UNIVERSITY OF CALIFORNIA, MERCED

Coevolution of leafhoppers and their microbial symbionts

by

Yumary M. Vasquez

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

Doctor of Philosophy

in

Quantitative and Systems Biology

Committee in Charge:
Professor Gordon Bennett, Research Advisor
Professor Emily Jane McTavish, Committee Chair
Professor Carolin Frank

Signature Page

The Dissertation of Yumary M. Vasquez, is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Gordon M. Bennett Ph.D.
Research Advisor

Emily Jane McTavish Ph.D.
Committee Chair

Carolin Frank Ph.D.

University of California, Merced

2023

Dedication

To my husband, Oscar, who continues to kill the insects that terrify me. Thank you for making me a better person. I love you.

To my parents, Carlos and Marieva, for giving me this chance. Thank you for your love and sacrifices growing up. Thank you for the continued support you show me to this day and for listening to all my science talk even though you didn't have a choice. Los quiero los dos.

To my in-laws, Oscar and Adela, for the love you gave me. I'm lucky to be a part of your family.

To my sister, and my family in Venezuela, Chile, Argentina, Canada and everywhere else. I dedicate this to you all, knowing you are all there to support me has given me strength.

To my chosen family in the United States, without any of you, I would not be the person I am today. Thank you for taking care of me, for being a second set of parents (I love you Heidy & Davy), for treating me like your own daughter, and giving me a family that continues to grow.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Gordon Bennett for being my mentor and advisor for the last five years. With your help I've learned so much about science and becoming a researcher. Your support and guidance have been invaluable to me both in my personal life and career goals. Thank you for allowing me to occasionally tear up in your office, to vent about teaching, and for helping me through my over-dramatic descriptions of lab failures. Thanks for always eating anything I made, I appreciate your sweet tooth.

I would like to thank the members of my committee, Dr. Emily Jane McTavish and Dr. Carolin Frank. Thank you for your feedback throughout this journey, I am lucky to have met both of you.

I would like to thank my mentors that have not been my advisor. Thank you to Dr. Katrina Hoyer, my lab mom even though I was not in your lab. Thanks for introducing me to good wine and being someone I felt comfortable talking to. Next, thank you to Dr. Dan Edwards. Thanks for being there when I went through one of my toughest times in grad school, thanks for suggesting yoga to me and thank you for continuing to be one of my biggest supporters. Finally, I would like to thank Dr. David Gravano from the Stem Cell Instrument Foundry. Thank you David for sitting long hours on the cell sorter and talking to me about the weather.

I would like to thank everyone in my lab. Thank you to Meng and Allen for being my friends and mentors. Thank you to DiemQunyh Nguyen who contributed the lovely drawing found on pg. iv. Thanks for spending days with me in lab and joining me in the last year of my PhD. I'm happy to have met you and I'm proud of the researcher you have become. Thank you to my lab mates, Reo Maynard, Heather Stever, and Ryan Torres. Thanks for the laughs and eating my baking so I don't have to eat it all myself.

Finally, for the friends I made in grad school. Each of you had a huge impact on me. Thanks for lunches, dinners, drinks, gossip, and making my experience in Merced better.

The studies in this dissertation were funded by Dr. Gordon Bennett and the BII-INSITE. Thank you to the fellowship opportunities on campus including the Intelligent Adaptive Systems group and the USDA-UC Merced FARMERS fellowship.

Curriculum Vitae

Education

PhD, Quantitative and Systems Biology | Aug 2023
University of California, Merced (Merced, CA)

Honors BS, Biotechnology, Computer Science Minor | May 2018
California State University, San Marcos (San Marcos, CA)

Experience

Graduate Research Scientist: University of California, Merced
2018 Aug – Present

- Conducted next generation sequencing analysis including genome, transcriptome and methylome assemblies, the use of BLAST to identify genes, differential expression analyses and synteny analyses.
- Performed fluorescent-based cell sorting for the isolation and downstream analysis of bacterial-infected cells.
- Owned experimental design, data collection, analysis, and interpretation for 3 long-term decision-making next generation sequencing research projects with multiple stakeholders, resulting in production of 3 papers.
- Coded R & Unix scripts to perform statistical analysis resulting in the completion of 3 projects
- Delivered 8 seminar presentations, including a well-received presentation on comparative genomics of microbial genomes.
- Mentored diverse team of direct reports (4 Undergraduates).
- Instructed RMarkdown pipelines for data wrangling & visualization to >100 undergraduates starting from no previous programming background.

Graduate Intern: Joint Genome Institute, CA
2022 Jun – 2022 Aug

- Conducted a large-scale data science project using >1000 microbial genomes.
- Implemented Python scripts to automate frequent tasks resulting in saving hours of manual labor.
- Demonstrated high-level technical, statistical and computational skills with the use of 2 high performance computing clusters.
- Delivered 2 presentations to general and specialized audiences within the company.

Undergraduate Research Scientist: California State University, San Marcos

2016 Sept – 2018 Jul

- Applied ANOVA, linear regression and principal component analysis in R for visualization & analysis to deepen insight into parasitoid-host relationships for one paper.
- Partnered with other scientific researchers across fields resulting in 3 publications.
- Acted as lab manager for 2 years, including the use of skills such as budgeting, organization, and product ordering for the successful execution of 4 projects.

Undergraduate Research Scientist: California State University, San Marcos
2014 Dec – 2015 Dec

- Maintained mice populations for the study of Alopecia Areata and assisted graduate students in various projects that included molecular skills such as dissection, imaging, and genotyping.
- Performed dissections and identification of T-cells within mammal intestinal tissue using microscopy and fluorescent imaging with antibody tagging. This work was done during the Summer Scholars Program at CSU San Marcos.

Publications

Gossett, J., **Vasquez, Y.M.,** Bennett, G.M. and Porter, M. (2023), Genomic comparisons reveal selection pressure and functional variation between nutritional endosymbionts of cave-adapted and epigean Hawaiian planthoppers. Accepted. *Genome Biology and Evolution.*

***Vasquez, Y.M.,** and Bennett, G.M. (2022), A complex interplay of evolutionary forces continues to shape ancient co-occurring symbiont genomes. *iScience.*

Sethuraman, A., Tovar, A., Welch, W., Dettmers, R., Arce, C., Skaggs, T., Rothenberg, A., Saisho, R., Summerhays, B., Cartmill, R. Grenier, C., **Vasquez, Y.,** Vansant, H., and Obrycki, J. (2022), Genome of the parasitoid wasp Dinocampus coccinellae reveals extensive duplications, accelerated evolution, and independent origins of thelytokous parthenogeny and solitary behavior. *G3.*

Waneka, G., **Vasquez, Y.M.,** Bennett, G.M. and Sloan, D.B. (2020), Mutational pressure drives differential genome conservation in two bacterial endosymbionts of sap feeding insects. *Genome Biology and Evolution.*

Vansant, H., ***Vasquez, YM.,** Obrycki JJ., Sethuraman A. (2019), Coccinellid host morphology dictates diversity of the parasitoid wasp Dinocampus coccinellae. *Biological Control.*

Sethuraman, A., Janzen, F. J., Rubio, M. A., **Vasquez, Y.** and Obrycki, J. J. (2017), Demographic histories of three predatory lady beetles reveal complex patterns of diversity and population size change in the United States. *Insect Science.*

* = first authorship papers

Certifications/Workshops

Google Project Management: Specialization
Coursera Mar 2023
- Foundations of Project Management
- Project Initiation: Starting a Successful Project
- Project Planning: Putting it All Together
- Project Execution: Running the Project
- Agile Project Management
- Capstone: Applying Project Management in the Real World

NERSC AI for Science Bootcamp
Joint Genome Institute NERSC Aug 2022

Applied Data Science with Python Specialization
Coursera May 2022
- Introduction to Data Science in Python
- Applied Plotting, Charting & Data Representation in Python
- Applied Machine Learning in Python
- Applied Text Mining in Python
- Applied Social Network Analysis in Python

Anti-Racist Pedagogy Discussion Series Center for Engaged Teaching and Learning
U. of California, Merced
- Anti-Racist Pedagogy Fall 2020, Spring 2021

CETL Workshop Series Center for Engaged Teaching and Learning
U. of California, Merced
- Preparing to Teach in the Online Environment Summer 2020
- Institute for Evidence-based Teaching Practices Spring 2021

Instructional Internship Program Center for Engaged Teaching and Learning
U. of California, Merced. Sept 2018
- Mastering the Classroom with 1st Generation College Students series:
  - Setting Expectations Workshop
  - Classroom Challenges Workshop
  - Grading Effectively Workshop
- The TA and the IOR Roles: Working Together Workshop

Affiliations/Service

USDA-UC Merced FARMERS
U. of California, Merced 2022 Aug – Aug 2023

RadioBio
U. of California, Merced
Marketing Director: May 2022 - current
President: May 2021 - May 2022
Vice President: May 2020 - May 2021
Treasurer: May 2019 - May 2020
Member: Sept 2018 - May 2019

Quantitative and Systems Biology Representative: Graduate Student Association
U. of California, Merced Mar 2020 - May 2021
Quantitative and Systems Biology: Diversity, Equity and Inclusion committee
U. of California, Merced Sept 2020 - May 2021
Reflecting on Anti-Racist Pedagogy Virtual Dialogue Series: Participant
U. of California, Merced Sept 2020 - Dec 2020
Data Carpentry Workshop: "Helper"
U. of California, Merced Aug 2020
NSF Louis Stokes Alliance for Minority Participation Program
California State Uni.,San Marcos 2013 Aug – 2018 May

Honors/Awards

USDA-UC Merced FARMERS Fellow,
U. of California, Merced 2022
The Ford Foundation Predoctoral Fellowship Honorable Mention 2020
NSF-NRT Training Program in Intelligent Adaptive Systems Fellowship,
U. of California, Merced 2019
NSF-NRT Interdisciplinary Computational Graduate Education Fellow,
U. of California, Merced 2019
Eugene V. Cota-Robles Fellowship (declined),
U. of California, Los Angeles 2018

Talks/Posters

Entomological Society of America Joint Annual Meeting
November 2022, Vancouver, Canada
Vasquez, Y., Bennett, G.. (2022) "Comparative genomics of endosymbionts across endemic Hawaiian leafhopper (Hemiptera: Cicadellidae: Nesophrosyne) species" - Virtual Talk
Entomological Society of America Joint Annual Meeting
November 2022, Vancouver, Canada
Vasquez, Y., Bennett, G.. (2022) "Comparative genomics of endosymbionts across endemic Hawaiian leafhopper (Hemiptera: Cicadellidae: Nesophrosyne) species" - Poster
Joint Genome Institute - UC Merced Internship Seminar: Joint Genome Institute
August 2022, Berkeley, CA
Vasquez, Y., Villada, J.. (2022) "Comparative genomics of novel symbionts"
Yosemite Symbiosis Workshop
May 2022, Sierra Nevada Research Institute
Vasquez, Y., Bennett, G.. (2022) "Leveraging a Hawaiian leafhopper adaptive radiation to understand bacterial symbiont genome evolution."
Student and Early Career Researcher Symposium: Hawaiian Entomological Society
March 2022, Virtual
Vasquez, Y., Bennett, G.. (2022) "Leveraging a Hawaiian leafhopper adaptive radiation to understand bacterial symbiont genome evolution."
3rd Place Award
Quantitative and Systems Biology Retreat: U. of California, Merced
May 2021, Virtual
Vasquez, Y., Bennett, G.. (2021) "The Co-evolution of Endemic Hawaiian Leafhoppers and Their Endosymbionts"

Teaching

Data Science Discussion
U. of California, Merced — Spring 2022

Guest Presenter: Data Science
U. of California, Merced — Spring 2022

Undergraduate Research Opportunities Center (UROC) Summer Grad Mentor
U. of California, Merced — Summer 2021

Evolution Lab
U. of California, Merced — Summer 2019, Fall 2020, Spring 2021, Fall 2021

Intro to Molecular Biology Lab
U. of California, Merced — Fall 2018, Spring 2019

Guest Presenter: Evolution
U. of California, Merced — Fall 2018, Spring 2019, Fall 2020

Dissertation Abstract

Coevolution of leafhoppers and their microbial symbionts

Yumary M. Vasquez

Doctor of Philosophy in Quantitative and Systems Biology
University of California, Merced, 2023

Graduate Advisor: Dr. Gordon M. Bennett
Committee Chair: Dr. Emily Jane McTavish

Nutritional symbionts are a vital component of many organisms' survival strategies, supplying essential nutrients, metabolizing substances that are otherwise indigestible, and in some cases, conferring resistance to environmental stressors. Insects have a close and complex association with nutritional symbionts. This dissertation explores critical aspects of leafhopper symbiosis and evolution through a three-part investigation into the Hawaiian leafhopper radiation (*Nesophrosyne)* and the aster leafhopper (*Macrosteles quadrilineatus*). In Chapter 2, I focused on the symbiont genome evolution in two of the smallest known bacterial genomes, *Sulcia* and *Nasuia*, in a Hawaiian leafhopper radiation (<5 million years). The work found that multiple evolutionary processes are continually shaping this multi-partner symbiosis in a relatively short time frame. In Chapter 3, I sequenced the aster leafhopper genome, a polyphagous pest with a symbiotic association with *Sulcia* and *Nasuia.* Using the chromosomal-level sequencing of this leafhopper, along with three other leafhopper genomes with different ecologies and symbiotic associations, I explored how pest species biology and symbiotic associations shape leafhopper genomes. I found that genes related to insecticide resistance and adaptation to environmental pressures are conserved in leafhoppers, with some undergoing positive selection. I also found that despite the difference in symbiotic associations between all leafhoppers, genes involved in symbiont maintenance and support are evolutionary conserved. In Chapter 4, I investigated the potential role of DNA methylation in the differential gene expression in two symbiont tissues of the aster leafhopper. I leverage fluorescence-based cell sorting methods to isolate symbiont-containing cells and compare their methylation properties with non-symbiotic host cell types. I showed that DNA methylation is not the primary driver of gene regulation in bacteriocytes, but may play a role in some symbiosis-related genes. Collectively, this dissertation extends our understanding of leafhopper-symbiont interactions, pest adaptation mechanisms, and the role of epigenetics in symbiotic partnerships, thus contributing significant insights to the broader field of insect evolutionary biology and genomics.

List of Abbreviations

ALF: aster leafhopper (also known as *Macrosteles quadrilineatus*)
GWSS: glassy winged sharpshooter (also known as Homalodisca *vitripennis*)
TGLH: tea green leafhopper (also known as *Empoasca onukii*)
GRHL: green rice leafhopper (also known as *Nephotettix cincticeps*)
Sulcia: Candidatus Sulcia muelleri
Nasuia: *Candidatus* Nasuia deltocephalinicola
Hodgkinia: Candidatus Hodgkinia cicadicola
Buchnera: Buchnera aphidicola
EAAs: essential amino acids

Chapter 1: Introduction

All animals engage in some form of symbiosis. These symbioses range from those that are obligately beneficial (mutualism) to those that are exploitative and harmful to host fitness (parasitism). Many organisms engage in beneficial symbiosis with microbial partners. In mammals and humans, these complex interactions can be difficult to disentangle due to their high diversity of microbial relationships. Thus, we turn to smaller organisms, such as insects, to answer fundamental questions about the intricacies of beneficial symbiosis.

1.1 Coevolution and symbiosis of insects

1.1.1 Ancient nutritional symbiosis in insects

In insects, many species engage in obligate and long-term symbiotic relationships with bacterial symbionts for essential nutrition and other essential services (Baumann, 2005; Douglas, 2009). It is estimated that over 20% of all known insect species depend on microbial symbionts that they harbor intracellularly (Douglas, 2011). These alliances with symbiotic microbes have permitted insects to diversify into niches that cannot sustain animal life, including sugar-rich or blood diets (Douglas, 2009; Feldhaar, 2011). In many cases, these relationships are ancient, spanning 10-100s of millions of years (Moran and Bennett, 2014; McCutcheon et al., 2019). This ancient association may lead to the formation of specialized host cells (bacteriocytes) and specialized host organs (bacteriomes) where bacteria are sustained in intracellular environments, receiving crucial resources and support from the insect host (Buchner, 1965; Baumann, 2005; Fronk and Sachs, 2022). Bacteriocytes allow for the propagation of bacteria in a stable environment and for the strict transmission from mother to offspring. Despite the known benefits of microbial symbionts to insect hosts and subsequent long coevolutionary history, these relationships are surprisingly dynamic (Bennett and Moran, 2015; Chong et al., 2019; Vasquez and Bennett, 2022). While stable endosymbiosis must be maintained for host fitness, symbiotic partners lose genomic functionality over time. Symbionts can lose over 90% of their free-living ancestral genes, even those deemed vital to both bacterium and the symbiotic relationship (McCutcheon and Moran, 2012a; McCutcheon et al., 2019). Occasionally, symbionts are lost or replaced by more capable partners as they shift to new ecological niches (Koga et al., 2013a; Koga and Moran, 2014a; Husnik and McCutcheon, 2016; Sudakaran et al., 2017; Mao and Bennett, 2020). Thus, symbionts pose significant evolutionary challenges to hosts as they must continually adapt to manage and integrate not only existing bacterial symbiont species, but also newly acquired ones essential to host health and fitness (Toenshoff et al., 2012; Koga et al., 2013a; Bennett and Moran, 2015; Mao and Bennett, 2020).

1.1.2 Genome evolution of endosymbionts

Insects have long histories with beneficial bacteria; however, these associations likely began from infecting pathogenic bacteria. Previously free-living bacteria infect host cells and avoid host immune systems by avoiding detection or destruction. Over time, these bacteria begin to rely on the host for a stable environment, and begin to lose redundant and unessential genes, a process that selection would favor. In turn, the host relies on the bacteria for the novel function that allows for host survival and proliferation. As bacteria begin to survive in an intracellular environment and are passed through host generations through vertical transmission, they lose their ability to gain new genes during an environmental phase. Then, strong genetic drift begins to dominate the evolution of symbiont genes through small populations and severe bottlenecks. The loss of genetic variation in bacteria contributes to the accumulation of deleterious mutations, a process that is aptly named “Muller’s Ratchet” (Moran, 1996a). Consequently, these deleterious mutations become fixed in the bacterial population repeatedly resulting in the loss of function, inactivation, and loss of genes from every cellular process, including those considered to be essential for bacterial cell functions (McCutcheon et al., 2019). Due to the loss of vital bacterial genes such as DNA repair mechanisms, intracellular symbiont genomes can exhibit accelerated sequence evolution and a mutation-driven bias favoring A+T nucleotide (Moran, 1996a; Wernegreen, 2002). This “symbiosis rabbit hole” between host compensatory mechanisms, deleterious mutations and selfish symbiont mutations has led to the occurrence of some of the smallest known bacterial symbiont genomes, as well as unique adaptive host compensation to support these tiny genomes (Bennett and Moran, 2015).

1.1.3 Genomic maintenance and support of endosymbionts

Obligate symbionts, by definition, are required for host fitness and reproduction (Dunbar et al., 2007; Kikuchi et al., 2012). Nevertheless, random and dynamic gene losses in symbiont genomes present significant challenges for their hosts (described above) (Gupta and Nair, 2020). As symbiont genomes degrade, novel evolutionary compensatory mechanisms must evolve to take their place (Hansen and Moran, 2011a; Douglas, 2016; Mao et al., 2018a; Ankrah et al., 2020). These compensatory mechanisms can be from the acquisition of another symbiotic partner, from adaptations from the symbiont genes themselves to moonlight as other functions, or for the host to evolve complex compensatory mechanisms to maintain their bacterial partner(s) (Nikoh et al., 2010; Husnik et al., 2013; Sloan et al., 2014a; Luan et al., 2015; Mao et al., 2018a). For example, in several leafhopper species, hosts express thousands of genes in bacteriocytes to support their symbionts’ individual needs (Mao et al., 2018a; Mao and Bennett, 2020). Hosts employ various strategies to maintain symbiosis, such as reassignment and duplication of mitochondrial support genes, acquisition of horizontally transferred genes from other infecting bacteria, or novel duplicates of host genes (Nakabachi et al., 2005a; Price et al., 2010; Hansen and Moran, 2011a; Price et al., 2011b; Husnik et al., 2013; Duncan et al., 2014; Sloan et al.,

2014a; Luan et al., 2015). Horizontal gene acquisition from other infecting bacteria to hosts can compensate for missing essential bacterial cellular functions when eukaryotic gene support is insufficient. These novel compensatory mechanisms distinctly complement the needs of symbionts, and in the case of multi-partner symbionts, the needs of both symbionts (Mao et al., 2018a; Ankrah et al., 2020). Symbiont gene loss is constrained to some extent by the host's gene repertoire and its capacity to complement deteriorating genomes, including the divergence in prokaryotic and eukaryotic cellular machinery that may not interact well (Sloan et al., 2014b; Wilson and Duncan, 2015). For example, insect hosts have evolved prokaryotic supports due to the interaction of eukaryotes with mitochondria (Gray et al., 2001; McCutcheon et al., 2019). Duplications, on the other hand, may facilitate host cell reconfigurations, but may have less to do with direct prokaryotic interactions (Price et al., 2011b; Duncan et al., 2014; Wilson and Duncan, 2015; Mao and Bennett, 2020).

1.1.4 Auchennorrhyncha and the multi-partner endosymbionts

Among insects, the pierce-sucking insect order Hemiptera engages in symbiosis to feed on a sugar-rich diet found in the phloem and xylem of plants (*e.g.*, cicadas, leafhoppers, aphids; Grimaldi et al., 2005). Phloem and xylem of plants are depauperate in the essential amino acids necessary for animals to make full proteins (Buchner, 1965; Moran et al., 2008). As a result, Hemipteran insects have acquired a diverse array of microbial symbionts for the synthesis of missing nutrition (Douglas, 1989; Moran and Telang, 1998; Bennett and Moran, 2013). Symbioses within the Hemiptera are ancient, although various lineages (*e.g.,* Sternorrhyncha and the Auchenorrhyncha) have acquired and replaced symbionts at different stages of their diversification (Figure 1; Sudakaran et al., 2017; Bennett, 2020). Following establishment, host and their symbionts generally coevolve for 10s-100s of millions of years (Wernegreen, 2002; Baumann, 2005; McCutcheon et al., 2009a; Moran et al., 2009).

Most species in the plant-sap feeding insect suborder, Auchenorrhyncha (Hemiptera: Suborder), have established mutualistic symbioses with two or more bacteria for over 300 million years (Takiya et al., 2006a; Moran et al., 2008). These partnerships generally provide essential amino acids and other nutrients that are absent in their hosts' phloem and xylem diets. Most species in the suborder harbor "*Candidatus* Sulcia muelleri" (*Bacteroidetes*; hereafter known as *Sulcia*) and a *Betaproteobacteria.* Both *Sulcia* and the *Betaproteobacteria* have been associated with Auchenorrhyncha since it emerged in the Permian (Moran et al., 2005; Bennett and Moran, 2013). The latter likely represents a single origin but has been given multiple names over the years: "*Ca.* Vidania fulgoroideae", "*Ca.* Zinderia insecticola", and "*Ca.* Nasuia deltocephalinicola" (hereafter *Nasuia*), are associated with the superfamilies, Fulgoroidea (planthoppers), Membracoidea (leafhoppers and treehoppers) and Cercopoidea (froghoppers), respectively.

Due to its more limited nutritional contributions, the *Betasymbiont* has been replaced several times throughout the suborder. For example, the sharpshooter leafhopper family appears to have replaced the more ancient symbiont, *Nasuia*, with "*Candidatus* Baumannia cicadellinicola" (hereafter *Baumannia*) >60 MYA in a transition between phloem and xylem feeding (Moran et al., 2003). Similar to *Nasuia*, *Baumannia* convergently evolved to provide the same two EAAs as *Nasuia* and few other nutritional resources (Wu et al., 2006). However, because *Baumannia* is relatively young, its genome encodes more functional capabilities requiring much less support from its partner symbionts (Wu et al., 2006; Bennett et al., 2014; Mao and Bennett, 2020). In other early evolutionary events, some leafhopper lineages, such as the Typhlocybinae subfamily, have lost all obligate symbionts as a consequence of shifting to a more nutrient-rich plant parenchyma diet (Günthardt and Wanner, 1981). Species in this group no longer retain *Sulcia* and *Nasuia*, nor the organs that house and support them (Buchner, 1965; Cao and Dietrich, 2022).

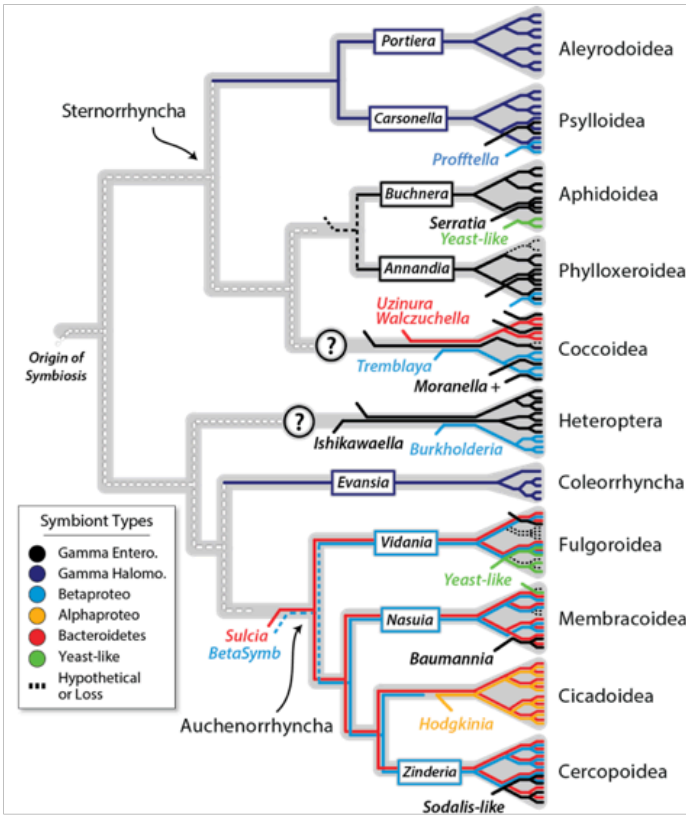


Figure 1: Variations in the bacterial associations across the Hemiptera order. Image from Bennett and Moran (2015).

