Title
Small RNAs in Small Genomes: The Characterization of Expressed Small RNAs in Obligate Bacterial Symbionts of Hemipterans

Permalink
https://escholarship.org/uc/item/2jk1x58x

Author
Thairu, Margaret

Publication Date
2019

Supplemental Material
https://escholarship.org/uc/item/2jk1x58x#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Small RNAs in Small Genomes: The Characterization of Expressed Small RNAs in Obligate
Bacterial Symbionts of Hemipterans

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

Doctor of Philosophy

in

Entomology

by

Margaret W. Thairu

June 2019

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

The Dissertation of Margaret W. Thairu is approved:

_____

_____

_____

_____
                                                    Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Small RNAs in Small Genomes: The Characterization of Expressed Small RNAs in Obligate
Bacterial Symbionts of Hemipterans

by

Margaret W. Thairu

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

Within hemipterans, the acquisition of bacterial symbionts has helped insects utilize the nutrient-limited resource of plant sap. Many of these bacterial symbionts have undergone a dramatic reduction in genome size, often losing key regulatory genes. It is therefore unclear how and if these symbionts regulate their gene expression. My dissertation explores the potential role that symbiont expressed small RNAs (sRNAs) have in gene expression. First, I use RNA-seq to characterize the sRNA expression profile of *Buchnera*, during two different life-stages (aphid ovarioles and maternal bacteriocytes) in which *Buchnera* has differential protein expression. The results from this experiement show that *Buchnera* sRNAs are differentially expressed between life-stages. My dissertation also provides *in vitro* evidence of the functionally of the *Buchnera* antisense sRNA *carB*. These results suggest that when *Buchnera* is in an extracellular state, free of the bacteriocytes, it can respond to changes in host nutritional demand. I then characterized *Buchnera's* sRNA expression when its aphid host fed on two different host-plants which have been shown to have different nutritional and plant defense profiles. The results from this

experiment show that *Buchnera* sRNA expression varies with aphid host-plant diet. These results suggest that *Buchnera* sRNAs can potentially impact the symbiosis in an adaptive nutritional manner, or stress response manner when aphids feed on a host-plant that is lower in nutrients. I also determined that sRNAs are expressed and conserved in one of the most reduced obligate insect symbionts, *Candidatus* Carsonella ruddii. Currently, many of the functional genomic tools that are optimized to work in a handful of model systems. As result, when working with non-model, unculturable systems, there is an additional challenge of optimizing and modifying current functional genomic tools. The last part of my dissertation tests the efficacy of novel RNAi delivery systems within three aphid species. This RNAi delivery system has resulted in successful gene knockdown in the soybean aphid. Overall, the findings of my dissertation strongly support the hypothesis that small bacterial genomes may utilize sRNAs to help regulate their own gene expression to help compensate for the loss of canonical regulatory proteins.

Dedication:
To Mom and Dad, thanks for everything.
To Ross, for sticking with me.

Table of Contents:

List of Tables:

List of Figures:

Chapter 1:

It's a small, small world: Unravelling the role and evolution of small RNAs in organelle and endosymbiont genomes

Introduction:

Bacterial symbionts of eukaryotes that are host-restricted, evolve distinctive genomic characteristics which include the re-organization and the reduction of their genomes (*reviewed in* Toft and Andersson, 2010; Sachs et al., 2011; Moran and Bennett, 2014). These architectural changes are hypothesized to primarily be due to non-adaptive evolutionary forces, however selection driven hypotheses have also been proposed to explain genome reduction in some symbiotic bacteria (*reviewed in* Toft and Andersson, 2010; Sachs et al., 2011; Moran and Bennett, 2014; Martínez-Cano et al., 2015). Host-restricted symbionts have lost many of their regulatory elements and genes, and therefore it has been postulated that bacteria in ancient symbioses have lost the ability to respond to changes in their environment compared to their free-living relatives (Hansen and Moran, 2014). However, in recent years, studies have revealed that bacterial symbionts with reduced genomes may in fact be able to respond to environmental changes by expressing regulatory sRNAs (*reviewed in* Kim et al., 2016).

Organelles, *i.e.* mitochondria and plastids, represent extreme examples of ancient host restricted bacterial symbioses (*reviewed in* Archibald, 2009; Gray, 2012; Toft and Andersson, 2010; McCutcheon, 2016). In this review, we compare and contrast sRNA expression patterns in organelle and symbiont genomes to gain a better understanding of sRNA evolution and their potential regulatory roles in small genomes. We also examine key genomic characteristics of organelles and host-restricted bacterial symbionts (now referred to as bacterial symbionts) that may influence the evolution and function of sRNAs. Given that comprehensive reviews are already available on eukaryotic host sRNAs (*e.g.* Moran et al. 2017; Bartel, 2018; Brant and Budak, 2018;

Chen et al., 2018; Singh et al. 2018), this review focuses primarily on sRNAs that are expressed by small host-restricted symbiont and organelle genomes. We broadly define 'symbiont' here as either mutualist, pathogenic, and/or commensal to its eukaryotic host.

A broad diversity of small RNA types and expression patterns identified from organelle and bacterial symbiont genomes

Bacterial derived sRNAs vary greatly in size and may or may not require accessory proteins to function (Wagner and Romby, 2015). In general, regulatory sRNAs in bacteria can either be *trans*-encoded or *cis*-encoded (Wagner and Romby, 2015). *Trans*-encoded sRNAs are transcribed at genomic locations that are distant from their target mRNA(s); as such they often have partial complementarity with their mRNA target(s) (Wagner and Romby, 2015). Many enterobacterial *trans*-encoded sRNAs require the RNA chaperone protein, Hfq, for proper functioning (Durand et al., 2015, Wagner and Romby, 2015). Other bacteria, such as *Staphylococcus aureus* or *Bacillus subtilis*, either do not require Hfq for sRNA function or utilize other proteins such as CshA and FbpA-C (Wagner and Romby, 2015). Organelles and many bacterial symbionts do not encode a known sRNA chaperone (Sun et al., 2002). In turn, if *trans*-encoded sRNAs are expressed in these small genomes they may not require a chaperone protein, or alternatively, they may utilize other stabilizing, protein recruiting, or mRNA-sRNA binding mechanisms (Hotto et al., 2012, Sobrero and Valverde, 2012). For example, Narra et al. (2016), demonstrated that a chaperone independent mechanism of sRNA-mRNA binding within the bacterial symbiont, *Rickettsia conorii,* may occur.

In contrast to *trans*-encoded sRNAs*, cis*-encoded or antisense sRNAs have perfect complementarity with their target RNAs and are encoded on the opposite DNA strand of their target mRNA (Wagner and Romby, 2015). Though some antisense sRNAs have been found to interact with Hfq, many can function without a protein chaperone (Bilusic et al. 2014). To date, many of

the sRNAs identified within organelles and bacterial symbionts are assumed to fall within the category of non-protein coding and *cis*-encoded (*discussed in detail below*).

It is of note that some sRNAs in mitochondria encode short polypeptides such as humanin, MOTS-c (mitochondrial open reading frame of the 12S rRNA-c), and small humanin like peptides (SHLPs) (Hashimoto et al. 2001; Lee et al. 2015; Cobb et al. 2016). There is evidence that these short polypeptides not only help regulate metabolic hemostasis, but also have protective roles and may be important in mito-nuclear communication (Hashimoto et al. 2001; Lee et al. 2015; Cobb et al. 2016; Kim et al. 2018). Within free-living bacteria, numerous short polypeptides are known to have important roles in regulating cell division, transporters and membrane bound enzymes (*reviewed in* Storz et al. 2014). Interestingly, some small transcripts were initially characterized as sRNAs and only later were revealed to encode small proteins; in some instances, both the sRNA and the short polypeptide each have distinct functional roles (*e.g.* sgrS, sgrT; Wadler and Vanderpool, 2007).

sRNAs can be expressed throughout the entire genome. This includes within both protein coding genes and intergenic regions, as well as, transfer RNAs (tRNA), and ribosomal RNA (rRNA) genes. Below, we synthesize literature on the diversity of organelle and symbiont sRNAs that are expressed from these genomic locations. Moreover, we highlight studies that demonstrate the potential regulatory role of these sRNAs.

sRNAs expressed within protein coding genes -sense and antisense:

sRNAs derived from the protein coding sequence of a gene can either be expressed in the same direction (sense) or the complementary direction (antisense) to the coding sequence. The majority of studies report primarily on the identification of antisense sRNAs however, because of the inability to distinguish between readthrough or processing of mRNA transcripts. Nevertheless,

in human mitochondria, the number of sense sRNAs identified varies widely, and depends on the experimental approach employed and the tissue type being characterized. For example, 409 sense sRNAs from mitochondria were identified from the human brain, heart, stomach, small intestine, and colon (Ro et al. 2013). In contrast, Sripada et al. (2012) identified only three to six sense sRNAs using HeLa and HEK293 cell lines. On the other hand, antisense sRNAs are generally thought to be more abundant within plastids (Wang et al., 2011; Hackenberg et al., 2015). For example, in *Salvia miltiorrhiza* (Chinese sage) plastids, 87% of the sRNAs identified were classified as antisense (Chen et al., 2014). The expression of these antisense sRNAs were positively correlated with the expression of their putative mRNA targets (Chen et al., 2014). Antisense sRNAs have also been identified in other genomic regions. For example, within *Arabidopsis thaliana* plastids, 70 antisense sRNAs were identified in the 5' and 3'-regions (Hotto et al., 2011). Plastid antisense sRNAs have also been identified complementary to important regulatory regions, such as the binding sites of group II introns (Zhelyazkova et al., 2012).

In free-living bacteria, antisense sRNAs are important in fine-tuning gene expression by regulating transcript stability or translational efficiency (Thomason and Storz, 2010). Within organelles, the functions of several antisense sRNAs have been tested. For example, by using Northern blots and real-time quantitative reverse transcription PCR, the conserved antisense sRNA-*ndhB* was shown to respond to changes in temperature in *A. thaliana*, and *Populus sp*. (poplar) hybrids (Georg et al., 2010). The sRNA-*ndhB* covers a temperature sensitive RNA editing site, as well as a group II intron splice acceptor site (Georg et al., 2010). In a second example, the antisense sRNA-*psbT* is co-transcribed with the *psbN* gene, which is controlled by the plastid sigma factor 3 (SIG3) within the *psbB* operon of plastids (Zghidi-Abouzid et al., 2011). Under photo-oxidative stress it has been shown that the antisense sRNA-*psbT* may prevent degradation of the *psbT* mRNA transcript. It has also been observed that the antisense sRNA, AS5, may regulate the processing

and accumulation of the 5S rRNA in both *A. thaliana* and *N. tabacum* plastids (Hotto et al., 2010). Antisense sRNAs have also been tested for functionality in animal mitochondria. For example, in the study by Ro et al. (2013), the authors identified 60 and 35 antisense sRNAs in mouse and human mitochondria, respectively. Using a mouse cell line, Ro et al. (2013), demonstrated that candidate antisense sRNAs inhibit mRNA expression, whereas sense sRNAs increase the expression of their mRNA targets (Ro et al., 2013).

Antisense sRNA expression is also observed in bacterial symbiont genomes. For example, within, *Mycoplasma*, antisense sRNAs are predicted to regulate genes that are involved in metabolism, DNA repair, and DNA replication (Güell et al., 2009). In another system, *Rickettsia prowazekii* and *R. conorii* were found to have unique sRNA expression profiles when infecting either human or tick, *Amblyomma americanum* (the lone star tick) cell lines (Schroeder et al., 2016; Narra et al., 2016; Schroeder et al., 2017). These differentially expressed sRNAs may be important in regulating genes responsible for symbiont host niche adaptation (Narra et al., 2016; Schroeder et al., 2017).

Antisense sRNA expression has also been observed in insect bacterial symbionts that provide nutrients to their host. For example, in one of the obligate nutritional bacterial symbionts of the glassy-winged sharpshooter (*Baumannia*)Bennett and Chong (2017) observed two antisense sRNAs that were co-expressed with their putative target genes in *Baumannia*. In the aphid, multiple lineages of its obligate nutritional symbiont, *Buchnera,* express the sRNA-*carB,* which is antisense to the gene *carB*. The gene *carB* is involved in arginine biosynthesis. Arginine along with other essential amino acids are synthesized by *Buchnera* and are required for aphid survival, because the aphid does not encode these pathways endogenously and feeds on a nitrogen deprived diet of plant sap (Bennett and Moran 2014). The antisense sRNA *carB* is hypothesized to be important in the regulation of this symbiosis, because it activates/stabilizes the *Buchnera* protein CarB when

5

Buchnera is extracellular inside of developing aphid embryos compared to when it is intracellular inside of symbiotic nymph cells (bacteriocytes) (Hansen and Degnan, 2014; Thairu et al., 2018). This sRNA-*carB* may be very important in regulating *Buchnera*'s arginine biosynthesis pathway during a crucial aphid life-stage (embryonic) where host regulation of *Buchnera*'s essential amino acid pathways may not be available (Hansen and Degnan 2014; Lu et al. 2016).

tRNA-derived sRNAs

tRNA-derived sRNAs, have been identified within eukaryotes, bacteria, archaea, and organelles (Fischer et al., 2011; Keam and Hutvagner, 2015; Martinez, 2018). In eukaryotes, some of these tRNA-derived sRNAs are known to have potential regulatory roles (*see* Keam and Hutvagner, 2015). Within plastids, some sRNA candidates that are derived from tRNA genes appear to respond to environmental stressors such as heat and nutrient deprivation. For example, plastid tRNA-derived sRNAs of *Brassica rapa* (common mustard), increase in expression during heat stress (Wang et al., 2011). In *Hordeum vulgare* (barley) plastids, sRNAs derived from tRNA genes increase in expression when the plant is deprived of phosphorous (Hackenberg et al., 2013). These tRNA-derived sRNAs are also observed to be differentially expressed in different plant tissues. For example, in *A. thaliana*, Cognat et al. (2017), observed that tRNA-derived sRNAs have tissue specific expression profiles and are up-regulated in photosynthetic tissues. Cognat et al. (2017) also found that ~ 25% of all tRNA-derived sRNAs expressed in plant cells are from organelle encoded tRNAs and accumulate outside of the organelle. Cognat et al. (2017) also found that the plasmid expressed sRNA fraction immunoprecipitated with the RNAi associated protein Argonaute1. These results suggest that these tRNA-derived sRNAs could be elements of a retrograde signaling pathway (Cognat et al., 2017).

Similar to plastids, tRNA-derived sRNA expression within animal mitochondria is known to respond to environmental stress. For example, when mitochondria within mice sperm are

chronically exposed to high levels of ethanol there is a shift in tRNA-derived sRNA expression profiles (Rompala et al., 2018). In addition to stress, tRNA-associated sRNA profiles of animal mitochondria shift throughout host development, as demonstrated by Ma et al. (2016), who looked at tRNA-derived sRNA expression during rainbow trout egg development. It is unknown how these sRNAs are regulated in response to stress or development. However, within mammalian cell line mitochondria, there is evidence that the expression of tRNA-associated sRNAs is reduced when *DICER,* a protein important in the RNAi pathway, is inactivated (Ro et al. 2013). These results suggest that indirect signaling between the mitochondria and the nuclear genome may be occurring because *DICER* does not directly interact with these sRNAs or even localize within the mitochondria (Ro et al., 2013).

tRNA-derived sRNAs have also been described within bacterial symbionts. For example, among five *Buchnera* lineages that diverged > 65 million years ago*,* the conserved expression of 12 antisense tRNAs has been observed (Hansen and Moran, 2012). Subsequent work demonstrated that six conserved and five lineage-specific antisense sRNAs derived from tRNAs display life-stage specific expression profiles (Thairu et al., 2018). The role if any of these conserved antisense sRNAs is currently unknown.

rRNA-derived sRNAs:

Small RNAs expressed from rRNA genes often represent the largest fraction of sRNAs identified within organelles (Gonzalez-Ibeas et al., 2011; Wang et al., 2011; Hackenberg et al., 2013). For example, in mouse and human tissue samples as much as ~40% of all sense sRNAs expressed were from mitochondrial rRNA genes (Sripada et al., 2012; Ro et al. 2013). The potential role, if any, of these highly abundant sRNAs in organelles is currently unknown. However, within plastids of *B. rapa* there is evidence that sRNAs derived from plastid rRNAs decrease in expression

by ~49% in response to heat stress (Wang et al., 2011). In the nuclear genome of eukaryotes there is emerging evidence that rRNA-derived sRNAs may have regulatory functions (Chen et al., 2017). For example, Chen et al. (2017), demonstrated that when a nuclear sRNA candidate was disrupted with RNAi there was an increase in cell death and the inhibition of cell proliferation in human H1299 cell lines. Currently, to our knowledge, no studies yet have described rRNA-derived sRNAs within bacterial symbiont systems .

Intergenic region-derived sRNAs:

In general, sRNAs expressed within the intergenic regions (IGR) of organelle and symbiont genomes make up a small percentage of the total amount of sRNAs identified from these genomes (Marker et al., 2002; Hansen and Degnan, 2014). Similar to antisense and tRNA-derived sRNA expression in symbionts and organelles, IGR-derived sRNAs have also been found to display differential expression between tissue types, developmental stages, and respond to environmental stressors. For example, Itaya et al. (2008), identified IGR-derived plastid and mitochondrial sRNAs of *Solanum lycopersicum* (tomato) that display differential expression profiles during the development of leaf and fruit tissues. In another study, Wang et al. (2011), observed that in *B. rapa* plastids, IGR-derived sRNAs increase by ~30% in response to heat stress. Regarding symbionts, numerous IGR-derived sRNAs have been identified in *Wolbachia* via RNA-sequencing. One of these sRNAs was found to have tissue and sex specific expression patterns (Woolfit et al., 2015). Within *R. prowazekii*, 35 and 26 IGR-derived sRNAs were identified when infecting either human or lone star tick cell lines*,* respectively (Schroeder et al., 2016; Schroeder et al., 2017). In another study, Nara et al. (2016)*,* identified 13 IGR-derived sRNAs being expressed by *R. conorii* in human cell lines.

Are organelle and bacterial symbiont small RNAs conserved?

Small RNA sequence and/or expression conservation has been observed across specific lineages of organelles and bacterial symbionts. For example, ~39 clusters of sRNAs are conserved between mouse and human mitochondria (Ro et al., 2013). Similarly, among plastid lineages some sRNA candidates are widely conserved within untranslated regions (Ruwe and Schmitz-Linneweber, 2012). Nevertheless, some sRNAs that are widely conserved (*e.g.* cobalamin, pfl and ykkC–yxkD riboswitches) across divergent free-living bacterial taxa, are generally not conserved in bacterial symbionts (Matelska et al., 2016). With this said, within and across specific lineages of *Buchnera* that span > 65 million years of divergence a high level of sequence and expression conservation is found among sRNA candidates (Hansen and Degnan, 2014; Thairu et al., 2018) and regulatory elements (Degnan et al., 2011). The conservation of specific sRNAs among divergent taxa supports the hypothesis that sRNAs may be maintained by selection for specific regulatory and/or structural functions.

Within free-living bacteria, ribonuclease polynucleotide phosphorylase (PNPase), an enzyme present in all domains except archaea, is important in the processing of bacterial non-coding RNAs (Leszczyniecka et al. 2004; Bandyra et al. 2016). Forms of this enzyme, which are encoded on the host's nuclear genome, have also been found within organelles (*see* Baginsky et al. 2001; Perrin et al. 2004; Viegas et al. 2007). Interestingly, Hotto et al. (2011) demonstrated that in a PNPase mutant of *A. thaliana,* sRNA expression profiles are significantly different compared to the wildtype. This finding suggests that PNPase's role in sRNA biogenesis and accumulation has been conserved within plastids and free-living bacteria.

It is important to note that the presence of conserved sRNAs does not automatically serve as evidence that the sRNA is functional. Indeed, conserved sRNAs may still be the by-product of transcriptional noise due to the presence of conserved genomic regions found among lineages,

especially if genomes have a high degree of synteny (Tamames 2001; van Ham et al., 2003; Degnan et al., 2011).

Are small RNAs in small genomes associated with genome architecture?

Genome architecture of organelles and bacterial symbionts may influence how sRNA expression and their putative functions evolve. Although genomes of organelles and bacterial symbionts share many architectural similarities, differences in genome structure are observed due to variation in evolutionary forces such as deletion bias and effective population sizes (N_e) (Lynch, 2007). In the following section, we compare and contrast genome architecture of these small genomes to illuminate how these structural similarities and differences may influence sRNA evolution and function.

Guanine-cytosine (GC) content can play a major role in sRNA interactions. On the one hand, high GC content can potentially hinder sRNA - target RNA binding and on the other hand, low GC content may reduce the stability of RNA secondary structures (Chan et al., 2009; Fallmann et al., 2017; Barik and Das, 2018). In free-living bacteria, sRNAs have an average GC content of ~49%, with Gram positive bacteria having a lower sRNA GC content (~44%) compared to Gram negative bacteria (~52%), reflecting the average GC composition of their genomes of ~42% and 54%, respectively (Barik and Das, 2018).

In contrast to free-living bacteria, organelle and bacterial symbiont genomes tend to have significantly lower GC content, though exceptions are known (McCutcheon and Moran, 2011; Smith and Keeling 2015). Interestingly, in free-living bacteria and bacterial symbionts, there is a positive relationship between average genome adenine-thymine (AT) richness and antisense sRNA expression (Lloréns-Rico et al., 2016). Lloréns-Rico et al. (2016), hypothesize that among bacteria with small AT rich genomes, spurious promoter sites can result in the production of antisense

sRNAs and create transcriptional noise. Although such antisense RNAs may initially emerge as transcriptional noise it is possible that beneficial RNAs could be retained due to purifying selection. Consequently, the more AT rich a genome is, potentially, the higher the frequency that raw material can be generated for the biosynthesis of regulatory sRNAs.

Within free-living bacteria, changes to genome organization can affect sRNA gene expression and regulation (Dorman, 2013; Lagomarsino et al., 2015). For example, recombination and mobile element insertion events can disrupt the proper functioning of existing sRNAs and/or provide new potential sites for sRNAs to emerge (Dutcher and Raghavan, 2018). This is illustrated by *Salmonella enterica* serovar Typhimurium str. 14028S, in which genome rearrangements have resulted in the loss of the sRNA EcsR1 compared to its relative *E. coli* (Raghavan et al., 2015). Small genomes that have lost the ability to recombine possess genomes that are highly stable, thus retaining syntenic gene order when compared to related taxa (Moran and Bennett, 2014). In turn, we hypothesize that sRNAs and their targets may be maintained longer in small genomes that are highly syntenic compared to dynamic genomes that rearrange more frequently.

In free-living bacteria, sRNAs have been identified in non-coding regions (Tsai et al., 2015). As such, the percent of non-coding DNA in a genome maybe an important determinant for the emergence of novel sRNAs in small genomes. In general, small symbiont genomes possess a high coding density with only 10 – 20% non-coding DNA compared to some organelles (Table 1.1), limiting the portion of their genomes where sRNAs can evolve in non-coding regions. Plant and animal mitochondria differ vastly in the size and percentage of non-coding DNA in their genomes (Table 1.1). Plants are hypothesized to have larger mitochondrial genomes due to lower mutation rates and smaller N_e compared to animal mitochondria, resulting in the accumulation of introns and fragmented genes (Lynch 2007; however, see Sloan et al., 2012; Christensen, 2013). Similar to mitochondria, plastids also show variation in genome size and range from 5% – 80% in

non-coding DNA (Table 1.1). Given that some organelles have large amounts of non-coding DNA, this may increase the probability of sRNAs evolving within non-coding sequences for these genomes. Furthermore, widespread transcription has been observed across non-coding regions of plant mitochondria and plastids, further increasing the chances of sRNA evolution within IGRs of these genomes (Lima and Smith, 2017). However, as of yet the increase in intergenic regions does not seem to result in more IGR-derived sRNAs. For example, within the large (~7Mb) *Silene noctiflora* (nightflowering silene) mitochondria, only four candidate sRNAs were identified within the intergenic spacer regions (Wu et al., 2015b).

Future Research

With improvements in sequencing and bioinformatics technology, a wide diversity of sRNAs have been identified in both organelle and bacterial endosymbiont genomes. This body of research provides us with the foundation to move forward and better understand the role of small RNAs in small genomes. To further understand the evolution of sRNAs in small genomes three main areas of sRNA research need to be explored further: *1*. The regulatory role of sRNAs in small genomes, *2*. The relative importance of symbiont sRNAs in the regulation of its symbiosis with its host, *3*. The role that genome architecture plays in sRNA evolution.

1. The regulatory role of sRNAs in small genomes:

One of the main challenges faced in both free-living and symbiotic bacterial systems is determining if identified sRNAs have a regulatory function, are by products of RNA degradation or processing, or if they are transcriptional noise (Georg and Hess, 2011; Jackowiak et al., 2011; Lloréns-Rico et al., 2016). Evidence that is commonly used to support the regulatory role of putative sRNAs, is the characterization of differential sRNA expression in response to

environmental or developmental changes. Though the characterization of sRNA expression is an important first step, experimental validation remains crucial to determine if differentially expressed sRNAs are regulatory (*i.e.* either interacting with an RNA, protein, or metabolite) and not artifacts of other processes such as RNA processing within the organism. This is illustrated in the recently described class of plant associated sRNAs; clustered organelle sRNA (cosRNAs). cosRNAs are expressed within the intergenic 5' and 3' regions of plastid and mitochondrial genes respectively (Ruwe et al., 2012; Ruwe et al., 2016; Cavaiuolo et al., 2017). While cosRNAs have been primarily identified in the organelles of *Arabidopsis thaliana* and *Chlamydomonas reinhardtii,* their size and location coincide with other previously identified sRNAs in other plant species' organelles (Ruwe et al., 2012; Ruwe et al., 2016; Cavaiuolo et al., 2017). Importantly, it has been shown that cosRNAs may be RNA binding protein footprints (Cavaiuolo et al., 2017) Though many of these small genome systems can be challenging to work with because many symbionts and organelles with small genomes are unculturable. In turn, we suggest the following experimental approaches to conduct on unculturable genomes to help improve our understanding sRNAs in these systems.

i. Parallel Proteomic and Transcriptomic Data Analysis

Though not a direct measure of sRNA functionality, a shift in mRNA expression and/or protein expression that correlates to changes in sRNA expression can be indicative of sRNA activity and potential sRNA targets. The characterization of proteomes can be especially helpful for identifying potential functional sRNAs that influence translation but not mRNA abundance. Though proteomic experiments present their own challenges, there are studies that have characterized the proteomes of organelles and bacterial symbionts in various environmental conditions (*e.g.* Calvo and Mootha, 2010; Kosmala et al., 2012; Morgenstern et al., 2017; Ramsey et al., 2017; Tamburino et al., 2017; Thagela et al., 2018). While the suggestion of the integration of these two data types is not novel, within the current literature, there are few cases of integration.

This approach of integrating sRNA expression and proteome data can be successful in understanding the potential functions of sRNAs as demonstrated by Noro et al. (2017), who used these datasets to help validate the function and target of artificial sRNAs. Hansen and Degnan (2014), also used this approach to help identify potential candidate sRNAs in the bacterial symbiont *Buchnera.*

ii. Characterizing the sRNA interactome:

Determining if expressed sRNAs can interact with their predicted targets is a vital step in understanding their functionality. In recent years, there has been significant innovation of high-throughput techniques that have helped describe the sRNA regulatory networks of various bacterial systems. These technologies include CLASH, RIL-seq, CLIP-seq, GRIL-seq, MAP-seq, RIL-seq and RIP-seq (reviewed in Saliba et al. 2017). As is the norm, these methods have been developed and widely used in model systems such as *E. coli,* however, we have not found any examples of these techniques being adapted to work in unculturable systems such as bacterial symbionts or organelles.

However, there are examples of low-throughput methods of sRNA interactome determination being used in unculturable or genetically intractable systems. A widely used method to determine sRNA function in free-living bacteria is the implementation of two plasmid translational fusion systems (*e.g.* Urban and Vogel., 2007; Gogol et al., 2011; Bobrovskyy and Vanderpool, 2016; Ivain et al., 2017). These systems entail cloning the candidate sRNA in a plasmid with an inducible promotor and its predicted target in a second *gfp* fusion plasmid. If the sRNA interacts with its predicted target, then there will be change in *gfp* florescence when compared to the controls. This system has been used to heterologously determine the functionality of sRNAs from various genetically intractable organisms (*e.g. Vibrio* Chang et al., (2015); *Sphingopyxis granuli* García-Romero (2017); the cyanobacterium *Synechocystis sp.* PCC 6803

Rübsam et al., (2018)). We predict that the more widespread application of this experimental method within organelle and bacterial symbiont systems will greatly help increase our understanding of sRNA function within these systems. Currently, there are promising results of the successful implementation of this experimental approach in the unculturable bacterial symbiont *Buchnera* (Thairu et al., 2018).

<u>iii. Characterization of sRNA-interacting proteins:</u>

Within free-living bacteria, sRNA binding proteins (*i.e.* Hfq, CsrA and ProQ) have been identified and subsequently used in various types of immunoprecipitation assays to help validate the functionality of predicated sRNAs (*e.g.* Faner and Feig, 2013; Holmqvist et al. 2016; Smirnov et al., 2016). However, within organelles and bacterial symbionts, sRNA binding proteins have not yet been characterized. The identification of a protein(s) that binds with predicted sRNAs will help provide further evidence for sRNA functionality. To help identify potential sRNA binding proteins we suggest that the utilization of global, non-targeted approaches such as the gradient profiling by sequencing (Grad-seq) pipeline will help better elucidate the sRNA-protein landscape in these small genomes (Smirnov et al., 2017). Within this pipeline cell lysates, which include RNA-protein complexes, are first fractionated and then each fraction undergoes characterization via RNA-seq and mass spectrometry (Smirnov et al., 2017). When these two data sets are combined, ideally, sRNAs that are interacting with similar proteins will cluster together (Smirnov et al., 2017). By grouping similarly behaving transcripts, not only can RNA binding proteins be identified, but any associated sRNAs. The Grad-seq pipeline has been used in *Salmonella* to successfully identify new ProQ-sRNA interactions (Smirnov et al., 2016).

2. Determine the relative importance of symbiont sRNAs in the regulation of its symbiosis with its host

Inter-domain crosstalk has been shown to occur via sRNAs, therefore understanding the role of sRNAs within these symbiotic relationships will help us further understand the evolution of bacterial symbioses (Knip et al., 2014; Zhou et al., 2017). If sRNAs in organelles and symbionts have evolved novel strategies for gene regulation in response to their symbiotic lifestyle this may redefine our understanding of the evolutionary potential of host-restricted organisms, especially in the face of genetic drift. This has profound implications for our understanding of symbiont and organelle evolution and the generation of adaptive traits for animal hosts.

3. Determine the role that genome architecture plays in sRNA evolution

The highly reduced genomes of organelles and bacterial symbionts allow us to ask various questions about sRNA evolution. For example, can the AT bias in most small genomes lead to an increase in the frequency of the emergence of antisense sRNA candidates compared to free-living bacteria? Also, are sRNAs maintained longer in evolutionary time in small genomes that are more stable in genome structure compared to bacteria with more dynamic genomes? By addressing these evolutionary questions, we can gain a better understanding of sRNA evolution within bacteria.

Conclusion

Small RNAs have emerged as vital regulators across all domains of life. Despite their highly reduced genomes, bacterial symbionts and organelles appear to be no exception encoding a wide diversity of transcribed putative sRNAs. As of yet the potential roles for these putative sRNAs in cellular or organellar regulation are poorly understood. This is driven in part by limitations of experimental methodologies but also the fact that in well studied free-living bacteria these small

metabolically inexpensive regulators not only have varied roles, ranging from regulating translation, to stabilizing mRNAs or initiating mRNA degradation; but also have varied sequence and structure. However, given the widespread purifying selection that occurs in reduced genomes within functional non-coding RNA sequences (Lambert and Moran 1998; Hansen and Moran 2012); we hypothesize that regulatory sRNAs can be maintained and potentially evolve in these small genomes in a way that regulatory proteins have not and cannot. We expect that ongoing research and application of novel "-omic" methodologies have the potential to reveal the evolution and function of these putative sRNAs.

References:

Archibald JM. The puzzle of plastid evolution. *Curr Biol* 2009; **19**: R81-8.

Baginsky S, Shteiman-Kotler A, Liveanu V, et al. Chloroplast PNPase exists as a homo-multimer enzyme complex that is distinct from the *Escherichia coli* degradosome. *RNA* 2001; **7**:1464–75.

Bandyra KJ, Sinha D, Syrjanen J, et al. The ribonuclease polynucleotide phosphorylase can interact with small regulatory RNAs in both protective and degradative modes. *RNA* 2016; **22:**360–372.

Barik A, Das S. A comparative study of sequence- and structure-based features of small RNAs and other RNAs of bacteria. *RNA Biol* 2018; **15**:95–103.

Barshad G, Marom S, Cohen T, et al. Mitochondrial DNA Transcription and its regulation: An evolutionary perspective. *Trends Genet* 2018; **9**:682-692.

Bartel DP. Metazoan MicroRNAs. *Cell* 2018; **173:**20–51.

Bennett GM, Chong RA. Genome-Wide transcriptional dynamics in the companion bacterial symbionts of the glassy-winged sharpshooter (Cicadellidae: *Homalodisca vitripennis*) reveal differential gene expression in bacteria occupying multiple host organs. *G3* 2017; **7**:3073–3082.

Bilusic I, Popitsch N, Rescheneder P, et al. Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* 2014; **11**:641–54.

Bobrovskyy M, Vanderpool CK. Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Mol Microbiol* 2016; **99**:254–273.

Bottje WG, Khatri B, Shouse, SA, et al. Identification and differential abundance of mitochondrial genome encoding small RNAs (mitosRNA) in breast muscles of modern broilers and unselected chicken breed. *Front Physiol* 2017; doi:10.3389/fphys.2017.00816.

