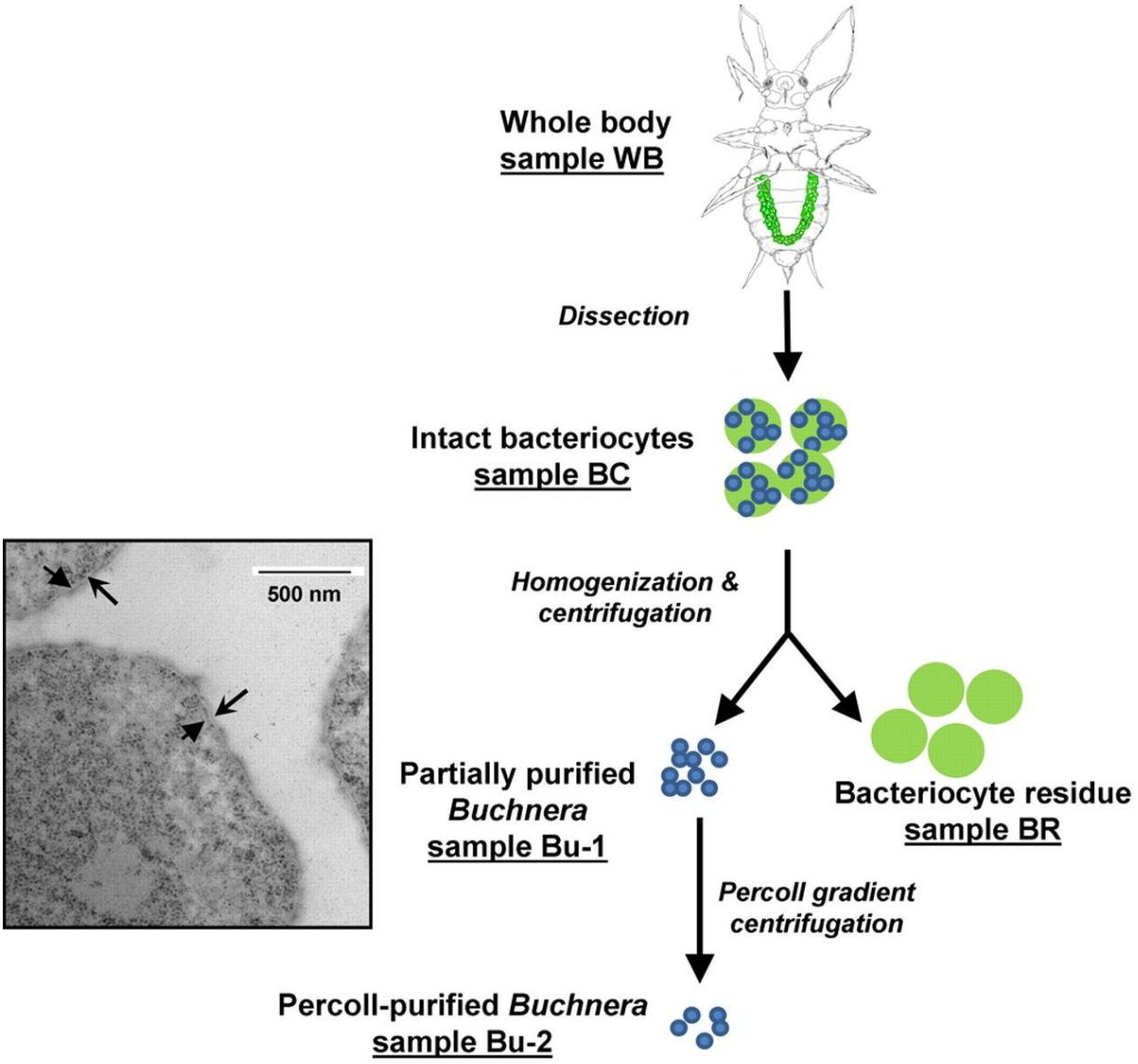


Figure 2.1 The five fractions used for the proteome analysis**.** The maternal bacteriocytes form a U-shaped organ in the body cavity of the insect (green in WB sample) and were dissected from the insects for preparation of the other samples. The aphids also contain bacteriocytes in their embryos (not illustrated), which contributed to the proteome of WB only. The *Buchnera* from Bu-1 sample (TEM image insert) are bounded by the inner bacterial membrane (↑) and outer bacterial membrane (↑), demonstrating that the symbiosomal membrane of aphid origin is lost during preparation. (Transmission electron microscopy performed by Harold Hoops)

2.2.2 Proteome Analysis by Nanolc-LTQ-Orbitrap

Proteins were separated by SDS-PAGE (10–14% acrylamide). Each gel lane was cut in 10 slices, proteins were digested with trypsin and the extracted peptides were analyzed by nanoLC-LTQ-Orbitrap (Thermoelectron) mass spectrometry using data-dependent acquisition and dynamic exclusion, similar as described in (Zybailov et al 2009). Peptide extracts were automatically loaded on a guard column (LC Packings; MGU-30-C18PM) via an autosampler, followed by separation on a PepMap C18 reverse-phase nano column (LC Packings nan75–15-03-C18PM), using 90-min gradients with 95% water, 5% acetonitrile (ACN), 0.1% formic acid (FA) (solvent A), and 95% ACN, 5% water, 0.1% FA (solvent B) at a flow rate of 200 nl/min. The acquisition cycle consisted of a survey MS scan in the Orbitrap with a set mass range from 350 *m/z* to 1800 *m/z* at the highest resolving power (100,000), followed by five data-dependent MS/MS scans acquired in the LTQ. Dynamic exclusion was used with the following parameters: exclusion size 500, repeat count 2, repeat duration 30 s, exclusion time 180 s, exclusion window ± 6 ppm. Target values were set at 5×10^5 and 10^4 for the survey and Tandem MS scans, respectively, and the maximum ion accumulation times were set at 200 ms in both cases. Regular scans were used both for the precursor and tandem MS with no averaging. The precursor isolation window was set at 2 *m/z* with mono-isotopic peak selection, and the FTMS preview option was used. The complete analysis comprised three independent biological replicates. In total 150 MS runs were carried out, with extensive blanks between each sample to avoid carry-over of peptides that could bias quantification.

2.2.3 Processing of the MS Data, Database Searches and Upload Into PRIDE

Peak lists (.mgf format) were generated using DTA supercharge (v1.19) software (http://msquant.sourceforge.net/) and searched with Mascot v2.2 (Matrix Science) against a

combined database containing the aphid genome with 34,834 protein-coding gene models (http://www.aphidbase.com/aphidbase), the *Buchnera* genome with 587 protein coding genes (including 14 predicted pseudogenes and a few plasmid genes) (Shigenobu et al 2000) and 187 sequences for known contaminants (*e.g.* keratin, trypsin) and concatenated with a decoy database where all the sequences were reversed; in total this database contained 71,216 protein sequences. For off-line calibration, first a preliminary search was conducted with the precursor tolerance window set at ±30 ppm. Peptides with the ion scores above 40 were chosen as benchmarks to determine the offset for each liquid chromatography (LC)-tandem MS (MS/MS) run. This offset was then applied to adjust precursor masses in the peak lists of the respective .mgf file for recalibration using a Perl script. The recalibrated peak lists were searched against the assembled *Buchnera*/aphid database. Each of the peak lists were searched using Mascot v2.2 (maximum *p* value of 0.01) for full tryptic peptides using a precursor ion tolerance window set at ±6 ppm, variable methionine oxidation and fixed cysteine carbamido-methylation and maximally one missed cleavage allowed. The maximum fragment ion tolerance (MS/MS) was 0.8 Da. The minimal ion score threshold was chosen such that a peptide false discovery rate (FDR) below 1% was achieved. Using an in-house written filter, the search results were further filtered as follows: For identification with two or more peptides, the minimum ion score threshold was set to 30. For protein identification based on a single peptide, the minimum ion score threshold was set to 33, and the mass accuracy of the precursor ion was required to be within ±3 ppm. The peptide FDR was calculated as: 2 × (decoy_hits)/(target + decoy hits) and was below 1%. The FDR of proteins identified with two or more peptides was zero. Peptides with less than seven amino acids were discarded. All mass spectral data (the mgf files reformatted as PRIDE XML files) are available via the Proteomics Identifications database (PRIDE) at http://www.ebi.ac.uk/pride/. For those

proteins identified with only one unique peptide (irrespective of charge state and PTM), for each accession the charge state, m/z value, mass error for the precursor ion, the peptide ion score and E-value as reported by MASCOT, as well as assigned MS/MS spectrum is provided.

The predicted *Buchnera* proteome is ~60-fold smaller than the predicted aphid proteome. Due to concern that searching a combined *Buchnera*/aphid database could potentially lead to masking some lower scoring *Buchnera* proteins by the large number of aphid sequences, the spectral data was also searched against separate aphid and *Buchnera* databases, but this did not result in a significant change in identification of *Buchnera* proteins and these results will not be further considered (not shown).

2.2.4 Selection of the Best Gene Models, Quantification and Creation of Protein Groups with a High Percentage of Shared Matched Spectra

Several aphid genes have more than one gene model, and in such cases the protein form with the highest number of matched spectra was selected; if two gene models had the same number of matched spectra, the model with the lower digit was selected. For quantification, each protein accession was scored for total spectral counts (SPC), unique SPC (uniquely matching to an accession) and adjusted SPC (Friso et al 2010). The latter is the sum of unique SPCs and SPCs from shared peptides across accessions with SPC distributed in proportion to their unique SPCs, if applicable. The normalized adjSPC (NadjSPC) for each protein was calculated through division of adjSPC by the sum of all adjSPC values for the proteins from that gel lane. To calculate the relative abundance for each protein sample type (per gel lane), the total adjSPC was divided by the predicted protein length, yielding the spectral abundance factor (SAF). The SAF values were then normalized to the total SAF of proteins identified in the gel lane, yielding

normalized spectral abundance factors (NSAFs). NadjSPC provides a relative protein abundance measure by mass, whereas NSAF estimates relative protein concentration within a particular sample. Proteins that shared more than ~80% of their matched adjusted peptides with other proteins across the complete data set were grouped into clusters by generating a similarity matrix through calculation of the dice coefficient between each pair of identified proteins as described in (Friso et al 2010). In all analyses the group was represented by a single member of the group with the highest value of adjSPC across all experiments and highest alphabetical order.

2.2.5 Functional Classification of Quantified Proteins

Aphid proteins were classified by Gene Ontology (GO) (www.geneontology.org/) using only GO terms within 'molecular function'. GO terms for each protein were collated, and then every GO term was ranked according to its abundance in the data set. All GO terms with ≥10 entries in the data set were assigned to one of seven functional groups: Metabolism, Signaling, Structural, Homeostasis, Transport, Information Transfer (*i.e.* associated with transcriptional and translational machinery), or "Other" for proteins without GO terms. All proteins described by the highest ranking GO term (*i.e.* with the most proteins) were assigned to the relevant functional group. The procedure was then repeated for the remaining proteins, using the second-ranking GO term; then the third, and so on. This procedure excluded multiple scoring of proteins. Multifun classification, developed for *Escherichia coli* (Serres and Riley 2000), is better suited than GO for the small and functionally biased *Buchnera* gene content. *Buchnera* proteins were classified into six Multifun categories, selected manually to accommodate the function of expressed proteins: Metabolism (Multifun category 1); Information Transfer (Multifun category 2, except 2.3.4); Transport (Multifun category 4), Flagellum (Multifun category 6.4), Homeostasis, comprising Protection (Multifun category 5.6) and Chaperones & Stress Response (Multifun

categories 2.3.4 & 5.5); and Other. To avoid double-counting proteins with multiple Multifun category entries, the arbitrary rule was adopted that any proteins entered as (a) Metabolism or Information Transfer, and (b) another category (Transport, Homeostasis) were omitted from the list of Metabolism and information Transfer proteins.

2.2.6 Hierarchical Clustering Analysis

Proteins with similar profiles across the samples in Figure 2.1 were grouped by hierachical clustering, a method that requires no prior assumptions about the number of clusters (Belacel et al 2006). Proteins with abundance <0.001 nadjSPC were excluded, after tests with different thresholds confirmed that this level was optimal (not shown). Using the Statistics toolbox of MATLAB version 7 (Mathworks, Inc.), the linear correlation (ρ) between every pair of proteins with NadjSPC distribution across the protein profiled of WB, BC, Bu-1, Bu-2, and BR:, $X_1,\ldots X_n$ and $Y_1,,\ldots Y_n$ where $n = 5$.

$$\rho_{XY} = \frac{\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^{n}(X_j - \bar{X})^2 \sum_{j=1}^{n}(Y_j - \bar{Y})^2}} \qquad \text{(Eq. 1)}$$

was derived. This was then converted into a distance measure $\Delta_{XY} = 1 - \rho_{XY}$. Protein pairs with similar profiles across the fractions have higher correlations and in turn, have smaller distance values. A linkage map based on the average distance among protein pairs was then constructed to yield a hierarchical cluster tree (dendrogram).

2.2.7 Transmission Electron Microscopy

Buchnera samples prepared as for Bu-1 samples (see above) were fixed on ice-cold 2% glutaraldehyde in 0.05 m cacodylate buffer and 4% sucrose for 2 h, postfixed on 1% osmium tetroxide for 1 h, and dehydrated in ethanol series. The pellet was embedded in Qyetol 651 epoxy

resin (Electron Microscopy Sciences, Hartfield, PA) and polymerized for 8 h at 60 °C. After sectioning and staining with lead citrate and uranyl acetate, the samples were examined hours at 60 °C. After sectioning and staining with lead citrate and uranyl acetate, the samples were examined with a Morgagni model 268 TEM (FEI Co., Hillsboro, OR) operated at 80 kV. Images were obtained on a model AMT XR-40 four megapixel CCD digital camera (Advanced Microscope Techniques Corp, Woburn, USA).

2.3 Results

2.3.1 Biological Properties of the Fractions Used for Proteome Analysis

We generated a series of samples from pea aphids (by combined dissection, homogenization, and differential centrifugation; three independent replicates) designed to obtain progressive enrichment of the symbiotic bacterium *Buchnera* from the whole body of the aphids (WB) through dissected bacteriocytes (the cells containing *Buchnera*, BC), partially purified from dissected maternal bacteriocytes (Bu-1) to *Buchnera* cells purified by Percoll gradient centrifugation (Bu-2) (Figure 2.1). The fifth set of samples comprised the residual fraction of dissected bacteriocytes (BR), which did not contain intact *Buchnera* cells detectable by light microscopy. Transmission electron microscopy revealed that the isolated *Buchnera* cells were intact and were bounded by two membranes (Figure 2.1), in contrast to three membranes bounding *Buchnera* cells in the bacteriocyte (and present in BC samples) (Lamb and Hinde 1967), suggesting that the outermost membrane, known as the symbiosomal membrane of aphid origin, was eliminated during preparation of Bu-1 and Bu-2.

2.3.2 Proteomics Workflow for Identification and Quantification of Aphid and Buchnera Proteins

The proteomics workflow is summarized in Figure 2.2. The proteins in each sample were separated by SDS-PAGE (Figure 2.3A), followed by processing of the gel lanes for data-dependent MS/MS analysis using a nanoLC-ESI-LTQ-Orbitrap mass spectrometer. Three independent biological replicates were used. Using MASCOT software, the spectral data were searched against the combined aphid genome sequence and the *Buchnera* sequenced genome. The search results were further processed to: (i) reduce false positive identification, (ii) avoid over-identification of members of protein families, and (iii) select the best gene model for each gene. This finally resulted in identification of 1940 aphid proteins (out of 34,616 predicted aphid proteins, counting only one model per gene) and 400 *Buchnera* proteins (out of 584 predicted proteins and pseudogenes). Thus ~68% of the predicted *Buchnera* proteome was identified. Using the predictor TM-HMM (http://www.cbs.dtu.dk/services/TMHMM/), and the signal peptide predictor SignalP (http://www.cbs.dtu.dk/services/SignalP/) to remove false positive single transmembrane domain (TMD) proteins, the theoretical *Buchnera* proteome is estimated to have about 85 integral α-helical TMD proteins. The proteome detected 37 of these TMDs proteins (43%) indicating that these TMD were well represented in the identified proteome even if they were somewhat underrepresented as can be expected.

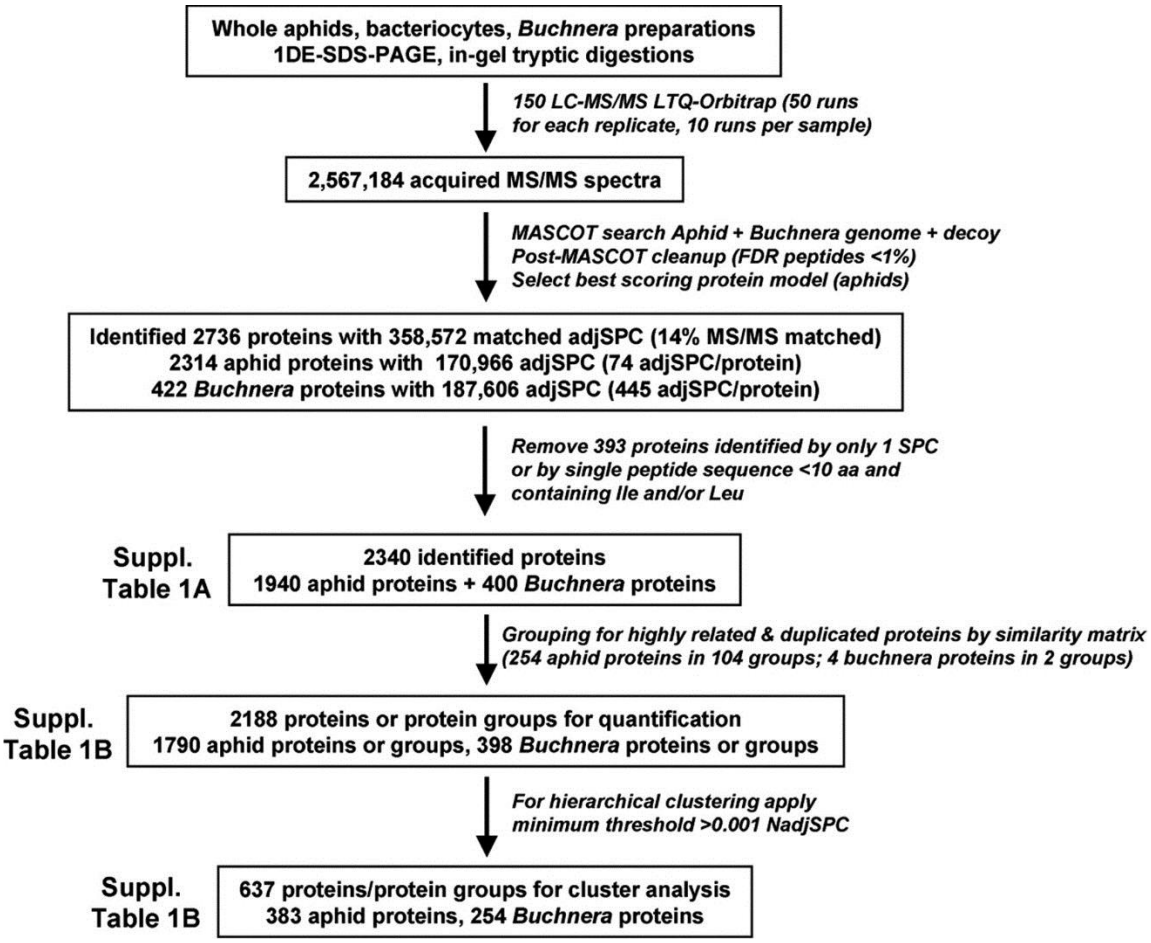


Figure 2.2. Experimental and bioinformatics workflow of the proteome analysis.

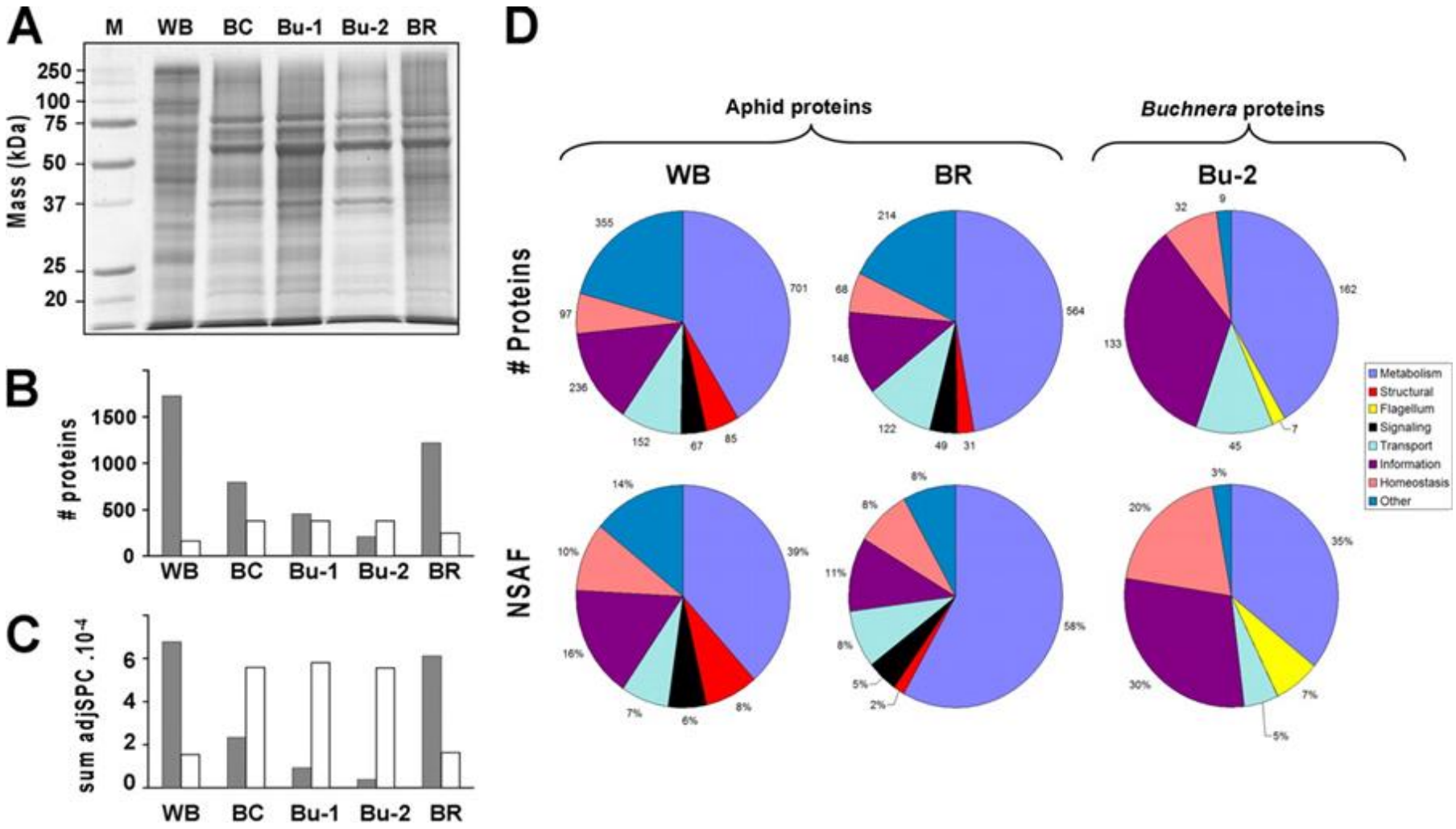

A
M
WB
BC
Bu-1
Bu-2
BR
Mass (kDa)
250
100
75
50
37
25
20
B
proteins
1500
1000
500
0
WB
BC
Bu-1
Bu-2
BR
C
sum adjSPC .10-4
6
4
2
0
WB
BC
Bu-1
Bu-2
BR
D
Aphid proteins
Buchnera proteins
WB
BR
Bu-2
Proteins
NSAF
355
701
97
236
152
67
85
214
564
68
148
122
49
31
32
9
162
133
45
7
14%
39%
10%
16%
7%
6%
8%
8%
58%
11%
8%
5%
2%
3%
35%
20%
30%
7%
5%
Metabolism
Structural
Flagellum
Signaling
Transport
Information
Homeostasis
Other

Figure 2.3 The proteome of the pea aphid and bacteriocyte fractions**. A** 1D-SDS-PAGE gel lanes of the five different types of fractions (WB, BC, Bu1-, Bu-2, and BR) stained with Coomassie Brilliant Blue. Only one of the three independent biological replicates is shown here. **BC**, Total number of proteins (**B**) and total mass of proteins (**C**) based on NadjSPC of aphid origin (gray) and *Buchnera* origin (white). **D** Functional classification of quantified proteins, expressed as number of proteins (*top row*) and abundance, expressed as mean NSAF normalized to total aphid or *Buchnera* proteins (*bottom row*).

To allow quantitative analysis of the distribution of identified proteins within and across the different fractions, the spectral count (SPC) information was used. The spectral counting technique is based on the observation that the number of successful MS/MS acquisitions of peptides coming from a protein shows a positive and linear correlation to the relative concentration of this protein in the studied sample (Zybailov et al 2009, Lin et al 2004, Old et al 2005). Spectral counting is particularly effective to detect large quantitative differences (much better than stable isotope labeling techniques), as expected in the study when comparing cellular fractions that are very different in function and composition. Meaningful spectral counting of complex proteomes requires a mass spectrometer with a high duty cycle and benefits from high resolution and high accuracy instrument such as the LTQ-Orbitrap as it reduces the false positive rate of peptide identification (Scigelora and Makarov 2006, Hu et al 2005). The SPC workflow was previously developed and tested for *Arabidopsis* and maize organelles, cell-types, and complexes (Zybailov et al 2009, Olinares et al 2010, Majeran et al 2010). The relative amount (mass) of each identified protein within each replicate was calculated based on the adjusted number of matched MS/MS spectra (adjSPC), normalized by the sum of adjusted SPC in the replicate, yielding NadjSPC. For quantification purposes, proteins that shared more than ~80% of their matched MS/MS spectra were grouped using the workflow developed in (Friso et al 2010) (Figure 2.2). The relative concentration for each identified protein was calculated as the NSAF, which was calculated from adjSPC weighted for protein length (Zybailov et al 2009). In total 254 aphid proteins were assigned into 104 groups.

Whereas there has now been ample demonstration that label-free spectral counting is a viable method for protein quantification, it was important to test the variation among biological replicates for each sample type. Such variation could result from the fractionation and isolated

procedures and/or the complete proteomics workflow from SDS-PAGE, in gel digests and MS/MS itself. Therefore, the reproducibility among the biological replicates within each of the five sample types (WP, BC, Bu1, Bu2, and BR) was determined by Pearson's linear correlation analysis, summarized in Table 2.1. The pair-wise correlations were calculated for the proteins present in each replicate pair and showed that the replicates were highly similar (coefficients between 0.87 and 0.98). This shows that the proteomics workflow was robust and reproducible. The correlation coefficient for groups of proteins within specific expression intervals (one order of magnitude each) was also calculated. This confirmed the notion that quantification is more reliable for proteins with higher numbers of matched spectra. In particular, the correlation for proteins with NadjSPC >0.001 (*i.e.* those that represent more than 0.1% of the total protein mass) showed good reproducibility (Table 2.1). see Figure 2.2); the grouping will decrease variation and therefore the test measured the worst-case scenario.

Table 2.1. Pearson's Linear Correlation analysis for the NadjSPC values across the three biological replicates for each of the five sample types*

avg(NadjSPC)>0 (i.e. ALL DETECTED PROTEINS)			
Sample	**REP(1,2)**	**REP(2,3)**	**REP(1,3)**
WB	0.911046 (0.000000) [1456]	0.935535 (0.000000) [1577]	0.915923 (0.000000) [1755]
BC	0.973189 (0.000000) [1026]	0.975794 (0.000000) [930]	0.968234 (0.000000) [1137]
Bu1	0.942316 (0.000000) [737]	0.949920 (0.000000) [650]	0.966941 (0.000000) [785]
Bu2	0.950022 (0.000000) [519]	0.956288 (0.000000) [521]	0.973451 (0.000000) [553]
BR	0.907301 (0.000000) [1156]	0.946602 (0.000000) [1236]	0.870057 (0.000000) [1408]

avg(NadjSPC)>=0.01			
Sample	**REP(1,2)**	**REP(2,3)**	**REP(1,3)**
WB	0.901750 (0.000362) [10]	0.968338 (0.000004) [10]	0.927791 (0.000109) [10]
BC	0.994483 (0.000000) [11]	0.996094 (0.000000) [11]	0.995430 (0.000000) [11]
Bu1	0.989705 (0.000000) [15]	0.990646 (0.000000) [15]	0.991988 (0.000000) [15]
Bu2	0.982754 (0.000000) [18]	0.990855 (0.000000) [18]	0.984292 (0.000000) [18]
BR	0.933994 (0.000001) [14]	0.952317 (0.000000) [14]	0.863387 (0.000069) [14]

avg(NadjSPC)>=0.001 and <0.01			
Sample	**REP(1,2)**	**REP(2,3)**	**REP(1,3)**
WB	0.676583 (0.000000) [190]	0.807994 (0.000000) [190]	0.722070 (0.000000) [190]
BC	0.651988 (0.000000) [185]	0.693290 (0.000000) [185]	0.495089 (0.000000) [185]
Bu1	0.743508 (0.000000) [178]	0.521708 (0.000000) [178]	0.557247 (0.000000) [178]
Bu2	0.779909 (0.000000) [170]	0.675159 (0.000000) [170]	0.674097 (0.000000) [170]
BR	0.558248 (0.000000) [188]	0.835215 (0.000000) [188]	0.619625 (0.000000) [188]

avg(NadjSPC)>=0.0001 and <0.001			
Sample	**REP(1,2)**	**REP(2,3)**	**REP(1,3)**
WB	0.211049 (0.000000) [821]	0.312124 (0.000000) [811]	0.370317 (0.000000) [835]
BC	0.159327 (0.000240) [527]	0.315368 (0.000000) [473]	0.137764 (0.001402) [535]
Bu1	0.121923 (0.023725) [344]	0.178438 (0.001819) [303]	0.014824 (0.783194) [347]
Bu2	0.195202 (0.003078) [228]	0.123246 (0.068064) [220]	0.071228 (0.286327) [226]
BR	0.099402 (0.011413) [647]	0.388469 (0.000000) [614]	0.135713 (0.000445) [666]

avg(NadjSPC)>=0.00001 and <0.0001			
Sample	**REP(1,2)**	**REP(2,3)**	**REP(1,3)**
WB	-0.606859 (0.000000) [435]	-0.362424 (0.000000) [525]	-0.558647 (0.000000) [679]
BC	-0.499333 (0.000000) [301]	-0.392403 (0.000000) [227]	-0.495410 (0.000000) [374]
Bu1	-0.547358 (0.000000) [197]	-0.540131 (0.000000) [129]	-0.566530 (0.000000) [221]
Bu2	-0.679944 (0.000000) [103]	-0.350547 (0.000141) [113]	-0.447838 (0.000000) [139]
BR	-0.591225 (0.000000) [304]	-0.379592 (0.000000) [372]	-0.539781 (0.000000) [495]

*Within each comparison, 0-0 pairs were removed to avoid artificial increase of the correlation values. For each pair is listed: the correlation coefficient, the p-value in paranthesis, and the number of proteins in brackets

Figures 2.3B and C display the number of aphid and *Buchnera* proteins and the total adjSPC in each of the five sample types. The aphid proteins were substantially depleted (in number and mass) in the BC samples and even more in Bu-1 and Bu-2 samples, whereas *Buchnera* proteins were enriched in BC samples, relative to WB. The number and identity of *Buchnera* proteins detected were very similar among the three preparations dominated by *Buchnera*, *i.e.* BC, Bu-1, and Bu-2 (Figure 2.3A), suggesting that the proteome analysis had been successful in capturing the great majority of *Buchnera* proteins (see next section). This analysis shows that the fractionation procedure and proteome analysis were successful technically, providing a basis for the analysis of aphid and *Buchnera* proteins, and their distribution across the different samples.