1.1.5 Leafhoppers as a model for symbiosis

Leafhoppers (Hemiptera: Auchennorrhyncha: Cicadellidae) comprise one of the largest Hemipteran families with over 20,000 described species (Grimaldi et al., 2005; Brambila and Hodges, 2008). While most leafhoppers are sap-feeding, their host range and feeding habits can vary widely. Some species are highly polyphagous and can feed/reproduce on a wide range of plant species, while others are more specialized and feed/reproduce exclusively on certain plant families or species (Weintraub and Beanland, 2006; Wilson and Weintraub, 2007). In addition to their symbiotic relationships with bacteria, leafhoppers are also known for their ability to pierce the phloem and xylem of plants, making them significant vectors of viral (*e.g.,* plant viruses) and bacterial (*e.g.,*

phytoplasmas) plant pathogens, causing enormous economic losses in agricultural and horticultural industries (Banttari and Zeyen, 1979; Nielson, 1979; Tsai, 1979; Weintraub and Beanland, 2006; Hogenhout et al., 2008a; Chasen et al., 2015; Greenway, 2022).

In some leafhopper species, *Sulcia* and *Nasuia* have among two of the smallest bacterial genomes known from any biological system. *Sulcia* (190 KB) has 190 predicted protein-coding genes and a GC content of about 24%, and *Nasuia* (112 KB) has 137 genes and a GC content of about 17% (Bennett et al., 2016a). Unlike *Sulcia*, *Nasuia* has undergone a major shift and uses an alternative genetic code in which the UGA stop codon now encodes for Trp (UGG) (McCutcheon et al., 2009b). *Sulcia* and *Nasuia* genomes maintain a core set of genes, but genetic drift and a lack of DNA repair mechanisms continue to diminish their already tiny genomes (Mao et al., 2017a; Bennett and Mao, 2018).

1.2 Lineage-specific symbiont gene losses in endemic Hawaiian leafhoppers

A long-standing question in symbiosis is how evolutionary processes shape ancient symbiont genome evolution in short time frames. Although we have a good picture of how symbiont genomes shrink on the scale of 10s-100s million years, we understand comparatively little about how this process continues to shape the tiny genomes of ancient symbionts among closely related host species (Wernegreen, 2002; McCutcheon et al., 2009a; Moran et al., 2009; Mao et al., 2017a; Chong et al., 2019). To close this knowledge gap, I leverage one of the largest insect radiations across the Hawaiian Archipelago, *Nesophrosyne* (Hemiptera: Cicadellidae) (Hembry et al., 2021).

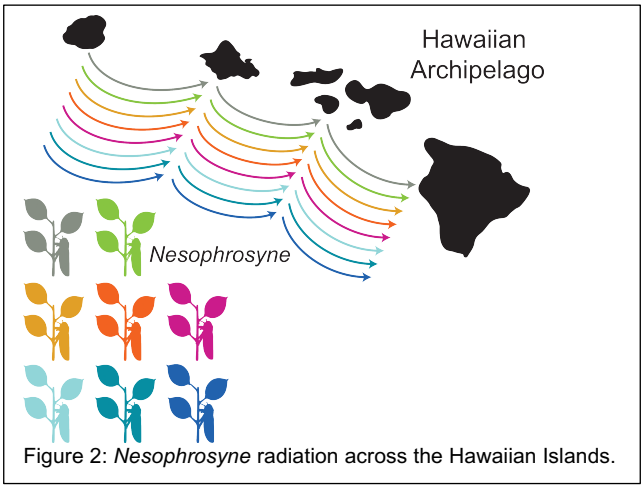


Figure 2: *Nesophrosyne* radiation across the Hawaiian Islands.

Like other Auchennorhynchan insects, *Nesophrosyne* has an association with two leafhoppers, *Sulcia* and *Nasuia*. The genus originated about 3.2 million years ago, rapidly diversifying into over 200 species that specialize across 28 host plant families in a one-to-one fashion (Zimmerman, 1948; Bennett and O'Grady, 2012, 2013). *Nesophrosyne* lineages established these host-plant relationships early in their diversification on the archipelago and then maintained their host-plant associations as species diversified across newly formed islands in a parallel and replicated fashion (Figure 2; Bennett and O'Grady, 2013). Thus, the multiple

independent specialization to host plants across multiple islands has shaped the diversification of insects. Hawaiian leafhoppers present a unique opportunity to ask fine-scale coevolutionary questions due to their isolated nature, ecological and temporal diversity, and replicated and rapid adaptive evolution. Moreover, they are host-plant specific and with limited dispersal capabilities, leading to high levels of local endemicity. The *Nesophrosyne* system permits the coevolutionary study of these complex symbioses to be contextualized in absolute timescales because we already know the historical biogeography, ecological interactions, and tempo of speciation in absolute time for all species in the genus.

In Chapter 2, to better understand the evolutionary processes that shape ancient symbiont genome evolution, I will analyze the genomes of *Sulcia* and *Nasuia* across *Nesophrosyne* lineages. Due to the natural evolutionary "experiments" inherent to island systems, the Hawaiian Archipelago presents an exciting avenue to study the genome evolution of symbionts in absolute time. Additionally, this island system yields a model for understanding how closely related species of hosts maintain symbiosis. I sequence *Sulcia* and *Nasuia* across 20 *Nesophrosyne* species, spanning 8 host plant groups and 5 Hawaiian Islands. Using comparative methods, I analyze the genomic evolution of this multi-partner symbiosis in a closely related species that have diversified across the islands. In this chapter, I leverage the *Nesophrosyne* leafhopper radiation and their endosymbionts to understand (i.) whether and how ongoing gene losses continue to shape ancient symbiont genomes, (ii.) how selection and drift lead to symbiont genome diversification, (iii.) whether these evolutionary forces are shared or distinct between multiple symbionts in a shared host. I further test our questions in absolute time provided by the *Nesophrosyne* radiation to gain general insights into the tempo and mode of bacterial symbiont genome evolution.

1.3 Pest ecology and symbiosis shape leafhopper genomes

The abilities of leafhoppers to feed on a wide range of plants in both agricultural and natural ecosystems and their dependence on beneficial bacteria likely exerts significant evolutionary pressures across their genomes (Francis et al., 2005; Després et al., 2007; Hogenhout et al., 2009; Wang et al., 2018; Zhang et al., 2022). Yet, we have a limited understanding of how leafhopper genomes evolve and how these specific evolutionary pressures influence that process. We further lack effective comparative tools to investigate such questions fully.

Chromosomal-level sequencing of leafhoppers and comparative genomics can provide insights into genetic adaptations that contribute to the evolutionary success of leafhoppers and the many pest species that exist within this group. To date, there are three sequenced leafhopper genomes: green rice leafhopper (Deltocephalinae: *Neophotettix cincticeps*; Yan et al., 2021), glassy-winged sharpshooter (Cicadellinae: *Homalodisca vitripennis*; Li et al., 2022), and the tea

green leafhopper (Typhlocybinae: *Empoasca onukii*; Zhao et al., 2022). These leafhoppers range in their symbiotic associations and host-plant feeding habits. *N. cincticeps* associates with the two ancient symbionts, *Sulcia* and *Nasuia*, while *H. vitripennis* still maintains *Sulcia* but has replaced *Nasuia* with *Baumannia*. Both species are polyphagous in their feeding habits. On the other hand, *E. onukii* is a monophagous pest and no longer retains any symbionts. Despite these different symbiont modalities and dietary patterns, it is possible they undergo similar selective pressures that shape their genomes.

In Chapter 3, I present the complete genome of the aster leafhopper (hereafter known as ALF), *Macrosteles quadrilineatus* (Hemiptera: Cicadellidae: Deltocephalinae), that feeds over 300 agriculturally important plants, including carrots, celery, wheat, barley, flax and lettuce (Wallis, 1962). ALF is the primary vector of the Aster Yellows phytoplasma, a bacterium of a plant disease that leads to stunting and deformation of crops (Kunkel, 1926). In an evolutionary framework, I compared ALF's genome to those of all other existing pest leafhopper genomes. These species come from different leafhoppers subfamilies that have overlapping and distinct symbiotic relationships. Our questions focus on understanding: (i.) how leafhoppers' genomes evolve in a global sense, (ii.) what evolutionary pressures do pest species biology and agricultural ecology place on leafhopper genomes, and (iii.) how do major transitions in symbioses shape the evolution of leafhopper genomes. Our results show remarkable conservation of leafhopper genomes but distinct signatures of pest ecology and symbioses in the expansion and retention of genes and molecular evolution of specific genes.

1.4 Cell sorting of bacteriocytes and its implications in DNA methylation

Cell sorting techniques have emerged as powerful tools for isolating and characterizing of specific cell populations overcoming the challenges of studying complex tissues and samples. To date, cell sorting in bacteriocytes has been used to identify the absolute number of symbiont cells within a bacteriocyte, as well as examine ploidy levels within bacteriocytes (Simonet et al., 2016; Nozaki and Shigenobu, 2022). Yet, fluorescence-based cell sorting has yet to be applied to separate bacteriocytes in a multi-partner symbiotic association. This approach enables the isolation of mostly pure populations of bacteriocytes, facilitating subsequent analyses of individual symbionts and their interactions within the host (*e.g.,* DNA methylation patterns within bacteriocyte cells).

DNA methylation is the process of an enzymatic addition of a methyl group to nucleic acid bases mediated by DNA methyltransferases (DMNTs). This methylation can lead to repression of transcription, often playing a crucial role in regulating gene expression (Suzuki and Bird, 2008; Keller et al., 2016; Zhang et al., 2018). In certain cases, it can contribute to gene silencing by interfering with the binding of transcription factors, inhibiting the transcriptional machinery (Eden

and Cedar, 1994). Epigenetic mechanisms in insects are vital for regulating a wide range of biological processes (Glastad et al., 2014; Bewick et al., 2016). Hemipteran insects have the capability to use DNA methylation as an important DNA modification that aids in the regulation of various biological processes including during nymphal development and adaptation to dietary change in the mono-symbiotic Aphid-*Buchnera* system (Bewick et al., 2016; Kim et al., 2018; Pers and Hansen, 2021). However, almost nothing is known about the role of epigenetic controls in *M. quadrilineatus* (ALF) and related insects in the auchenorrhyncha. Therefore, if methylation plays a role in the expression of thousands of genes between two distinct symbiotic bacteriocyte types is still unknown.

In Chapter 4, I leverage fluorescence-based cell sorting to identify and isolate bacteriocytes, then use it in practice to show its power in downstream analysis such as DNA methylation. In this study, we utilize the observation that bacteriocytes often exhibit higher ploidy levels to isolate bacteriocytes. We focus our study on understanding: (i). what is the best method to dissociate bacteriocyte tissues, (ii). is it possible to isolate bacteriocytes for cell sorting based on ploidy levels, and (iii). how can cell sorting be used to perform other downstream analyses. We use DNA methylation to indicate the power of cell sorting by comparing methylation patterns of isolated bacteriocyte populations to non-symbiotic host cell types. We identified and correlated differential hyper- and hypo-methylated genes within and between bacteriocytes compared to non-symbiotic host cell types with gene expression patterns. Finally, we compared methylation and gene expression patterns of putative symbiosis support genes for each bacteriocyte type to understand the role of DNA methylation in a multi-partner symbiosis.

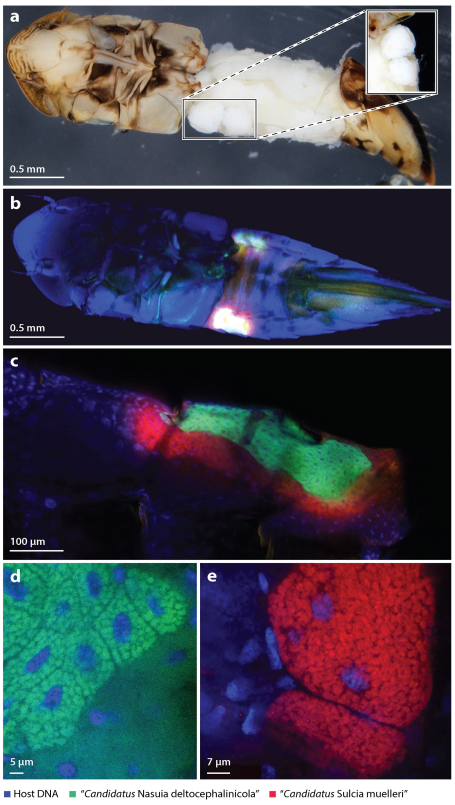


Figure 3: Fluorescent *in situ* hybridization of *Sulcia* and *Nasuia* in the leafhopper, *Nesophrosyne montium*. a-c: bacteriome d-e: bacteriocytes. Image from Moran and Bennett.

Chapter 2: Comparative genomics of endosymbionts across endemic Hawaiian leafhopper (Hemiptera: Cicadellidae: Nesophrosyne) species

2.1 ABSTRACT

Many insects depend on ancient associations with intracellular bacteria for essential nutrition. The genomes of these bacteria are often highly reduced. Although drift is a major driver of symbiont evolution, other evolutionary forces continue to influence them. To understand how ongoing molecular evolution and gene loss shape symbiont genomes, we sequenced two ancient symbionts, *Sulcia* and *Nasuia*, from 20 Hawaiian *Nesophrosyne* leafhoppers. We leveraged the parallel divergence of *Nesophrosyne* lineages throughout Hawaii as a natural experimental framework. *Sulcia* and *Nasuia* experience ongoing—but divergent—gene loss, often in a repeatable fashion. While some genes are under relaxed selection, purifying and positive selection are also important drivers of genome evolution, particularly in maintaining certain nutritional and cellular functions. Our results further demonstrate that symbionts experience dramatically different evolutionary environments, as evidenced by the finding that *Sulcia* and *Nasuia* have one of the slowest and fastest rates of molecular evolution known.

2.2 INTRODUCTION

Bacterial symbionts have enabled many animal groups to take advantage of unsuitable ecological niches leading to their biological diversification (Takiya et al., 2006b; Moran, 2007; Sudakaran et al., 2015; Hendry et al., 2016; Sogin et al., 2020; Myers et al., 2021). In insects, bacterial endosymbionts are a key source of essential nutrition for many species, and even entire orders, that specialize on diets limited in essential nutritional resources (Moran, 1996b; Baumann, 2005; Douglas, 2009). These symbionts are generally restricted to within specialized insect organs (bacteriomes) and cells (bacteriocytes) that enable host-bacterial interaction and strict vertical transmission from mother to offspring (Buchner, 1965; Koga et al., 2012). However, due to their intracellular lifestyle, bacteria often lose over 90% of their genes (McCutcheon and Moran, 2012b; McCutcheon et al., 2019). As symbioses age, bacteria continue to experience ongoing gene losses from their most basic—and essential—cellular processes and metabolisms (*e.g.,* DNA replication and repair). Thus, endosymbionts require extensive resources from their hosts and other bacterial partners to function (Nakabachi et al., 2005b; Hansen and Moran, 2011b; McCutcheon and von Dohlen, 2011; Sloan et al., 2014b; Luan et al., 2015; Mao et al., 2018b; Weglarz et al., 2018; Kobiałka et al., 2018). While we have a good picture of how symbiont genomes shrink on the scale of 10s-100s millions of years, we understand comparatively little about how this process continues to shape the tiny genomes of ancient symbionts among closely related host species (Wernegreen, 2002; McCutcheon et al., 2009; Moran et al., 2009; Mao et al., 2017; Chong et al., 2019).

In many cases, our understanding and inference of the evolutionary processes that influence the structure and function of symbiont genomes is derived from a single, or just a few, representative genomes (McCutcheon et al., 2009; Bennett and Moran, 2013; Koga and Moran, 2014; Bennett and Mao, 2018; Michalik et al., 2021). These processes include selection to retain essential functions, selection to adapt to changing host and environmental conditions, and strong genetic drift due to small population sizes and strong intergenerational bottlenecks (Moran, 1996; Wernegreen, 2002; Woolfit and Bromham, 2003; Campbell et al., 2015; Perreau et al., 2021). These processes are ongoing among related symbiont lineages that are separated into distinct host species and their populations. The intensity of these forces likely varies depending on the bacterial symbiotic roles and cellular environments, as well as the biology and ecology of their different host insect species (Wernegreen, 2002; Sabater-Muñoz et al., 2017; Chong et al., 2019). As a result, there is likely to be tremendous variation among the symbiont genomes of closely related host species. For example, the process of drift is known to cause independent symbiont lineages to differ widely in their genetic capabilities even between closely related host species (McCutcheon and Moran, 2010; Patiño-Navarrete et al., 2013; Husnik et al., 2013; Campbell et al., 2015; Bennett et al., 2016b; Husnik and McCutcheon, 2016; Campbell et al., 2017; Boscaro et al., 2017; Łukasik et al., 2018; Chong et al., 2019; Monnin et al., 2020; Santos-Garcia et al., 2020). It is less clear how drift and selection work together to shape and maintain the genes and functions of symbionts as they diversify along with their hosts. Thus, an investigation into the patterns of gene loss and molecular evolution (*e.g.,* the relative roles of selection vs. drift in gene evolution) among closely related host sister species has the potential to illuminate the fine-scale evolutionary processes that underlie ongoing symbiont genome evolution and diversification.

To better understand the evolutionary processes that shape ancient symbiont genome evolution, we analyzed the genomes of "*Candidatus* Sulcia muelleri" (*Bacteroidetes*; hereafter referred to as *Sulcia*) and "*Ca.* Nasuia deltocephalinicola" (*Betaproteobacteria*; hereafter *Nasuia*) lineages from insect sister species belonging to the endemic Hawaiian leafhopper genus, *Nesophrosyne* (Hemiptera: Auchenorrhyncha: Cicadellidae). *Nesophrosyne* is one of the largest insect adaptive radiations in Hawaii (Hembry et al., 2021). It originated ~3.2 million years ago, rapidly diversifying into over 200 species that specialize on endemic plant species in a one-to-one fashion (Zimmerman, 1948; Bennett and O'Grady, 2012, 2013). *Nesophrosyne* lineages established these host-plant relationships early in their diversification on the archipelago. These lineages then maintained their host-plant associations as they diversified across newly formed islands in a parallel and replicated fashion (Bennett and O'Grady, 2013). Thus, *Nesophrosyne* can provide a natural evolutionary experiment to test questions of how evolutionary processes shape symbiont genome evolution

across diverging insect lineages and to further understand to what extent evolutionary processes are predictable.

Sulcia and *Nasuia*, like most other symbionts in the Auchenorrhyncha suborder, complement each other to provide their hosts with the 10 essential amino acids (EAAs) that are lacking in their xylem and phloem plant sap diets (Bennett and Moran, 2013). Genomic evidence suggests that both bacteria are ancient, having partnered with insects ~300 million years ago (Moran et al., 2005; Bennett and Mao, 2018). As a result, *Sulcia and Nasuia* have highly reduced genomes of 190 kilobases (kb) and 112 kb, respectively (Bennett and Moran, 2013). Both genomes maintain a core set of essential nutritional genes, but are lacking genes in most other essential functions that include translation and transcription, energy synthesis, and DNA replication and repair (Bennett and Moran, 2013; Bennett et al., 2014; Mao et al., 2017).

Here, we leverage the *Nesophrosyne* leafhopper radiation and their endosymbionts to understand (i.) whether and how ongoing gene losses continue to shape ancient symbiont genomes, (ii.) how selection and drift lead to symbiont genome diversification, (iii.) whether these evolutionary forces are shared or distinct between multiple symbionts in a shared host. We further test our questions in absolute time provided by the *Nesophrosyne* radiation to gain general insights into the tempo and mode of bacterial symbiont genome evolution.

2.3 METHODS

2.3.1 Sample collection

We sampled a targeted set of 20 species that span the ecological and phylogenetic diversity of the genus. Adult female and male leafhoppers were field-collected and stored in ethanol to be used for downstream analysis. The selected species are also specific to eight different host-plant species that encompass the diversity of host-plant families and genera that the genus is associated with (see Table 2.1).

2.3.2 Genome sequencing

For each target species, ten field-collected individuals were pooled to obtain enough DNA for genomic sequencing. DNA was extracted with a DNAeasy kit (Qiagen) and concentration quantified with a Quibit 3.0 fluorometer (ThermoFisher). Libraries and Illumina MiSeq sequencing were conducted at UC Berkeley qB3 Functional Genomics Lab for 4 million 2x300 base pair (bp), paired-end reads.

2.3.3 Genome assembly and annotation

Raw reads were quality filtered and cleaned of adapters using Trimmomatic v0.39 and checked with FastQC v0.11.9 (Andrews, 2010; Bolger et al., 2014). Assembly of symbiont genomes was done using SPAdes v3.14 (program settings: -k 127, --only-assembler, --meta; Bankevich et al., 2012). Since extracts contain both host and bacterial DNA, symbiont and mitochondrial genomes were manually extracted by using features unique to each symbiont (*i.e.,* high relative coverage and high AT content). Extracted contigs were verified and confirmed with BLAST searches of open reading frames predicted with GLIMMER v3.02 in Geneious Pro (Altschul et al., 1990; Drummond et al., 2011). To confirm consistent assembly coverage and circularization of bacterial genomes, quality filtered reads were aligned to the completed symbiont genome using Bowtie2 v2.3.5.1 (program settings: --local; Langmead and Salzberg, 2012). Linear chromosomes were circularized by breaking contigs and attaching ends. High, consistent coverage across these ends were verified to confirmed closure of the circular bacterial chromosomes. No plasmids were identified, as is to be expected for tiny symbiont genomes (Bennett and Moran 2013).

Initial genome annotations were done with RAST v2 (Overbeek et al., 2014). Annotations were then verified with GLIMMER v3.02 gene predictions that were checked with BLASTP searches against the nr database (Altschul et al., 1990; Delcher et al., 2007). Bacterial RNA genes were further identified with RNAmmer v1.2 and tRNAscan-SE v2.0 (Lagesen et al., 2007; Lowe and Chan, 2016). Mitochondrial genes were identified with MITOS (Bernt et al., 2013).

2.3.4 Phylogenetic tree construction

To construct a phylogenetic tree in absolute time for downstream analysis, and to verify relationships with the *Nesophrosyne* genus, we extracted and aligned complete mitochondrial genomes of our sampled insect species. We included the mitochondrial DNA from *Orosius sp.* as the known outgroup for *Nesophrosyne* (Bennett and O'Grady, 2012; Fletcher et al., 2017). We also tested co-cladogenesis between host and symbiont by extracting and aligning all *Nasuia* and *Sulcia* genes. For the mitochondria and symbiont genomes, each gene was individually aligned with MAFFT v7.455 using the L-INS-i model (Katoh and Standley, 2013). Genes that did not occur in all genomes, or that were difficult to align with confident site homology, were omitted.

The resulting phylogenetic data sets included concatenated protein coding and ribosomal genes for a total of 184 genes (181,781 sites) from *Sulcia*, 99 genes (86,095 sites) from *Nasuia*, and 15 genes (14,304 sites) from leafhopper mitochondria. Best-fit models of nucleotide substitutions and partitioning schemes were determined using PartitionFinder 2 (program settings:

branchlengths = unlinked, models = all, model_selection = bic; Lanfear et al., 2017). Bayesian time calibrated phylogenies were then inferred using BEAST v.1.10.4, using the generated partition scheme and corresponding molecular substitution models (Drummond and Rambaut, 2007). The tree prior included the yule process speciation with a random starting tree. Five internal node calibrations were selected following our previous phylogenetic study of the *Nesophrosyne* (Bennett and O'Grady, 2013). Briefly, internal node calibrations were determined from *Nesophrosyne* species divergences that match the sequential geological formation of the Hawaiian Islands (*i.e.,* progression rule). Calibrations were applied with a normal prior distribution since absolute species divergence could have occurred earlier or after island formation (see Bennett and O'Grady, 2012 for additional information; Bennett and O'Grady, 2013). Multiple BEAST chains were run per genome alignment and sampled every 1000 generations following Bayesian recommendations (two chains with four million generations; Huelsenbeck et al., 2002). Runs were performed with an uncorrelated relaxed clock with a lognormal distribution. Convergence and stationarity of chains were verified with ESS values were >200 using Tracer v1.7.1 and RWTY v1.0.2 (Rambaut et al., 2018; Warren et al., 2017).

2.3.5 Ancestral genome reconstruction and ancestral state reconstruction

To determine patterns of gene loss (*i.e.,* unique vs convergent gene losses) among symbionts between host lineages, we reconstructed the ancestral gene retention across *Nesophrosyne's* symbionts. We further included symbiont genomes from previously sequenced lineages including the aster leafhopper (*Macrosteles quadrilineatus*) and the keeled treehopper (*Entylia carinata*) (Bennett and Moran, 2013; Mao et al., 2017). Maximum likelihood ancestral state reconstructions were with phytools v.1.0-1 package (Revell, 2012; RStudio Team, 2018). We used a custom model that allows for the loss of genes to occur (no gene gain) to account for the inability of symbionts to recombine with other environmental or symbiotic bacteria. Posterior density of stochastic character maps was generated by simulating 100 trees.

2.3.6 Patterns of molecular evolution

To test for genome wide substitution rates across *Sulcia* and *Nasuia* genes, as well as host mitochondria, we used MEGAX v.10.2.4 (Kumar et al., 2018). Model selection for pairwise evolutionary distances were selected with JModelTest2 v2.1.10 with constricted model selection to those available for MEGAX analyses (*e.g.,* Jukes-Cantor, Tamura-Nei, etc.; Darriba et al., 2012). To test the rate of substitutions between islands within the same plant family, two pairwise analyses were done: (i.) the oldest diverging species (*e.g.,* Kauaʻi and Hawaiʻi species) and (ii.) the closest diverging species (*e.g.,* Maui and Hawaiʻi species). These ages ranged from the most recent divergence (0.351 MYA in *Broussaisia*) to the most ancient divergence (2.4239 MYA in *Pipturus;* see Fig. 2.4A). Rates of

substitutions were graphed across absolute time of divergence. To test whether *Sulcia* and *Nasuia* show correlation in their evolutionary rates with *Nesophrosyne* genes, we performed a Kruskal-Wallis test and a pairwise comparison with a Wilcoxon rank sum test and a Benjamini-Hochberg correction for multiple tests (Wilcoxon, 1945; Benjamini and Hochberg, 1995; McKight and Najab, 2010; RStudio Team, 2018). Additionally, in our analysis, we consider the rate of evolution in absolute time by testing across host-plant groups (specifically *Broussaisia, Coprosma, Pipturus* and *Myoporum)* and within host-plant clades (Fig. 2.4).