Brant EJ, Budak H. Plant Small Non-coding RNAs and Their Roles in Biotic Stresses. *Front Plant Sci* 2018; doi:10.3389/fpls.2018.01038.

Calvo S, Mootha VK. The Mitochondrial Proteome and Human Disease. *Annu Rev Hum Genet* 2010; **11**:25–44.

Cavaiuolo M, Kuras R, Wollman FA, et al. Small RNA profiling in *Chlamydomonas*: Insights into chloroplast RNA metabolism. *Nucleic Acids Res* 2017; **45**:10783–10799.

Chan CY, Carmack CS, Long DD, et al. A structural interpretation of the effect of GC-content on efficiency of RNA interference. *BMC Bioinformatics* 2009; doi: 10.1186/1471-2105-10-S1-S33.

Chang H, Replogle JM, Vather N, et al. A cis-regulatory antisense RNA represses translation in *Vibrio cholerae* through extensive complementarity and proximity to the target locus. *RNA Biol* 2015; **12**:136–148.

Charles H, Balmand S, Lamelas A, et al. A genomic reappraisal of symbiotic function in the aphid *Buchnera* symbiosis: Reduced transporter sets and variable membrane organizations. *PLoS One* 2011; doi: 10.1371/journal.pone.0029096.

Chen C, Zeng Z, Liu Z, et al. Small RNAs, emerging regulators critical for the development of horticultural traits. *Hortic Res* 2018; doi: 10.1038/s41438-018-0072-8.

Chen H, Zhang J, Yuan G, et al. Complex interplay among DNA modification, noncoding RNA expression and protein-coding RNA expression in *Salvia miltiorrhiza* chloroplast genome. *PLoS One* 2014; doi: 10.1371/journal.pone.0099314.

Chen Z, Sun Y, Yang X, et al. Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS One* 2017; e0176458 doi: 10.1371/journal.pone.0176458.

Chi W, He B, Mao J, et al. Plastid sigma factors: Their individual functions and regulation in transcription. *Biochim Biophys Acta - Bioenerg* 2015; **1847**:770–778.

Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol Evol* 2013; **5**:1079–1086.

Cobb LJ, Lee C, Xiao J, et al. Naturally occurring mitochondrial-derived peptides are agedependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging* 2016; **8**:796–808.

Cognat V, Morelle G, Megel C, et al. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. *Nucleic Acids Res* 2017; **45**:3460–3472.

Cook H, Ussery DW. Sigma factors in a thousand *E. coli* genomes. *Environ Microbiol* 2013; **15**:3121–3129.

Degnan PH, Ochman H, Moran NA. Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont *Buchnera*. *PLoS Genet* 2011; doi: 10.1371/journal.pgen.1002252.

Dietrich A, Wallet C, Iqbal RK, et al. Organellar non-coding RNAs: Emerging regulation mechanisms. *Biochimie* 2015; **117**:48–62.

Dorman CJ. Genome architecture and global gene regulation in bacteria: Making progress towards a unified model? *Nat Rev Microbiol* 2013; **11**:349–355.

Durand S, Tomasini A, Braun F, et al. sRNA and mRNA turnover in Gram-positive bacteria. *FEMS Microbiol Rev* 2015; **39**:316–30.

Dutcher HA, Raghavan R. Origin, Evolution, and Loss of Bacterial Small RNAs. *Microbiol. Spectr* 2018; doi: 10.1128/microbiolspec.RWR-0004-2017.

Fallmann J, Will S, Engelhardt J, et al. Recent advances in RNA folding. *J. Biotechnol.* 2017; **261**: 97–104.

Faner MA, Feig AL. Identifying and characterizing Hfq-RNA interactions. *Methods* 2013; **63:**144–159.

Fischer, S. Benz J, Späth B, et al. Regulatory RNAs in *Haloferax volcanii. Biochem Soc Trans* 2011; **39:**159–62.

Gabaldón T, Huynen MA. From endosymbiont to host-controlled organelle: The hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput Biol* 2007; **3**:2209–2218.

García-Romero I, Förstner KU, Santero E, et al. SuhB, a small non-coding RNA involved in catabolite repression of tetralin degradation genes in *Sphingopyxis granuli* strain TFA. *Environ Microbiol* 2018; **20:**3671–3683.

Georg J, Hess WR. cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* 2011; **75**:286–300.

Georg J, Honsel A, Voß B, et al. A long antisense RNA in plant chloroplasts. *New Phytol* 2010; **186**:615–622.

Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci* 2011; **108,** 12875–12880.

Gonzalez-Ibeas D, Blanca J, Donaire L, et al. Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. *BMC Genomics* 2011; doi: 10.1186/1471-2164-12-393.

Gray MW. Mitochondrial Evolution. *Cold Spring Harb Perspect Biol* 2012; doi: 10.1101/cshperspect.a011403.

Gualberto JM, Mileshina D, Wallet C, et al. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 2014; **100**:107–120.

Gualberto JM, Newton KJ. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu Rev Plant Biol* 2017; **68**:225–252.

Güell M, van Norrt V, Yus E, et al. Transcriptome complexity in a genome-reduced bacterium. *Science* 2009; **326**:1268–71.

Hackenberg M, Gustafson P, Langridge P, et al. Differential expression of microRNAs and other small RNAs in barley between water and drought conditions. *Plant Biotechnol J* 2015; **13**:2–13.

Hackenberg M, Huang PJ, Huang CY, et al. A Comprehensive expression profile of microRNAs and other classes of non-coding small RNAs in barley under phosphorous-deficient and-sufficient conditions. *DNA Res.* 2013; **20**:109–125.

Hansen AK, Degnan PH. Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J* 2014; **8**:2490–502.

Hansen AK, Moran NA. Altered tRNA characteristics and 3′ maturation in bacterial symbionts with reduced genomes. *Nucleic Acids Res*. 2012; **40**:7870–7884.

Hansen AK, Moran NA. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol* 2014; **23**:1473–96.

Hashimoto Y, Niikura T, Tajima H, et al. A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and A. *Proc. Natl. Acad. Sci*. 2001; **98**:6336–6341.

Himmelreich R, Plagens H, Hilbert H, et al. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res* 1997; **25**:701–712.

Holmqvist E, Wright PR, Li L, *et al.* Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* 2016; **35,** 991–1011.

Hotto AM, Germain A, Stern DB. Plastid non-coding RNAs: Emerging candidates for gene regulation. *Trends Plant Sci* 2012; **17**:737–744.

Hotto AM, Huston ZE, Stern DB. Overexpression of a natural chloroplast-encoded antisense RNA in tobacco destabilizes 5S rRNA and retards plant growth. *BMC Plant Biol* 2010; doi: 10.1186/1471-2229-10-213.

Hotto AM, Schmitz RJ, Fei Z, et al. Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the *Arabidopsis* transcriptome. *G3* 2011; **1**:559–70.

Hua C, Zhao JH, Guo HS, Trans-Kingdom RNA Silencing in Plant–Fungal Pathogen Interactions. *Mol Plant* 2018; **11:**235–244.

Hutchison CA, Chuang RY, Noskov VN, et al. Design and synthesis of a minimal bacterial genome. *Science* 2016; doi: 10.1126/science.aad6253.

Itaya A, Bundschuh R, Archual AJ, et al. Small RNAs in tomato fruit and leaf development. *Biochim Biophys Acta* 2008; **1779**:99–107.

Ivain L, Bordeau V, Eyraud Alex, et al. An in vivo reporter assay for sRNA-directed gene control in Gram-positive bacteria: Identifying a novel sRNA target in *Staphylococcus aureus. Nucleic Acids Res* 2017; **45:**4994–5007.

Jackowiak P, Nowacka M, Strozycki PM, Figlerowicz M. RNA degradome-its biogenesis and functions. *Nucleic Acids Res* 2011; **39**:7361–7370.

Keam S, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life* 2015; **5:**1638–1651.

Kim KH, Son JM, Benayoun BA, et al. The Mitochondrial-Encoded Peptide MOTS-c Translocates to the Nucleus to Regulate Nuclear Gene Expression in Response to Metabolic Stress. Cell *Metab* 2018; **28**:516–524.

Kim D, Thairu MW, Hansen AK. Novel Insights into Insect-Microbe Interactions-Role of Epigenomics and Small RNAs. *Front Plant Sci* 2016; doi: 10.3389/fpls.2016.01164.

Knip M, Constantin ME, Thordal-Christensen H. Trans-kingdom cross-talk: small RNAs on the move. *PLoS Genet.* 2014; doi: 10.1371/journal.pgen.1004602.

Kosmala A, Perlikowski D, Pawłowicz I, et al. Changes in the chloroplast proteome following water deficit and subsequent watering in a high- and a low-drought-tolerant genotype of *Festuca arundinacea*. *J. Exp. Bot.* 2012; **63:**6161–6172.

Lagomarsino MC, Espéli O, Junier I. From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. *FEBS Lett* 2015; **589**:2996–3004.

Lambert JD, Moran NA. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc Natl Acad Sci* 1998; **95:**4458–4462.

Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: Mt-Genome Organization and Evolution in Nonbilaterian Lineages. *Genome Biol Evol* 2016; **8**:2896–2913.

Lee C, Zeng J, Drew B, et al. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 2015; **21**, 443–454.

Leszczyniecka M, DeSalle R, Kang DC, et al. The origin of polynucleotide phosphorylase domains. *Mol Phylogenet Evol* 2004; **31**:123–130.

Liere K, Weihe A, Börner T. The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol* 2011; **168**:1345–1360.

Lima MS, Smith DR. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biol Evol* 2017; **9**:2650–2657.

Liu W, Fang L, Li M, et al. Comparative genomics of *Mycoplasma*: Analysis of conserved essential genes and diversity of the pan-genome. *PLoS One* 2012; doi: 10.1371/journal.pone.0035698.

Lloréns-Rico V, Cano T, Kamminga T, et al. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv* 2016; doi: 10.1126/sciadv.1501363.

Lluch-Senar M, Vallmitjana M, Querol et al. A new promoterless reporter vector reveals antisense transcription in *Mycoplasma genitalium*. *Microbiology* 2007; **153**:2743–2752.

Lu H, Price DRG, Wikramanayake A, et al. Ontogenetic differences in localization of glutamine transporter ApGLNT1 in the pea aphid demonstrate that mechanisms of host/symbiont integration are not similar in the maternal versus embryonic bacteriome. *Evodevo* 2016; **7**:1 doi: 10.1186/s13227-015-0038-y.

Lung B, Zemann A, Madej MJ, et al. Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res* 2006;**34**, 3842–52.

Lynch, M. *The Origins of Genome Architecture.* Massachusetts: Sinauer Associates, Inc, 2007

Ma H, Weber GM, Wei H, et al. Identification of mitochondrial genome-encoded small RNAs related to egg deterioration caused by postovulatory aging in rainbow trout. *Mar Biotechnol* 2016; **18**:584–597.

Marker C, Zemann A, Terhörst T, et al. Experimental RNomics: Identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol* 2002; **12**:2002–2013.

Martinez G. tRNA-derived small RNAs: New players in genome protection against retrotransposons. *RNA Biology* 2018; **15:**170–175.

Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, et al. Evolution of small prokaryotic genomes. *Front Microbiol* 2015; doi: 10.3389/fmicb.2014.00742

Matelska D, Kurkowska M, Purta E, et al. Loss of conserved noncoding RNAs in genomes of bacterial endosymbionts. *Genome Biol Evol* 2016; **8**:426–438.

Mayoral JG, Hussain M, Joubert D, et al. *Wolbachia* small noncoding RNAs and their role in cross-kingdom communications. *Proc Natl Acad Sci U S A* 2014; **111**:18721–6.

McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2011; **10**:13–26.

McCutcheon JP. From microbiology to cell biology: when an intracellular bacterium becomes part of its host cell. *Curr Opin Cell Biol* 2016; **41**:132–136.

Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol* 2014; **68**:195–215.

Moran Y, Agron M, Praher D, et al. The evolutionary origin of plant and animal microRNAs. *Nature Ecology and Evolution* 2017; doi: 10.1038/s41559-016-0027.

Morgenstern, M. Stiller SB, Lübbert Philipp, et al. Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. *Cell Rep* 2017; **19:**2836–2852.

Narra HP, Schroeder CLC, Sahni A, et al. Small regulatory RNAs of *Rickettsia conorii*. *Sci. Rep.* 2016; doi: 10.1038/srep36728.

NCBI National Center for Biotechnology Information.
https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/

Noro E, Mori M, Makino G, et al. Systematic characterization of artificial small RNA-mediated inhibition of Escherichia coli growth. *RNA Biol* 2017; **14:**206–218.

Ortelt J, Link G. Plastid gene transcription: Promoters and RNA polymerases. *Methods Mol Biol* 2014; **1132**:47–72.

Parks DH, Chuvovhina M, Waite DW, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018; doi: 10.1038/nbt.4229/

Perrin R, Lange H, Grienenberger JM, Gagliardi D. AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in *Arabidopsis thaliana* mitochondria. *Nucleic Acids Res* 2004; **32:**5174–5182.

Pett W, Ryan JF, Pang K, et al. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. *Mitochondrial DNA* 2011; **22**, 130–42.

Quesada V. The roles of mitochondrial transcription termination factors (MTERFs) in plants. *Physiol Plant* 2016; **157**, 389–399.

Raghavan R, Kacharia FR, Millar JA, et al. Genome rearrangements can make and break small RNA genes. *Genome Biol Evol* 2015; **7**:557–566.

Ramsey JS, Chavez JD, Johnson R, et al. Protein interaction networks at the host–microbe interface in *Diaphorina citri*, the insect vector of the citrus greening pathogen. *R. Soc. Open Sci.* 2017; doi: 10.1098/rsos.160545.

Ro S, Ma H, Park C, et al. The mitochondrial genome encodes abundant small noncoding RNAs. *Cell Res* 2013; **23**:759–74.

Rompala GR, Mounier A, Wolfe CM, et al. Heavy Chronic Intermittent Ethanol Exposure Alters Small Noncoding RNAs in Mouse Sperm and Epididymosomes. *Front Genet* 2018; **9**: 1–14.

Rübsam H, Kirsch F, Reimann V, et al. The iron-stress activated RNA 1 (IsaR1) coordinates osmotic acclimation and iron starvation responses in the cyanobacterium *Synechocystis sp.* PCC 6803. *Environ Microbiol* 2018; **20:**2757–2768.

Ruwe H, Schmitz-Linneweber C. Short non-coding RNA fragments accumulating in chloroplasts: Footprints of RNA binding proteins? *Nucleic Acids Res* 2012; **40**:3106–3116.

Ruwe H, Wang G, Gusewski, S, et al. Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res* 2016; **44**:7406–7417.

Sachs JL, Essenberg CJ, Turcotte MM. New paradigms for the evolution of beneficial infections. *Trends Ecol Evol* 2011; **26**:202–209.

Saliba AE, Santos SC. Vogel J. New RNA-seq approaches for the study of bacterial pathogens. *Current Opinion in Microbiology* 2017; **35**:78–87.

Schroeder CLC, Narra HP, Rojas M, et al. Bacterial small RNAs in the Genus *Rickettsia*. *BMC Genomics* 2015; doi: 10.1186/s12864-015-2293-7.

Schroeder CLC, Narra HP, Sahni A, et al. Identification and characterization of novel small RNAs in *Rickettsia prowazekii*. *Front. Microbiol* 2016; doi: 10.3389/fmicb.2016.00859.

Schroeder CLC, Narra HP, Sahni A, et al. Transcriptional profiling of *Rickettsia prowazekii* coding and non-coding transcripts during in vitro host-pathogen and vector-pathogen interactions. *Ticks Tick Borne Dis* 2017; **8**:827–836.

Shigenobu S, Watanabe H, Hattori M, et al. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp*. APS. *Nature* 2000 **407:**81–6.

Singh A, Gautam V, Singh S, et al. Plant small RNAs: advancement in the understanding of biogenesis and role in plant development. *Planta* 2018; **248:**545–558.

Siqueira FM, de Morais GL, Higashi S, et al. *Mycoplasma* non-coding RNA: Identification of small RNAs and targets. *BMC Genomics* 2016; doi:10.1186/s12864-016-3061-z.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol Evol* 2012; **4**:294–306.

Smalheiser NR, Lugli G, Thimmapuram J, et al. Mitochondrial small RNAs that are up-regulated in hippocampus during olfactory discrimination training in mice. *Mitochondrion* 2011; **11**:994–995.

Smirnov A, Förstner KU, Holmqvist E, et al. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci USA* 2016; **113,** 11591–11596.

Smirnov A, Schneider C, Hör J, et al. Discovery of new RNA classes and global RNA-binding proteins. *Curr Opin in Micro* 2017; **39:**1–9.

Smith DR, Hua J, Lee RW. Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. *Curr Genet* 2010; **56**:427–438.

Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci USA* 2015; **112**:10177–84.

Smith DR. Plastid genomes hit the big time. *New Phytol* 2018; **219**:491–495.

Smith DR. Updating our view of organelle genome nucleotide landscape. *Front Genet* 2012; doi:.10.3389/fgene.2012.00175

Sobrero P, Valverde C. The bacterial protein Hfq: Much more than a mere RNA-binding factor. *Crit Rev in Micro* 2012; **38:**276–299.

Sripada L, Tomar D, Prajapati P, et al. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS One* 2012; doi: 10.1371/journal.pone.0044873.

Sun X, Zhulin I, Wartell RM. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* 2002 **30**:3662–3671.

Tamames J. Evolution of gene order conservation in prokaryotes. *Genome Biol.* 2001; **2** doi 10.1186/gb-2001-2-6-research0020

Tamburino R, Vitale M,Ruggiero A, et al. Chloroplast proteome response to drought stress and recovery in tomato (*Solanum lycopersicum L.*). *BMC Plant Biol.* 2017; doi: 0.1186/s12870-017-0971-0.

Thagela, P. Yadav, Ravindra K, et al. Salinity induced changes in the chloroplast proteome of the aquatic pteridophyte *Azolla microphylla*. *Symbiosis* 2018; **75:** 61–67.

Thairu MW, Cheng S, Hansen AK. A sRNA in a reduced mutualistic symbiont genome regulates its own gene expression. *Mol Ecol* 2018; **27**:1766-1776.

Thomason MK, Storz G. Bacterial Antisense RNAs: How many are there, and what are they doing?*. *Annu Rev Genet* 2010; **44**:167–188.

Thorpe HA, Bayliss SC, Hurst LD, et al. Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. *Genetics* 2017; **206:**363–376.

Toft C, Andersson SGE. Evolutionary microbial genomics: Insights into bacterial host adaptation. *Nat Rev Genet* 2010; **11**:465–475.

Touchon M, Hoede C, Tenaillon O, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* 2009; doi:10.1371/journal.pgen.1000344.

Tsai CH, Liao R, Chou B, et al. Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. *J Bacteriol* 2015; **197**:40–50.

Urban JH, Vogel J. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.* 2007; **35:**1018–1037.

van Ham RCHJ, Kamerbeek J, Palacios C, et al. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci* 2003; **100:**581–586.

Viegas SC, Pfeiffer V, Sittka A, et al. Characterization of the role of ribonucleases in *Salmonella* small RNA decay. *Nucleic Acids Res* 2007; **35**:7651–7664.

Wagner EGH, Romby P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet* 2015; **90:**133–208.

Wadler CS, Vanderpool CK. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci.* 2007; **104**:20454–20459.

Wang L, Yu X, Wang H, et al. A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). *BMC Genomics* 2011; doi 10.1186/1471-2164-12-289.

Weiberg A, Bellinger M, Jin H. Conversations between kingdoms: small RNAs. *Curr Opin Biotechnol* 2015; **32:**207–15.

Woolfit M, Algama M, Keith JM, et al. Discovery of putative small non-coding RNAs from the obligate intracellular bacterium *Wolbachia pipientis*. *PLoS One* 2015; doi: 10.1371/journal.pone.0118595.

Wu Z, Cuthbert JM, Taylor DR, et al. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc Natl Acad Sci* 2015a; **112**:10185–10191.

Wu Z, Stone JD, Štorchová H, et al. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC Genomics* 2015b; doi: 10.1186/s12864-015-2155-3.

Zghidi-Abouzid O, Merendino L, Buhr F,et al. Characterization of plastid *psbT* sense and antisense RNAs. *Nucleic Acids Res* 2011;**39**: 5379–87.

Zhelyazkova P, Sharma CM, Forstner KU, et al. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 2012; **24**:123–136.

Zhou G, Zhou Y, Chen X. New insight into inter-kingdom communication: horizontal transfer of mobile small RNAs. *Front Microbiol* 2017; doi: 10.3389/fmicb.2017.00768.

Table 1.1: Overview of organelle and host-restricted symbiont genome characteristics and evidence of sRNA expression and function in these small genomes. ("-" indicates unknown/variable)

	Plastids	Mitochondria		Host Restricted Bacterial Symbionts (examples)		Host Restricted Bacterial Symbionts (general)	Free living Bacteria (example)
		Plant	**Animal**	**Pathogenic** *Mycoplasma sp.*	**Mutualistic** *Buchnera sp.*		*Escherichia coli* K-12
Common Genome size range	~75Kb – ~200Kb[1]	~13Kb – ~700Kb[22]	~14Kb – ~20Kb[24]	~513Kb – ~1.4Mb[36,37]	~416 – ~642 Kb[44]	~112Kb – ~5.85Mb[49,50]	~4.5Mb - ~5.2Mb[58]
Lower limit example	~11,348nt *Pilostyles aethiopica*[1]	<13Kb *Polytomella*[2]	~10,326nt *Mnemiopsis leidyi*[24]	~580Kb *Mycoplasma genitalium*[37]	~416Kb *Buchnera aphidicola-Cc*[44]		~4.5Mb *E. coli* strain: UMN026[58]
Upper limit example	~1.13Mb *Corynoplastis japonica*[1]	~11.3Mb *Silene conica*[2]	~51Kb *Clathrina clathrus*[24]	~1.3Mb *Mycoplasma penetrans*[37]	~641Kb *Buchnera aphidicola-Ap*[44]		~5.2Mb *E. coli* strain: ED1a[58]
Average percent coding	~5% – ~80%[2]	~1% – ~99%[2]		~80% – ~90%[38]	~85%[44]	~80%–~90%[38]	~87%[58]
Average number of protein coding genes	~20–~250[2]	~10[25]–~54[26]	~33[27]–~66[2]	~475–~1037[37]	~357–~564[45]	~137–~5509[49,50]	~4306~5129[58]
Average AT richness	~45%–~87%[3]	~32%–~78%[3]	~48%–~84%[3]	~68%[37]	~73%–~79%[44]	~32%–~83%[49,50]	~49%~50%[58]
Number of retained Transcription/Sigma factors in symbiont/bacterial genome	0[4,5]	0[28,29]		Variable, at least >8[39,40]	2 sigma factors[46]	-	4-7 known sigma factors[59]

28

	Plastids	Mitochondria		Host Restricted Bacterial Symbionts (examples)		Host Restricted Bacterial Symbionts (general)	Free living Bacteria (example)
		Plant	**Animal**	**Pathogenic**	**Mutualistic**		
				Mycoplasma sp.	*Buchnera sp.*		*Escherichia coli* K-12
Number of retained Transcription/Sigma factors in nuclear genome	1 – 7 sigma factors [4,5] Multiple mitochondrial transcription factors (mTERFs) [6]	Multiple mitochondrial transcription factors (mTERFs) [6,28]	2 – 4 mitochondrial transcription factors (mTERFs) [6,29]	0	0	-	-
Organisms/tissues demonstrating sRNA expression	*Arabidopsis thaliana,*[7-12] *Brassica rapa,*[13] *Chlamydomonas reinhardtii,*[14] *Cucumis melo,*[15] *Hordeum vulgare,*[16,18] *Nicotiana tabacum,*[19] *Populus sp.,*[9] *Salvia miltiorrhiza,*[20] *Silene noctiflora*[21]	*Arabidopsis thaliana*[11,12]	Human and Mouse Tissues,[7,30] Human Cell lines[31]	*M. genitalium,*[41] *M. pneumoniae,*[42] *M. hyopneumoniae, M. flocculare and M.hyorhinis*[43]	*B. aphidicola Ap,Ak,Ua,Sg*[47]	*Baumannia,*[51] *Rickettsia sp.,*[52-55] *Wolbachia*[56,57]	-
Studies demonstrating that sRNAs candidates are regulatory	PNPase mutants,[9] Response to Drought,[13] Nutrient stress[17,18]		Behavior,[32] Tissue development,[33] Cancer phenotype,[34] Alcohol exposure[35]	Environmental Stress[43]	Development[48]	Host infection[56,57]	Numerous

1. Smith, 2018; **2.** Smith and Keeling, 2015; **3.** Smith, 2012; 4. Ortelt and Link, 2014; 5. Chi et al., 2015; 6.Quesada, 2016; 7. Lung et al., 2006 8. Georg et al., 2010; 9. Hotto et al., 2011, **10.** Zghidi-Abouzid et al., 2011;**11.** Cognat et al., 2017; **12.** Ruwe et al., 2016; **13.** Wang et al., 2011; **14.** Cavaiuolo et al., 2017; **15.** Gonzalez-Ibeas et al., 2011; **16.** Zhelyazkova et al., 2012; **17.** Hackenberg et al., 2013; **18.** Hackenberg et al., 2015; **19.** Hotto et al., 2010; **20.** Chen et al., 2014; **21.** Wu et al., 2015b; **22.** Gualberto & Newton, 2017; **23.** Gualberto et al., 2014; **24.** Lavrov and Pett, 2016; **25.** Smith et al., 2010; **26.** Wu et al., 2015a; **27.** Pett et al., 2011; **28.** Liere, et al., 2011; **29.** Barshad et al., 2018; **30.** Ro et al., 2013; **31.** Sripada et al., 2012; **32.** Smalheiser et al., 2011; **33.** Ma et al., 2016; **34.** Bottje et al., 2017; **35.** Rompala et al., 2018; **36.** NCBI 2018; **37.** Liu et al., 2012; **38.** Parks et al., 2018; **39.** Himmelreich et al., 1997; **40.** Hutchison et al., 2016; **41.** Lluch-Senar et al., 2007; **42.** Güell et al., 2009; **43.** Siqueira et al., 2016; **44.** McCutcheon and Moran 2011; **45.** Charles et al., 2011; **46.** Shigenobu et al., 2000; **47.** Hansen and Degnan, 2014; **48.** Thairu et al., 2018; **49.** Moran and Bennett, 2014; **50.** Toft and Andersson, 2010; **51.** Bennet and Chong, 2017; **52.** Schroeder et al., 2015; **53.** Narra et al., 2016; **54.** Schroeder et al.,2016; **55.** Schroeder et al., 2017; **56.** Mayoral **et** al., 2014; **57.** Woolfit et al., 2015, **58.** Touchon et al., 2009, **59.** Cook et al. 2013**.**

Chapter 2: A sRNA in a reduced mutualistic symbiont genome regulates its own gene expression

Abstract: Similar to other nutritional endosymbionts that are obligate for host survival, the mutualistic aphid endosymbiont, *Buchnera,* has a highly reduced genome with few regulatory elements. Until recently, it was thought that aphid hosts were primarily responsible for regulating their symbiotic relationship. However, we recently revealed that *Buchnera* displays differential protein regulation, but not mRNA expression. We also identified a number of conserved small RNAs (sRNAs) that are expressed among *Buchnera* taxa. In this study, we investigate if differential protein regulation in *Buchnera* is the result of post-transcriptional gene regulation via sRNAs. We characterize the sRNA profile of two *Buchnera* life-stages: *1.* When *Buchnera* is transitioning from an extracellular proliferating state in aphid embryos, *2.* When *Buchnera* is in an intracellular non-proliferating state in aphid bacteriocytes (specialized symbiont cells). Overall, we identified 90 differentially expressed sRNAs, 97% of which were up-regulated in aphid embryos. Of these sRNAs, the majority were predicted to be involved in the regulation of various metabolic processes, including arginine biosynthesis. Using a heterologous dual expression vector, we reveal for the first time that a *Buchnera* antisense sRNA can post-transcriptionally interact with its cognate *Buchnera* coding sequence, *carB*, a gene involved in arginine biosynthesis. These results corroborate our *in vivo* RNAseq and proteomic data, where the candidate antisense sRNA *carB* and the protein CarB are significantly up-regulated in aphid embryos. Overall, we demonstrate that *Buchnera* may regulate gene expression independently from its host by utilizing sRNAs.

Introduction:

Bacterial symbionts are widespread across animal lineages and have played an important role in animal evolution and diversification (McFall-Ngai *et al.*, 2013; Bennett & Moran, 2015). However, the factors that contribute to the maintenance of intracellular symbiont and host relationships remain unclear. Conventional wisdom has held that intracellular symbionts with reduced genomes are incapable of directly regulating these relationships. Instead their hosts are responsible for regulating cellular activities such as shared metabolic pathways in which both the symbiont and host contribute to the formation of key metabolites (*reviewed in* Hansen & Moran, 2014). Contrary to this notion there is emerging evidence that suggests that endosymbionts and eukaryotic organelles with reduced genomes can in fact utilize post-transcriptional methods of gene regulation (*reviewed in* Kim *et al.,* 2016).

The aphid-*Buchnera* endosymbiosis is a well-established model that is used to study nutritional symbioses. Within this aphid-*Buchnera* symbiosis, each partner depends on the other for the production of essential amino acids (Hansen & Moran, 2011; Poliakov *et al.*, 2011). The aphid host provides *Buchnera* with non-essential amino acids and other metabolites, which are then converted into essential amino acids by *Buchnera*'s biosynthetic pathways. During *Buchnera's* co-evolution with its aphid host it has experienced dramatic gene loss due to genetic drift; like many other intracellular symbionts (Moran, 1996; Wernegreen, 2002; *reviewed in* Bennett & Moran, 2015). For example, *Buchnera* has lost many genes that are associated with gene regulation, yet still retains genes that are associated with its nutritional role for its host (Shigenobu *et al.*, 2000).

Buchnera's gene regulation at the mRNA level is generally assumed to be negligible (*reviewed in* Hansen & Moran, 2014). This is in part because it has lost many transcription factors (Shigenobu *et al.*, 2000), and many of its operons have been fragmented when compared to its free-living relatives such as *Escherichia coli* (Moran & Mira, 2001). Moreover, microarray experiments

measuring *Buchnera'*s gene expression reveal that genes underlying essential amino acid biosynthesis are not differentially regulated at the mRNA level in response to nutritional demand (Moran *et al.*, 2003; 2005). To this end, the prevailing hypothesis is that the integrated metabolism shared between *Buchnera* and its aphid host is regulated primarily by the aphid through aphid-encoded transporters and genes that complement *Buchnera*'s essential amino acid pathways (Wilson *et al.*, 2010; Hansen & Moran, 2011; Poliakov *et al.*, 2011; Price *et al.*, 2014). Recently, *Buchnera* was found to exhibit differential expression of proteins without a concomitant change in mRNA expression between two distinct *Buchnera* life-stages: aphid maternal bacteriocytes and embryos (Hansen & Degnan, 2014). These results suggest that post-transcriptional regulation occurs, however the exact mechanism(s) that facilitates this form of regulation is unclear.

In recent years, small RNAs (sRNAs) have emerged as important post-transcriptional regulatory factors in Bacteria, Archaea, Eukaryotes, and eukaryotic organelles (Thomason & Storz, 2010; Babski *et al.*, 2014; *reviewed in* Kim *et al.,* 2016). In general, sRNAs are molecules that can affect translation directly or indirectly (Thomason & Storz, 2010). This includes post-transcriptional regulation via altering the targets of endonucleases and exonucleases (Thomason & Storz, 2010). Among four *Buchnera* taxa that began diverging from one another ~65 million years ago, 636 highly conserved sRNAs have been identified (Hansen & Degnan, 2014). Sixty-three percent of these sRNAs were associated with proteins that were differentially expressed between the two *Buchnera* life-stages (Hansen & Degnan, 2014).

In this study, we determine if a symbiont with a highly reduced genome, such as *Buchnera*, regulates its gene expression using sRNAs. To address this question, first we identify differentially expressed sRNAs between two different *Buchnera* life-stages, which previously showed differential protein expression but not mRNA expression (Hansen & Degnan, 2014). Second, we used a dual expression vector system in *E. coli* to heterologously express a *Buchnera* sRNA

candidate and its cognate *Buchnera* coding sequence (CDS) to determine if it regulates its predicted *Buchnera* target protein.

Methods:

Identification and categorization of Buchnera sRNA

sRNA sample preparation and sequencing

Three sub-lines of *Acyrthosiphon pisum* (LSR1) that were established over 100 generations ago from a single female were reared in a growth chamber at 20°C under a 16-h light/8-h dark regime and maintained on *Vicia faba* (fava bean). The *Buchnera* life-stage treatments consisted of: [1] aphid embryos, where *Buchnera* is in its extracellular proliferating state and [2] maternal bacteriocytes, where *Buchnera* is in an intracellular non-proliferating state (Koga *et al.,* 2012). Both these life-stages were co-collected from the same 4th instar aphid nymph individual via dissection. Roughly 200 aphids from each sub-line (n=3 sub-lines per life-stage) were dissected for each life-stage as in Hansen & Moran (2011) for maternal bacteriocytes and Hansen & Degnan (2014) for aphid embryos. All tissues were immediately stored in RNAprotect Bacteria Reagent (Qiagen, Germantown, MD).