2.3.3 The *Buchnera* Proteome

The most appropriate preparation for quantitative analysis of *Buchnera* proteins within the *Buchnera* cells was the most highly purified *Buchnera* preparation, Bu-2. The relative concentration (NSAF) of the 388 *Buchnera* proteins in the Bu-2 samples spanned nearly four orders of magnitude from 0.00015% to 8.7%). All identified *Buchera* proteins were assigned to one of seven functional groups. The distribution of the proteome across these seven groups Fig. 3D), showed that primary metabolism and information transfer (DNA, RNA and protein transformations) accounted together for 76% of the number of identified proteins and 65% of the total protein concentration.

The most abundant *Buchnera* protein was GroEL (NP_239860.1), a molecular chaperone previously demonstrated as an abundant protein of *Buchnera* (Sato and Ishikawa 1997). GroEL accounted for 8.7% of the protein concentration in the Bu-2 samples; and it was the most abundant protein in every other sample type, with NSAF values of 2.9% for WB, 10% for BC,

and 4.9% for BR. GroES (NP_239859.1), the co-protein of this chaperone, was also abundant in Bu-2 samples (ranked #12 in relative concentration, 1.1% NSAF). This first quantitative proteome study of the aphid-*Buchnera* symbiosis offers the strongest confirmation of many studies over the last 25 years that GroEL is the dominant *Buchnera* protein (Baumann et al 1996, Humphreys and Douglas 1997, Ishikawa et al 1985). GroESL is believed to play a crucial role in the stabilization of the proteins coded by the many *Buchnera* genes that have accumulated deleterious mutations as a result of repeated population bottlenecks at vertical transmission (Fares et al 2004). Other high-ranking chaperones detected were the Hsp70 family member DnaK (NP_239985.1, #5, 1.8%), and HtpG (NP_240294.1, #6, 1.5%) (in parentheses are provided the gene name, and the rank number and %NSAF in Bu2). The second most abundant *Buchnera* protein in Bu-2 samples was AhpC (NP_240013.1), alkyl hydroperoxide reductase (4.3%), with a predicted role in eliminating hydrogen peroxide. Overall, the *Buchnera* proteins classified to a protective role, both against protein denaturation and against environmental stress (Multifun categories 2, 3, 4, and 5) accounted for 20% NSAF (Figure 3D).

It is well established that the core function of *Buchnera* in the aphid symbiosis is the synthesis of EAAs, which are supplied to the aphid host (Douglas 1998). All enzymes in amino acid metabolism annotated in the *Buchnera* genome were detected in the proteome. The relative concentration of proteins contributing to EAA synthesis ranged from 0.002% (ArgE, NP_239884.1) to 1.28% (IlvC, NP_240398.1). The median NSAF for proteins varied significantly among the various EAA biosynthetic pathways (Kruskal Wallis test: $H_8 = 16.73$, $p < 0.05$), but the significance was lost when the arginine biosynthetic pathway was omitted from the data (Kruskal Wallis test: $H_7 = 12.69$, $p > 0.05$). The cumulative NSAF for EAA synthesis proteins was 16%. This value represented nearly half of the total NSAF (35%) in Bu-2 for

proteins with metabolic function. Proteins involved in DNA, RNA and protein transformations are classified separately from the metabolic enzymes and collectively represent NSAF of 30% (Figure 3D).

Buchnera has the genetic capacity to synthesize the structural components of the flagellum, except for flagellin, FliC, and flagellar motor proteins; and the flagellar bases are uniformly distributed over the outer surface of *Buchnera* cells (Maezawa et al 2006). Twenty flagellar proteins could be quantified, including the nine proteins reported previously (Maezawa et al 2006), collectively accounting for 3.5% of NSAF (Figure 2.3D).

Buchnera has 14 genes with predicted capacity to transport small molecules across its inner membrane. Six transporter proteins were detected: two channels, OmpF (NP_240177.1) at high relative concentration (#4, 1.95% NSAF) and MscS (NP_240264.2, #394 0.017%), but not the third predicted channel protein, the glyceroaquaporin GlpF (NP_240129.1); both of the phosphotransferase systems PtsG (NP_240174.2, #279, 0.05%) and MtlA (NP_240376.1, #346, 0.029%); the phosphate transporter PitA (NP_240390.1, #543, 0.0022%); and one subunit, MdlB (NP_240291.1, #558, 0.0013%), of the ABC transporter MdlAB. The *Buchnera* genes *secABEGY* code for the Sec protein export system of *Buchnera*, of which SecAEGY were detected in Bu-2 samples, and SecB in Bu-1 (but not Bu-2) samples. The *Buchnera* proteins with transport function in Bu-2 samples represented 5% of the total protein concentration.

There are 14 predicted pseudogenes in the *Buchnera* genome, but the possibility that at least some may code for proteins is raised by the evidence for transcriptional slippage in Buchnera (Tamas et al 2008). One of the 14 pseudogenes with transcriptional slippage was detected in the proteome: DdlB (d-alanine-d-alanine ligase B (BB0079). Insertion of an extra A by transcriptional slippage would mediate amino acid sequence switch from frame-1 to frame-3

and the reconstructed DdlB protein (307 amino acids) includes the predicted site of transcriptional slippage and the residues contributing to the active site, E15, S150, and S281 (Figure 2.4). The proteome analysis detected most of the amino acid residues of this 307 aa protein, including protein sequence for both frame-1 and frame -3 (Figure 2.4). However, I cannot exclude that also a truncated protein accumulated resulted from just frame 1 translation and full functionality of the stop codon.

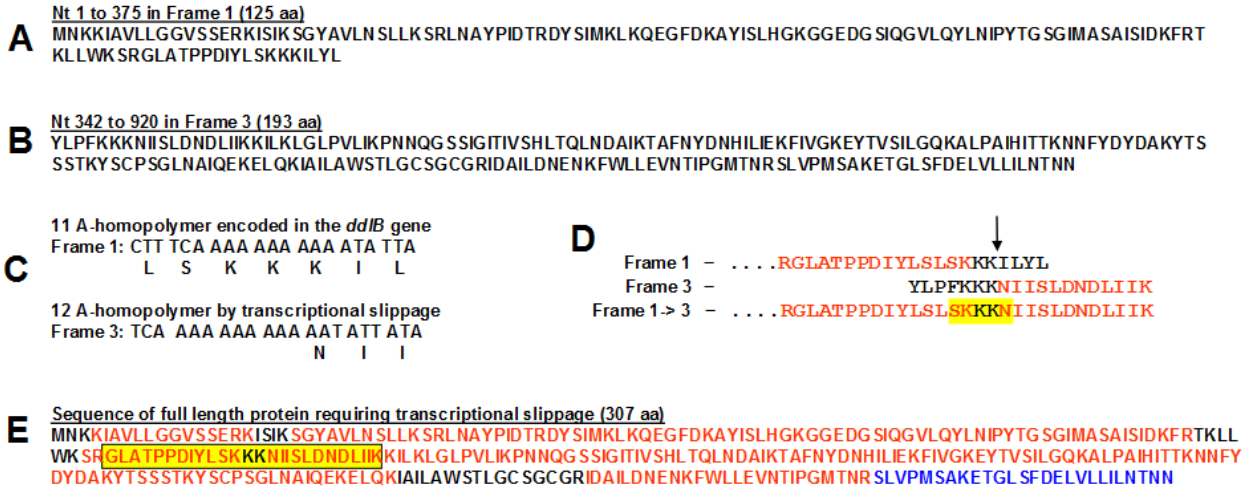


Figure 2.4 Proteome analysis of the Buchnera protein DdlB. The Buchnera ddlB gene has two long reading frames. (A) Amino acid sequence in frame-1 (stop codon at nt 376-379), and (B) Amino acid sequence in frame-3 (stop codon at nt 921-923). (C) Insertion of an extra A at a string of 11 As (nt 354-364) by transcriptional slippage* can be invoked to mediate amino acid sequence switch from frame-1 to frame-3. (D) The C-terminal residues of the protein generated from frame 1, the N-terminal residues generated from frame 3 and the aa sequence region of combined frame 1 and frame 3 if transcriptional slippage had occurred, including the minimal diagnostic sequence SKKKN (yellow). The arrow indicates the switchfrom frame 1 to 3. (E) The reconstructed DdlB protein (307 amino acids) including the predicted site of transcriptional slippage and the residues contributing to the active site, E15, S150,S281. The proteome analysis detected most of the amino acid residues of this 307 aa protein, but not the tryptic peptide (with 3 missed cleavages; in yellow) that included the minimal diagnostic sequence SKKKN.

2.3.4 The Aphid Proteome

A total of 1943 individual aphid proteins were detected, representing the largest aphid proteome set to date, and assigned to one out of seven functional groups (Figure 2.3D). Of these, 1762 (90%) were present in the WB samples and 1414 (73%) were detected in one or more of the bacteriocyte samples (*i.e.* BC, BR, Bu-1, Bu-2). The very abundant proteins were generally common across all samples. For example, the ten most abundant aphid proteins in the WB samples were mostly ranked in the top 30 aphid proteins in the BR and BC samples, and *vice versa*. Among the four sample types derived from dissected bacteriocytes (Figure 2.1), the greatest number of aphid proteins (1,171) were detected in the BR samples (Figure 2.3B). The aphid proteins detected in bacteriocyte samples other than BR were generally of low abundance (median 0.0003% NSAF). A detailed functional analysis of the proteome in the bacteriocyte and whole aphid was conducted on the WB and BR samples. The most strongly represented functional class in both WB and BR was metabolism, whether expressed per number of proteins or protein mass (Figure 2.3D) The representation of functional classes was similar in WB and BR samples on the protein number basis, but metabolism was relatively over-represented and structural proteins under-represented in BR on the quantitative (NSAF) basis.

2.3.5 Cluster Analysis of Protein Partitioning Between Aphid and Buchnera Compartments

To investigate the partitioning of host and symbiont proteins to the bacteriocyte and *Buchnera* fractions, the proteins with abundance >0.001 NadjSPC were subjected to hierarchical clustering, generating the dendrogram shown in Figure 2.5. This cutoff was based on the correlation analysis presented above (Table 2.1) and also after testing clustering with different cutoffs. Two main clusters were detected, with cluster-A containing proteins that are depleted in

Bu-1 and Bu-2 samples, and cluster-B comprising proteins enriched in these samples. Cluster-A could be divided into several subclusters, including subcluster A2 containing proteins enriched in bacteriocytes, and subcluster A1b-1 comprising proteins depleted in bacteriocytes (Figure 2.5).

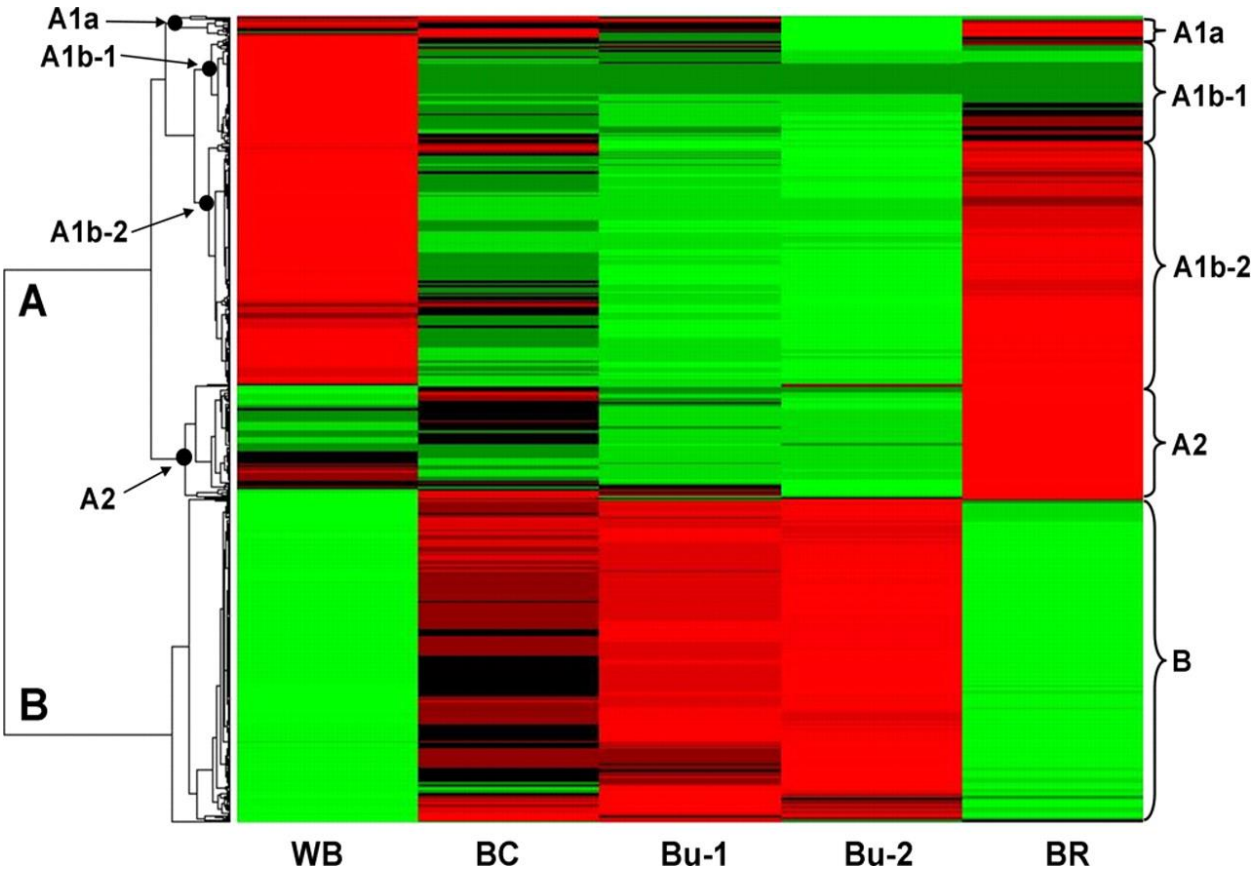


Figure 2.5: Hierarchical cluster analysis of protein distribution across the five fractions**.** The abundance is indicated by the colors, expression increased from green to black to red. There are two main clusters (A and B), with cluster A divided in four subclusters. None of the *Buchnera* proteins were found in cluster A.

The contribution of aphid and *Buchnera* proteins to the different clusters was analyzed. All the *Buchnera* proteins were restricted to cluster B. In other words, no evidence for the selective export of proteins from *Buchnera* cells into aphid cells was found. Even so, *Buchnera* proteins were detected in the BR samples (which lacked *Buchnera* cells, as determined by light

microscopical analysis). Consistent with the interpretation that these proteins were derived from *Buchnera* cell breakage during sample preparation (and not selective secretion), the abundance of *Buchnera* proteins present in both the Bu-2 and BR samples was significantly correlated (Spearman's Rank Test: $r_s = 0.536$, $p < 0.001$).

Only four aphid proteins were allocated to cluster B representing proteins highly enriched in the purified *Buchnera* cells indicating that the great majority of aphid proteins were not associated selectively with *Buchnera* cells, as expected. Thus except for possibly these four proteins, no aphid proteins were detected that appeared synthesized specifically for sorting into the *Buchnera* cells. These four proteins are: ACYPI001025 (mitochondrial α-ketoglutarate/malate carrier protein), ACYPI002559 (monocarboxylate transporter), ACYPI0004040 (a conserved hypothetical protein and ortholog, by reciprocal top BLAST hit, to *Drosophila* CG2206), and ACYPI004434 (elongation factor 1α). Careful manual evaluation of the distribution and abundance, as well as the experimental variation (*i.e.* CVs of NadjSPC) of these four proteins across the four fractions, did not show robust evidence that these four proteins were indeed specifically enriched in *Buchnera* cells. Rather these four proteins were contaminants in BR samples. This proteome analysis did not support the hypothesis that a subset of aphid proteins is specifically synthesized to carry out functions within the *Buchnera* cells.

2.3.6 Aphid Proteins Enriched in Bacteriocyte Samples

Of the 84 aphid proteins in subcluster-A2, enriched in bacteriocyte fractions (Figure 2.5), 60 (71%) had an annotated function in metabolism. These proteins and the five enriched transporters are considered below. Among the remaining proteins in this category, three classes were of particular interest. First were 8 of the 13 annotated subunits of vacuolar-type H^+-ATPase (V-ATPase), including ACYPI002584 and ACYPI010077, ranked #4 and #9 and with NSAF

1.33% and 1.11%, respectively. Second were the two subunits of phenoloxidase, ACYPI001367 (ranked #115th NSAF 0.028%) and ACYPI004484 (#650, 0.016%), the sole proteins among the pea aphid gene products with an annotated defensive function (Gerardo et al 2010) enriched in bacteriocytes. Finally, one protein, ACYPI005979 (#277, 0.087%) was one of the five rare lipoproteins (Rlp A_5) that has been acquired by lateral gene transfer from a bacterial donor (not *Buchnera*) and is known to be transcribed in aphid bacteriocytes (Nikoh et al 2010).

2.3.7 Aphid Enzymes in Amino Acid Metabolism Enriched in Bacteriocytes

Many of the proteins with metabolic function enriched in bacteriocyte fractions (subcluster-A2) were related to amino acid transformations, and their enrichment is predicted to be linked to the biosynthesis of amino acids by *Buchnera*. Metabolic network analysis (Thomas et al 2009) has revealed that *Buchnera* requires an input of four amino acids to support its amino acid biosynthetic function: glutamate, glutamine, aspartate, and serine. The enzymes mediating the biosynthesis of these amino acids (and no other amino acids) were all enriched in bacteriocytes.

Glutamine synthetase (GS: ACYPI001461,#148, 0.16%) and glutamate synthetase (GOGAT: ACYPI000303, #180, 0.13%) were strongly represented (in parentheses are provided the gene name, and the rank number and mean %NSAF normalized to total mean aphid proteins in BR samples). GS/GOGAT mediates the synthesis of glutamate and, where ammonia is not limiting and the demand for glutamate is not high, net glutamine is also generated (Figure 2.6A). A candidate asparaginase, which mediates the hydrolysis of asparagine to aspartate, was also enriched (ACYPI005215, #14, 0.88%) (Figure 2.6B). Finally, all three enzymes in serine biosynthesis from 3-phospho-d-glycerate were represented in subcluster A2: d-3-phosphoglycerate dehydrogenase (ACYPI006664, rank #10, 1.04%), phosphoserine

aminotransferase (ACYPI004666, #64, 0.3%), and phosphoserine phosphatase (ACYPI000304, #83, 0.24%) (Figure 2.6*C*).

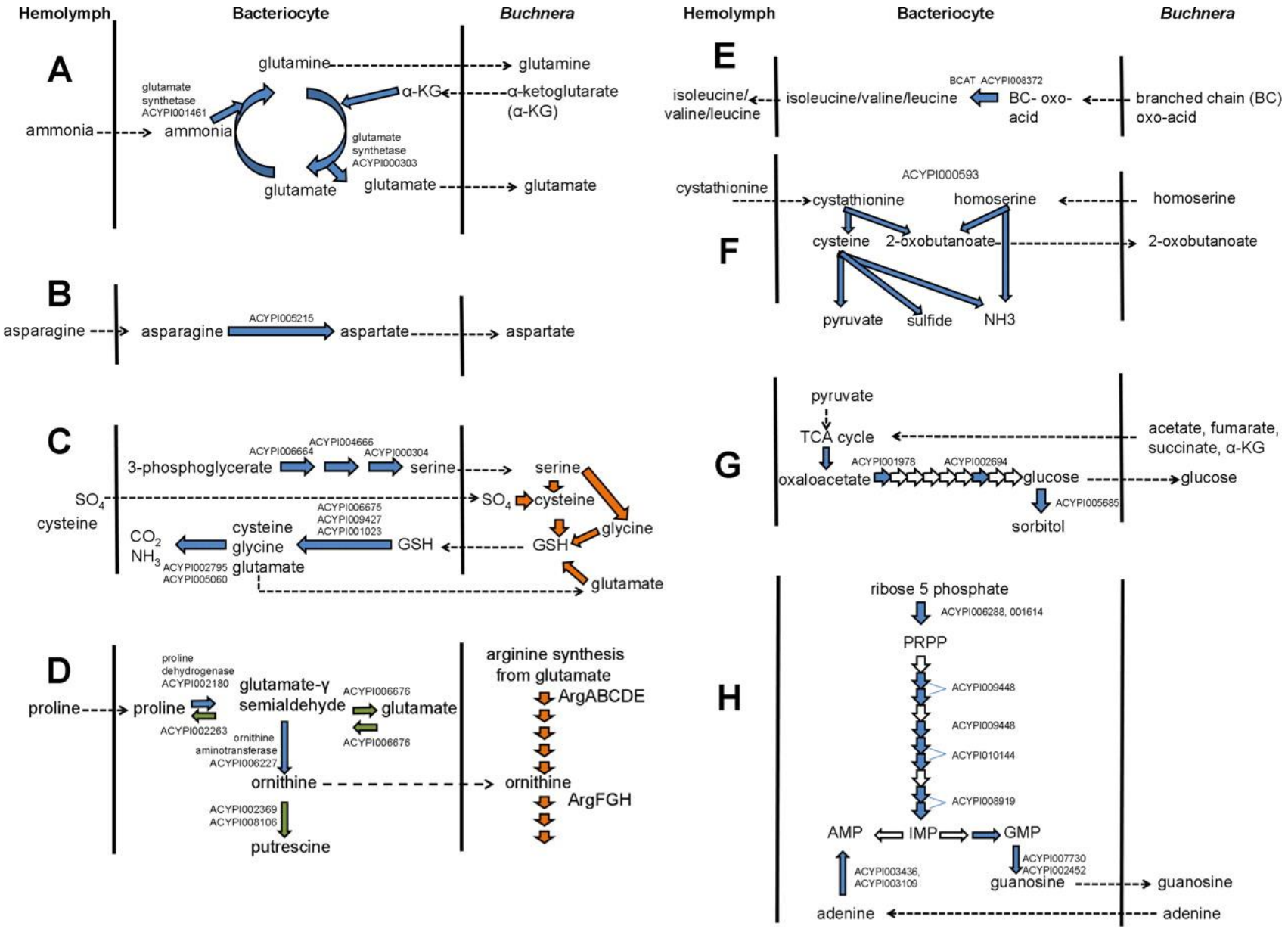

Hemolymph
Bacteriocyte
Buchnera
A
glutamine
glutamine
glutamate synthetase ACYPI001461
α-KG
α-ketoglutarate (α-KG)
ammonia
ammonia
glutamate synthetase ACYPI000303
glutamate
glutamate
glutamate
B
asparagine
asparagine
ACYPI005215
aspartate
aspartate
C
ACYPI004666
ACYPI006664
ACYPI000304
3-phosphoglycerate
serine
serine
SO4
SO4
cysteine
cysteine
ACYPI006675
ACYPI009427
ACYPI001023
CO2
NH3
cysteine
glycine
GSH
GSH
glycine
ACYPI002795
ACYPI005060
glutamate
glutamate
D
proline dehydrogenase ACYPI002180
glutamate-γ semialdehyde
ACYPI006676
proline
proline
glutamate
ACYPI002263
ACYPI006676
ornithine aminotransferase ACYPI006227
ornithine
ornithine
ACYPI002369
ACYPI008106
putrescine
arginine synthesis from glutamate
ArgABCDE
ArgFGH
Hemolymph
Bacteriocyte
Buchnera
E
isoleucine/ valine/leucine
isoleucine/valine/leucine
BCAT ACYPI008372
BC- oxo-acid
branched chain (BC) oxo-acid
F
ACYPI000593
cystathionine
cystathionine
homoserine
homoserine
cysteine
2-oxobutanoate
2-oxobutanoate
pyruvate
sulfide
NH3
G
pyruvate
TCA cycle
acetate, fumarate, succinate, α-KG
ACYPI001978
ACYPI002694
oxaloacetate
glucose
glucose
ACYPI005685
sorbitol
H
ribose 5 phosphate
ACYPI006288, 001614
PRPP
ACYPI009448
ACYPI009448
ACYPI010144
ACYPI008919
AMP
IMP
GMP
ACYPI007730
ACYPI002452
ACYPI003436, ACYPI003109
guanosine
guanosine
adenine
adenine

Fig. 2.6 Reactions and transport of metabolites across the aphid-*Buchnera* interface, deduced from the proteome of bacteriocytes**.** *Blue arrows*: reactions mediated by aphid proteins enriched in bacteriocytes (subcluster A2); *green arrows*: reactions mediated by aphid proteins depleted (subcluster A1b-1) or undetectable in bacteriocytes; *white arrows*: reactions mediated by aphid proteins that are not differentially expressed between whole body and bacteriocyte samples; *orange arrows*: reactions mediated by *Buchnera* proteins (cluster B); *dotted arrows*: transport reactions. **A** Synthesis of glutamate and glutamine. **B** Synthesis of aspartate. **C** Synthesis of serine and reductive assimilation of inorganic sulfate into cysteine. **D** Synthesis of ornithine from proline. **E** synthesis of branched chain amino acids. **F** Reactions mediated by cystathionine-γ-lyase (ACYPI000593). **G** gluconeogenesis. **H** Purine synthesis and recycling.

Subcluster A2, comprising aphid proteins enriched in the bacteriocytes, also included enzymes in the degradation of the tripeptide glutathione (γ-l-glutamyl-l-cysteinylglycine) to its constituent amino acids: specifically three isoforms of the peptidase PepN (ACYPI006675, #96, 0.22%; ACYPI009427, #156, 0.15%; ACYPI001203, #360, 0.065%) and 5-oxoprolinase (ACYPI004211, #405, 0.057%). Components of the glycine cleavage system, generating CO_2 and ammonia were also enriched (ACYPI002795, #116, 0.19%; ACYPI005060, #175, 0.14%). Taken together with the presence of glutathionine synthase and enzymes for utilization of host-derived serine in cysteine and glycine synthesis in the *Buchnera* proteome, these data raise the possibility that *Buchnera* mediates the reductive assimilation of inorganic sulfate into cysteine, which is then delivered back to the aphid in the form of glutathione; with subsequent PepN-dependent recovery of cysteine from the glutathione, together with the degradation of glycine and recycling of glutamate. This putative cycle is illustrated in Figure 2.6C.

The most abundant protein in the bacteriocyte proteome was ornithine aminotransferase (OAT) (ACYPI009480, #2, 1.67%), which mediates the reversible interconversion of ornithine and glutamate-γ-semialdehyde. In many animals, OAT is responsible for ornithine degradation, a key step in the degradation of arginine to glutamate, but the aphid apparently lacks an arginase (IAGC 2010), which mediates the first step in arginine degradation, and a catabolic role of OAT is, therefore, unlikely. Other data are consistent with the alternative interpretation that OAT mediates ornithine synthesis, specifically from proline (Figure 2.6D). Specifically, the bacteriocytes were enriched for proline dehydrogenase (ACYPI002180, #221, 0.11%), the first step in the proline degradation pathway, yielding glutamate-γ-semialdehyde, the substrate for OAT; but the protein catalyzing the degradation of glutamate-γ-semialdehyde to glutamate

(ACYPI006676) and reverse reaction generating glutamate-γ-semialdehyde from glutamate (ACYPI000665) were under-represented in the bacteriocyte samples. The metabolic fate of ornithine may include the provision of ornithine to *Buchnera*, supplementing the endogenous supply of this substrate for arginine synthesis. Importantly, ornithine decarboxylase (ACYPI002369 and ACYPI008106) which mediates the synthesis of putrescine, the substrate for polyamine synthesis by both aphid and *Buchnera*, was not detected. A role in arginine synthesis is further indicated by the significantly greater abundance of *Buchnera* proteins mediating the synthesis of arginine from ornithine than those mediating the synthesis of ornithine from glutamate ($t_6 = 3.44$, $p < 0.05$) (Figure 2.7).