2.3.7 Selection among symbiont genomes

In order to test for selection among symbiont genes, we measured rates of synonymous (d*S*) and nonsynonymous (d*N*) substitutions with codeml v4.8 using the mitochondrial time-calibrated phylogeny (Yang, 2007). The ratio of nonsynonymous to synonymous substitutions (ω = d*N*/d*S*) were calculated for each aligned *Sulcia* and *Nasuia* gene. Selection is calculated by measuring the ratio of nonsynonymous to synonymous substitutions (denoted by ω), where $\omega > 1$, $\omega = 0$ and $\omega < 1$ indicate positive, neutral and purifying selection, respectively (Yang and Bielawski, 2000).

We estimated rates of synonymous and nonsynonymous substitutions using three models. The M0 model was used to test for selection (ω) across all codon sites. This generated a single ω value that was evaluated further. Additionally, we used two nested models, M1a-M2a and M7-M8, to determine strong support for selection in codon sites (Anisimova et al., 2001). Models M1a and M7 are constrained and disallow positive selection while the M2a and M8 models are unconstrained, allowing for positive selection. The M7-M8 models offer a more stringent test of positive selection (Anisimova et al., 2001). However, by using multiple nested models, we verify positive selection in genes that are highly supported in both models (Anisimova et al., 2001). The likelihood scores were compared within paired models (chi-squared test; $p <= 0.05$) to indicate significant positive selection for genes. To confirm the specificity of our results, we only consider genes that were identified as being under significant positive selection by both nested models for downstream analysis. Genes were further separated into Clusters of Orthologous Genes (COG) categories to test for functional group enrichment within functional categories, using a Fisher's exact test with Benjamini-Hochberg Procedure multiple-testing correction (Benjamini and Hochberg, 1995; Fisher, 1992).

2.4 RESULTS AND DISCUSSION

2.4.1 Host-symbiont taxon sampling and nomenclature for genomic analysis

To compare the genomes of symbionts associated with *Nesophrosyne*, we strategically sub-sampled 20 species that span the ecological and phylogenetic diversity of the leafhopper genus. Our selected species comprise monophyletic groups that specialize on eight distinct endemic Hawaiian plants, encompassing the diversity of host-plant associations known in the *Nesophrosyne* (Bennett and O'Grady, 2012; see Figs. 1 and 2). The sister species of each host lineage diversified across the Hawaiian Archipelago, with distinct species occurring on each island and even on a volcanic mountain (Bennett and O'Grady, 2013).

Sulcia and *Nasuia* strains are hereafter identified by the island and host plant genera associated with the leafhopper host as follows: the initial for the island location (*e.g.,* Hawai'i Island = HI) and the first two letters of the plant genus (*e.g., Pipturus* = PI). For example, the *Sulcia* strain associated with the *Nesophrosyne* species restricted to Hawai'i Island and the plant genus *Pipturus* is referred to as *Sulcia*-HIPI (Table 2.1).

2.4.2 Phylogenetic relationships among Nesophrosyne and their symbionts are congruent

To test our questions in absolute time, we generated a time calibrated phylogeny from complete *Nesophrosyne* mitochondrial genomes (15 genes, 14,304 sites) using absolute time calibration points determined previously by Bennett and O'Grady (2013). To compare phylogenetic topologies of the host and symbionts, we also reconstructed phylogenies for *Nasuia* (99 genes, 86,095 sites) and *Sulcia* (184 genes, 181,781 sites) independently using their complete genomes. The relationships among our sub-sampled host species agree with previous phylogenetic work and there is strong support for the monophyly of leafhoppers associated with their host-plant groups (Bennett and O'Grady, 2013, 2012; see Figure S1). However, two *Nesophrosyne* species are weakly supported in their placements and vary between trees derived from the symbiont and mitochondrial data sets (KIDO and MIPS; see Figures S1-S3). Previous work also observed a similar mid-depth polytomy, suggesting a rapid early diversification in the *Nesophrosyne* that is unable to be split by available genetic and genomic data (*i.e.,* a hard polytomy; see Bennett and O'Grady 2012), or the possible introgression of symbiont lineages between hybridizing hosts. Hybridization events are known to occur among rapidly diversifying, host-plant restricted auchenorrhynchan lineages in Hawai'i (Roesch Goodman et al., 2012). However, we do not currently have suitable sampling of host species and their populations to thoroughly test this hypothesis in the *Nesophrosyne*.

2.4.3 Gene content of Nesophrosyne's symbiont genomes varies despite their highly reduced size

From the Illumina sequenced *Nesophrosyne* host species, we recovered 20 *Sulcia* and 18 *Nasuia* genomes (see Table S1). All symbiont genomes are

complete and circularized. Even coverage across all circularized genomes was verified with read mapping that verified complete, high-quality assemblies (see Methods). Two *Nasuia* genomes (*Nasuia*-HIPI and *Nasuia*-OIDO) were omitted because sequencing coverage was too low to assemble reliable contigs and complete genomes required for our downstream molecular assays.

The average genome size of *Sulcia* is 190 kilobases (kb; range = 190.3-190.9 kb) with an average of 190 protein coding genes (range = 188-192 genes). These genomes further retain a single conserved 16S/23S/5S rRNA operon and 30 tRNAs. In contrast, *Nasuia* exhibits more variation between host species. Its average genome size is 112 kb (range = 107.7-116.1 kb) with an average of 132 protein coding genes (range = 125-139 genes), a single 16S/23S/5S rRNA operon, and 18-21 tRNAs (see Table S1). *Nesophrosyne's Sulcia* genomes are highly conserved, varying by two genes (~1% of its genome). In contrast, *Nasuia's* genome is highly variable among the *Nesophrosyne*, differing by up to 12 genes (~10% of its genome). In contrast, *Nasuia's* genome is highly variable among the *Nesophrosyne*, differing by up to 24 genes (>20% of its genome).

Globally, patterns of genome evolution and gene retention among *Sulcia* and *Nasuia* are similar to patterns observed in other Auchenorrhyncha lineages. Both retain complementary essential amino acid (EAA) pathways in an 8 + 2 arrangement for *Sulcia* and *Nasuia*, respectively, as observed in other leafhoppers and related insects (Chang et al., 2015; Bennett et al., 2016a; Mao et al., 2017; see also Bennett and Moran, 2013 for a list and pathways in *Nasuia* and *Sulcia*). The highly conserved nature of *Sulcia*'s genome has been widely observed across the other major auchenorrhynchan lineages that retain it (*e.g.,* sharpshooter leafhoppers, cicadas, spittlebugs; McCutcheon and Moran, 2007, 2010; Koga and Moran, 2014; Campbell et al., 2015; Łukasik et al., 2018; Matsuura et al., 2018). The most dramatic differences observed among *Sulcia* genomes occurs between the major infraorders Fulgomorpha (planthoppers) and Cicadomorpha (leafhoppers and kin). Among the planthoppers, *Sulcia* genomes are much smaller than in the cicadomorphan lineages (<149 kb in Fulgomorpha vs. an avg. of 251 kb in Cicadomorpha [range = 179 kb - 288 kb]; Bennett and Mao, 2018; Michalik et al., 2021). Among other gene losses, planthopper *Sulcia* lineages provide only 3 of the 7-8 EAAs typically retained in strains found in cicadomorphan hosts (McCutcheon and Moran, 2010a; Michalik et al., 2021).

Nasuia's genomes, in contrast to *Sulcia*, exhibit more variation in the genes they retain. The number of gene losses involve ~10% of *Nasuia*'s genome (n = 12 genes) among the *Nesophrosyne*. This variation is significant considering that *Nasuia*'s genome is among the smallest known of any bacterium. The loss of any single gene that *Nasuia* lineages still retain likely requires direct adaptation by the host, or its partner symbionts, to support lost genetic and functional capabilities (Moran and Bennett, 2014; McCutcheon et al., 2019). The diversity of co-primary symbionts associated with *Sulcia* also show similarly higher genomic

variation. For example, the cicada co-primary symbiont, "*Ca.* Hodgkinia" (hereafter *Hodgkinia*), has the most dramatic genomic variation among a symbiont yet observed. While some cicada species harbor *Hodgkinia* with typical circular chromosomes (avg. size = 142 kb), in other hosts its genome is broken into mini circles of varying size and complexity (*e.g.,* fragments range from 71-150 kb in 13-year *Magicicadas*; Van Leuven et al., 2014; Campbell et al., 2015; Łukasik et al., 2018). Similarly, the lineages of the co-primary symbiont *"Ca.* Baumannia", which replaced *Nasuia* in sharpshooter leafhoppers >60 million years ago, vary by the loss of large chunks of its genome spanning >100 kb and >100 protein coding genes (Wu et al., 2006; Bennett et al., 2014, 2016b).

2.4.4 Ancestral gene losses are ongoing and to some extent evolutionarily convergent

To determine patterns of gene loss (*i.e.,* unique vs. convergent losses) among *Sulcia* and *Nasuia* genomes, we reconstructed ancestral patterns of gene loss and retention with maximum likelihood approaches (see Figures S4-S5 and Methods). We also used *Sulcia* and *Nasuia* lineages from two previously sequenced species from the Membracoidea superfamily, the aster leafhopper (*Macrosteles quadrilineatus*) and the keeled treehopper (*Entylia carinata*), to determine ancestral patterns of gene loss leading to the *Nesophrosyne* lineage (Bennett and Moran, 2013; Mao et al., 2017b). In general, the membracoidean *Sulcia* and *Nasuia* lineages are structurally conserved and perfectly syntenic, aside from differences in their patterns of individual gene losses (reviewed by Mao et al., 2017). Below we summarize gene losses in both symbionts. We caution that gene loss counts presented here are minimums, as we did not sequence the symbiont genomes from all of *Nesophrosyne*'s 200+ species.

Sulcia in the common ancestor to Membracoidea likely retained at least 210 protein coding genes and 30 tRNAs (reviewed by Mao et al., 2017). Prior to the diversification of the Deltocephalinae leafhoppers (*Nesophrosyne*'s leafhopper subfamily), *Sulcia*'s genome was reduced to 192 genes and 30 tRNAs (Shcherbakov, 2002; Mao et al., 2017). Among the *Nesophrosyne*, it has further undergone two instances of gene loss (*bamAD*; see Fig. 1). The loss of these genes impacts a range of key cellular functions in *Sulcia,* which have significant implications for how the symbiosis functions and is maintained (Wilson and Duncan, 2015; Bublitz et al., 2019; Mao and Bennett, 2020). One of these gene losses appear to be a unique independent event, while the other has been gently lost multiple times during *Nesophrosyne* diversification.

One gene removed from *Sulcia*'s genome was lost multiple times in a convergent evolutionary event: *bamA* (Fig. 1). The outer membrane protein assembly factor gene (*bamA*) was lost at least two independent times among our sampled *Nesophrosyne* species (Fig. 1). It is part of a multi-gene complex essential for bacterial outer membrane assembly and metabolite exchange (Malinverni et al.,

2006; Charles et al., 2011). Interestingly, the loss of *bamA* in *Sulcia*-KISC co-occurs with the loss of its interacting partner protein (*bamD*) and may be linked (Wu et al., 2005). Thus, some gene losses may instigate the loss of others in a domino-like fashion, as has been proposed to occur in the *Blattabacterium*-cockroach system (Kinjo et al., 2021). It is important to note that in small symbiotic genomes, there is the possibility of gene functions to be rescued despite frameshift deletions (Tamas et. al., 2008). However, this requires experimental confirmation which has not been done in this scope of work. *Nasuia* in the common ancestor to Membracoidea retained at least 163 genes and 29 tRNAs (Mao et al., 2017). Early on in the divergence of the Deltocephalinae leafhoppers, *Nasuia's* genome was further reduced to a mere ~142 genes and just over 112 kb in size (Bennett and Moran, 2013). Across the Nesophrosyne, Nasuia has undergone at least 12 instances of gene loss (~10% of its genome). Seven of these genes have been lost once among our sampled *Nesophrosyne* host species, including genes involved in ribosome function (*rpsKR* and *rpmB*) and tRNA synthesis (*tilS* in *Nasuia*-MICO). The latter gene is essential for EAA and general protein synthesis (Soma et al., 2003, Van Leuven et al., 2019). These lost genes could undergo function rescue and should be experimentally tested (Tamas et. al., 2008).

Half of the gene losses in Nasuia (n = 6) have been convergently lost in at least two or more insect species (Fig. 2.2). The most extreme case of convergent loss is that of a gene cassette that includes four complete genes: 30S ribosomal subunit protein S21 (*rpsU*), tyrosine-tRNA ligase (*tyrS)*, tRNA-specific 2-thiouridylase (*mnmA)*, and flavodoxin/ferrodoxin-NADP+ reductase (*fpr*) (see ancestral state reconstructions in Fig. 2.3B). Additionally, an essential component of the DNA replication holoenzyme, *dnaN*, has been lost at least six times and is missing from more than half of our *Nasuia* genomes (Johanson and McHenry, 1980; Fig. 2.3C).

The ongoing loss of 10% of *Nasuia'*s genes among these lineages similarly presents major challenges to its hosts and partner symbionts, particularly because some of these genes are essential to its cellular functions (*e.g*., *tilS*). Although the number of genes Nasuia is capable of losing stands in stark contrast to *Sulcia*, both symbionts require the host or companion symbionts to adapt stabilizing support mechanisms (Mao et al., 2018). However, *Nasuia* is likely a more demanding partner requiring independent host lineages to innovate novel support strategies. The reason for *Nasuia'*s more exaggerated rates of gene losses over *Sulcia* is likely due to its rapid rates of molecular evolution discussed in the sections below (Bourguignon et al., 2020).

2.4.5 Sulcia and Nasuia have among the slowest and fastest rates of symbiont molecular evolution

To determine the underlying drivers of symbiont genome evolution among *Nesophrosyne*, we estimated genome-wide substitution rates in absolute time for *Sulcia* and *Nasuia*. We further compared these against host mitochondrial rates (Fig. 2.4). The average substitution rates for *Sulcia* genomes are 3.24×10^{-9} substitutions/site/year. In contrast, *Nasuia* genomes exhibit a 34.8-fold higher rate of molecular evolution (avg. = 1.13×10^{-7} substitutions/site/year). The evolutionary rates of both symbiont genomes do not exceed that of the mitochondria (avg. = 2.15×10^{-7} substitutions/site/year: 1.9-fold from *Nasuia* and 66-fold from *Sulcia*). *Nesophrosyne* mitochondrial rates of evolution are in-line with observations from other hemipteran insects, which are generally elevated relative to other insects (Dowton et al., 2009; Song et al., 2012; Cui et al., 2013; Cameron, 2014; Li et al., 2017).

To compare the rate of molecular evolution in *Sulcia* and *Nasuia* to available symbionts from other insect hosts, we estimated nonsynonymous (dN) and synonymous (dS) substitutions over their divergence times. We used the general M0 model, which averages substitutions rates across whole genes and phylogeny. We further converted rates to dN/time (dN/t) and dS/time (dS/t) (Yang, 2007; Silva and Santos-Garcia, 2015). The rates for *Sulcia* among the *Nesophrosyne* are 9.46×10^{-9} dN/t and 4.21×10^{-8} dS/t (time = 3.2 MYA). In contrast, *Nasuia's* rates are highly elevated, averaging 1.65×10^{-7} for dN/t and 3.76×10^{-6} for dS/t. The average dN/t in *Nasuia* is 17.5-fold higher than in *Sulcia*, while the differences in average dS/t are even higher (89.3-fold).

Compared to other insect symbionts, for which data is available, *Nasuia* has among the highest evolutionary rates yet identified. The well-known symbiont of aphids, *Buchnera*, has an average rate of 2.58×10^{-9} dN/t and 1.43×10^{-8} dS/t, which is 64.2 and 262-fold less than *Nasuia's* rates, respectively (using 20 *Buchnera* protein coding genes; Clark et al., 1999). Several of the highest rates of molecular evolution previously documented for insect symbionts belong to *Baumannia*, which replaced *Nasuia* in sharpshooter leafhoppers >60 million years ago, and *Blochmannia* found in carpenter ants (Silva and Santos-Garcia, 2015). *Nasuia* exceeds these, with 42 and 341-fold higher for dS/t and 20 and 165-fold higher for dN/t than *Blochmannia* and *Baumannia,* respectively.

One factor that may influence the rates of evolution in *Nasuia* and *Sulcia*, as well as other symbionts more broadly, is differences in host and symbiont generation times (*e.g.*, bacteria with shorter, more frequent generations can incur more mutations and substitutions per some unit of time; Degnan et al., 2005; Silva and Santos-Garcia, 2015). However, we do not have a clear understanding of generation times in our insects for several reasons. First, they are difficult to rear due to their highly restricted habitat ranges and species further exhibit differences in their rain-fall associated seasonality (Bennett and O'Grady, 2012; see also Degnan et al., 2005). In addition, it is not known whether symbiont replication rates are even coupled with host generations in our system, nor

among most other insect symbioses. It is, however, worth noting that *Sulcia* and *Nasuia* exhibit dramatically different rates of evolution despite sharing the same host lineages. Thus, host generation time alone cannot explain observed differences.

In contrast to *Nasuia*'s highly elevated rates of molecular evolution, *Sulcia* has one of the most depressed rates of any biological system (McCutcheon et al., 2009; Bennett et al., 2014; see Fig. 2.4). It has been widely observed among the Auchenorrhyncha (*e.g.,* in cicadas and spittlebugs), that *Sulcia* has a nearly inert rate of molecular evolution even across divergences spanning 100s of millions of years (Takiya et al., 2006b; Koga et al., 2013b; Bennett et al., 2014; Bennett and Mao, 2018; Waneka et al., 2021; Arab and Lo, 2021; Michalik et al., 2021). *Sulcia's* depressed evolutionary rates is an enigmatic biological phenomenon. One possible explanation for *Sulcia*'s reduced rates of molecular evolution may be its retention of mutation repair systems. In a *Macrosteles* leafhopper, *Sulcia* was found to have an overall low rate of mutagenesis compared to *Nasuia,* possibly due to its retention of the DNA mismatch repair protein, *mutS* (Waneka et al., 2021). The *mutS* gene recognizes and initiates the repair of mismatched bases and small indels, which can lower substitution rates (Dettman et al., 2016; Long et al., 2018; Waneka et al., 2021). Most *Sulcia* lineages in the Auchenorrhyncha still retain *mutS*. This gene retention pattern may further explain *Sulcia's* universally conserved rates of molecular evolution (McCutcheon and Moran, 2007; McCutcheon et al., 2009; McCutcheon and Moran, 2010; Woyke et al., 2010; Bennett and Moran, 2013; Koga and Moran, 2014; Bennett and Mao, 2018). In contrast, the *mutS* gene is widely lost from most of *Sulcia*'s co-primary symbiont partners, which have characteristically elevated rates of molecular evolution—with the exception of "*Candidatus* Zinderia insecticola" in spittlebugs, which still maintains the gene (Takiya et al., 2006; McCutcheon and Moran, 2012; Bennett et al., 2014; Campbell et al., 2015; Bennett et al., 2016b; Arab and Lo, 2021). These higher rates may drive genomic volatility and variation among partner symbionts, which are more frequently lost or replaced among the Auchenorrhyncha (Bennett and Moran, 2015; Sudakaran et al., 2017).

2.4.6 Rates of molecular evolution are uncorrelated between each symbiont

To further test whether *Sulcia* and *Nasuia* show correlated rates of evolution with each other and the host mitochondria, suggesting a shared evolutionary environment (Arab and Lo, 2021), we estimated substitution rates for both symbionts and mitochondria globally across and within the *Nesophrosyne*'s monophyletic host-plant associated clades (*Broussaisia, Coprosma, Pipturus,* and *Myoporum*; see Fig. 2.4a). Correlations between substitutions rates among host and symbiont genomes could be explained if similar forces of selection are acting on the genomes, such as shared population bottlenecks and dependence on shared genes (*e.g.,* host mutation repair genes taking over for those lost in symbiont genes; McCutcheon and Moran, 2012; Mao et al., 2018; McCutcheon

et al., 2019) Overall, substitution rates between *Nesophrosyne's Sulcia* and *Nasuia,* and the mitochondria are not correlated across host-plant affiliated clades (Fig. 2.4b). Rates between symbionts and mitochondrial genes within host-plant clades are also uncorrelated (Fig. 2.4b). Our results support a recent analysis of *Sulcia* from more widely divergent auchenorrhynchan clades that found similar decoupling of rates (see Arab and Lo, 2021). However, both of these findings are in contrast to mono-symbiont systems (*Blochmannia*-Carpenter Ants, Blattabacterium-Cockroaches, and *Buchnera*-Aphids) that tend to show significant correlation between the rate of evolution in mitochondrial and symbiont genes (Degnan et al., 2004; Arab et al., 2020; Arab and Lo, 2021).

The disparity in rates of molecular evolution between *Sulcia* and its partners strongly suggests that different molecular and cellular processes, and evolutionary pressures, are likely shaping symbiont genomes (see Fig. 2.4; Takiya et al., 2006; Bennett et al., 2014; Campbell et al., 2015; Arab and Lo, 2021). While there is evidence that more ancient mitochondrial and plastid symbiont genomes do tend to show correlated rates of molecular evolution, these genomes are highly integrated into the general biology of most eukaryotic cellular and metabolic processes (Smith and Lee, 2010; Sloan et al., 2012; Hua et al., 2012). In contrast, the biological roles of insect symbionts are arguably less integrated into system-wide biological functions of their host insects. Nutritional symbionts are sequestered to distinct organs in the host, retaining relatively enriched genetic autonomy, distinct population sizes, and distinct cellular replication and repair capabilities (Buchner, 1965; Mira and Moran, 2002; Koga et al., 2012; Bennett et al., 2014; Chong and Moran, 2016; Mao and Bennett, 2020; Stever et al., 2021). As a result, they likely do not experience the same patterns and processes of molecular evolution as their partner symbionts, mitochondria, or host nuclear genes. Nevertheless, more fine-scale analyses that include a broadscale sampling of host nuclear and symbiont genes, as well as focus on protein domains, may further find rate correlations on interacting genes.

2.4.7 Sulcia and Nasuia experience differential selection patterns across their genomes

To understand how different modes of selection are shaping the evolution and function of *Sulcia's* and *Nasuia's* genome, we tested for selection both across genes and across sites using the ratio of nonsynonymous to synonymous substitutions (denoted by ω; see Methods). To determine whether symbiont genes are generally under strong purifying selection ($\omega < 0.1$), relaxed purifying selection ($0.95 < \omega > 0.1$), or positive selection ($\omega > 1$), we initially used M0 model in codeml (ω estimated across the whole gene; Yang et al., 2000; Z. Yang, 2007; Sloan and Moran, 2012; Sabater-Muñoz et al., 2017; Perreau et al., 2021). In *Sulcia*, a small subset of genes are undergoing strong purifying selection (avg. $\omega = 0.0615$ [range = 0.0001 – 0.0970, n = 34]), while most genes are undergoing relaxed purifying selection (avg. $\omega = 0.331$ [range = 0.112– 0.931, n = 133]).

Additionally, three genes in *Sulcia* show signatures of positive selection (*atpH, putA, trpC;* $\omega > 1$). In *Nasuia*, most genes are undergoing strong purifying selection (avg. $\omega = 0.0449$ [range = 0.0162 – 0.0988, n = 73]), while comparatively few genes are experiencing relaxed purifying selection (avg. $\omega = 0.243$ [range = 0.105 – 0.555, n = 12]). This approach did not detect any *Nasuia* genes under positive selection.

Although it is useful to obtain an average ω value for a gene, it is not a sufficiently realistic model to detect signatures of positive selection that operate at finer scales (Anisimova et al., 2001). Specific codons related to intrinsic protein function may be under positive selection, while the majority of the gene can experience relaxed purifying selection (Yang et al., 2000; Anisimova et al., 2001). Thus, to test for positive selection on different codon sites within symbiont genes, we used two nested models: M1a-M2a and M7-M8 in codeml (Yang, 2007). We interpret consistent results between the two models (and also the M0 model from above) as strong global support for positive selection operating on sites within a gene (Anisimova et al., 2001; Padhi et al., 2009; Price et al., 2011a; Alves et al., 2013). We applied this approach to *Sulcia* and *Nasuia* across our sampled *Nesophrosyne*.

Sulcia site-based selection analyses recovered the same genes undergoing positive selection from the M0 model, as well as an additional 94 genes (97 in total, $p < 0.05$ with BH correction). Genes under positive selection are functionally enriched for the Clusters of Orthologous Categories (COG) Translation (J; 31 genes; (fisher exact test, $p = 0.0008$; Fig. 2.5) and Energy Production and Conversion (C; 14 genes; $p = 0.0486$; Fig. 2.5). It is notable that, although not significant, 27 genes under positive selection are involved in essential amino acid synthesis (E; $p = 0.1030$; Fig. 2.5).

Sulcia's genes under positive selection are primarily involved in essential amino acid synthesis, buffering degraded protein function, and other cellular processes. Some examples of proteins with positive sites include protein chaperonins (*groEL* and *dnaK*), EAA metabolite synthesis (*e.g.,* arginine [*argBDEG*], lysine [*asd, dapBD,* and *lysAC*], valine [*ilvBCEN*], and phenylalanine and tryptophan [*aroABCEGK* and *trpABCE*]), transcription and translation (*rplA* and *rpsDFMNPT*), transcription release factors (*prfAB*), aminoacyl tRNA synthetases (*glnS, leuS, serS, tyrS,* and *valS*), and energy synthesis (*gapA* and *atpBCFGH*). In addition to genes involved in the incomplete TCA cycle of *Sulcia* (*aceF, acoA, lpdA, korB*, and *sucA*) and the pathway for the conversion of glutamine to carbamoyl phosphate (*carAB*). While some of these genes are clearly important to the host (*e.g.,* amino acid related genes), others are key components of *Sulcia*'s independent cellular stability. For example, *Sulcia's groEL* gene, which assists in the folding of damaged proteins (reviewed in Kupper et al., 2014), exhibits an overall pattern of strong purifying selection, but some sites within it are under positive selection. As suggested previously in pea-aphids, the sites

within *groEL* under positive selection are involved in gene interaction domains that likely improve its ability to bind with other rapidly evolving proteins (Fares et al., 2002; Fares et al., 2004).