For each replicate, total RNA was extracted using the miRNAeasy kit (Qiagen, Germantown, MD). The small RNA fraction, for each replicate were then size selected on a 6% PAGE gel at a size range of ≤ 200 nt by the University of Illinois at Urbana-Champaign Roy J. Carver Biotechnology Center. Library preparation was then performed on this RNA fraction (≤ 200 nt) using the Illumina mRNA directional sequencing protocol starting from the phosphatase treatment step by the UIUC- Roy J. Carver Biotechnology Center. Each library, which consisted of 100 nt single-end reads, was then sequenced on the Illumina (San Diego, CA, USA) x at the UIUC- Roy J. Carver Biotechnology Center.

sRNA identification

First, reads were trimmed and quality screened using Trimmomatic (Bolger *et al.,* 2014), and then aligned using Bowtie2 (Langmead *et al.*, 2009) to remove aphid reads. The resulting mapping file generated for each sample was then re-mapped using Rockhopper (McClure *et al.,* 2013) to identify putative sRNAs (*see,* Supporting Information for more details). To determine if sRNAs were differentially expressed between the different life-stages, a significance criterion of $\geq$ 2-fold change between samples and with $q < 0.05$ was established. The boundaries of the predicted sRNAs were then manually determined using Artemis (Rutherford *et al.,* 2000). The identified sRNAs were then cross-referenced to the conserved sRNAs identified previously by Hansen & Degnan (2014). We binned sRNAs into three different categories depending on where they were expressed relative to coding sequence(s). These three categories were: sRNAs expressed antisense to the gene, sRNAs expressed within the untranslated regions (UTR) of genes, and sRNAs identified within the intergenic spacer regions. We did not focus on antisense RNAs expressed in the same direction as a possible UTR or co-transcribed operon.

For sRNAs that were differentially expressed, the putative secondary structure was predicted with RNAalifold (Bernhart *et al*., 2008) using the same methods as in Hansen & Degnan (2014) (*see* Supporting Information). Differentially expressed sRNAs that had a single predicted gene target, because of their direct base-pairing interaction with the coding sequence (i.e. antisense sRNAs and UTR sRNAs) were functionally characterized by their predicted target gene's ontology (GO) classification using PANTHER v.12 (Mi *et al.,* 2016). We also used PANTHER to determine if any of the GO classifications were overrepresented. A Fisher's exact test was run to determine if the GO process classification results were significantly different between all the CDS in the *Buchnera* genome compared to only the differentially expressed sRNA predicted target CDS.

In vitro green fluorescent protein (GFP) reporter assays were conducted in _E. coli,_ the culturable, genetic model relative of _Buchnera._ These experiments allowed us to determine if sRNAs expressed from _Buchnera_ result in post-transcriptional regulation of its predicted _Buchnera_ target gene. These assays utilize a dual vector system in which we heterologously expressed the sRNA and its predicted target coding sequence (CDS) as in Bobrovskyy and Vanderpool (2016). The selection criteria for choosing the sRNA and CDS candidate for our _in vitro_ validation experiment were the following: [1] the sRNA expression pattern and genomic location had to be conserved in two or more _Buchnera_ lineages, [2] the free energy of thermodynamic ensemble, a measure of sRNA structural stability had to be significantly lower than a random control, as determined in Hansen & Degnan (2014), [3] the predicted target CDS has to be differentially expressed at the protein level between the two different _Buchnera_ life-stages in aphid embryos and maternal bacteriocytes (Hansen & Degnan, 2014), [4] the sRNA candidate must be assigned to a single predicted CDS that it directly base pairs with, and [5] the candidate sRNA must have a predicted target CDS that is directly involved in the symbiosis (i.e., the biosynthesis of essential amino acids).

Plasmid cloning

First, to clone the predicted CDS that the sRNA candidate is targeting, _Buchnera_ total DNA was extracted from _A. pisum_ (LSR1)'s whole body using the DNeasy blood and tissue extraction kit (Qiagen, Germantown, MD). Next, Hot Start PCR (KAPA Biosystems, Wilmington, MA) was used to amplify the putative _Buchnera_ sRNA and its corresponding predicted CDS target. Primers used to amplify both the putative _Buchnera_ sRNA and its corresponding predicted CDS target were

obtained from Integrated DNA Technologies (Table S 2.1). Each primer pair used contained restriction digestion cut sites at the 5' ends to facilitate directional cloning into each vector.

The plasmid, containing the target CDS, pZEMB1 (Bobrovskyy & Vanderpool, 2016), was constructed by PCR amplifying a ~200 base pair fragment of the *Buchnera* CDS that included that region that was predicted to directly base pair with the candidate sRNA. The PCR product containing the predicted target CDS and the vector pZEMB1 was digested with KpnI and EcoRI (New England Biolabs, Ipswich, MA) and ligated with DNA Ligase (New England Biolabs, Ipswich, MA). The predicted target CDS was directionally cloned downstream of the Isopropyl-β-D-1-thiogalactopyranoside (IPTG) inducible PL_{lac} promoter, and in-frame with the superfolder GFP (Fig. S 2.2A).

The plasmid containing the sRNA, pZAMB1 (Bobrovskyy & Vanderpool, 2016) was constructed by PCR amplifying the putative *Buchnera* sRNA. Putative *Buchnera* sRNAs were directionally cloned downstream of an anhydrotetracycline (aTc) inducible promoter (PL_{tet}). Both the PCR products and pZAMB1 were digested with NdeI and BamHI (New England Biolabs, Ipswich, MA) restriction endonucleases and ligated with DNA Ligase (New England Biolabs, Ipswich, MA) (Fig S 2.2b).

A complete list of primers and vectors used for this study are provided in Table S2.1. Complete maps of the plasmids and their sequences are included in the Supporting Information. All cloning experiments were conducted in *E. coli* strain DH-5α. Both the plasmid carrying the sRNA and the predicted target CDS were then co-transformed into *E. coli* (DB166), a modified strain of *E. coli* DJ480 for the GFP-reporter fluorescence assay.

Bacterial culturing conditions:

Bacteria were cultured in Luria–Bertani (LB) broth medium or on LB agar plates at 37°C, except for the GFP-reporter fluorescence assay where bacterial strains were cultured in MOPS (morpholine-propanesulfonic acid) rich defined medium (Teknova). The following antibiotics were used: ampicillin (Amp) (100 μg/ml) for the vector pZAMB1 and chloramphenicol (Cm), (25 μg/mL) for the vector pZEMB1. For the GFP-reporter fluorescence assay, strains were grown overnight in MOPS media. The overnight culture was then sub-cultured 1:100 to fresh medium with appropriate inducers in 96-well plates. Two independent rounds of assays were carried out for each experiment in biological and technical triplicate. IPTG was used at 1 mM to induce expression from the PL_{lac} promoter, and 75 ng/ml of anhydrotetracycline (aTc) was used to induce expression from the PL_{tet} promoter.

GFP-reporter fluorescence assay

For the GFP-reporter fluorescence assay experiment three treatments were conducted with a double plasmid transformant: *1.* empty sRNA carrying plasmid + predicted target CDS carrying plasmid (empty sRNA control), *2.* sRNA carrying plasmid + empty predicted target CDS carrying plasmid (empty target CDS control), *3.* sRNA carrying plasmid + predicted target carrying plasmid (experimental double transformant) (*see* Fig. S 2.1 for experimental schematic). GFP fluorescence, measured as relative fluorescence units (RFUs) and optical density at 600 nm (OD_{600}) were measured for 12hrs using the FLUOstar Omega microplate reader (BMG Labtech). *E. coli* natively expresses sRNAs and it is possible that these natively expressed sRNAs and/or other *E. coli* expressed gene products may interact with *Buchnera's* candidate sRNA and CDS target. If such an interaction were occurring between a natively expressed *E. coli* product and the expressed

Buchnera sRNA/predicted gene target this interaction would be observed in one or both of the control experimental treatments.

<u>*Statistics*</u>

To determine if there was a difference in normalized RFUs among the treatments, an ANOVA with repeated measures was performed with PROC MIXED (SAS 9.4) with a Bonferroni adjustment for least square means comparison (α <0.05). All statistical analyses were conducted on data collected after the first hour of growth i.e. until all colonies reached stationary phase, the period in which *E. coli* enters a quiescent period (OD_{600} ~1.0). To determine if there were differences in RFUs during the early and mid-log growth (OD_{600} 0.1 and 0.4) phase a one-sided Welches-Test (α < 0.05) was used.

Results:

<u>Buchnera *sRNAs are differentially expressed between two different life-stage treatments:*</u>

To determine if *Buchnera* sRNAs have a functional role in gene regulation, we first characterized the sRNA profiles of two distinct *Buchnera* life-stages, the aphid embryo and the maternal bacteriocytes, using directional RNA-seq. We mapped over 96% of the high quality filtered reads (332,987,365 reads) to both the *Buchnera* and aphid genomes (Table S2.2). Using Rockhopper, we identified 90 differentially expressed sRNAs, including tRNAs, ($q \leq 0.05$, fold change ≥ 2) between the two life-stages. Three sRNAs were up-regulated in the maternal bacteriocytes and 87 were up-regulated in the aphid embryos (Table 2.1; Fig. S2.3).

Buchnera retains 32 tRNA genes in its genome. Twenty-five of the 90 differentially expressed sRNAs were either tRNAs or sRNAs expressed antisense to the tRNA gene and were up-regulated in the aphid embryo (Table S2.3). Forty percent of these sRNAs were expressed in

the antisense direction of the tRNA gene. Over half (6/10) of the tRNAs expressed in antisense were previously identified and display conserved antisense expression across various *Buchnera* lineages (Hansen & Moran, 2012).

The 65 remaining differentially expressed sRNAs comprise sRNAs expressed within the intergenic spacers, UTRs, and antisense to coding sequences. Thirty-two of these differentially expressed sRNAs were newly identified when compared to previous data (Hansen & Degnan 2014) (Table 2.2).

Antisense sRNAs represented the majority (54/ 65) of the differentially expressed sRNAs (Table S2.4). As in Hansen & Degnan (2014), over half (53%) of the antisense sRNAs demonstrated significant predicted thermodynamic stability when compared to the random control (one-tailed t-test, d.f.99, P<0.05, Table S2.4). Of the differentially expressed antisense sRNAs, 50% of the sRNAs identified in this study were conserved within one or more divergent *Buchnera* lineages. We also identified 27 antisense sRNAs that were not previously identified by Hansen & Degnan (2014), which may be specific to the *Buchnera A. pisum*-LSR1 lineage (Table 2.2).

Only three of the differentially expressed sRNAs were identified to be sRNAs expressed within the UTR. Specifically, two were expressed in the 3' UTR of *sirA* and *rnpB,* and only *sirA*'s 3' UTR sRNA was significantly thermodynamically stable when compared to the random control (one-tailed t-test, d.f.99, P<0.05, Table S2.5). The third UTR sRNA was differentially expressed in the 5' UTR of *amiB*. This sRNA was also significantly thermodynamically stable compared to a random control (one-tailed t-test, d.f. 99, P<0.05, Table S2.5). The sRNAs identified as thermodynamically stable (*sirA* and *amiB*) were conserved in expression across *Buchnera* lineages, whereas the sRNA that was associated with *rnpB* is not (Table 2.2).

Eight of the differentially expressed sRNAs were identified within the intergenic spacer regions. All but one of these sRNAs were predicted to be significantly thermodynamically stable

when compared to the random control (7/8; one-tailed t-test, d.f.99, P<0.05, Table S2.6). Of these sRNAs, 63% were conserved across *Buchnera* lineages (Table 2.2).

Functional categories of predicted target CDS

To determine if differentially expressed sRNAs were associated with differential protein expression patterns between the two different *Buchnera* life-stages, we compared the protein expression profiles characterized by Hansen & Degnan (2014) to the predicted target CDS of the differentially expressed sRNAs. For antisense sRNAs the predicted target CDS was determined by identifying the CDS on the complementary strand. Eight out of 54 predicted proteins that were associated with the up-regulated antisense sRNAs were also up-regulated in either aphid embryos or maternal bacteriocytes (Table 2.3). To determine the predicted target gene for UTR sRNAs we identified the genes that these UTRs were associated with. None of the target proteins that were associated with the differentially expressed UTR sRNAs were differentially expressed. It is difficult to predict the target CDS of intergenic sRNAs since they can influence adjacent genes or act *in trans* affecting one or more genes. Given the uncertainty in assigning targets of intergenic sRNAs we did not assign them to a predicted target CDS for this analysis

Using GO functional gene analysis we determined that 53% of the predicted CDS targets that were associated with differentially expressed sRNA's (antisense sRNAs and UTR sRNAs) were categorized within the metabolic processes GO category (Table S2.9). Several metabolic GO processes identified for predicted target CDS within this study overlapped similar GO processes identified previously (Hansen & Degnan 2014) for differentially expressed proteins. Specifically, the peptidoglycan, *de novo* pyrimidine, *de novo* purine and arginine biosynthesis pathway (Table S2.10).

sRNAs may not only affect the expression of the gene they directly bind to but also the expression of additional genes within their operon. Consequently, we broadened our analysis to the operon level when examining the potential proteins that are regulated by differentially expressed sRNAs. First, we found that no operons were identified with significant up-regulation or down-regulation of both protein and UTR sRNA expression or intergenic sRNA expression (Table 2.3). In contrast, we identified six operons that shared significant up-regulation or down-regulation of both protein and antisense sRNA expression, suggesting that these antisense sRNAs may facilitate differential protein expression within these operons (Table 2.3). Five of these operons were associated with the up-regulation of both antisense sRNAs and proteins in the embryo, suggesting that these antisense sRNAs may be involved in the activation/stabilization of these proteins (Table 2.3). The sixth operon containing the genes *ilvI* and *ilvH,* had the protein IlvH down-regulated and the antisense sRNA *ilvI* up-regulated in the embryo; suggesting that the antisense sRNA *ilvI* may be involved in the repression of *ilvH* (Table 2.3). Four of the seven differentially expressed antisense sRNAs that were associated with operons with differential protein expression were significantly thermodynamically stable when compared to the random control (one-tailed t-test, d.f. 99, P<0.001): *carB, ftsL, murC* and *nuoG* (Table S2.4).

Heterologous functional validation of a candidate Buchnera *sRNA:*

Based on our selection criteria for our life-stage data (*see methods*) we selected the antisense sRNA *carB* and its predicted target CDS *carB* for our *in vitro* expression assays. The putative protein target CarB is involved in *Buchnera*'s essential amino acid biosynthesis pathway for arginine, and therefore this protein is important for the aphid-*Buchnera* mutualism. In this study, we found that the antisense sRNA *carB* is significantly up-regulated in embryos compared to maternal bacteriocytes. Moreover, the target protein CarB was previously determined to be up-

regulated in the aphid embryos compared to maternal bacterioyctes (Hansen & Degnan, 2014) (Fig. 2.1A). In addition, the antisense sRNA *carB* is conserved in two divergent *Buchnera* taxa (*Buchnera-A. pisum* and *Buchnera- S. graminum)* and parsimony would suggest that this specific antisense sRNA has been conserved for over 65 MY in *Buchnera* (Hansen & Degnan, 2014). In this study, we found that the antisense sRNA *carB* was significantly more thermodynamically stable when compared to the random control (Fig. 2.1B; one-tailed t-test, d.f.99, P<0.001).

For the GFP-reporter fluorescence assay experiment we expected to observe a change in normalized RFUs if the RNA-target CDS interaction (both activating and repressing) was regulatory in function. If the sRNA was activating and/or stabilizing gene expression, we expected to see an increase in normalized RFUs compared to the empty sRNA control. In contrast, if the sRNA was repressing gene expression we expected to see a decrease in normalized RFUs compared to the empty sRNA control. The empty target control was expected to display negligible, background fluorescence.

Overall, we observed a significantly higher expression of RFUs in the experimental treatment (*carB* antisense sRNA+ *carB* CDS) than the empty antisense sRNA control (*carB* CDS only) (df=374, T=-3.23, Bonferroni adjusted P ≤ 0.01) (Table 2.4), suggesting that the *carB* antisense sRNA facilitates the activation or stabilization of *carB*. As expected, the empty target control (*carB* antisense sRNA only) had a low level of expression of RFUs compared to the other two treatments (*carB* antisense sRNA+ *carB* CDS vs *carB* CDS only df=374, T=8.87, Bonferroni adjusted P ≤ 0.001; *carB* antisense sRNA vs *carB* antisense sRNA+ *carB* CDS df=374, T=12.1, Bonferroni adjusted P ≤ 0.001; Table 2.4).

The RFU level of the *carB* antisense sRNA + *carB* CDS treatment was significantly different compared to the empty target control (*carB* antisense sRNA only) during the early- mid exponential growth phase of *E. coli* (OD_{600} 0.1 and 0.4), the most metabolically active periods of

E. coli growth (early exponential growth phase OD_{600} 0.1: df = 6.73, t = 1.90, P= 0.05: mid exponential growth phase OD_{600} 0.4: df = 6.73, t = 1.90, P= 0.05; Fig. 2.2; Table S2.8). These results further suggest that the *Buchnera carB* antisense sRNA is activating and/or stabilizing *Buchnera*'s CarB protein expression.

Discussion:

Until recently it has been unclear how and if mutualistic symbionts with highly reduced genomes, such as *Buchnera,* regulate their own gene expression. In this study, we reveal for the first time that an obligate mutualistic endosymbiont with a reduced-genome has the potential to utilize a sRNA to regulate protein expression, similar to pathogens/parasites that also possess reduced-genomes, such as *Wolbachia* (Mayoral *et al.,* 2014) and *Mycoplasma pneumoniae* (Güell *et al.*, 2009). Specifically, our *in vitro* assays demonstrate that the *Buchnera* antisense sRNA *carB* may activate/stabilize the *Buchnera* coding sequence (*carB*). Moreover, our *in vivo* assays corroborate these results because the antisense sRNA *carB* is differentially expressed in the same direction as its predicted target protein CarB between two different *Buchnera* life-stages. Our *in vivo* assays also reveal other potential sRNA candidates that may be involved in protein regulation, because they are also differentially expressed between these life-stages, possess stable secondary structures, and are associated with putative proteins targets that are differentially regulated as well.

The protein CarB is involved in the arginine biosynthesis pathway (Fig. 2.1). In *E. coli*, a free-living relative of *Buchnera*, all arginine biosynthetic enzymes are repressed by its end-product arginine, via the repressor ArgR (Caldara *et al.,* 2008). *Buchnera* has lost the repressor gene *argR* through endosymbiont evolution, and therefore the regulation of this pathway has been modified from its free-living relatives. The genes *carA* and *carB* form two sub-units that biosynthesize the metabolite Carbamyl phosphate (CP). Within the arginine biosynthesis pathway CP is a key

44

intermediate metabolite, and its biosynthesis is rate-limiting (Caldara *et al.,* 2008), thus making *carA* and/or *carB* an ideal target of regulation. sRNAs can act in various ways that would lead to increased protein expression, including preventing the formation of inhibitory secondary structures, or increasing mRNA stability (Storz *et al.*, 2011). In our study, we observed an increase of GFP expression when the antisense sRNA *carB* was co-expressed with its predicted gene target. Since mRNA levels are not significantly different between *Buchnera* life-stages (Hansen & Degnan, 2014), this observation suggests that the antisense sRNA *carB* may increase gene regulation at the post-transcriptional level, leading to an increased rate of CP production when this sRNA is up-regulated. This hypothesis is further supported by the observed up-regulation of the two proteins CarA and CarB in the same life-stage in which we observed the up-regulation of the antisense sRNA *carB* (Hansen & Degnan, 2014). Since CarA and CarB are within the same operon, the antisense sRNA *carB* may play an important regulatory role in increasing both CarA and CarB expression in the embryonic stage of aphids.

Another key intermediate metabolite in this pathway is ornithine, which is synthesized by *argE*. Interestingly, we identified a sRNA that was differentially regulated antisense to argE however a previous study did not find the protein ArgE differentially regulated between the same two *Buchnera* life-stages (Fig.2.1) (Hansen & Degnan 2014). It has been hypothesized previously that the aphid host may play a role in biosynthesizing and thus regulating ornithine for *Buchnera*, and in fact many *Buchnera* taxa have lost the genes responsible for ornithine biosynthesis (Hansen & Moran 2014). Currently it is unclear if the antisense sRNA *argE* has a regulatory function based on our data for this life-stage comparison.

During aphid development, aphid embryos have lower levels of free arginine compared to later developmental stages, suggesting that these life-stages may use differential regulation to respond to nutrient demand (Rabatel *et al.* 2013). Arginine has been found to play an important

role in various insect developmental pathways, including the target of rapamycin (TOR) signaling pathway (Zheng *et al.,* 2016). The TOR signaling pathway is important in regulating cell growth in response to nutrients (Loewith & Hall, 2011). Increases in dietary arginine has been found to increase signaling of the TOR pathway, leading to increased cell growth (Zheng *et al.,* 2016). *Slimfast* is an amino acid transporter that is found within the TOR signaling pathway, which can sense and transport arginine (Colombani *et al.,* 2003; Boudko *et al.,* 2015). Within developing bacteriocytes of aphid embryos, one of the most highly expressed amino-acid transporters, *APC-8904*, is orthologous to *Drosophila*'s *slimfast* (Lu *et al.,* 2016). A second pathway that requires arginine that is especially important during insect development is the nitric oxide signaling (NOS) pathway. Arginine serves as the original substrate needed for nitric oxide production (Davis 2000). NOS has been found to be important in regulating gene expression during insect development (Yamanaka and O'Connor 2011). NOS also has a key role in the development of neuronal pathways (Davis 2000). It is therefore possible that during aphid embryo development, *Buchnera* produces more arginine to activate pathways that are vital for rapidly growing aphid tissues. Our experimental data provides tangential evidence for this hypothesis. It is also interesting to note that the majority of sRNAs (this study) and proteins (Hansen & Degnan, 2014) were up-regulated in aphid embryos compared to maternal bacteriocytes. Potentially if some of these sRNA candidates are regulatory the enrichment of *Buchnera* proteins may be more important in the embryonic stage of insect development where host control is not yet possible, because bacteriocytes have not fully formed yet around *Buchnera* cells (Skidmore & Hansen, 2017).

Regulatory antisense sRNAs have also been found in the culturable human pathogen, *M. pneumoniae*, which has a reduced genome size of ~816 Kb, (Himmelreich, 1996). Approximately 13% of the coding genes in *M. pneumoniae* have a corresponding antisense sRNA (Güell *et al.*, 2009). For at least two of these *M. pneumoniae* antisense sRNA candidates there is evidence that

they can regulate genes that are involved in metabolism, DNA repair, and DNA replication (Güell *et al*., 2009). Within *Buchnera,* antisense sRNAs are conserved across various *Buchnera* taxa, similarly, several of the identified *M. pneumoniae* antisense sRNAs are conserved in closely related taxa, such as *Mycoplasma genitalium* (Güell *et al*., 2009). Previous studies predict that most antisense sRNAs are simply transcriptional noise, especially in bacteria that are AT rich such as *Buchnera* (Raghavan *et al.*, 2012; Llorens-Rico *et al.*, 2016). However, the data here and in the *M. genitalium* studies suggest that if purifying selection is strong enough in the face of genetic drift antisense sRNAs may be conserved across divergent taxa, especially if they are essential for the symbiont/pathogen's gene regulation.

Overall, we found twenty-eight antisense sRNAs, one UTR sRNA and three intergenic sRNAs that are specific to the *Buchnera- A. pisum* (LSR1) line compared to other *Buchnera* taxa in this study (Table 2.2). The *Buchnera* strains LSR1 and 5A, both originate from the same aphid species *A. pisum,* though most likely diverged between ~7-20 MYA (Degnan *et al*., 2011). Identification of sRNAs that are uniquely expressed within *Buchnera –A. pisum* (LSR1)*,* and those that have conserved expression across taxa, indicate that the maintenance/recruitment of functional sRNAs can dynamically occur across evolutionary time scales. We hypothesize that the sRNAs that are unique to specific *Buchnera* lineages may be important in the symbiont's adaption to its specific aphid host lineage; whereas, sRNAs conserved across taxa may be important for the general maintenance of the nutritional symbiosis, or general bacterial function. Currently it is unclear how these sRNAs are regulated. The regulation of several sRNAs has been elucidated in only a few free-living bacterial taxa (Wagner & Romby, 2015), and much is still unknown for taxa with small genomes. Future empirical studies are required in *Buchnera* to identify putative promoter sites that are not easily identified with computational approaches, because of *Buchnera*'s

AT rich genome. By further elucidating the mechanism of sRNA regulation within organisms with small genomes, we will gain a deeper understanding of their function and evolution

sRNAs may evolve in *Buchnera*'s genome through adaptive processes. However, genome evolution in *Buchnera* is generally hypothesized to be a non-adaptive process due to genetic drift reducing the efficacy of selection (Moran, 1996; Wernegreen, 2002; Bennett & Moran, 2015). Alternatively, *Buchnera* may use compensatory mechanisms to maintain critical regulatory functions for the symbiosis in the face of genome erosion (Moran, 1996). For example, regarding regulatory RNAs, *Buchnera* was hypothesized to compensate for the truncation of the 3' CCA sequence on its tRNAs, which is required for amino acid activation, by co-opting the CCA-adding enzyme (Hansen & Moran, 2012). Overall, our data also supports the non-adaptive, compensation hypothesis. In summary, we hypothesize that *Buchnera* maintains the expression of critical sRNAs through purifying selection to compensate for the loss of canonical regulatory proteins, thus reverting to the "RNA world" of regulation (Kim *et al.* 2016).

References:

Babski, J. Maier, L-K., Heyer, R., Jaschinski, K., Prasse, D., Jäger, D., … Soppa, J. (2014). Small regulatory RNAs in Archaea. *RNA biology*, *11*, 484–93. doi: 10.4161/rna.28452

Bennett, G.M. & Moran, N.A. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 201421388. doi: 10.1073/pnas.1421388112

Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R. & Stadler, P.F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474. doi: 10.1186/1471-2105-9-474

Bobrovskyy, M. & Vanderpool, C.K. (2016). Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Molecular Microbiology, 99,* 254–273. doi:10.1111/mmi.13230

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

Boudko, D.Y., Tsujimoto, H., Rodriguez, S.D., Meleshkevitch, E.A., Price, D.P., Drake, L.L. & Hansen, I. A. (2015). Substrate specificity and transport mechanism of amino-acid transceptor Slimfast from Aedes aegypti. *Nature communications*, 6, 8546. doi: 10.1038/ncomms9546

Caldara, M., Dupont, G., Leroy, F., Goldbeter, A., De Vuyst, L. & Cunin, R. (2008). Arginine biosynthesis in Escherichia coli: experimental perturbation and mathematical modeling. The Journal of biological chemistry, 283, 6347–58. doi:10.1074/jbc.M705884200

Colombani, J., Raisin, S., Pantalacci, S., Radimerski, T., Montagne, J. & Léopold, P. (2003). A nutrient sensor mechanism controls Drosophila growth. *Cell*, 114, 739–749. doi: 10.1016/S0092-8674(03)00713-X

Davies, S. (2000). Nitric oxide signaling in insects. *Insect biochemistry and molecular biology,* 30, 1123–38. doi: 10.1016/S0965-1748(00)00118-1

Degnan, P.H., Ochman, H. & Moran, N.A. (2011). Sequence Conservation and Functional Constraint on Intergenic Spacers in Reduced Genomes of the Obligate Symbiont Buchnera. *PLoS Genetics*, 7, e1002252. doi: 10.1371/journal.pgen.1002252

Güell, M., van Noort, V., Yus E., Cheng, W-H., Leigh-Bell, J., Michalodimitrakis, K., … Serrano, L. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science (New York, N.Y.)*, 326, 1268–71. doi:10.1126/science.1176951

Hansen, A. K. & Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2849–54. doi: 10.1073/pnas.1013465108

Hansen, A.K. & Degnan, P.H. (2014). Widespread expression of conserved small RNAs in small symbiont genomes. *The ISME journal,* 8, 2490–502. doi: 10.1038/ismej.2014.121

Hansen, A.K. & Moran, N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular ecology*, 23, 1473–96. doi: 10.1111/mec.12421

Hansen, A.K. & Moran, N.A. (2012). Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes. *Nucleic acids research*, 40, 7870–84. doi: 10.1093/nar/gks503

Himmelreich, R., Hilbert, H., Plagens H., Pirkl, E., Li, B-C.C. & Herrmann, R. (1996). Complete Sequence Analysis of the Genome of the Bacterium Mycoplasma Pneumoniae. *Nucleic Acids Research*, 24, 4420–49. doi: 10.1093/nar/24.22.4420

Kim, D., Thairu, M.W. & Hansen, A.K. (2016). Novel Insights into Insect-Microbe Interactions- - Role of Epigenomics and Small RNAs. *Frontiers in Plant Science*, 7, 1164. doi: 10.3389/fpls.2016.01164

Koga, R., Meng, X.Y., Tsuchida, T. & Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proceedings of the National Academy of Sciences of the United States of America,* 109, E1230-7. doi: 10.1073/pnas.1119212109

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10, R25. doi: 10.1186/gb-2009-10-3-r25

Llorens-Rico, V., Cano, J., Kamminga T., Gil, R., Latorre, A., Chen, W.-H., … Lluch-Senar, M. (2016). Bacterial antisense RNAs are mainly the product of transcriptional noise. *Science Advances,* 2, e1501363–e1501363. doi: 10.1126/sciadv.1501363

Lluch-Senar, M., Vallmitjana, M., Querol, E., Piñol, J. (2007). A new promoterless reporter vector reveals antisense transcription in *Mycoplasma genitalium*. *Microbiology (Reading, England)*, 153, 2743–52. doi: 10.1099/mic.0.2006/007559-0

Loewith, R., & Hall, M.N. (2011). Target of rapamycin (TOR) in nutrient signaling and growth control. Genetics, 189, 1177–201. doi: 10.1534/genetics.111.133363

Lu, H-L., Chang, C-C. & Wilson A.C.C. (2016). Amino acid transporters implicated in endocytosis of *Buchnera* during symbiont transmission in the pea aphid. *EvoDevo*, 7, 24. doi: 10.1186/s13227-016-0061-7

Mayoral, J.G., Hussain, M., Joubert, D.A., Iturbe-Ormaetxe, I., O'Neill, S.L. & Asgari, S. (2014). Wolbachia small noncoding RNAs and their role in cross-kingdom communications. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 18721–6. doi: 10.1073/pnas.1420131112

McClure, R., Balasubramanian, D., Sunm Y., Bobrovskyy, M., Sumby, P., Genco, C. A., … Tjaden, B. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research*, 41, e140. doi: 10.1093/nar/gkt444

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A., …, Wernegreen, J.J. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 3229–36. doi: 10.1073/pnas.1218525110

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. (2015). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic acids research*, 44, D336-42. doi: 10.1093/nar/gkv1194

Moran, N.A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 2873–8. doi: 10.1073/pnas.93.7.2873

Moran, N.A. & Mira, A. (2001). The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. *Genome biology*, 2, RESEARCH0054. doi: 10.1186/gb-2001-2-12-research0054

Moran, N.A., Dunbar, H.E. & Wilcox, J.L. (2005). Regulation of transcription in a reduced bacterial genome: Nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *Journal of Bacteriology*, 187, 4229–4237. doi: 10.1128/JB.187.12.4229-4237.2005

Moran, N.A., Plague, G.R., Sandström, J.P. & Wilcox, J.L. (2003). A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proceedings of the National Academy of Sciences of the United States of America*, 100 Suppl, 14543–8. doi: 10.1073/pnas.2135345100

Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H. J., Sun, Q., Douglas, A.E. & van Wijk, K.J. (2011). Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Molecular & Cellular Proteomics: MCP*, 10, M110.007039. doi: 10.1074/mcp.M110.007039

Price, D.R.G., Feng, H., Baker, J.D., Bavan, S., Luetje, C.W. & Wilson, A.C.C. (2014). Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 320–5. doi: 10.1073/pnas.1306068111

Rabatel, A., Febvay, G., Gaget, K., Duport, G., Baa-Puyoulet, P., Sapountzis, P., … Vincens, P. (2013). Tyrosine pathway regulation is host-mediated in the pea aphid symbiosis during late embryonic and early larval development. *BMC Genomics*, 14, 235. doi: 10.1186/1471-2164-14-235

Raghavan, R., Sloan, D.B. & Ochman, H. (2012). Antisense transcription is pervasive but rarely conserved in enteric bacteria. *mBio*, 3. doi: 10.1128/mBio.00156-12

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)*, 16, 944–5. doi: 10.1093/bioinformatics/16.10.944

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, 81–6. doi: 10.1038/35024074

Skidmore, I.H. & Hansen, A.K. (2017). The evolutionary development of plant-feeding insects and their nutritional endosymbionts. *Insect Science*, 1–19. doi: 10.1111/1744-7917.12463

Storz, G., Vogel, J. & Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell*, 43, 880–91. doi: 10.1016/j.molcel.2011.08.022

Thomason, M.K. & Storz, G. (2010). Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *. *Annual Review of Genetics*, 44, 167–188. doi: 10.1146/annurev-genet-102209-163523

Wagner, E.G.H. & Romby, P. (2015) Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It; Editor(s): Friedmann T., Dunlap, J.C., & Goodwin, S.F. in *Advances in Genetics*, Academic Press, 90:133-208. doi.org/10.1016/bs.adgen.2015.05.001.

Wernegreen, J.J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 3229–36. doi: 10.1073/pnas.1218525110

Wernegreen, J.J. (2002). Genome evolution in bacterial endosymbionts of insects. Nature reviews. *Genetics,* 3, 850–61. doi: 10.1038/nrg931

Wilson, A.C.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G., … Douglas, A.E. (2010). Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect molecular biology*, 19 Suppl 2, 249–58. doi: 10.1111/j.1365-2583.2009.00942.x

Yamanaka, N. & O'Connor, M.B. (2011). Nitric oxide directly regulates gene expression during Drosophila development: need some gas to drive into metamorphosis? *Genes & development*, 25, 1459–63. doi: 10.1101/gad.2080411

Zheng, L., Zhang, W., Zhou, Y., Li, F., Wei, H. & Peng, J. (2016). Recent Advances in Understanding Amino Acid Sensing Mechanisms that Regulate mTORC1. *International Journal of Molecular Sciences*, 17, 1636. doi: 10.3390/ijms17101636

Tables and Figures:

Table 2.1:

Summary of *Buchnera* sRNAs (excluding tRNAs) that were differentially expressed significantly (*q*<0.05, fold change >2) between two different *Buchnera* life-stages (embryos and maternal bacteriocytes).