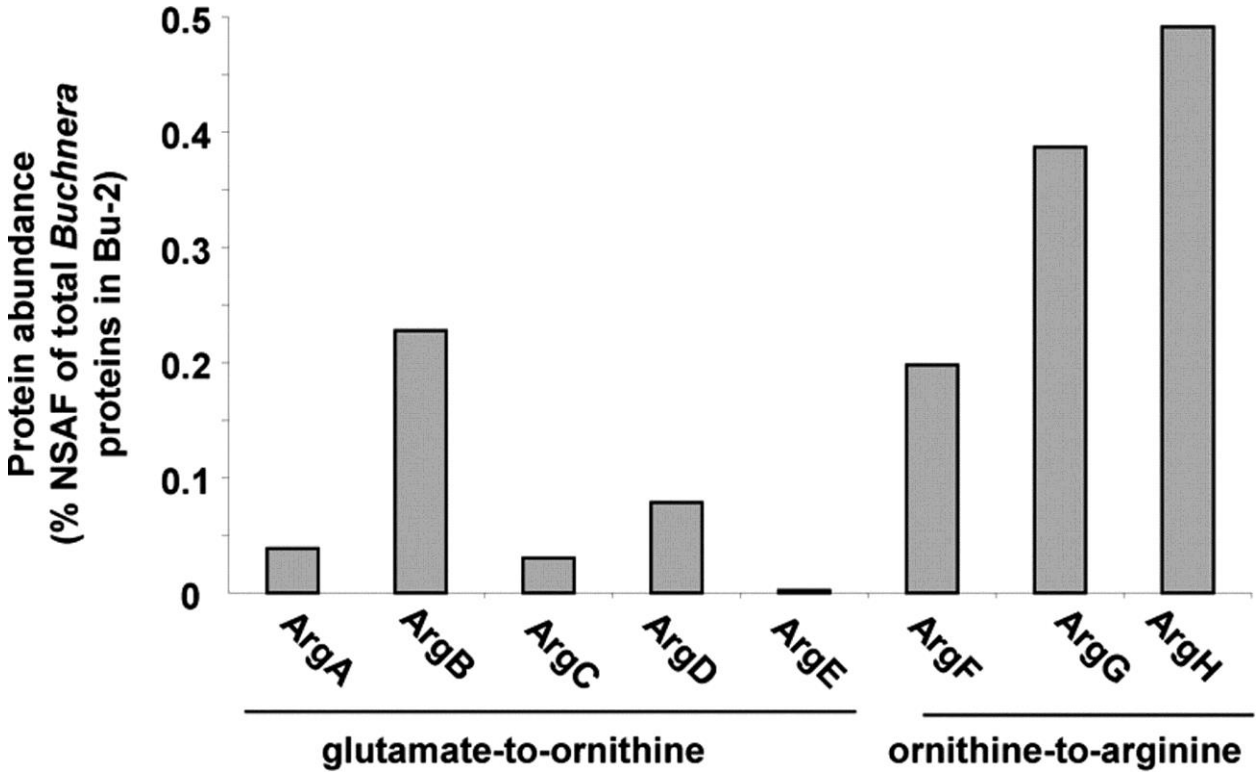


Fig. 2.7 Abundance of individual *Buchnera* enzymes involved in the synthesis of arginine within the purified *Buchnera* cells.

Branched chain amino acid aminotransferase (BCAT) (ACYPI008372, #110, 0.2%) was enriched (Fig. 2.6E). BCAT in animals generally serves as the first step in the degradation of branched chain amino acids (BCAs, comprising leucine, isoleucine and valine), but it has been suggested that, in the pea aphid, this enzyme may mediate the final step in BCA synthesis, compensating for the absence of the equivalent gene, *ilvE*, in *Buchnera* (Wilson et al 2010). Other enzymes in BCA degradation were undetectable or depleted in bacteriocytes, including ACYPI002372 (acyl CoA dehydrogenase) mediating the equivalent intermediate step in isoleucine and valine degradation (from S-2-methyl-butyryl-CoA to tiglyl-CoA, and from isobutyryl-CoA to methylacrylyl-CoA, respectively), and ACYPI002843 (acetyl-CoA acetyltransferase), which catalyzes the final step in isoleucine degradation to acetyl CoA (ACYPI002843 also mediates the final step in the degradation of a second amino acid, lysine, which is not a BCA). These data support the hypothesis that BCAT has a predominantly biosynthetic function in the bacteriocyte.

Buchnera additionally lacks the gene for a second enzyme in the biosynthesis of one BCA, isoleucine: threonine dehydratase (*ilvA*), which mediates the production of the 2-oxobutanoate from threonine. It has been proposed that this reaction is mediated by the aphid threonine dehydratase (ACYPI006784), with the transfer of threonine from *Buchnera* to aphid, and of 2-oxobutanoate in the reverse direction (Shigenobu et al 2000). Contrary to this hypothesis, ACYPI006784 was not detected in the aphid proteome. A candidate alternative source of 2-oxobutanoate is cystathionine-γ-lyase (CGL) (ACYPI000593, #71, 0.27%), which is included in subcluster A2. Of the three reactions that CGL is annotated to mediate (Figure 2.6F), two yield 2-oxobutanoate: from the substrates cystathionine and homoserine. Cystathionine could be derived from hemolymph (various insects have appreciable hemolymph cystathionine

levels, although this has not been investigated in aphids (Wyatt 1961, Terra et al 1973). Cystathionine is an intermediate in methionine degradation, but this pathway is apparently not functional in bacteriocytes because all aphid proteins annotated to this pathway were undetected in BC, Bu-1, Bu-2, and BR samples, apart from ACYPI004816 (present at low abundance in BR samples). The alternative substrate yielding 2-oxobutanoate, homoserine, is synthesized by *Buchnera*. CGL activity is also potentially a source of ammonia substrate for GS/GOGAT (Figure 2.6A) and pyruvate for gluconeogenesis (see below).

Two enzymes in synthesis of the aromatic amino acid tyrosine are also assigned to subcluster A2: phenylalanine hydroxylase (ACYPI007803, #18, 0. 74%) mediating the conversion of phenylalanine to tyrosine, and dihydropteridine reductase (ACYPI006909, #33, 0.43%), in the recycling of tetrahydrobiopterin, the reductant in the phenylalanine hydroxylase reaction. These results confirm the evidence from the bacteriocyte transcriptome (Nakabachi et al 2005) that the bacteriocyte is an important source of tyrosine.

2.3.8 Aphid Enzymes in Gluconeogenesis Enriched in Bacteriocytes

Both of the two rate-limiting steps in gluconeogenesis were included in subcluster A2: phosphoenolpyruvate carboxykinase (ACYPI001978, #203, 0.12%) and fructose-1,6-bisphosphatase (ACYPI002694, #121, 0.18%). The glucose generated by this route may be translocated to *Buchnera*, for which glucose is the major carbon source (Thomas et al 2009). Glucose might also be used in sorbitol synthesis, as indicated by the enrichment of aldose reductase (ACYPI005685, #153, 0.16%), which mediates the NADPH-dependent reduction of glucose to sorbitol (Figure 2.6G). Trehalose-6-phosphate synthase (ACYPI006164) was assigned to subcluster A1b-1, depleted in bacteriocytes, indicating that the major hemolymph sugar trehalose is not a major fate of glucose synthesized in the bacteriocyte. The noncarbohydrate

carbon substrates for gluconeogenesis may include acetate, fumarate, succinate, and α-ketoglutarate, all predicted net products of *Buchnera* metabolism (Thomas et al 2009) fed into the TCA cycle and converted to oxaloacetate; and the degradation of glycine, which is a glucogenic amino acid (*i.e.* can be converted to glucose via gluconeogenesis). Glycerol, derived from the degradation of triglycerides, is unlikely to contribute to gluconeogenesis because four proteins with a role in β-oxidation of fatty acids were in subcluster A1b-1, depleted in bacteriocyte samples: ACYPI001113 and ACYPI003232 (acyl-CoA dehydrogenases), ACYPI006589-PA (3-hydroxyacyl-CoA dehydrogenase), and ACYPI008366-PA (acetyl-CoA acyltransferase). One of these proteins, ACYPI006589, is also annotated to mediate a key step in the degradation of the glucogenic branched chain amino acid valine. This and the data above suggest that amino acids are also not major substrates for gluconeogenesis in bacteriocytes.

2.3.9 Aphid Enzymes in Purine Metabolism Enriched in Bacteriocytes

Two of the three isozymes of ribose phosphate pyrophosphokinase, which mediates the synthesis of 5-phosphoribosyl-1-pyrophosphate (PRPP) from ribose-5-phosphate, were assigned to subcluster A2 (ACYPI006288, #79, 0.25%, and ACYPI001614, #162, 0.15%); the third isozyme was detectable but considerably less abundant (ACYPI006561, #444, 0.051%). PRPP contributes to the synthesis of purines, pyrimidines and NAD(P). Of these pathways, purine synthesis is probably particularly important in bacteriocytes (Figure 6H) because three of six annotated aphid proteins mediating the synthesis of inosine monophosphate (IMP) from PRPP were also in subcluster A2 (ACYPI009448, #81, 0. 24%; ACYPI010114, #36, 0.42; and ACYPI008919, #88, 0.22%). Of the remaining proteins in this pathway, amidophosphoribosyltransferase (ACYPI001885, #711, 0.02%) and one of the two isoforms of adenylosuccinate lyase (ACYPI001812, #399, 0.058%) were readily detectable in the

bacteriocyte proteome (the second isozyme of adenylosuccinate lyase, ACYPI007825, was detectable, but was below the threshold for cluster analysis). The other purine metabolism enzymes enriched in bacteriocyte samples relate to the predicted purine metabolite exchange with *Buchnera* (Homma et al 2001): GMP synthase (ACYPI006177, #278, 0.087%) and two 5′nucleotidases (ACYPI007730, #41, 0.40%; ACYPI002452, #454, 0.049%), for the production of guanosine, delivered to *Buchnera*, and both of the isozymes of adenine phosphoribosyltransferase, which metabolize adenine, received from *Buchnera*, to AMP (ACYPI003436, #24, 0.53%; ACYPI003109, #52, 0.35%) (Figure 6H). A further protein enriched in bacteriocyte samples was ACYPI001365 (#214, 0.11%). Although annotated as "insect-derived growth factor," the homolog of this protein has been reported to have adenosine deaminase activity in the fleshfly *Sarcophaga peregrine* (Homma et al 2001), raising the possibility that this protein mediates the transformation of adenosine to inosine.

2.3.10 Aphid Transporters Enriched in Bacteriocytes

Five proteins with transporter function were enriched in bacteriocytes. They are annotated as a candidate Na^+-dependent phosphate transporter (ACYPI003986, #172, 0.14%), a low affinity cationic amino acid transporter (ACYPI008904, #187, 0.13%), a K^+/Cl^- symporter (ACYPI000507, #281, 0.086%), an ABC transporter (ACYPI009528, #578, 0.031%), and the mitochondrial ATP/ADP translocase (ACYPI000784, #31, 0.44%). ACYPI003986 (phosphate transporter) is one of two putative inorganic ion transporters detected in the bacteriocyte proteome, the other being ACYPI001177 (#926, 0.01%), annotated as a sulfate transporter; and these two transporter may contribute to the delivery of phosphate and sulfate to the *Buchnera* cells. ACYPI008904 (cationic amino acid transporter) may mediate the transport of the EAAs arginine and lysine from *Buchnera* cells to the surrounding bacteriocyte cytoplasm. The enriched

ABC transporter is the ortholog of the human ABCA3 (NP_001080.1, member of the ABC1 subfamily), which is annotated as implicated in lipid (including cholesterol) transport and programmed cell death.

The proteome of BR samples also included 18 further aphid transporters, some of which are candidates for mediating the transfer of metabolites between the host and *Buchnera* symbiont. These transporters include two further amino acid transporters, ACYPI000536 (#412, 0.051%), the ortholog of the *Drosophila* protein CG8785 and ACYPI008971 (#599, 0.029%), mediating H^+ coupled amino acid transport; and a sugar transporter ACYPI001077 (#613, 0.028%), designated ApST9 (Price et al 2010). Four further members of the MFS superfamily identified in BR samples are probably not sugar-transporters (ACYPI002278_ApST38, ACYPI010074_ApST42, ACYPI007742_ApST45, and ACYPI009892, which has no ApST designation) (Price et al 2010). The bacteriocyte samples had two monocarboxylate transporters (ACYPI002559, #692, 0.021%, and ACYPI002787, #669, 0.023%), the candidate substrates for which could include glutamate and aspartate transferred from aphid to *Buchnera* (Figure 2.6AB), branched-chain oxoacid precursors for BCA synthesis by the aphid BCAT (Figure 2.6E), 2-oxobutanoate provided to *Buchnera* for isoleucine synthesis (Figure 2.6F), and *Buchnera*-derived organic acids delivered to the bacteriocyte as substrates for gluconeogenesis (Figure 2.6G). Three organic cation transporters (ACYPI007237, #949, 0.0095%; ACYPI007903, #1032, 0.0073%; ACYPI005500, #1204, 0.0037%) were detected. The organic cations in bacteriocytes likely include polyamines, as well as cationic amino acids considered above. Five ABC transporters (additional to ACYPI009528 enriched in bacteriocyte samples considered above): two members of the MDR/TAP family (ACYPI003751 and ACYPI008583), and one member of the GCN20, White and ABC1 sub families (ACYPI009003, ACYPI007666, and ACYPI001137,

respectively). Finally, a nucleoside transporter (ACYPI006094) which mediates the movement of nucleosides across membranes down their concentration gradient, could be responsible for the guanosine transport from bacteriocyte to *Buchnera* as a candidate substrate Fig. 2.6H).

2.3.11 Aphid Proteins Depleted in Bacteriocytes

A total of 84 proteins were assigned to subcluster A1b-1, representing the proteins depleted in bacteriocyte samples relative to the whole body (Figure 2.5). These included proteins involved in protein synthesis (*e.g.* ribosomal proteins), associated with endoplasmic reticulum function (*e.g.* Ca^{2+}-transporting ATPase, ACYPI008535; perlecan, ACYPI010019; and perlecan, ACYPI010019) and both cytoskeletal and muscle proteins (*e.g.* β-tubulin, ACYPI008874, which is the most abundant aphid protein in the whole aphid samples; myosin proteins ACYPI008178, ACYPI000027; ACYPI006043; α-actinin, ACYPI000667). These data reflect the specialization of the bacteriocyte to house the *Buchnera* cells, which occupy >60% of the cytoplasmic volume (Whitehead and Douglas 1993), in contrast to, for example, myosin-rich muscle, ER-rich fat body cells, and the embryos with high protein synthesis rates that contribute to the whole body samples. The depletion of proteins involved in protein synthesis and ER-processing provides information on the fate of EAAs released from *Buchnera* cells, specifically that the EAAs received by the bacteriocyte are not incorporated substantially into protein before export to the rest of the aphid body. Just 14 (17%) of the proteins in cluster A1b-1 had a metabolic function.

2.4 Discussion

2.4.1 Absence of Selective Protein Transfer Between Aphid and Buchnera

A key purpose of this study was to investigate the partitioning of host and symbiont proteins across aphid tissues and *Buchnera* cells, in order to test whether specific proteins are transferred selectively between the two partners. In principle, proteins transferred in either

direction could be effector molecules regulating gene expression, metabolism, growth rates, and other biological functions that regulate or support the symbiosis. Such interactions are central to both bacterial pathogenesis (Nakabachi et al 2005) and the function of various symbioses (Hueck 1998, Dale et al 2002).

No evidence for the selective transfer of proteins among the partners was obtained from the cluster analysis, which revealed near-perfect separation of the aphid and *Buchnera* proteins across the cellular localizations. This result has major implications for an understanding of *Buchnera* function. In the absence of the selective import of proteins from the host (as occurs on organelles), the gene content of *Buchnera* encompasses the functional capability of this bacterium, thereby validating previous analyses of *Buchnera* function based on the *Buchnera* gene content alone (Shigenobu et al 2000, Thomas et al 2009, Vinuelas et al 2007). A further implication is that *Buchnera* differs from mitochondria and plastids (in plants and green algae), the proteomes of which are much greater than their respective genomes as a result of the specific targeting of many proteins synthesized by the nucleocytoplasm. Taken together with the evidence that functional *Buchnera* genes have not been translocated to the aphid genome (Nikoh et al 2010), these data strongly suggest that *Buchnera* has greater genetic independence than organelles. Furthermore, if this trait is general among the bacteriocyte symbionts of insects, these data offer some assurance that the widespread use of bacterial gene content to describe the function of these bacteria *e.g.* (McCutcheon and Moran 2007, Perez-Brocal 2006, Sabree et al 2009) is biologically valid.

The proteome analysis additionally offers no evidence for the between-partner transfer of effector proteins that function in partner communication in this symbiosis. This interpretation should, however, be considered as tentative, because (i) effector proteins that are present in both

host and symbiont may escape detection because they do not necessarily show a strong enrichment pattern, and (ii) only those 637 proteins/protein groups that were expressed above the set threshold level were part of the quantitative cluster analysis. Proteins below this expression threshold, or even proteins that were not detected, may include such effectors.

2.4.2 Metabolite Exchange Between Bacteriocytes and Buchnera

Biosynthesis and metabolite exchange are central to the function of the bacteriocyte and its complement of *Buchnera* cells. The growth and survival of the aphid depend on the sustained synthesis of EAAs by *Buchnera*; and because *Buchnera* is intracellular, all its metabolic requirements for growth and EAA production are derived from the surrounding bacteriocyte cytoplasm, and ultimately from the hemolymph (insect blood) bathing the bacteriocytes. The *Buchnera* proteome included all enzymes coded by the *Buchnera* genome that contribute to EAA synthesis, validating the genomic and transcriptomic evidence that *Buchnera* can produce EAAs (Shigenobu et al 2000, Reymond et al 2006, Wilson et al 2006). Furthermore, the abundance of EAA biosynthesis enzymes does not vary substantially among different biosynthetic pathways even though the amounts predicted from budget analyses and modeling studies vary among EAAs by an order of magnitude (Thomas et al 2009, Gunduz and Douglas 2009). This discrepancy suggests that variation in the rates of synthesis among different EAAs is unlikely to be shaped by variation in abundance of the enzymatic machinery, raising the possibility that other factors, such as the supply of precursors from the host, might determine the amount of each EAA that the *Buchnera* cells produce and release to the host.

Prior to this analysis, understanding the metabolites exchanged between *Buchnera* and the bacteriocyte was based primarily on the reconstruction of the *Buchnera* metabolism and *in silico* modeling of flux through the metabolic network (Shigenobu et al 2000, Thomas et al 2009,

Ramsey et al 2010). These studies predicted that the host metabolites taken up by *Buchnera* cells include glucose as the dominant carbon source, four nonessential amino acids (nEAAs: serine, aspartate, glutamate, and glutamine) as precursors of EAAs, and the nucleoside guanosine as precursor of *Buchnera* purines. The finding here that enzymes involved in the production of these precursors are enriched in the aphid bacteriocyte offers empirical validation of the models.

These results, however, raise the question why many of the core precursors for *Buchnera* metabolism are synthesized in the bacteriocyte, when the bacteriocyte could depend entirely on import of the compounds from the hemolymph bathing the bacteriocytes. In particular, insect organs in contact with the hemolymph generally derive their sugar requirement by uptake and trehalase-mediated hydrolysis of the dominant hemolymph sugar, trehalose (Thompson 2003). Remarkably, the trehalase protein was barely detectable in the bacteriocyte, despite the high predicted glucose demand of the *Buchnera* cells (Thomas et al 2009). Instead, the two rate-limiting steps in gluconeogenesis were enriched (Figure 2.6G), suggesting that bacteriocytes generate glucose by gluconeogenesis. The inclusion of enzymes mediating amino acid and lipid degradation in the protein cluster depleted in bacteriocytes (subcluster A1b-1 in Fig.2.5) suggests that host-derived substrates do not make an important contribution to gluconeogenesis. This leaves the organic acids (acetate, fumarate, and succinate), that are the dominant "overflow metabolites" of the *Buchnera* metabolic network (Thomas et al 2009), as likely substrates for bacteriocyte gluconeogenesis. In other words, the bacteriocyte can potentially recycle *Buchnera* waste carbon compounds to glucose, which is delivered back to the *Buchnera* cells. In this way, the availability of glucose for uptake by *Buchnera* varies according to the metabolic activity of the complement of *Buchnera* cells within the bacteriocyte.

The importance of enzymatic reactions in the bacteriocyte as the source of inputs to *Buchnera* metabolism is further illustrated by the *Buchnera* amino acid and purine nutrition. *Buchnera* derives its supply of the eight nEAAs that it cannot synthesize (all but glycine and cysteine) from the host. This proteome study confirms the interpretation from genomic and transcriptomic analyses (Shigenobu et al 2000, Nakabachi et al 2005) that one nEAA, tyrosine, is generated in the bacteriocyte from the *Buchnera*-derived EAA phenylalanine by the aphid phenylalanine hydroxylase. Of the remaining 7 nEAAs, four contribute to *Buchnera*-mediated synthesis of EAAs, and the enzymes generating these nEAAs (aspartate, glutamate, glutamine, and serine) are enriched and very abundant in the bacteriocyte proteome (Figures 2.6A-C). Routes for the synthesis of the other three nEAAs (alanine, asparagine and proline) are not enriched in the bacteriocyte, and these amino acids are presumably derived principally from the hemolymph. Similarly, *Buchnera* lacks the genetic capacity to synthesize purines *de novo*, but can derive its nucleotide requirements from host-derived nucleosides, principally guanosine (Shigenobu et al 2000, Ramsey et al 2010). The detection in the bacteriocyte of every enzyme contributing to *de novo* purine synthesis, and enrichment of several enzymes, including those contributing directly to the synthesis of guanosine (GMP synthase and 5′nucleotidase) (Figure 2.6H) suggests that, as with glucose and nEAAs contributing to *Buchnera* metabolism, guanosine is synthesized by the bacteriocyte and not derived from the hemolymph.

The hypothesis that the major metabolites taken up by *Buchnera* are synthesized in the bacteriocyte, and not imported from the hemolymph, is based on protein abundance. It requires testing by analysis of enzyme activities and metabolite flux *in vivo*. Even so, this likely pattern of host cell metabolism would facilitate precise, coordinated metabolic control by the host cell over the concentrations of the precursor pools for *Buchnera* metabolism. By controlling the supply of

key metabolites to *Buchnera*, the host can regulate *Buchnera* metabolism, including its capacity for growth and production of EAAs. The one major exception to this generality relates to polyamines. *Buchnera* is dependent on the supply of putrescine from the host for its synthesis of spermidine (Shigenobu et al 2000), which is a very abundant *Buchnera* transcript (Nakabachi et al 2005). In the absence of evidence for *de novo* synthesis of putrescine by ornithine decarboxylase in the bacteriocytes (Figure 2.6D), this crucial substrate for *Buchnera* is probably derived directly from the hemolymph.

Metabolite exchange between host and symbiont is also dependent on transporters of both *Buchnera* and aphid origin. This is because each *Buchnera* cell is bounded by three membranes, the inner and outer bacterial membranes typical of a Gram-negative bacterium, and the outermost symbiosomal membrane of aphid origin (Lamb and Hinde 1967). Although further research is required to identify the subcellular localization of the transporter proteins detected in this proteome analysis, the transporters of *Buchnera* origin were most likely located in *Buchnera* membranes (they were assigned to cluster-B) whereas many of the aphid transporters were probably symbiosomal because this is a very abundant membrane type in the bacteriocyte. For example, the surface area of the symbiosomal membrane is four times the surface area of the bacteriocyte cell membrane in the 7-day-old larvae used in this study [calculated from data in (Douglas and Dixon 1987, Whitehead and Douglas 1993)]. The annotated specificities of many of the transporters detected are consistent with the predicted metabolite transfer between the partners (*e.g.* phosphate, sulfate, glucose, amino acids, and nucleosides). Nevertheless, these annotations should be treated with great caution because their function is extrapolated from *E. coli* (for *Buchnera* transporters) and other insects or even vertebrates (for the aphid transporters).

To date, the function of none of the transporter proteins identified in the proteome of the bacteriocyte or *Buchnera* has been studied empirically.

A further candidate route for exchange between the partners is the hundreds of flagellar bases that cover the surface of the *Buchnera* cells (Shigenobu et al 2000). These structures have been suggested to mediate the selective transfer of *Buchnera* proteins and possibly small molecules to the bacteriocyte cytoplasm (Maezawa et al 2006, Reymond et al 2006). The reasoning is that *Buchnera* flagellar bases cannot confer motility because the flagellum and motor proteins are not encoded in the *Buchnera* genome (Shigenobu et al 2000), and that they function as Type III secretion systems (to which bacterial flagellum is evolutionarily related (Gophna et al 2003, Pallen et al 2006). In various pathogenic and symbiotic bacteria, type III secretion systems mediate the selective transfer of proteins from bacteria into eukaryotic cells (Kambara et al 2009). The identification of 20 flagellar proteins in this proteome study confirms and extends the previous proteomic and microscopical analysis (Shigenobu et al 2000). Nevertheless, the function of these remarkable structures remains unresolved, beyond the conclusion based on the cluster analysis (see above) that if they have any role in protein transfer, it is either nonselective or quantitatively small.

2.4.3 The Bacteriocyte as a Habitat for Symbiotic Bacteria

The intracellular habitat is widely regarded as an extreme environment because cells are colonized by relatively few microorganisms. Cellular defenses include efficient trafficking of intracellular microbes to lysosomes, apoptosis of cells infected by microbes, and the production of humoral immune effectors, such as anti-microbial peptides (Diacovich and Gorrel 2010, Radtke and O'Riordan 2006, Ray et al 2009). Aphid bacteriocytes conform to this generality in some respects. They are occupied by very few bacterial species: *Buchnera*, and occasionally

other bacteria ("secondary symbionts" *e.g. Hamiltonella defensa*, *Regiella insecticola*) that occur facultatively in aphids (Chandler et al 2008, Moran et al 2005, Oliver et al 2010).

The nonmetabolic proteins enriched in the bacteriocyte proteome offer insight into the bacteriocyte as a habitat for bacteria. In particular, multiple subunits of the V-type H^+ transporting ATPase (V-ATPase) are very abundant and enriched in the bacteriocyte proteome. The membrane localization of this ATPase has yet to be determined, but likely includes the lysosomes. Its abundance suggests that it may also be localized to the symbiosomal membranes, which are a major membrane fraction of bacteriocytes (see above), with the implication that the immediate environment of the *Buchnera* cells may be acidic.

Among the annotated immune-related genes, the only ones enriched in bacteriocytes are the two subunits of phenoloxidase. Phenoloxidase catalyzes the hydroxylation of monophenols to *o*-diphenols and ultimately to highly reactive quinines that react with oxygen to generate hydrogen peroxide and other reactive oxygen species (Nishikori et al 2009). These reactions contribute to the sclerotization of insect cuticle and synthesis of melanin, which is important in wound healing and encapsulation of fungi, parasitic wasps etc., especially in the hemolymph (Dittmer and Kanost 2010, Bilda et al 2005). The site of activity of the bacteriocyte phenoloxidase remains to be established. It may be released into the hemolymph, where it contributes to the hemolymph-based defenses, with the implication that the function of bacteriocytes extends beyond housing the *Buchnera* symbionts to contribute to the systemic immunity of the insect. Additionally or alternatively, phenoloxidase-mediated production of hydrogen peroxide may impose oxidative stress within the bacteriocyte. An independent indication that *Buchnera* is exposed to oxidative stress is that alkyl hydroperoxide reductase (AhpC) is the second most abundant protein in *Buchnera* cells studied here. The *E. coli* AhpC

scavenges endogenous hydrogen peroxide, generated in particular by flavoenzymes, notably NADH dehydrogenase (NuoA-N) also present in *Buchnera* (Seaver and Imlay 2001, Shigenobu et al 2000). The role of AhpC in protecting *Buchnera* from reactive oxygen species of endogenous origin and potentially generated by the host phenoloxidase is predicted to be particularly important because the *Buchnera* genome lacks homologs of the *E. coli* catalase genes *katE* and *katG* (Shigenobu et al 2000).

The bacteriocidal protein lysozyme has been reported previously as a major bacteriocyte transcript (Nakabachi et al 2005), but it was detected at low abundance in the proteome of just one of the three replicates studied here. Given that the proteome analysis was conducted on precisely-aged larval aphids in which *Buchnera* degradation is minimal (Whitehead and Douglas 1993, Douglas and Dixon 1987), this result suggests that lysozyme is not important in regulation of the functional symbiosis. As reported previously (Nishikori et al 2009), lysozyme may contribute to the orchestrated symbiosis breakdown in older insects.

A striking feature of this study is the lack of known immune effectors (Gerardo et al 2010) that are enriched or abundant in the bacteriocyte proteome in 7-day old larvae, even though *Buchnera* cells occupy more than 60% of the volume of these cells. Aphid proteins without sequence homology to previously described immune-related proteins are also most unlikely to mediate immune responses in the aphid bacteriocyte because none of the proteins enriched in bacteriocytes are coded by either hypothetical or orphan genes. This condition is in stark contrast to various other animal-microbial symbioses, including independently-evolved bacteriocyte symbioses in other insects, where the host immune system plays a central role in determining the abundance and proliferation of the microbial symbionts (Heddi et al 2005, Round and Mazmanian 2009, Wang et al 2009). These considerations raise the possibility that,

with the possible exception of the aphid phenoloxidase, the immune system may play little or no direct role in controlling the abundance and traits of *Buchnera* cells in the functional symbiosis of 7-day-old larvae. The dominance of enriched bacteriocyte proteins by metabolic enzymes involved in the supply of metabolites to the *Buchnera* raises the alternative hypothesis, that the *Buchnera* function, including population increase that is tightly coordinated with the host and sustained nutrient release, is regulated principally by the metabolic relations of the bacteria with the host cell. The important implication is that the specialized function of the bacteriocyte, which evolved *de novo* at the evolutionary inception of the aphid-*Buchnera* symbiosis, may not be underpinned by a suite of novel gene products unique to aphids. Instead, this novel cell type is founded on a unique expression profile of host and bacterial genes, many of which are individually widely distributed among animals and bacteria, respectively, and beyond.