In *Nasuia*, only four genes are predicted to be under positive selection with site-based models. This result is congruent with results from the M0 model that show most *Nasuia* genes are under strong purifying selection (Fig. 2.5). None of which are associated with amino acid synthesis and are too few for COG enrichment analyses. Given *Nasuia's* elevated rates of molecular evolution, selection seems to be operating to maintain function, even though this symbiont has a higher rate of gene loss compared with its partner. The ongoing widespread gene losses from *Nasuia's* genome may generally result from its high rates of molecular evolution. In this scenario, the probability of a random mutation disabling a gene that can become fixed through genetic drift is much higher (Wernegreen, 2015). Similar results have also been demonstrated in the genomes of *Blattabacterium* from cockroaches and *Buchnera* from pea-aphids, where the rate of genome reduction is also associated with an increased mutation rate rather than selection acting upon genes (Bourguignon et al., 2020; Kinjo et al., 2021).

2.5 CONCLUSION

Our sampling of *Nesophrosyne* leafhoppers and their symbionts provides a fine-scale look into the processes that underlie symbiont genome evolution. Gene losses among the *Nesophrosyne's Sulcia* and *Nasuia* lineages demonstrate that there is still ongoing volatility, even among symbionts with two of the smallest known genomes. However, these patterns vary widely between these symbiont species. *Sulcia* has one of the slowest evolving genomes known, while *Nasuia* has an exceptionally fast evolving one. *Sulcia* also has far more genes under positive or relaxed purifying selection (80% of genes tested) than *Nasuia's* genome, which is largely under strong purifying selection (86% of genes tested). Taken together, our results indicate that the two symbionts experience independent cellular, metabolic, and evolutionary pressures. These differences may further lead to a high level of retention of *Sulcia* among the Auchenorrhyncha and the relatively high turnover of its partner symbionts. (reviewed in Koga et al., 2013; Bennett and Moran, 2015; Sudakaran et al., 2017; Bourguignon et al., 2020).

Regardless of the differences between *Sulcia* and *Nasuia*, our results show that there is repeatability in gene losses that may be more easily accommodated if mutations render them non-functional. Both exhibit convergent gene loss events, particularly regarding genes involved in transcription and translation. These findings suggest a pre-adapted genetic or cellular host environment that permits these genes to be repeatedly lost (see Figs. 1 and 2). The host, or partner symbionts, may be capable of filling gaps in these bacterial cellular processes, as is predicted to occur in a wide range of hemipteran insect systems (Hansen and

Moran, 2011; Sloan et al., 2014b; Luan et al., 2015; Mao et al., 2018; Van Leuven et al., 2019; Mao and Bennett, 2020).

Finally, even though ancient symbiont genomes of insects show some level of conservation among host lineages, a closer look among host sister species reveals complex patterns of gene loss and modes of selection. The evolutionary processes acting on even the tiniest symbiont genomes are clearly dynamic and highly variable. This understanding can be overlooked when comparing lineages among disparately related taxonomic groups.

2.6 FIGURES

Host Species	**Island Location**	**Habitat Type**	**Host-Plant Group**	**Shorthand Naming**
N. sp. 295	East Hawai'i Island	Rain Forest	*Broussaisia*	EHBR
N. ogradyi	East Maui Island			EMBR
N. kanawao	North Hawai'i Island			NHBR
N. makaihe	West O'ahu Island			WOBR
N. sp. 23	Kaua'i Island	Rain Forest	*Clermontia*	KICL
N. haleakala	Haleakalā Mtn, Maui	Spanning all habitats	*Coprosma*	HMCO
N. sp. 302	Hawai'i Island			HICO
N. sp. 58	Maui Island			MICO
N. sp. 29	Kaua'i Island	Dry-Mesic	*Dodonea*	KIDO
N. maratima	O'ahu Island			OIDO
N. sp. 281	Hawai'i Island	Spanning all habitats	*Myoporum*	HIMY
N. sp. 126	Kaua'i Island			KIMY
N. sp. 242	Moloka'i Island			MIMY
N. sp. 246	O'ahu Island			OIMY
N. montium	Hawai'i Island	Rain Forest	*Pipturus*	HIPI
N. sp. 17	Kaua'i Island			KIPI
N. sp. 48	Maui Island			MIPI
N. ponapona	O'ahu Island			OIPI
N. sp. 15	Moloka'i Island	Dry-Mesic	*Psychotria*	MIPS

N. sp. 21	Kaua'i Island	Spanning all habitats	*Scaevola*	KISC

Table 2.1: Shorthand nomenclature used for bacterial symbiont strains. Names are the first letters of the island location and the first two letters of the host-plant group. Habitat type of each host-plant group is added as well. (*N.* = *Nesophrosyne*, sp. = species)

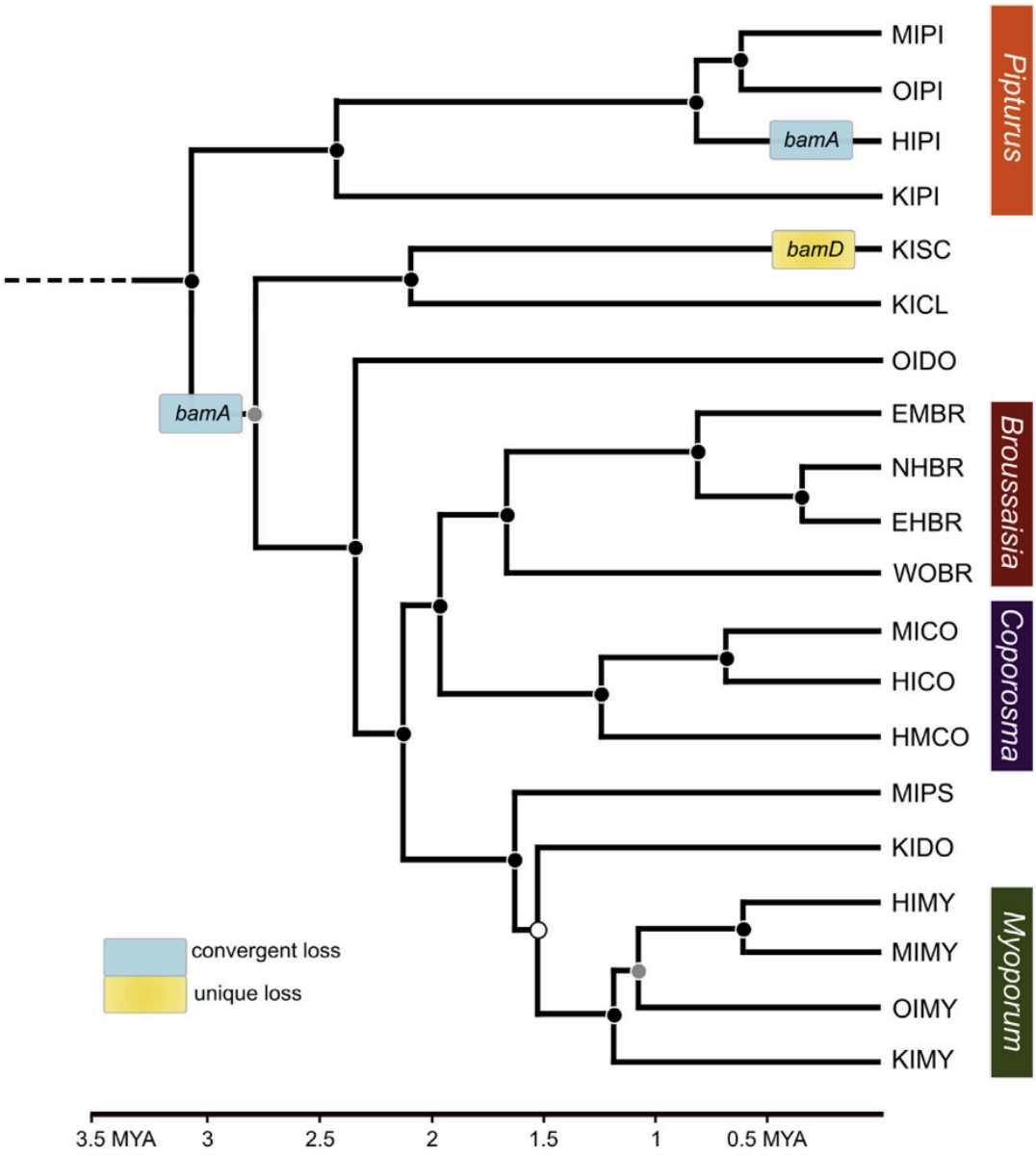


Figure 2.1: Convergent and unique gene loss among *Sulcia* genomes from endemic Hawaiian leafhoppers (*Nesophrosyne*).

Patterns of gene loss were estimated with maximum likelihood ancestral state reconstruction on an absolute time calibrated phylogenetic tree in millions of years, using complete host mitochondrial genes. Convergent gene losses (i.e., multiple repeated losses) are indicated by blue boxes. Unique gene losses (i.e., losses that occurred once) are indicated by yellow boxes. Shorthand gene names are provided in each box. Posterior support for each node is shown as colored circles as follows: black > 95, grey = 90–95 and white < 90. See section methods for details on tree search parameters. See Table 2.1 for species shorthand nomenclature. MYA = million years ago.

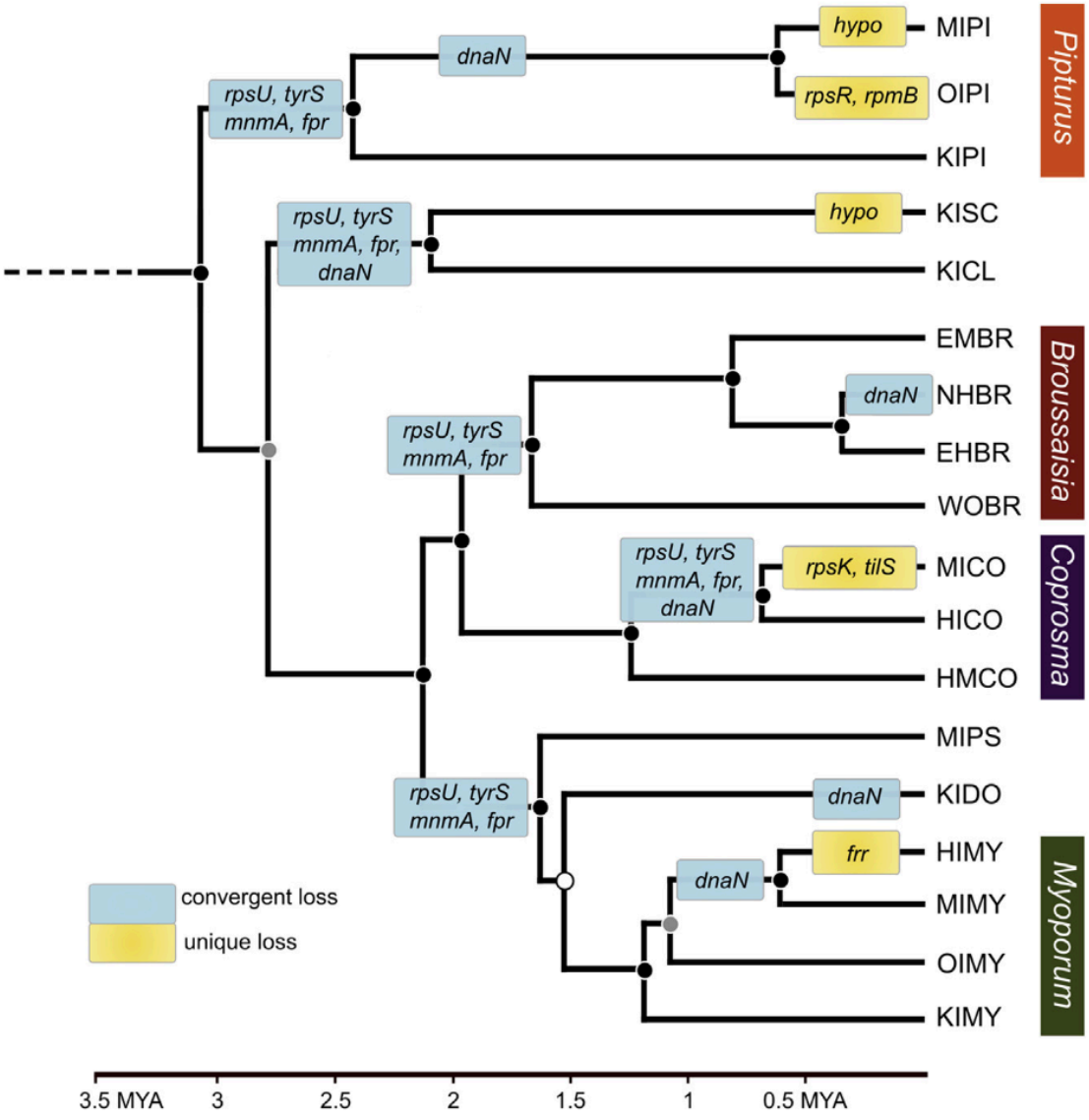


Figure 2.2: Convergent and unique gene loss among *Nasuia* genomes from endemic Hawaiian leafhoppers (*Nesophrosyne*).
Patterns of gene loss were estimated with maximum likelihood ancestral state reconstruction on an absolute time calibrated phylogenetic tree in millions of

years, using complete host mitochondrial genes. Convergent gene losses (i.e., multiple repeated losses) are indicated by blue boxes. Unique gene losses (i.e., losses that occurred once) are indicated by yellow boxes. Shorthand gene names are provided in each box. Posterior probability support for each node is shown as colored circles as follows: black > 95, grey = 90–95 and white < 90. See section methods for details on tree search parameters. See Table 2.1 for species shorthand nomenclature. MYA = million years ago. hypo = hypothetical protein.

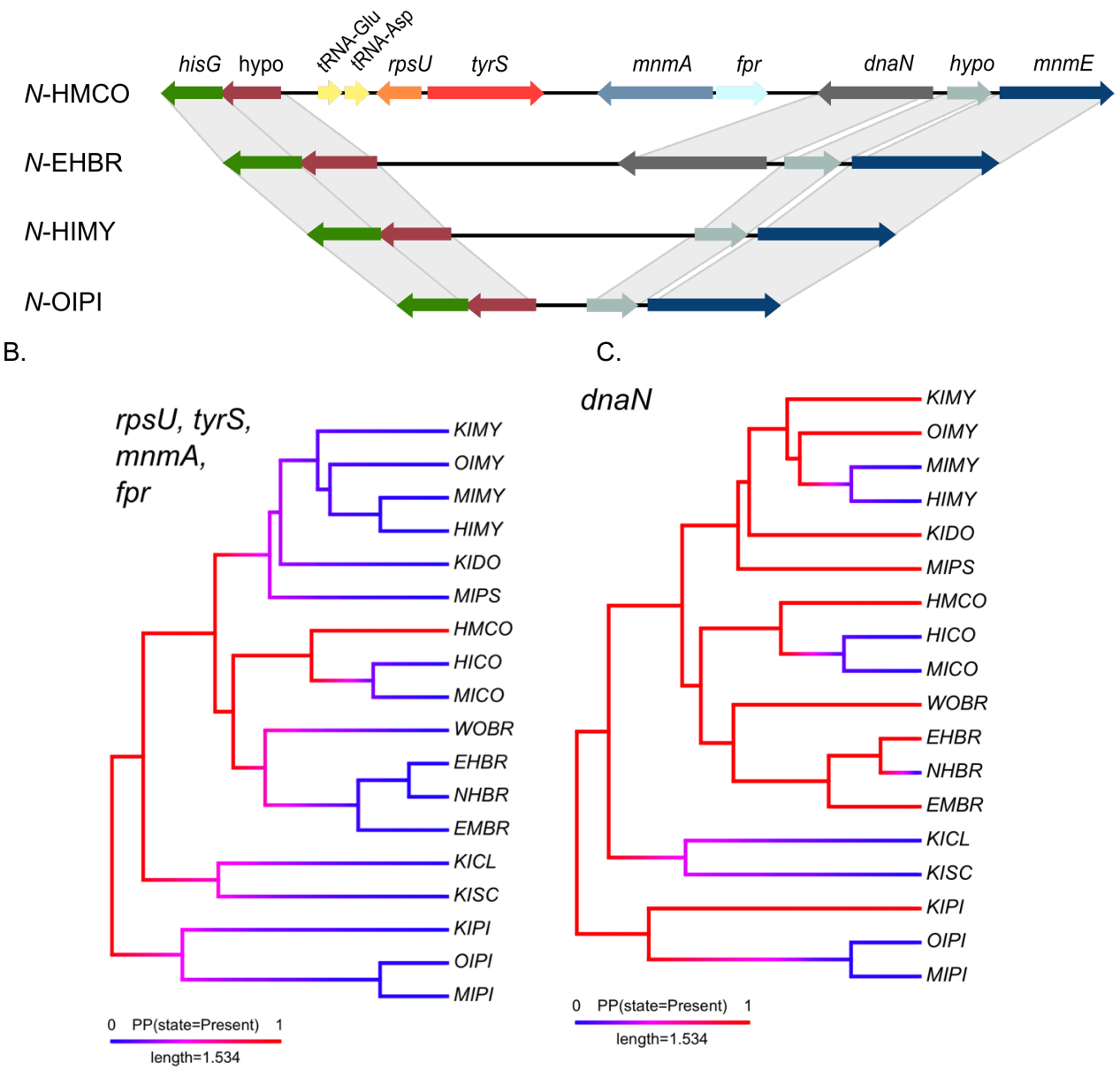
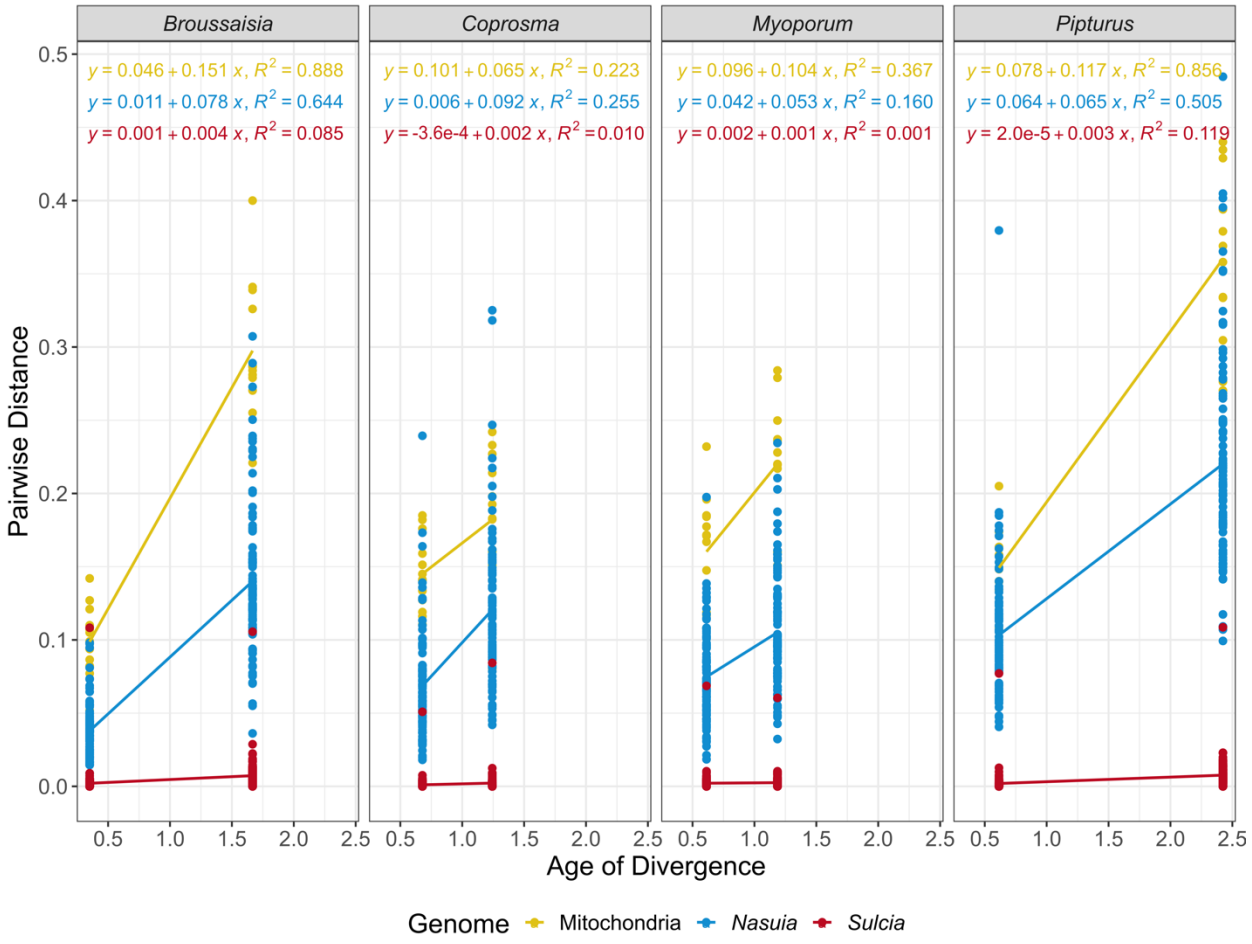


Figure 2.3: Convergent loss of gene segment in *Nasuia*.

A) *Nasuia* gene losses of *rpsU, tyrS, mnmA, fpr* and *dnaN* cassette for three representative genomes, as well as the genes maintained in *Nasuia*-HMCO (*N*-HMCO) sequence. The genes *rpsU, tyrS, mnmA* and *fpr* have been lost in all genomes except *N*-HMCO (top). The *dnaN* gene has been convergently lost in nine sequences. Ancestral state reconstruction of *rpsU, tyrS, mnmA* and *fpr* (B) suggest that these genes have been lost at least five times. Ancestral state reconstruction of *dnaN* (C) suggest that these genes have been lost at least five times. Colors correspond to the same gene in each genome segment (hypo = hypothetical protein). See Table 2.1 for species shorthand nomenclature. B & C) Ancestral state reconstruction of *rpsU, tyrS, mnmA* and *fpr*, as well as *dnaN* from phytools v.1.0-1 package on genes convergently lost in *Nasuia* (Revell, 2012). The loss of *rpsU, tyrS, mnmA* and *fpr* may have instigated the loss of *dnaN.* See section methods for details on ancestral state reconstruction.

A.



B.

	Summary of *p*-values				
	Across All Groups	*Myoporum*	*Coprosma*	*Pipturus*	*Broussaisia*
Mitochondria - *Nasuia*	<2e-16	4.20E-13	6.80E-11	2.80E-07	2.30E-14
Mitochondria - *Sulcia*	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
Sulcia - Nasuia	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

Figure 2.4: Summary of uncorrelated rates of evolution between symbiont and host genes.
A) Linear regression between pairwise distances and age of divergence among all *Nasuia, Sulcia,* and mitochondrial genes for Hawaiian leafhoppers (*Nesophrosyne*). Colors indicate pairwise distance of protein coding genes in each genome. Genomes are separated by the host-plant group that the host species has specialized on. A linear regression line is mapped between pairwise distance values from the closest related species to the most divergent species in each host-plant group. The regression equation and the coefficient of determination (R^2) are also reported and colorized by genome. Outliers with a pairwise distance >0.5 were removed (4 from *Nasuia*, 3 from Mitochondria). B) Associated *p*-values of statistical tests between evolutionary rates of genomes across all plant groups and within plant groups. Significant *p*-value between groups indicates no correlation between means of evolution rate.

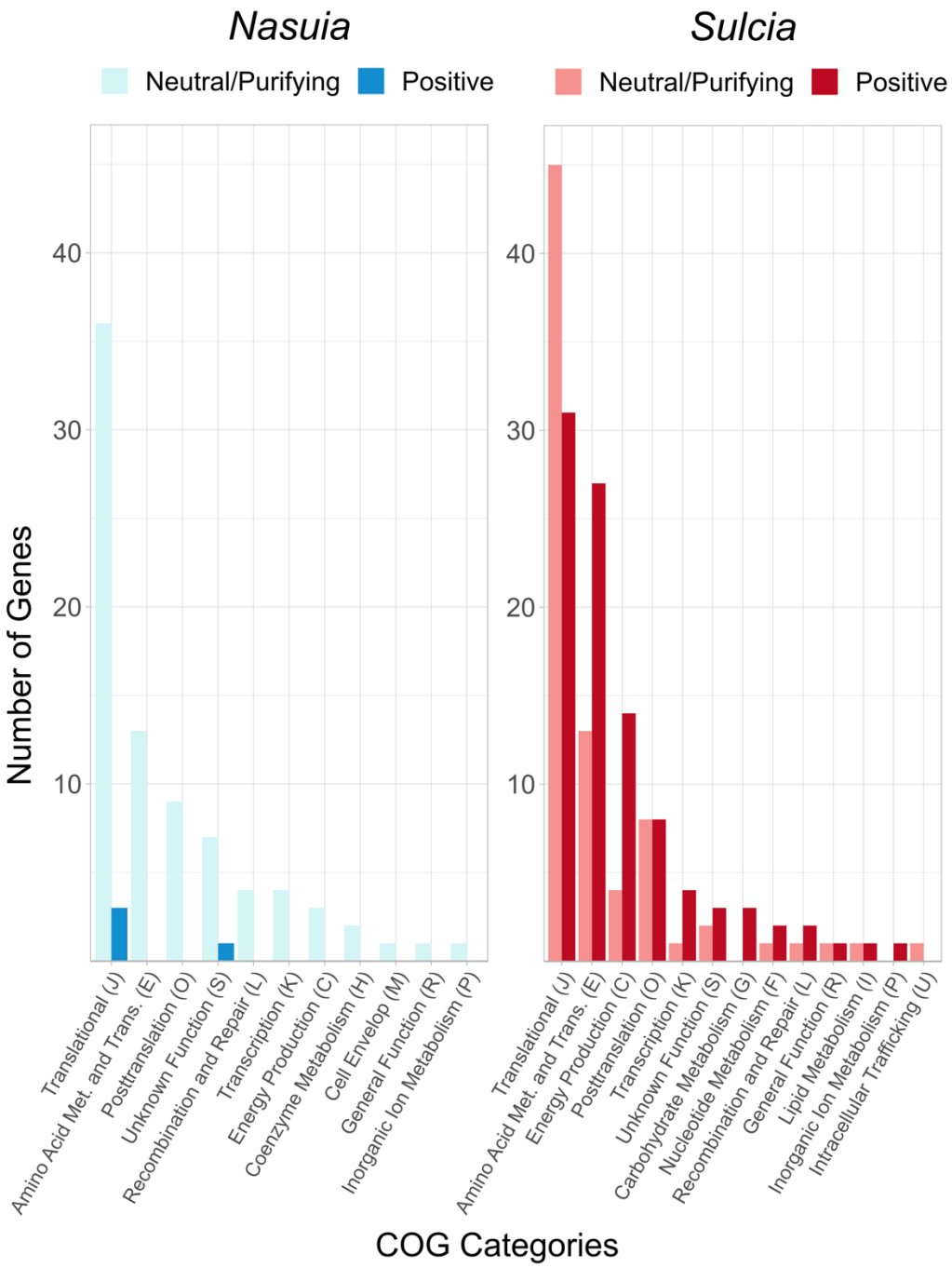


Figure 2.5: Bar chart showing genes undergoing positive selection or purifying/neutral selection across Sulcia and Nasuia genomes from Hawaiian leafhoppers (Nesophrosyne).
Genes are binned into their clusters of orthologous groups (COGs) functional categories (Tatusov et al., 2000). Bars are color coded according to their genome (*i.e., Sulcia* or *Nasuia*) and selection (*i.e.,* Positive or Neutral/Purifying). We used two nested models, M1a-M2a and M7-M8, to determine overall support for

selection among genes (See section methods for further explanation; (Anisimova et al., 2001). The likelihood scores were compared within paired models (chi-squared test; $p \leq 0.05$)

Chapter 3: Chromosomal-level assembly of the aster leafhopper agricultural pest (Hemiptera: Cicadellidae: *Macrosteles quadrilineatus* leafhopper)

3.1 ABSTRACT

Leafhoppers comprise over 20,000 plant-sap feeding species, many of which are important agricultural pests. Most species rely on two ancestral bacterial symbionts, *Sulcia* and *Nasuia*, for essential nutrition lacking in their phloem and xylem plant sap diets. To understand how pest leafhopper genomes evolve and are shaped by microbial symbioses, we completed a chromosomal-level assembly of the aster leafhopper's genome (ALF; *Macrosteles quadrilineatus*). We compared ALF's genome to three other pest leafhoppers, *Neophotettix cincticeps, Homalodisca vitripennis,* and *Empoasca onukii,* that have distinct ecologies and symbiotic relationships. Despite diverging ~155 million years ago, leafhoppers have high levels of chromosomal synteny and conservation of gene families. Conserved genes include those involved in plant chemical detoxification, resistance to various insecticides, and defense against environmental stress. Positive selection acting upon these genes further points to ongoing adaptive evolution in response to agricultural environments. In relation to leafhoppers' general dependence on symbionts, species that retain the ancestral symbiont, *Sulcia*, displayed gene enrichment of metabolic processes in their genomes. Leafhoppers with both *Sulcia* and its ancient partner, *Nasuia,* showed genomic enrichment in genes related to microbial population regulation and immune responses. Finally, horizontally transferred genes (HTGs) associated with symbiont support of *Sulcia* and *Nasuia* are only observed in leafhoppers that maintain symbionts. In contrast, HTGs involved in non-symbiotic functions are conserved across all species. The high-quality ALF genome provides deep insights into how host ecology and symbioses shape genome evolution and a wealth of genetic resources for pest control targets.