Differentially Expressed sRNAs	
Up-regulated *Buchnera* sRNAs from Maternal Bacteriocytes	
5' UTRs sRNAs	1
3' UTR sRNAs	0
antisense sRNAs	2
intergenic sRNAs	0
Up-regulated *Buchnera* sRNAs from Aphid Embryos	
5' UTRs sRNAs	0
3' UTR sRNAs	2
antisense sRNAs	52
intergenic sRNAs	8

Table 2.2:

Conservation of *Buchnera* - *A. pisum* LSR1 sRNAs (excluding tRNAs) that were differentially expressed significantly ($q<0.05$, fold change >2) between two different *Buchnera* life-stages (embryos and maternal bacteriocytes).

Type of sRNA	Total sRNAs differentially expressed in *Buchnera*- *A. pisum*-LSR1[†]	Total sRNAs differentially expressed in *Buchnera* (*A. pisum*-LSR1) [†] that were not previously identified[††]	Number of conserved sRNAs identified previously[††] that overlap with the sRNAs that were differentially expressed in *Buchnera*- *A. pisum*-LSR1[†]				
			Buchnera (*A. pisum*-LSR1)	*Buchnera* (*A. pisum*-5A)	*Buchnera* (*A. kondoi*)[†]	*Buchnera* (*U. ambrosiae*)	*Buchnera* (*S. graminum*)
antisense sRNA	54	27	26	25	12	20	15
UTR sRNA	3	1	2	2	1	0	1
Intergenic sRNA	8	4	3	2	1	4	1

[†]Differentially expressed sRNA data from this study.
[††]sRNA data from Hansen and Degnan (2014)

Table 2.3: Operons that have both *Buchnera* antisense sRNA and protein up-regulation in aphid embryos or maternal bacteriocytes.

Operon[†]	Up-regulated protein[‡]/predicted CDS	Gene Description	Life-stage that the antisense sRNA/protein is up-regulated in			
			Aphid embryo		Maternal bacteriocyte	
			Protein[†]	antisense sRNA[§]	Protein[‡]	antisense sRNA[§]
rpoBC	*rpoC*	DNA-directed RNA polymerase subunit beta'	X	X		
carAB	*carA*	carbamoyl-phosphate synthase small subunit	X			
	carB	carbamoyl-phosphate synthase large chain	X	X		
nuoGHIJKL	*nuoG*	NADH dehydrogenase gamma subunit		X		
	nuoI	NADH dehydrogenase subunit I	X			
pth, ychF	*ychF*	GTP-binding protein	X			X
rsmH, ftsLIW, murEFCDG	*murC*	UDP-N-acetylmuramate-alanine ligase		X		
mraY, ddlB	*murG*	UDP-N-acetylglucosamine-N-acetylmuramyl transferase	X	X		
	ftsL	cell division protein	X	X		
ilvIH	*ilvH*	acetolactate synthase small subunit		X	X	
	ilvI	acetolactate synthase large subunit		X		
flgH	*flgH*	flagellar L-ring protein precursor	X	X		
leuS, holA, nadD	*leuS*	leucyl-tRNA synthetase	X	X		
def, fmt	*def*	polypeptide deformylase		X	X	
rpsR, rplI, cysQ, dut	*rpsR*	30S ribosomal protein S18	X			
	cysQ	CysQ protein		X		
mtlAD, mutL, miaA	*mtlA*	PTS system mannitol-specific IIABC component	X			
	mtlD	mannitol-1-phosphate 5-dehydrogenase		X		

[†] Operons were determined by Hansen & Degnan (2014) [‡] Proteomic data derived from Hansen & Degnan (2014)
[§] Antisense sRNAs identified in this study

Table 2.4:

Difference of Least Square Means from the repeated measures ANOVA for *carB* antisense sRNA

validation experiments. Standard Error (SE), Degrees of freedom (DF).

Comparison	SE	DF	T-value	Bonferroni Adjusted p-value
carB CDS VS *carB* CDS + *carB* antisense sRNA	3172.63	374	-3.23	<0.001
carB CDS VS *carB* antisense sRNA	3172.63	374	8.87	<0.001
carB CDS + *carB* antisense sRNA VS *carB* antisense sRNA	3172.63	374	12.1	<0.001
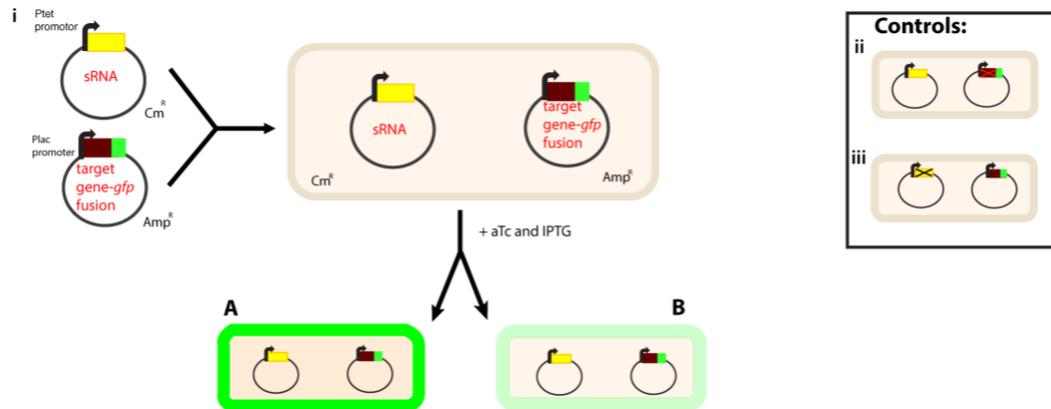
Figure 2.1:

GFP- report assay experimental overview. In vitro assays were conducted in the culturable relative of *Buchnera, Escherichia coli* to determine whether a sRNA expressed from *Buchnera* results in post-transcriptional regulation of its predicted *Buchnera* target gene. Once the antisense sRNA and predicted target coding sequence were cloned into their respective plasmids, the following treatments were used: (i)antisense sRNA carrying plasmid+predicted target carrying plasmid (experimental double transformant), (ii) antisense sRNA carrying plasmid+empty predicted target CDS carrying plasmid (empty target CDS control) and (iii) empty antisense sRNA carrying plasmid+predicted target CDS carrying plasmid (empty antisense sRNA control). The predicted target CDS was directionally cloned downstream of the Isopropyl-b-D-1-thiogalactopyranoside (IPTG)-inducible PLlacpromoter, and in-frame with the superfolder GFP and theBuchneraantisense sRNAs wasdirectionally cloned downstream of an anhydrotetracycline (aTc)-inducible promoter (PLtet). Strains were then grown overnight in MOPS media. The overnight culture was then subcultured 1:100 to fresh medium with appropriate inducers in 96-well plates. Two independent rounds ofassays were carried out for each experiment in biological and technical triplicate. IPTG was used at 1 mMto induce expression from the PLlacpromoter, and 75 ng/ml of anhydrotetracycline (aTc) was used to induce expression from the PLtetpromoter. If the

putative antisense sRNAhas a regulatory role, we expect sRNA–target CDS interactions (both activating and repressing) to alter the magnitude of GFP fluorescence. Ifthe antisense sRNA is activating/stabilizing gene expression, we expect to see an increase in normalized relative fluorescence units (RFUs)compared to the empty antisense sRNA control (a); if the antisense sRNA is repressing gene expression, we expect to see a decrease innormalized RFUs compared to the empty antisense sRNA control (b). The empty target control is expected to display negligible, backgroundfluorescence. RFUs and optical density (OD) 600 were measured for 12 hr
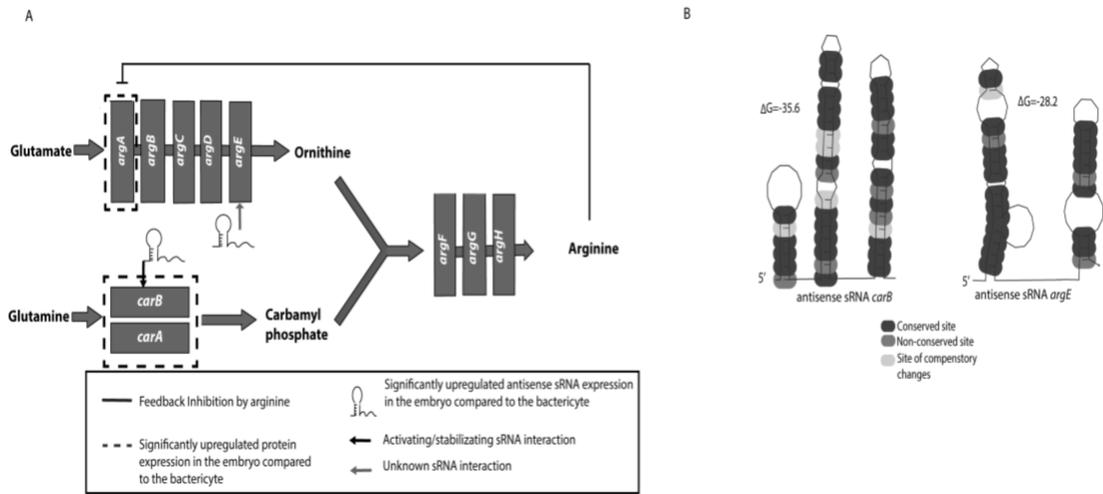
Figure 2.2:

Our model of the regulation of *Buchnera's* arginine biosynthesis pathway, highlighting significantly upregulated sRNAs from RNAseq and heterologous expression assays (this study) and significantly upregulated proteins (from Hansen and Degnan (2014)) between two different *Buchnera* life stages (ovary and bacteriocytes). Within *Buchnera*, the following genes are monocistronic: *argA, argE* and *argF*. The remaining genes are found within two operons: *argBCGH* and *carAB*.(b) Predicted thermodynamic stability and sequence covariation in sRNA secondary structures of antisense sRNAs *carB* and *argE* using RNAalifold. Both structures were significantly stable based on random control sequences (p<.05). These sRNAs were upregulated significantly in embryos compared to bacteriocytes and are associated with the arginine biosynthesis pathway. See Supporting Information for predicted structures using R-scape.
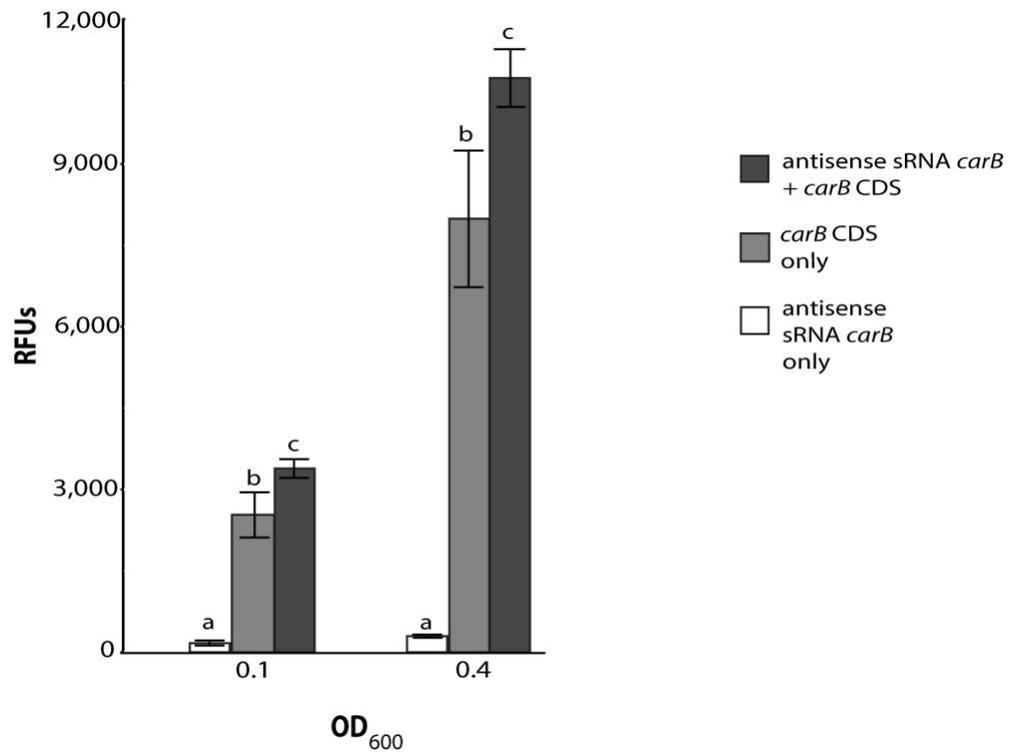
Figure 2.3:

Results from the sRNA validation experiments using the GFP reporter fluorescence assay. RFUs at two specific timepoints where colonies were at OD (600) 0.1 and 0.4, early and mid-log growth, two metabolically active periods of *E. coli* growth. Error bars indicate ±SEfrom the mean. Bars with different letters aboverepresent significant differences among treatments at p<.05

Chapter 3: Changes in aphid host-plant diet influences the small RNA expression patterns of its obligate nutritional symbiont, *Buchnera*

Abstract:

Plants are a difficult food resource to utilize and herbivorous insects have therefore evolved compensatory mechanisms that allow them to fully exploit this poor nutritional resource. One such mechanism is the maintenance of bacterial symbionts that aid in host plant feeding and development. Many obligate nutritional symbionts of plant feeding insects occur within the insect order Hemiptera. The majority of these intracellular symbionts have highly eroded genomes that lack many key regulatory genes. Consequently, it is unclear if these symbionts can respond to changes in the insect's diet to facilitate host-plant use. There is emerging evidence that symbionts with highly eroded genomes express small RNAs (sRNAs), some of which potentially regulate gene expression. In this study we determine if the reduced genome of the nutritional symbiont (*Buchnera*) in the pea aphid (*Acyrthosiphon pisum*) responds to changes in the aphid's host plant diet. Using RNA-seq, *Buchnera* sRNA expression profiles were characterized within two *Buchnera* life-stages (aphid ovarioles and maternal bacteriocytes) when pea aphids fed on either alfalfa or fava bean. Overall, this study demonstrates that *Buchnera* sRNA expression changes not only with life-stage, but also with changes in aphid host-plant diet. Of the 321 sRNAs characterized in this study, 47% were previously identified and 22% of the 321 sRNAs showed evidence of conservation in two or more *Buchnera* taxa. Functionally, 11 differentially expressed sRNAs were predicted to target genes related to pathways involved in essential amino acid biosynthesis. Overall, results from this study reveal that host plant diet influences the expression of conserved and lineage specific sRNA in *Buchnera,* and these sRNAs display distinct host-plant specific expression profiles among biological replicates. Future studies are needed to further determine if these sRNA candidates are regulatory and involved in host-plant interactions.

Introduction:

Herbivorous insects are faced with the challenge of using a food resource that contains noxious defensive compounds and varies in nutritional profiles both spatially and temporally (Schoonhoven et al., 2005). As such, insect herbivores have evolved a diversity of mechanisms that facilitate host plant use (Simpson & Simpson, 1990; Schoonhoven et al., 2005). One mechanism that has aided various phytophagous insects in using plants as nutrient resources is the acquisition of microbial symbionts (Buchner, 1965 Janson et al., 2008; Hansen and Moran, 2014; Sugio et al., 2015). Some of these bacterial symbionts aid in the detoxification of plant chemicals, such as gut microbes found in the large pine weevil, *Hylobius abietis,* and the coffee berry borer beetle, *Hypothenemus hampeaides* (Ceja-Navarro et al., 2015; Berasategui et al., 2017). Other beneficial bacteria help insects by providing nutrients that are deplete or missing in the plant diet. Many examples of such bacterial symbioses are widely found throughout the insect order Hemiptera (Sudakaran et al., 2017). In the hemipteran sub-order Heteroptera, many of these bacteria are extracellular, and are housed in specialized structures in the gut (Glasgow, 1914; Goodchild, 1963; Buchner, 1965; Kikuchi, 2009; Sudakaran et al., 2017). This is in contrast to the sub-orders Auchenorrhyncha and Sternorrhyncha that are primarily populated with obligate, nutritional symbionts that are housed intracellularly in specialized insect cells (bacteriocytes) within the insect's body (Buchner, 1965; Sudakaran et al., 2017). Obligate intracellular symbionts have highly eroded genomes that lack many key regulatory genes in contrast to the majority of extracellular symbionts (Moran and Bennett, 2014). Consequently, it is unclear if intracellular symbionts can respond to changes in insect diet to facilitate host plant use; especially in comparison to extracellular symbionts that often directly interface with the plant material and have larger, more dynamic genomes.

Within Hemiptera, the partnership between *Acyrthosiphon pisum* (pea aphid) and the bacteria *Buchnera* is one of the best characterized nutritional, intracellular symbioses within insects.

In this symbiosis, each partner depends on the other for the production of amino acids (Shigenobu & Wilson, 2011; Hansen & Moran, 2011; Poliakov et al., 2011). The symbiont *Buchnera* encodes the majority of genes within pathways for essential amino acid biosynthesis (Shigenobu et al., 2000), which are not encoded *de novo* by the aphid (Wilson et al., 2010). The aphid then encodes genes that are predicted to be vital in the regulation of this microbial symbiosis. These aphid genes include those that are involved in amino acid transport, ammonia recycling, synthesis of amino donors and metabolite intermediates, as well as the terminal steps for essential amino acid biosynthesis that are missing from the *Buchnera* genome (Nakabachi et al., 2005; Wilson et al., 2010; Hansen and Moran 2011; Poliakov et al., 2011; Price et al., 2014). Homologs of these aphid genes are also up-regulated in bacteriocytes from other Hemipteran taxa, suggesting that host control of these intracellular symbioses is widespread among Hemipterans (*see* Husnik et al., 2013; Sloan et al., 2014; Luan et al., 2015; Moa et al., 2018). These insect genes potentially can respond to host plant diet, for example, Kim et al., (2018) observed the up-regulation of several of these collaborating aphid genes in bacteriocytes when aphids fed on their sub-optimal, specialized host-plant, alfalfa, compared to their universal host-plant, fava bean.

Currently, there is limited evidence that supports the role of intracellular symbionts responding to changes in insect diet to facilitate host plant use. For example, *Buchnera* displays negligible gene expression responses at the mRNA level when aphids feed on an artificial diet or plant material that was spiked with different concentrations of amino acids (Moran et al., 2005; Reymond et al., 2006). However, Viñuelas and colleagues (2011) demonstrated that *Buchnera's* pLeu plasmid can respond to changes in the aphid's diet by increasing plasmid copy number and up-regulating the genes on the plasmid (*leuABCD, repA1, repA2, and yqhA*) in response to limited concentrations of leucine in an artificial diet. These results suggest that in the face of genome reduction *Buchnera* displays limited transcriptional control on its chromosome in response to diet

variation, however other regulatory strategies may still be maintained via plasmids. Currently, it is unknown if post-transcriptional regulation is important in *Buchnera* in response to host plant diet.

Post-transcriptional strategies of gene regulation, such as small RNAs (sRNAs), have been widely observed throughout all domains of life (Kim et al. 2009; Babski et al., 2014; Brant & Budak, 2018; Hör et al., 2018; Ozata et al., 2019). Emerging evidence supports the role of regulatory small RNAs (sRNAs) in intracellular bacterial symbiont gene regulation (Thairu et al. 2019). For example, sRNAs expressed from the tick symbionts, *Rickettsia prowazekii* and *Rickettsia conorii,* are hypothesized to be important in facilitating host-niche adaptation (Narra et al., 2016; Schroeder et al., 2017). In *Buchnera,* sRNAs are hypothesized to be important in regulating genes at the post-transcriptional level when *Buchnera* transitions between different life-stages (Hansen & Degnan, 2014; Thairu et al. 2018). In turn, there is potential for *Buchnera* sRNAs to aid in regulating essential amino acid and vitamin biosynthesis pathways in response to nutrient demand. Nutrient demand in aphids may occur when aphids feed on host plants that vary in free amino acid content as well as defensive compounds, which may inhibit nutrient uptake by the aphid (Sandström & Pettersson, 1994; Sandström & Moran, 1999; Sanchez-Arcos et al. 2016; 2019; Yuan et al., 2019).

In this study, we determine if sRNAs expressed by the intracellular symbiont, *Buchnera,* respond to aphid host-plant diet. Specifically, we determine if *Buchnera* sRNAs are differentially regulated when *A. pisum* feeds on fava bean (*Vicia faba*) compared to alfalfa (*Medicago sativa*). These host-plants were chosen because they vary in amino acid profiles and host-plant defenses (Sandström & Pettersson, 1994; Sandström & Moran, 1999; Sanchez-Arcos et al. 2016; 2019; Yuan et al., 2019). Moreover, the *A. pisum* (LSR1) sub-lines used in this study display higher fitness when they feed on their "universal" host-plant fava bean (FB) compared to their specialized host-plant alfalfa (ALF), and several aphid genes involved in the nutritional symbioses are differentially expressed between host-plant diets (Kim et al., 2018). In this study we use RNA-seq to characterize

Buchnera sRNA expression between the aphid's universal (FB) and specialized (ALF) host-plant diets for two different *Buchnera* life-stages: [1] aphid ovarioles (referred to thereafter as embryos (EMB)), where early in aphid development *Buchnera* is in its extracellular proliferating state and [2] maternal bacteriocytes (BAC), where *Buchnera* is in an intracellular state (Koga et al., 2012).These two *Buchnera* life-stages were collected separately because it was previously shown that *Buchnera* displays differential sRNA (Thairu et al., 2018) and protein expression profiles between these life-stages (Hansen & Degnan, 2014).

Methods:

Small RNA sample preparation and sequencing

Three sub-lines of *A. pisum* (LSR1) that were established in Kim et al. (2018), were allowed to independently develop and feed for >100 generations on either *V. faba* (fava bean, FB) or *M. sativa* (alfalfa, ALF) (N= 3 biological replicates per host plant species treatment). These six sub-lines were reared in a growth chamber at 20°C under a 16-h light/8-h dark regime.

For each sub-line treatment, two life-stage samples were co-collected from the same 4th instar aphid nymph individual via dissection similar to Thairu et al. (2018). These two samples represent two different *Buchnera* life-stages: [1] aphid ovarioles (EMB), and [2] maternal bacteriocytes (BAC). Approximately, 200 aphids from each sub-line were dissected for each life-stage and pooled. All tissues were immediately stored in RNAprotect Bacteria Reagent (Qiagen, Germantown, MD) and stored at -80C.

For each sample, RNA was extracted using the miRNAeasy kit (Qiagen, Germantown, MD). Library preparation and sequencing was then performed on the small RNA enriched fraction (≤ 250 nt) using the Illumina mRNA directional sequencing protocol by the University of California, San Diego, Institute for Genomic Medicine Genomics Center (UCSD IGM Genomics Center). Each

library was then sequenced as 75 nt single-end reads on the Illumina Hi-seq 4000 (San Diego, CA, USA) at the UCSD IGM Genomics Center. A total of 12 samples were sequenced: three bacteriocyte biological replicate samples from aphids feeding on alfalfa (ALF-BAC), three bacteriocyte biological replicate samples from aphids feeding on fava bean (FB-BAC), three embryo biological replicate samples from aphids feeding on alfalfa (ALF-EMB), and three embryo biological replicate samples from aphids feeding on fava bean (FB-EMB).

Identification and categorization of *Buchnera* sRNAs

First, reads were quality screened using Trimmomatic v.0.33 (Bolger *et al.,* 2014). Adapters were then removed using Cutadapt v2.1 (Martin, 2011). To remove aphid reads, sequences were aligned to the aphid genome using Bowtie2 v.2.2.9 (Langmead *et al.*, 2009). Once aphid reads were removed, Bowtie2 was used to map the remaining reads to the *Buchnera* genome. Rockhopper v.2.0.3 (McClure *et al.,* 2013) was then used to identify putative *Buchnera* sRNAs. The same Rockhopper parameters optimized for identifying *Buchnera* sRNAs in Hansen & Degnan (2014), were used to identify sRNAs in this study. Reads were normalized by the upper-quartile method in Rockhopper. sRNA boundaries were manually determined by inspecting directional coverage curves in Artemis v.16 (Rutherford et al., 2000). sRNAs were then binned into three different categories similar to Hansen &Degnan (2014), and Thairu et al. (2018): sRNAs expressed antisense to the gene (antisense sRNAs), sRNAs expressed within the untranslated regions of genes (UTR sRNAs), and sRNAs identified within the intergenic spacer regions (intergenic sRNAs). As in Hansen &Degnan (2014) and Thairu et al. (2018), antisense and UTR sRNAs are named after their predicted target coding sequence (CDS) based on direct base-pairing interactions. For example, the antisense sRNA *aroC,* which is expressed antisense to the *aroC* coding sequence, is predicted to target the CDS *aroC.* If multiple sRNAs are predicted to target different regions of the same CDS then a number

will follow (e.g. antisense sRNAs *ilvI_1,* and *ilvI_2)*. For sRNAs expressed within the intergenic spacer region the sRNA name contains both the up-stream and down-stream CDS names (e.g. intergenic sRNA *argH-yibN*).

Using the read counts for each of the identified sRNAs from Rockhopper, a Principle Component Analysis (PCA) was conducted in R v.3.5.2 (R Core Team, 2018) using the package DEbrowser v.1.10.6 (Kucukural et al., 2019) to compare how similar *Buchnera* sRNA expression profiles are across all treatments. The reads were filtered using the default DEbrowser settings and were normalized by the upper-quartile method, the same method used by Rockhopper(McClure *et al.,* 2013). The following four groups were compared: [1] ALF-BAC, [2] ALF-EMB, [3] FB-BAC, and [4] FB-EMB. To determine if sRNA profiles were significantly different between treatments, multi-response permutation procedure (MRPP) was used using the vegan v.2.5-3 package in R (Oksanen et al., 2019).

Rockhopper was also used to determine if sRNAs were differentially expressed between host-plant treatments and life-stage categories. sRNA expression comparisons include: [1] bacteriocytes from aphids feeding on alfalfa (ALF-BAC) compared to bacteriocytes from aphids feeding on fava (FB-BAC), [2] embryos from aphids feeding on alfalfa (ALF-EMB) compared to embryos from aphids feeding on fava (FB-EMB), [3] bacteriocytes compared to embryos from aphids feeding on alfalfa (ALF-BAC and ALF-EMB, respectively), and [4] bacteriocytes compared to embryos from aphids feeding on fava (FB-BAC and FB-EMB, respectively). A significance criterion of ≥ 1.5-fold change between samples and a *q* < 0.05 was used to determine if a sRNA was differentially expressed. The putative secondary structures were predicted for all differentially expressed sRNAs using RNAalifold v.2.1 (Hofacker et al., 2002; Bernhart *et al*., 2008), following the same methods as in Hansen & Degnan (2014).

Using PANTHER GO functional gene list analysis (Mi et al. 2019), the GO pathways and GO biological processes were identified for the predicted CDS targets of sRNAs that displayed both significant differential expression and predicted thermodynamic stability. Only putative *cis*-acting sRNAs (e.g. antisense and UTR sRNAs, which have hypothetical direct base-pairing interactions with their CDS target) were included in this analysis, because it is uncertain whether putative intergenic sRNAs target one or both neighboring genes or act in *trans* and target other genes located distantly in the genome.

Results:

Buchnera *sRNA expression profiles are influenced by aphid host-plant diet and Buchnera life-stage:*

For all RNA-seq samples [1] ALF-BAC, [2] ALF-EMB, [3] FB-BAC, and [4] FB-EMB, which contain three biological replicates each (N= 12 RNAseq samples total), ~ 95% of all high-quality reads (6.89×10^8) mapped to either the *Buchnera* or aphid genome (Supplemental Table 3.1). Within the ALF and FB host-plant treatments, 1.60×10^8 reads and 1.87×10^8 reads mapped to the *Buchnera* genome respectively.

A total of 321 unique sRNAs were identified among all 12 samples and consisted of 253 antisense sRNAs, 17 UTR sRNAs, and 52 intergenic sRNAs. Many of these sRNAs were identified previously where 51% of antisense sRNAs, 82% of UTR sRNAs, and 12% of intergenic sRNAs are conserved in two or more *Buchnera* lineages (Hansen & Degnan 2014; Supplementary Tables 3.2-3.7,3.11-3.16).

After identifying expressed sRNAs using Rockhopper, a PCA and MRPP analysis was run to determine how similar the *Buchnera* sRNA expression profiles were across treatments. Using the MRPP analysis, the sRNA expression profiles of the *a prior* groups: [1] ALF-BAC, [2] ALF-EMB, [3] FB-BAC, and [4] FB-EMB were found to be significantly different from each other ($p \leq 0.01$;

A=0.4418). When grouped by *Buchnera* life-stage, the sRNA expression profiles were significantly different from each other (p ≤ 0.05; A= 0.167) with similar amounts of dispersion within groups (BAC delta=0.021, EMB delta=0.029). When grouped by host-plant treatment, the sRNA profiles were also significantly different (p ≤ 0.05; A= 0.2162), however there was more within group dispersion among the ALF samples (delta=0.02655) compared to the FB samples (delta=0.01979). This segregation of sRNA expression profiles by treatment was supported by the PCA ordination results (Figure 3.1). The first three PCA axes explained 73% of the variation in the dataset, with the first PCA axis (PC1) accounting for 35% (±13.21 standard deviation (SD) of the variation), the second axis (PC2) accounting for 22% (± 10.43 SD), and the third axis (PC3) accounting for 16% (± 8.90SD). Combined, these results suggest that both aphid host-plant and *Buchnera* life-stages have a significant effect on sRNA expression profiles.

Aphid host-plant diet significantly affects Buchnera *sRNA expression for genes involved in essential amino acid biosynthesis:*

To determine how sRNA expression profiles differ between host-plant treatments for each life-stage, Rockhopper identified 131 differentially expressed sRNAs between the bacteriocyte (BAC) samples (ALF-BAC and FB-BAC) and 176 sRNAs between the embryo (EMB) samples (ALF-EMB and FB-EMB), including tRNAs (q≤0.05; 1.5-fold change; Table 1). Within the ALF-BAC *vs.* FB-BAC comparison, 103 sRNAs were up-regulated in the ALF-BAC treatment group and 28 sRNAs were up regulated in FB-BAC treatment group. Within the ALF-EMB *vs.* FB-EMB comparison, 99 sRNAs were up-regulated in the ALF-EMB treatment group and 77 sRNAs were up-regulated in the FB-EMB treatment group (q≤0.05; 1.5-fold change; Table 3.1).

For the functional gene analysis, in the BAC comparison (ALF-BAC *vs.* FB-BAC) a total of 11 PANTHER GO pathways associated with 11 putative CDS targets of differentially expressed

sRNAs were identified (Table 3.2). For the EMB comparison (ALF-EMB *vs.* FB-EMB), 11 CDSs were predicted to be associated with the same 11 PANTHER GO pathways found in the BAC comparisons (Table 3.2). There was notable overlap between the differentially expressed sRNAs identified in the EMB comparison with those identified in the BAC comparison, however, sRNAs unique to both comparisons were also found. Specifically, sRNAs unique to the ALF-EMB *vs.* FB-EMB comparison comprised of the antisense sRNAs *ilvI_1,* and *thrB_2,3* (Table 3.2). The sRNAs unique to the ALF-BAC *vs.* FB-BAC comparison were, antisense sRNAs, *aroA_1, ilvI_2,* and *thrC_2.*

In the ALF-BAC *vs.* FB-BAC comparison, seven of the 11 sRNAs associated with the PANTHER GO pathways identified were up-regulated in FB-BAC (Table 3.2). The antisense sRNAs, *aroA_1, ilvC_2,* and *murD_1,2* were up-regulated in ALF-BAC. For the differentially expressed sRNAs associated with PANTHER GO pathways in the ALF-BAC *vs.* FB-BAC comparison, only antisense sRNA *ilvD* was found to be conserved among two or more *Buchnera* strains (Hansen & Degnan, 2014; Supplemental Table 3.2). The remaining sRNAs are unique to the *A. pisum Buchnera* strains, having been found in both *Buchnera A. pisum*, strains (5A and LSR1) (Hansen & Degnan, 2014; Supplemental Table 3.2), except for the antisense sRNAs *aroC* and *prsA* which are unique to this study. Eight of the 11 sRNAs predicted to target CDSs in the ALF-EMB *vs.* FB-EMB comparison, were up-regulated in the FB-EMB samples, with the antisense sRNAs *ilvC_2* and *murD_1,3* being up-regulated in the ALF-EMB samples. The antisense sRNAs *ilvI_2,* and *thrB_2,3* which were found in the ALF-EMB *vs.* FB-EMB comparison were identified for the first time in this study. Of these 11 PANTHER GO pathways predicted to be associated with differentially expressed sRNAs found in both the ALF-BAC *vs.* FB-BAC and ALF-EMB *vs.* FB-EMB, five are associated with essential amino acid biosynthesis, specifically, arginine, chorismate, isoleucine, threonine, and valine (Table 3.2). The antisense *thrC_3* sRNA, which was up-regulated

in FB-BAC is predicted to target the CDS in both the threonine biosynthesis pathway and the vitamin B6 metabolism pathway.

Based on the PANTHER GO biological processes analysis (Supplementary Table 3.8) the antisense sRNA *prsA,* which was up-regulated in the ALF-BAC samples and ALF-EMB samples, was identified for the first time in this study, is part of the 5-phosphoribose 1-diphosphate biosynthetic process (GO:0006015); a component of the pentose phosphate pathway that produces phosphoribosyl pyrophosphate (PRPP) a necessary precursor to the histidine biosynthesis pathway (Figure 3.2). The antisense sRNA *dapD_2,* (which was identified for the first time in this study, Supplementary Table 3.3) was found to be up-regulated only in the ALF-EMB samples, and is predicted to target the CDS *dapD,* which synthesizes diaminopimelate, a precursor for lysine biosynthesis.