CHAPTER 3

SHARED AMINO ACID PATHWAYS IN THE PEA APHID-*BUCHERA* SYMBIOSIS

3.1 Introduction

The capacity of various insect groups to utilize nutritionally unbalanced diets is correlated with possession of vertically-transmitted, obligate symbiotic microorganisms with reduced genomes (Shigenobu et al 2000, Akman et al 2002, Nasir and Noda 2003). Unexpectedly, the genomes of bacterial symbionts of aphids and other Sternorrhynchan insects (aphids, mealybugs, whiteflies, psyllids) are missing genes in the biosynthetic pathways for essential amino acids (EAAs) required by the host (Shigenobu et al 2000, McCutcheon and von Dohlen 2011, Jiang et al 2013). The pea aphid genome, like other insect genomes (IAGC 2010), contains genes coding for enzymes with equivalent activity to the genes lost by the obligate bacterial symbionts. The resulting proposal that metabolic pathways are shared between the pea aphid *Acyrthosiphon pisum* and its symbiotic bacterium *Buchnera aphidicola* was unprecedented (Wilson et al 2010), as metabolic pathways are traditionally defined as the property of individual organisms.

Buchnera lack the genes coding for the terminal reactions of the branched chain amino acids (BCAAs) and phenylalanine biosynthetic pathways, and proximal reactions in isoleucine and methionine biosynthesis. The terminal reactions for the branched chain amino acids (BCAAs) and phenylalanine are predicted to be mediated by host enzymes, branched chain

---

* Presented with minor modifications from the originally submitted article Russell CW, Bouvaine S, Newell PD, Douglas AE. 2013. Shared Metabolic Pathways in a Coevolved Insect-Bacterial Symbiosis. Submitted to Applied and Environmental Microbiology. All protocols and experiments conducted by Calum Russell, except for Western blots (Sophie Bouvaine).

aminotransferase (BCAT) and an aspartate aminotransferase (GOT2), respectively (Wilson et al 2010). Consistent with the predicted aphid contribution to the synthesis of the BCAAs and phenylalanine, the transcript and protein of the candidate aphid enzymes contributing to these metabolic pathways are enriched in the specialized host cells (bacteriocytes) that only house the bacterial symbiont, *Buchnera* (Poliakov et al 2011, Hansen and Moran 2011, Macdonald et al 2012). In contrast to the consensus among genomic, transcriptomic and proteomic data sets for the enzymes performing the terminal reactions, there are multiple candidate precursors and host enzymes for the production of isoleucine and methionine (Poliakov et al 2011, Hansen and Moran 2011). Furthermore, these predictions do not exclude the alternative explanation for the sustained synthesis of EAAs: that the reactions coded by the missing *Buchnera* gene products are mediated by other *Buchnera* enzymes (and not host enzymes) with a greater or different substrate range than indicated by their annotation (Shigenobu et al 2000, Kelkar and Ochman 2013).

Resolving the question of whether some EAAs are synthesized by shared metabolic pathways or entirely by *Buchnera* is central to the understanding of the coevolutionary interactions in this relationship because the *Buchnera*-derived EAAs are required by the aphid to utilize plant phloem sap, a diet grossly deficient in these nutrients (Gunduz and Douglas 2009). In this study, I sought to test whether the host contributes to EAA production. The discovery that lysates of bacteriocytes, containing intact *Buchnera* cells, display a net production of certain EAAs at linear rates allowed for an opportunity to test the hypothesis of shared EAA biosynthetic pathways. Isolated *Buchnera*, which solely infect the bacteriocytes, were incubated with soluble host factors obtained from the bacteriocytes, recombinant enzymes, and various precursors to measure the effect on essential amino acid release. These experiments demonstrate that *Buchnera* require exogenous enzymes to produce certain essential amino acids, and that

some predictions about aphid candidate enzymes from genome annotation and gene expression studies (Wilson et al 2010, Poliakov et al 2011, Hansen and Moran 2011) were correct, while others were erroneous. This study further illustrates how genome-based reconstructions of organismal function are not facts, but hypotheses that require testing by empirical analysis.

3.2 Materials and Methods

3.2.1 Aphid rearing

Aphids were reared from a single, parthenogenetic female collected from an alfalfa field in Freeville, NY in June 2009. The aphid line, CWR09/18, was screened by PCR and microscopy for bacterial symbionts and found to contain *Buchnera aphidicola* and no secondary symbionts. The line was maintained on pre-flowering *Vicia faba* cv. Windsor at 20˚C with a 16:8 light:dark cycle.

3.2.2 Amino acid release and analysis

Bacteriocytes were dissected from 7-day-old larval aphids in extraction medium (28 mM glucose, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.25 M sucrose, 50 mM NaH_2PO_4, 13 mM $K_2H_2PO_4$, pH 7.5), lysed by pipetting 4-6 times, and centrifuged at 1000 *g* for 5 minutes at 4˚C to separate the *Buchnera*-free supernatant (referred to as the host fraction, HF) from the pellet containing *Buchnera* cells. The *Buchnera* cells in the pellet were quantified by hemocytometer counts at x400 magnification, and diluted to 4 x 10^8 *Buchnera* ml^{-1} in extraction medium. To initiate the release experiment, 5.5 µl reaction medium was added to 8 replicate samples of 5.5 µl bacterial suspension. The standard reaction medium comprised the extraction medium supplemented with glutamate, glutamine, serine, aspartate, and 2-oxobutanoate, each at 2 mM, but 2-oxobutanoate was replaced by 2 mM homoserine, threonine, cystathionine, cysteine, or homocysteine where indicated. *Buchnera* with access to the host fraction was

exposed to standard extraction medium versus reaction medium to determine if glutamate, glutamine, serine, and aspartate promoted release of certain essential amino acids. For determination of branched chain aminotransferase and phenylalanine transaminase activity in the host fraction, the standard reaction medium was supplemented with 2 mM 4-methyl-2-oxopentanoate or phenylpyruvate and added to HF. At 5 minute intervals over 40 minutes, one tube was centrifuged at 1000 *g* for 70 seconds, and 11 µl supernatant was immediately flash-frozen in liquid nitrogen and stored at -80^{o}C. The experiments were conducted at 22.5^{o}C, and were repeated 3-5 times on different days using different sets of aphids.

The amino acid content of the supernatant was quantified using the AccQ Tag derivatization kit (Waters) by UPLC with PDA detector (Waters Acquity). An equal volume of 40 mM HCl was added to 10 µl of each supernatant. Following incubation on ice for 30 minutes, the sample was centrifuged at 18000 *g* for 10 minutes at 4˚C, and the supernatant was filtered through a 0.45 µm filter plate (Millipore) by centrifugation at 1500 *g* for 10 min. The filtrate (2.5 µl) was derivatized with AccQ Tag (Waters), following manufacturer's protocol, and injected into Waters Acquity UPLC with PDA detector and AccQ-Tag Ultra 2.1 x 100 mm column. The gradient was: 0-0.54 min, 99.9% A 0.1% B; 0.54-5.74 min, 90.9% A and 9.1% B; 5.74-7.74 min, 78.8% A 21.2% B; 7.74-8.04 min, 40.4% A 59.6% B; 8.04-8.64 min, 10% A 90% B; 8.05-8.64 min 10% A 90% B; 8.64-8.73 min 99.9% A 0.1% B; 8.73-9.50 min, 99.9% A 0.1% B (linear between each time point), where A is 10% AccQ-Taq Ultra Eluent A in water, and B is Accq-Taq Ultra Eluent B. Amino acids were determined by comparison to standards: 1, 5, 10, 50 and 100 pmol amino acids µl^{-1} (Waters amino acid hydrolysate standard #088122, supplemented with asparagine, tryptophan and glutamine).

3.2.3 Immunoblots

Custom-made polyclonal antibodies against glutamate oxaloacetate transaminase 2 (GOT2, Gene ID: 100144899), branched chain aminotransferase (BCAT, **Gene ID: 100167587)** and threonine dehydratase (TD, Gene ID: **100165866)** were produced by GenScript (Piscataway, NJ, USA). Rabbits were immunized four times using the purified peptides CNPTGVDPKPEQWKE (GOT-2) CVDERPHLYESQNYK (BCAT) and CMEHGSPITVDGKST (TD) conjugated with Keyhole limpet hemocyanin (KLH). The polyclonal antiserum obtained from the last bleed was affinity purified and tested by ELISA, including confirmation that each pre-immune serum did not react to aphid proteins. The specificity of the antibodies was assessed using mass spectrometry. Briefly, the reacting band in the western blot was manually excised from the corresponding acrylamide gel, digested and submitted to a HPLC system (Dionex Ultimate 300 configured for nanobore), in line connected to a hybrid triple quadropole linear ion trap mass spectrometer, 4000 Q Trap (ABI/MDS Sciex) equipped with Ion Spray Head ion source. The resulting MS data were submitted for database searching using MASCOT search engine version 2.3 against a database containing the pea aphid genome with 34,834 protein-coding gene models (AphidBase.com).

For protein analysis, whole bodies and dissected bacteriocytes of 7-day-old larval aphids were homogenized in ice-cold buffer containing 35 mM Tris, 25 mM KCl amd 10 mM $MgCl_2$, pH 7.4. Following centrifugation at 10,600 *g* for 5 minutes, the protein content of the supernatant was quantified using RC/DC protein assay kit (Bio-Rad, Hercules, Ca, USA) following the manufacturer's instructions with bovine serum albumin as a standard. SDS-PAGE gels were performed as described by Laemmli (1970), with 2.5 µg protein per sample on a 12 % polyacrylamide gel containing 0.1% SDS. For western blots, the proteins were transferred to a

nitrocellulose membrane and blocked in PBS containing 0.5 % Tween (PBS-T) and 5% milk powder. The membrane was incubated successively in PBS-Tween containing a polyclonal antibody diluted either at 1/200 (BCAT and GOT2) or 1/1000 (TD), with an anti-rabbit IgG conjugated to peroxidase (Sigma) diluted 1/20,000, and then ECL substrate (Bio-Rad), followed by visualization with Molecular Imager ChemiDoc XRS (Bio-Rad). The intensity of the bands was analyzed by Quantity One software (Bio-Rad).

3.2.4 Generation of Recombinant Enzymes

E. coli was used as a source of enzymes that mediate the terminal reactions in BCAA synthesis (IlvE), phenylalanine synthesis (TyrB) and the synthesis of homocysteine precursor of methionine (MetC). The genes were amplified from *E. coli* JM109 genomic DNA with the following primers: ilvE_forward: 5' TTTAGGATCCACCACGAAGAAAGCTG 3'; ilvE_ reverse: 5' TTTAGGTACCGTGTCTGTCTCGTAAA 3'; tyrB_forward: 5' TTTAGGATC CCAAAAAAGTTGACGCCT 3'; tyrB_reverse: 5' TTTAGGTACCTTACATCACCGCAGCA 3'; metC_forward: 5' TTTAGGATCCGCGGACAAAAAGCTTG 3'; metC_reverse: 5' TTTA GGTACCTTATACAATTCGCGCAA 3'. Each forward primer includes a *BamH1* site, and each reverse primer a *Kpn1* site, for ligation into the expression vector pProExHT, to yield p6His-IlvE, p6His-TyrB and p6His-MetC. Successful plasmid construction was confirmed by restriction digests and Sanger sequencing.

Recombinant enzyme was expressed and purified from *E. coli* BL21 cells, essentially as described (Newell et al 2011). Briefly, 1 ml of overnight culture was added to 350 ml LB, and grown shaking at 120 rpm, 37˚C for 4.5 hours. After 4.5 hours, IPTG was added to a concentration of 0.5mM, and the culture shifted to 25˚C for an additional 18-20 hours of incubation. Cells were pelleted, flash frozen and stored at -80˚C until enzyme purification.

For purification, 5 ml equilibration buffer (20 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 7.4) and 50 ul Halt Protease inhibitor cocktail (Thermo Scientific) were added to the frozen cell pellet, and cells lysed by sonication. Cellular debris was removed by centrifugation at 4500 g for 10 minutes at 4˚C. The supernatant was added to 1 ml equilibrated HisPur Ni-NTA Resin (Thermo Scientific), and incubated at 4˚C for 1.5 hours with gentle rocking to allow binding of the His-tagged protein. The resin was then washed 5 times with 5 ml wash buffer (20 mM NaH_2PO_4, 300 mM NaCl, 25 mM Imidazole, pH 7.4) for 10 minutes at 4˚C. Resin-bound protein was eluted in three steps with elution buffer (20 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, pH 7.4) and the third elution was dialyzed overnight at 4˚C in a *7000MWCO* membrane against 1 l glucose-free extraction buffer (see *(b)* above). The presence of pure protein was confirmed by SDS-PAGE, and enzymatic activity was demonstrated by incubation with substrate and detection of the predicted product by UPLC (IlvE: isoleucine, leucine, valine from the cognate oxo-acid; TyrB: phenylalanine from phenylpyruvate; and MetC: homocysteine from cystathionine).

To quantify the effect of recombinant enzyme on amino acid production by *Buchnera*, the reaction medium (in *(b)* above) was supplemented with 0.2 mM pyridoxal-5-phosphate and 20 µg recombinant enzyme protein ml^{-1}. The dilute protein concentration was to ensure that activity was due to the extracted protein, and not a contaminant. As a result, activity was low for all experiments. The reaction medium for MetC experiments also contained 2 mM cystathionine.

3.2.5 Statistical Analyses

The release rate of amino acids from *Buchnera* preparations was quantified from the regression of amino acid content in the medium over time. The variation in release rates with treatment was analyzed by ANOVA, t-test, or paired t-test (with Holm-Bonferroni correction),

following logarithmic transformation to obtain normally distributions (Shapiro-Wilk test) with homogenous variance (Levene's test), with the LSD post-hoc to determine significant pair-wise differences within the ANOVA. Data that failed the Levene's test were analyzed with a Kruskal-Wallis test, with pairwise differences determined with Mann-Whitney U tests with adjusted p-values (bonferroni correction).

3.3 Results

3.3.1 Impact of substrate on release of isoleucine, leucine, phenylalanine, and valine

Due to the decreased metabolic potential of *Buchnera*, metabolic maps predict that *Buchnera* require 33-host derived compounds to promote growth and to synthesize essential amino acids (Macdonald et al 2011). The medium was supplemented with 5 of these substrates to determine impact on essential amino acid release: aspartate, glutamate, glutamine, serine, and 2-oxobutanoate. As expected, supplementation of medium with substrate increases the release rate of certain essential amino acids by *Buchnera* with access to the host fraction (Figure 3.1). Isoleucine, and leucine all had a significant increase in the release rate, with valine and phenylalanine showing a similar trend. For the purpose of this paper, these essential amino acids were focused on as they had the highest release rate and were predicted to have host contribution towards their synthesis.

3.3.2 Metabolic source of 2-oxobutanoate, the precursor of isoleucine

The supplementation of the medium with 2-oxobutanoate was to provide the carbon skeleton for isoleucine production. Production of isoleucine is an essential function of the aphid *Buchnera* symbiosis, yet the source of this key intermediate, 2-oxobutanoate, remains to be defined. Predictions for the source of 2-oxobutanoate based upon genomic, transcriptomic, and proteomic datasets have failed to produce a consensus (Wilson et al 2010, Poliakov et al 2011,

Hansen and Moran 2011). One possible source of 2-oxobutanoate is threonine degradation, mediated by threonine dehydratase (Figure 3.2A) (Wilson et al 2010, Hansen and Moran 2011). Contrary to this prediction, threonine dehydratase was readily detected by Western blot in the whole aphid, but undetectable in the bacteriocytes (Figure 3.2B). An alternative source of 2-oxobutanoate could be from the conversion of homoserine by cystathionine-γ-lyase, a possibility supported by the enrichment of cystathionine-γ-lyase in the bacteriocyte proteome (Poliakov et al 2011).

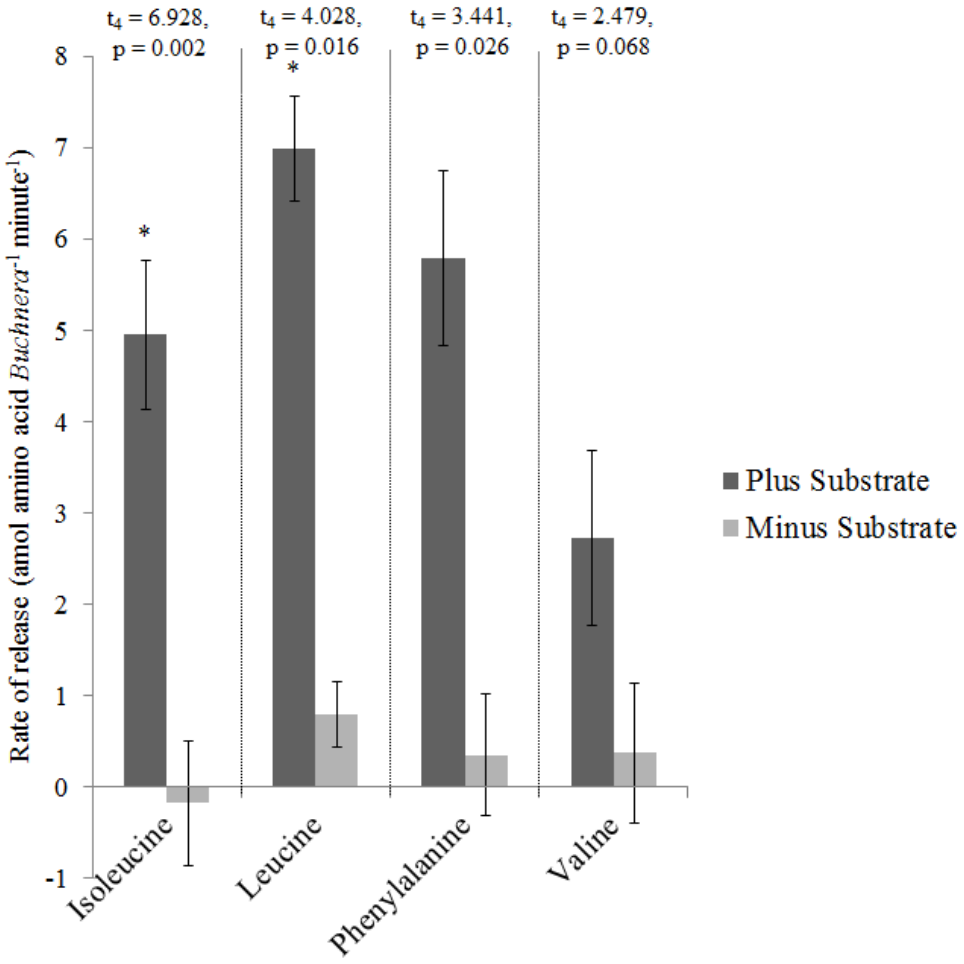


Figure 3.1 Rate of amino acid release dependent on substrates provided in medium. Isoleucine, leucine, phenylalanine, and valine release by preparations of *Buchnera* with HF with and without substrate (aspartate, glutamate, glutamine, serine, and 2-oxobutanoate). (mean +/- s.e., paired t-test, critical p-value determined by Holm-Bonferroni correction)

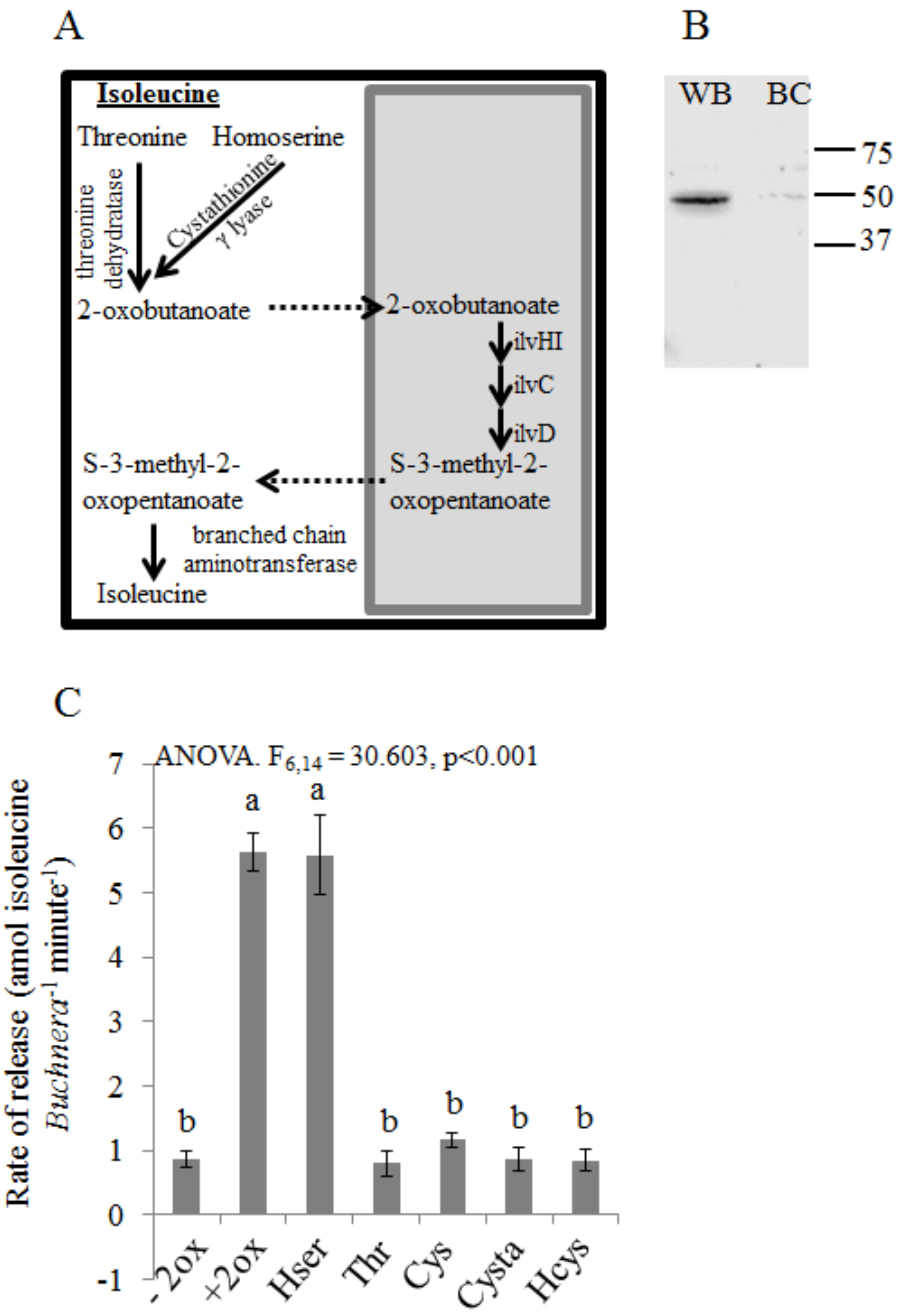


Figure 3.2. Rate of isoleucine release by preparations of *Buchnera* with HF. A) A metabolic map illustrating the potential sources for 2-oxobutanoate based upon candidate enzymes: threonine and homoserine (shaded box represents *Buchnera*) B) Western blot of 5 ug of protein from whole aphid (WB) and bacteriocytes (BC) using polyclonal antibody against threonine dehydratase C) Isoleucine release from *Buchnera* + HF incubated with different precursors, -2ox = minus 2-oxobutanoate (no substrate), +2ox = 2-oxobutanoate, Hser = homoserine, Thr = threonine, Cys = cysteine, Cysta = cystathionine, Hcys = homocysteine (mean +/- s.e., LSD post hoc test, $p < 0.001$)

Isoleucine release appears to be dependent on supplementing the reaction medium with aspartate, glutamate, glutamine, serine and 2-oxobutanoate. To test whether 2-oxobuanoate, or an earlier precursor, leads to production of isoleucine, *Buchnera* preparations with HF were supplemented with candidate pathway intermediates (including 2-oxobutanoate) and isoleucine production was measured. Only 2-oxobutanoate and homoserine supported a significantly elevated release rate of isoleucine, when compared to medium with no source of 2-oxobutanoate (Figure 3.2C). These findings suggest that the precursor of 2-oxobutanoate is homoserine and not threonine, which did not stimulate isoleucine release.

3.3.3. Terminal transamination reactions in synthesis of leucine and phenylalanine

Modifying the medium has allowed us to determine possible substrates for production of isoleucine and validate hypothesized pathways for 2-oxobutanoate synthesis. Building upon this, and previous work fractioning the bacteriocytes into host and bacterial components (Poliakov et al 2011), will allow us to determine whether the host contributes to production of the branched chain amino acids and phenylalanine. As well as lacking the genes coding for enzymes responsible for the production of 2-oxobutanoate, *Buchnera* lacks the genetic capacity for the terminal reactions in the synthesis of the branched chain amino acids (BCAAs) and phenylalanine. The aphid enzymes branched chain aminotransferase (BCAT) and aspartate aminotransferase (GOT2 – annotated with phenylalanine aminotransferase activity) have been hypothesized to mediate the missing reactions in BCAAs and phenylalanine biosynthesis, respectively (Figure 3.3A) (Wilson et al 2010, Poliakov et al 2011, Hansen and Moran 2011). Both BCAT and GOT2 were detected in the bacteriocyte when tested by immunoblotting with polyclonal antibodies (Figure 3.3BC). To test whether exogenous enzymes mediate these *Buchnera* missing reactions under physiological conditions, leucine and phenylalanine

production by *Buchnera* preparations incubated with and without HF were quantified. HF significantly promoted leucine and phenylalanine release, by 7-fold and 4-fold, respectively (Figure 3.3DE).

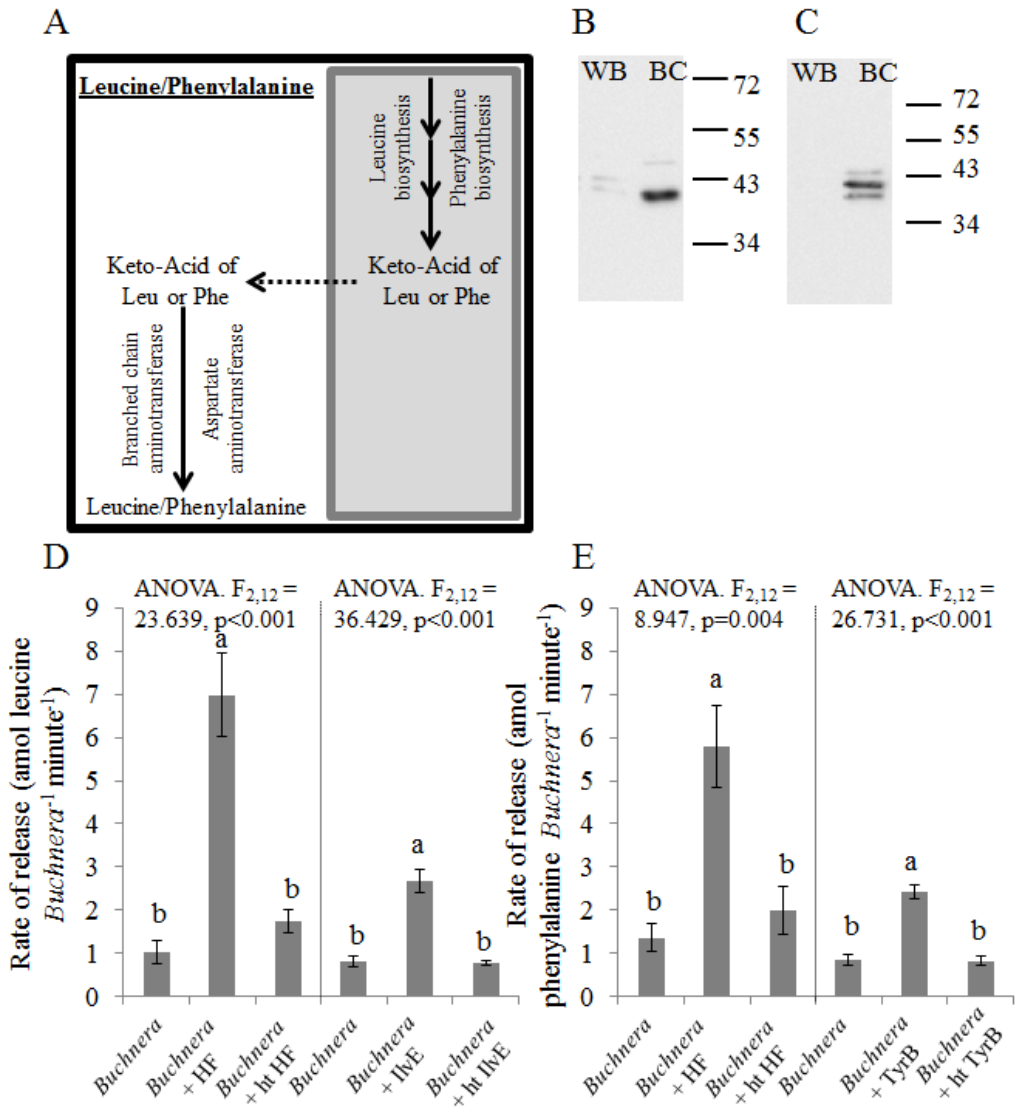


Figure 3.3. Rate of leucine and phenylalanine release by isolated *Buchnera* preparations. A) Metabolic map of the terminal reactions for leucine and phenylalanine biosynthesis (shaded box represents *Buchnera*), B) Western blot of 5 ug of protein from whole aphid (WB) and bacteriocytes (BC) using polyclonal antibody against BCAT, C) Western blot of 5 ug of protein from whole aphid (WB) and bacteriocytes (BC) using polyclonal antibodies against GOT2, D) Leucine release from *Buchnera* +/- HF and release from *Buchnera* +/- recombinant IlvE, E) Phenylalanine release from *Buchnera* +/- HF and release from *Buchnera* +/- recombinanat TyrB (HF = host fraction, ht = heat treated. mean +/- s.e., LSD post hoc test, p <0.01)

To further test whether the active constituent in HF-stimulated production of BCAAs and phenylalanine was transaminase activity, *Buchnera* preparations were incubated with recombinant transaminases (IlvE and TyrB). Consistent with the predicted role of host enzyme-

mediated transamination, these treatments resulted in significant increases in the release rate of leucine and phenylalanine (3.3 and 2.9-fold increase respectively, Figure 3.3DE). Rather than rely solely on recombinant protein to support exogenous activity, HF was tested for branched chain aminotransferase and phenylalanine aminotransferase activities. The HF showed a significant increase in leucine and phenylalanine release (9-fold and 135-fold, respectively) when the medium was supplemented with either 4-methyl-2-oxopentanoate or phenylpyruvate, with activity being lost when the HF is heat treated (Figure 3.4).