3.2 INTRODUCTION

Leafhoppers (Hemiptera: Auchennorrhyncha: Cicadellidae) are one of the largest Hemipteran families, encompassing >20,000 described species (Grimaldi et al., 2005; Dietrich, 2005; Brambila and Hodges, 2008). They rely on plants for food and reproduction, with most species exclusively feeding on phloem or xylem plant sap. The feeding range of leafhoppers can vary considerably, with some species exhibiting a high degree of polyphagy, while others specialize exclusively on only one or a few plant species (Weintraub and Beanland, 2006; Wilson and Weintraub, 2007). Leafhoppers are also primary vectors for many viral (*e.g.,* plant viruses) and bacterial (*e.g.,* phytoplasmas) plant pathogens, causing enormous economic losses in agricultural and horticultural industries (Banttari and Zeyen, 1979; Nielson, 1979; Tsai, 1979; Weintraub and Beanland, 2006; Hogenhout et al., 2008a; Chasen et al., 2015; Greenway, 2022). Remarkably,

however, leafhoppers depend on non-pathogenic, beneficial bacteria to feed on plants in the first place (Buchner, 1965). Due to the essential nutritional deficiencies in their primary diet (xylem and phloem), most leafhopper species have evolved ancient and complex nutritional relationships with bacteria. The abilities of leafhoppers to feed on a wide range of plants in both agricultural and natural ecosystems and their dependence on beneficial bacteria likely exerts significant evolutionary pressures across their genomes (Francis et al., 2005; Després et al., 2007; Hogenhout et al., 2009; Wang et al., 2018; Zhang et al., 2022). Yet, we have a limited understanding of how leafhopper genomes have evolved and how these specific evolutionary pressures influenced that process. We further lack effective comparative tools to investigate such questions fully.

Here, we present the complete genome of the aster leafhopper (hereafter known as ALF), *Macrosteles quadrilineatus* (Hemiptera: Cicadellidae: Deltocephalinae). ALF is a widespread pest that feeds on over 300 agriculturally important plants, including carrots, celery, wheat, barley, flax, and lettuce (Wallis, 1962). It is the primary vector of the Aster Yellows phytoplasma, a bacterium that causes crop stunting, deformation, and ultimately loss (Kunkel, 1926). Given ALF's ability to feed and reproduce on multiple plant species, and dependence on beneficial microbial symbionts, ALF is an emerging model system for understanding vector biology and beneficial symbioses (Hogenhout, Ammar, et al., 2008; Hogenhout, Oshima, et al., 2008; Bennett & Moran, 2013; Mao et al., 2018).

ALF's ability to feed on such a diverse array of plants is owed in part to its dependency on two intracellular bacteria, "*Candidatus* Sulcia muelleri" (*Bacteroidetes*; hereafter *Sulcia*) and "*Candidatus* Nasuia deltocephalinicola" (*Betaproteobacteria*; hereafter *Nasuia*). Both bacteria complement each other to provide ALF with the ten essential amino acids (EAAs) that are depauperate in its phloem diet and that no animal can make (Moran et al., 2008; McCutcheon & Moran, 2012; Bennett & Moran, 2013; Douglas, 2017). ALF, and many other related auchenorrhynchan insects, house symbionts within specialized cells (bacteriocytes) and organs (bacteriomes) and exclusively transmit bacteria transovarially (Buchner, 1965; Baumann, 2005; Fronk and Sachs, 2022). As a result, *Sulcia* and *Nasuia* are ancient, having been vertically transmitted within lineages and across generations for >300 million years (Moran et al., 2005; Bell-Roberts et al., 2019). These conditions have led to the streamlining and severe reduction of their genomes to <10% of those of their free-living ancestors (190 kb and 112 kb in *Sulcia* and *Nasuia*, respectively; McCutcheon and Moran, 2012; McCutcheon et al., 2019). To maintain these highly degraded symbionts, the host or partner symbionts must compensate for incomplete genomic functions (Hansen and Moran, 2011a; Douglas, 2016; Mao et al., 2018a; Ankrah et al., 2020). To accomplish this, ALF has acquired 100s-1000s of support genes that differentially support *Sulcia* or *Nasuia*. These support mechanisms evolved from the reassignment of mitochondrial support genes, ancient horizontal gene acquisitions from other infecting bacteria, and widespread gene duplications

(Mao et al., 2018a). However, how these evolutionary processes have structured ALF's genome, as well as other leafhoppers with different symbiotic relationships (i.e., losses and replacements), remains unclear.

While ALF and related species, such as the treehoppers from the Membracidae family and leafhoppers from the Deltocephalinae subfamily, generally retain both ancient symbionts, other leafhopper groups have replaced or lost *Sulcia*, *Nasuia*, or both (Mao et al., 2017a; Sudakaran et al., 2017; Bennett and Mao, 2018; Bell-Roberts et al., 2019; Michalik et al., 2021). For example, the sharpshooter leafhopper family appears to have replaced the more ancient symbiont, *Nasuia*, with "*Candidatus* Baumannia cicadellinicola" (hereafter *Baumannia*) >60 MYA in a transition between phloem and xylem feeding (Moran et al., 2003). Similar to *Nasuia*, *Baumannia* convergently evolved to provide the same two EAAs as *Nasuia* and few other nutritional resources (Wu et al., 2006). However, because *Baumannia* is relatively young, its genome encodes more functional capabilities requiring much less support from its partner symbionts (Wu et al., 2006; Bennett et al., 2014; Mao and Bennett, 2020). In other early evolutionary events, some leafhopper lineages, such as the Typhlocybinae subfamily, have lost all obligate symbionts as a consequence of shifting to a more nutrient-rich plant parenchyma diet (Günthardt and Wanner, 1981). Species in this group no longer retains *Sulcia* and *Nasuia*, nor the organs that house and support them (Buchner, 1965; Cao and Dietrich, 2022). In contrast to sharpshooter leafhoppers, we have a more limited understanding of how these kinds of dramatic symbiotic transitions influence the evolution of host genomes.

To better understand how pest ecology and obligate symbioses shape ALF and other leafhopper genomes more generally, we used PacBio HiFi long-read sequencing and Omni-C long-range proximity ligation to generate a high-quality chromosome-level genome assembly of ALF. In an evolutionary framework, we compared ALF's genome to those of all other existing pest leafhopper genomes. These species come from different leafhoppers subfamilies that have overlapping and distinct symbiotic relationships. Our questions focus on understanding: (i.) how leafhoppers' genomes evolve in a global sense, (ii.) what evolutionary pressures do pest species biology and agricultural ecology place on leafhopper genomes, and (iii.) how do major transitions in symbioses shape the evolution of leafhopper genomes. Our results show remarkable conservation of leafhopper genomes but distinct signatures of pest ecology and symbioses in the expansion and retention of genes and molecular evolution of specific genes.

3.3 MATERIALS AND METHODS

3.3.1 Insect rearing and material preparation

Lab-reared *Macrosteles quadrilineatus* (ALF) insect lines for this analysis were collected in New Haven, Connecticut in 2013. A single mated female and male of

ALF was used to establish an inbred line to reduce genetic heterozygosity. This line was maintained on barley grass for four generations at 25 °C with 12L:12D light/dark. Inbred female and male adults were starved for 6 hours, immediately flash-frozen in liquid nitrogen, and stored at -80 °C.

3.3.2 Genome sequencing and assembly

A chromosome-level genome assembly was generated with the Omni-C proximity ligation technique developed by Dovetail Genomics (Santa Cruz, CA, USA). The assembly process involved scaffolding assembled genome contigs into chromosomes using a combination of long-read (Pacific Biosciences: PacBio) and short-read (Illumina HiSeqX) sequences. Briefly, a draft genome was assembled using 58.2 giga base-pairs (GB) of PacBio HiFi circular consensus sequencing reads and the de-novo assembler Hifiasm v0.15.4-r347 with default parameters. Scaffolds identified as possible contamination by blobtools v1.1.1 were removed (Laetsch and Blaxter, 2017). Haplotigs and contig overlaps were removed using purge_dups v1.2.5 (Guan et al., 2020). The dovetail Omni-C library was sequenced on an Illumina HiSeqX platform for ~30x coverage. The *de novo* PacBio assembly and Dovetail OmniC library reads were used as the inputs for HiRise proximity ligation assembly (Putnam et al., 2016). To evaluate genome completeness, BUSCO v4.0.5 was used on the chromosome-level assembly using the eukaryote_odb10 lineage dataset (Simão et al., 2015; Zdobnov et al., 2017).

3.3.3 Genome Annotation

The NCBI Eukaryotic Genome Annotation Pipeline was used for genome annotation (NCBI 2017). Repeat families were identified and masked using WindowMasker (Morgulis et al., 2006). For gene predictions using Gnomon, RNA-seq data from five previously sequenced ALF samples and high-quality protein coding sequence alignments from six closely related insects were aligned to the genome with STAR v2.7.10b and ProSplign v3.8.2 (Kiryutin et al., 2007; Dobin et al., 2013). The final annotation quality of ALF's *genome* was assessed with BUSCO v4.1.4 (Simão et al., 2015; Zdobnov et al., 2017; Manni et al., 2021).

3.3.4 Comparative genomics and orthologous gene analysis

Chromosomal conservation and shared gene families across leafhoppers was inferred using comparative genomic analysis with the green rice leafhopper (Deltocephalinae: *Neophotettix cincticeps*; hereafter known as GRLH; Yan et al., 2021), glassy-winged sharpshooter (Cicadellinae: *Homalodisca vitripennis*; hereafter known as GWSS; Li et al., 2022), and the tea green leafhopper (Typhlocybinae: *Empoasca onukii*; hereafter known as TGLH; Zhao et al., 2022) (Table 1). Prior to all comparative analyses, annotated gene isoforms and

transcript duplicates were consolidated into the single longest gene using a custom script. Total orthogroups and single-copy orthologs between all species were identified using OrthoFinder v2.5.4 (Emms and Kelly, 2019). For further clustering of orthologous groups and gene enrichment analyses across the four leafhopper species, we used OrthoVenn2 (settings: e-value = 1e-5, inflation value = 1.5) (Xu et al., 2019). We identified Gene Ontology (GO) categories enriched in conserved gene clusters across leafhopper lineages and those exclusive to ALF. Functional gene enrichment analyses were conducted across three levels: (i.) shared core genes across all species, (ii.) between different symbiotic modalities or relationships (leafhoppers with *Sulcia*, and leafhoppers with both *Sulcia* and *Nasuia*), and (iii.) unique gene clusters to ALF. Among all levels of functional enrichment analyses, we identified gene clusters that contribute to pest ecology, as well as symbiosis, in order to understand how these pressures may be shaping leafhopper genomes. We assessed chromosomal synteny among the leafhopper species using BLASTP (settings: e-value = 1e-5) and MCScanX with default parameters (Altschul et al., 1990; Wang et al., 2012). Finally, SynVisio was used for the visualization of chromosomal synteny across the genomes (Bandi and Gutwin, 2020).

3.3.5 Gene family expansion and contraction analysis in ALF

To investigate genome-wide family expansions and contractions within the ALF genome, we analyzed protein-coding sequences from 20 publicly available insect genomes (Table S1). This dataset comprised 18 species from the Hemipteran order, as well as two from the Hemipteran sister order, Thysanoptera (Johnson et al., 2018). Annotated gene isoforms and transcript duplicates were consolidated into the single longest gene in all 20 species using a custom script. OrthoFinder v2.5.4 was used to infer gene orthology between species. The resulting single-copy orthologs were aligned with MAFFT v7.52 (settings: -m L-INS-I model) (Katoh and Standley, 2013; Emms and Kelly, 2019). Ambiguously aligned regions were trimmed using BMGE v1.12 (settings: -m BLOSUM90, -h 0.4) (Criscuolo and Gribaldo, 2010). The concatenated alignments were then used to construct a maximum likelihood (ML) phylogenetic tree with IQ-TREE v2.2.03 using the best-fit partition model for each gene set (settings: -m MFP+MERGE, -B 1000) (Minh et al., 2020). The resulting ML tree was used to estimate divergence times between insect species. MCMCTree (PAML v4.10.6) was used to place 95% confidence intervals for the six node calibrations as soft bounds between Thysanoptera and Hemiptera (373.3, 451.2 million years ago (MYA)), Sternorrhyncha and Auchennorrhyncha/Heteroptera (353.8, 427.3 MYA), Psylloidea and Aleyrodidae (322.4, 396.6 MYA), Aphidomorpha and Psylloidea (303.9, 377.2 Ma), Fulgoromorpha and Cicadomorpha (275.3, 348.6 MYA) and finally between the Deltocephalinae tribes Macrostelini and Chiasmini (45.0, 95.0 Ma) (Yang, 2007; Johnson et al., 2018; Cao et al., 2022).

Gene family evolution (*e.g.*, expansions and contractions) was inferred using CAFÉ v5 (setting: --pvalue 0.01) (Han et al., 2013). Gene families that were identified as significant expansion or contraction events in ALF were functionally annotated using eggNOG-mapper . To perform enrichment at the KEGG functional level for significantly expanded families, we used the R package “clusterProfiler” v4.6.2 (settings: pvalueCutoff = 0.01, pAdjustMethod = "BH", qvalueCutoff = 0.01, minGSSize = 10) (Yu et al., 2012).

3.3.6 Selection analysis of leafhopper genes

Evolutionary selection operating on leafhopper genes was estimated using the ratio of nonsynonymous to synonymous substitutions ($\omega = d_N/d_S$) across sites in genes. This analysis was first conducted on genes from ALF, GWSS, TGLH, and GRLH to provide a broader perspective on evolutionary patterns in leafhoppers. A narrower analysis was then performed between only the Deltocephalinae leafhoppers (ALF and GRLH) to reduce the evolutionary divergence, potential for substitution saturation, and to determine the influence of both *Sulcia* and *Nasuia* that are retained in ALF and GLRH.

Single-copy orthologs were identified from the gene dataset with isoforms and transcript variant duplicates removed using OrthoFinder v2.5.4 (Emms and Kelly, 2019). To ensure accurate alignment and avoid out-of-frame indels, we used PAL2NAL v14 and TranslatorX v1.1 (Suyama et al., 2006; Abascal et al., 2010). Estimated rates of ω among gene sites were conducted using two nested site-specific models (M1a-M2a and M7-M8) from CODEML in PAML v4.10.6 (Nielsen and Yang, 1998; Yang et al., 2000; Wong et al., 2004; Yang et al., 2005). These nested models test for selection at any site in the gene with differences in the number of site classes (M1a: 2; M2a: 3; M7: 10; M8:11). The M2a and M8 models include an additional class of sites under positive selection. We performed nested likelihood ratio and chi-squared tests ($p < 0.05$) based on likelihood scores from each model. Specifically, we compared null models (M1a and M7), which do not allow for any sites with $\omega>1$, against the alternative models (M2a and M8), which permit positive selection. Genes identified as undergoing positive selection were clustered into euKaryotic Orthologous Groups (KOG) and used to perform enrichment of positive selection at the GO functional level using the R package “clusterProfiler” v4.6.2 (settings: pvalueCutoff = 0.01, pAdjustMethod = "BH", qvalueCutoff = 0.01, minGSSize = 10) (Tatusov et al., 2003; Yu et al., 2012; Huerta-Cepas et al., 2019).

3.3.7 Localization of symbiosis support genes

To identify the prevalence of putative symbiosis-related genes (Table S3) in our leafhopper genome evolution analyses, we used genes identified in Mao et al. (2018). Mao et al., 2018 identified 118 genes that are involved in nutrition synthesis, information processing, population regulation, and metabolite transport

in *Nasuia* and *Sulcia* bacteriocytes. We mapped these genes to the shared gene clusters identified between leafhoppers (see Comparative genomics and orthologue analysis) and to genes under positive selection (see Selection analysis in leafhopper genes) using NCBI-tBLASTX (setting: e-value = 0.001).

Finally, one major evolutionary modality to support ancient obligate symbionts is the acquisition of horizontally transferred genes (HTGs) from various bacterial families. To investigate evolutionary conservation of HTGs from Mao et al. (2018) across all four leafhopper species, we used an NCBI-tBLASTX search (setting: e-value = 0.001) of ALF HTG transcripts to each of the leafhopper chromosomal-level sequences to find homologous genes (Table S2). We identified conserved genes across the four species as matches with >40% identity and >100-bit score. We also identified possible HTG remnants as matches with >40% identity and a bit score between 90–100.

3.4 RESULTS

3.4.1 Assembly and Annotation of ALF's Genome

PacBio sequencing of ALF resulted in 6,333,492 reads and 58.2 GB of PacBio HiFi circular consensus sequencing reads at 45x coverage. After scaffolding with the PacBio assembly and OmniC reads, the final assembly included 1,164 scaffolds (total size = 1.3 GB) with an N50 score of 116.5 MB (Table 2). All 1,164 scaffolds were larger than 1 KB and consisted of nine scaffolds larger than 78.3 MB (total = 1.12 GB). These nine scaffolds are interpreted to be the complete ALF chromosomes (Li et al., 2022).

The final BUSCO score for the ALF assembly was 96.08% (83.92% single, 12.16% duplicated, 1.57% fragmented, and 2.35% missing). Using the NCBI RefSeq annotation pipeline, 43.54% of the genome was masked due to repetitive elements by WindowMasker. A total of 24,178 genes (21,979 protein-coding genes and 103 pseudogenes) were annotated across all 1,164 scaffolds. Among the nine chromosome scaffolds, 19,395 protein-coding genes were identified, ~88% of all annotated protein-coding genes.

3.4.2 Comparative genomics and ortholog analysis

To understand general patterns of leafhopper genome evolution, we investigated genomic synteny and gene content. At a structural level, leafhopper chromosomal structures are generally syntenic, with some large exceptions (Figure 1). For example, there are two notable chromosomal rearrangements between the Deltocephalinae leafhoppers (ALF and GRLH). First, there is the large-scale chromosomal rearrangement between two ALF chromosomes (ALF chr1 and ALF chr2) and two GRLH chromosomes (GRLH chr6 and GRLH chr8). Second, ALF chr6, ALF chr8, and GRLH chr3 have had a chromosomal

fusion/fission event between ALF chr6, ALF chr8, and GRLH chr3. There is also evidence of a chromosomal fusion/fission event between TGLH and GWSS, as well as chromosomal rearrangements between the two species, among other shuffling of smaller chromosomal segments.

At a gene level, ALF, GRLH, TGLH, and GWSS have 14,567 orthogroups comprising 68,553 genes (89.1%). All four leafhoppers share 3,450 single-copy gene (1:1:1) orthologs. ALF exhibited the highest number of species-specific genes (2,675) and the lowest number of unassigned genes (1,129) compared to TGLH (1,848 species-specific genes and 4,024 unassigned genes), GWSS (2,300 species-specific genes and 1,895 unassigned genes), and GRLH (812 species-specific genes and 1,301 unassigned genes).

Shared gene analysis across the four leafhoppers species was extended to consider gene clusters and includes gene duplications. Using OrthoVenn, we found 7,723 gene clusters shared between all four leafhoppers (Figure 2). These shared gene clusters showed enrichment for 10 Gene Ontology (GO) processes (peptidoglycan catabolic process, GO:0009253; DNA integration, GO:0015074; transposition DNA-mediated, GO:0006313; oxidoreductase activity, GO:0016705; trehalose transport, GO:0015771; RNA-directed DNA polymerase activity, GO:0003964; translation, GO:0006412; response to bacterium, GO:0009617; proteolysis, GO:0006508; and, telomere maintenance, GO:0000723) (Figure 3a; Table S4). Interestingly, the oxidoreductase activity category consisted of cytochrome P450 enzyme duplications (103 genes). Similarly, the trehalose transport category exhibited multiple duplications across species of the facilitated trehalose transporter Tret-1.

TGLH (Subfamily: Typhlocybinae) is the only sequenced leafhopper species in this analysis that does not have known obligate symbiotic associations. Therefore, we identified the shared gene clusters between the three leafhopper species with *Sulcia*: ALF, GWSS, and GRLH. In the three *Sulcia*-associated leafhopper species, we found 986 unique gene clusters enriched for four GO categories (nucleoside triphosphate biosynthetic process, GO:0009142; carboxylic ester hydrolase activity, GO:0052689; tRNA metabolic process, GO:0006399; and, succinate metabolic process, GO:0006105) (Figure 3b; Table S4). We further identified shared gene clusters between the two Deltocephalinae leafhoppers, ALF and GRLH, that have a symbiotic association with *Sulcia*'s ancestral symbiont partner, *Nasuia*. In these two host species, we found 789 unique gene clusters enriched for nine GO categories (peptidoglycan catabolic process, GO:0009253; positive regulation of cytolysis in other organism, GO:0051714; response to bacterium, GO:0009617; regulation of circadian sleep/wake cycle, sleep, GO:0045187; peptidoglycan metabolic process, GO:0000270; Toll signaling pathway, GO:0008063; regulation of inflammatory response, GO:0050727; microtubule-based process, GO:0007017; and, apoptotic DNA fragmentation, GO:0006309) (Figure 3c; Table S4). Finally, we

identified 598 unique gene clusters specific to ALF that are significantly enriched for four GO categories (telomere maintenance, GO:0000723; receptor-mediated endocytosis, GO:0006898; DNA integration, GO:0015074; response to peptide hormone, GO:0043434) (Figure 3d; Table S4).

3.4.3 Gene family expansion and contraction in ALF's genome

We investigated gene family expansions and contractions across evolutionary time using a time-calibrated phylogenetic tree generated from 469 shared single-copy orthologs between 20 insect species. Briefly, the estimated divergence time for the common ancestor of all four species is ~150 MY (Confidence Interval (CI): 44-282 MY). Between species, the estimated divergence time for ALF and GRLH was ~67 MY (CI: 39-95 MY), and between GWSS and TGLH was ~57 MY (CI: 0-155 MY) (Figure S1). These divergence times are similar to estimated times in previous studies (Moran et al., 2005; Johnson et al., 2018; Cao et al., 2022).

Gene family evolution analyses identified a total of 1,834 and 663 gene families underwent expansions and contractions, respectively, in the ALF genome (Figure S1). When compared to the other 19 insect species, 67 gene families in ALF were expanded (p-value < 0.05), and 32 gene families were contracted (p-value < 0.05) (Figure S2). ALF exhibited the highest number of significant gene family expansions and the lowest number of contractions compared to GRLH (25 expansions, 82 contractions), GWSS (64 expansions, 53 contractions), and TGLH (52 expansions, 60 contractions) (Figure S2). Expanded families in ALF were enriched for genes involved in the KEGG categories: G-quadruplex DNA unwinding, major facilitator superfamily sugar transporter family, DDE superfamily endonuclease, cathepsin propeptide inhibitor domain, E3 ubiquitin-protein ligase, K02A2.6-like, and baculoviral inhibition of apoptosis protein repeat (Figure S3; Table S5).

3.4.4 Leafhopper genome-wide selection analysis

To understand how selection is broadly shaping leafhopper genome evolution, we investigated site-specific selection across shared single-copy orthologs. Estimated rates of ω among gene sites were conducted using two neutral models (M1a and M7) and two models with an additional class of sites under positive selection (M2a and M8). The stringent M1a-M2a nested model approach identified seven orthologs under positive selection (chi p-value < 0.05; Table S6). These genes include two uncharacterized genes, an IDLSRF-like protein, inhibitory POU protein (POU4F2), synaptotagmin-7 (SYT7), voltage-dependent calcium channel type A subunit alpha-1 (cac), and homeobox protein SI-like (SIX6). In contrast, the M7-M8 analysis identified 1,103 orthologs undergoing positive selection (Table S6). The M7-M8 approach also identified all seven genes found in the M1a-M2a models.