Within both the ALF-BAC *vs.* FB-BAC and ALF-EMB *vs.* FB-EMB comparisons, antisense tRNAs were found to be differentially expressed (Table 3.1; Supplemental Table 3.9, 3.10). Antisense tRNA expression has been previously observed within *Buchnera* (Hansen & Moran, 2011; Thairu et al. 2018) and within this current study, ≥60% of the differentially expressed tRNAs were expressed in the antisense direction (Table 3.1; Supplemental Table 3.9, 3.10). All antisense tRNAs were up-regulated in the FB host-plant treatments for both ALF-BAC *vs.* FB-BAC and ALF-EMB *vs.* FB-EMB comparisons and of these antisense sRNAs, four (antisense tRNAs Asn, Glu, His and Met) are conserved in multiple *Buchnera* strains (Hansen & Moran, 2011, Supplemental Table 3.9-3.10).

Buchnera sRNAs are differentially expressed between life-stages when aphids feed on either host-plant:

To determine if _Buchnera_ sRNA expression is different between life-stages we analyzed the following sRNAs expression profiles for each host plant treatment: _i_. ALF-BAC compared to ALF-EMB and _ii_. FB-BAC compared to FB-EMB. For the ALF host-plant treatment, one sRNA was up-regulated in the ALF-BAC life-stage and the remainder (29) were up-regulated in the ALF-EMB life-stage (q≤0.05; 1.5-fold change; Table 3.3). This pattern of more sRNAs being up-regulated in the EMB compared to the BAC life-stage was also observed in the FB host-plant treatment, where 222 sRNAs were up-regulated in the FB-EMB life-stage and 15 were up-regulated in the FB-BAC life-stage (q≤0.05; 1.5-fold change; Table 3.3). Notably, there was a high level of overlap for the sRNAs that were differentially expressed between life-stages for both host-plant comparisons. Of the 30 sRNAs differentially expressed in the ALF-BAC _vs._ ALF-EMB comparison, 83% were also differentially expressed, in the same direction for the FB-BAC _vs._ FB-EMB comparison (Supplemental Tables 3.11-3.16, 3.18,3.19).

Using PANTHER GO functional gene list analysis, For the ALF-BAC _vs._ ALF-EMB comparison, of the differentially expressed antisense or UTR sRNAs only the antisense sRNAs _ilvI_1_ and _ilvI_5_ were predicted to target CDSs within any PANTHER GO pathways, specifically the isoleucine, and valine biosynthesis pathways. However, in the FB-BAC _vs._ FB-EMB comparison, 15 antisense sRNAs which were all up-regulated in the FB-EMB samples were predicted to target CDSs within 16 PANTHER GO pathways (Table 3.4) Of these 15 antisense sRNAs, three (_carB-1, ilvD_ and _pta_) are conserved across two or more _Buchnera_ strains (Hansen & Degnan, 2014; Supplemental Table 3.12). The remaining 12 antisense sRNAs are specific to the _A. pisum Buchnera_ strains, with antisense sRNAs _aroA_1, ilvI_1,5,_ and _murD_3,_ being conserved in both the _Buchnera A. pisum_ strains 5A and LSR1 (Hansen & Degnan, 2014; Supplemental Table

3.12). The following antisense sRNAs were predicted to target CDSs within the same five essential amino acids biosynthesis pathways which were identified in the host-plant comparisons (arginine, chorismate, isoleucine, threonine, and valine): *aroA_1, aroC, carB_1, ilvD and ilvI_1,5* (Table 3.4). The antisense sRNA *serC_2,* was predicted to target three GO pathways including the vitamin B_6 pathway. The remaining associated GO pathways include acetate utilization, *de novo* purine biosynthesis, *de novo* pyrimidine ribonucleotides biosynthesis, *de novo* pyrimidine deoxyribonucleotide biosynthesis, peptidoglycan biosynthesis, pyridoxal-5-phosphate biosynthesis, Salvage pyrimidine deoxyribonucleotides, Salvage pyrimidine deoxyribonucleotides and serine glycine biosynthesis. Based on the PANTHER complete GO biological processes analysis (Supplemental Table 3.17) the antisense sRNAs *prsA* and *dapD_2* were predicted to target the CDSs important in the histidine and lysine biosynthesis processes (respectively) were also up-regulated in the EMB life-stage (Supplemental Table 3.17).

Antisense tRNAs made up the majority of the differentially expressed tRNAs identified in both life-stage comparisons. Specifically, within the ALF-BAC *vs.* ALF-EMB comparison, three differentially expressed antisense tRNAs were found, and all three were antisense and up-regulated in the ALF-EMB samples (Supplemental Table 3.18). Of these three antisense tRNAs, the antisense Glu tRNA which was also identified in the ALF-BAC *vs.* FB-BAC and ALF-EMB *vs.* FB-EMB comparisons, is conserved among multiple *Buchnera* strains (Hansen & Moran, 2011). In the FB-BAC *vs.* FB-EMB comparison, 14 antisense sRNAs were differentially expressed and up-regulated in the FB-EMB samples (Supplemental Table 3.19). Four of the antisense tRNAs (antisense Asn, His, Glu and Met tRNAs) identified in the FB-BAC *vs.* FB-EMB comparison, are conserved among *Buchnera* strains (Hansen & Moran, 2011, Supplemental Table 3.19). These four conserved antisense tRNAs were also identified in the ALF-BAC *vs.* FB-BAC and ALF-EMB *vs.* FB-EMB comparisons.

Discussion:

This study demonstrates for the first time that *Buchnera* sRNA expression changes in response to aphid host-plant diet. Many of the differentially expressed sRNAs expressed between host-plant diet treatments are conserved and share similar profiles across life-stage treatments for essential amino acid biosynthesis pathways. It is unclear at this time if *Buchnera* sRNA expression observed in this study is associated with post-transcriptional gene regulation of the predicted gene targets, however accumulating evidence from previous studies show that some *Buchnera* sRNAs are conserved among strains and respond in a consistent and predictable manner depending on aphid host-plant diet and life-stage (Hansen & Degnan, 2014, Thairu et al. 2017).

Buchnera sRNA expression is more heterogenous among biological replicates in both life-stages when *A. pisum* feeds on its specialized host plant alfalfa (ALF) compared to its universal host-plant fava bean (FB) (Figure 3.1). It is uncertain if this heterogenous response is driven by one of the aphid sub-lines (Figure 3.1) or is a diverse response to its specialized host-plant, alfalfa (ALF). Nevertheless, despite this diffuse response in sRNA expression profiles there was a significant difference in sRNA expression observed between host plant treatments for both life-stages. In a previous study, *A. pisum* aphids displayed lower mass on alfalfa (ALF), their specialized host-plant, compared to fava bean (FB), which is suggestive that they are obtaining lower amounts of nutrients and/or are expending more energy (Kim et al., 2018). These two host plants are also known to vary significantly in essential amino acid profiles (Sandström & Pettersson, 1994) and metabolite profiles, including host plant defense compounds (Sanchez-Arcos et al., 2016, 2019). When comparing the metabolite profiles among the *A. pisum* host-plants alfalfa (ALF), fava bean (FB), pea, and red-clover, Sanchez-Arcos et al., (2019) found that alfalfa (ALF) contains the highest number of unique metabolites, especially when compared to fava bean (FB). Sanchez-Arcos et al., (2019) also hypothesized that the relatively simple metabolic profile of fava bean (FB), which had

the fewest unique metabolites compared to the other host-plants included in the study, may contribute to fava bean (FB) acting as *A. pisum's* universal host-plant. Aphid stylectomy experiments have also shown that fava bean (FB) and alfalfa (ALF), have different concentrations of arginine, isoleucine, leucine, lysine, phenylalanine, tryptophan, tyrosine, valine, and histidine; with alfalfa (ALF) generally having lower concentrations of essential amino acids compared to fava bean (FB) (Sandström and Pettersson 1994).

We hypothesize that the potential variation of metabolites (which include amino acids) in alfalfa (ALF), may contribute to the heterogeneity observed in *Buchnera* sRNA expression profiles when aphids feed on alfalfa (ALF). For example, regardless of *Buchnera* life-stage, sRNAs predicted to target genes within the arginine, chorismate (phenylalanine, tryptophan, tyrosine), isoleucine, threonine, and valine biosynthesis pathways are differentially expressed between aphid host-plant treatments (Table 3.2, Figure 3.2, Supplemental Figure 3.1). Alternatively, *Buchnera* sRNAs may be responding sporadically, in a non-directed manner if *Buchnera* is stressed when the aphids fed on a metabolically complex diet such as alfalfa (Kim et al. 2018). Of the 321 differentially expressed sRNAs identified in this study, 18% are predicted to target genes that are associated with stress responses in *Escherichia coli* (Jozefczuk et al., 2010). Future work is needed to determine if the observed aphid host-plant specific response of *Buchnera* sRNAs results in the differential expression of predicted *Buchnera* protein targets, which potentially can impact this symbiosis in either an adaptive nutritional manner or a non-adaptive, random stress response.

In the life-stage comparisons of ALF-BAC *vs.* ALF-EMB and FB-BAC *vs.* FB-EMB, there was notable overlap in the differentially expressed antisense sRNAs identified between host plant treatments (Supplemental Table 3.11-3.12). The differentially expressed antisense sRNAs found in both life stage comparisons may potentially be crucial in the regulation of important life-stage specific CDSs regardless of host-plant diet. Of these 22 overlapping antisense sRNAs, six sRNAs

(antisense sRNAs *ansA, alaS, leuS, ilvI_1,5* and *truB*) had significantly thermodynamically stable secondary structure predictions. Other than the antisense sRNAs *ilvI_1* and *ilvI_5,* which were predicted to target CDS within the isoleucine and valine biosynthesis PANTHER GO pathways, the remaining four antisense sRNAs were predicted to target genes in eight PANTHER GO biological processes which include mRNA pseudouridine and pseudouridine processes (Supplemental Table 3.17). *Buchnera* is unable to produce its own uracil, a precursor for pseudouridine synthesis and therefore it depends on the host for uracil biosynthesis (Shigenobu et al. 2000; Figure 3.2). Pseudouridine is a widespread and a functionally important post-transcriptional modification of RNAs within eukaryotes (Karijolich et al., 2015). Within bacteria, pseudouridine modification is not considered as important or widespread in cell function relative to eukaryotes (Marbaniang & Vogel, 2016; O'Connor et al. 2018) nevertheless, within *E. coli*, pseudouridine modification of mRNAs can influence translation, reducing protein expression (Hoernes et al., 2016) and the presence of pseudouridine in a stop codon can result in readthrough occurring (Fernández et al. 2016). tRNA pseudouridine synthase TruB, the predicted target CDS of the antisense sRNA *truB* identified in this study*,* also acts as a tRNA chaperone in *E. coli*; helping re-fold misfolded tRNAs (Keffer-Wilkes et al. 2016). It is therefore possible, that the chaperone activity of *truB,* and/or pseudouridine modification is important in *Buchnera* to facilitate different translation rates between life-stages*.*

Insects, like many other animals cannot synthesize B vitamins *de novo,* and so they obtain these necessary micro-nutrients from their diet or their association with various microorganisms (Douglas, 2017). Within aphids, *Buchnera* encodes several genes for the biotin (B_7), folic acid (B_9), pantothenate (B_5), pyridoxine (B_6) riboflavin (B_2), and thiamine (B_1) biosynthesis pathways (Shigenobu et al., 2000; Serbus et al., 2017). Vitamin B_6, is an important bacterial and eukaryotic co-factor especially in the amino acid metabolism (John 1995; Daub and Ehrenshaft 2000); and neither *Buchnera* nor *A. pisum,* encodes all the genes needed for its biosynthesis. Previous work

looking at aphid gene expression has shown that within the bacteriocytes of *A. pisum* that have fed

on alfalfa (ALF)*,* the vitamin B$_6$ pathway was significantly enriched (Kim et al., 2018). In this study

it was found that in the ALF-BAC *vs.* FB-BAC comparison, the antisense sRNA predicted to target

thrC, one of the two *Buchnera* genes vital for vitamin B$_6$ biosynthesis, was up-regulated in the ALF-

BAC samples (Supplemental Table 3.2). In addition, a second antisense sRNA which was up-

regulated in the FB-EMB samples compared FB-BAC samples was predicted to target *serC,* the

other *Buchnera* encoded gene vital for vitamin B$_6$ biosynthesis. Further studies that directly

manipulate vitamin B nutrition, in conjunction with proteomics will help illuminate whether or not

the antisense sRNAs *thrC* and *serC* differently regulate their predicted CDSs.

One of the most important questions in the field of bacterial sRNA research is determining

if a particular putative sRNA has a functional role. Working in non-model, unculturable systems

such as *Buchnera,* increases this challenge because many of the techniques used to validate and

determine functionality are not easily implemented or possible (Thairu & Hansen, 2019).

Nevertheless, the results from this study are consistent with previous *Buchnera* sRNA experiments

which provide some support for the potential functional role identified sRNAs. Specifically, Thairu

et al., (2018), looked at *Buchnera* sRNA expression in bacteriocytes and embryos of aphids that fed

on fava bean (FB). They identified 90 differentially expressed sRNAs, 27 of which were also

identified in the FB-BAC *vs.* FB-EMB comparison done in this study (Thairu et al., 2018;

Supplemental Tables 3.11-3.16, 3.18-3.19). Both studies also identified sRNAs predicted to target

CDSs within similar PANTHER GO Pathways. Specifically, the acetate utilization, arginine

biosynthesis, de novo purine biosynthesis, de novo pyrimidine ribonucleotide biosynthesis,

isoleucine biosynthesis, peptidoglycan biosynthesis, and valine biosynthesis pathways. Though both

studies used the same aphid-lines, methodological differences, such as differences in sequencing

platforms, library preparation, and centers could account for some the differences observed. In

Thairu et al. (2018), the authors also provided evidence for the functionality of the antisense sRNA *carB* (classified as antisense sRNA *carB_1,* in this study), by heterologously expressing the sRNA in *E. coli*. This study not only detected this sRNA, but also found that it had the same expression pattern that was described by Thairu et al. (2018) between different *Buchnera* life-stages. There was also an overlap between this study and the proteomic study done by Hansen & Degnan (2014), who characterized the differentially expressed *Buchnera* proteins between *Buchnera* life-stages when aphids feed on fava bean. Specifically, 27 of the 54 differentially expressed proteins identified by Hansen & Degnan (2104), were either directly or indirectly (*i.e.* within an operon) associated with a differentially expressed sRNA identified in this study (Supplemental Table 3.20). This study also found 71 (53 antisense sRNAs, 15 UTR sRNAs and 4 intergenic sRNAs) differentially expressed conserved sRNAs which were also characterized by Hansen & Degnan (2014) (Supplemental Table 3.2-3.7,3.11-3.16). The presence of conserved sRNAs across four *Buchnera* strains (*Buchnera A. pisum, Acyrthosiphon kondoi, Uroleucon ambrosiae,* and *Schizaphis graminum*) that have diverged over 65 million years, strongly suggests that some sRNAs may be important in *Buchnera's* gene regulation.

Chong et al. (2019), recently compared the genomes of 39 *Buchnera* strains and found that across these taxa, 29 genes are under strong positive selection within the aphid subfamily Aphidinae. Interestingly, of these 29 genes there is evidence that 24 of them are predicted targets of conserved antisense or UTR sRNAs (Hansen & Degnan, 2014). Twenty of 29 *Buchnera* genes under positive selection identified by Chong et al., (2019), were predicted to be either direct or indirect targets of one or more differentially expressed antisense sRNA identified in this study. Four of these genes (*asps, mtlA, rnr,* and *serC*)*,* were also found to be differentially expressed as proteins in *Buchnera* embryos and maternal bacteriocytes which correspond to differentially expression patterns of antisense sRNA identified in this and other studies (Hansen & Degnan, 2014; Thairu et al., 2018).

In recent years, increasing numbers of studies have begun to find sRNAs being expressed in highly reduced genomes, including organelles (Dietrich et al., 2015; Thairu & Hansen, 2019). Though the roles of sRNAs within these reduced genomes are still poorly understood; within mitochondria and chloroplasts they have been described to respond to stress (*e.g.* Georg et al., 2010; Zghidi-Abouzid et al., 2011; Hackenberg et al., 2013; Rompala et al., 2018) and host development and/or tissue type (*e.g.* Itaya et al., 2008; Ro et al., 2013; Ma et al., 2016; Cognat et al., 2017). Based on the trends observed in other reduced genomes and the patterns that we observed within the *Buchnera* system, we hypothesize that sRNAs can be maintained and evolve in reduced genomes potentially to help compensate for the loss of regulatory proteins. Host-restricted bacterial symbionts with highly eroded genomes such as *Buchnera,* face the challenge of having to compensate for the loss of key genomic material and until recently many of the compensatory methods that have been described have been host meditated. Moving forward, based on this and other studies, we predict that more examples of symbiont mediated compensatory mechanisms of control will emerge.

References:

Babski, J. *et al.* 2014. Small regulatory RNAs in Archaea. *RNA Biol.* 11:484–493.

Berasategui A. *et al.* 2017. Gut microbiota of the pine weevil degrades conifer diterpenes and increases insect fitness. *Mol. Ecol.* 26:4099–4110 2017.

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474.

Brant EJ, Budak H. (2018). Plant Small Non-coding RNAs and Their Roles in Biotic Stresses. *Front. Plant Sci.* 9:1038.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience, New York, NY

Ceja-Navarro JA. *et al.* 2015. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nat. Commun.* 6:7618.

Chong RA, Park H, Moran NA. 2019. Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol*. doi:10.1093/molbev/msz082

Cognat, V. *et al.* 2016.The nuclear and organellar tRNA-derived RNA fragment population in Arabidopsis thaliana is highly dynamic. *Nucleic Acids Res.* 45:3460–3472 .

Daub ME, Ehrenshaft M. 2000. The Photoactivated Cercospora Toxin Cercosporin : Contributions to Plant Disease and Fundamental Biology. *Annu. Rev. Phytopathol.* 38:461–490.

Dietrich A, Wallet C, Iqbal RK, Gualberto JM, Lotfi, F. 2015.. Organellar non-coding RNAs: Emerging regulation mechanisms. *Biochimie* 117: 48–62 2015.

Fernández, I. S. *et al.* 2013.Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* 500:107–110.

Georg J, Honsel A, Voß B, Rennenberg H, Hess WR. 2010. A long antisense RNA in plant chloroplasts. *New Phytol.* 186:615–622.

Glasgow H. 1914. The gastric cæca and the cæcal bacteria of the Heteroptera. *Biol. Bull.* 26:101 .

Goodchild AJP.1963. Studies on the functional anatomy of the intestines of Geteroptera. *Proc. Zool. Soc.* London 141:851–910 .

Hackenberg M. *et al.* 2013. A Comprehensive expression profile of micrornas and other classes of non-coding small RNAs in barley under phosphorous-deficient and-sufficient conditions. *DNA Res.* 20:109–125.

Hansen AK, Degnan PH. 2014. Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J.* 8:2490–2502.

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

Hansen AK, Moran NA. 2012.Altered tRNA characteristics and 3′ maturation in bacterial symbionts with reduced genomes. *Nucleic Acids Res.* 40:7870–7884.

Hansen AK, Moran, NA. 2014.The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* 23:1473–1496.

Hoernes, T. P. *et al.* 2016. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* 44:852–62.

Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319:1059–1066 .

Hör J, Gorski SA, Vogel J.(2018) Bacterial RNA Biology on a Genome Scale. *Molecular Cell* 70:785–799.

Husnik F. *et al.* 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567.

Itaya, A. *et al.* 2008. Small RNAs in tomato fruit and leaf development. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1779:99–107.

Janson EM, Stireman JO, Singer MS, Abbot P. 2008. Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. *Evolution* 62:997–1012.

John RA.1995. Pyridoxal phosphate-dependent enzymes. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular* 1248:81–96.

Jozefczuk S. *et al.* 2010. Metabolomic and transcriptomic stress response of *Escherichia coli. Mol. Syst. Biol.* 6, 364..

Karijolich J, Yi C, Yu YT. 2015.Transcriptome-wide dynamics of RNA pseudouridylation. *Nature Reviews Molecular Cell Biology* 16:581–585.

Keffer-Wilkes LC, Veerareddygari GR, Kothe U. 2016.RNA modification enzyme TruB is a tRNA chaperone. *Proc. Natl. Acad. Sci.* 113:14306–14311.

Kim D, Minhas BF, Li-Byarlay H, Hansen AK. 2018.Key Transport and Ammonia Recycling Genes Involved in Aphid Symbiosis Respond to Host-Plant Specialization. *G3 (Bethesda).* 8:2433–2443.

Kim VN, Han J, Siomi MC. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10:126–139.

Kikuchi, Y. 2009. Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes Environ.* 24:195–204 .

Koga R, Meng XY, Tsuchida T, Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Natl. Acad. Sci. U. S. A.* 109:E1230-7.

Kucukural A, Yukselen O, Ozata DM, Moore MJ, Garber M. 2019. DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics* 20:6.

Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Luan JB. *et al.* 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-Feeding insects. *Genome Biol. Evol.* 7: 2635–2647.

Ma H, Weber GM, Wei H, Yao, J.2016. Identification of Mitochondrial Genome-Encoded Small RNAs Related to Egg Deterioration Caused by Postovulatory Aging in Rainbow Trout. *Mar. Biotechnol.* 18:584–597.

Mao M, Yang X, Bennett GM. 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc. Natl. Acad. Sci.* 115:E11691–E11700.

Marbaniang CN, Vogel J. (2016) Emerging roles of RNA modifications in bacteria. *Curr. Opin. Microbiol.* 30, 50–57.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17:10-12.

McClure, R. *et al.* 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 41:e140.

Mi, H. *et al.* (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14:703–721.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68:195–215.

Moran NA, Dunbar HE, Wilcox JL. 2005. Regulation of transcription in a reduced bacterial genome: Nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *J. Bacteriol.* 187:4229–4237.

Nakabachi A. Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. 2005.Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera. Proc. Natl. Acad. Sci.* 102:5477–5482.

Narra, H. P. *et al.* 2016. Small Regulatory RNAs of Rickettsia conorii. *Sci. Rep.* 6:36728.

O'Connor M, Leppik M, Remme J. 2018. Pseudouridine-free Escherichia coli ribosomes. *J. Bacteriol.* 200:e00540-17 2018.

Oksanen, J. *et al.* 2019. Package "vegan" Title Community Ecology Package. *Community Ecol. Packag.* 2.

Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019.PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics* 20:89–108.

Poliakov, A. *et al.* 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Mol. Cell. Proteomics* 10:M110.007039.

Price DRG. *et al.* 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proc. Natl. Acad. Sci.* 111:320–325.

Reymond, N. *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.

Ro, S. *et al.* 2013.The mitochondrial genome encodes abundant small noncoding RNAs. *Cell Res.* 23:759–74 .

Rompala GR, *et al.* 2018. Heavy chronic intermittent ethanol exposure alters small noncoding RNAs in mouse sperm and epididymosomes. *Front. Genet.* 9:1–14.

Rutherford, K. *et al.* 2000. Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944–945.

Salem, H. *et al.* 2017. Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. *Cell* 171:1520–1531.e13.

Sanchez-Arcos C, Kai M, Svatoš A, Gershenzon J, Kunert, G. 2019. Untargeted Metabolomics Approach Reveals Differences in Host Plant Chemistry Before and After Infestation With Different Pea Aphid Host Races. *Front. Plant Sci.* 10:188.

Sanchez-Arcos C, Reichelt M, Gershenzon J, Kunert G. 2016. Modulation of Legume Defense Signaling Pathways by Native and Non-native Pea Aphid Clones. *Front. Plant Sci.* 7:1872.

Sandström J, Moran N. 1999. How nutritionally imbalanced is phloem sap for aphids? *Entomologia Experimentalis et Applicata* 91:203–210..

Sandström J, Pettersson J. 1994. Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. *J. Insect Physiol.* 40:947–955.

Schoonhoven LM, Loon JJA, Dicke M. 2005.Insect-Plant Biology. Oxford University Press, Oxford, UK.

Schroeder, C. L. C. *et al.* 2017.Transcriptional profiling of Rickettsia prowazekii coding and non-coding transcripts during in vitro host-pathogen and vector-pathogen interactions. *Ticks Tick. Borne. Dis.* 8:827–836.

Serbus LR, Rodriguez BG, Sharmin Z, Momtaz AMZ, Christensen S. 2017. Predictive Genomic Analyses Inform the Basis for Vitamin Metabolism and Provisioning in Bacteria-Arthropod Endosymbioses. *G3 (Bethesda)*. 7:1887–1898.

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.

Shigenobu S, Wilson ACC. 2011. Genomic revelations of a mutualism: The pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences* 68:1297–1309.

Simpson SJ, Simpson CL. 1990. The mechanism of nutritional compensation by phytophagous insects. In Insect Plant Interactions, vol.II (ed. E. A. Bernays), pp.111 -160.Boca Raton: CRC Press

Sloan, D. B. *et al.* 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol. Biol. Evol.* 31:857–871.

Sudakaran S, Kost C, Kaltenpoth M. 2017. Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends in Microbiology* 25:375–390.

Sugio A, Dubreuil G, Giron D, Simon JC. 2015. Plant-insect interactions under bacterial influence: ecological implications and underlying mechanisms. *J. Exp. Bot.* 66:467–478.

Thairu MW, Cheng S, Hansen AK. 2018. A sRNA in a reduced mutualistic symbiont genome regulates its own gene expression. *Mol. Ecol.* 27, 1766–1776.

Thairu MW, Hansen AK. 2019.It's a small, small world: Unravelling the role and evolution of small RNAs in organelle and endosymbiont genomes. *FEMS Microbiol. Lett.* 366.

Viñuelas, J. *et al.* 2011. Multimodal dynamic response of the *Buchnera aphidicola* pLeu plasmid to variations in leucine demand of its host, the pea aphid *Acyrthosiphon pisum*. *Mol. Microbiol.* 81:1271–1285.

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.

Wilson ACC. *et al.* 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.* 19:249–258.

Yuan, E. *et al.* 2019.Increases in Genistein in Medicago sativa Confer Resistance against the Pisum Host Race of *Acyrthosiphon pisum*. *Insects* 10:97.

Zghidi-Abouzid O, Merendino L, Buhr F, Malik Ghulam M, Lerbs-Mache S. 2011). Characterization of plastid psbT sense and antisense RNAs. *Nucleic Acids Res*. 39:5379–87.

Tables and Figures:

Table 3.1:

Differentially expressed sRNAs in aphid host-plant comparisons (q≤0.05; 1.5-fold change)

Comparison: ALF-BAC vs. FB-BAC		
Type of sRNA	**Up-regulated in ALF**	**Up-regulated in FB**
Antisense sRNA	69	17
Intergenic sRNA	19	2
UTR sRNA	11	0
tRNAs		
Sense	4	1
Antisense	0	8
Total	103	28

Comparison: ALF-EMB vs. FB-EMB		
Type of sRNA	**Up-regulated in ALF**	**Up-regulated in FB**
Antisense sRNA	63	56
Intergenic sRNA	22	7
UTR sRNA	10	3
tRNAs		
Sense	4	2
Antisense	0	9
Total	99	77

Table 3.2:

GO PANTHER Pathways of the predicted CDS for differentially expressed sRNAs with significantly stable secondary structure predictions in the aphid-host plant comparisons. Bolded pathways are related to essential amino acid biosynthesis.

Pathway associated with predicted sRNA target	Differentially expressed sRNA	Host-plant treatment that the sRNA is up-regulated	
		Comparison: ALF-BAC *vs.* FB-BAC	Comparison: ALF-EMB *vs.* FB-EMB
Arginine biosynthesis	antisense *carB_2*	FB	FB
Chorismate biosynthesis	antisense *aroA_1*	ALF	n.s.
	antisense *aroC*	FB	FB
De novo pyrimidine ribonucleotides biosynthesis	antisense *carB_2*	FB	FB
De novo pyrimidine deoxyribonucleotide	antisense *yfhC_2*	FB	FB
Isoleucine biosynthesis	antisense *ilvC_2*	ALF	ALF
	antisense *ilvD*	FB	FB
	antisense *ilvI_1*	FB	n.s.
	antisense *ilvI_2*	n.s.	FB
	antisense *ilvI_5*	FB	FB
Peptidoglycan biosynthesis	antisense *murD_1,3*	ALF	ALF
Salvage pyrimidine deoxyribonucleotides	antisense *yfhC_2*	FB	FB
Salvage pyrimidine ribonucleotides	antisense *yfhC_2*	FB	FB
Threonine biosynthesis	antisense *thrB_2,3*	n.s.	FB
	antisense *thrC_3*	FB	n.s.
Valine biosynthesis	antisense *ilvC_2*	ALF	ALF
	antisense *ilvD*	FB	FB
	antisense *ilvI_1*	FB	n.s.
	antisense *ilvI_2*	n.s.	FB
	antisense *ilvI_5*	FB	FB
Vitamin B6 metabolism	antisense thrC_3	FB	n.s.

n.s. indicates that he sRNA was not differentially expressed in this host-plant treatment

Table 3.3:

Differentially expressed sRNAs in aphid life-stage comparisons (q≤0.05; 1.5-fold change)

Comparison: ALF-BAC vs. ALF-EMB		
Type of sRNA	**Up-regulated in BAC**	**Up-regulated in EMB**
Antisense sRNA	0	22
Intergenic sRNA	1	3
UTR sRNA	0	1
tRNAs		
Sense	0	0
Antisense	0	3
Total	1	29

Comparison: FB-BAC vs. FB-EMB		
Type of sRNA	**Up-regulated in BAC**	**Up-regulated in EMB**
Antisense sRNA	12	171
Intergenic sRNA	1	27
UTR sRNA	1	6
tRNAs		
Sense	1	4
Antisense	0	14
Total	15	222

Table 3.4:

GO PANTHER Pathways of the predicted CDS for differentially expressed sRNAs with significantly stable secondary structure predictions in the life-stage comparisons. Bolded pathways are related to essential amino acid biosynthesis.

Comparison: FB-BAC *vs.* FB-EMB		
Pathway associated with predicted sRNA target	**up-regulated sRNA**	**Life-stage group that the sRNA is up-regulated in**
Acetate utilization	antisense *pta*	EMB
Arginine biosynthesis	antisense *carB_1*	EMB
Chorismate biosynthesis	antisense *aroA_1*	EMB
	antisense *aroC*	EMB
De novo purine biosynthesis	antisense *purA*	EMB
De novo pyrimidine ribonucleotides biosynthesis	antisense *carB_1*	EMB
De novo pyrimidine deoxyribonucleotide biosynthesis	antisense *dut_1*	EMB
De novo pyrimidine deoxyribonucleotide	antisense *yfhC_2*	FB
Isoleucine biosynthesis	antisense *ilvD*	EMB
	Antisense *ilvI_1,5*[1]	EMB
Peptidoglycan biosynthesis	antisense *murD_1,3*	EMB
Salvage pyrimidine deoxyribonucleotides	antisense *yfhC_2*	EMB
Salvage pyrimidine deoxyribonucleotides	antisense *yfhC_2*	EMB
Pyridoxal-5-phosphate biosynthesis	antisense *serC_2*	EMB
Serine glycine biosynthesis	antisense *serC_2*	EMB
Threonine biosynthesis	antisense *thrB_3*	EMB
Valine biosynthesis	antisense *ilvD*	EMB
	antisense *ilvI_5**	EMB
Vitamin B6 metabolism	antisense *serC_2*	EMB

[1]Antisense sRNAs *ilvI_1,5* were also detected in the ALF-BAC *vs.* ALF-EMB comparison and were up-regulated in the EMB life-stage group.
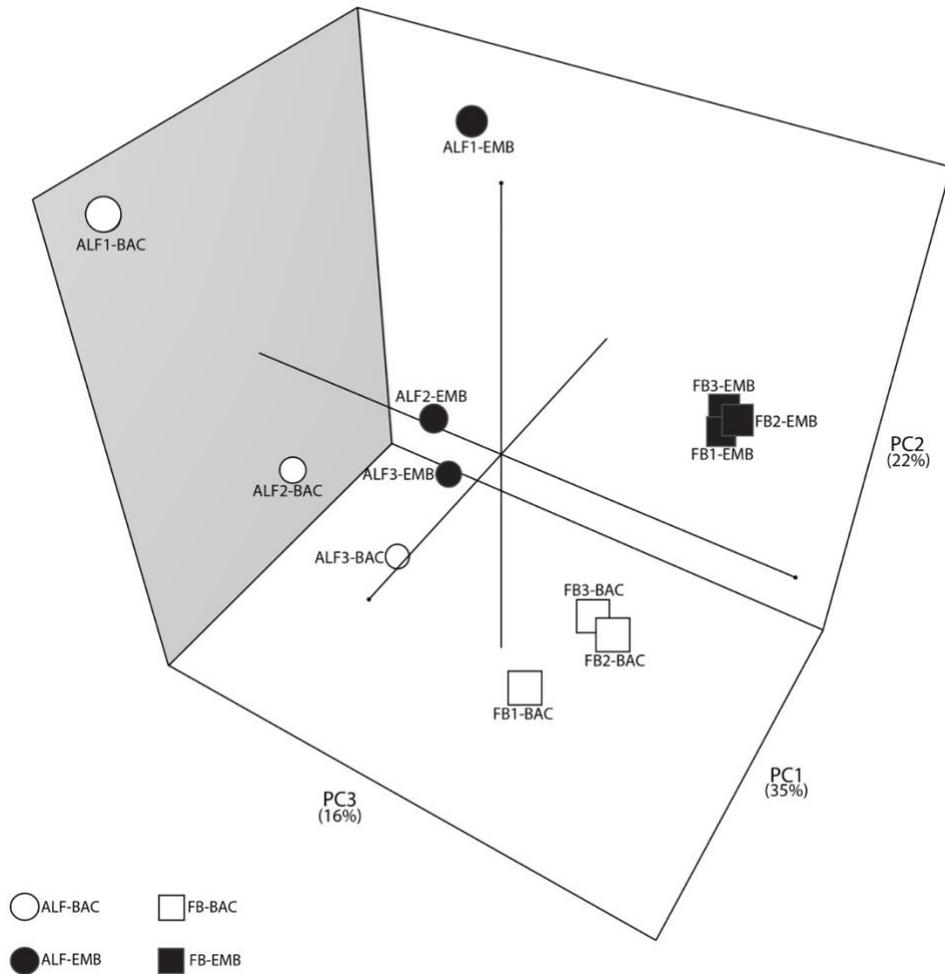
Figure 3.1:

Principal Component Analysis (PCA) of sRNA expression profiles for each aphid host-plant treatment (fava bean (FB) and alfalfa (ALF)) and aphid life-stage group (bacteriocytes (BAC) and embryo (EMB)). ALF-BAC1-3, ALF-EMB1-3, FB-BAC1-3 and FB-EMB1-3, represents the 3 biological replicates for each aphid host-plant treatment and life-stage group.