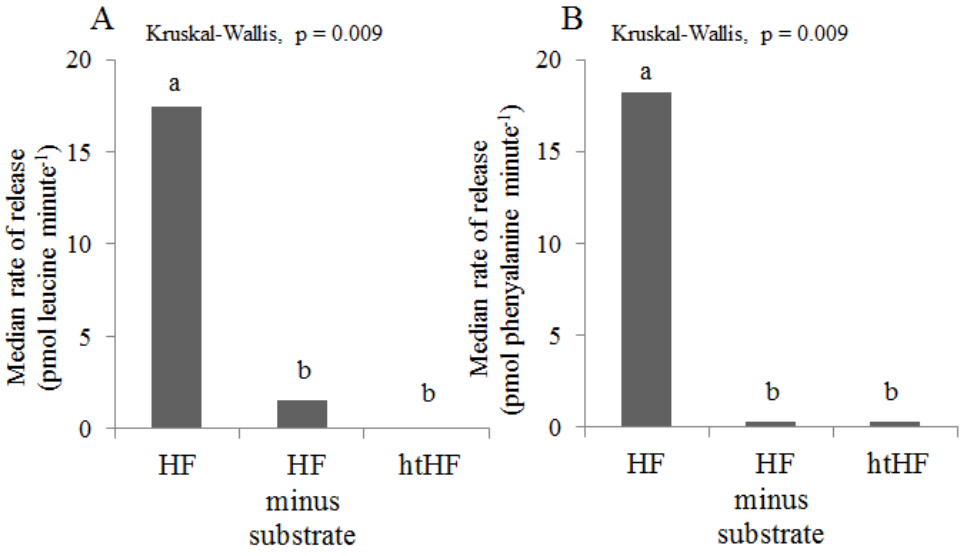


Figure 3.4. Enzymatic activities in the host fraction (HF). A) Release of leucine from HF with and without 4-methyl-2-oxopentanoate B) Release of phenyalanine from HF with and without phenylpyruvate (ht = heat treated. Kruskal-Wallis test, $p = 0.009$, Mann-Whitney U tests $p < 0.016$).

3.3.4 Metabolic source of homocysteine, the precursor of methionine

The immediate precursor for methionine synthesis is homocysteine, to which a methyl group is added to form methionine. *Buchnera* has retained the gene for the terminal enzyme, MetE, but lacks genes for the other steps in the methionine biosynthesis. As a result of this genome erosion, *Buchnera* is predicted to require an exogenous source of homocysteine to support methionine production. It has been hypothesized that the aphid provides *Buchnera* with homocysteine by a reversal of the transulphuration pathway, utilizing cystathionine-γ-lyase and cystathionine-β-synthase to form cystathionine from cysteine and homocysteine from cystathionine respectively (Figure 3.5A) (Wilson et al 2010, Hansen and Moran 2011). The reversal of the transulphuration pathway in animals has never been experimentally shown in any system to date. To investigate the metabolic source of homocysteine, *Buchnera* preparations with HF were incubated with several possible precursors for homocysteine, including homocysteine as a positive control. As expected, homocysteine promoted methionine production (73-fold increase over the medium-only control), yet only cystathionine, and not cysteine, promoted methionine production (25-fold increase vs. no increase; Figure 3.5B). These findings confirm that aphid homocysteine is the precursor for *Buchnera*-mediated methionine synthesis, but are inconsistent with the prediction that the transulphuration pathway functions in reverse (Wilson et al 2010), with methionine ultimately being produced from cysteine. Instead, they suggest that cystathionine is the metabolic precursor for methionine.

To determine which steps in the methionine pathway occur outside of the bacterium, *Buchnera* were supplemented with homocysteine or cystathionine and incubated with or without host fraction, and the rate of production of methionine was measured. Addition of HF to the *Buchnera* preparation in homocysteine-supplemented medium showed no significant increase in

methionine production relative to the same treatment without HF (Figure 3.5C), as predicted with the terminal enzyme being a *Buchnera* enzyme. In cystathionine-supplemented medium, however, HF caused a significant increase in the methionine production by *Buchnera* (14-fold increase, Figure 3.5D). To test whether the enzymatic activity generating homocysteine is external to the bacterium, *Buchnera* preparations were incubated with recombinant MetC (converts cystathionine to homocysteine). The *Buchnera* treated with MetC and cystathionine showed a significant increase in methionine production (8-fold increase) when compared to enzyme-free medium (Figure 3.5D). Taken together, these findings indicate that homocysteine is likely being produced from exogenous cystathionine, which in turn is converted to methionine by *Buchnera* MetE.

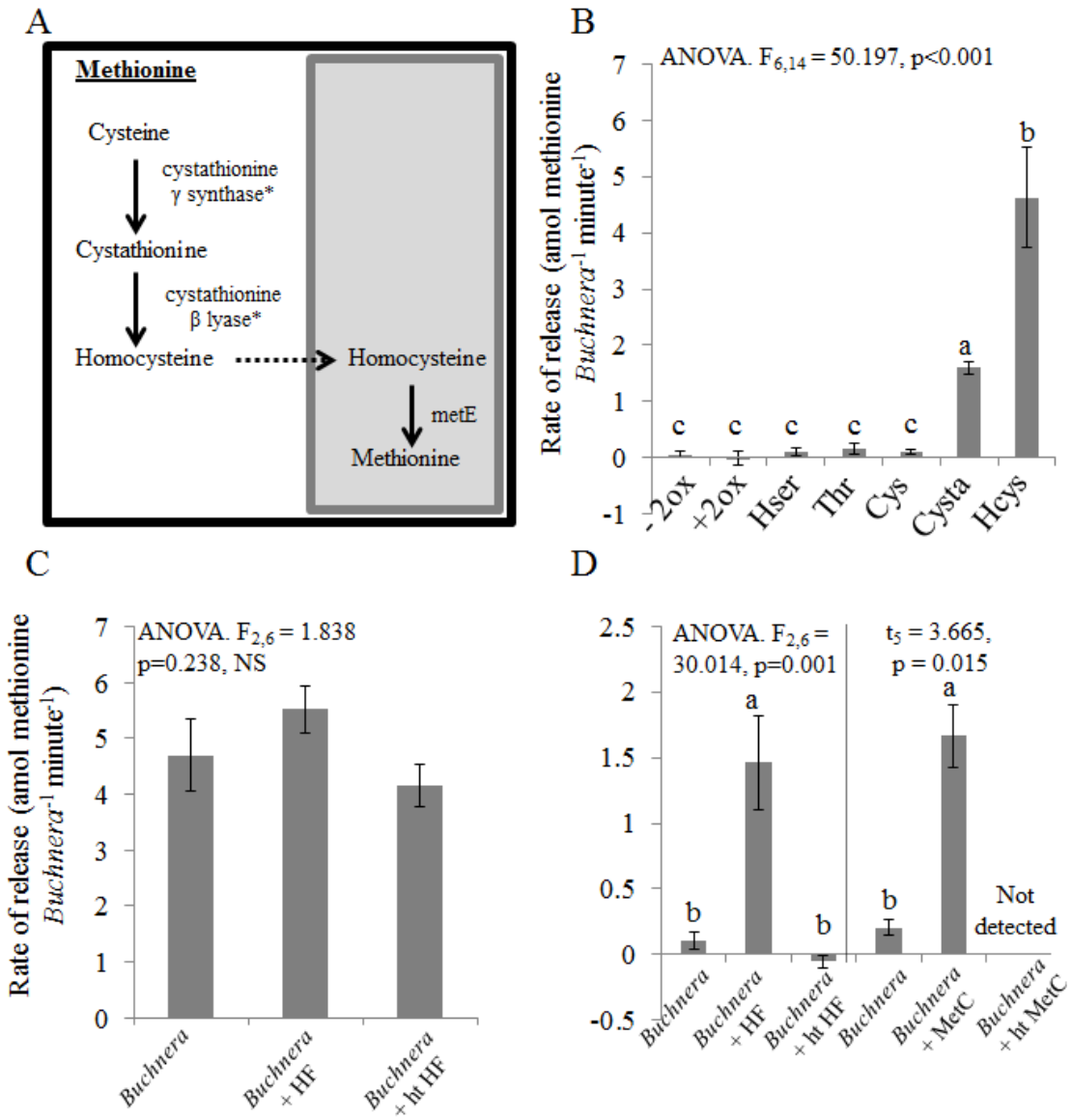


Figure 3.5. Rate of methionine release by isolated *Buchnera* preparations. A) Metabolic map of the reversal of the transsulfuration pathway to produce homocysteine (shaded box represents *Buchnera*, * annotated as cystathionine-γ-lyase and cystathionine-β-synthase, names were changed to reflect theorized activities) B) Methionine release from *Buchnera* + HF incubated with different precursors, -2ox = minus 2-oxobutanoate (no substrate), +2ox = 2-oxobutanoate, Hser = homoserine, Thr = threonine, Cys = cysteine, Cysta = cystathionine, Hcys = C) Methionine release from *Buchnera* +/- HF in homocysteine medium D) Methionine release from *Buchnera* +/- HF in cystathionine medium, and methionine release from *Buchnera* +/- recombinant MetC. (HF = host fraction, ht = heat treated. mean +/- s.e., LSD post hoc test, p <0.01)

3.4 Discussion

Linked to its small gene content, *Buchnera* is nutritionally fastidious. *Buchnera* growth is inferred to require 33 host-derived metabolites (Macdonald et al 2011) and there is no reasonable prospect of culturing this bacterium. Despite this, viable and metabolically active *Buchnera* cells persist in defined medium for some hours (Douglas et al 2011, Sasaki and Ishikawa 1995, Whitehead and Douglas 1993), facilitating direct investigation of their metabolic capabilities. In this study, the metabolic functions of isolated *Buchnera* preparations was probed and generated direct evidence that metabolic pathways are shared between host and symbiont. Specifically, the experiments demonstrated (a) net production of leucine and phenylalanine by *Buchnera* supplemented with either recombinant enzyme mediating the terminal biosynthetic reaction (IlvE and TyrB, respectively) or HF containing functionally equivalent aphid enzymes; and (b) that *Buchnera* preparations require the exogenous supply of either the substrate or product of host reactions for production of isoleucine and methionine, indicating the host's compensation for missing *Buchnera* IlvA and MetC.

In addition to validating the host-mediated terminal reactions in leucine and phenylalanine synthesis, this study provides metabolism-based identification of the substrates utilized by the host-mediated proximal reactions. This study supports that the host-derived 2-oxobutanoate precursor of isoleucine synthesis is synthesized from homoserine and not threonine. This runs contrary to the prediction from the aphid genome annotation (IAGC 2010), but is fully congruent with metabolic evidence that pea aphids do not metabolize dietary ^{14}C-threonine to isoleucine (Febvay et al 1995); and cystathionine-γ-lyase, an enzyme that is enriched in the pea aphid bacteriocytes, has previously been suggested to mediate the synthesis of 2-oxobutanoate from homoserine (Poliakov et al 2011).

Somewhat unexpectedly, and counter to predictions based on pathway metabolic pathway reconstruction (Wilson et al 2010), these results suggest that the homocysteine substrate for *Buchnera*-mediated methionine synthesis may be derived from cystathionine, and not cysteine. The predicted host enzyme mediating homocysteine synthesis from cystathionine, cystathionine-β-synthase (Gene ID: 100166111), is undetectable in the aphid bacteriocytes (Poliakov et al 2011), and the enzyme mediating this reaction remains to be identified; detailed analysis of the bacteriocyte proteome has identified no strong candidates (Russell, unpub. data).

An important consideration in interpreting experiments on isolated symbiotic microorganisms is their relevance to function in the symbiosis. When certain symbionts are separated from their hosts, key symbiotic traits, notably the selective release of nutrients advantageous to the host, are abolished (Douglas 2010). Several lines of evidence suggest that the isolated *Buchnera* preparations retain symbiosis-relevant metabolic properties. In particular, the sustained release of the penultimate metabolites in the BCAA and phenylalanine biosynthesis pathways can be inferred with confidence because these EAAs are produced by *Buchnera* preparations incubated with recombinant enzymes mediating the terminal reactions, but an exogenous supply of these precursors are required for the EAA production by *Buchnera*-free HF. The isolated *Buchnera* cells apparently do not release the substrate of the proximal host reactions for isoleucine and methionine synthesis (homoserine and cystathionine, respectively), suggesting that these substrates are either derived from the host or that their release from *Buchnera* cells is lost on isolation. Another consideration is the transport of these compounds between the *Buchnera* and the pea aphid. The pea aphid has an expanded set of amino acid transporters that could fulfill the role of transport of intermediates, yet the *Buchnera* has relatively few transporters that cannot account for the wide variety of compounds that would cross the

periplasm (Shigenobu et al 2000). Future studies will need to be performed to determine the transporters responsible, and permeability, of the *Buchnera* membrane.

Metabolic cooperation occurs widely among microorganisms, generally in response to selection imposed by resource-poor habitats (Wintermute and Silver 2010, Klitgord and Segre 2011, Freilich et al 2011). The coevolution of shared metabolic pathways in the aphid-*Buchnera* symbiosis most probably has a different evolutionary basis. Because the vertically-transmitted *Buchnera* cells have a small effective population size, they are subject to gene loss through genomic deterioration (Moran 1996, Moran and Mira 2011), selecting for evolutionary changes in host gene expression networks to recruit enzymes that compensate for the missing bacterial enzymatic reactions in the host cell.

The incidence of shared metabolic pathways in animal-microbial symbioses is largely unstudied, but may be widespread, albeit not universal, in associations involving microorganisms with reduced genomes. For example, independently-evolved bacterial symbionts in phloem-feeding whiteflies and mealybugs of the same sub-order as the aphids, the Sternorrhyncha, have incomplete genetic capacity for EAA biosynthesis (McCutcheon and von Dohlen 2011, Sloan and Moran 2012, Sabree et al 2013), but the biosynthetic pathways are apparently complete in the bacterial symbionts of insects (e.g. cicadas, spittlebugs, leafhoppers) in the related sub-order, the Auchenorryncha (McCutcheon et al 2009, McCutcheon and Moran 2007, McCutcheon and Moran 2010). Turning to different symbioses, the stimulation of EAA production from *Buchnera* preparations by the host cell fraction (HF) is reminiscent of enhanced photosynthate release by symbiotic algae isolated from corals or other marine invertebrates and incubated with homogenate of their hosts (Trench 1971). The similarity may, however, be superficial because the effect of host homogenate on algal cells is generally interpreted as a chemical signal that

induces the synthesis and export of specific photosynthesis-derived nutrients (Grant et al 2006, Wang and Douglas 1997).

In conclusion, the shared metabolic pathways in the pea aphid-*Buchnera* symbiosis are the product of metabolic coevolution, involving the loss of *Buchnera* genes mediating reactions also present in animals, and the compensatory enrichment of expression of the animal enzymes in host cells housing the symbionts. Associated coevolutionary changes in the host and symbiont transporters mediating the transfer of metabolic intermediates between the partners are also anticipated (Price et al 2011). Importantly, the host enzymes implicated in symbiotic EAA biosynthesis are localized to the host cell (Poliakov et al 2011), and not derived by lateral gene transfer from the ancestral *Buchnera* or other bacteria (Nikoh et al 2010). In this respect, the evolutionary trajectory of these symbionts displaying metabolic coevolution differs markedly from that of the bacterial-derived organelles whose metabolism is sustained by the products of organelle-derived genes transferred to the host nucleus (McCutcheon 2010).

CHAPTER 4

RESPONSE OF THE MATERNAL APHID BACTERIOCYTE TO DIETARY ESSENTIAL AMINO ACIDS

4.1 Introduction

A common adaptation to overcoming dietary restriction is to form stable, long-term associations with microbial partners. These partners range from bacterial, to protozoan, and to fungal microorganisms that have associations with their animal host (Shigenobu et al 2000, Hungate 1950, Nasir and Noda 2003). These partners are commonly vertically transmitted, undergo genome reduction due to relaxed selective pressure and genomic decay, and contain genes coding for metabolic pathways not found in the host genome (McCutcheon and Moran 2010). These metabolic pathways tend to mediate the synthesis of nutrients lacking in the diet (Akman et al 2002, Kirkness et al 2010, Shigenobu et al 2000, Tamas et al 2002, van Ham et al 2003, Perez-Brocal et al 2006, Lamelas et al 2011), implicating these microbes in nutrient provisioning to the host. Dietary supplementation of the microbe-free host with these essential nutrients either completely or partially rescues the phenotype of the untreated host bearing its symbiotic microorganisms (Puchta 1956, Nogge 1976, Mittler 1971).

In aphids, which require the resident bacterium *Buchnera aphidicola* to provide essential amino acids, the metabolic integration of the host and the microbe is more extensive than previously observed in other host-microbe systems. *Buchnera* has lost several key genes that code for enzymes in essential amino acids biosynthesis (Wilson et al 2010). These missing genes code for the terminal reactions of the branched chain amino acids (isoleucine, leucine, and valine) and phenylalanine, and the proximal reactions for isoleucine and methionine synthesis. Previous work (Chapter 3) has shown that the aphid mediates these missing reactions with animal enzymes of equivalent substrate specificity outside of *Buchnera*, but within the

specialized cells that house the bacteria (bacteriocytes). This integration of metabolic capacities leads to one important question: how does the host, or the microbe, respond to an increase in demand for essential amino acids?

The *Buchnera* genome does not contain regulatory elements and is believed to be unable respond to a wide range of biochemical cues. The lack of transcriptional response to changes in dietary nutrients has been experimentally validated for essential amino acid synthesis in several studies (Moran et al 2003, Moran et al 2005, Reymond et al 2006) though one study finds a downregulation of *Buchnera* leucine biosynthetic genes when the aphid was provisioned with leucine at high concentrations (Viñuelas et al 2011). Aside from the mixed results from transcriptomic studies, metabolic models for *Buchnera* show that varying the levels of certain substrates (notably 2-oxobutanoate, homocysteine, glutamine, glutamate, asparagine, and aspartate) should have a demonstrable impact on the production of essential amino acids (Macdonald et al 2011). These substrates are all host derived, so ultimately the host should be able to vary the production of essential amino acids by varying the production rate of the substrates. The following study is an exploratory attempt to determine whether *Buchnera* and their hosts can respond to changes in diet. This exploration was approached by determining whether isoleucine and methionine biosynthetic pathways can vary their output based upon the concentration of input, whether deletion of isoleucine or methionine from the diet has an appreciable impact on the amino acid composition of bacteriocytes, and whether dietary deletion has any impact of the protein composition of bacteriocytes.

4.2 Materials and Methods

4.2.1 Aphid rearing

Aphids were reared as previous described in Chapter 3.

4.2.2 Diet composition and tissue extraction

Aphids were reared from 2-days-old until 7-days-old on chemically defined diets prepared following the procedure and composition of Prosser and Douglas (1992), with 0.5 M sucrose and 0.15 M amino acids. Single deletions of amino acids were made where noted with a complete amino acid diet as a control. Bacteriocytes from 30 aphids were dissected in phosphate buffered saline (PBS – 0.15 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, pH 7.4), immediately flash frozen in liquid nitrogen and stored at -80ºC. Whole aphid extractions were performed on 5 aphids which were flash frozen and immediately stored at -80ºC. Prior to preparation of samples for UPLC analysis, samples were thawed on ice.

4.2.3 Proteomic analysis by nanoLC-LTQ-Orbitrap

Aphid bacteriocytes (dissected in PBS from 115 aphids reared on complete diet or amino acid deletion diets – cysteine, isoleucine, leucine, methionine, phenylalanine/tyrosine, lysine, valine) were solubilized in SDS-PAGE loading buffer (125 mM Tris-HCl pH 6.8, 10% v/v ß-mercaptoethanol, 20% v/v glycerol, 4% w/v SDS, few grains of bromophenol blue). All samples in loading buffer were incubated at 90^{o}C for 5 min prior to separation by SDS-PAGE (10-14% acrylamide). Each gel lane was cut in 10 slices, proteins were reduced, alkylated and digested with trypsin and peptide extracted as in Zybailov et al (2009) and Poliakov et al (2011). The proteomic pipeline is outlined in Chapter 2 and was followed for this analysis of bacteriocytes.

4.2.4 Amino acid production with varying substrate concentration

Bacteriocytes were dissected from 7-day-old larval aphids in extraction medium (28 mM glucose, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.25 M sucrose, 50 mM NaH_2PO_4, 13

mM $K_2H_2PO_4$, pH 7.5), lysed by pipetting 4-6 times, and centrifuged at 1000 *g* for 5 minutes at 4˚C and then resuspended. The *Buchnera* cells were quantified by hemocytometer counts at x400 magnification, and diluted to 4 x 10^8 *Buchnera* ml^{-1} in extraction medium. To initiate the release experiment, 5.5 µl reaction medium was added to 6 replicate samples of 5.5 µl bacterial suspension. The standard reaction medium comprised the extraction medium supplemented with 2 mM glutamate and glutamine and cystathionine (at 0.25, 0.1, 1, and 3 mM) or homoserine (at 10, 20, and 30 µM) for analysis of methionine or isoleucine flux, respectively. At 5 minute intervals over 30 minutes, one tube was centrifuged at 1000 *g* for 70 seconds, and 11 µl supernatant was immediately flash-frozen in liquid nitrogen and stored at -80^{o}C. The experiments were conducted at 22.5^{o}C, and were repeated 5 times at different times on different sets of aphids.

4.2.5 UPLC analysis

Amino acid content was performed using methods outlined in Chapter 3, with the addition of homocysteine, cystathionine, and homoserine to the standards.

4.2.6 Statistical Analyses

Amino acid composition was analyzed by t-tests with a Bonferroni correction for 23 comparisions ($p = 0.0021$). The dependence on substrate concentration for product formation was analyzed by linear regression (ANOVA with the slope = 0 as the null, $p < 0.05$). For analysis of the proteome datasets the adjSPC from each sample were normalized by total number of spectral counts (NadjSPC). Pair-wise correlation coefficients between NadjSPC values of different biological replicates were calculated. The correlation coefficients were high (>0.98) between all replicates within each treatment (diet). The correlation coefficients decreased when the comparison involved only subsets of values in the lower range of NadjSPC, but they are still

considerably high (around 0.7) for NadjSPC values in the range of 0.001 to 0.01 (i.e. proteins that contribute 0.1-1% of bacteriocyte protein mass). This indicated that the reproducibility of the data is high, suitable for more advanced statistical analysis using NadjSPC values. Normalized values (NadjSPC) were used for statistical analysis using multidimensional ANOVA using the statistical platform "R".

4.3 Results

4.3.1 Pathways responsive to varying substrate concentration

The supplementation of cystathionine promotes methionine release (Chapter 3). The concentration of cystathionine in the bacteriocyte (0.2 +/- 0.1 nmol/mg protein) is much lower than other amino acids and is severely depleted in the bacteriocyte when compared to the whole body (9 +/- 0.6 nmol/mg protein in whole aphids). This is suggestive of a very small pool that is almost immediately shunted down the methionine biosynthetic pathway (methionine = 9 +/- 0.2 nmol/mg protein). By supplementing with cystathionine in the medium to above physiological levels (starting concentration without exogenous supplementation is 1.58 µM cystathionine) the study was able to demonstrate that the amount of product released is dependent on the amount of substrate provided (Figure 4.1A, $r^2 = 0.607$).

The requirement for *Buchnera* preparations for exogenous homoserine to promote the production of isoleucine have been previously addressed (Chapter 3). The concentration of homoserine is exceedingly low in the bacteriocyte (9 +/- 0.7 nmol/mg protein), especially when compared to the predicted precursors asparate (64 +/- 2.5 nmol/mg protein) and asparagine (66 +/- 2.7 nmol/mg protein). This relatively small pool (starting concentration without exogenous supplementation is 2.9 µM) is suggestive that the homoserine pool has a fast turnover and is readily converted to 2-oxobutanoate. The supplementation of exogenous homoserine above

physiological levels (Figure 4.1B, $r^2 = 0.376$) demonstrated that the amount of product release is dependent on the amount of substrate provided.

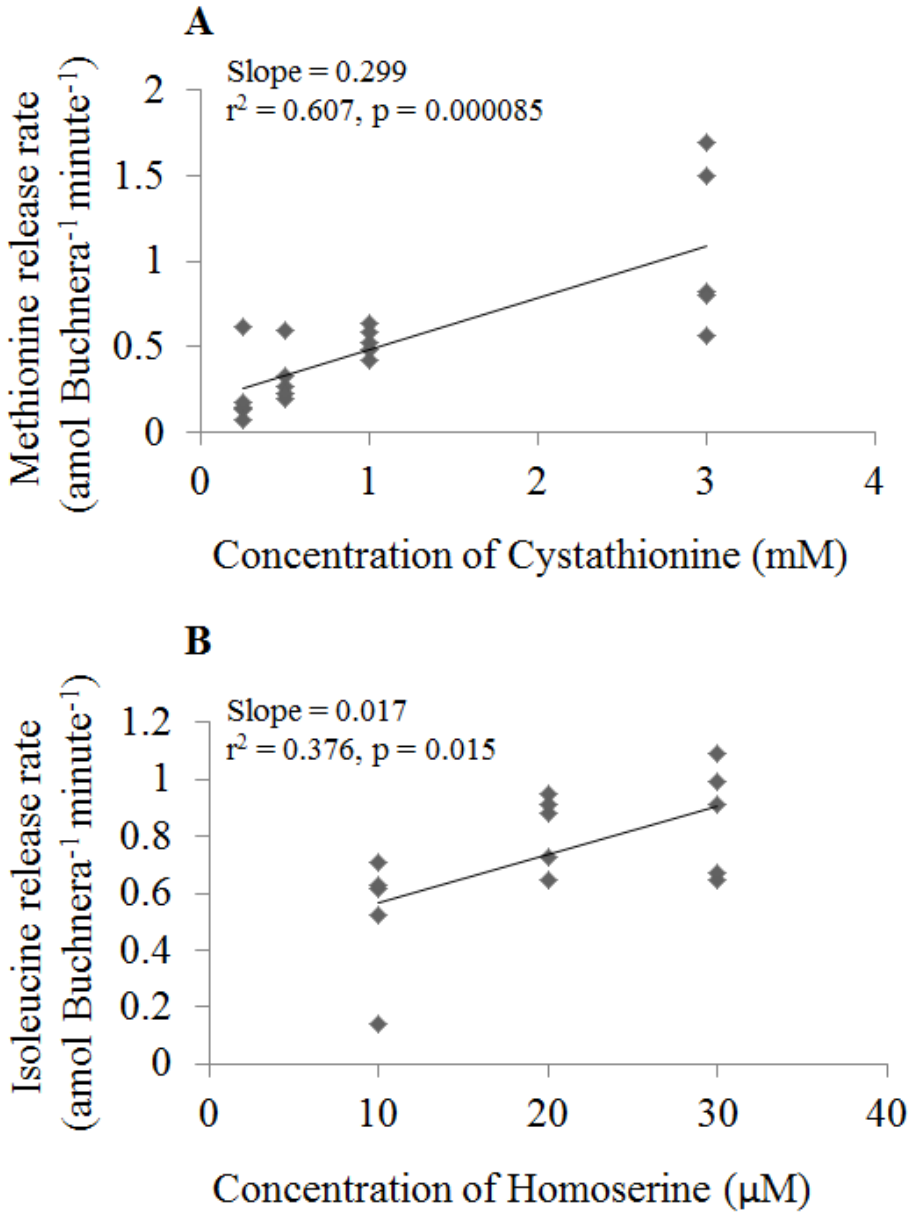


Figure 4.1 Methionine and Isoleucine release from *Buchnera* with access to the host fraction of the bacteriocyte given varying concentrations of cystathionine (A) and homoserine (B). (Linear regression, ANOVA $p < 0.05$)

4.3.2 Amino acid Composition with Dietary Deletions

The impact of single deletions of either isoleucine or methionine on the amino acid composition of bacteriocytes and entire aphids was observed. The greatest differences were observed at the bacteriocyte level, with multiple non-essential and essential amino acids showing differential concentrations. Bacteriocytes from aphids reared without dietary isoleucine showed a significant increase in the asparagine, glutamine, leucine, methionine, and homocysteine (hcys) concentrations relative to protein (Table 4.1, t-tests, $p < 0.0021$). The only amino acid with a significantly lower concentration was isoleucine itself (Table 4.1, t-tests, $p < 0.0021$), indicating that a portion of this amino acid may be of dietary origin rather than symbiotic origin. The increased concentration of glutamine and asparagine is especially important as these two non-essential amino acids are predicted to be important substrates for essential amino acid biosynthesis (Thomas et al 2009) and leucine and methionine are essential amino acids predicted to require host enzymatic input. Bacteriocytes from aphids reared without dietary methionine follow similar, yet slightly distinct, trends as bacteriocytes from aphids reared without isoleucine. Bacteriocytes from aphids reared without dietary methionine showed a significant increase in the aspartate, glutamine, isoleucine, leucine, methionine, threonine, and homocysteine concentrations (Table 4.2, t-tests, $p < 0.0021$). The increase in aspartate and glutamine, like the previous experiment, is important as these amino acids are predicted to be important substrates for essential amino acid biosynthesis (Thomas et al 2009). The increased concentration of isoleucine, leucine and methionine indicates that bacteriocytes from aphids reared without methionine may be responsive across divergent pathways to a drop in the methionine content of the diet. This response focuses on non-essential amino acids required for essential amino acid synthesis, as well as essential amino acids predicted to have shared

metabolism between the aphid and *Buchnera*. Aside from the amino acids predicted to require host enzymes (isoleucine, leucine, homocysteine, methionine), the essential amino acid threonine has an increased concentration relative to protein content. As threonine, like isoleucine, is synthesized from aspartate, the increased concentration of this non-essential amino acid may lead to an increased production rate of threonine.