Genes undergoing positive selection were binned by euKaryotic Orthologous Groups (KOG) to infer function . Genes with “unknown function” (KOG Category S) showed the highest number of genes under positive selection (Figure S4). Of genes with known functions, signal transduction mechanisms (KOG Category T) and post-translational modification, protein turnover, and chaperones (KOG Category O) had the highest number of genes under positive selection. In the enrichment analysis, generation of neurons (GO:0048699) exhibited the highest gene ratio (204/1103) of genes undergoing positive selection (Figure 4).

Finally, we investigated recent signatures of positive selection between the 6,198 single-copy orthologs shared between ALF and GRLH, which share *Sulcia* and *Nasuia*. A total of 131 and 2048 genes show signatures of positive selection with the M1a-M2a and M7-M8, respectively (Table S7). Of these genes, 129 were jointly identified by both models. Gene function binning identified unknown function (KOG Category S), signal transduction mechanisms (KOG Category T), and post-translational modification, protein turnover, and chaperones (KOG Category O) as the three most enriched categories (Figure S5). While enrichment analysis at the GO level identified cell morphogenesis (GO:0000902) exhibiting the highest gene ratio (259/2048) of genes undergoing positive selection (Figure 4).

3.4.5 Genome-wide evolution of symbiosis support genes

A major question among symbiotic insects is how their genomes evolve to support obligate symbionts living in their bodies. We mapped the 118 putative symbiosis-related genes identified by Mao et al. (2018) (Table S3) to gene clusters shared between all leafhopper species and found 53 symbiosis-related genes (Table S8). In ALF, these genes are upregulated in either *Sulcia* or *Nasuia* tissues, as well as in body tissues. Generally, their function falls within the categories of DNA replication & repair and aminoacyl tRNA formation. Symbiosis-related genes involved in COA synthesis, NH3 recycling, and PEP synthesis are only found in shared gene clusters from all four species. Within shared gene clusters between *Sulcia*-associated leafhopper species (ALF, GRLH, and GWSS), we found 12 symbiosis genes (Table S8). As expected, these genes are upregulated in *Sulcia* tissues in ALF. In the gene clusters shared by Deltocephalinae leafhoppers (ALF and GRLH), we found 18 symbiosis-related genes (Table S8). Most of these 18 genes are upregulated in *Nasuia* and include population regulation genes. In the analysis of selection across the four leafhoppers, we found nine genes involved in symbiosis to have support for positive selection (Table S9). In the two Deltocephalinae leafhoppers, 20 symbiosis-related genes are undergoing positive selection (Table S10). These genes occur across multiple categories, with eight genes occurring in amino acid transport and aminoacyl-trna formation.

Finally, an important evolutionary mechanism is the incorporation of HTGs, which is widely used in leafhopper systems. Here, we investigated whether HTGs found in the ALF genome by Mao et al. (2018) were conserved across all four leafhopper species (Table S2). Only three HTGs are conserved among all four leafhoppers lineages (*cel-1, cel-2, pel*) (Figure 5). In ALF, *cel-1* and *cel-2* are upregulated in body tissues, and the *pel* gene is upregulated in both bacteriocytes (Mao et al., 2018a). Between the Deltocephalinae species that retain *Sulcia* and *Nasuia*, GRLH retains a total of 27 of 30 HTGs found in ALF (Figure 5). GRLH is missing two *Nasuia* symbiont support genes and one non-symbiotic gene. The sharpshooter genome, GWSS that has replaced *Nasuia* with *Baumannia*, retains 11 *Sulcia* and non-symbiotic support HTGs, as was found previously (Mao et al., 2018a; Li et al., 2022). Finally, in the TGLH genome, which has lost all ancestral symbionts, we identified remnants of three genes in TGLH (*gh25-2* (40% identity, 85 amino acids)*, ileS* (60% identity, 60 amino acids)*,* and *yebC-1* (43% identity, 94 amino acids) and one gene in GWSS (*tmk* (59% identity, 67 amino acids)).

3.5 DISCUSSION

We present the first fully sequenced genome of the aster leafhopper (ALF), *Macrosteles quadrilineatus*. To better understand how leafhopper genomes evolve, particularly in relation to evolutionary pressures from their pest ecology and symbiotic biology, we compared the genomes of all available leafhopper species in an evolutionary framework. To date, only three other leafhopper chromosomal-level genomes exist: *Neophotettix cincticeps* (GRLH), *Homalodisca vitripennis* (GWSS), and *Empoasca onukii* (TGLH) (Yan et al., 2021; Li et al., 2022; Zhao et al., 2022). These species are all plant pests, spanning over 150 million years of leafhopper evolution. They further have distinct symbiotic modalities, ranging from retention of leafhoppers' ancestral symbionts *Sulcia* and *Nasuia* (ALF and GRLH), symbiont replacement of *Nasuia* (GWSS), and complete loss of both obligate symbionts (TGLH) (Table 1) (Moran et al., 2003; Noda et al., 2012; Bennett & Moran, 2013; Cao & Dietrich, 2022). This taxon sampling uniquely places an understanding of ALF genome evolution in a framework spanning deep divergences among leafhoppers, symbiotic interactions, and parallel pest species biology and ecology.

3.5.1 Genome evolution of leafhoppers

The inferred size of the ALF genome is 1.3 Gb with a BUSCO score of 96.08%, suggesting a complete high-quality assembly (Table 2) (Yan et al., 2021; Li et al., 2022; Feron and Waterhouse, 2022). ALF's genome encodes 21,979 protein-coding genes, which is in alignment with previous leafhopper genomes (*e.g.,* GWSS has 19,904 genes, and TGLH has 19,642 genes; Li et al., 2022; Zhao et al., 2022). This genome further contains a relatively high number of repetitive elements (*e.g.,* tandem repeats and interspersed repeats) in the genome (43.54%), aligning with the range observed in other hemipteran insects

(leafhoppers: 39.4-46.1%; psyllids: 43.3%) (Yan et al., 2021; Li et al., 2022; Kwak et al., 2023).

At the chromosomal level, ALF and the other sequenced leafhoppers exhibit architectural conservation as evidenced by their relatively high level of synteny (Figure 1). This pattern differs substantially from those observed in other Hemipteran groups, such as aphids, that exhibit extensive autosomal rearrangements despite shorter divergence times (<33 MY) (Li et al., 2020; Mathers et al., 2021). Nevertheless, leafhoppers, like sequenced planthoppers and aphids, demonstrate large-scale conservation of the X chromosome that has little to no rearrangements with autosomes (Li et al., 2020; Ma et al., 2021; Mathers et al., 2021). Strong selection pressures against sex chromosome rearrangements may arise from constraints to maintain dosage compensation mechanisms and avoid X chromosome elimination errors in XO male determination (Sharp et al., 2002; Pal and Vicoso, 2015). In the context of leafhoppers that display this XO sex determination, these selection pressures may lead to a lack of sex chromosome rearrangement (Halkka, 1960; Hu et al., 2022). Despite distinctive genomic conservation in leafhoppers, their chromosomes still exhibit some level of architectural plasticity. For example, the relatively closely related ALF and GRLH species exhibit large-chromosomal fusion or fission events. Comparison among the other species shows similarly large chromosome structure and organization changes.

At the gene level, leafhopper genomes exhibit widespread functional conservation, likely due to fundamentally shared biological and ecological leafhopper traits (Figure 2). For instance, among shared Gene Ontology (GO) categories, translation is most enriched in unique gene clusters, or gene orthologs that include duplications (88 gene clusters; 385 genes; Table S4), indicating a highly conserved biological need for these cellular functions in these species. Leafhoppers also exhibit enrichment for functional categories involved in pest ecology, including the detoxification of insecticides (*e.g.,* cytochrome P450 monooxygenases and carboxylesterases), as well as in categories that have a role in maintaining symbionts (Figure 3). The conserved enrichment of these categories points to the ways in which shared evolutionary pressures, the agricultural environment, and dependence on obligate symbioses have the potential to influence gene content in genomes.

Leafhopper genomes have also experienced extensive proliferation of selfish genetic elements. Several categories associated with the movement and integration of genetic elements within genomes, such as transposition (23 genes), DNA integration (191 genes), and RNA-directed DNA polymerase activity (56 genes), are enriched. Genetic elements can contribute to gene duplications and the general expansion of other gene families (Finnegan, 1989). Thus, they can enable duplication of advantageous genes, including those that respond to environmental changes such as host plant chemical defenses and climate

variation, particularly insecticide resistance (Aminetzach et al., 2005; González et al., 2008, 2010; Rostant et al., 2012; Schrader et al., 2014; Stapley et al., 2015; Rech et al., 2019; Gupta and Nair, 2022).

At the molecular level, positive selection is acting upon multiple single-copy orthologs across various functional categories (Table S6; Figure S4). One such category is signal transduction mechanisms (KOG Category T), with many genes undergoing positive selection (170 out of 401 analyzed signal transduction genes). Signal transduction mechanisms include functions such as odorant receptors, gustatory receptors, and toll-like receptors (Benton et al., 2007; Leulier and Lemaitre, 2008; Engsontia et al., 2014). These genes are likely evolving in response to environmental pressures, possibly in association with leafhopper food-plant finding and oviposition (Smadja et al., 2009).

3.5.2 The potential influence of agricultural pest ecology on leafhopper genome evolution

The lifestyle of plant pests, such as leafhoppers, exposes them to strong selective pressures, including various insecticides, fluctuating environmental conditions, and changes in plant chemical defenses (Nauen and Denholm, 2005; Després et al., 2007; Khaliq et al., 2014; Dumas et al., 2019; Spencer and Hughson, 2023). One evolutionary strategy to adapt to these pressures is the duplications or expansions of genes involved in stress responses (Stark and Wahl, 1984; Tabashnik, 1990; Rostant et al., 2012; Schrader et al., 2014). These genes can play dual roles as insect detoxification enzymes that break down plant toxins and insecticides (Després et al., 2007). Several specific genes and gene family expansions, discussed below, have clear roles in selective pressures placed on leafhoppers from their agricultural pest ecology.

Among the leafhopper species examined in this study, their shared genes showed enrichment for functions involved in oxidoreductase activity (103 genes) and trehalose transport (101 genes), both of which enable adaptation to environmental stressors (Kikawada et al., 2007; Heidel-Fischer and Vogel, 2015) (Figure 3a). For example, genes involved with oxidoreductase activity (*e.g.,* cytochrome P450) contribute to the detoxification of xenobiotics, such as insecticides, environmental pollutants, and plant secondary compounds (Peng et al., 2017). Specifically, these detoxification genes play dual roles in aiding the insect's response to plant defense compounds and resistance to insecticides like pyrethroids and neonicotinoids (Schuler, 1996; Heidel-Fischer and Vogel, 2015; Lu et al., 2021). Leafhoppers further show enrichment for genes associated with trehalose transport (*e.g.,* facilitated trehalose transporter *Tret1*) to different insect tissues (Kanamori et al., 2010). Trehalose is the primary sugar in insect hemolymph and is highly regulated in order to maintain sugar levels (Wyatt and Kalf, 1957; Becker et al., 1996). It is involved in energy production and insect growth, as well as an important bioprotectant under various stressors, including

heat, oxidation, cold, dryness, and hypoxia (Elbein et al., 2003; Kikawada et al., 2007; Kanamori et al., 2010; Liu et al., 2013; Tellis et al., 2023).

At the molecular level, >1000 shared single-copy genes show evidence of positive selection, including genes associated with insecticide resistance. For example, the enriched GO category undergoing positive selection, generation of neurons, includes two voltage-dependent sodium channel genes and four voltage-dependent calcium channel subunit genes (Figure 4a). These ion channels are targeted by various insecticide classes like dichlorodiphenyltrichloroethane (DDT), diamides (*e.g.,* chlorantraniliprole), and pyrethroids (*e.g.,* permethrin) (Silver et al., 2014; ffrench-Constant et al., 2016). While DDT has been banned in the United States since 1972, diamide insecticides (*e.g.,* chlorantraniliprole and flubendiamide) and pyrethroid insecticides (*e.g.,* permethrin, cypermethrin, deltamethrin) are still approved for use in the United States (US EPA, 2014). Moreover, detoxification gene families involved in insecticide resistance, such as cytochrome P450 enzymes, esterases, and carboxylesterases, are also under positive selection (Montella et al., 2012; Cui et al., 2015). Beyond agriculture-specific stressors, leafhopper genes show adaptation to broader environmental pressures that include heat shock proteins (*i.e., HSPA12A* and *HSPA5)* and peroxiredoxin gene (*i.e., PRDX6*) protective against thermal and oxidative extremes (Feder and Hofmann, 1999; Radyuk et al., 2001). Taken together, these results suggest that pest leafhopper genomes adapt to agricultural and environmental challenges across multiple scales.

3.5.3 The potential influence of microbial symbiosis on leafhopper genome evolution

Symbioses with obligate, ancient symbionts has had significant impacts on leafhopper and related insect genomes (Nikoh et al., 2010; Hansen and Moran, 2011; Kim et al., 2011; McCutcheon and von Dohlen, 2011; Nygaard et al., 2011; Price et al., 2011; Husnik et al., 2013; Ratzka et al., 2013; Mao et al., 2018; Couret et al., 2019). The impact of symbioses on host genomes depends on the interaction modality, symbiont identity, and genetic capabilities of partner microbes. Leafhoppers have relied on two ancient obligate partners, but these relationships have changed within and between lineages over evolutionary time (Moran et al., 2005; Koga et al., 2013a; Łukasik et al., 2018; Bennett, 2020). As such, we identified 53 shared gene clusters among all four leafhopper species that contain genes associated with symbiont support and integration (Table S8). Among the leafhopper species compared in this study, however, there have been significant changes in symbiont relationships outlined below.

Two leafhopper species in this study, ALF and GLRH, share the ancestral symbionts, *Sulcia's* and *Nasuia*, which may have been present with *Sulcia* since the origin of leafhoppers and possibly the Cicadomorpha and Auchenorrhyncha

(Koga et al., 2013; Bennett & Mao, 2018). Due to their age, *Sulcia* and *Nasuia* genomes are among the smallest known genomes from any system, requiring extensive host support (Bennett & Moran, 2013). To provide this support, ALF distinctly expresses thousands of previously identified host genes in symbiont-containing bacteriocytes and bacteriome organs (Mao et al., 2018a). ALF and GRLH genomes are enriched for the GO categories peptidoglycan catabolic process (208 genes), positive regulation of cytolysis in another organism (47 genes), and response to bacterium (59 genes). Concurrently, there is enrichment in the functions peptidoglycan metabolic process (12 genes), apoptotic DNA fragmentation (4 genes), and Toll signaling pathway immune response (15 genes) for defense against pathogenic microbes (Hoffmann et al., 1999; Schauvliege et al., 2007; Bao et al., 2014). Furthermore, among these ALF and GLRH genes, some show evidence of recent positive selection, including enrichment of multiple morphogenesis-related functions (Figure 4b). We identified 20 genes undergoing positive selection, which may be an underestimate due to the evolutionary distance and saturation among evolving genes in these species. Nevertheless, the highest count of positively selected symbiosis-related genes fell within the amino acid transport (5 genes) and aminoacyl-tRNA formation categories (3 genes), underling their key role in the evolution of these leafhoppers' symbiotic relationships (Table S10).

Similarly to the deltocephaline species (i.e., ALF and GLRH), the GWSS leafhopper shares the ancestral *Sulcia* symbiont (Moran et al., 2005; Johnson et al., 2018). However, this lineage has completely replaced *Nasuia* with *Baumannia* >60 MYA (Moran et al., 2003; Wu et al., 2006). Several recent projects have identified 1000s of host support genes expressed among the bacteriocytes and bacteriomes that house *Sulcia* and *Baumannia* (Mao et al., 2018a; Li et al., 2022). Given *Baumannia*'s relative youth, it requires far fewer support genes than *Sulcia* and *Nasuia* (reviewed in Li et al., 2022). Nevertheless, among these *Sulcia*-associated leafhopper species, we confirmed global enrichment of GO categories linked to multiple metabolic processes, such as nucleoside triphosphate biosynthetic process (31 genes), tRNA metabolic process (10 genes), and succinate metabolic process (10 genes) (Table S4). Enrichment in these genes highlights the shared dependence of *Sulcia* symbionts on their hosts. *Sulcia* symbionts generally have among the smallest, most dependent genomes (Moran et al., 2005; Takiya et al., 2006a; McCutcheon and Moran, 2010b; Chang et al., 2015; Łukasik et al., 2018; Shih et al., 2019). The same mechanisms of support are required to sustain this ancient and indispensable association among leafhoppers and likely also the Auchenorrhyncha (Campbell et al., 2018; Michalik et al., 2021; Gossett et al., 2023).

Finally, species in the typhlocybinae leafhopper lineage have gone further and purged all obligate bacterial symbionts (Buchner, 1965; Cao and Dietrich, 2022). It has remained an open question as to what happens to host genomes when

they lose symbionts and can genomic paleontology reveal remnants of these associations. Remarkably, despite having lost its obligate symbionts, the TGLH species still indeed retains remnants of genes inferred to support *Sulcia* and *Nasuia* in other leafhoppers (*e.g.*, Mao et al., 2018; Mao and Bennett, 2020; Li et al., 2022). For example, symbiosis-related genes involved in aminoacyl-tRNA formation (13 genes), amino acid transport (12 genes), and DNA repair and replication (9 genes) are retained in all four leafhoppers (Table S8). The retention of these genes in TGLH might indicate the retention of vestigial genes or the functional adaptation and reassignment of these genes for alternate host functions. Their retention further provides direct support for the early loss of *Sulcia* and *Nasuia* from the Typhlocybinae lineage when it transitioned to a more nutrient-rich food source (Buchner, 1965; Moran et al., 2005; Bennett, 2020; Sudakaran et al., 2017).

3.5.4 Horizontally transferred genes in leafhoppers

Horizontally transferred genes (HTGs) present the opportunity for an organism to gain novel traits (Husnik and McCutcheon, 2018). This appears to be a common mechanism to support symbioses in insects, particularly leafhoppers and other hemipterans (Nikoh et al., 2010; Husnik et al., 2013; Sloan et al., 2014a; Luan et al., 2015). ALF has 30 HTGs from various bacterial origins, most of which were inferred to support either *Sulcia* or *Nasuia* (Mao et al., 2018). GRLH retains most of the HTGs present in ALF, with notable differences (Figure 5). By contrast, GWSS retains only 11 HTGs that are also present in ALF, most of which are predicted to support *Sulcia* (Mao & Bennett, 2020; Li et al., 2022).

A major focus of HTG analysis was the fate of these genes in species that have lost (TGLH) and replaced more ancient symbionts (GWSS). TGLH and GWSS retain orthologous HTGs involved in other host functions, including cell wall degradation (*cel-2, cel-1,* and *pel*). However, while GWSS has lost those related to *Nasuia*, TGLH has lost all full-length symbiosis support genes found in the other leafhopper species. To investigate whether remnants of lost ancestral HTGs in GWSS and TGLH, we scanned their genomes for gene fragments (Queffelec et al., 2022). Remarkably, we found remnants of three genes in TGLH used to support *Sulcia* (*ileS and yebC-1*) and *Nasuia* (*gh25-1*) in ALF (Mao et al., 2018a). We further found a remnant of the cell growth gene, thymidylate kinase (*tmk*), in GWSS, which also supports *Sulcia* in ALF (Chaperon, 2006). We note that the specific function of these HTGs in leafhoppers remains untested and is inferred from their identities and specific expression patterns in ALF.

3.5.5 ALF's distinct genomic traits highlight its polyphagous ecology

Species-specific gene groups and gene family expansions in ALF can help to identify adaptations to its unique biology and ecology. As an illustrative point, ALF's genome has uniquely experienced gene family expansion involved in

sugar transport and protease genes (Rispe et al., 2008; Price et al., 2010). Enrichment in these specific genes likely underly the differences between ALF's host plant associations and the other species investigated here. ALF is a massively polyphagous pest that feeds on a sugar-rich phloem diet across 100s of plants. GRLH feeds narrowly on some monocot species (*i.e.,* rice and grasses), while GWSS is restricted to the sugar-depauperate xylem saps (Turner and Pollard, 1959; Pathak, 1968). TGLH is a monophagous pest that feeds on leaf parenchyma cell contents in tea plants (Günthardt & Wanner, 1981; Kawai, 1997; Qin et al., 2015). We speculate that the feeding habit of ALF as a polyphagous pest on dicot plants may lead to an expansion of these sugar-related genes.

3.5.6 Conclusion

The completion and analysis of the ALF genome assembly has unveiled new insights into leafhopper genomics and the broad evolutionary history of this group. Our study underscores the intricate balance between genomes, symbionts, and the environment in the field of insect pest evolution and adaptability. Comparisons among the genomes of four pest species highlights the environmental pressures, including from insecticide treatments, farm cropping strategies (*e.g.,* mono-cropping seasonality and cover crops), climate and its changes, and a range of plant secondary defensive compounds (Trenbath, 1993; Nauen and Denholm, 2005; Després et al., 2007; Khaliq et al., 2014; Dumas et al., 2019; Bai et al., 2022; Spencer and Hughson, 2023). The ability of these species to persist in these environments further depends on their obligate associations with a diversity of bacterial symbionts. These relationships similarly place significant evolutionary pressures on host genomes and further are likely to shape the environmental interactions of insects (Brodbeck et al., 1990, 2014; Bennett and Moran, 2015). Major transitions in symbiotic relationships among leafhoppers has left strong signatures of evolution, affecting the expansion and contraction of gene families, as well as the fate of ancient HTGs.

Taken together, the addition of the ALF genome to the insect genome resources contributes to our broader understanding of insect genomics, their evolutionary adaptations, and potential impacts on their ecology and symbiotic relationships. Since ALF is a major pest of North American agriculture (Wallis, 1962; Beanland et al., 2005), these results can be used to inform and develop modern pest management strategies such as RNAi gene expression interference (Baum and Roberts, 2014; Jain et al., 2021). ALF's shared and unique genomic features identified in this study, including HTGs, selection events, and symbiosis-related genes, can potentially be targeted for disrupting the survival of these insects.

3.6 TABLES

Species Name	Shorthand Naming	Genome Size	Subfamily	Feeding	Symbionts
Macrosteles quadrilineatus	ALF	1.3 Gb	Deltocephalinae	Polyphagous	*Sulcia*, *Nasuia*
Neophotettix cincticeps	GRLH	746 Mb	Deltocephalinae	Polyphagous	*Sulcia*, *Nasuia*
Homalodisca vitripennis	GWSS	2.3 Gb	Cicadellinae	Polyphagous	*Sulcia*, *Baumannia*
Empoasca onukii	TGLH	599.3 Mb	Typhlocybinae	Monophagous	None

Table 3.1: Four publicly available leafhopper chromosomal-level genomes used for comparative genomics analysis.

Feature	Value
Assembly size	1,317,891,973 bp
Assembly size of chromosomes	1,127,407,133 bp
Scaffold N50	116,548,684 bp
Scaffold L50	5
Number of scaffolds	1,164
Number of total genes	24,178
Number of protein-coding genes	21,979
Number of pseudogenes	103

Table 3.2: Genome statistics of the *Macrosteles quadrilineatus* genome.

3.7 FIGURES

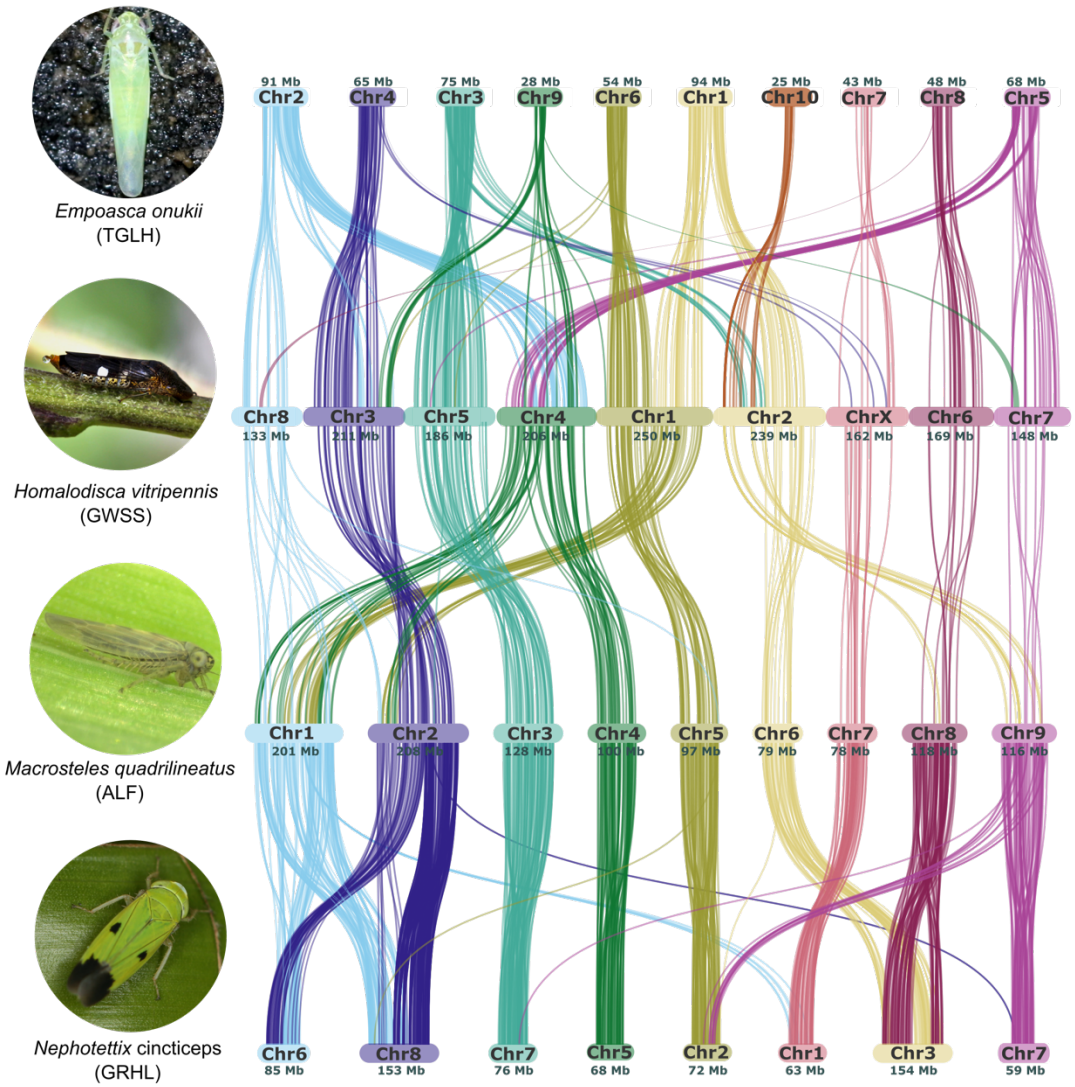


Figure 3.1: Chromosomal synteny between the four leafhopper species. Images of *H. vitripennis* and *M. quadrilineatu*s courtesy of Dr. Ziheng Li and Dr. Xiushuai Yang, respectively. Images of *E. onukii* and *N. cincticeps* are from www.inaturalist.org (from the following users: straybird726 and bob15noble, respectively. All photos have CCBY-NC license).