Figure 3.2: Integration of amino acid biosynthetic pathways of the aphid and *Buchnera* within the aphid bacteriocyte. Yellow boxes indicate the predicted CDS targets of differentially expressed *Buchnera* sRNAs identified in the ALF-BAC *vs.* FB-BAC comparison. Green and purple areas represent the cytosol of the bacteriocyte and of the *Buchnera* cell, respectively. Green and purple lines represent aphid and *Buchnera* cell membranes, respectively. Amino acids are represented by blue disks or by yellow (glutamine) or white (glutamate) disks with blue outlines. Aphid gene expression data adapted from Kim et al. (2018). *Buchnera* sRNA expression data are from this study.

Chapter 4: Identification of four conserved sRNAs in the nutritional endosymbionts of psyllids, *Candidatus* Carsonella ruddii

Abstract

During the evolution of symbiotic relationships, the genomes of obligate host-restricted bacteria become greatly reduced, with symbionts losing many key regulatory genes due to genetic drift. As a result, it is unclear how and if these bacterial symbionts regulate their gene expression. There is emerging evidence that the aphid symbiont, *Buchnera*, may utilize small RNAs (sRNAs). However, it is unclear if all obligate host-restricted bacteria express regulatory sRNAs. To address this question, this study characterizes the sRNA expression profile of one of the most reduced obligate insect symbiont genomes, *Candidatus* Carsonella ruddii. In this study, the sRNA expression profiles of two strains of *Candidatus* Carsonella ruddii, one from the psyllid *Bactericera cockerelli (Carsonella-BC) and another from the psyllid Diaphorina citri (Carsonella-DC),* are characterized and compared using RNA-seq. Two life-stages of *B. cockerelli* are also used to identify differentially expressed Carsonella-BC sRNAs. The sRNA expression profile of *D. citri's* defensive symbiont, *Candidatus* Profftella armature, is also determined. Overall, within *Carsonella-BC,* 37 antisense sRNAs were identified *and within Carsonella-DC,* 32 antisense sRNA*s* were identified. F*our sRNAs* predicted to target the coding sequences (CDSs), *aroC, clpX, carB* and *prfA* were conserved between these two *Carsonella* taxa. Three of these four CDSs*, clpX, carB* and *prfA,* are also the targets of conserved antisense sRNAs in the endosymbiont of aphids, *Buchnera aphidicola.* The results from this study also demonstrate that the expressed *Carsonella-BC* sRNAs may be functional as they are differentially expressed between the two *B. cockerelli* life-stages investigated. Within *D. citri's* defensive symbiont *Candidatus* Profftella armature, 190 antisense and intergenic sRNAs were expressed – 16 of which were predicted to target genes in the polyketide synthase biosynthetic gene clusters. The results of this experiment provide further

93

evidence in support of the hypothesis that highly reduced genomes of obligate host-restricted bacteria may utilize sRNA-mediated gene regulation to compensate for the loss of regulatory proteins.

Introduction:

Many obligate intracellular symbionts of insects are characterized by having highly reduced genomes which lack key regulatory proteins (Moran & Bennett, 2014). As such, it has been hypothesized that within these symbiotic relationships the host is primarily responsible for regulating the symbiosis (Hansen & Moran, 2014). However, recently it has been shown that the aphid symbiont, *Buchnera* expresses putative regulatory small RNAs (sRNAs) that are conserved across *Buchnera* lineages that diverged >65 million years ago (Hansen & Degnan, 2014; Thairu et al., 2018). Bacterial sRNAs serve as important post-transcriptional regulatory mediators that are metabolically cheap and allow the organism to rapidly respond to changes in the environment (Beisel & Storz, 2010). Within the highly reduced genomes of host-restricted bacterial symbionts, there is evidence that suggests that putative sRNAs have regulatory roles however, many of these examples come from organelles and only a few bacterial symbiont systems (Thairu & Hansen, 2019). As such, the goal of this study is to expand our understanding of sRNA regulation in small genomes by characterizing sRNA expression in a greater diversity of symbiont taxa.

Within the Hemipteran Superfamily, Psylloidea, all psyllid members have evolved an obligate symbiotic relationship with the Gamma-proteobacteria, *Candidatus* Carsonella ruddii (hereafter *Carsonella*) (Thao et al., 2000; Hall et al., 2016). Like many obligate, host-restricted symbionts, *Carsonella* has a very small genome of approximately (~166 kb), making it one of the smallest insect symbionts (Moran and Bennett, 2014; NCBI, 2019). In this study we analyzed two different lineages of *Carsonella.* The first from the potato psyllid, *Bactericera cockerelli (Carsonella-BC),* a pest of Solanaceous crops and the vector of *Candidatus* Liberibacter psyllaurous, which is associated with psyllid yellows (Hansen et al., 2008). The second *Carsonella* taxa is harbored in the Asian citrus psyllid, *Diaphorina citri* (*Carsonella-DC*), the vector of *Candidatus* Liberibacter asiaticus, which is associated with citrus greening disease (Jagoueix et al.,

1994). *Carsonella-DC,* unlike many other *Carsonella* taxa sequenced to date (including *Carsonella-BC*), has complete amino acid biosynthesis pathways for tryptophan and histidine (Nakabachi et al., 2006; Sloan and Moran, 2012; Nakabachi et al., 2013; Riley et al., 2017). *D. citri* also houses the obligate bacterial symbiont *Candidatus* Profftella armature (hereafter *Profftella-DC*) in the syncytial region of the bacteriocyte (Subandiyah et al., 2000; Nakabachi et al., 2013). Though larger than *Carsonella, Profftella-DC* still has a reduced genome (~464KB) and serves as a defensive symbiont producing the toxin diaphorin (Nakabachi et al., 2013; Szebenyi al., 2018; Yamada et al., 2019). Both *Carsonella,* and *Profftella-DC,* are vertically transmitted transovarially (Dan et al., 2017).

I use RNA-seq to characterize and compare the sRNA expression profile of the two *Carsonella* lineages, *Carsonella-BC and Carsonella-DC*. Here, I also determine if there is conservation of sRNAs between these two taxa and if sRNAs are differentially expressed between different life-stages of *B. cockerelli*. Finally, the expressed sRNAs of *Profftella-DC* are also identified.

Methods:

Small RNA sample preparation and sequencing

Bactericera cockerelli and *D. citri* psyllids were reared at ~27°C under a 16-h light/8-h dark regime on tomato (*Solanum lycopersicum*) and ~1 year old curry leaf (*Murraya koenigii*) plants, respectively. For *B. cockerelli,* three samples were collected: 1. Dissected and isolated adult bacteriocytes (BC-A), 2. Whole body 5th instar nymphs (BC-N) and 3. A mix of whole body 1st-5th instar nymphs and adults (B-All). For the first and second samples (BC-A and BC-N), three biological replicates of ~60 psyllids (approx. 30 males and 30 females per sample) were collected. For the third *B. cockerelli* sample (B-All), a mixed population of ~20 psyllids per life-stage were

collected and combined into a single sample. For the *D. citri* sample, similar to the *B. cockerelli* mixed population sample, ~20 *D. citri* psyllids of each life-stage were collected and combined in a single sample (DC-All). All tissues were immediately placed in RNAprotect Bacteria Reagent (Qiagen, Germantown, MD) and stored at -80C.

For each sample, RNA was extracted using the Quick-RNA Microprep (Zymo, Irvine, CA). Library preparation and sequencing was then performed on the sRNA enriched fraction (≤ 250 nt) using the Illumina mRNA directional sequencing protocol by the University of California, San Diego, Institute for Genomic Medicine Genomics Center (UCSD IGM Genomics Center). Each library was then sequenced as 75 nt single-end reads on the Illumina Hi-seq 4000 (San Diego, CA, USA) at the UCSD IGM Genomics Center. A total of seven *B. cockerelli* samples were sequenced: three adult bacteriocyte biological replicate samples (BC-A1-3), three 5th instar nymph biological replicate samples (BC-N1-3), and one sample with a mix of life stages (1st instar-adult) (BC-All). For *D. citri,* one sample containing a mix of life stages (1st instar-adult) was sequenced (DC-All).

Identification and categorization of Carsonella and Profftella sRNAs

For all samples, reads were quality screened and adapters were removed using Trimmomatic v.0.33(Bolger *et al.,* 2014) and Cutadapt v2.1 (Martin, 2011). For *B. cockerelli* samples, reads mapping to *Carsonella-BC*, were aligned using Bowtie2 v.2.2 (Langmead *et al.*, 2009). Bowtie2 v.2.2 was also used to map reads from the *D. citri* sample to either *Carsonella-DC* or *Profftella-DC.* Rockhopper v.2.0.3 (McClure *et al.,* 2013) was then used to identify putative *Carsonella* and *Profftella-DC* sRNAs. To maximize detection of conserved sRNAs between the *Carsonella* lineages, the default parameters of Rockhopper were modified; specifically, the "minimum expression of untranslated regions (UTR) and non-coding RNAs (ncRNA)" parameter was modified to 0.3. All other parameters were left at the default settings for strand-specific reads.

Using the same default parameters for strand-specific reads in Rockhopper putative sRNAs were identified in *Profftella-DC* as well. All symbiont sRNAs were then binned into three different categories: sRNAs expressed antisense to the gene (antisense sRNAs), sRNAs expressed within the untranslated regions of genes (UTR sRNAs), and sRNAs identified within the intergenic spacer regions (intergenic sRNAs). For conserved sRNAs, boundaries were manually determined by inspecting directional coverage curves in Artemis v.16 (Rutherford et al., 2000). RNAalifold (Bernhart *et al*., 2008) was then used to predict the secondary structure following the methods of Hansen & Degnan (2014).

Rockhopper was also used to determine if *Carsonella-BC* sRNAs were differentially expressed between the two life-stages, adult (samples: BC-A1-3) and 5th instar nymphs (BC-N1-3). Rockhopper normalizes reads among samples using upper-quantile normalization (McClure *et al.,* 2013). Because RNAseq coverage was 66X lower for BC-N1-3 samples compared to the BC-A1-3 samples only sRNAs that were expressed highly in both treatments could be analyzed after normalization for differential expression.

Results:

Sequencing summary

To determine if sRNAs are conserved between *Carsonella* lineages, the sRNA profiles of the two distinct taxa were characterized using directional RNA-seq. Across all *B. cockerelli* samples, an average of 6.10×10^6 high quality reads mapped to *Carsonella-BC* (Table 4.1). Within the BC-A1-3 samples, an average of 1.99×10^6 reads mapped to *Carsonella-BC,* which resulted in a depth of coverage of ~860X (Table 4.1). For the BC-N1-3, an average of 3.28×10^4 reads mapped to *Carsonella-BC,* which resulted an average depth of coverage of ~13X. For the mixed life-stage (BC-All) sample 4.69×10^4 reads mapped to *Carsonella-BC,* which resulted an average depth of

coverage of ~20X (Table 4.1). For the DC-All sample, 1.06×10^4 (~46X depth of coverage) and 2.26×10^6 (~365X depth of coverage) high quality reads mapped to the *Carsonella-BC* and *Profftella-DC* genomes, respectively (Table 4.1).

Carsonella expresses conserved sRNAs

Overall, within both lineages of *Carsonella,* only antisense sRNAs were found to be expressed. For *Carsonella-BC,* all seven samples (BC-A1-3, BC-N1-3, and BC-All) were used to initially determine strain-specific expression of sRNAs. From these samples, 36 antisense sRNAs which are predicted to target 27 coding sequences (CDSs) were found (Supplementary Table 4.1). PANTHER GO functional gene list analysis (Mi et al. 2019) was then used to determine GO pathways associated with the predicted target CDS of identified sRNAs. For the predicted CDSs of expressed antisense sRNAs found within *Carsonella-BC,* 13 GO pathways were identified, seven of which were associated with the biosynthesis of the essential amino acids: arginine, chorismate, histidine, isoleucine, leucine, lysine, and valine (Table 4.2).

Thirty-two antisense sRNAs were found to be expressed in *Carsonella-DC*. These 32 antisense sRNAs were predicted to target 27 CDSs (Supplementary Table 4.2). Similar to *Carsonella-BC* samples, the majority (7/10) of the identified GO pathways were associated with of the biosynthesis of essential amino acids; specifically, arginine, chorismate, isoleucine, leucine, lysine, threonine and valine (Table 4.2).

In both *Carsonella-BC* and *Carsonella-DC,* antisense sRNAs were predicted to target the CDSs: *aroC, atpA, atpF, carAc/carA-carB, carB, clpX, dnaK, gidA, grepE, leuC,* and *prfA* (Supplementary Table 4.1-2). To identify conserved sRNAs between *Carsonella* taxa the following criteria from Hansen & Degnan, 2014 were used: *1*. the sRNA is a discreet transcript at a specific location in the genome, and *2.* the sRNA transcript meets the Rockhopper optimized thresholds as

described above. Using these criteria, four antisense sRNAs predicted to target *aroC, clpX, carB,* and *prfA* were found to be conserved between both *Carsonella-BC* and *Carsonella-DC* (Figure 1).

Carsonella-BC antisense sRNAs are Differentially expressed between B. cockerelli life-stages

For *Carsonella-BC,* sRNA expression was characterized in two different life-stages; adults (BC-A1-3) and 5th instar nymphs (BC-N1-3). Rockhopper was then used to identify differentially expressed sRNAs between these two life-stages. Six antisense sRNAs were found to be up-regulated in BC-A1-3 samples and seven were up-regulated in BC-N1-3 samples (Supplemental Table 4.3). These differentially expressed sRNAs were predicted to target CDSs in six PANTHER GO pathways (Table 4.3), including two amino acid biosynthesis pathways. Three of the antisense sRNAs up-regulated in the BC-N1-3 samples were predicted to target the CDSs in the following GO pathways: 5-Hydroxytryptamine degradation, ATP synthesis and pentose phosphate (Table 4.3). Two antisense sRNAs up-regulated in the BC-A1-3 samples, were predicted to target CDSs in the three GO pathways of arginine, leucine, and *de novo* pyrimidine ribonucleotide biosynthesis (Table 4.3).

Profftella-DC expresses both antisense and UTR sRNAs

Rockhopper identified 190 sRNAs expressed within *Profftella*-DC (Supplementary Table 4.4). These sRNAs comprise of 186 antisense sRNAs and four intergenic sRNAs. No UTR sRNAs were observed. The antisense sRNAs are predicted to target 111 unique CDSs. Notably, sixteen antisense sRNAs (antisense sRNAs, *dipE_1-2, dipJ, dipO, dipP_1-2, dipQ, dipR_1-2,* and *dipT_1-7*) were predicted to target seven CDSs in the polyketide synthase biosynthetic gene clusters (Supplementary Table 4.4). The polyketide synthase biosynthetic genes are primarily responsible for the production of the toxin, diaphorin (Nakabachi et al., 2013).

100

Discussion:

This study reveals for the first time that *Carsonella,* an insect symbiont with a tiny genome, expresses conserved sRNAs; four of which are conserved between two divergent lineages of *Carsonella*, which are harbored in the psyllids *B. cockerelli* and *D. citri* Percy et al., 2018). This study also provides evidence that some of these sRNAs are differentially expressed between life-stages, which provides additional evidence that expressed sRNAs may be functional. Future studies with higher sRNA coverage for the nymphal stage may detect even more sRNAs with differential expression due to higher sensitivity. Within the defensive symbiont, *Profftella-DC,* both antisense and intergenic sRNAs were found to be expressed. Overall, these results provide the basis for further studies to investigate the functionality of *Carsonella* and *Profftella* expressed sRNAs in gene regulation.

The aphid-*Buchnera* endosymbiosis is a well-established model that is used to study nutritional symbioses. *Buchnera* like *Carsonella* has a reduced genome, albeit larger than *Carsonella* at ~559KB (Shigenobu et al., 2000; Chong et al., 2019; NCBI, 2019). Within *Buchnera*, Hansen & Degnan (2014), characterized the expressed sRNAs of five divergent lineages of *Buchnera,* and within each taxa they found ~236 antisense sRNAs, 115 of which were conserved in two or more taxa (Hansen & Degnan, 2014). Overall, this study found an average of 34 antisense sRNAs per *Carsonella* taxa, which is similar to the relative number identified in *Buchnera,* given the difference in genome of size and gene number; the genome sizes and gene number of the *Carsonella* taxa used in this study are ~1/4 the size of the *Buchnera* taxa that were analyzed by Hansen & Degnan (2014) (Table 4.1). *Carsonella* genomes are characterized by having high gene density with very few intergenic regions (Sloan & Moran, 2012). As a result, the lack of intergenic sRNAs observed is not surprising. There is also evidence that suggests that UTR-encoded sRNAs

are lost as genomes shrink (Matelska et al., 2017), which would corroborate with the observations made in this study.

A total of four sRNAs that are expressed antisense to the CDSs *aroC, clpX, carB,* and *prfA* are conserved between *Carsonella* taxa. Interestingly, three of these conserved sRNAs that are expressed antisense to *clpX, carB* and *prfA* are also conserved among *Buchnera* taxa. If these sRNAs are functional, this finding suggests that these different symbiont species, which have co-evolved in two different insect super-families may have independently evolved regulatory sRNA to target the same CDSs. These sRNAs are viable targets for future functional studies. Notably, the conserved *Buchnera* antisense sRNA *carB* was shown to activate/or stabilize its predicted gene target when heterologously expressed in *Escherichia coli* (Thairu et al., 2018). These results were also corroborated *in vivo* as the *Buchnera* antisense sRNA *carB i*s up-regulated in aphid ovarioles, the same life-stage that the protein, CarB, is up-regulated in (Hansen & Degnan, 2014; Thairu et al., 2018).

With the increase of "omics" based experiments, there is emerging evidence that sRNAs are expressed within highly reduced bacterial genomes, and these sRNAs potentially have functional roles in gene regulation (Dietrich et al., 2015, Thairu & Hansen, 2019). The main challenges faced in both free-living and symbiotic bacterial systems is determining if expressed sRNAs are functional, transcriptional noise, or the products of RNA degradation/processing (Georg and Hess, 2011; Jackowiak et al., 2011; Lloréns-Rico et al., 2016). Now that sRNAs have been observed in the psyllid-*Carsonella* system proteomic and functional genomic experiments are needed to further determine if the expressed sRNAs are regulatory. Based on evidence from other systems where bacterial symbionts have reduced genomes, we predict that a fraction of the sRNAs identified are functional.

References:

Beisel CL, Storz G.2010. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev*. 34:866–882.

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Chong RA, Park H, Moran NA. 2019. Genome Evolution of the Obligate Endosymbiont *Buchnera aphidicola. Mol. Biol. Evol.* (2019). doi:10.1093/molbev/msz082

Dan H, Ikeda N, Fujikami M, Nakabachi A. 2017. Behavior of bacteriome symbionts during transovarial transmission and development of the Asian citrus psyllid. *PLoS One* **12:**e0189779.

Dietrich A, Wallet C, Iqbal RK, Gualberto JM, Lotfi, F. 2015. Organellar non-coding RNAs: Emerging regulation mechanisms. *Biochimie* 117:48–62.

Georg J, Hess WR. 2011. cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.* **75:** 286–300.

Hall AAG, et al. 2016. Codivergence of the primary bacterial endosymbiont of psyllids versus host switches and replacement of their secondary bacterial endosymbionts. *Environ. Microbiol.* **18:**2591–2603.

Hansen AK, Degnan PH. 2014.Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J.* 8:2490–502.

Jackowiak P, Nowacka M, Strozycki PM, Figlerowicz M. 2011. RNA degradome-its biogenesis and functions. Nucleic Acids Res. 39:7361–7370.

Jagoueix S, Bove JM, Garnier M.1994 The Phloem-Limited Bacterium of Greening Disease of Citrus Is a Member of the Subdivision of the Proteobacteria. *Int. J. Syst. Bacteriol.* 44:379–386.

Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Llorens-Rico V, et al. 2016.Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv.* **2:**1501363–e1501363.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17:10-12.

Matelska D, Kurkowska M, Purta E, Bujnicki JM, Dunin-Horkawicz, S. 2016. Loss of conserved noncoding RNAs in genomes of bacterial endosymbionts. *Genome Biol. Evol*. 8:426–438.

McClure, R. *et al.* 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 41:e140.

Mi, H. *et al.* (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14:703–721.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68:195–215

Nakabachi A, et al. 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science.* **314:** 267-267.

Nakabachi A, et al. 2013. Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* **23:**1478–1484.

NCBI Genomes 2019: Candidatus Carsonella ruddii and *Buchnera aphidicola.* https://www.ncbi.nlm.nih.gov/genome/genomes/1470

Percy DM, et al. 2018. Resolving the psyllid tree of life: phylogenomic analyses of the superfamily Psylloidea (Hemiptera). *Syst. Entomol.* **43:**762–776.

Riley AB, Kim D, Hansen AK. 2017. Genome Sequence of " *Candidatus* Carsonella ruddii" Strain BC, a Nutritional Endosymbiont of *Bactericera cockerelli. Genome Announc.* 5

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. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. in *Molecular Biology and Evolution* **29:**3781–3792.

Subandiyah S, Nikoh N, Tsuyumu S, Somowiyarjo S, Fukatsu, T. 2004. Complex Endosymbiotic Microbiota of the Citrus Psyllid *Diaphorina citri* (Homoptera: Psylloidea). *Zoolog. Sci.* **17:**983–989.

Szebenyi DM, et al. 2018. Crystal structure of diaphorin methanol monosolvate isolated from Diaphorina citri Kuwayama, the insect vector of citrus greening disease. *Acta Crystallogr. Sect. E Crystallogr. Commun.* **74:**445–449.

Thairu MW, Cheng S, Hansen AK. 2018. A sRNA in a reduced mutualistic symbiont genome regulates its own gene expression. *Mol. Ecol.* 27:8: 1766-1776

Thairu MW, Hansen AK. 2019. It's a small, small world: Unravelling the role and evolution of small RNAs in organelle and endosymbiont genomes. *FEMS Microbiol. Lett.* 366

Thao ML, et al. 2000. Co-speciation of psyllids and their primary prokaryotic endosymbionts. *Appl. Environ. Microbiol.* 66:2898–2905.

Yamada T, Hamada M, Floreancig P, Nakabachi A. 2019. Diaphorin, a polyketide synthesized by an intracellular symbiont of the Asian citrus psyllid, is potentially harmful for biological control agents. *PLoS One* **14,** e0216319 (2019).

Tables and Figures

Table 1: Summary of RNA-seq data. BC-A1-3 and BC-N1-3 represents *B. cockerelli* life-stage samples.

	DC-All	BC-A1	BC-A2	BC-A3	BC-N1	BC-N2	BC-N3	BC-All
Total number of reads	4.64×10^7	3.81×10^7	4.32×10^7	4.42×10^7	4.49×10^7	5.92×10^7	5.49×10^7	5.75×10^7
Reads after quality screen and adapter trimming	2.92×10^7	3.30×10^7	3.59×10^7	3.63×10^7	2.96×10^7	2.97×10^7	3.78×10^7	3.54×10^7
Number of reads aligning to *Carsonella-BC*	1.06×10^5	2.24×10^6	1.80×10^6	1.91×10^6	4.00×10^4	2.61×10^4	2.56×10^4	4.69×10^4
Carsonella-BC genome coverage (genome size: 173,802 bp)	46	974	782	829	17	11	11	20
Number of reads aligning to *Profftella-DC*	2.26×10^6							
Profftella-DC genome coverage (genome size: 174,014 bp)	365							

Table 4.2: GO PANTHER pathways of the predicted CDSs for sRNAs found in *Carsonella-BC* and *Carsonella-DC.* Bolded pathways are related to essential amino acid biosynthesis

Pathway associated with predicted sRNA target	Predicted CDS of expressed sRNA	
	Carsonella-BC	*Carsonella-DC*
5-Hydroxytryptamine degradation	*putA*	-
Alanine biosynthesis	*ilvE*	-
Arginine biosynthesis	*carA-carB*	*argH*
	carB	*carAc*
	-	*carB*
ATP synthesis	*atpA*	*atpA*
Chorismate biosynthesis	*aroA*	-
	aroC	*aroC*
De novo purine biosynthesis	*purA*	-
De novo pyrimidine ribonucleotides biosynthesis	*carA-carB*	*carAc-carB*
	carB	*carB*
Histidine biosynthesis	*hisD*	-
Isoleucine biosynthesis	*ilvE*	*ilvD*
Leucine biosynthesis	*ilvE*	*leuD*
	leuC	*leuC*
Lysine biosynthesis	*ilvE*	*lysC*
	lysA	-
	dapF	-
Pentose phosphate pathway	*tktA*	*tktA*
Threonine biosynthesis	-	*lysC*
Valine biosynthesis	*ilvE*	*ilvD*

"-" not targeted in *Carsonella* taxa

Table 4.3: GO PANTHER pathways of the predicted CDSs for sRNAs of *Carsonella-BC* that are differentially expressed in adult (BC-A1-3) and nymph (BC-N1-) samples. Bolded pathways are related to essential amino acid biosynthesis

	Predicted CDS of expressed sRNA	
Pathway associated with predicted sRNA target	**Predicted CDS of differentially expressed sRNA *Carsonella-BC***	**Life-stage that sRNA is up-regulated**
5-Hydroxytryptamine degradation	*putA*	Nymph (BC-N1-3)
Arginine biosynthesis	*carA-carB*	Adult (BC-A1-3)
ATP synthesis	*atpA*	Nymph (BC-N1-3)
De novo pyrimidine ribonucleotides biosynthesis	*carA-carB*	Adult (BC-A1-3)
Leucine biosynthesis	*leuC*	Adult (BC-A1-3)
Pentose phosphate pathway	*tktA*	Nymph (BC-N1-3)
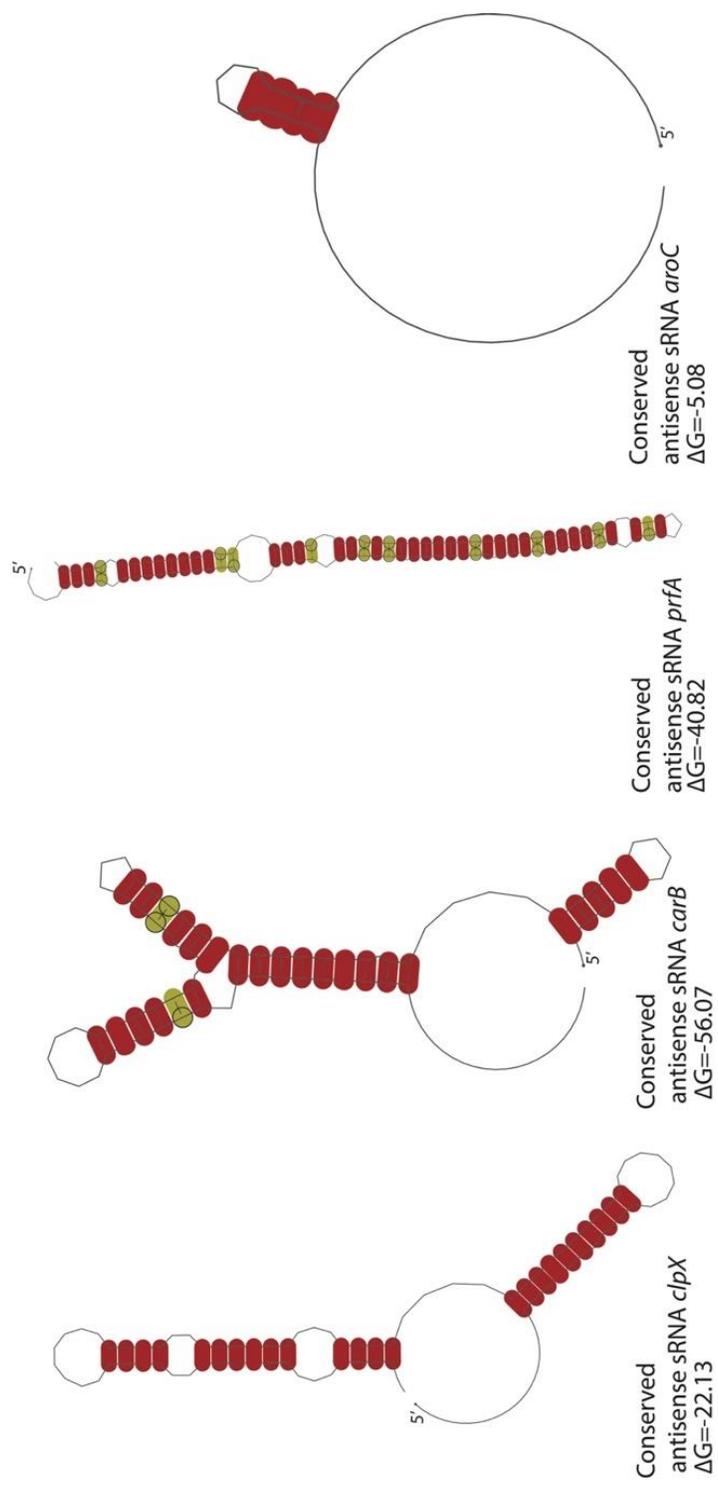
Figure 4.1:

Structural conservation diagrams based on RNAalifold and predicted thermostability of the four antisense sRNAs conserved between

Carsonella strains. Red: present in both genomes. Yellow: Compensatory changes.

Chapter 5:

Efficacy of RNAi knockdown using aerosolized siRNAs bound to nanoparticles in three divergent aphid species

Abstract

RNAi has emerged as a promising method for validating gene function; however, its utility in non-model insects has proven problematic, with delivery methods being one of the main obstacles. This study investigates a novel method of RNAi delivery in aphids, the aerosolization of a siRNA-nanoparticle complexes. By using nanoparticles as a siRNA carrier, the likelihood of cellular uptake is increased, when compared to methods previously used in insects. To determine the efficacy of this RNAi delivery system, siRNAs were aerosolized with and without nanoparticles in three aphid species*: Acyrthosiphon pisum, Aphis glycines,* and *Schizaphis graminum.* The genes targeted for knockdown were carotene dehydrogenase (*tor*), which is important for pigmentation in *A. pisum*, and branched chain-amino acid transaminase (*bcat*), which is essential in the metabolism of branched-chain-amino acids in all three aphid species. Overall, we observed modest gene knockdown of *tor* in *A. pisum* and moderate gene knockdown of *bcat* in *A. glycines* along with its associated phenotype. We also determined that the nanoparticle emulsion significantly increased the efficacy of gene knockdown. Overall, these results suggest that the aerosolized siRNA-nanoparticle delivery method is a promising new high-throughput and non-invasive RNAi delivery method in some aphid species.

Introduction

Currently, the biggest challenge for the field of genomics is the functional characterization of genes and linking them to organismal phenotypes. Over the last decade, there has been substantial progress in the sequencing of eukaryotic genomes. Presently, nearly 720 animal genomes including numerous insect, mammal, and avian species have been fully sequenced and/or are in draft form (NCBI, 2017). Gene prediction and annotation for these eukaryotic genomes has relied primarily on bioinformatics approaches utilizing *ab initio* gene models and homology. However, many of these computational predictions remain to be functionally validated. In non-model organisms, there is the additional challenge of characterizing lineage specific genes. Consequently, there is a need for functional genomic techniques that can be easily used in both model and non-model systems.

One promising approach for validating gene function in animals is RNA interference (RNAi), a type of post-transcriptional gene silencing that targets a specific mRNA. By introducing, double stranded RNA (dsRNA) or short interfering RNA (siRNA) the RNAi pathway can be induced to silence a gene of interest (Fig. 3.S1). The RNAi pathway is highly conserved across eukaryotes and within insects, including aphids (Shabalina and Koonin, 2008; Bellés 2010).

Unfortunately, reliable and widespread success of RNAi gene knockdown is limited, especially in some insect taxa (Scott et al., 2013; Yu et al., 2013; Wyant et al., 2014). Numerous factors influence RNAi's performance including the inherent characteristics of the animal tested, the tissue target, target gene and the delivery methods employed (Scott et al., 2013; Yu et al., 2013; Wyant et al., 2014). To this end, more research is desperately needed to troubleshoot different RNAi delivery methods, and to tease apart which mechanisms ultimately determine RNAi success and failure in different non-model systems.