Table 4.1. Amino acid composition (nmol mg^{-1} protein) of bacteriocytes from aphids reared on complete and isoleucine deletion diets (mean +/- s.e.m, n = 10, t-tests, * = $p < 0.0021$)

Amino acid	Minus Isoleucine	Control	p-value
Ala	137 ± 5.3	123 ± 2.2	p = 0.0226
Asn	90 ± 5.5	66 ± 2.7	p = 0.0011*
Asp	79 ± 4.3	64 ± 2.5	p = 0.0068
Cys	20 ± 0.9	19 ± 1.2	p = 0.5226
Glu	236 ± 9.8	205 ± 4.4	p = 0.0109
Gln	480 ± 30.0	339 ± 14.1	p = 0.0005*
Gly	42 ± 1.2	39 ± 1.0	p = 0.0453
Pro	5 ± 0.3	5 ± 0.2	p = 0.0179
Ser	66 ± 2.7	58 ± 1.1	p = 0.0160
Tyr	28 ± 3.2	28 ± 4.3	p = 0.9062
Arg	96 ± 8.9	105 ± 4.0	p = 0.3449
Ile	10 ± 0.3	18 ± 0.9	p = 0.0001*
Leu	37 ± 1.5	29 ± 0.6	p = 0.0001*
Val	4 ± 0.3	4 ± 0.3	p = 0.7362
His	87 ± 5.0	77 ± 3.1	p = 0.0899
Lys	183 ± 8.1	169 ± 5.8	p = 0.1669
Met	11 ± 0.3	9 ± 0.2	p < 0.0001*
Phe	5 ± 0.1	5 ± 0.4	p = 0.8125
Thr	43 ± 4.3	29 ± 1.1	p = 0.0054
Trp	16 ± 1.5	16 ± 0.6	p = 0.7276
Cysta	1 ± 0.2	0.2 ± 0.1	p = 0.0507
Hcys	20 ± 0.8	16 ± 0.4	p < 0.0001*
Hser	11 ± 1.7	9 ± 0.7	p = 0.2120

Table 4.2. Amino acid composition (nmol mg^{-1} protein) of bacteriocytes from aphids reared on complete and methionine deletion diets (mean +/- s.e.m, n = 10, t-tests, * = $p < 0.0021$)

Amino acid	Minus Methionine	Control	p-value
Ala	126 ± 6.1	123± 2.2	p = 0.5897
Asn	73 ± 5.3	66 ± 2.7	p = 0.2663
Asp	88 ± 5.2	64 ± 2.5	p = 0.0006*
Cys	16 ± 0.9	19 ± 1.2	p = 0.0377
Glu	214 ± 12.2	205 ± 4.4	p = 0.5277
Gln	487 ± 36.5	339 ± 14.1	p = 0.0013*
Gly	45 ± 2.1	39 ± 1.0	p = 0.0111
Pro	6 ± 0.6	5 ± 0.2	p = 0.0117
Ser	58 ± 3.3	58 ± 1.1	p = 0.9397
Tyr	46 ± 5.6	28 ± 4.3	p = 0.0169
Arg	97 ± 5.0	105 ± 4.0	p = 0.1936
Ile	27 ± 1.3	18 ± 0.9	p < 0.0001*
Leu	38 ± 1.6	29 ± 0.6	p < 0.0001*
Val	4 ± 0.1	4 ± 0.3	p = 0.9418
His	88 ± 8.5	77 ± 3.1	p = 0.2250
Lys	148 ± 6.3	169 ± 5.8	p = 0.0299
Met	16 ± 1.1	9 ± 0.2	p < 0.0001*
Phe	5 ± 0.2	5 ± 0.4	p = 0.9220
Thr	49 ± 2.5	29 ± 1.1	p < 0.0001*
Trp	15 ± 1.3	16 ± 0.6	p = 0.4788
Cysta	1 ± 0.2	0.2 ± 0.1	p = 0.0977
Hcys	32 ± 2.2	16 ± 0.4	p < 0.0001*
Hser	14 ± 1.7	9 ± 0.7	p = 0.0078

4.3.3 Proteomic analysis of Bacteriocytes

To determine which proteins were involved with the changes in amino acid composition in bacteriocytes from aphids reared on diets with either methionine or isoleucine deleted, bacteriocytes were isolated from aphids reared on deletion diets (cysteine, methionine,

isoleucine, leucine, valine, phenylalanine/tyrosine, lysine) and compared to bacteriocytes from aphids reared on a complete diet. The workflow for the comparative proteome analysis of isolated bacteriocytes is summarized in Figure 4.2A. The proteins were separated by SDS-PAGE (Figure 4.2B), in-gel digested and analyzed by data-dependent tandem mass spectrometry (MS/MS) analysis using a nanoLC-ESI-LTQ-Orbitrap mass spectrometer. The MASCOT search results were processed as in Poliakov et al (2011). Briefly, the highest scoring gene model for each identified protein was selected and filtering to reduce false positive identifications was performed. This resulted in identification of 1030 aphid proteins (out of 34616 predicted aphid proteins) and 406 *Buchnera* proteins (out of 579 predicted proteins). The proteome data obtained in this work was compared with the bacteriocyte proteome of plant-reared aphids published in previous work (Poliakov et al 2011). The overall number of identified proteins is higher in this work, which is likely a reflection of the fact that 8 times more bacteriocyte samples were analyzed compared to the previous work. However, the relative mass of aphid and *Buchnera* proteins (expressed as NadjSPC) were very similar for bacteriocytes from diet-reared aphids (this study) and plant-reared aphids (Poliakov et al 2011) (Figure 4.3), indicating that protein composition, as assessed from the proteome, is not substantially affected by diet. This is important to note, as previous studies have shown that diet reared aphids have distinct transcript profiles from their plant-reared counterparts (Bouvaine et al 2012).

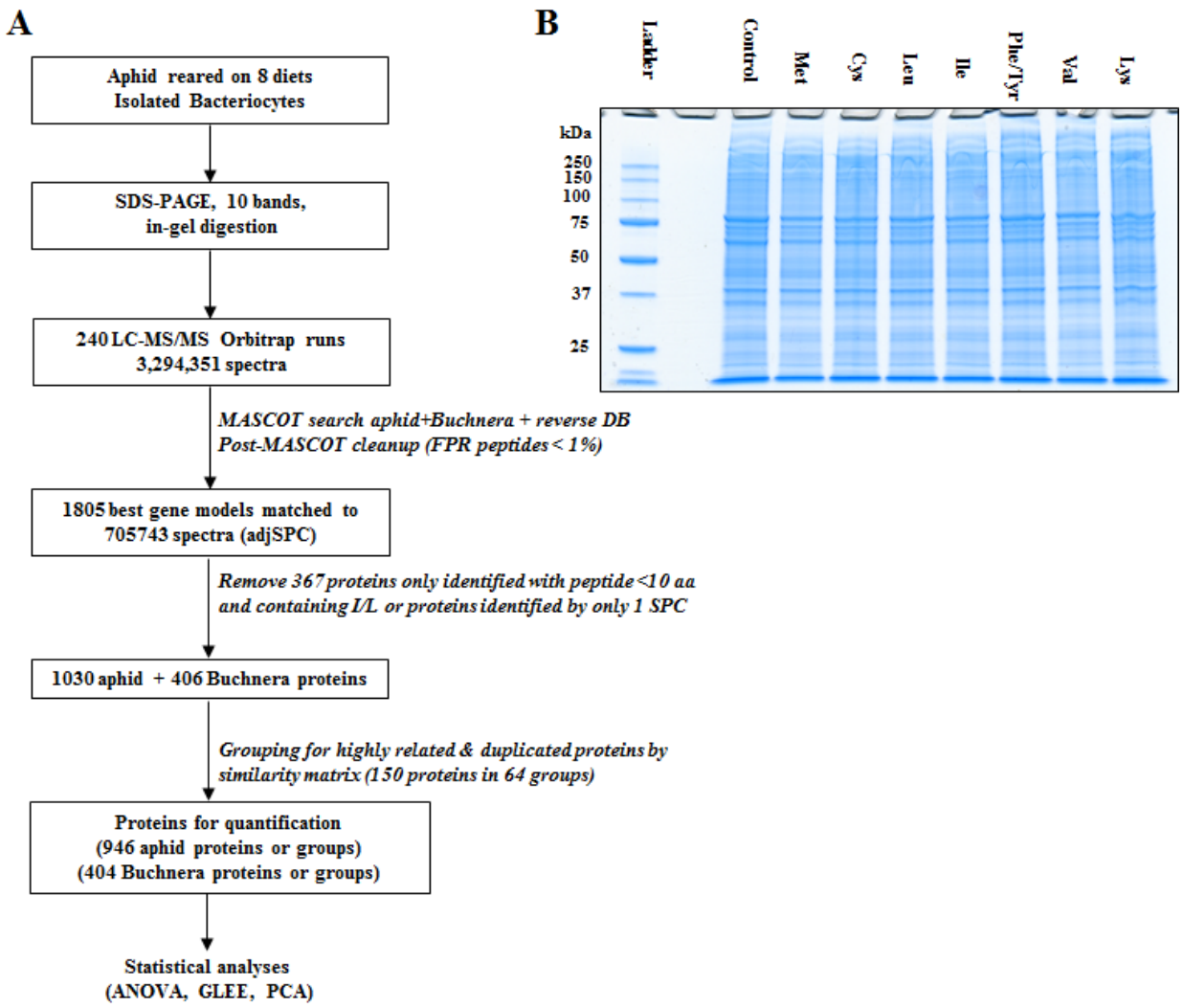


Figure 4.2. Workflow for proteomic experiments. A) Workflow from dissections to statistical analyses B) SDS-PAGE of protein samples from different treatments (Met = methionine, cys = cysteine, leu = leucine, ile = isoleucine, phe/tyr = phenyalalnine/tyrosine, val = valine, and lys = lysine)

A

Bacteriocyte Proteomes

Parameter	Diet reared	Plant reared*
Total # of proteins	1436	1213
# of aphid proteins	1030	827
# of Buchnera proteins	406	386
Total NadjSPC aphid proteins	0.3	0.3
Total NadjSPC Buchnera proteins	0.7	0.7

*Poliakov et al 2011

B

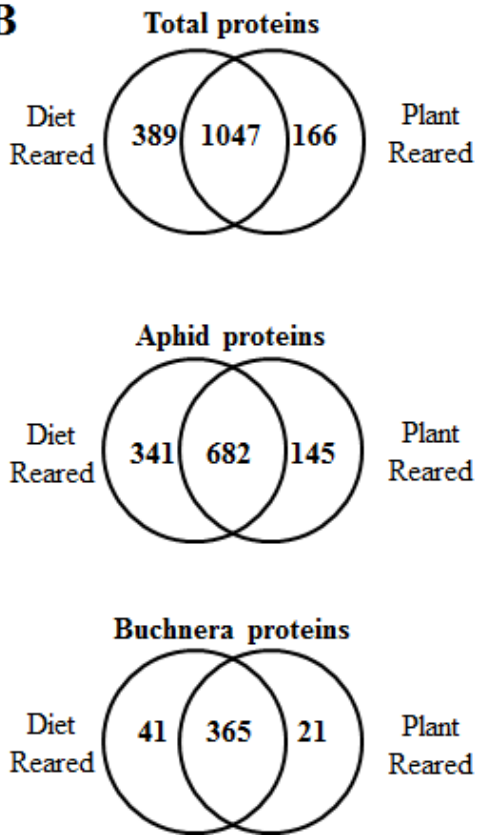


Figure 4.3. Comparison of bacteriocytes from diet- and plant-reared aphids. A) Number of proteins and portion of NadjSPC attributed to aphid and *Buchnera* proteins B) Overlap of proteins in bacteriocytes between diet- and plant-reared aphids.

Our quantitative analysis of the data was based on spectral counting (SPC), which is an established technique for relative mass spectrometric quantification (Old et al 2005, Zybailov et al 2005). The relative amount (mass) of each identified protein within each replicate was calculated based on the adjusted number of matched MS/MS spectra (adjSPC), normalized by the sum of adjusted SPC in the replicate, yielding NadjSPC. To remove ambiguity associated with shared peptides, proteins that shared more than ~80% of their matched MS/MS spectra were grouped using the workflow developed in (Friso et al 2010). In total, 146 aphid proteins were assigned into 64 groups, 140 of which were aphid proteins (the few *Buchnera* proteins needing to be grouped is consistent with the low proteome complexity and absence of gene duplications in *Buchnera*). The resulting number of proteins for quantitation was 1350 proteins or protein groups (Figure 4.1A), of which 946 were aphid proteins and 404 *Buchnera* proteins.

To test for variation in protein abundance across diet treatments, multidimensional ANOVA was applied. At the probability threshold of 0.01, two of the 1,435 proteins with statistically significant variation across diet treatments were identified: LOC100163029 ($p=10^{-6}$, heterogenous nuclear ribonucleoprotein H, hnRNP H), which contributes to the hnRNP complex in mRNA processing, and LOC100165171 (p=0.008, adenylyl cyclase-associated protein 1: p=0.009), contributing to cAMP signaling. These differences are considered likely to have arisen by chance, and neither explains the differences in amino acid composition of bacteriocytes from aphids reared on deletion diets.

4.4 Discussion

Our preliminary work towards determining whether the *Buchnera* and the host are responsive to changes in diet met with mixed results. Similar to other experiments, there is no

change in the *Buchnera* corresponding to changes in amino acid composition of the diet. Due to previous findings (Chapter 3) that exogenous host enzymes mediate certain reactions in essential amino acid synthesis, the expectation was that these aphid enzymes would increase in abundance to drive these pathways to producing more essential amino acids. In this regard, the hypotheses did not bear fruit. There were no changes in the abundance of these enzymes (branched chain aminotransferase, cystathionine-γ-lyase, and aspartate aminotrasferase) in the bacteriocytes on amino acid deletion diets. This suggests that the translational responses of the host portion of the bacteriocyte are also non-responsive to changes in dietary signals.

Using isoleucine and methionine production from bacteriocyte preparations exposed to variable substrate concentrations illustrates that these pathways can perform (in a linear manner) over a wide range of substrate concentrations. Once absolute concentration of these substrates is determined for bacteriocytes, whether these pathways are underperforming in a typical bacteriocyte can be determined. The amino acid composition of bacteriocytes from aphids reared on deletion diets was determined in order to observe substrate concentration. For methionine, the significant increase in homocysteine (and methionine) fits with the expectation that the production of methionine is tightly linked to the amount of precursor generated in the system. For isoleucine, the increase in asparagine (which can be converted to aspartate by an asparaginase) may indicate that early precursors for this pathway lead to an increased production of isoleucine. These findings will need to be verified by further labeled experiments showing the production of essential amino acids increases as amino acids are deleted from the diet.

These pilot experiments were important for a) ruling out responsiveness of candidate aphid enzymes to dietary deletion and b) verifying that metabolite concentrations change in bacteriocytes depending on diet. Prior to these studies,it was not known whether aphid enzymes

localized to the bacteriocyte were responsive to the diet and whether the amino acid profiles of bacteriocytes shifted based upon dietary deficiency.

CHAPTER 5

RESEARCH SUMMARY AND FUTURE DIRECTIONS

5.1 Research summary

The research undertaken in this thesis provided a unique understanding of the metabolic interactions between the pea aphid and their resident bacteria *Buchnera aphidicola*. The proteomic study of the bacteriocyte (Chapter 2) validated candidate enzymes laid out in Wilson et al (2010) and hinted at a greater degree of metabolic integration between the host and symbiotic partner. Enzymes believed to be pivotal for mediating missing reactions in the *Buchnera*, as well as producing the substrate required for essential amino acid production, were enriched in the bacteriocyte relative to the whole aphid. The medium that was developed to test the metabolic capacity of the host and bacterial components of the bacteriocyte was generated from this work, as well as work outlined in Thomas et al (2009). It was further shown that exogenous enzymes were required by the *Buchnera* to produce the branched chain amino acids (using leucine as a proxy), phenylalanine, and methionine. The candidates were narrowed down for proximal enzymes in isoleucine and methionine synthesis by determining which substrate was required for production of their respective amino acids (homoserine for isoleucine, cytathionine for methionine). The final chapter of my thesis approached the subject of how the symbiosis responds to dietary deletions of certain amino acids. Like other studies, this study found that the symbiosis does not respond readily by transcriptional or translational means, but unlike other researchers an increase in the concentration of precursor (aspartate, asparagine, glutamine, homocysteine) in bacteriocytes from aphids facing dietary deletions was noticed. This, coupled with an experiment showing increasing the levels of substrate in the system to above physiological levels yields more product, support the hypothesis that the production of

essential amino acids is regulated by the provisioning of the bacteriocytes with the appropriate substrate.

5.2 Future directions

There are multiple directions the project can take from here, yet three are readily available for pursuit. I would determine the identity of the enzyme that is converting cystathionine to homocysteine by fractioning the protein of the bacteriocyte and looking for activity in each of the fractions. Once activity has been establishedthe identity of the enzyme could be determined by MS. As animals have no known genes that can mediate this reaction, and *Buchnera* doesn't contain any genes that can either, I potentially have a novel annotation. Furthering the study of shared metabolism between insects and their bacterial symbionts, other bacterial symbionts of Sternorrhynchan insects are missing the same genes as *Buchnera*. Now that the techniques have been developed, a broad assay of these insects and symbionts can be undertaken, and I can determine whether convergent evolution has occurred at the metabolic level. I predict that these other insects mediate these missing reactions in a similar manner as the pea aphid. In terms of responding to diets depleted of certain amino acids, I would check fo changes in transcription/protein content in tissues other than the bacteriocytes (as other tissues may cause the changes I saw in metabolite concentration) as well as assay the bacteriocytes for essential amino acid releasal and compare rates. I would expect bacteriocytes reared on isoleucine or methionine depleted diets would have higher rates of release of certain essential amino acids.

APPENDIX A

NITROGEN EXCRETION AND ASSIMLATION IN PEA APHIDS

A.1 Introduction

Nitrogen homeostasis is of key importance to aphids, which feed on a diet low in essential amino acids. Several studies have implicated ammonia, glutamine, or arginine as the major nitrogenous waste excreted by aphids(Wilkinson and Douglas 1995, Prosser and Douglas 1992) . A replication of certain findings from these studies (increase nitrogen via ammonium chloride supplementation and testing responsiveness of aphids to nitrogen supplementation; Wilkinson and Douglas 1995)) was performed and found that the addition of ammonia to the diet does not increase the global nitrogen content of the aphid, and that only increasing the concentrations of amino acids in the diet has this effect.

A.2 Materials and Methods

A.2.1 Aphid rearing

Aphids were reared according to protocols laid out in Chapter 3.

A.2.2 Diet and honeydew extraction

Unless otherwise stated, aphids were reared from 2 days old until 7-days old on chemically defined diets prepared following the procedure and composition of Prosser and Douglas (1992), with 0.5 M sucrose and 0.15 M amino acids. Honeydew was extracted by suspending diet-feeding aphids over water-saturated mineral oil for 24 hours (day 6 – day 7). Diet was extracted from between the membranes aphids feed from on day 7 to reflect the nutritional quality of the diet at the same time as the honeydew. Assimilation efficiency was calculated by dividing the concentration of amino acids in the honeydew by the concentration of amino acids in the diet.

A.2.3 Supplementation of Aphid diet with Nitrogen

The aphid diet was supplemented with extra nitrogen in two forms: diet supplemented with 10 mM NH_4Cl (as this was not deleterious to aphid growth – Fig. 2) and varying the amino acid content (50, 100, 150, 200 mM with arginine deleted). For the NH_4Cl diets, honeydew was extracted as previously stated, as well as the aphid tissues from 20 aphids (in PBS) and whole aphids to compare nitrogen content. Hemolymph was also extracted by submerging aphids in water-saturated mineral oil, cutting their legs off, and allowing to bleed out for 1 hour.

A.2.4 Glutamine synthetase activity

The glutamine synthetase activity was determined by a transferase assay that measures the synthesis of γ-glutamyl hydroxamate from glutamine spectrophotometrically. This assay was modified from Pahel et al (1982) to accommodate small samples. Fat body, bacteriocytes, and embryos were dissected in Buffer A (0.25 M sucrose, 20 mM imidazole-HCl pH 7.4) from 60 aphids that had either been fed on control diet or 10 mM NH_4Cl supplemented diet. The tissue samples, as well as samples containing single aphids from each diet group (3 aphids per replicate) were diluted to 225 µl with Buffer A and hand-homogenized and centrifuged at 13000 rpm for 5 minutes at 4°. A 25 µl aliquot of the supernatant was set aside for a protein assay (Bio-RAD), and the remaining 200 µl was added to 200 µl of Buffer B (0.6 $MnCl_2 * 4H_2O$, 0.25 M sucrose, 20 mM dithiothreitol, 20 mM imidazole-HCl pH 7.4, and 100 µg CTAB/ml) and vortexed for approximately 10 seconds. Sample was then split up into 3 tubes (experimental, heat treated control, and minus substrate (ADP and glutamine) control), each at 125 µl. To the 125 µl of extract 125 µl of the reaction buffer (0.3 mM $MnCl_2$, 40 mM potassium arsenate, 40 mM NH_2OH, 0.8 mM ADP, 40 mM glutamine, 50 mM imidazole-HCl pH 7, and 100 µg CTAB/ml) was added. The reacting mixture was incubated at 30°C for 60 minutes at which point the

reaction was stopped with the 0.5 ml stop buffer (0.13 M HCl, 3.4% $FeCl_2$, 2% trichloroacetic acid). The samples were compared to standards of 0, 20, 40, 60, 80, and 100 µg/ml L-glutamic acid gamma monohydroxamate in a total volume of 750 µl (with 500 µl of the total volume as the stop buffer) at 540 nm. A unit of activity is equivalent to a nanomole of product (γ-glutamyl hydroxamate) formed per minute.

A.2.5 UPLC analysis

The amino acid content of the samples were quantified using the AccQ Tag derivatization kit (Waters) by UPLC with PDA detector (Waters Acquity). Tissues, and aphids, were homogenized in PBS (recipe) to which an equal volume of 40 mM HCl was added. To 5 µl of either honeydew, hemolymph, or diet, 15 µl of 30 mM HCl was added. Following incubation on ice for 30 minutes, the sample was centrifuged at 18000 *g* for 10 minutes at 4˚C, and the supernatant was filtered through a 0.45 µm filter plate (Millipore) by centrifugation at 1500 *g* for 10 min. The filtrate was measured for protein content (Bio-Rad) where applicable to standardize values after UPLC analyses. The filtrate (2.5 µl) was derivatized with AccQ Tag (Waters), following manufacturer's protocol, and injected into Waters Acquity UPLC with PDA detector and AccQ-Tag Ultra 2.1 x 100 mm column. The gradient was: 0-0.54 min, 99.9% A 0.1% B; 0.54-5.74 min, 90.9% A and 9.1% B; 5.74-7.74 min, 78.8% A 21.2% B; 7.74-8.04 min, 40.4% A 59.6% B; 8.04-8.64 min, 10% A 90% B; 8.05-8.64 min 10% A 90% B; 8.64-8.73 min 99.9% A 0.1% B; 8.73-9.50 min, 99.9% A 0.1% B (linear between each time point), where A is 10% AccQ-Taq Ultra Eluent A in water, and B is Accq-Taq Ultra Eluent B. Amino acids were determined by comparison to standards: 1, 5, 10, 50 and 100 pmol amino acids μl^{-1} (Waters amino acid hydrolysate standard #088122, supplemented with asparagine, tryptophan, and glutamine).

A.3 Results

A.3.1 Assimilation efficiency of amino acids

The assimilation of amino acids, which includes absorption and digestion, vary amongst the different amino acids. Overall, most amino acids have assimilation efficiencies greater than 92%, with the greatest efficiency observed for aspartate with an assimilation efficiency of 99.7% (Figure A.1). Surprisingly, essential amino acids believed to be at low abundance in the phloem sap (Isoleucine, leucine, lysine, methionine, phenylalanine, threonine, and valine) have a relatively high assimilation rate, indicative that the gut of the pea aphid may have relatively non-specific transporters or digestive enzymes. The few compounds with low assimilation efficiencies (ammonia, arginine, histidine, and tryptophan) are relatively unsurprising due to the fact that these compounds should not be routinely observed in phloem sap. Yet arginine appears to increase in concentration as it is excreted, as opposed to all the other compounds assayed. As arginine has a relatively high amount of nitrogen for an amino acid, it has been implicated with nitrogen excretion in phloem feeders. The observation of a nearly 50% increase in the concentration of arginine in the honeydew compared to the diet gives a unique opportunity to determine whether arginine is an excretory product, or merely fails to assimilate into the aphid.

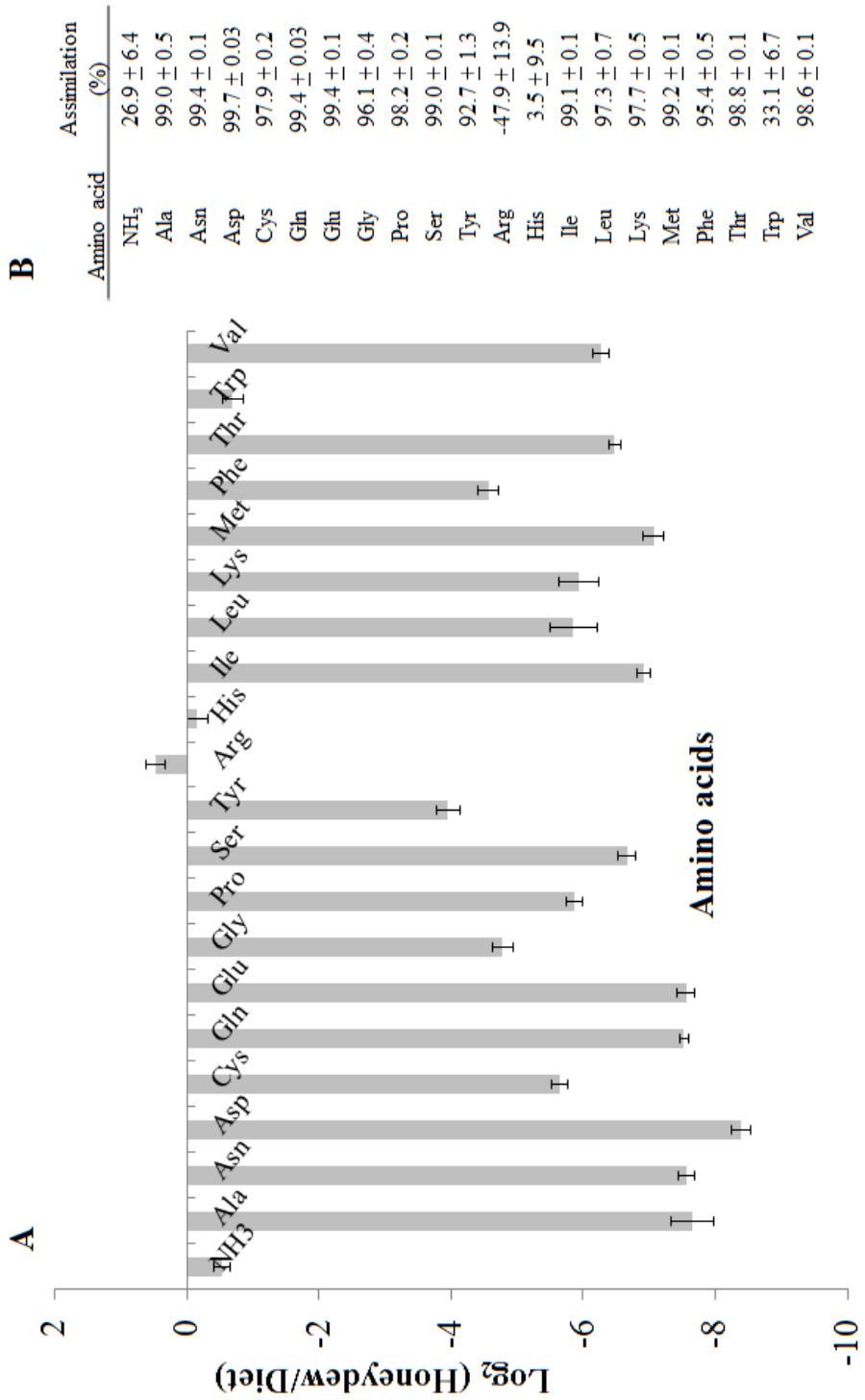


Figure A.1. Assimilation efficiency of amino acids and ammonia in the pea aphid. A) The log2 of assimilation efficiency (concentration of AA in honeydew/concentration of AA in diet), B) The percent assimilation of each amino acid (concentration of AA in honeydew/concentration of AA in diet * 100)

A.3.2 Ammonia assimilation and glutamine synthetase activity

An attempt to increase the global nitrogen content of the aphid and measure excretion was attempted via two distinct methods: supplementing the diet with ammonium chloride and increasing the concentration of amino acids. The diet was supplemented with 10 mM NH_4Cl, as this concentration was not deleterious to the growth of the aphid (Figure A.2) and had no effect on protein concentration of the aphid relative to wet weight (data not shown). Surprisingly, the supplementation of the diet with NH_4Cl caused no significant change in any amino acid, and ammonia, in any aphid tissue (Table A.1). To determine whether the low assimilation efficiency of ammonia (26.9 $\pm$ 6.4%, Figure A.1) prevents ammonia from entering into the hemocoel of the aphid, honeydew was assayed from aphids fed on control diet and diet supplemented with 10 mM NH_4Cl. Ammonia had a significantly higher concentration in the honeydew than the control honeydew (3-fold increase, Table A.2). This finding is consistent with an inability of ammonia to cross the gut barrier, yet does not account for the 26.9% assimilation observed in previous experiments. However, the concentration of glutamine significantly increases, with the concentration of glutamate significantly decreasing, when the diet is supplemented with NH_4Cl (Table A.2). An enzyme, glutamine synthetase, fixes an ammonium molecule to glutamate to make glutamine. When aphid tissues were tested for glutamine synthetase activity, the gut (and bacteriocytes) had the highest glutamine synthetase activity of any of the tissues, and that this activity was not responsive to NH_4Cl supplementation of the diet, rather that the activity was constitutive (Figure A.3). This evidence suggests that the aphid is not responsive to exogenous nitrogen in the form of ammonia, and that it cannot be directly utilized by the aphid.