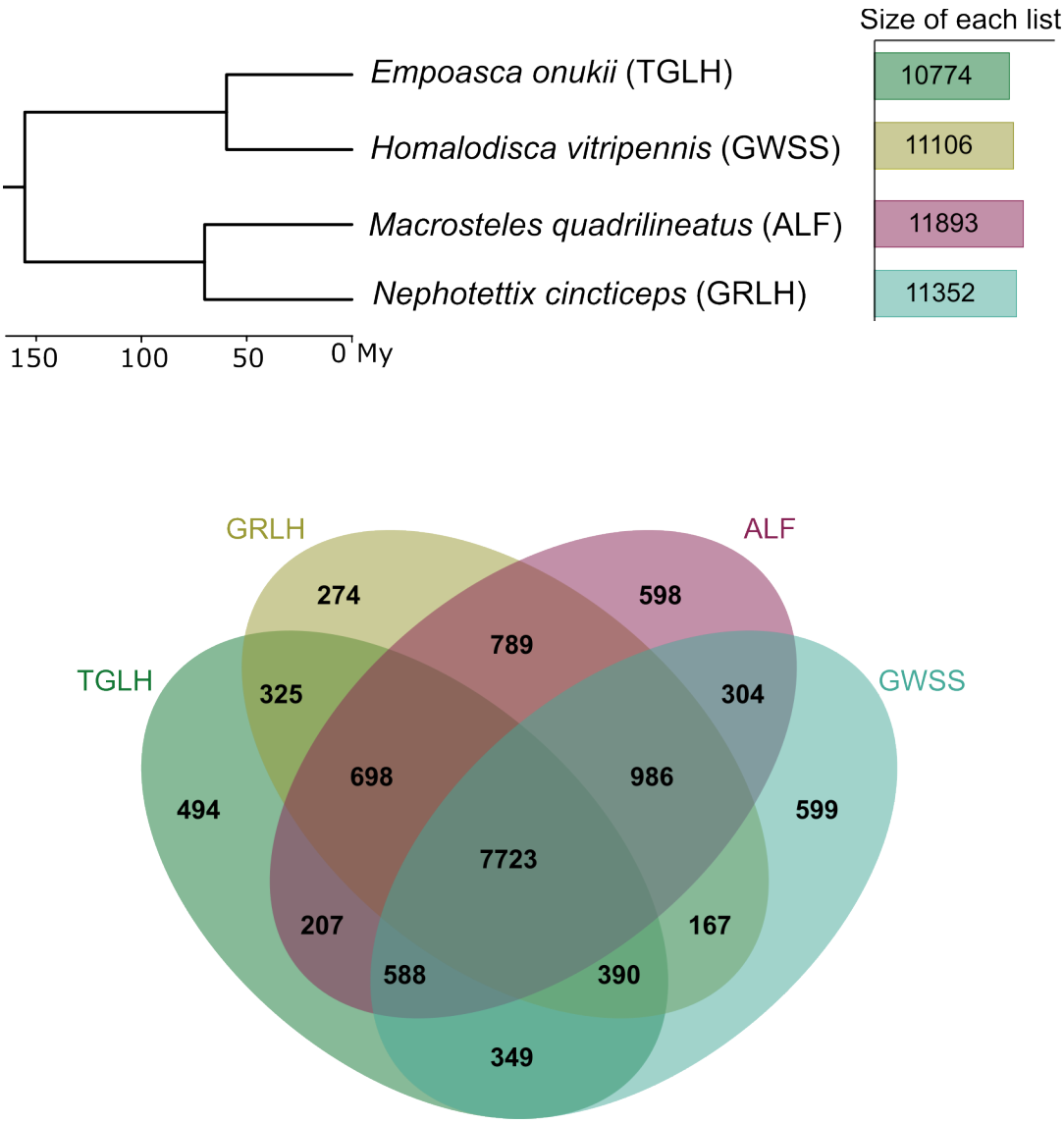


Figure 3.2: Shared gene clusters between leafhopper species.
Venn diagram of shared gene clusters, that is clusters that include duplications of genes, among the four species. My = million years.

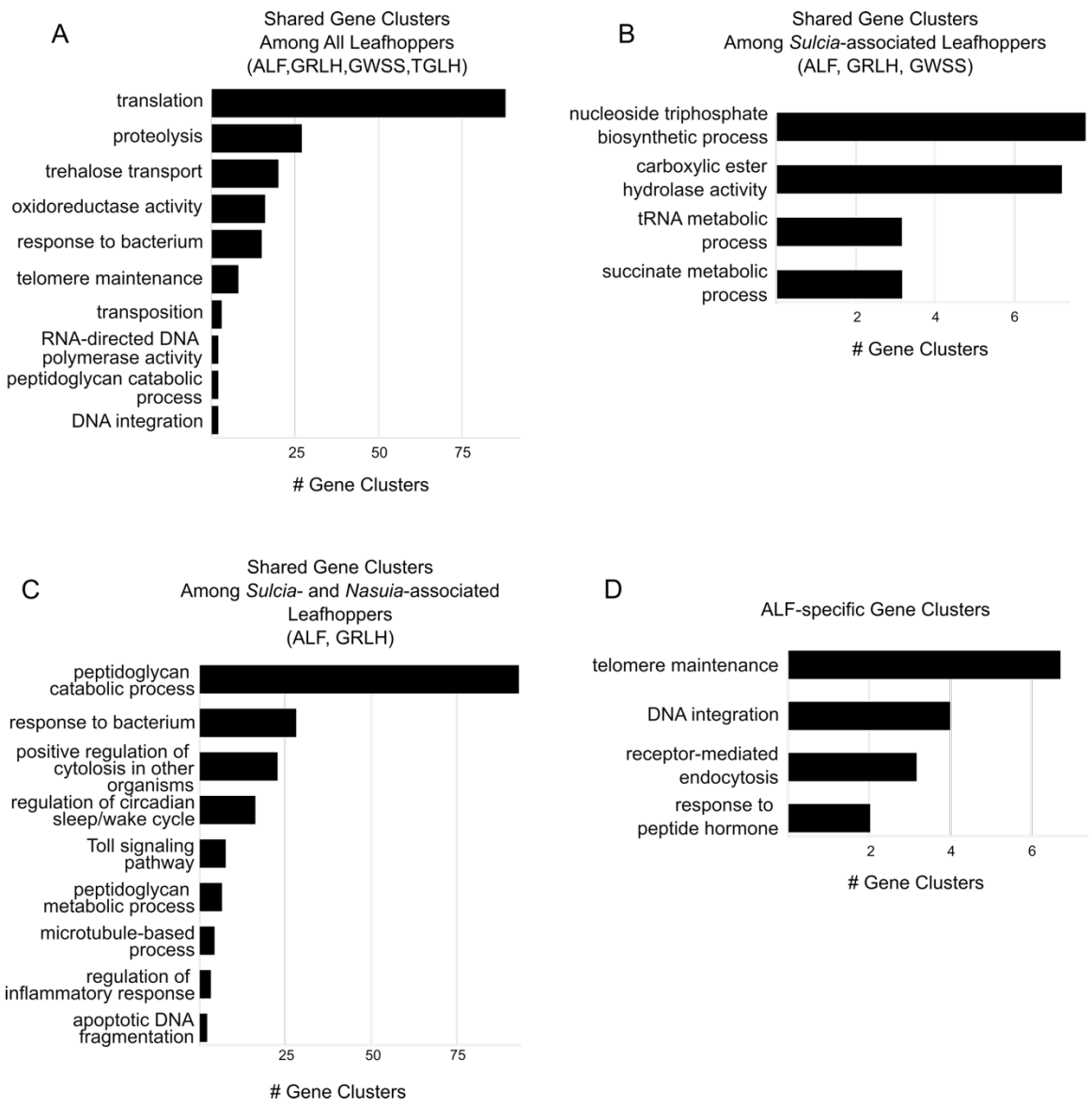


Figure 3.3: Functional enrichment among shared gene clusters.
A) Enrichment among gene clusters shared between all leafhopper species in this study. B) Enrichment among gene clusters shared between leafhoppers that have an association with the symbiont, Sulcia. C) Enrichment among gene clusters shared between Deltocephalinae leafhoppers, which have an association with both ancient symbionts Sulcia and Nasuia. D) Enrichment in gene clusters unique to ALF. GO categories and number of unique gene clusters can be found in Supplemental Table 4. ALF = Macrosteles quadrilineatus, GRLH = Nephotettix cincticeps, GWSS = Homalodisca vitripennis, TGLH = Empoasca onukii.

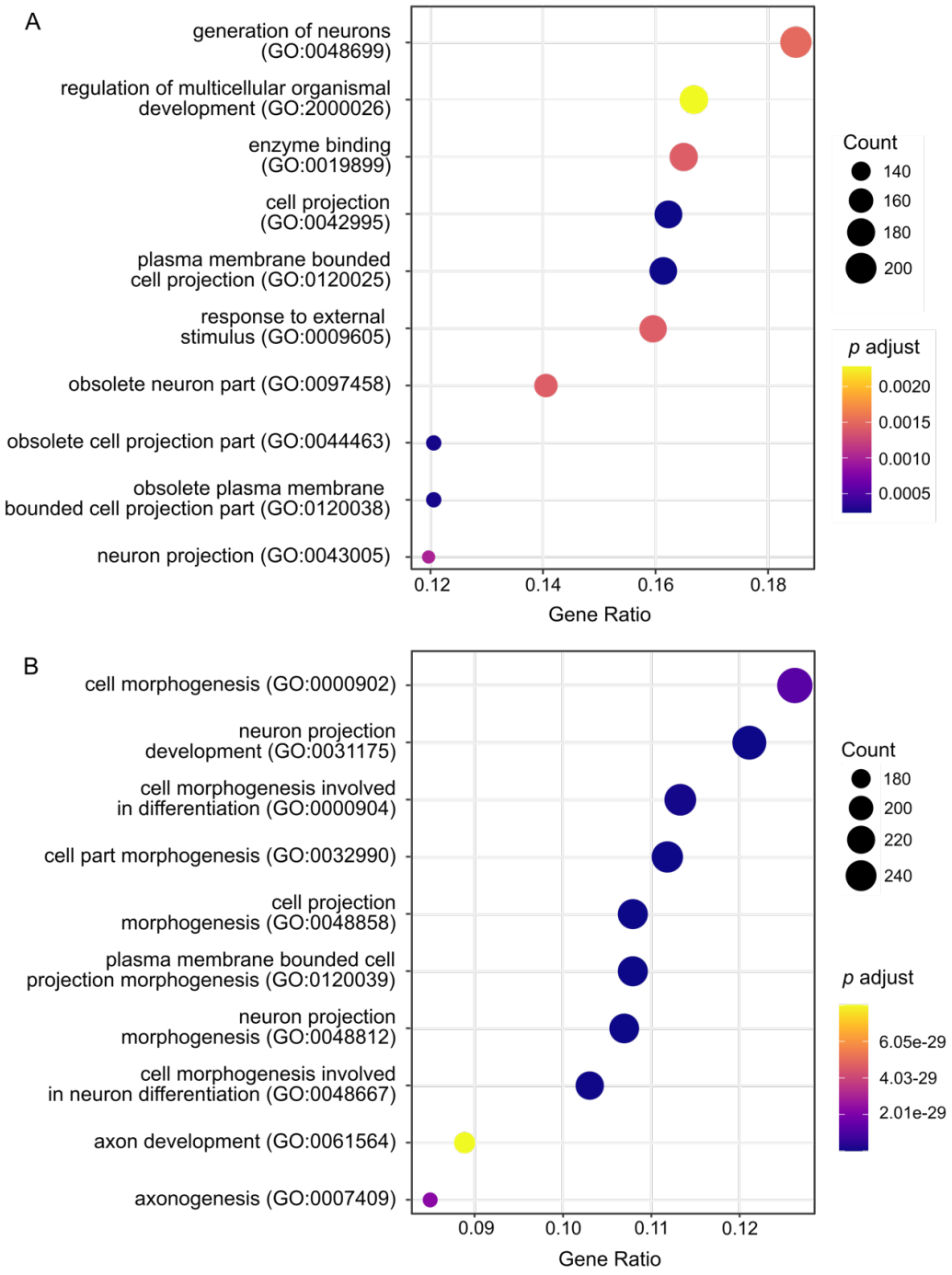


Figure 3.4: Enriched GO categories of positively selected genes.
A) Top 10 GO categories that exhibit enrichment in positive selection in the analysis for four leafhoppers (ALF, TGLH, GRLH, and GWSS). B) Top 10 GO categories that exhibit enrichment in positive selection in the analysis for two leafhoppers (ALF and GRLH).

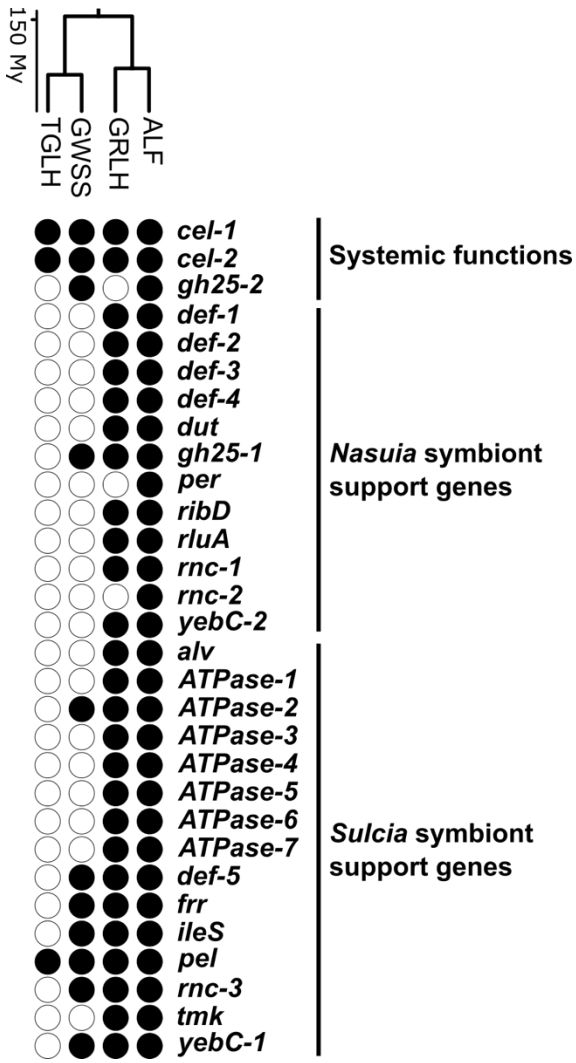


Figure 3.5: Conserved horizontally transferred genes (HTG) among leafhoppers. HTGs are grouped by their role in the symbiosis between ALF, Sulcia and Nasuia. Empty circles indicate missing gene, while enclosed circles indicate presence of genes. Gene products can be found in Supplemental Table 3. ALF: Macrosteles quadrilineatus, GRLH: Nephotettix cincticeps, GWSS: Homalodisca vitripennis, TGLH: Empoasca onukii.

Chapter 4: Fluorescent-based cell sorting and epigenetic regulation of symbiont bacteriocytes

4.1 ABSTRACT

Insects, such as leafhoppers, often rely on beneficial symbiotic microbes for essential nutrients missing from their plant-sap diets. These microbes are typically housed within host cells known as bacteriocytes. Many insect species have evolved to accommodate multiple bacterial partners within their own unique and distinct bacteriocyte types. In these cases, hosts tailor their cellular and genetic support within these cells to meet the needs of each symbiont. An emergent model of such symbioses is the aster leafhopper, *Macrosteles quadrilineatus* (Hemiptera: Cicadellidae), which harbors two symbiont species, *Sulcia* and *Nasuia*. The insect host differentially regulates thousands of nuclear-encoded support genes to maintain its symbionts. However, the mechanisms through which a single host genome differentially regulates gene expression between distinct cell types remain poorly understood. To address this question, we developed a fluorescence-based cell sorting technique. We apply this method to test whether differential DNA methylation sequencing is a mechanism by which ALF differentially regulates gene expression among bacteriocyte types. This approach leverages the high ploidy level of bacteriocytes to isolate them based on their differentially enriched fluorescence signal over other insect cell types. We confirmed the successful isolation of bacteriocytes with reverse transcription quantitative-PCR and horizontally transferred genes in the *M. quadrilineatus* genome that exhibit differential expression in *Sulcia* and *Nasuia* bacteriocytes. Using this approach, we discovered that bacteriocytes experience differential methylation compared to non-symbiotic host cell types. We show that the distinct bacteriocyte types that house *Sulcia* or *Nasuia* have distinct methylation patterns; however, the relationship between these methylation patterns and gene expression is complex. Methylation at promoter regions is not linked to gene expression patterns, but increased intronic methylation could indicate interactions with other elements of gene regulation, such as alternative splicing, non-coding RNAs, or chromatin modifications. Nevertheless, our ability to sort and determine differential methylation properties of bacteriocyte genomes demonstrates the power of this approach as applied to complex symbiotic systems.

4.2 INTRODUCTION

Many insect species rely on bacterial symbionts to synthesize missing nutrition from deficient diets. Notably, these insects frequently form close obligate relationships with one or more symbiotic partners (Buchner, 1965; Moran et al., 2008; Bennett and Moran, 2013; Douglas, 2017). This type of symbiosis generally leads to the formation of specialized host cells (bacteriocytes) and host-derived organs (bacteriomes) (Buchner, 1965). Bacteriocytes house bacteria in a stable intracellular environment that provides essential resources

and support to bacteria (Buchner, 1965; Baumann, 2005; Nakabachi et al., 2005b; Hansen and Moran, 2011b; McCutcheon and von Dohlen, 2011; Sloan et al., 2014b; Luan et al., 2015; Mao et al., 2018b; Weglarz et al., 2018; Kobiałka et al., 2018; Fronk and Sachs, 2022). Symbionts, despite their essential nutritional roles, undergo extreme genomic streamlining, losing >90% of genes typically found in their free-living relative—even genes vital to both symbionts and their hosts (McCutcheon and Moran, 2012a; McCutcheon et al., 2019). To support their symbionts, insect hosts often regulate the expression of thousands of genes in each bacteriocyte to meet the needs of their resident bacteria (Hansen and Moran, 2011b; Douglas, 2016; Mao et al., 2018; Ankrah et al., 2020). However, the means by which the host genome can modulate the expression of these genes in two distinct symbiotic cell types remains unclear. One potential mechanism for regulating gene expression involves epigenetic modifications, such as DNA methylation, which can alternately inhibit or promote transcriptional activity without altering the underlying DNA sequence.

Many animals, including hemipteran insects, have the capability to use DNA methylation as an important DNA modification that aids in the regulation of various biological processes (Bewick et al., 2016). In insects, epigenetic molecular mechanisms contribute to invertebrate systems, including the caste system in ants (Bonasio et al., 2012), as well as resilience to different environmental stresses (*e.g.,* pesticides and temperature increase; Dombrovsky et al., 2009; Brevik et al., 2021). In the case of bacterial symbiosis in hemipterans, methylation may play a regulatory role in bacteriocytes during nymphal development and in response to ecological conditions such as different food plant species with distinct plant-sap nutritional profiles (Kim et al., 2018; Pers and Hansen, 2021). However, our understanding of the role that DNA methylation may play in the general differential support of bacterial partners is less known. DNA methylation may inhibit transcriptional activity leading to the regulation of thousands of genes that exhibit differential gene expression in separate cell types (Eden and Cedar, 1994; Ehrlich and Lacey, 2013; Lowdon et al., 2016; Zhang et al., 2018). This epigenetic modification could allow for specific symbiont support expression patterns from a single gene that is differentially used to support different cell types and symbiont species. Understanding differential patterns of host support to symbionts requires the isolation of bacteriocytes from non-symbiotic host cell types. However, the dissection and isolation of these cells can be technically challenging, as they are often embedded deep within host tissues and are often closely associated with other bacteriocyte types within a bacteriome organ (Buchner, 1965; Hand, 1987; Koga et al., 2012; Zhu et al., 2017; Ribeiro Lopes et al., 2021). To better leverage high throughput genomic and gene expression technologies (*e.g.,* whole genome bisulfite sequencing), there is a strong need to develop novel methods for single-cell isolation of host bacteriocyte types.

Cell sorting techniques have emerged as powerful tools for different fields (Barlogie et al., 1983; Rieseberg et al., 2001; Comas-Riu and Rius, 2009). These methods are generally applied to isolate and characterize (*e.g.,* gene expression and immunophenotyping) specific cell populations, thereby overcoming the challenges of studying complex tissues and samples (Walberg et al., 1996; Nakahara et al., 2009; Bhagwat et al., 2018; Rolton et al., 2020; Andrade et al., 2021). To date, cell sorting in bacteriocytes has been used to identify the absolute number of symbiont cells within a bacteriocyte, as well as examine ploidy levels within bacteriocytes (Simonet et al., 2016; Nozaki and Shigenobu, 2022). Despite prior knowledge of the application of flow cytometry in these cells, fluorescence-based cell sorting has yet to be applied to separate bacteriocytes in a multi-partner symbiotic association. Such approaches can permit questions aimed at disentangling the regulation, contributions, and evolution of each partner in tightly entwined multi-partner symbioses.

One advantageous application of single-cell sorting techniques is their ability to naively sort cell types based on the fluorescence of stained DNA rather than distinct fluorescence tagging of specific protein markers (Darzynkiewicz et al., 1999; Ibrahim and van den Engh, 2007; McKinnon, 2018). Differences in the ploidy of some cells provide a unique opportunity to successfully isolation of selected populations (Metcalfe et al., 2019). Given that polyploid cells contain multiple copies of a genome within a single cell, their increased DNA content can be exploited as a distinguishing feature by which they can be sorted. In insects, bacteriocytes often exhibit higher ploidy levels (Braendle et al., 2003; Woyke et al., 2010b; Nozaki and Shigenobu, 2022; Nakabachi and Moran, 2022). By utilizing sorting methods that rely on DNA staining, it becomes possible to differentiate polyploid bacteriocytes from diploid non-symbiotic host cells in insects with multiple symbionts. Approaches such as RT-qPCR of known expressed genes, or total RNA-seq expression profiling, can then be used to validate sorted populations. This combined approach enables the isolation of body and bacteriocyte types, facilitating subsequent analyses of individual symbionts and their interactions within the host.

The aster leafhopper (hereafter known as ALF), *Macrosteles quadrilineatus* (Hemiptera: Cicadellidae), has a complex relationship with two symbiont species, "*Candidatus* Sulcia muelleri" and "*Candidatus* Nasuia deltocephalinicola" (hereafter known as *Sulcia* and *Nasuia*, respectively). These symbionts supplement the leafhopper's nutritionally limited plant-sap diets with essential amino acids that animals are unable to make. *Sulcia* (190 KB) and *Nasuia* (112 KB) have undergone severe genome reduction over deep evolutionary time due to a lack of an external environmental phase, small population sizes, and intergenerational population bottlenecks during vertical transmission (McCutcheon and Moran, 2012b; McCutcheon et al., 2019). As a result, both symbionts have lost genes from all essential cellular processes. They no longer retain genes to synthesize their own cell membrane (Bennett & Moran 2013). In

ALF and similar species, there is a strict separation of *Sulcia* and *Nasuia* within distinct bacteriocyte types. Bacteriocytes have distinct gene-expression patterns tailored to the specific cellular and genomic needs of *Sulcia* or *Nasuia* (E.g., ALF differentially expresses 3,064 and 4,201 genes in *Sulcia* and *Nasuia* bacteriocytes relative to host insect body tissues (Mao et al., 2018). Additionally, multiple genes are found to be up-regulated in both bacteriocyte types (Mao et al., 2018). However, how the host modulates the expression of these genes between the two distinct symbiotic bacteriocyte types is unknown. Additionally, almost nothing is known about the role of epigenetic controls in *M. quadrilineatus* (ALF) and related insects in the auchenorrhyncha.

In this study, we use the unique ploidy properties of bacteriocytes in ALF to identify and isolate cells using fluorescence-based cell sorting. We demonstrate the utility of this technique in downstream analysis by sequencing DNA methylation patterns within isolated populations. Applying our newly developed technique, we compared expression patterns from previous transcriptomic data (Mao et al., 2018) with hyper- and hypo-methylated genes. This combined approach has the ability to identify whether DNA methylation is a mechanism by which insect hosts may differentially control gene expression and which specific genes between cell types.

4.3 MATERIALS AND METHODS

4.3.1 Leafhoppers

The *Macrosteles quadrilineatus* (ALF) insect line used in this study was initially collected in New Haven, Connecticut in 2013 and reared at the University of California, Merced. The established line was maintained on barley grass at 25 °C and under a stable 12L:12D light/dark.

4.3.2 Trypsin isolation

Bacteriomes of insects were isolated on ice and placed in RNAse-free Buffer A solution (25mM potassium chloride, 35 mM Tris, 100 mM EDTA, 250 mM sucrose, vacuum filtered through a 0.22-micrometer filter and stored at 4 °C) during dissection.