Within the insect order Hemiptera, there are several key economic agricultural pests with sequenced genomes in which RNAi technology has had varied success. Aphids, in particular, represent a group of hemipteran pests in which there is a wealth of genomic information making them key candidates in functional genomic studies. Many critical RNAi genes are present in aphids suggesting that this group of insects would be ideal for RNAi gene knockdown (e.g. Jaubert-Possamai et al., 2010; Ortiz-Rivas et al., 2012; Bansal et al. 2013). However, in aphids there has been a large amount of variation in the success of using RNAi technology for gene knockdown. For example, plant-mediated RNAi within the green peach aphid, *Myzus persicae (*Sulzer), has high efficacy (Pitino et al., 2011; Guo et al., 2014; Coleman et al., 2015). On the other hand, in the pea aphid, *Acyrthosiphon pisum* (Harris) plant-mediated RNAi is more difficult to develop, because *A. pisum's* host plants are not easily transformable. Instead, siRNA or dsRNA have been delivered orally via artificial diets or by injection into the aphid. However, inconsistent and modest RNAi gene knockdown has been reported from these two delivery methods in *A. pisum* (Mutti et al., 2011; Jaubert–Possamai et al., 2007; Shakesby et al., 2009; Mao et al., 2012; Christiaens et al., 2014). Christiaens et al. (2014) reported that dsRNAs delivered either orally or by injection are rapidly degraded in both aphid salivary secretions and hemolymph (insect blood) by putative RNAses, which may help explain why RNAi may not work efficiently in *A. pisum* (Christiaens et al., 2014).

A novel approach that has been successfully used in a few insect systems is the delivery of siRNA and dsRNA using nanoparticle carriers (Zhang et al., 2010; Li-Byarlay et al., 2013; Das et al., 2015). A recent study in the honeybee demonstrated successful gene knockdown of the targeted gene by aerosolizing a siRNA- perfluorocarbon nanoparticle emulsion through the honeybee's spiracles (Li-Byarlay et al., 2013). Nanoparticle carriers, such as the perfluorocarbon nanoparticle emulsion used by Li-Byarlay et al. (2013) facilitate the transfer of nucleic acids across biological

barriers by enhancing their cellular uptake and by increasing the stability of dsRNA and siRNA (Ross et al., 2015).

In this current study, we determined the rate of dsRNA degradation after aerosolized nanoparticle delivery compared to microinjection in *A. pisum*. By aerosolizing the dsRNA with nanoparticles, we hope to limit the dsRNAs degradation because of the exposure to the hemolymph as shown in Christiaens et al., (2014) by directly targeting cellular tissues via tracheoles. Using similar methods and time points as Christiaens et al. (2014) we tested if double-stranded green fluorescent protein (GFP) RNA degrades rapidly inside of the aphid when using the aerosolized nanoparticle delivery method compared to microinjection.

We also investigated the efficacy of the aerosolized siRNA-nanoparticle technique in three aphid species the pea aphid, *A. pisum,* the soybean aphid, *Aphis glycines* (Matsumura) and the greenbug, *Schizaphis graminum* (Rondani). We first determined if we could knockdown the gene carotene dehydrogenase, *tor*, (LOC100169245). In *A. pisum*, *tor*, is a carotenoid gene that results in pink aphid pigmentation through torulene production (Moran & Jarvik, 2010). This gene was chosen because if *tor* is successfully knocked-down, the aphid cuticle is expected to display a green-yellow phenotype, compared to the native pink color displayed by our control *A. pisum* line. *Aphis glycines* and *S. graminum* were not included in the *tor* knockdown experiments because *S. graminum* does not produce torulene (Nováková et al., 2012) and the green-yellow pigmentation of *A. glycines* indicates a similar carotenoid profile to *S. graminum*.

For the second knockdown experiment, the gene *bcat,* (branched chain-amino acid transaminase) (LOC100167587) was selected because it is encoded in all aphid species in this study. Moreover, aphid encoded *bcat* is hypothesized to be important in the regulation of the terminal step of the branch-chain amino acid biosynthesis pathways (e.g. Leucine (Leu), Valine (Val), and Isoleucine (Ile)) (Wilson et al. 2010; Hansen & Moran, 2011; Poliakov et al., 2011),

113

because it is not encoded within the aphid's bacterial symbiont's (*Buchnera*) genome (Shigenobu et al., 2000). We predict that when *bcat* is successfully knocked-down there will be a reduction in the biosynthesis of these essential amino acids Leu, Val, and Ile. *Bcat* is also important for the degradation of Leu, Val, and Ile in other body tissues. We expect that a reduction in the biosynthesis and recycling of these branched-chain amino acids will result in lower aphid body mass compared to control aphids, because of the dysregulation of these essential nutrients.

In this study, we determined that *in vivo* dsRNA degradation was minimal when we delivered it via aerosolization, especially with the nanoparticle emulsion. Modest gene knockdown of *tor* in *A. pisum* was observed at lower concentrations of siRNA, but no phenotypic effect was observed. We did not achieve successful knockdown of *bcat* in *A. pisum* or *S. graminum.* However, in *A. glycines,* we observed that aerosolized siRNA-nanoparticles significantly knocked-down the target gene, *bcat,* and that there was a corresponding reduction in adult mass when compared to the control treatment. Overall, our findings indicate that aerosolized siRNA-nanoparticle delivery is a promising new high-throughput method of targeted gene knockdown in some aphid species. This method marks an improvement over traditional RNAi delivery methods (microinjection and artificial diets) because it can be widely applied to various non-model organisms in which the tradition RNAi delivery methods are not feasible or easily applied.

Results

Green fluorescent protein (GFP) dsRNA degradation trials

Degradation of dsRNA *in vivo* was tested using two different dsGFP (double-stranded Green Fluorescent Protein RNA) delivery techniques: *1-* direct microinjection, and *2-* a non-invasive aerosolization of siRNA-nanoparticle complexes using a nebulizer/compressor. For these experiments we used the *A. pisum* strain, LSR1.

dsRNA degradation trials: Microinjection

Our results indicate that in dsGFP microinjected aphids, dsGFP does not significantly degrade over time *in vivo*. However, there was a trend of degradation (ANOVA; F =2.28, d.f. = 4, P = 0.093; Fig. 5.1).

dsRNA degradation trials: Non-invasive aerosolization of nanoparticle complexes

In contrast to injection treatments, levels of intact dsGFP do significantly decrease over time for aphids that were treated with aerosolized dsGFP alone without nanoparticles (Treatment 1, F = 3.10, d.f. = 4, P = 0.04) (Fig. 5.2). Levels of intact dsGFP significantly decreased 5 hours after the aerosolized dsGFP treatment. However, 0.5, 5, and 24h treatments are not significantly different at α ≤ 0.05 (Fig. 5.2, Table 5.1). In comparison for both aerosolized dsGFP treatments with nanoparticles (Treatment 2 and Treatment 3), which varied 5-fold in dsGFP concentration, dsGFP does not significantly degrade over time (ANOVA; F = 2.07, d.f. = 4, P = 0.12; F = 0.66, d.f. = 4, P = 0.63, respectively, Fig. 5.2).

Overall, these results indicate that dsGFP delivered either through microinjection or aerosolization, especially with nanoparticles, is highly stable inside of the *A. pisum* (LSR1) over time.

<u>*siRNA- nanoparticle aerosolization trials*</u>

During each experiment, there were three treatments: *1*-target siRNA and nanoparticle emulsion (siRNA + nanoparticles), *2*-control siRNA and nanoparticle emulsion (control siRNA + nanoparticles) and *3*- siRNA only. For gene knockdown experiments, the sample size for each treatment was six aphids. The control siRNA was a randomized siRNA sequence that had no predicted target within the aphid.

Knockdown of the tor *gene*:

For the *tor* gene knockdown experiments, we used the *A. pisum* strain LSR1. For trial TOR-1, which tested the siRNA concentration of 100nM, we found a significant decrease in *tor* gene expression for the *tor* siRNA + nanoparticle experimental treatment compared to the control treatment (t-value = -2.08, P = 0.03; Fig. 5.3 A). This *tor* siRNA + nanoparticle treatment resulted in a ~12% knockdown compared to the control. Significant knockdown was not found for the siRNA only treatment compared to the control (t-value = -1.69, P = 0.06; Fig. 5.3 A) indicating that nanoparticles are important for *tor* knockdown in *A. pisum* (LSR1). For trial TOR-2, which used a 2-fold higher concentration of *tor*-siRNA, a significant difference in *tor* gene expression was observed in the *tor* siRNA + nanoparticles compared to the control (t-value= -2.90, P ≤ 0.01) but the percent knockdown was less than in TOR-1 at ~8% (Fig. 5.3B). Like the TOR-1 treatment, significant knockdown was not found for the siRNA only treatment compared to the control (t-value = 0.64, P = 0.64; Fig. 5.3 B) indicating that nanoparticles are important for *tor* knockdown in *A. pisum* (LSR1) at a siRNA concentration of 200nM. When we increased the concentration of siRNA in trials TOR-3 (siRNA concentration 500nM) and TOR-4 (siRNA concentration 1000nM) in *A. pisum* (LSR1) (Fig. 5.3C- D, Table 5.S3) we observed no significant gene knockdown. We also evaluated aphid pigment changes after all treatments prior to isolating RNA because *tor* is responsible for the aphid's pink pigmentation. Regardless of trial or treatment, we did not observe any visible change in aphid pigmentation, compared to the corresponding trial's control treatment in *A. pisum* (LSR1).

Knockdown of the bcat *gene*:

siRNA concentrations for the *bcat* experiments were optimized based on results from the first set of experiments with *tor*. Consequently, for *bcat* knockdown we tested the concentrations of siRNA at 100nM (BCAT-100) and 200nM (BCAT-200). We found no significant difference in

bcat expression in *S. graminum* and the two strains of *A. pisum* 5A and LSR1, between all three treatments (*1-* siRNA + nanoparticles, *2-* control siRNA + nanoparticles and *3-* siRNA only (Fig. 5.4 A- B, Tables S4.5-S5.5)). For these aphid species, we also observed no significant change in aphid mass between all three treatments (Fig.5.4 C-D, Tables S5.4-S5.5).

In contrast to *A. pisum* (5A and LSR1) and *S. graminum*, in *A. glycines* we observed both a significant gene knockdown and an effect on aphid mass for the aerosolized target siRNA nanoparticle trials (BCAT-200). Specifically, in the BCAT-200 *A. glycines* trials we observed a ~30% knockdown in the siRNA-nanoparticle treatment (t-value = 3.67 P ≤ 0.001; Fig. 5.5 B) and ~19% in the siRNA only treatment (t-value = 2.35, P=0.02; Fig. 5 .5B) when compared to the control. This gene knockdown was accompanied by a decrease in mass in the siRNA+ nanoparticle trial when compared to the control (t-value = 2.69, P = 0.01; Fig. 5.5 E). A significant decrease in mass was not found for the siRNA only treatment compared to the control (t-value = 0.91, P = 0.19; Fig. 5.5 E). These results strongly suggest that nanoparticles are important for the successful delivery of siRNAs in *A. glycines*. We were able to replicate a similar level of significant gene knockdown in the siRNA + nanoparticle treatment compared to the control in an independent trial and observed a ~30% knockdown (t-value = 1.92, P = 0.05; Fig. 5.5 C). In the replicate trial, we also observed a reduction in overall aphid mass in the *bcat* siRNA + nanoparticles treatment when compared to the control treatment (t-value = 3.62, P ≤ 0.001; Fig. 5.5 F), once more indicating that nanoparticles are important for the successful delivery of siRNAs in *A. glycines*.

In the *A. glycines* BCAT-100 trial we observed a significant decrease in aphid mass in the *bcat* siRNA treatment + nanoparticle treatment compared to the control (t-value = 2.13 P = 0.02; Fig. 5 .5A). However, this was not reflected in the relative gene expression results. Based on the RT-qPCR results we observed no significant gene knockdown in the *bcat* siRNA + nanoparticle treatment (t-value = -0.75, P = 0.24; Fig. 5.5 B), and no significant knockdown in the *bcat* siRNA

only treatment compared to the control treatment (t-value = -1.62, P = 0.08; Fig. 5.5 B). This suggests that lower concentrations of siRNA does not result in the significant knockdown of *bcat*.

Nanoparticles have no effect on aphid mass or mortality:

In the trials with *A. glycines* and *S graminum,* we observed a greater level of aphid mortality after the aerosolized spray treatments when compared to treatments conducted on *A. pisum.* To understand aerosolization treatment effects on these two species we exposed *A. glycines* and *S graminum* aphids to two additional experimental treatments: water +200pM nanoparticles, and water only.

Overall, aerosolized water with or without nanoparticles has no effect on aphid mass in *A. glycines,* or *S. graminum* (*S. graminum:* t-value = 0.97, P = 0.17; *A. glycines:* t-value = 0.44, P = 0.67; Fig. S5.2A). In *S. graminum* and *A. glycines* we observed no significant differences in mortality when aphids were exposed to aerosolized water with or without nanoparticles (*S. graminum* :z-score = 0.40, P = 0.35;*A. glycines* : z-score = 0.64 P = 0.26; Fig. S5.2B). It is important to note that the mortality that we observed when aphids were exposed to aerosolized water with nanoparticles was not significantly different from background mortality (*S. graminum:* z-score = 0.40, P = 0.35; *A. glycine:* z-score = 0.64 P = 0.2; Fig. S5.2B) . Background mortality observed in *A. glycines* was similar to what has been observed by McCornack et al. (2004).

Discussion

Results from our study suggest that a new RNAi delivery technique used in aphids that aerosolizes siRNA-nanoparticles into insect spiracles can result in successful, targeted gene knockdown in some aphid species depending on siRNA concentrations and gene targets. In our study, gene knockdown was observed in *A. glycines* and to a lesser extent in *A. pisum* (LSR1) when

aerosolized siRNAs and nanoparticles were administered to aphids compared to aerosolized siRNAs without nanoparticles, or a control treatment consisting of aerosolized control-siRNA-nanoparticles (Figs. 5.2 A-B, Fig. 5.4, and Fig. 5.5). These results indicate that siRNAs in combination with nanoparticles are required for successful target gene knockdown when using this particular mode of aerosolized delivery into aphids.

Variable success in RNAi knockdown has been a major obstacle curtailing widespread implementation of RNAi in insects and other organisms. In this study, we also observed lineage-specific responses to RNAi knockdown. We attempted to knockdown gene expression of two gene targets in *A. pisum*: *bcat* and *tor* and were able to achieve only modest gene knockdown of *tor* and no gene knockdown of *bcat*. Within the *tor* trials, we observed a modest gene knockdown at the lower concentrations tested of 100nM and 200nM (trials TOR-1 and TOR-2, Fig. 5.2 A-B). However, when we increased siRNA concentrations we did not see a concomitant increase in target gene knockdown (trials TOR-3 and TOR-4). However, the higher siRNA concentration of 1000nM achieved efficient knockdown in the honeybee, where a 30% reduction in the expression of *dnmt3* was observed using aerosolized siRNA-nanoparticles (Li-Byarlay et al., 2013). A possible explanation for this variation may be that, at higher concentrations the charge of the siRNA-nanoparticle complex is altered, so that cellular uptake into aphid cells is disrupted. Both this current study and the honeybee study (Li-Byarlay et al., 2013) used the same nanoparticle emulsion, which utilizes the lipid–raft transport trafficking into the cell and requires a net positive charge (Kaneda et al., 2010). This positive charge is achieved when the siRNAs bind to the nanoparticle, however, different siRNA concentrations may result in different net charges of the complex so that the rate of uptake is negatively impacted, in different tissue types (He et al. 2010). The *tor* carotenoid gene is responsible for the pink pigmentation in *A. pisum* through torulene production (Moran & Jarvik, 2010). Though we were able to measure gene knockdown in the TOR-1 and

TOR-2 trials we did not observe a color change in aphid pigmentation, compared to the corresponding trial's control treatments. However, even if *tor* is knocked-down a change in pigmentation may not occur because carotenoid proteins are extremely stable in many organisms after their production (Yahia et al., 2009).

We were able to observe successful knockdown of *bcat* in *A. glycines,* but not in *S. graminum* or *A. pisum.* The aphid gene *bcat,* is hypothesized to be important in the production of branched-chain amino acids in the aphid by utilizing metabolic intermediates produced by the aphid symbiont's (*Buchnera*) branched-chain amino acid pathways (Wilson et al. 2010; Hansen & Moran, 2011; Poliakov et al. 2011). Therefore, by targeting this aphid gene in bacteriocytes (specialized aphid cells that contain the aphid symbiont *Buchnera*), we would be disrupting a key gene that is involved in complementing *Buchnera's* essential amino acid pathways. We hypothesize that when *bcat* is successfully knocked-down in aphids, a reduction in the biosynthesis of these essential amino acids (Leu, Val, and Ile) will result. In consequence, the predicted reduction of these essential amino acids will result in aphid starvation reducing aphid body mass compared to wild-type. In the *A. glycines* BCAT-200 trials, we were able to observe not only successful gene knockdown of *bcat* in the aerosolized siRNA and nanoparticle treatment but also a concomitant change in aphid mass compared to the control, which matched our predictions (Fig 5). It is important to note that this effect on aphid mass can also be associated with the disruption of branched-chain amino acid degradation in other body tissues as well. We were able to determine that any changes in aphid mass was not associated with nanoparticles or water, since our water only, water + nanoparticles (Fig. S2A), and our control siRNA + nanoparticle treatment did not result in a body mass reduction.

The successful gene knockdown observed in this study may be due in part to the fact that nanoparticles bound to siRNA molecules tend to be more stable and more likely to undergo cellular

uptake into cells (Ross et al., 2015). However, we found that dsRNA, which are double stranded like siRNAs, do not degrade rapidly in the pea aphid *in vivo* using either microinjection or aerosolization, with or without nanoparticles as a carrier (Figs. 5.1 & 5.2). These results are in contrast to Christiaens et al. (2014). Our dsRNA degradation results may differ from Christiaens et al. (2014) because we used a different pea aphid strain, GFP plasmid, and/or aphid life stage. Regardless, no rapid degradation of dsRNA was observed in our *A. pisum* strain via either microinjection or aerosolized dsRNA-nanoparticle delivery. As such, we do not expect RNA degradation of dsRNA to impact RNAi in this new mode of delivery using the aerosolization of nanoparticles. In sum, our results suggest that *in vivo* siRNA degradation may not be an important factor affecting the success of RNAi for this delivery method or aphid line. In general results from our study indicate that aerosolized siRNA-nanoparticle complexes may be more efficient at undergoing cellular uptake compared to siRNA molecules alone, suggesting that nanoparticles are key for increasing the efficiency of RNAi using this delivery technique.

The collection time after the siRNA treatment seems to play an important role on whether or not gene knockdown is observed. Previous RNAi studies conducted on pea aphids using microinjection or ingestion observed gene knockdown that ranged from 27-46% on day five or seven post-RNAi treatment, regardless of the delivery method (Mutti et al., 2006; Jaubert-Possamai et al., 2007; Shakesby et al., 2009; Mao et al., 2012). We observed a similar trend in *A. glycines* in which no gene knockdown was observed when aphids were screened earlier than five days post-siRNA exposure (Fig. S5.3).

When compared to *A. pisum, S. graminum* and *A. glycines* are smaller in body size and mass. For example, *S. graminum* is ~2 times smaller by body length and ~7 times smaller in mass, and *A. glycines* is ~3 times smaller in body length and ~ 27 times smaller in mass compared to *A. pisum* (Blackman & Eastop, 2006; Figs 5.5C-D,5.6C-D, S5.2A). *Aphis glycines,* was the smallest

121

aphid that we tested in this study and its small size may have increased the efficacy of gene knockdown given our selected experimental concentrations and volumes. For example, it may be possible that we were able to expose *A. glycines'* internal tissues to a greater volume of siRNA-nanoparticles because this species has a larger surface-area-to-volume ratio when compared to *A. pisum* and *S. graminum*.

Overall, this new mode of aerosolized siRNA-nanoparticle delivery is a promising high-throughput and non-invasive RNAi technique to optimize further and investigate gene functions in aphids and other insect systems. Once optimized, this method marks an improvement over traditional RNAi delivery methods (microinjection and artificial diets) because it can be widely applied in systems where the traditional RNAi delivery methods are not feasible or easily applied. For example, to implement the feeding technique an artificial diet must be created, however it may not be available for all insect systems. The development of an artificial diet is not always easy, and for the soybean aphid a successful artificial diet has not been fully developed (R. Bansal *personal communication*). Alternatively, a transgenic plant must be made, which is time consuming and much easier to implement in some model plants systems compared to others. Microinjection though successful for various insect species, has the disadvantage of being highly invasive and difficult for small insects such as the soybean aphid. It also has the disadvantage of being a relatively slow technique and requires expensive, specialized equipment and experience. Via aerosolization, ~20 aphids can be exposed to a RNAi treatment in 5 minutes and only requires a nebulizer.

In this study, we observed modest to moderate levels of gene knockdown; however, it is important to note that these levels were for whole body samples. The efficiency of gene knockdown may be higher in specific tissue types using this technique, as was demonstrated by Li-Byarlay et al. (2013). This highlights the need for future studies to determine if this new method of gene

silencing in aphids can be optimized further, and which aphid cells and tissues successfully uptake siRNA-nanoparticle complexes. Many aphid species are major economic pests and in the long term this technology may potentially be used in their management. Results generated from such studies will ultimately augment research productivity and progress in the exploding field of functional genomics.

Experimental Procedures

Aphid Colonies

    The *A. pisum* strains 5A and LSR1, were reared on broad bean (*Vicia faba*), *S. graminum* was reared on barley (*Hordeum vulgare L*), and *A. glycines* biotype 1 (Kim et al., 2008; Cooper et al., 2015) was reared on susceptible soybean seedlings (*Glycine max*, cultivar Williams82). *Acyrthosiphon pisum* and *S. graminum* aphid colonies were maintained at 20ºC with a 16h light: 8h dark cycle (Hansen & Moran, 2012). *Aphis glycines* was maintained at 24°C with a 14h light: 10h dark cycle which was similar to the condition used by Bansal et al. (2014). Before each experiment, even-age cohorts were established. To establish these cohorts, 25-30 adults were placed on a plant; within 24 hours all adults were removed from the plant leaving the 1-day-old nymphs to develop on the plant.

Green fluorescent protein (GFP) dsRNA degradation trials

    Degradation of dsRNA *in vivo* was tested using two different dsGFP delivery techniques: *1-* direct microinjection, and *2-* a non-invasive aerosolization of siRNA-nanoparticle complexes using a nebulizer/compressor (Probasics, PMI, Marlboro, NJ, USA). For both the direct microinjection and aerosolization assays the degradation process was measured at five time points after the dsGFP delivery (30 minutes, 1 hour, 2 hours, 5 hours and 24 hours) (Table S5.2). For both techniques, 30 six-day old LSR1 *A. pisum* aphids (4th instar- early adult) were exposed to each

treatment. At each time point, six individuals were collected, snap-frozen at -80°C on dry ice, and preserved in RNA*later*-ICE (Ambion Life Technologies Corporation, Grand Island, NY, USA) (see Fig. 5.6 A for experimental design).

dsRNA preparation

dsRNA was used in this experiment because unlike siRNA, it is long enough to test for degradation using reverse transcriptase-quantitative PCR (RT-qPCR). Synthesized double-stranded green fluorescent protein RNA (dsGFP) was used for degradation assays on aphids *in vivo*. First, DNA template of 424 bp of superfold green fluorescent protein (sfGFP) was amplified from the pZEMB8 plasmid that we kindly received as a gift from the C.K. Vanderpool lab in the Department of Microbiology at University of Illinois, Urbana-Champaign. Specific primers (dssfGFP-F/R), containing the T7 promotor sequence at each 5' end were used for the initial PCR reaction (Table S5.1). Double stranded RNA for sfGFP was synthesized *in vitro* using Megascript RNAi Kit (Ambion, Life Technologies Corporation, Grand Island, NY, USA) from 1 μg of amplified PCR template. To maximize dsRNA yield, the composition of the binding mix in the purification step was altered and three volumes of unpurified dsRNA were used. The concentration of dsGFP was determined using a NanoDrop Lite Spectrophotometer (Thermo-Fisher Scientific, Wilmington, DE, USA). Furthermore, dsGFP stability in the supplied elution buffer (Ambion, Life Technologies Corporation, Grand Island, NY, USA), and water was assessed after 24 hours using 1% agarose gel electrophoresis per the manufacturers' protocol (data not shown). Based on this quality check, synthesized dsGFP appeared to be of high quality, and in the correct size range.

dsRNA degradation trials: Microinjection

For the microinjection assay, ~40ng of dsGFP was administered into each individual using a Discovery V8 stereo-microscope (Zeiss, Oberkochen, Germany) with the Micro 4 MicroSyringe pump controller micro-injector (World Precision Instruments, Sarasota, FL, USA) using a 34 GA, 1.97", 30° microinjection needle (Hamilton, Reno, NV, USA).

dsRNA degradation trials: Non-invasive aerosolization of nanoparticle complexes

The nanoparticle emulsion used in this study was designed by the Wickline laboratory (Kaneda *et al.*, 2010) and kindly provided to us. The nanoparticle emulsion used was the same as the one used by Li-Byarlay *et al.* (2013), to successfully knockdown gene expression in the honeybee. Li-Byarlay *et al.* (2013) also demonstrated that the siRNA-nanoparticle complexes successfully penetrate the spiracles on the thorax and abdomen and travels through the tracheal respiratory system. For the dsRNA degradation aerosolization assays, aphids were administered with 3 mL of aerosolization mixture which contained 200pM of the nanoparticle emulsion, varying in concentrations of dsRNA and water. Treatment 1 contained 100nM dsGFP only, Treatment 2 contained 100nM dsGFP + 200pM nanoparticles, and trial Treatment 3 contained 500nM +200pM nanoparticles.

<u>*siRNA-nanoparticle emulsion aerosolization trials*</u>

For all experiments, 20- 30 aphids were placed in an enclosed container and a 3-mL siRNA solution was aerosolized using a nebulizer/compressor (Probasics, PMI, Marlboro, NJ, USA). The 3mL solution contained 200pM of the nanoparticle emulsion, varying concentrations of siRNA and water. During each experiment, there were three treatments: *1*-target siRNA and nanoparticle emulsion (siRNA+nanoparticles), *2*-control siRNA and nanoparticle emulsion (control

siRNA+nanoparticles) and *3*- siRNA only. The control siRNA was a randomized siRNA sequence that had no predicted target within the aphid. The concentration of siRNA varied between trials, however, the concentration of nanoparticles remained constant (200pM) for all trial. This was the same concentration that was used in the degradation trials. All siRNAs were ordered through Sigma (St. Louis, MO, USA). Five day old aphids were exposed to the aerosolized siRNA mixture for 5 minutes; after which aphids were moved back to their host plant until they were collected. Aphids were then collected five days after spray treatment, snap-frozen on dry ice and preserved in RNA*later*-ICE (Ambion Life Technologies Corporation, Grand Island, NY, USA) (see Fig.5.6 B for experimental design). These time points and aphid ages were selected based on preliminary experiments (data not shown) and time points used in other aphid RNAi experiments (e.g. Mutti et al., 2006; Jaubert-Possamai et al., 2007; Shakesby et al., 2009; Mao et al., 2012).

Knockdown of the tor *gene*:

A control siRNA was designed by randomizing the *tor* nucleotide sequence (Sigma, St. Louis, MO, USA). Control and target *tor* siRNA oligo sequences are given in Table S1. Four trials for the *tor* gene that vary in siRNA concentrations, hereafter referred to as (TOR-1 to TOR-4) were performed in *A. pisum* (LSR1). Each trial had a specific siRNA concentration: TOR-1: 100nM siRNA; TOR-2: 200nM siRNA; TOR-3: 500nM siRNA; and TOR-4: 1000nM siRNA. Sample sizes for each trial are detailed in Table 5.2.

Knockdown of the bcat *gene*:

For the second set of *siRNA-nanoparticle aerosolization* experiments, knockdown of *bcat* was tested on *A. pisum* (5A and LSR1), *A. glycines* and *S. graminum.* Species-specific *bcat* siRNAs were designed for each species. Control and target *bcat* siRNA oligonucleotide sequences are

presented in Table S5.1. siRNA concentrations for the *bcat* experiments were optimized based on results from the first set of experiments with *tor*. Consequently, for *bcat* knockdown we tested the concentrations of siRNA at 100nM (BCAT-100) and 200nM (BCAT-200) (Table 5.3). Aphid masses were recorded using a UMX2 microscale (Mettler Toledo, Columbus, OH, USA) (see Fig. S5.1B for an experimental schematic).

After observing greater aphid mortality in the *A. glycines* and *S graminum* trials when compared to *A.pisum,* we exposed *A. glycines* and *S graminum* aphids to two additional experimental treatments: water +200pM nanoparticles, and water only. These experiments would help us understand aerosolization treatment effects on these two species. We then recorded aphid mortality and aphid mass (Fig. S5.4). We also determined background mortality rates for these two species (Fig. S5.4B)

RNA extractions and Reverse transcriptase-quantitative PCR (RT-qPCR)

To determine if there was a difference in GFP dsRNA abundance for dsRNA degradation and target gene mRNA abundance for nanoparticle-siRNA aerosolization trials detailed above, total RNA was extracted from individual aphid whole bodies using the Qiagen RNeasy Mini kit (Qiagen, Valencia, CA, USA) with the Qiagen RNase-Free DNase treatment (Qiagen, Valencia, CA, USA). Aphids that displayed low quality/quantity RNA (determined by spectrometry -NanoDrop lite, Thermo-Fisher Scientific, Wilmington, DE, USA), which prevented cDNA synthesis, were removed from the analysis.

For cDNA synthesis, the All-in-One cDNA Synthesis SuperMix kit (Biotool, Houston, TX, USA) was used with 200ng of total RNA for each sample following the manufacture's protocols. Reverse transcriptase-quantitative PCR was conducted on each sample using the SYBR Fast Universal qPCR reagents (KAPA Biosystems, Woburn, MA, USA) and iTaq Universal SYBR

127

Green Supermix (Bio-Rad, Hercules, CA, USA) on the BioRad CFX96 Real-Time PCR Detection System (Hercules, CA, USA). No aphid samples were pooled. All qPCR experiments were performed on individual aphids.

For the GFP dsRNA degradation trials, the presence of intact, non-degraded dsGFP was detected using RT-qPCR with the primers qsfGFP-F/R 9 (Table S5.1), which amplifies a 102 bp region of our 424 bp dsRNA target. Untreated aphids served as negative controls for RT-qPCR detection of dsGFP, because dsGFP is not endogenously expressed in the aphid. For control, non-GFP treated aphids, expression of the aphid housekeeping *Elongation factor 1-alpha* (*EF-1α*) was measured with RT-qPCR to verify RNA quality and quantity of aphid extractions.

For the siRNA-nanoparticle emulsion aerosolization trials, gene expression values were calculated using the standard curve method for relative quantification (Bookout et al.,2006) and normalized to the housekeeping gene *EF-1α* (Table S5.1) for *A. pisum* (Dunbar et al., 2007) and *S. graminum. A. glycines* gene expression was normalized to the housekeeping gene *RPS9*, ribosomal protein S9 (Bansal et al., 2012; Table S5.1).

Statistical Analysis

For the GFP dsRNA degradation trials analysis of variance (ANOVA) was used to determine if there was a statistically significant difference in normalized expression between treatments (Bookout et al., 2006). Post-hoc multiple comparison analyses of normalized expression values were conducted between treatments using Least Significant Difference (LSD) Tests. The statistical program R (R Core Team; 2016) was used for all statistical analyses and an α of 0.05 or less was chosen *a priori* as a significance threshold for treatment differences.

For the nanoparticle spray trials, statistical analysis was performed using unpaired, one-tailed Welch's T-test comparing each experimental treatment (siRNA with nanoparticles and

siRNA alone) to the control treatment (control siRNA with nanoparticles). Note, trials in which we observed a significant gene knockdown at an α of 0.05 or less and a corresponding change in phenotype were repeated. For the experiments in which we tested the effect of aerosolization on aphid mortality in *S. gramminum* and *A. glycines;* a two-proportion z-test was used to determine if there are differences in percent aphid mortality. A z-test was used because we were making comparisons using proportion data.

References

Bansal, R., Mamidala, P., Mian, M.A.R., *et al*. (2012) Validation of Reference Genes for Gene Expression Studies in *Aphis glycines* (Hemiptera: Aphididae). *J. Econ. Entomol.*, **105** (4), 1432–1438.

Bansal, R., Mian, M.A.R., Mittapalli, O., and Michel, A.P. (2014) RNA-Seq reveals a xenobiotic stress response in the soybean aphid, *Aphis glycines*, when fed aphid-resistant soybean. *BMC Genomics*, **15** (1), 972.

Bansal, R., and Michel, A. (2013) Core RNAi Machinery and Sid1, a Component for Systemic RNAi, in the Hemipteran Insect, *Aphis glycines*. *Int. J. Mol. Sci.*, **14** (2), 3786–3801.

Bellés, X. (2010). Beyond Drosophila: RNAi *in vivo* and functional genomics in insects. *Annu. Rev. Entomol.*, **55**(1), 111–128.

Blackman, R. L., & Eastop, V. F. (2006). *Aphids on the World's Herbaceous Plants and Shrubs*. Wiley.

Bookout, A.L., Cummins, C.L., Mangelsdorf, D.J., et al. (2006) High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol.*, **Chapter 15**, Unit 15.8.

Christiaens, O., Swevers, L., and Smagghe, G. (2014) DsRNA degradation in the pea aphid (*Acyrthosiphon pisum*) associated with lack of response in RNAi feeding and injection assay. *Peptides*, **53**, 307–14.

Coleman, A.D., Wouters, R.H.M., Mugford, S.T., and Hogenhout, S.A. (2015) Persistence and transgenerational effect of plant-mediated RNAi in aphids. *J. Exp. Bot.*, **66** (2), 541–8.

Cooper, S.G., Concibido, V., Estes, R., et al. (2015) Geographic Distribution of Soybean Aphid Biotypes in the United States and Canada during 2008–2010. *Crop Sci.*, **55** (6), 2598.