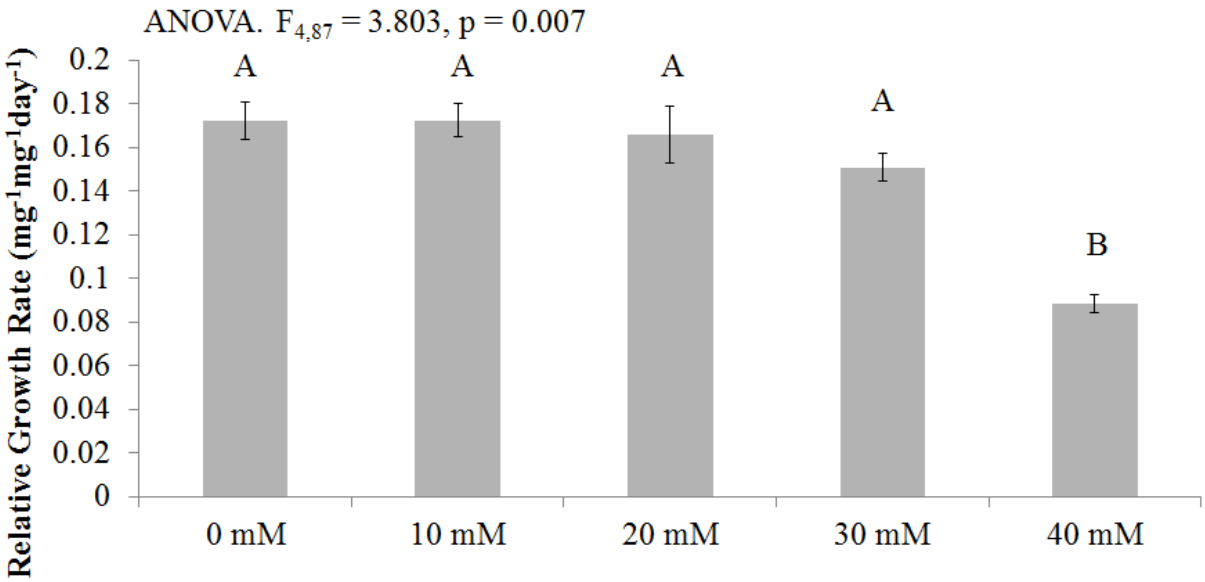


Figure A.2. The relative growth rate of aphids reared on diets with different concentrations of NH_4Cl.

Table A.1. Amino acid concentration in tissues from aphis reared on control diet or 10 mM NH_4Cl diet (nmol amino acid/mg protein, t-tests)

	Aphid			Gut			Fat body		
	Control	NH_4Cl	p-value	Control	NH_4Cl	p-value	Control	NH_4Cl	p-value
NH_3	7.5 ± 0.5	7.4 ± 0.6	0.903488	19.9 ± 2.5	20.9 ± 2.1	0.7659296	29.1 ± 3.6	30.0 ± 4.5	0.88076
Ala	68.0 ± 6.3	65.9 ± 7.6	0.752833	30.0 ± 2.0	26.5 ± 1.5	0.1656869	13.8 ± 1.5	12.6 ± 1.4	0.559875
Asn	89.9 ± 7.0	81.0 ± 7.4	0.873936	8.9 ± 0.5	8.2 ± 0.5	0.3665167	24.2 ± 1.3	22.5 ± 1.8	0.477527
Asp	17.7 ± 1.6	22.1 ± 2.2	0.121408	3.4 ± 0.3	2.8 ± 0.3	0.121098	4.8 ± 0.8	4.2 ± 0.6	0.513514
Cys	1.1 ± 0.1	1.1 ± 0.1	0.792292	2.1 ± 0.3	2.3 ± 0.4	0.5680682	ND	ND	
Gln	119.0 ± 9.9	121.1 ± 12.1	0.581633	14.4 ± 1.1	12.2 ± 1.0	0.1526542	30.6 ± 2.4	32.7 ± 3.4	0.337715
Glu	98.2 ± 9.8	94.5 ± 10.2	0.460388	19.8 ± 1.3	19.3 ± 1.3	0.780756	17.1 ± 1.4	16.4 ± 1.7	0.759799
Gly	17.9 ± 1.5	17.0 ± 1.5	0.680222	12.9 ± 1.0	11.7 ± 0.7	0.3641446	11.9 ± 0.8	10.5 ± 0.8	0.203988
Pro	73.3 ± 6.1	70.0 ± 6.4	0.571225	10.6 ± 0.5	9.5 ± 0.6	0.1989589	13.8 ± 0.9	12.7 ± 1.0	0.418979
Ser	15.5 ± 1.1	16.0 ± 1.0	0.754367	5.6 ± 0.3	5.1 ± 0.3	0.2420516	5.7 ± 0.5	5.2 ± 0.4	0.38593
Tyr	51.2 ± 4.2	54.0 ± 5.4	0.088669	3.0 ± 0.4	2.4 ± 0.2	0.2334298	50.3 ± 5.4	52.9 ± 4.0	0.695151
Arg	85.7 ± 10.3	95.0 ± 12.0	0.086417	9.3 ± 0.8	8.9 ± 0.9	0.7644745	54.1 ± 4.8	59.1 ± 6.0	0.520853
His	46.5 ± 3.5	47.7 ± 3.1	0.198114	14.3 ± 0.6	14.3 ± 0.6	0.9468821	18.8 ± 1.5	19.5 ± 1.9	0.774995
Ile	15.8 ± 1.4	19.9 ± 1.7	0.07099	5.2 ± 0.3	4.6 ± 0.3	0.1804497	4.5 ± 0.4	5.4 ± 0.6	0.192615
Leu	14.1 ± 1.2	17.1 ± 1.4	0.123989	5.2 ± 0.4	4.5 ± 0.3	0.1714726	5.1 ± 0.4	5.9 ± 0.6	0.259222
Lys	54.5 ± 6.7	55.5 ± 5.5	0.493775	5.9 ± 0.6	5.7 ± 0.4	0.7240635	51.8 ± 6.4	59.7 ± 6.4	0.321553
Met	3.4 ± 0.3	4.5 ± 0.4	0.039499	1.4 ± 0.1	1.3 ± 0.1	0.4032363	1.6 ± 0.1	1.8 ± 0.2	0.469347
Phe	7.1 ± 0.5	8.4 ± 0.4	0.070266	1.7 ± 0.1	1.5 ± 0.1	0.2498207	6.9 ± 0.5	7.7 ± 0.6	0.323579
Thr	18.9 ± 2.2	22.1 ± 1.9	0.290834	5.4 ± 0.3	4.9 ± 0.4	0.2446076	6.8 ± 0.6	6.7 ± 0.7	0.914952
Trp	23.3 ± 1.6	29.1 ± 2.0	0.031638	2.4 ± 0.2	2.4 ± 0.1	0.9839592	25.5 ± 1.7	28.1 ± 2.1	0.359758
Val	25.2 ± 2.1	29.1 ± 2.4	0.233355	7.2 ± 0.4	6.4 ± 0.4	0.2105246	6.7 ± 0.6	7.4 ± 0.8	0.519659

	Bacteriocytes			Embryos			Hemolymph		
	Control	NH_4Cl	p-value	Control	NH_4Cl	p-value	Control	NH_4Cl	p-value
NH_3	49.4 ± 6.3	54.5 ± 7.3	0.603	9.6 ± 0.9	8.3 ± 1.0	0.3308835	468.5 ± 28.3	576.7 ± 53.5	0.088445
Ala	12.8 ± 1.5	12.0 ± 1.1	0.662765	51.0 ± 2.7	46.6 ± 4.2	0.3874776	2038.5 ± 108.4	1902 ± 174.1	0.512969
Asn	6.4 ± 0.6	10.8 ± 5.0	0.377844	33.7 ± 2.5	27.4 ± 2.8	0.1039301	936.2 ± 43.3	976.8 ± 66.8	0.61501
Asp	8.8 ± 0.8	8.4 ± 0.7	0.651269	43.6 ± 1.9	40.7 ± 3.1	0.4217211	25.7 ± 1.5	29.5 ± 2.0	0.143507
Cys	2.5 ± 0.2	2.1 ± 0.2	0.128974	0.6 ± 0.1	0.5 ± 0.1	0.4748437	ND	ND	
Gln	9.9 ± 1.0	9.9 ± 0.8	0.957012	42.4 ± 4.2	32.1 ± 3.9	0.0841084	1313.2 ± 61.1	1480.2 ± 115.6	0.215741
Glu	39.0 ± 3.4	36.4 ± 2.6	0.544375	75.9 ± 3.4	60.8 ± 8.5	0.117702	113.3 ± 4.1	114.5 ± 7.5	0.887683
Gly	6.1 ± 0.7	5.4 ± 0.5	0.452705	9.0 ± 0.6	7.5 ± 0.8	0.1345019	231.5 ± 10.0	216.0 ± 18.2	0.466297
Pro	4.1 ± 0.3	3.7 ± 0.3	0.323067	16.6 ± 1.1	14.4 ± 1.5	0.2515035	991.3 ± 39.9	879.6 ± 64.9	0.156322
Ser	7.2 ± 0.9	7.0 ± 0.8	0.891331	9.7 ± 0.5	8.7 ± 0.8	0.2983011	262.1 ± 10.1	269.6 ± 20.7	0.747956

Tyr	1.8 ± 0.4	1.9 ± 0.2	0.954067	5.5 ± 0.4	5.5 ± 0.5	0.9145529	ND	ND	
Arg	11.1 ± 0.9	9.7 ± 0.9	0.268263	3.7 ± 0.2	4.2 ± 0.1	0.3088143	707.1 ± 83.4	507.3 ± 60.0	0.063111
His	8.7 ± 1.0	8.8 ± 0.8	0.935355	12.3 ± 1.9	11.1 ± 1.2	0.4682547	681.9 ± 43.5	585.4 ± 37.2	0.103245
Ile	2.5 ± 0.2	2.4 ± 0.2	0.816669	10.6 ± 0.7	9.7 ± 0.8	0.3634692	622.0 ± 26.8	663.5 ± 58.7	0.528153
Leu	3.7 ± 0.3	3.5 ± 0.2	0.607135	9.6 ± 0.7	8.7 ± 0.7	0.3630296	440.0 ± 21.5	439.2 ± 37.9	0.984092
Lys	23.2 ± 2.4	19.6 ± 2.0	0.265709	6.1 ± 0.5	5.3 ± 0.6	0.3086981	421.5 ± 20.4	418.4 ± 33.3	0.938736
Met	1.5 ± 0.2	1.8 ± 0.5	0.546392	1.7 ± 0.1	1.6 ± 0.2	0.6467034	81.4 ± 3.7	78.5 ± 8.4	0.759945
Phe	1.0 ± 0.2	1.1 ± 0.1	0.95066	1.3 ± 0.1	1.3 ± 0.1	0.9612499	180.8 ± 27.7	145.2 ± 10.8	0.245021
Thr	4.4 ± 0.4	4.0 ± 0.3	0.475783	9.8 ± 0.6	8.3 ± 0.7	0.1072755	557.9 ± 26.4	515.3 ± 49.8	0.45826
Trp	2.3 ± 0.2	2.3 ± 0.2	0.975973	1.9 ± 0.1	1.9 ± 0.1	0.9369178	150.2 ± 10.2	132.1 ± 10.8	0.234998
Val	4.3 ± 0.3	3.8 ± 0.3	0.340681	17.5 ± 1.2	14.8 ± 1.2	0.1311056	745.6 ± 31.0	758.0 ± 59.4	0.8551

Table A.2. Amino acid concentration of honeydew from aphids reared on control or 10 mM NH_4Cl diet (nmol amino acid/ ul honeydew, * = $p < 0.0024$)

	Honeydew		
	Control	NH_4Cl	p-value
NH_3	4447.1 ± 197.1	13128.9 ± 909.5	4.4E-10*
Ala	109.4 ± 60.9	66.2 ± 15.0	0.49153
Asn	173.3 ± 19.3	191.7 ± 18.5	0.49706
Asp	83.6 ± 9.3	291.1 ± 199.2	0.30701
Cys	74.3 ± 5.9	81.2 ± 6.6	0.44127
Gln	219.4 ± 10.9	373.5 ± 25.0	1.0204E-05*
Glu	76.4 ± 7.1	42.5 ± 4.2	0.000306453*
Gly	103.2 ± 10.9	120.1 ± 5.6	0.2309
Pro	196.9 ± 17.1	184.3 ± 15.0	0.58375
Ser	107.8 ± 9.3	116.5 ± 11.1	0.55544
Tyr	63.7 ± 11.3	43.7 ± 2.8	0.09741
Arg	42726.1 ± 4024.8	43403 ± 3620.0	0.90126
His	20906.7 ± 7991.4	22708 ± 2041	0.53975
Ile	133.1 ± 9.2	121.4 ± 9.9	0.39537
Leu	440.4 ± 112.0	610.0 ± 175.1	0.42151
Lys	412.9 ± 98.3	420.3 ± 123.1	0.96289
Met	38.6 ± 4.5	55.5 ± 7.0	0.0525
Phe	247.6 ± 28.6	610.0 ± 175.1	0.42151
Thr	185.8 ± 11.6	176.4 ± 13.3	0.59758
Trp	3414.33 ± 341.5	3409.2 ± 280.9	0.99086
Val	210.9 ± 18.5	200.5 ± 14.2	0.65815

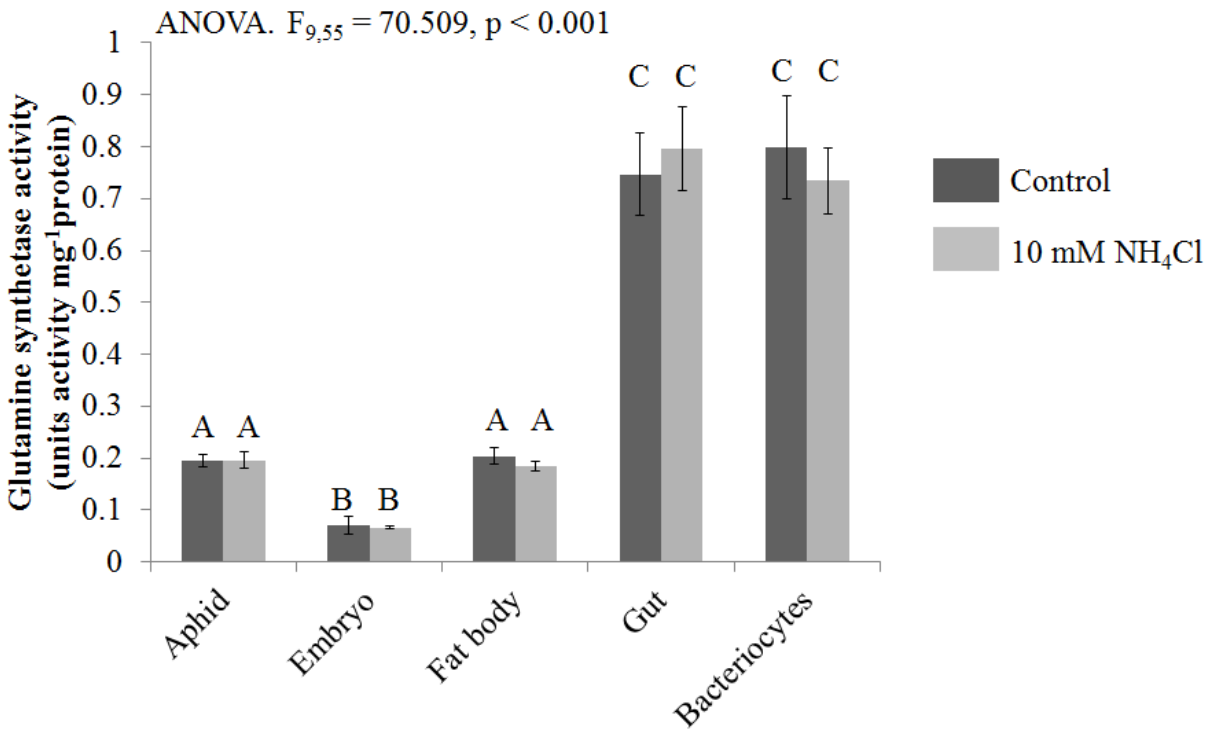


Figure A.3. Glutamine synthetase activity of tissue and aphids reared on control or 10 mM NH_4Cl diets (ANOVA $p < 0.001$, LSD posthoc tests).

A.3.3 Impacts of varying amino acid concentration on hemolymph and honeydew

Supplementing the diet with NH_4Cl had no impact on the free amino acid and ammonia concentrations of the aphid. The amino acid (and ammonia) concentrations of hemolymph and honeydew from aphids reared on 50, 100, 150, and a 200 mM amino acid diet were measured. To determine whether arginine is truly an excretion product, this experiment was conducted with a complete amino acid profile and a diet solely missing arginine. Arginine concentration in the hemolymph does not appear to correlate with amino acid concentration in the diet, indicating that arginine is either not taken up (in a linear fashion) or is immediately broken down when it is assimilated into the aphid (Figure A.4A). The honeydew from the two treatments showed very divergent patterns. When arginine is deleted from the diet (any of the concentrations of amino acids) the honeydew has a barely detectable level of arginine that does not correlate with amino acid concentration of the diet (Figure A.4B). This is indicative that arginine is not excreted from the aphid as nitrogenous waste and is not taken up from the diet. When arginine is present in the diet, the concentration of arginine in the honeydew correlates strongly with the amino acid concentration in the diet, which is to be expected if it is merely passing through the gut.

The correlation of other amino acids (and ammonia) in the hemolymph and honeydew with increasing concentrations of amino acids in the diet was observed to try and understand how an aphid moderates their nitrogen content (Table A.3). As aphids have lost the ability to make uric acid, leaving only ammonia as a potential nitrogenous waste, an expectation that the ammonia concentration will increase with an increased abundance of nitrogen (in the form of amino acids) is presumed. There is no significant correlation between ammonia concentration in either the hemolymph or honeydew and amino acid concentrations in the diet. The other compounds with low assimilation efficiencies (arginine, histidine, and tryptophan), showed a

significant correlation between amino acid content of the diet and concentration in the honeydew. The rates of increase for these amino acids in the honeydew are orders of magnitude higher than any other amino acid, which are indicative that these amino acids are not taken up or digested in the gut but a large proportion are merely passed through. Several of the remaining amino acids exhibited high assimilation rates (Figure A.1), yet do not exhibit a strong increase in their hemolymph concentration when dietary amino acids increase: alanine, cysteine, glutamate, tyrosine, and methionine. Conversely, there are amino acids with high assimilation rates that exhibit an extremely strong increase in their hemolymph concentrations: asparagine and glutamine. This may indicate that certain amino acids are broken down and used to make glutamine and asparagine, potentially as stores for excess nitrogen in the form of these two non-essential amino acids.

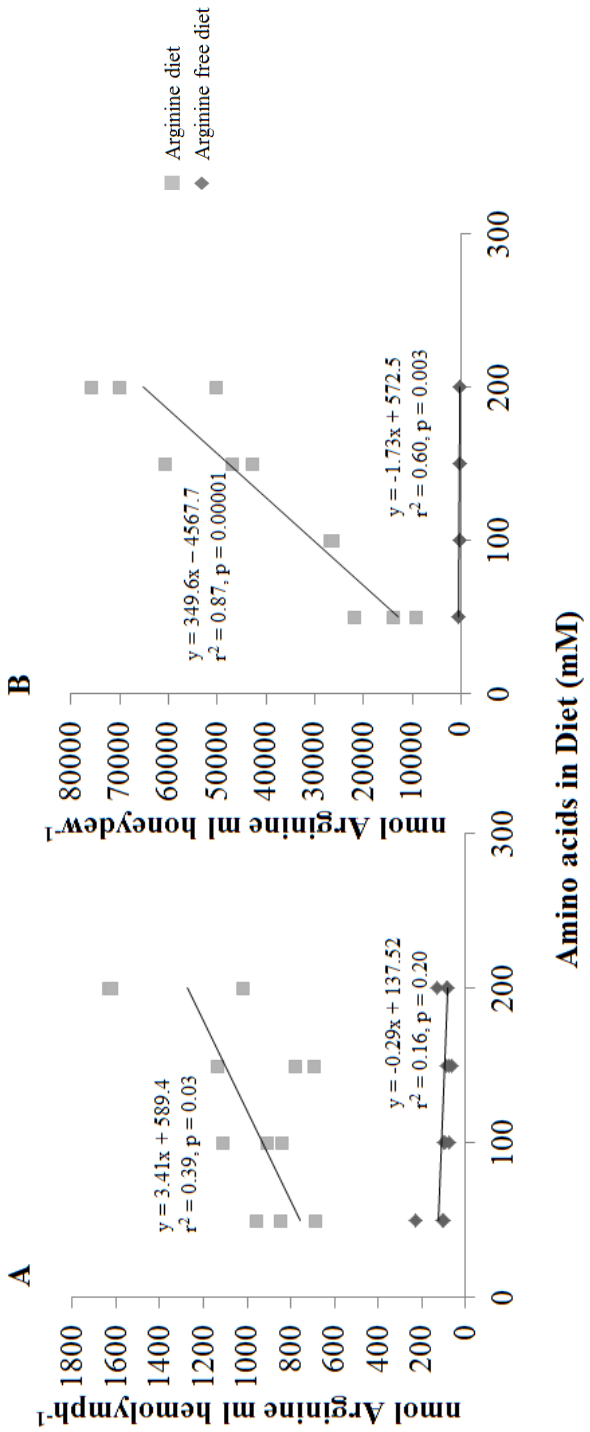


Figure A.4. Concentration of arginine in a) hemolymph and b) honeydew when aphids were reared on diet containing and missing arginine. (ANOVA, $p < 0.05$)

Table A.3. Correlations between amino acid concentration of the diet and amino acid concentrations in the honeydew and hemolymph (slope = nmol amino acid ml hemolymph or honeydew^{-1} mM amino acid increase in the diet^{-1}, ANOVA, * = p-value < 0.0025)

Amino acid	Hemolymph/Honeydew	slope	r^2	p-value
NH3	Hemolymph	0.79	0.05	0.3
	Honeydew	2.15	0.01	0.64
Ala	Hemolymph	0.53	0.001	0.87
	Honeydew	2.77	0.38	0.001*
Asn	Hemolymph	14.65	0.63	3.43E-06*
	Honeydew	3.5	0.58	1.61E-05*
Cys	Hemolymph	-0.01	0.01	0.71
	Honeydew	1.91	0.6	8.4E-06*
Gln	Hemolymph	22.73	0.67	9.42E-07*
	Honeydew	4.93	0.8	1.32E-06*
Glu	Hemolymph	0.45	0.19	0.04
	Honeydew	2.57	0.31	0.005
Gly	Hemolymph	6.88	0.41	0.0007*
	Honeydew	3.84	0.5	0.0002*
Pro	Hemolymph	8.33	0.45	0.0003*
	Honeydew	5.2	0.41	0.0008*
Ser	Hemolymph	2.2	0.45	0.0003*
	Honeydew	2.75	0.29	0.007
Tyr	Hemolymph	-0.3	0.0045	0.76
	Honeydew	0.71	0.83	8.3E-10*
Arg	Hemolymph	3.41	0.39	0.03
	Honeydew	349.59	0.87	1.01E-05*
His	Hemolymph	0.73	0.07	0.2
	Honeydew	127	0.88	1.5E-11*
Ile	Hemolymph	5.68	0.5	0.0001*
	Honeydew	10.09	0.28	0.008
Leu	Hemolymph	4.32	0.58	1.36E-05*
	Honeydew	20.87	0.52	7.27E-05*
Lys	Hemolymph	8.71	0.74	0.0003*
	Honeydew	19.25	0.43	0.02
Met	Hemolymph	0.53	0.25	0.01
	Honeydew	0.86	0.8	4.55E-09*
Phe	Hemolymph	0.66	0.36	0.002*
	Honeydew	6.55	0.86	8.1E-11*
Thr	Hemolymph	3.8	0.41	0.0008*
	Honeydew	4.19	0.28	0.008
Trp	Hemolymph	1.48	0.41	0.0008*
	Honeydew	77.46	0.93	7.13E-14*
Val	Hemolymph	6.59	0.69	5.25E-07*
	Honeydew	5.25	0.3	0.005

A.4 Discussion

The exploration of various methods to increase the endogenous nitrogen content (in the form of free amino acids and ammonia) yielded some surprising results. Dietary ammonia is not a viable method for increasing the nitrogen of the system as the gut has a low assimilation efficiency and glutamine synthetase makes glutamine from the free ammonia and glutamate. The alternate method to increase the amino acid content of the diet was the most consistent method to increase the amino acid content of the aphid. The patterns varied by amino acid with glutamine and asparagine showing the highest rate of increase, and amino acids with low assimilation efficiencies showing some of the least. The deletion of arginine highlighted the fact that the high arginine concentration of honeydew is due to dietary arginine not being utilized by the aphid.

APPENDIX B

BRANCHED CHAIN AMINOTRANFERASE AND PHENYALANINE AMINOTRANSFERASE ACTIVITY IN THE BACTERIOCYTE

B.1 Introduction

The contribution of the host to essential amino acid has been addressed for the branched chain amino acids, phenylalanine, and methionine (Chapter 3). Yet the activities for transamination were only addressed for leucine and phenylalanine aminotransferase (in the host tissue). The obvious question is does the transaminase activity follow similar patterns for isoleucine and valine as they do for leucine and phenylalanine? The prediction is that they will, as only one enzyme, the aphid branched chain aminotransferase (BCAT), mediates all 3 of the terminal reactions of branched chain amino acid synthesis.

B.2 Materials and Methods

B.2.1 Aphid rearing

Aphids were reared from a single, parthenogenetic female collected from an alfalfa field in Freeville, NY in June 2009. The aphid line, CWR09/18, was screened by PCR and microscopy for bacterial symbionts and found to contain *Buchnera aphidicola* and no secondary symbionts. The line was maintained on pre-flowering *Vicia faba* cv. Windsor at 20˚C with a 16:8 light:dark cycle.

B.2.2 Branched chain and phenylalanine aminotransferase assays

Bacteriocytes were dissected from 7-day-old larval aphids in extraction medium (28 mM glucose, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.25 M sucrose, 50 mM NaH_2PO_4, 13 mM $K_2H_2PO_4$, pH 7.5), lysed by pipetting 4-6 times, and centrifuged at 1000 *g* for 5 minutes at 4˚C to separate the *Buchnera*-free supernatant (referred to as the host fraction, HF) from the pellet containing *Buchnera* cells. The *Buchnera* cells in the pellet were quantified by

hemocytometer counts at x400 magnification, and diluted to 4 x 10^8 *Buchnera* ml^{-1} in extraction medium. The host fraction was diluted to the same volume as the *Buchnera* preparations, so a direct comparison could be made. To initiate the release experiment, 5.5 µl reaction medium was added to 8 replicate samples of 5.5 µl bacterial or host fraction suspension (with a heat treated, reaction missing substrate, and blank control). The standard reaction medium comprised the extraction medium supplemented with glutamate, glutamine, serine, aspartate, and 2-oxobutanoate, each at 2 mM. For branched chain aminotransferase activity, the reaction medium was supplemented with 2 mM 3-methyl-2-oxopentanoate, 3-methyl-2-oxobutanoate, or 4-methyl-2-oxopentanoate (oxo-acids for isoleucine, valine, and leucine synthesis respectively). For phenylalanine aminotransferase activity, the reaction medium was supplemented with 2mM phenylpyruvate (oxo-acid for phenylalanine synthesis). At 5 minute intervals over 40 minutes, one tube was centrifuged at 1000 *g* for 70 seconds, and 11 µl supernatant was immediately flash-frozen in liquid nitrogen and stored at -80^{o}C. The experiments were conducted at 22.5^{o}C, and were repeated 5 times at different times on different sets of aphids.

The amino acid content of the supernatant was quantified using the AccQ Tag derivatization kit (Waters) by UPLC with PDA detector (Waters Acquity). An equal volume of 40 mM HCl was added to 10 µl of each supernatant. Following incubation on ice for 30 minutes, the sample was centrifuged at 18000 *g* for 10 minutes at 4˚C, and the supernatant was filtered through a 0.45 µm filter plate (Millipore) by centrifugation at 1500 *g* for 10 min. The filtrate (2.5 µl) was derivatized with AccQ Tag (Waters), following manufacturer's protocol, and injected into Waters Acquity UPLC with PDA detector and AccQ-Tag Ultra 2.1 x 100 mm column. The gradient was: 0-0.54 min, 99.9% A 0.1% B; 0.54-5.74 min, 90.9% A and 9.1% B; 5.74-7.74 min, 78.8% A 21.2% B; 7.74-8.04 min, 40.4% A 59.6% B; 8.04-8.64 min, 10% A 90%

B; 8.05-8.64 min 10% A 90% B; 8.64-8.73 min 99.9% A 0.1% B; 8.73-9.50 min, 99.9% A 0.1% B (linear between each time point), where A is 10% AccQ-Taq Ultra Eluent A in water, and B is Accq-Taq Ultra Eluent B. Amino acids were determined by comparison to standards: 1, 5, 10, 50 and 100 pmol amino acids μl^{-1} (Waters amino acid hydrolysate standard #088122, supplemented with asparagine, tryptophan and glutamine).

B.2.3 Immunoblots

Custom-made polyclonal antibodies against branched chain aminotransferase (BCAT, **Gene ID: 100167587)** were produced by GenScript (Piscataway, NJ, USA). Rabbits were immunized four times using the purified peptide CVDERPHLYESQNYK (BCAT) conjugated with Keyhole limpet hemocyanin (KLH). The polyclonal antiserum obtained from the last bleed was affinity purified and tested by ELISA, including confirmation that each pre-immune serum did not react to aphid proteins. The specificity of the antibody was assessed using mass spectrometry. Briefly, the reacting band in the western blot was manually excised from the corresponding acrylamide gel, digested and submitted to a HPLC system (Dionex Ultimate 300 configured for nanobore), in line connected to a hybrid triple quadropole linear ion trap mass spectrometer, 4000 Q Trap (ABI/MDS Sciex) equipped with Ion Spray Head ion source. The resulting MS data were submitted for database searching using MASCOT search engine version 2.3 against a database containing the pea aphid genome with 34,834 protein-coding gene models (AphidBase.com).

For protein analysis, dissected bacteriocytes of 7-day-old larval aphids were fractioned into *Buchnera* and host fractions (Poliakov et al 2011). Following centrifugation at 10,600 *g* for 5 minutes, the protein content of the supernatant was quantified using RC/DC protein assay kit (Bio-Rad, Hercules, Ca, USA) following the manufacturer's instructions with bovine serum

albumin as a standard. SDS-PAGE gels were performed as described by Laemmli (1970), with 2.5 µg protein per sample on a 12 % polyacrylamide gel containing 0.1% SDS. For western blots, the proteins were transferred to a nitrocellulose membrane and blocked in PBS containing 0.5 % Tween (PBS-T) and 5% milk powder. The membrane was incubated successively in PBS-Tween containing a polyclonal antibody diluted either at 1/200 (BCAT and GOT2) or 1/1000 (TD), with an anti-rabbit IgG conjugated to peroxidase (Sigma) diluted 1/20,000, and then ECL substrate (Bio-Rad), followed by visualization with Molecular Imager ChemiDoc XRS (Bio-Rad). The intensity of the bands was analyzed by Quantity One software (Bio-Rad).

B.3 Results

B.3.1 Transaminase activity

The host fraction of the bacteriocytes was directly compared to *Buchnera* to determine whether the terminal reactions for isoleucine, leucine, valine, and phenyalanine were likely to occur in the host or in the bacterial portion of the bacteriocyte. As the only enzyme annotated within the bacteriocyte with branched chain aminotransferase activity is the aphid branched chain aminotransferase (BCAT) the activity for the transmination of these three amino acids (isoleucine, leucine, and valine) was expected to follow similar patterns. Yet oddly enough, they followed divergent patterns of transaminase activity (Figure B.1ABC). Most of the transaminase activity for synthesizing isoleucine and valine (Figure B.1AC) is associated with the *Buchnera*, whereas most of the transaminase activity for synthesizing leucine is associated with the host fraction of the bacteriocyte. This was highly unexpected, and suggests that there may be multiple enzymes (or an enzyme with multiple locations) that are mediating the terminal reactions for branched chain amino acid synthesis. The terminal reaction for phenylalanine synthesis (phenylpyruvate -> phenylalanine), much like the terminal reaction for leucine synthesis, is

predominately observed in the host fraction. The activity for leucine and phenylalanine transaminations observed associated with *Buchnera* are believed to be the result of contamination, as the separation process is crude at best.

B.3.2 BCAT localization

To verify previous findings that BCAT is host associated (via cluster analysis in Poliakov et al 2011) different fractions of the bacteriocyte were analyzed for BCAT presence. Westerns dispute the previous findings that BCAT is in the host fraction of the bacteriocyte, and suggest that this aphid enzyme associates with the *Buchnera* (Figure B.2). Though this directly contradicts the previous study (Chapter 3), this finding does give a candidate enzyme that could be mediating the terminal reaction of isoleucine and valine synthesis (at the expense of a candidate for leucine synthesis).

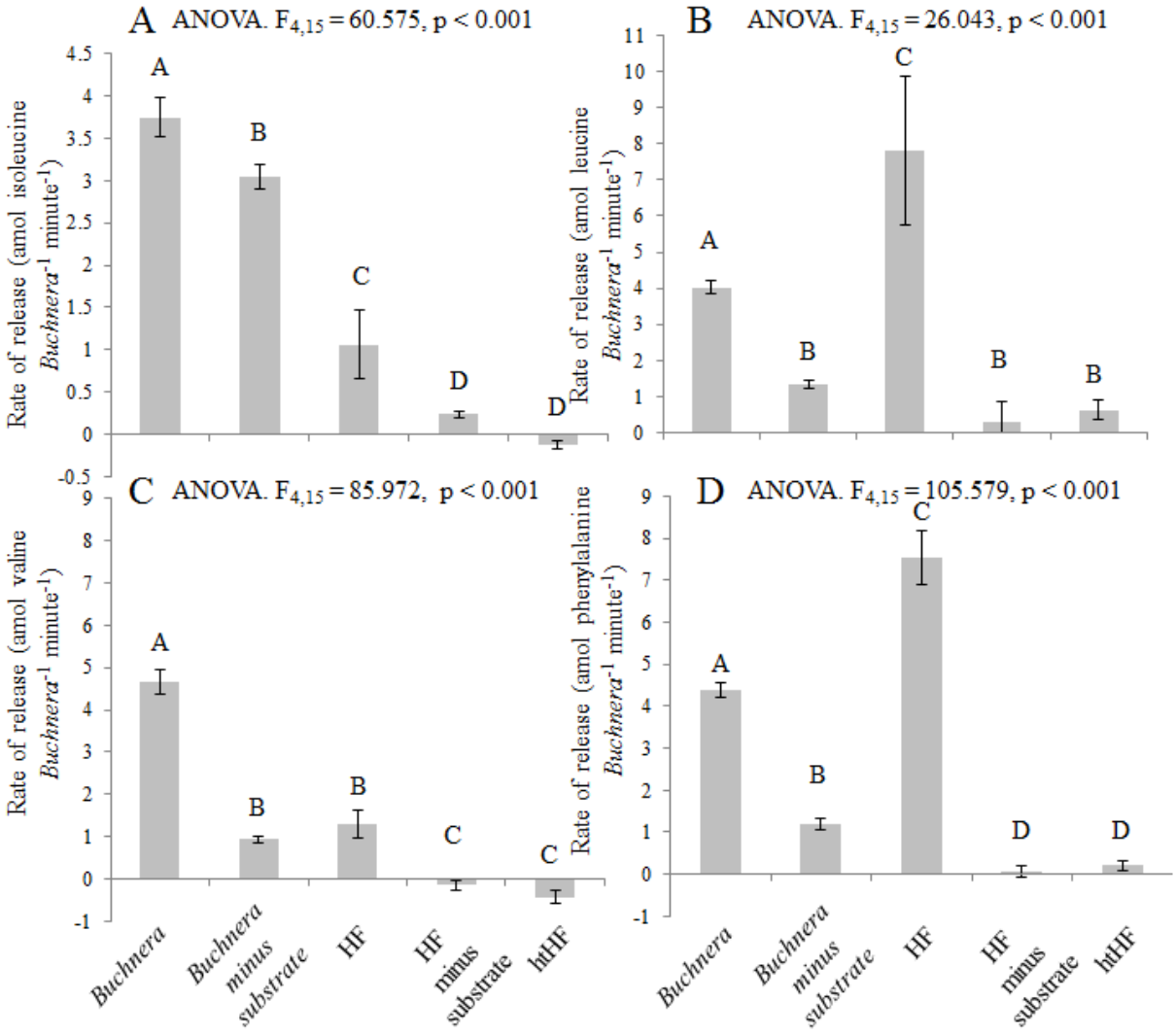
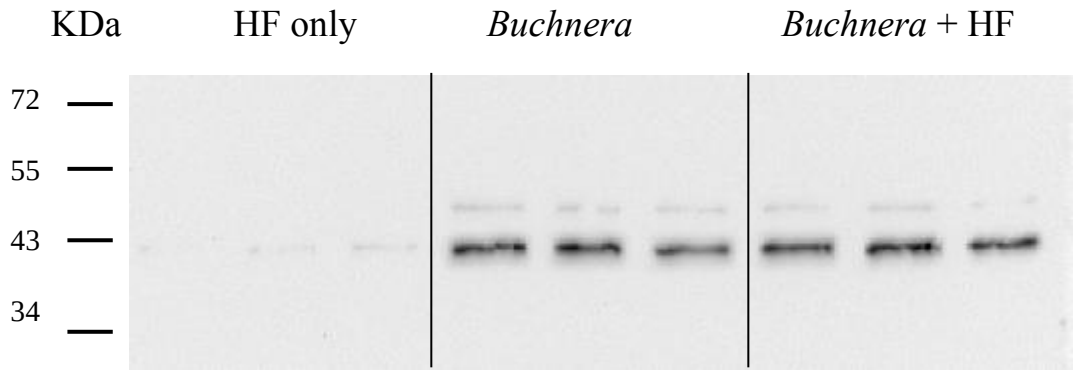


Figure B1. Activity assays for terminal transamination reactions of A) isoleucine, B) leucine, C) valine, and D) phenylalanine**.** (HF = host fraction, htHF = heat-treated host fraction, ANOVA $p < 0.001$, LSD posthoc tests $p < 0.05$)

Figure B.2. General localization of BCAT in the bacteriocyte. Western blot of 5 ug of protein the host fraction (HF), isolated *Buchnera*, and isolated *Buchnera* + HF. (Bouvaine and Russell, unpublished data).



B.4 Discussion

As *Buchnera* lack the genes coding for enzymes responsible for the terminal reactions of the branched chain amino acids and phenyalanine, the reigning hypothesis is that these reactions are mediated by the host. Chapter 3 of this thesis explored this phenomenon, but did not adequately explain the terminal reactions for isoleucine and valine. These reactions are associated with *Buchnera*. If I follow the model that only one enzyme is responsible for the terminal transamination of all branched chain amino acids, this data set is not compatible. If I accept that there may be multiple enzymes, one associated with the *Buchnera* that mediates the terminal reaction of isoleucine and valine synthesis and one associated with the host that mediates the terminal reaction of leucine, this data set begins to make sense. This is also congruent with the finding that the aphid BCAT associates with the *Buchnera* (Bouvaine, personal correspondence). These findings illustrate that exploring shared metabolic pathways rarely follows simple, predictable patterns.

REFERENCES

Akman Gunduz E., Douglas AE. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. R. Soc. Lond. B* 276, 987-991.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy A. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics*. 32, 402-407

An Nguyen TT, Michaud D, Cloutier C. 2007. Proteomic profiling of aphid *Macrosiphum euphorbiae* responses to host-plant-mediated stress induced by defoliation and water deficit. *J Insect Physiol* 53, 601-611.

Bak AL, Black FT, Christiansen C, Freundt EA. 1969. Genome size of mycoplasmal DNA. *Nature*. 224, 1209–1210

Baumann P, Baumann L, Clark MA. 1996. Levels of *Buchnera aphdiicola* chaperonin GroEL during growth of the aphid *Schizaphis graminum*. *Curr Microbiol* 32, 279-285.

Belacel N., Wang Q, Cuperlovic-Culf M. 2006. Clustering methods for microarray gene expression data. *Omics* 10, 507-531

Bidla G, Lindgren M, Theopold U, Dushay MS. 2005. Hemolymph coagulation and phenoloxidase in *Drosophila* larvae. *Dev Comp Immunol* 29, 669-679.

Bouvaine S, T Behmer S, Lin GG, Faure ML, Grebenok RJ, Douglas AE. 2012. The physiology of sterol nutrition in the pea aphid Acyrthosiphon pisum. *J Insect Physiol*. 58(11):1383-9. doi: 10.1016/j.jinsphys.2012.07.014

Brodbeck BV, Mizell RF, Andersen PC. 1993. Physiological and behavioral adaptations of three species of leafhoppers in response to the dilute nutrient content of xylem fluid, *J Insect Physiol*, 39 (1): 73-81, doi: 10.1016/0022-1910(93)90020-R.

Buchner P. 1965. *Animal Endosymbioses with Plant Microorganisms*. Chichester, UK: John Wiley & Sons.

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

Cavalier Smith T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos Trans R Soc Lond B* 358, 109-133.

Chandler SM, Wilkinson TL, Douglas AE. 2008. Impact of plant nutrients on the relationship

between a herbivorous insect and its symbiotic bacteria. *Proc Royal Soc London B* 275, 565-570.

Cristofoletti PT, Ribeiro AF, Deraison C, Rahbé Y, Terra WR. 2003. Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrthosiphon pisum. J Insect Physiol*, 49: 11–24

Dale C, Plague GR, Wang B, Ochman H, Moran NA. 2002. Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc Natl Acad Sci U S A* 99, 12397-12402.

Diacovich L, Gorvel JP. 2010. Bacterial manipulation of innate immunity to promote infection. *Nat Rev Microbiol* 8, 117-128.

Dittmer NT, Kanost MR. 2010. Insect multicopper oxidases: diversity, properties, and physiological roles. *Insect Biochem Mol Biol* 40, 179-188.

Douglas AE. 1989. Mycetocyte symbiosis in insects. *Biol Rev* 64, 409-434.

Douglas AE. 1998. Host benefit and the evolution of specialization in symbiosis. *Heredity* 81, 599-603.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Ann Rev Entomol* 43, 17-37.

Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. *J Exp Bot* 57, 747-754.

Douglas AE. 2010. *The Symbiotic Habit*. Princeton, New Jersey: Princeton University Press.

Douglas AE and Prosser WA. 1992. Synthesis of the essential amino acid tryptophan in the pea aphid (*Acyrthosiphon pisum*) symbiosis. *J Insect Physiol* 38, 565-568.

Douglas AE, Bouvaine S, Russell RR. 2011. How the insect immune system interacts with an obligate symbiotic bacterium. *Proc R Soc Lond. B* 278, 333-338.

Douglas AE, Dixon AFG. 1987. The mycetocyte symbiosis of aphids - variation with age and morph in virginoparae of *Megoura viciae* and *Acyrthosiphon pisum*. *J Insect Physiol* 33, 109-113.

Douglas AE. 1988. Sulphate utilisation in an aphid symbiosis. *Insect Biochemistry* 18, 599-605.

Douglas AE. 2009. The microbial dimension in insect nutritional ecology. *Functional Ecology* 23, 38-47

Edwards JS, Palsson BO. 2000. Metabolic flux balance analysis and the in silico analysis of

Escherichia coli K-12 gene deletions. *BMC Bioinformatics*. 2000;1:1.

Edwards M. A., Kaufman M. L., Storvick C. A. 1957 Microbiologic assay for the thiamine content of blood of various species of animals and man. *Am J Clin Nutr.* 5, 51–55

Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature* 440, 623-630.

Fares MA, Moya A, Barrio E. 2004. GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet* 20, 413-416.

Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* 417, 398.

Febvay G, Liadouze I, Guillaud J, Bonnot G. 1995. Analysis of energetic amino acid metabolism in *Acyrthosiphon pisum*: A multidimensional approach to amino acid metabolism in aphids. *Arch Insect Biochem Physiol*, 29: 45–69. doi: 10.1002/arch.940290106

Febvay G, Rahbe Y, Rynkiewicz M, Guillaud J, Bonnot G. 1999. Fate of dietary sucrose and neosynthesis of amino acids in the pea aphid, *Acyrthosiphon pisum*, reared on different diets. *J Exp Biol*. 202 (19):2639-52.

Francis F, Guillonneau F, Leprince P, De Pauw E, Haubruge E, et al. 2010. Tritrophic interactions among *Macrosiphum euphorbiae* aphids, their host plants and endosymbionts: investigation by a proteomic approach. *J Insect Physiol* 56, 575-585.

Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E. 2011. Competitive and cooperative metabolic interactions in bacterial communities. *Nature Comm*. 2, 589.

Friso G, Majeran W, Huang M, Sun Q, van Wijk K J. 2010. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol* 152, 1219-1250

Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, et al. 2010. Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome Biol* 11, R21.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Hölldobler B, van Ham R, Gross R, Moya A. 2003. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes *Proc. Natl Acad. Sci. USA* 100 (16): 9388-9393; doi:10.1073/pnas.1533499100

Gophna U, Ron EZ, Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* 312, 151-163.

Grant AJ, Trautman DA, Menz I, Hinde R. 2006. Separation of two cell signalling molecules from a symbiotic sponge that modify algal carbon metabolism. *Biochem. Biophys. Res. Comm.* 348, 92-98.

Gündüz EA and Douglas AE. 2009. Symbiotic bacteria enable insect to utilise a nutritionally-inadequate diet. *Proc Roy Soc B.* 276, 987-991

Hall TR, Wallin R, Reinhart GD, Hutson SM. 1993. Branched chain aminotransferase isoenzymes. Purification and characterization of the rat brain isoenzyme. *J. Biol. Chem*. 268: 3092-3098.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA* 108, 2849-2854.

Heddi A, Grenier A, Khatchadourian C, Charles H, Nardon P. 1999. Four intracellular genomes direct weevil biology: Nuclear, mitochondrial, principal endosymbiont, and *Wolbachia. Proc Natl Acad Sci USA.* 96 (12) 6814-6819; doi:10.1073/pnas.96.12.6814

Heddi A, Vallier A, Anselme C, Xin H, Rahbe Y, et al. 2005. Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. *Cell Microbiol* **7**, 293-305.

Hocking PJ. 1980. The Composition of Phloem Exudate and Xylem Sap from Tree Tobacco (*Nicotiana glauca* Grah.) *Ann Bot*. 45(6): 633-643

Homma KJ, Tanaka Y, Matsushita T, Yokoyama K, Matsui H, et al. 2001. Adenosine deaminase activity of insect-derived growth factor is essential for its growth factor activity. *J Biol Chem* 276, 43761-43766.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. 2005. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 40, 430-443

Huang CY, Lee CY, Wu HC, Kuo MH, Lai CY. 2008. Interactions of chaperonin with a weakly active anthranilate synthase from the aphid endosymbiont *Buchnera aphidicola*. *Microb Ecol*. 56(4):696-703. doi: 10.1007/s00248-008-9389-4.

Hueck CJ. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 62, 379-433.

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

Hungate RE. 1950. Mutualisms in protozoa." *Annual Reviews in Microbiology* 4.1: 53-66.

The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid

Acyrthosiphon pisum. *PLoS Biol*. 8, e1000313.

Ishikawa H, Yamaji M, Hashimoto H. 1985. Symbionin, an aphid endosymbiont-specific protein. 2. Diminution of symbionin during post-embryonic development of aposymbiotic insects. *Insect Biochem*15, 165-172.

Jiang Z-H, Xia F, Johnson KW, Brown CD, Bartom E, et al. 2013. Comparative genome uences of the primary endosymbiont Candidatus Portiera aleyrodidarum from the whiteflyBemisia tabaci B and Q biotypes. *Appl. Env. Microbiol*., in press.

Kambara K, Ardissone S, Kobayashi H, Saad MM, Schumpp O, et al. 2009. Rhizobia utilize pathogen-like effector proteins during symbiosis. *Mol Microbiol* 71, 92-106.

Karley AJ, Ashford DA, Minto LM, Pritchard J, Douglas AE. 2005. The significance of gut sucrase activity for osmoregulation in the pea aphid, *Acyrthosiphon pisum*. *J Insect Physiol*. 51(12): 1313-1319, doi: 10.1016/j.jinsphys.2005.08.001.

Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. *Trends in Ecology & Evolution* 20:670-676

Kelkar YD, Ochman H. 2013. Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics* 193, 303-307.

Kerney R, Kim E, Hangarter RP, Heiss AA, Bishop CD, Hall BK. 2011. Intracellular invasion of green algae in a salamander host. *Proc Natl Acad Sci USA* 108 (16) 6497-6502

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JM, Ribeiro JM, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA* 107(27):12168-73. doi: 10.1073/pnas.1003379107

Klitgord N, Segre D. 2011. Ecosystems biology of microbial metabolism. *Curr. Op.Biotech.* 22, 541-546

Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 11(5):209.

Laemmli UK. 1970. Cleave of structural protein during the assembly of the head of

bacteriophage T4. *Nature* 227:680

Lain-Guelbenzu B, Muñoz-Blanco J, Cárdenas J. 1990. Purification and properties of L-Aspartate aminotransferase of *Chlamydomonas reinhardtii*. *Eur J Biochemi* 188*:* 529–533.

Lamb KP, Hinde R. 1967. Structure and development of the mycetome in the cabbage aphid, *Brevicoryne brassicae*. *J Invert Pathol* 9, 3-11.

Lamelas A, Gosalbes MJ, Andres M, Latorre A. 2011. The genome of *Buchnera aphidicola* from the aphid *Cinara tujafilina* provides new clues about the evolutionary history of metabolic losses in bacterial endosymbionts. *Appl. Environ. Biol.*

Le Trionnaire G, Francis F, Jaubert-Possamai S, Bonhomme J, De Pauw E, et al. 2009. Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics* 10. 456.

Liu H, Sadygov RG, Yates JR, 3rd. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76, 4193-4201.

Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proc. R. Soc. Lond. B* 279, 2965-2973.

MacDonald SJ, Thomas GH, Douglas AE. 2011. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Mol. Ecol.* 20, 2073-2084.

Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, et al. 2006. Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp. strain APS. *J Bacteriol* 188, 6539-6543.

Majeran M, Friso G, Ponnala L, Connolly B, Huang M, Reidel E, Zhang C, Asakura Y, Bhuiyan NH, Sun Q, Turgeon R, van Wijk KJ. 2010. Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics. *The Plant Cell* prepublication on-line Nov 16

Margulis, L. 1970. Origin of Eukaryotic Cells. Yale Univ. Press, New Haven, CT, 1970.

Martin W. 2010. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philos Trans R Soc Lond B* 365, 847-855.

McCutcheon JP. 2010. The bacterial essence of tiny symbiont genomes. *Curr Opin Microbiol* 13, 73-78.

McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl Acad. Sci. USA* 106, 15394-15399.

McCutcheon JP, Moran NA. 2007). Parallel genomic evolution and metabolic interdependence

in an ancient symbiosis. *Proc Natl Acad Sci U S A* 104, 19392-19397.

McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol.* 2, 708-718.

McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21, 1366-1372.

Mittler TE. 1971. Dietary amino acid requirements of the aphid Myzus persicae affected by antibiotic uptake. *J Nutr.* 101(8):1023-8.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* 93, 2873-2878.

Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci U S A* 104 Suppl 1, 8627-8633.

Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. *Genome Biol*. 2, R0054.

Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond*. 253: 167-171.

Moran NA, Russell JA, Koga R, Fukatsu T. 2005. Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. *Appl Environ Microbiol* 71, 3302-3310.

Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, et al. 2005. Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc Natl Acad Sci U S A* 102, 5477-5482.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006.The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science*. 13;314(5797):267.

Nappi AJ, Carton Y, Frey F. 1991. Parasite-induced enhancement of hemolymph tyrosinase activity in a selected immune reactive strain of *Drosophila melanogaster*. *Arch Insect Biochem Physiol* **18**, 159-168.

Nasir H, Noda H. 2003. Yeast-like symbiotes as a sterol source in anobiid beetles (Coleoptera, Anobiidae): Possible metabolic pathways from fungal sterols to 7-dehydrocholesterol. *Arch. Insect Biochem. Physiol.* 52, 175–182.

Newell PD, Boyd CD, Sondermann H, O'Toole GA. 2011. A c-di-GMP effector system controls cell adhestion by inside-out signaling and surface protein cleavage. *PLoS Biol*. 9

Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, et al. 2010. Bacterial genes

in the aphid genome: absence of functional gene transfer from Buchnera to its host. *PLoS Genet* 6, e1000827.

Nishikori K, Morioka K, Kubo T, Morioka M. 2009. Age- and morph-dependent activation of the lysosomal system and *Buchnera* degradation in aphid endosymbiosis. *J Insect Physiol* 55, 351-357.

Nogge, G. 1976. Sterility in tsetse fly (*Glossina morsitans* Westwood) caused by loss of symbionts. *Experientia*, 32, 995.

Okazaki S, Okabe S, Higashi M, Shimoda Y, Sato S, et al. 2010. Identification and functional analysis of type III effector proteins in *Mesorhizobium loti*. *Mol Plant Microbe Interact* 23, 223-234.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, et al. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4, 1487-1502.

Olinares PD, Ponnola L, van Wijk KJ. 2010. Megadalton complexes in the chloroplast stroma of arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering. *Mol Cell Proteomics* 9.7, 1594-1615

Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* 55, 247-266.

Pais R, Lohs C, Wu Y, Wang J, Aksoy S. 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl Environ Microbiol* 74, 5965-5974.

Pallen MJ, Bailey CM, Beatson SA. 2006. Evolutionary links between FliH/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the FoF1 and vacuolar ATPases. *Protein Sci* 15, 935-941.

Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* 314, 312-313.

Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ. 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Molecular and Cellular Proteomics* 10: M10.007039

Price DR, Duncan RP, Shigenobu S. Wilson AC. 2011. Genome expansion and differential expression of amino acid transporters at the aphid/Buchnera symbiotic interface. *Mol Biol Evol.* 28, 3113-3126.

Price DR, Tibbles K, Shigenobu S, Smertenko A, Russell CW, et al. 2010. Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional

characterization. *Insect Mol Biol* 19 Suppl 2, 97-112.

Prosser WA, Simpson SJ and Douglas AE. 1992. How an aphid (*Acyrthosiphon pisum*) symbiosis responds to variation in dietary nitrogen. *J Insect Physiol* 38, 301-307.

Puchta, O. 1956. Zuchtungsversuche an den Symbionten von *Pediculus vestimenti* Burm. Nebst physiologischen und morphologischen Beobachtungen. *Zeitschrift für Morphologie und Oekologie der Tiere,* 44, 416–441.

Radtke AL, O'Riordan MX. 2006. Intracellular innate resistance to bacterial pathogens. *Cell Microbiol* 8, 1720-1729.

Ramsey JS, MacDonald SJ, Jander G, Nakabachi A, Thomas GH, et al. 2010. Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrthosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol* 19 Suppl 2, 241-248.

Ray K, Marteyn B, Sansonetti PJ, Tang CM. 2009. Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nat Rev Microbiol* 7, 333-340.

Reymond N, Calevro F, Vinuelas J, Morin N, Rahbe Y, et al. 2006. Different levels of transcriptional regulation due to trophic constraints in the reduced genome of *Buchnera aphidicola* APS. *Appl Environ Microbiol* 72, 7760-7766.

Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9, 313-323.

Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. 2013. The nutrient supplying capabilities of Uzinura, an endosymbiont of armoured scale insects. *Environ Microbiol*. doi: 10.1111/1462-2920.12058

Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* 106, 19521-19526.

Santos-Garcia D, Farnier PA, Beitia F, Zchori-Fein E, Vavre F, Mouton L, Moya A, Latorre A, Silva FJ. 2012. Complete genome sequence of "Candidatus *Portiera aleyrodidarum*" BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to Bemisia tabaci. *J Bacteriol*. 194(23):6654-5. doi: 10.1128/JB.01793-12.

Sasaki T, Ishikawa H. 1995. Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, Acyrthosiphon pisum. *J. Insect Physiol*. 41, 41-46.

Sato S, Ishikawa H. 1997. Expression and control of an operon from an intracellular symbiont which is homologous to the groE operon. *J Bacteriol* 179, 2300-2304.

Scigelova M, Makarov A. 2006. Orbitrap mass analyzer - overview and applications in

proteomics. *Proteomics* 6 Suppl 2, 16-21

Seaver LC, Imlay JA. 2001. Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in *Escherichia coli*. *J Bacteriol* 183, 7173-7181.

Serres MH, Riley M. 2000. MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb Comp Genomics* 5, 205-222.

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

Sloan DB, Moran NA. 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol Lett.* 8, 986-989.

Tamas I, Klasson L, Canback B, Naslund A, Eriksson A, Wernegreen J, Sandstrom J, Moran NA, Andersson SGE. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Nature*. 296: 2376-2379

Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, et al. 2008. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc Natl Acad Sci U S A* 105, 14934-14939.

Terra WR, de Bianch AG, Gambarini AG, Lara FJS. 1973. Haemolymph amino acids and related compounds during cocoon production by the larvae of the fly, *Rhynchosciara americana*. *J Insect Physiol* 19, 2097-2106.

Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, et al. 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Systems Biol* 3, 24.

Thompson SN. 2003. Trehalose - the insect 'blood' sugar. *Adv Insect Physiol* 31, 205-285.

Toft C, Andersson SG. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 11, 465-475.

Trench RK. 1971. Physiology and biochemistry of zooxanthellae symbiotic with marine coelenterates. III. Effect of homogenates of host tissues on excretion of photosynthetic products in vitro by zooxanthellae from two marine coelenterates. *Proc R Soc Lond. B* 177, 251-269.

van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Moran F, Moya A. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci USA* 100: 581-586.

Vinuelas J, Calevro F, Remond D, Bernillon J, Rahbe Y, et al. 2007. Conservation of the links

between gene transcription and chromosomal organization in the highly reduced genome of *Buchnera aphidicola*. *BMC Genomics* 8, 143.

Wang J, Wu Y, Yang G, Aksoy S. 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proc Natl Acad Sci U S A* 106, 12133-12138.

Wang JT, Douglas AE. 1997. Nutrients, signals, and photosynthate release by symbiotic algae. *Plant Physiol*. 114, 631-636.

Wang Y, Carolan JC, Hao F, Nicholson JK, Wilkinson TL, et al. 2010. Integrated metabonomic-proteomic analysis of an insect-bacterial symbiotic system. *J Proteome Res* 9, 1257-1267.

Whitehead LF, Douglas AE. 1993. Populations of symbiotic bacteria in the parthenogenetic pea aphid (*Acyrthosiphon pisum*) symbiosis. *Proc Roy Soc London B* 254, 29-32.

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

Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol Microbiol* 48, 1491-1500.

Wilkinson TL and Douglas AE, 1995. Why aphids lacking symbiotic bacteria have elevated levels of the amino acid glutamine. *J Insect Physiol* 41, 921-927.

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

Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan P, Macdonald SA, Schwartz J, Thomas GH and Douglas AE. 2010. Genomic insight into the amino acid relations of the pea aphid *Acyrthosiphon pisum* with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Bio* 19 (2): 249-258

Wintermute EH, Silver PA. 2010. Dynamics in the mixed microbial concourse. *Genes Dev.* 24, 2603-2614.

Wyatt GR. 1961. The biochemistry of insect hemolymph. *Ann Rev Entomol* 6, 75-102.

Zybailov B, Friso G, Kim J, Rudella A, Rodriguez VR, et al. 2009. Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol Cell Proteomics* 8, 1789-1810.