Das, S., Debnath, N., Cui, Y., *et al*. (2015) Chitosan, Carbon Quantum Dot, and Silica Nanoparticle Mediated dsRNA Delivery for Gene Silencing in *Aedes aegypti*: A Comparative Analysis. *ACS Appl. Mater. Interfaces*, **7** (35), 19530–5.

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

Guo, H., Song, X., Wang, G., *et al.* (2014) Plant-generated artificial small RNAs mediated aphid resistance. *PLoS One*, **9** (5), e97410.

Hansen, A.K., and Moran, N.A. (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. U. S. A.*, **108** (7), 2849–54.

He, C., Hu, Y., Yin, L., *et al.* (2010). Effects of particle size and surface charge on cellular uptake and biodistribution of polymeric nanoparticles. *Biomaterials*, **31**(13), 3657–3666.

Jaubert-Possamai, S., Le Trionnaire, G., Bonhomme, J., *et al.* (2007) Gene knockdown by RNAi in the pea aphid *Acyrthosiphon pisum*. *BMC Biotechnol.*, **7** (1), 63.

Jaubert-Possamai, S., Rispe, C., Tanguy, S., *et al.* (2010) Expansion of the miRNA pathway in the hemipteran insect *Acyrthosiphon pisum*. *Mol. Biol. Evol.*, **27** (5), 979–87.

Kaneda, M.M., Sasaki, Y., Lanza, G.M., *et al.* (2010) Mechanisms of nucleotide trafficking during siRNA delivery to endothelial cells using perfluorocarbon nanoemulsions. *Biomaterials*, **31** (11), 3079–3086.

Kim, K.-S., Hill, C.B., Hartman, G.L*., et al*. (2008) Discovery of Soybean Aphid Biotypes. *Crop Sci.*, **48** (3), 923.

Li-Byarlay, H., Li, Y., Stroud, H., *et al.* (2013) RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U. S. A.*, **110** (31), 12750–5.

Mao, J., and Zeng, F. (2012) Feeding-based RNA interference of a gap gene is lethal to the pea aphid, *Acyrthosiphon pisum*. *PLoS One*, **7** (11), e48718.

McCornack, B. P., Ragsdale, D. W., & Venette, R. C. (2004). Demography of soybean aphid (Homoptera: Aphididae) at summer temperatures. *J. Econ. Entomol.*, **97**(3), 854–61.

Moran, N.A., and Jarvik, T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, **328** (5978), 624–7.

Mutti, N.S., Park, Y., Reese, J.C., and Reeck, G.R. (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. *J. Insect Sci.*, **6**, 1–7.

NCBI, 2017. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/genome/browse/. Accessed -1-12-17.

Nováková, E., and Moran, N.A. (2012) Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Mol. Biol. Evol.*, **29** (1), 313–23.

Ortiz-Rivas, B., Jaubert-Possamai, S., Tanguy, S., *et al.* (2012) Evolutionary study of duplications of the miRNA machinery in aphids associated with striking rate acceleration and changes in expression profiles. *BMC Evol. Biol.*, **12** (1), 216.

Pitino, M., Coleman, A.D., Maffei, M.E., *et al.* (2011) Silencing of aphid genes by dsRNA feeding from plants. *PLoS One*, **6** (10), e25709.

Poliakov, A., Russell, C. W., Ponnala, L., *et al.* (2011) Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Mol. Cell Proteomics*, **10**(6), M110.007039.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ross, K.A., Brenza, T.M., Binnebose, A.M., *et al.* (2015) Nano-enabled delivery of diverse payloads across complex biological barriers. *J. Control. Release*, **219**, 548–59.

Scott, J.G., Michel, K., Bartholomay, L.C., *et al.* (2013) Towards the elements of successful insect RNAi. *J. Insect Physiol.*, **59** (12), 1212–21.

Shabalina, S. A., & Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. *Trends in Ecol. Evol.*, **23**(10), 578–87.

Shakesby, A.J., Wallace, I.S., Isaacs, H. V, *et al.* (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem. Mol. Biol.*, **39** (1), 1–10.

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.

Wilson, A.C.C., Ashton, P.D., Calevro, F., *et al.* (2010) Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.*, **19 Suppl 2**, 249–58.

Wynant, N., Santos, D., and Vanden Broeck, J. (2014) Biological mechanisms determining the success of RNA interference in insects. *Int. Rev. Cell Mol. Biol.*, **312**, 139–67.

Yahia EM, Ornelas-Paz J de J Chemistry, Stability, and Biological Actions of Carotenoids. Fruit and Vegetable Phytochemicals. Oxford, UK: Wiley-Blackwell.(2009); pp. 177–222. doi:10.1002/9780813809397.ch7.

Yu, N., Christiaens, O., Liu, J., *et al.* (2013) Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Sci.*, **20** (1), 4–14.

Zhang, X., Zhang, J., and Zhu, K.Y. (2010) Chitosan/double-stranded RNA nanoparticle-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Mol. Biol.*, **19** (5), 683–93.

Tables and Figures:

Table 5.1. Pair-wise comparison of time points from Treatment 1 dsRNA degradation aerosolization assays. Significantly different treatments are **bolded**

	T1 (0.5h)	T2 (1h)	T3 (3h)	T4 (5h)	T5 (24h)
T1 (0.5h)	-	0.539	0.133	0.425	0.063
T2 (1h)	-	-	0.344	0.165	**0.022**
T3 (3h)	-	-	-	**0.029**	**0.004**
T4 (5h)	-	-	-	-	0.207
T5 (24h)	-	-	-	-	-

[*]*A. pisum* aphids in trial D1 were aerosolized with 100 nM of dsGFP without nanoparticles

[1]Least Significant Difference (LSD) tests were conducted on normalized expression values for GFP standardized to an aphid housekeeping gene

Table 5.2: Aerosolized siRNA –nanoparticles trials and treatments testing the knockdown of the *tor* gene in *A. pisum* (LSR1)

Trial	NP concent ration	siRNA concentration	Aphid age during spray treatment	Spray treatment (sample size for RT-qPCR)	Sample collection time after spray	Cuticle color	Aphid species (Line)
TOR 1	200pM	100nM	5 day	Control- NP & ctrl siRNA (N=6)	120 hr	pink	*A.pisum* (LSR1)
				NP + *tor*-siRNA (N=6)		pink	*A.pisum* (LSR1)
				tor- siRNA (N=6)		pink	*A.pisum* (LSR1)
TOR 2	200pM	200nM	5 day	Control- NP + ctrl siRNA (N=6)	120 hr	pink	*A.pisum* (LSR1)
				NP + *tor*-siRNA (N=6)		pink	*A.pisum* (LSR1)
				tor- siRNA (N=6)		pink	*A.pisum* (LSR1)
TOR 3	200pM	500nM	5 day	Control- NP + ctrl siRNA (N=6)	120 hr	pink	*A.pisum* (LSR1)
				NP + *tor*-siRNA (N=6)		pink	*A.pisum* (LSR1)
				tor- siRNA (N=6)		pink	*A.pisum* (LSR1)
TOR 4	200pM	1000nM	5 day	Control- NP + ctrl siRNA (N=6)		pink	*A.pisum* (LSR1)
				NP + *tor*-siRNA (N=6)	120 hr	pink	*A.pisum* (LSR1)
				tor- siRNA (N=6)		pink	*A.pisum* (LSR1)

Table 5.3: Aerosolized siRNA-nanoparticles trials testing the knockdown of the *bcat* gene in three aphid species

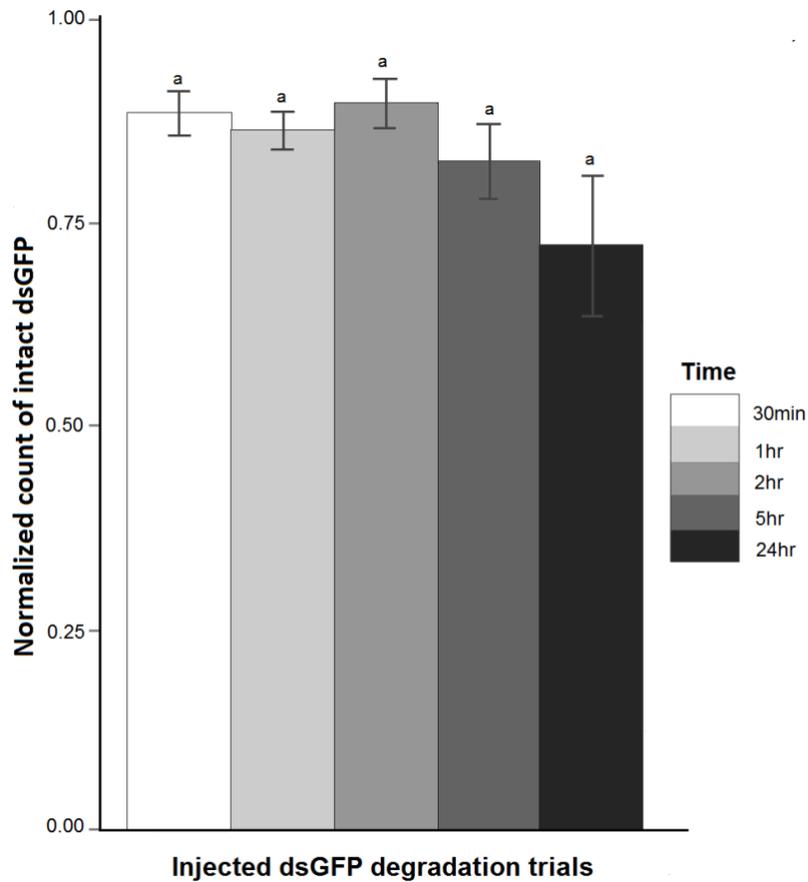| Trial	NP concentration	siRNA concentration	Aphid age during spray treatment	Spray treatment (sample size)	Sample collection time after spray	Aphid species (Line)
BCAT 1	200pM	100nM	5 day	Control- NP + ctrl siRNA (N=6)	120 hr	*A.pisum* (5A, LSR1), *A. glycines, S. graminum*
				NP + *bcat*-siRNA (N=6)		*A.pisum* (5A, LSR1), *A. glycines, S. graminum*
				bcat- siRNA (N=6)		*A.pisum* (5A, LSR1), *A. glycines, S. graminum*
BCAT 2	200pM	200nM	5 day	Control- NP + ctrl siRNA (N=6)	120 hr	*A.pisum* (5A, LSR1), *A. glycines, S. graminum*
				NP + *bcat*-siRNA (N=6)		*A.pisum* (5A, LSR1), *A. glycines, S. graminum*
				bcat- siRNA (N=6)		*A.pisum* (5A, LSR1), *A. glycines, S. graminum*

Figure 5.1:

Normalized count of intact double-stranded green fluorescent protein (dsGFP) RNA observed in microinjected double-stranded RNA degradation trials. The counts of GFP transcripts were standardized to an aphid housekeeping gene (elongation factor 1-alpha). Error bars indicate ± 2 SD from the mean. Bars with different letters above represent significant treatment differences within a trial at $P < 0.05$ based on post-hoc comparisons (least significant difference tests). Non-GFP treated aphids served as negative controls and dsGFP RNA was not present as expected (bars not shown).
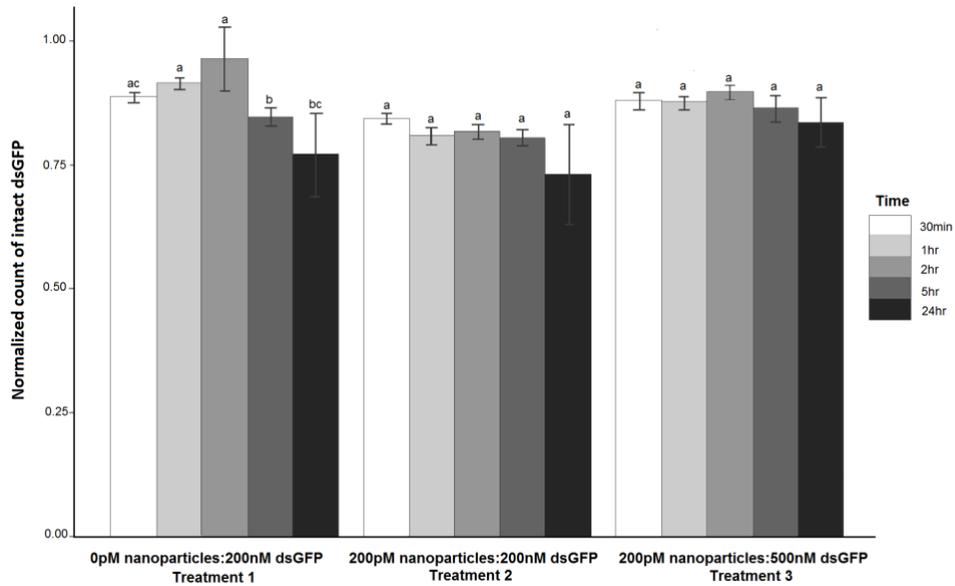
Figure 5.2:

Normalized count of intact double-stranded green fluorescent protein (dsGFP) RNA observed in aerosolization double-stranded RNA degradation trials. The counts of GFP transcripts were standardized to an aphid housekeeping gene (elongation factor 1-alpha). Error bars indicate ± 2 SD from the mean. Bars with different letters above represent significant treatment differences within a trial at $P < 0.05$ based on post-hoc comparisons (least significant difference tests). Non-GFP treated aphids served as negative controls and dsGFP RNA was not present as expected (bars not shown).
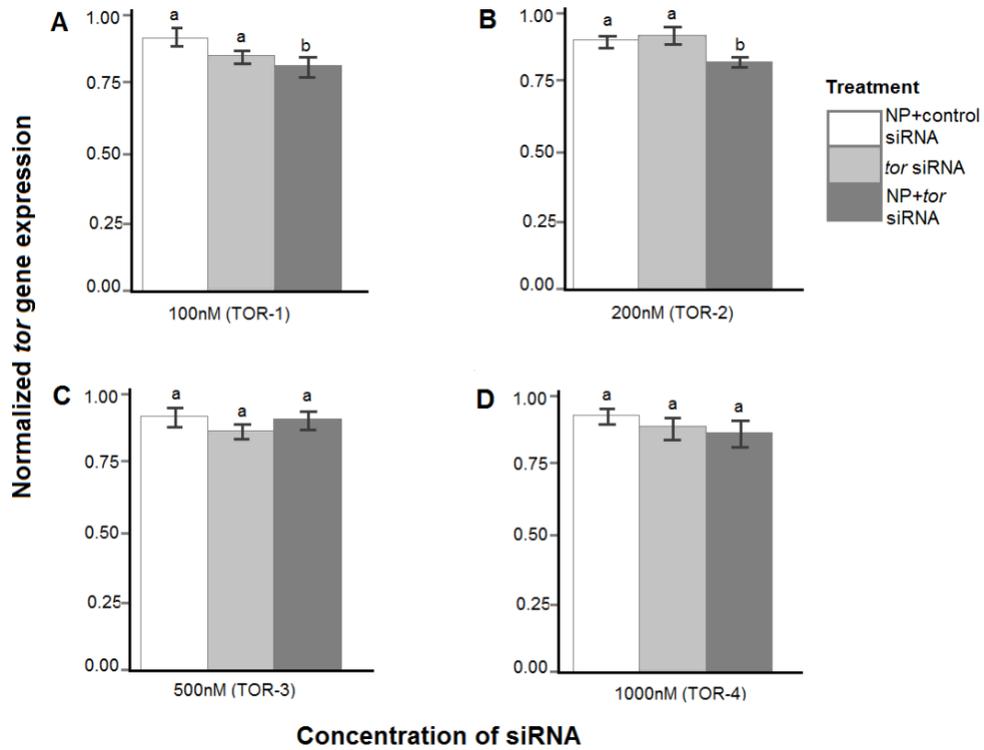
Figure 5.3:

Normalized carotene dehydrogenase (tor) gene expression levels for the aerosolized siRNA-nanoparticle (siRNA-NP) trials (TOR-1, TOR-2, TOR-3 and TOR-4) for Acyrthosiphon pisum (LSR1). All expression values for tor were standardized to an aphid housekeeping gene (elongation factor 1-alpha). See Table 1 for sample sizes. Error bars indicate ± SEM. Bars with different letters above represent significant differences within a trial at $P < 0.05$. Controls for all trials were aerosolized NP–control siRNA treatment
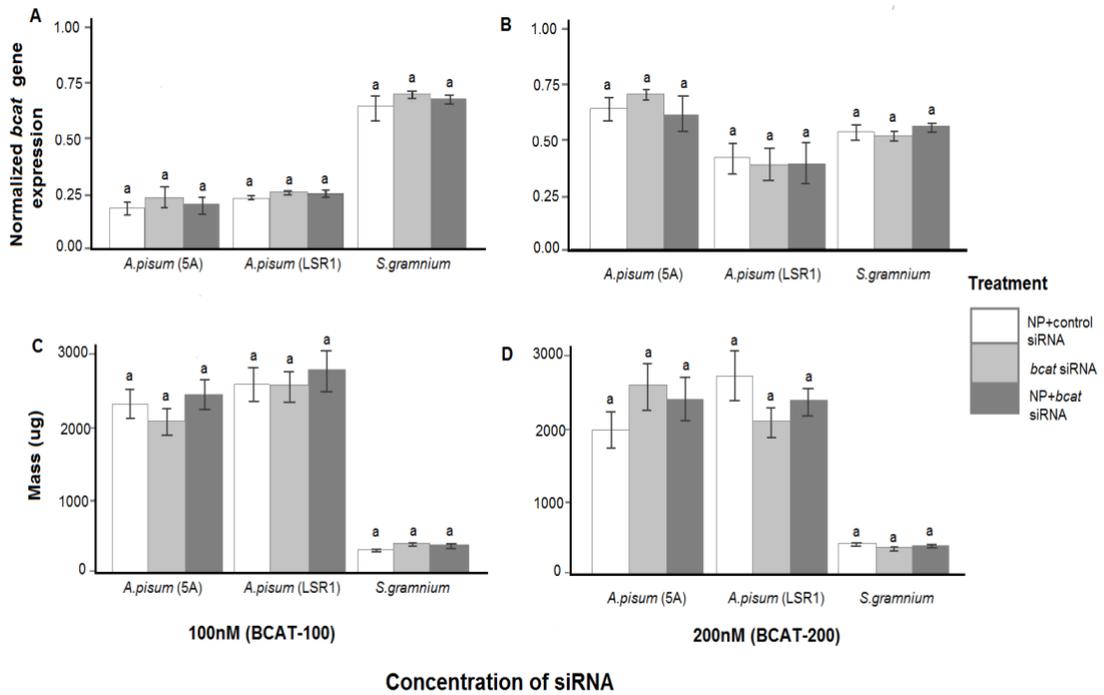
Figure 5.4:

Normalized branched-chain amino acid transaminase (bcat) gene expression levels and aphid mass for the aerosolized siRNA-nanoparticle (siRNA-NP) trials (BCAT-100 and BCAT-200) for Acyrthosiphon pisum (5A, LSR1) and Schizaphis graminum. See Table 2 for gene expression experiment sample sizes and Tables S4 and S5 for aphid mass experiment sample sizes. All expression values for bcat were standardized to an aphid housekeeping gene (elongation factor 1-alpha). Error bars indicate $\pm$ SEM from the mean. Bars with different letters above represent significant differences within a trial at $P < 0.05$.
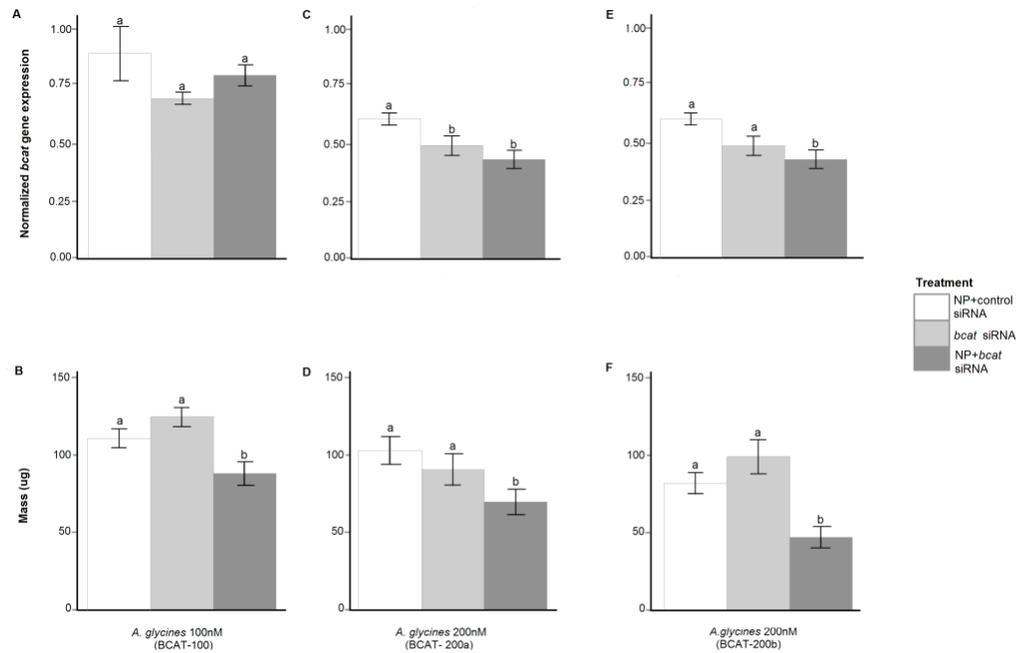
Figure 5.5:

Normalized branched-chain amino acid transaminase (bcat) gene expression levels and aphid mass for the aerosolized siRNA- nanoparticle trials for *Ap. glycines*. All expression values for bcat were standardized to the housekeeping gene ribosomal protein S9 (RSP9). BCAT- 200a and BCAT-200b represent two independent trials: See Table 2 for gene expression experiment sample sizes and Tables S4 and S5 for aphid mass experiment sample sizes. Error bars indicate ± SEM. Bars with different letters above represent significant differences within a trial at $P < 0.05$.
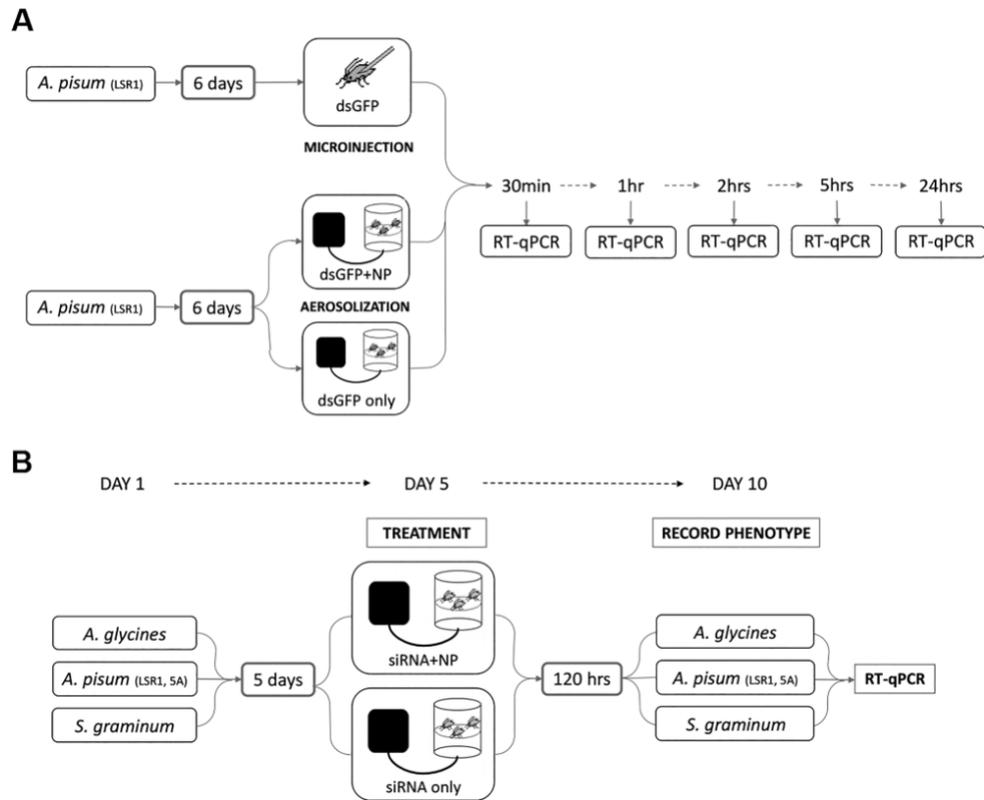
Figure 5.6:

Overview of double-stranded RNA (dsRNA) degradation and short interfering RNA (siRNA)–nanoparticle emulsion aerosolization experiment methodology. (A) dsRNA degradation experiments: 6-day-old aphids were exposed to dsRNA either via microinjection or aerosolization. dsRNA degradation was calculated via reverse transcriptase-quantitative PCR (RT-qPCR) at five time points after the treatment. (B) siRNA–nanoparticle emulsion aerosolization experiments: 5-day-old aphids were exposed to siRNAs with or without the nanoparticle emulsion. The control treatment consisted of a randomized siRNAs + nanoparticles. After 120 h any change in phenotype was recorded (colour or mass) and gene expression levels were determined via RT-qPCR. Abbreviations: *A. glycines*: *Aphis glycines*; *A. pisum: Acyrthosiphon pisum;* dsGFP, double stranded green fluorescent protein RNA; NP, nanoparticle; *S. graminum: Schizaphis gramim*

Chapter 6: Conclusion

Bacterial symbionts have significantly shaped the evolution of many animal lineages (McFall-Ngai et al., 2013). Within the most diverse group of animals, class Insecta, the acquisition of bacterial symbionts has often resulted in adaptive radiations of insect species into ecological niches that were otherwise inaccessible (Hansen & Moran, 2014). Within hemipterans, the acquisition of bacterial symbionts has helped insects utilize the nutrient-limited resource of plant sap (Hansen & Moran, 2014; Sudakaran et al., 2017). During the evolution of symbiotic relationships, obligate, host-restricted intracellular bacteria often experience dramatic genome reduction, where they lose many key regulatory genes (Moran & Bennett, 2014). It has been hypothesized that these host-restricted intracellular bacteria with reduced genomes are incapable of directly regulating the symbiosis. Their insect hosts primarily regulate cellular activities, such as shared metabolic pathways, in which both the symbiont and host contribute to the formation of key metabolites (Hansen & Moran, 2014). However, within host-restricted intracellular bacterial symbionts, small RNAs (sRNAs) are emerging as potential regulators of gene expression (Chapter 1). As such, the goal of the research presented in my dissertation was to better understand the potential functional role that symbiont expressed sRNAs have in post-transcriptional gene regulation.

The research presented in my dissertation uses the well characterized aphid-*Buchnera* nutritional symbiosis. The symbiotic relationship between aphids and *Buchnera* originated ~ 160–280 million years ago (Moran et al., 1993). Like other bacterial symbionts, *Buchnera* has experienced significant genome reduction, loosing many of the genes involved in transcriptional regulation, however it has retained most of its housekeeping genes and genes related to the production of essential amino acids (Shigenobu et al., 2000). Hansen & Degnan (2014), were the first to show that within two life-stages of *Buchnera*, differential protein expression occurred with

no concomitant changes in mRNA expression which was suggestive that post-transcriptional regulation was occurring. Hansen & Degnan (2014), also characterized 636 highly conserved sRNAs in four aphid species that diverged over 65 million years ago.

Though Hansen & Degnan (2014), were able to identify conserved sRNAs being expressed within *Buchnera*, their functionality remained in question. Within highly reduced, adenine-thymine rich genomes, such as *Buchnera*, many sRNAs – especially antisense sRNAs – can be by-products of transcriptional noise (Raghavan et al., 2012; Llorens-Rico et al., 2016). As a result, working to validate the functionality of expressed sRNAs is a vital step in understanding their biological role within the symbiont. In the first part of my dissertation I was able to find 90 differentially expressed sRNA within the life-stages that Hansen & Degnan (2014) observed differential protein expression (chapter two). I also found that within 11 operons there was some overlap between the differentially expressed sRNAs and differentially expressed proteins that Hansen & Degnan (2014) identified (chapter two). The presence of differential sRNA expression in the same life-stages in which differential protein expression was observed suggests that at least some of the expressed sRNAs identified are functional. I also provided direct experimental evidence that *Buchnera* sRNAs can be functional by heterologously expressing the the antisense sRNA *carB,* in *Escherichia coli.*

Buchnera is a nutritional symbiont which synthesizes essential amino acids which are deplete in plant sap (Shigenobu et al., 2000; Douglas, 2006). Since I was able to demonstrate that *Buchnera* may use sRNAs to regulate gene expression independently from its host during development, I wanted to understand if this type of regulation occurred with changes in aphid diet. Chapter three of my dissertation provides insight into the potential adaptive role that sRNA regulation may have in the aphid-*Buchnera* symbiosis. I found that *Buchnera* sRNA expression not only changes with life-stage as was demonstrated in chapter two, but also with host-plant diet. Many of the sRNAs identified were conserved and, most importantly, their expression patterns were

congruent with previous work (Hansen & Degnan, 2014; chapter 2). Overall, the data from chapters two and three support the overarching hypothesis that *Buchnera* sRNAs have an important role in symbiont gene regulation.

The field of sRNA research in small genomes is still quite young, as such sRNA expression has only been characterized in a handful of beneficial bacterial symbionts (chapter 1). To increase our understanding of the evolution of sRNA gene regulation in highly reduced symbiont genomes, comparative experiments across different taxa are necessary. As such, chapter four of my dissertation characterized the sRNA expression profile of the psyllid nutritional symbionts *Candidatus* Carsonella ruddii (hereafter referred to as *Carsonella*), one of the tiniest insect nutritional symbiont genomes being on average 166kb large with ~216 genes (Moran and Bennett, 2014; NCBI genomes, 2019), in two psyllid species, the tomato psyllid *Bactericera cockerelli* and the Asian citrus psyllid *Diaphorina citri*. The psyllid *D. citri,* also houses a second intracellular symbiont, *Candidatus* Profftella armature a defensive symbiont that produces a polyketide toxin (Nakabachi et al., 2013). Overall, both symbionts were found to express sRNAs (chapter 4). Four antisense sRNAs were found to be conserved within *Carsonella*. I also demonstrated that within *Carsonella- B. cockerelli* sRNAs are differentially expressed between psyllid life-stages. Notably, within *Buchnera,* conserved antisense sRNAs have been found within the genes *clpX, carB* and *prfA,* the same genes that conserved *Carsonella* sRNAs were found. If these sRNAs are functional, this suggests sRNA dependent gene regulation mechanisms have independently evolved within these important genes for the maintaining the nutritional symbioses.

When working with non-model, unculturable systems such as the aphid-*Buchnera* system*,* there is a lack of functional genomic tools adapted for working in these challenging systems. In chapter five, I sought to develop an efficient RNAi delivery system that could work in aphids. Using three aphid species: the pea aphid, *Acyrthosiphon pisum,* the soybean aphid, *Aphis glycines*,

and the greenbug, *Schizaphis graminum*, the effectiveness of aerosolizing siRNA-nanoparticles as a RNAi delivery system was determined. Overall, aerosolizing nanoparticles with siRNAs is a viable, high-throughput method of gene-knockdown that can be used in functional genomic studies in aphids. Though only modest levels of gene knockdown was observed in the pea aphid, significant levels of gene-knockdown were observed in the *A. glycines* and was accompanied with an expected phenotype. Moving forward, this technique can be used to knock down key aphid genes predicted to be important in regulating the symbiotic relationship with *Buchnera*. Not only will these experiments be important in understanding the role that these predicted aphid genes have, but they can also provide insight into how *Buchnera,* responds to changes in its aphid host.

Across all domains of life, our understanding of the prevalence and importance of sRNA regulation is rapidly increasing. The research in this dissertation has highlighted that within the reduced genomes of insect bacterial symbionts that have lost many of their regulatory proteins, it is likely that sRNAs have a regulatory role. A major challenge in the field of bacterial symbioses is the fact that many of the systems are non-model, unculturable, and difficult to manipulate. However, as was shown in chapters two and five, as experimental techniques are adapted and modified to work in non-model systems, our understanding of how symbiotic relationships are regulated will increase.

References:

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci.* 112, 201421388.

Hansen AK, and Degnan PH. 2014. Widespread expression of conserved small RNAs in small symbiont genomes. *ISME J.* 8:2490–502.

Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* 23:1473–96.

Kim D, Thairu MW, Hansen AK. 2016. Novel Insights into Insect-Microbe Interactions-Role of Epigenomics and Small RNAs. *Front. Plant Sci.* 7:1164.

Llorens-Rico V, et al. 2016. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv.* 2: e1501363–e1501363.

McFall-Ngai, M, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U. S. A.* 110:3229–36.

Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol* 2014; 68:195–215.

Moran NA, Munson MA, Baumann, P, Ishikawa, H. 1993. A Molecular Clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. B-Biological Sci.* 253:167–171

Nakabachi A. et al. 2013. Defensive bacteriome symbiont with a drastically reduced genome. Curr. Biol. 23:1478–1484.

NCBI Genomes 2019: Candidatus Carsonella ruddii
https://www.ncbi.nlm.nih.gov/genome/genomes/1470

Raghavan R, Sloan DB, Ochman H. 2012. Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio* 3.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp. APS*. *Nature* 407, 81–6.

Sudakaran S, Kost C, Kaltenpoth M. 2017. Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends in Microbiology* 25:375–390.

Appendix: List of supplementary materials

Chapter 2:
2.0 Supplemental Information

2.1 Supplemental Tables: 2.1-2.10

*Chapter 3***:**
3.0. Supplemental Results

3.1 Supplemental Figure 1:

3.2 Supplemental Tables: 3.1-3.20

Chapter 4:
4.0 Supplemental Tables: 4.1-4.3

Chapter 5:
5.0 Supplemental Tables and Figures