We attempted enzymatic digestions using trypsin, which is a common method for obtaining the single-cell suspensions needed for flow cytometry (Stewart et al., 1995). For trypsin isolation trials, three techniques were used. First, we confirmed the optimal trypsin concentration using varying percentages (0.05%, 0.125%, 0.25%, 0.5%, and 1%) of trypsin diluted with 1x PBS at room temperature. In this trial, we used one bacteriome for each percentage. We pipetted up and down gently and visualized cells under a Keyence BZ-X710 microscope every 10 minutes. In the second method, we added 40 ul of 1%

trypsin diluted with 1x PBS to the isolated bacteriomes (n=10). Samples were added to a water bath at 37 °C and gently pipetted up and down every 10 minutes to help dissociation. After 30 minutes in the water bath, cells were spun at 200g for 5 min (4 °C). The supernatant was removed, and 80 ul of buffer (1x PBS 1% BSA) was used to inactivate trypsin and resuspend the pellet. Cells were centrifuged at 200g for 5 min (4 °C) again and resuspended in 0.5 mL of buffer (1x PBS 1% BSA). Cells were placed on ice to be visualized under a Keyence BZ-X710 microscope at 40x. Third, isolated bacteriomes (n=10) and 10 mL of 1% trypsin diluted with 1x PBS was added to a 15-mL Falcon tube. Tubes were shaken at 35 rpm for 30 minutes at room temperature. The reaction was pipetted every 10 minutes to help dissociation. After incubation, cells were washed twice using 5 mL of buffer (1x PBS 1% BSA) and centrifuged at 1000g for 5 min (4 °C). The resulting pellet was resuspended in 1 mL of buffer (1x PBS 1% BSA) and transferred to a 1.5 mL Eppendorf tube. Cells were placed on ice to be either visualized under a Keyence BZ-X710 microscope at 40x or were used to test with ethanol fixation.

4.3.3 Ethanol Fixation

Cells from trypsin dissociation were placed in a 15 mL Falcon Tube, and 3 mL of 70% ice-cold ethanol was added dropwise while vortexing gently. Samples were incubated at 4 °C overnight. For post-fixation recovery, fixed cells were centrifuged at 300g for 10 minutes (4 °C). Cells were washed twice using 5 mL of Galbraith's buffer (45 mM MgCl2, 20 mM MOPS, 30 mM sodium citrate, 0.1% (vol/vol) TritonX-100; Buffer was adjusted pH to 7.0 with 1M NaOH, vacuum filtered through 0.22 um filter and stored at -20 °C in 10 mL aliquots) and centrifuged at 300g for 10 minutes (4 °C) (Galbraith et al., 1983). The supernatant was removed, and cells were resuspended in 500 uL of Galbraith's buffer. Cells were visualized under a Keyence BZ-X710 microscope at 40x.

4.3.4 Host cell and bacteriocyte identification using flow cytometry

To identify fluorescence signals between bacteriocytes and host cells, we first analyzed DAPI fluorescence with non-symbiotic host cell types. Isolated heads of ALF were ground using the frosted label of two microscope slides in 1mL of Galbraith's Buffer (Galbraith et al., 1983). Cells were filtered through a 70-micrometer filter and stained with DAPI (5mg/mL) for 15 minutes on ice. As a control, *Drosophila* heads were isolated and stained as well. DAPI fluorescence was recorded on a FACSAria III (BD). In order to determine the global signal of fluorescence in insect abdomen cells compared to ALF host cells, we implemented the same technique above. We found four distinct cell populations in isolated insect abdomens. Using the knowledge of increased ploidy signal in bacteriocytes, we further investigated two cell populations that exhibited increased DAPI fluorescence compared to non-symbiotic host cell types.

4.3.5 Galbraith's Buffer Triton X-100 isolation

Triton X-100 is a common detergent used to permeabilize the membranes of cells prior to fluorescent staining. Due to its ability to lyse cells at high percentages, we also tested varying percentages of Triton X-100 in Galbraith's buffer (0%, 0.01%, 0.05%, 0.1%). Using ALF heads and isolated abdomens, we ground tissues with the frosted label of two microscope slides in 1mL of Galbraith's Buffer (0%, 0.01%, 0.05%, 0.1% Triton X-100). Cells were stained with DAPI (5mg/mL) for 15 minutes on ice, and fluorescence was recorded for each percentage on a FACSAria III (BD).

4.3.6 Bacteriocyte isolation

After finding the optimum Triton X-100 percentage (0.01%) for Galbraith's buffer isolation and identifying controls of DAPI fluorescence of non-symbiotic host cells, we isolated the bacteriomes of ALF for cell sorting. Isolated bacteriomes were ground using the frosted label of two microscope slides in 1mL of Galbraith's Buffer (Galbraith et al., 1983). Cells were filtered through a 100-micrometer filter and stained with DAPI (5mg/mL) for 15 minutes on ice. Populations exhibiting increased DAPI fluorescence compared to non-symbiotic host cell types were sorted into individual 1.5 mL Eppendorf tubes to isolate bacteriocytes. Population isolation was performed in triplicate. Cells were either photographed under a microscope to determine if bacteriocytes could be identified by their morphological characteristics or undergone DNA extraction for methylation sequencing.

4.3.7 Reverse transcription quantitative-PCR (RT-qPCR) confirmation

We confirmed the presence/absence of *Sulcia*- and *Nasuia*-specific bacteriocyte cells using RT-qPCR with custom primers. RNA was isolated using the Direct-zol RNA kit (Zymo), quantified (Quibit), converted to cDNA, and tested with primers targeting genes up-regulated in *Sulcia* or *Nasuia* bacteriocytes following previously established work (Mao et al., 2018). To identify *Sulcia* bacteriocytes, we used the horizontally transferred gene (HTG) *frr,* which is highly expressed in *Sulcia* bacteriocytes. To identify *Nasuia* bacteriocytes, we used the HTG *ribD*, which is highly expressed in *Nasuia* bacteriocytes. We used the host Ribosomal protein L7 (*rpl7*) to normalize gene expression.

4.3.8 Methylation genome assembly and identification of CpG sites

DNA from assumed bacteriocyte cell populations (Pop3 and Pop4; Figure #), head tissue, and leg tissue was extracted using the Qiagen DNAeasy protocol. Triplicates of each DNA sample were quantified (Quibit) and sent for whole genome bisulfite sequencing at the QB3 Genomics Facility at the University of

California, Berkeley. Sequencing was done using an Illumina NovaSeq S1 150PE.

Using the insect genome generated by Vasquez et al. (in prep), methylation data was aligned to the genome with Bowtie2 and Bismark (Krueger and Andrews, 2011; Langmead and Salzberg, 2012b). Site-specific CpG methylation data for each sample were calculated using the MethylKit package in R (Akalin et al., 2012). Coverage outliers (above the 99.9 percentile) and bases with <10 coverage were removed using *filterByCoverage()*. Coverage values between samples were scaled using the median and the *normalizeCoverage()* function. Samples were pooled, and a logistic regression test was carried out per CpG position per sample group using an overdispersion correction and *p*-values corrected using the Benjamini-Hochberg procedure (overdispersion = “MN”, adjust = “BH”; Benjamini and Hochberg, 1995). A site was classified as differentially methylated between the treatment and control samples if the Benjamini-Hochberg corrected p-value was <0.05 and the overall methylation difference was >10%.

For a global analysis of epigenetic changes across the genome, we identified CpG islands in the chromosomal ALF genome using EMBOSS cpgplot (-window 200 -minlen 200 -minpc 50 -minoe 0.6; Madeira et al., 2022). Flanking sequences 2 KB away from CpG islands were defined as CpG island shores using *readFeatureFlank()* in Methylkit (Akalin et al., 2012). We used a paired t-test to determine whether there is significant percent methylation of CpG sites in CpG islands and shores between bacteriocyte samples and body samples. MethylKit was also used to place differentially methylated sites into a specific genomic context by annotating sites with a genomic annotation of promoter, exon, intron, or intergenic region of the chromosomal level DNA sequence and its distance from the nearest Transcription Start Site (Akalin et al., 2012).

4.3.9 Global patterns of genome methylation

To understand global DNA methylation patterns, we identified differentially methylated CpG sites within promoter regions between both isolated bacteriocyte populations (Pop3 and Pop4) and non-symbiotic host cell types. We further linked differential methylation in promoter regions to gene expression using raw RNA-sequencing files from previously sequenced data (Mao et al., 2018). Briefly, RNA-seq reads were mapped to the chromosomal-level insect genome using HISAT v2.2.1 (Kim et al., 2019). We estimated the number of reads mapped to exons using featureCounts (Liao et al., 2014). Counts were normalized, and differentially expressed genes (>1.5 log fold change between bacteriocytes and non-symbiotic host cells) were identified in R using DESeq2 (test = "Wald", Love et al., 2014; RStudio Team, 2018). To identify whether DNA methylation and gene expression are globally correlated, we used a Pearson’s Chi-squared test with Yates’ continuity correction.

4.3.10 Differential gene-specific DNA methylation regulation of bacteriocytes

Following cell sorting and qPCR confirmation, we identified Pop3 as containing primarily *Nasuia* bacteriocytes. In contrast, Pop4 contained both *Nasuia* and *Sulcia* bacteriocytes (see "qPCR cell sorting confirmation" below for more details). To identify DNA methylation patterns in *Nasuia* bacteriocytes, Pop3 samples were treated as the treatment condition, while body samples were treated as the control. For DNA methylation in *Sulcia* bacteriocytes, Pop4 samples were treated as the treatment condition, while Pop3 and body samples were treated as the control condition. This method removed the signal of *Nasuia* bacteriocytes from Pop4, allowing the specificity of the *Sulcia* bacteriocyte methylation signal. Each pairwise analysis was filtered and normalized using the same approach as above. To identify whether DNA methylation and gene expression in each bacteriocyte type are globally correlated, we used a Pearson's Chi-squared test with Yates' continuity correction. To identify DNA methylation-based regulation of bacterial support genes, we mapped putative symbiont support genes found in Mao et al. (2018) to genes associated with differentially methylated promoter regions. We used a Pearson correlation between genes with an inverse relationship between expression and methylation, that is, hypo-methylation/up-regulation (hypo-up) and hyper-methylation/down-regulation (hyper-down).

4.4 RESULTS AND DISCUSSION

Symbioses between animals and bacteria are often highly complex ,involving multiple partners with distinct roles and evolutionary histories (Moran, 2007; McFall-Ngai et al., 2013). In the most extreme cases, as are common in insects, bacteria are permanently held within specially evolved bacteriocyte host cells (Buchner, 1965; Fronk and Sachs, 2022). To fully understand how these systems function and evolve into tightly integrated systems, new approaches are needed to disentangle them (Woyke et al., 2010b; Kim et al., 2018; Pers and Hansen, 2021). Here, we use the *Macrostel*es *quadrilineatus* system to develop a single-cell sorting technique. We validate this approach and, as a proof of concept, use it to address the question of what role DNA methylation may play in the regulation of insect symbiosis support genes.

4.4.1 Chemical trypsin dissociation is not appropriate for bacteriocyte isolation

Trypsin dissociation, or trypsinization, is a common method for obtaining the single-cell suspensions needed for flow cytometry (Stewart et al., 1995). However, this proteolytic enzyme can lead to cell stress and alter gene expression (Huang et al., 2010). Despite its known effects on cell viability and functionality, trypsinization remains widely used due to its effectiveness in dissociating cells from tissues and organs (Brown et al., 2007; Nozaki and Shigenobu, 2022). Here, we conducted multiple dissociation experiments to find

the optimal percentage of trypsin (0.05%, 0.125%, 0.25%, 0.5%, and 1%) and dissociation time (0-50 minutes) to separate bacteriocyte cells. However, these observations were made on a single bacteriome with cells being isolated, dissociated, and imaged within a short time frame. Given the small size of the leafhopper's bacteriomes (~0.6mm), substantial downstream analysis (*e.g.,* whole genome bisulfite sequencing and RNA-sequencing) may require the isolation of bacteriomes from more than 50 insects (Noda et al., 2012; Pers and Hansen, 2021). We refined our approach for cell dissociation from multiple bacteriomes.

Trypsinization and ethanol fixation methods require multiple washing steps with buffer to effectively remove any residual trypsin or ethanol from cells before proceeding with downstream experiments. Repeated washing and centrifugation can lead to extensive loss of cells, making it difficult to obtain a high yield of viable and intact cells, which is crucial for subsequent flow cytometry analysis (Krishan, 1975; Xu et al., 1982; Bauman and Bentvelzen, 1988). In our work, microscopic examination after each trypsin dissociation method and ethanol fixation revealed cell debris rather than intact cells. Any surviving cells were observed to be a perfect circular shape (Figure 1), which we presumed to be sheath cells present in bacteriomes (Kobiałka et al., 2015). Thus, we were unable to isolate bacteriocyte cells from host organs using this approach.

Taken together, we found that trypsin dissociation and ethanol fixation proved to be a difficult process that did not lead to the successful isolation of bacteriocytes. These methods may cause significant damage to bacteriocyte structures, resulting in the loss of crucial cellular components. Therefore, we found that it is necessary to develop an alternative approach to isolate host cells inclusive of bacteriocytes.

4.4.2 Physical bacteriocyte isolation and validation for fluorescence

Given the challenges with trypsin dissociation of bacteriocytes in ALF, we switched to using a common physical cell dissociation technique. Tissues were ground between two frosted microscope slides in order to achieve a single-cell suspension for downstream flow cytometry (DeSalle et al., 2005; Sinha and Bhatnagar, 2010; Wing et al., 2017). We identified the fluorescence signal for ALF and *D. melanogaster* (control) using isolated insect heads (Figure 2). After identifying the fluorescence signal of ALF cells from head tissues, we tested fluorescence sorting signals from other body tissues, and their cells in the abdomen in ALF abdomens showed a differential fluorescent signal compared to host tissues. Our logic is that since *Sulcia* and *Nasuia* bacteriocytes are integrated into the abdomen, we can separate them based on their enriched fluorescence signal compared with host head and abdominal somatic cells (Woyke et al., 2010b; Nozaki and Shigenobu, 2022). Indeed, we were able to identify four unique peaks of cell populations in our abdominal tissue samples

(Figure 3). One peak corresponded with host cells (Pop2), while a lower fluorescence signal was hypothesized to represent cell debris (Pop1) (Figure 3). Given that bacteriocytes have higher ploidy levels (Braendle et al., 2003; Woyke et al., 2010b; Nozaki and Shigenobu, 2022; Nakabachi and Moran, 2022) and would therefore exhibit heightened DAPI signals, we hypothesized that two fluorescence peaks higher than that of the host cells contained bacteriocytes (Pop3 and Pop4). Both populations (Pop3 and Pop4) exhibited 2-4x higher fluorescence signals than the host cell population.

To validate our novel cell-sorting technique, we used reverse transcription quantitative-PCR (RT-qPCR) to confirm the presence of symbiont-specific bacteriocytes in our sorted populations. Following the approach established by Mao et al. (2018), we used custom primers targeting genes up-regulated in *Sulcia* or *Nasuia* bacteriocytes. For the identification of *Sulcia* bacteriocytes, we utilized the primer for the horizontally transferred gene (HTG) *frr*. The ribosome recycling factor gene (*frr*) exhibits a ~16x higher expression fold change in *Sulcia* bacteriocytes compared to *Nasuia* bacteriocytes. In a similar vein, we identified *Nasuia* bacteriocytes by targeting the HTG *ribD*, which has an ~19x higher expression fold change when compared to *Sulcia* bacteriocytes. Excitingly, we found low expression of the *Sulcia* bacteriocyte gene, *frr*, in Pop3 and approximately a 6-fold higher difference in Pop4. For *ribD*, we found a ~2-fold difference in Pop3 and Pop4, with Pop4 showing a higher expression of the *Nasuia* bacteriocyte gene. Our results show that Pop4 contained a mix of *Nasuia* and *Sulcia* bacteriocytes, while Pop3 is enriched with a population of *Nasuia* bacteriocytes. As evident in our flow cytometry results, Pop3 has a more uniform signal with higher cell counts, whereas Pop4 had a higher dispersion within its cell population, suggesting a greater variability in the fluorescent signal of these cells. This variability in fluorescent signal most likely occurs due to inconsistencies between the number of genomes in a single bacteriocyte. *Sulcia* bacteriocytes can contain approximately 200-900 genome copies per cell (Woyke et al., 2010b); however, the number of genomes within *Nasuia* bacteriocytes is still unknown. Given that there are variations of bacterial ploidy levels within the same bacteriocyte type, it is possible that some *Nasuia* bacteriocytes exhibited a higher fluorescence signal due to an increased number of genome copies leading to similar signals as other *Sulcia* bacteriocytes.

Tissue grinding presents a viable method for bacteriocyte isolation that removes the need for harsh enzymatic digestion and increases speed and simplicity (DeSalle et al., 2005; Sinha and Bhatnagar, 2010; Hare and Johnston, 2011; Bingham et al., 2015; Wing et al., 2017). It minimizes cell stress and maximizes cell viability, leading to more reliable and reproducible results in downstream analyses. An advantage of this approach is that it can be used to avoid potential cell loss due to repeated washing and longer incubation times. In addition, by circumventing the enzymatic digestion step, tissue grinding also mitigates the

introduction of unwanted enzymes or contaminants that could interfere with subsequent molecular or functional studies.

4.4.3 Genome-wide patterns of methylation

To demonstrate the application of bacteriocyte splitting and sorting, we performed DNA methylation sequencing on bacteriocyte populations (i.e., Pop3 and Pop4) and non-symbiotic host tissues. We compared methylation profiles of both bacteriocyte populations and non-symbiotic host cells. We matched methylation patterns with previously determined expression patterns between bacteriocyte types and whole-body cells (see Mao et al., 2018) to address the question of whether ALF has evolved epigenetic mechanisms to control symbiont interactions. Moreover, almost nothing is known about epigenetic controls in *M. quadrilineatus* (ALF) and related insects in the auchenorrhyncha.

We found 2,871 CpG sites undergoing hyper-methylation, with 663 CpG sites undergoing hypo-methylation between the two groups (Figure 4). Across the ALF's genome, differentially methylated CpG sites were found in 5% of CpG islands (CpGi) and 24% of CpGi shores (Figure 5A). CpGi are often associated with gene regulatory elements, such as promoters, playing a crucial role in modulating gene expression patterns. In vertebrates, while the majority of CpG dinucleotides are heavily methylated, CpGi are typically unmethylated (Bird et al., 1985; Suzuki and Bird, 2008). In comparison, CpG dinucleotides in invertebrates generally lack widespread DNA methylation signatures and instead display a "mosaic" methylation pattern (Suzuki and Bird, 2008; Feng et al., 2010). Unlike the uniform methylation patterns of vertebrates, mosaic methylation is defined by interspersed methylated and unmethylated regions throughout the invertebrate genome. Most invertebrate methylated CpGs are concentrated on gene bodies, which are correlated with genomic regions that are actively transcribed (Foret et al., 2009; Bonasio et al., 2012; Sarda et al., 2012; Wang et al., 2013; Keller et al., 2016). Given this overall mosaic pattern of DNA methylation in invertebrates, understanding the specificities of where our differentially methylated CpG sites occur may provide insights into the role of DNA methylation in symbiosis.

In the ALF system, we found globally higher number of differential methylated CpG sites in introns (66%) compared to intergenic regions (24%), exons (7%), and promoter regions (3%) (Figure 5B). The higher percentage of methylated CpG sites in introns may indicate a potential role for intronic methylation in regulated gene expression in this host-symbionts system. For example, it is possible that intronic methylation could be interacting with other elements of gene regulation, such as alternative splicing, non-coding RNAs, or chromatin modifications (Ball et al., 2009; Cedar and Bergman, 2009; Luco et al., 2010). While the role of intronic methylation in regulating gene expression remains less understood, methylation among promoter regions has been primarily linked to the repression of transcriptional activity (Zhang et al., 2006; Weber et al., 2007;

Zhang, 2008). In insects, promoter methylation is yet to have evidence of correlation with gene expression levels (Lyko et al., 2010; Xiang et al., 2010; Bonasio et al., 2012).

Despite the known mosaic methylation characteristic of insects, such as ALF, our understanding remains limited regarding the correlation between specific genomic regions and gene expression levels within symbiotic host cells. To address this gap, we tested whether gene expression levels corresponded with genomic regions in a comparative analysis between symbiotic and non-symbiotic host cells. In the case of promoter regions, we found 1,305 gene promoters with differentially methylated CpG sites (836 hyper-methylated, 469 hypo-methylated) (Figure 6A). Using a Pearson chi-squared test, we found a significant association between hyper-methylation in promoter regions and down-regulation (*p*-value = 0.007), but no significant association between hypo-methylation and up-regulation (*p*-value = 0.365). We identified 47 promoter genes that show an inverse expression of hypo-methylation/up-regulation (hypo-up), and 91 promoter genes show an inverse expression of hyper-methylation/down-regulation (hyper-down) (Figure 6B,C). Using a Pearson correlation test, neither inverse expression pair showed significance in correlation (hypo-up: *p*-value = 0.26; hyper-down: *p*-value = 0.3).

We further tested whether methylation in other genomic regions may be correlated with gene expression. Exons (hypo-up: *p*-value = 0.18; hyper-down: *p*-value = 0.23) and introns (hypo-up: *p*-value = 0.11; hyper-down: *p*-value = 0.53) do not exhibit a significant correlation between methylation and gene expression. These results suggest that DNA methylation at promoter regions, exons, or introns may not directly correspond with gene expression changes in a simple inverse manner, at least in the bacteriocytes and non-symbiotic tissue context. The lack of a strong correlation between differential methylation and gene expression indicates that other regulatory mechanisms may be at play. These might include histone modifications, non-coding RNAs, or more complex methylation interactions with these and other elements of the gene regulatory machinery (Ball et al., 2009; Cedar and Bergman, 2009; Luco et al., 2010).

4.4.4 Differential methylation M. quadrilineatus (ALF) bacteriocytes

One of the major aims of this study is to demonstrate how fluorescence-based cell sorting can be utilized to isolate separate bacteriocytes to understand differential host support to each bacterial partner. A major question in symbiotic systems is how hosts might differentially regulate the expression of genes within and between bacteriocytes. In systems like ALF and related auchenorrhyncha, symbionts are differentially sequestered to specific bacteriocyte types. Despite having the same genome, gene expression varies by thousands of genes, many with unique expression patterns (Bennett and Chong, 2017; Mao et al., 2018a; Mao and Bennett, 2020). Additionally, a substantial quantity of genes is found to

be up-regulated in both bacteriocyte types, highlighting the intricate complexity of this regulatory process (Mao et al., 2018a). To test questions surrounding the role of methylation in gene regulation in different symbiotic host cells, we analyzed DNA methylation patterns between *Nasuia* bacteriocytes (Pop3) and *Sulcia* bacteriocytes (Pop4) bacteriocytes.

In our results from *Nasuia* bacteriocytes*,* we found 2,979 CpG sites undergoing hyper-methylation and 501 CpG sites undergoing hypo-methylation as compared to non-symbiotic host cell types. *Sulcia* bacteriocytes, on the other hand, exhibited 430 and 519 CpG sites undergoing hyper-methylation and hypo-methylation as compared to non-symbiotic host cell types, respectively. In the context of methylation at promoter regions, we found that *Sulcia* and *Nasuia* bacteriocytes share 161 hyper-methylated and 135 hypo-methylated genes (Figure 7). It is important to note that several genes are up-regulated in both bacteriocyte types (Mao et al., 2018a), suggesting a common, potentially conserved mechanism of epigenetic regulation in maintaining both symbiont relationships in ALF. There is no association between genes that are undergoing hyper-methylation and down-regulation (*Sulcia* bacteriocytes: *p*-value = 0.940, *Nasuia* bacteriocytes: *p*-value = 0.075), as well as hypo-methylation and up-regulation in either bacteriocyte types *(Sulcia* bacteriocytes: *p*-value = 0.1492, *Nasuia* bacteriocytes: *p*-value = 0.6172). In a more focused analysis, there are 146 promoter regions that exhibit an inverse hypo- and hyper-methylation signal in *Nasuia* and *Sulcia* bacteriocytes, respectively, with no significant correlation (*p*-value = 0.75). However, there is a significant correlation between the 190 promoters that exhibit an inverse hyper- and hypo-methylation signal in *Nasuia* and *Sulcia* bacteriocytes, respectively (*p*-value = 0.0024). The discrepancy between the methylation status of these two sets of promoter regions suggests that the direction of differential methylation might influence gene expression in a species-specific manner in *Nasuia* and *Sulcia* bacteriocytes. Although both bacteriocytes house symbiotic bacteria, the regulation of genes in each bacteriocyte type through methylation can be quite different. These results indicates that it is possible that the species-specific methylation patterns of *Nasuia* and *Sulcia* bacteriocytes could be the result of unique co-evolutionary dynamics between ALF and each symbiotic partner.

In a narrowed investigation, we contrasted specific bacteriocyte types with non-symbiont host cells to determine if the global methylation patterns previously established between bacteriocytes and non-symbiont host cells persist. As shown with global patterns, there is no correlation between methylation and gene expression in promoters (hypo-up: *p*-value = 0.075; hyper-down: *p*-value = 0.14), exons (hypo-up: *p*-value = 0.57; hyper-down: *p*-value = 0.63) and introns (hypo-up: *p*-value = 0.24; hyper-down: *p*-value = 0.29) in *Sulcia* bacteriocytes. In *Nasuia,* this pattern is also exhibited in exons (hypo-up: *p*-value = 0.72; hyper-down: *p*-value = 0.14) and introns (hypo-up: *p*-value = 0.12; hyper-down: *p*-value = 0.99); however, there is a correlation between methylation and gene

expression in promoters in hyper-down genes (*p*-value = 0.012), but not in hypo-up genes (*p*-value = 0.27). This lack of correlation between methylation and gene expression in various genomic regions in *Sulcia* bacteriocytes, as well as in exons and introns in *Nasuia* bacteriocytes, emphasizes the complex and distinct regulatory mechanisms that are likely at play in regulating this complex symbiotic interaction in ALF. Interestingly, the observed correlation between promoter methylation and downregulated gene expression in *Nasuia* bacteriocytes hints towards a possible role of DNA methylation in silencing or downregulating gene expression. However, the loss of significant correlation in hypo-up methylation patterns still suggests that other regulatory factors may be involved in the regulation of host gene expression in *Nasuia*. DNA methylation may not be involved in the repression of transcription, but our results suggest that it most likely plays other regulatory roles in symbiotic host cells.

Finally, there are hundreds of genes involved in symbiosis that support several key functional categories, including genes involved in nutrition synthesis, information processing, population regulation, and metabolite transport. Individual genes show differential expression patterns between *Nasuia* bacteriocytes and *Sulcia* bacteriocytes, generally matching the specific needs of individual bacterial species (Mao et al., 2018) (Figure 8). We tested whether methylation may play a role in the differential expression of at least some of these genes. Among the 148 symbiont-support genes, only 33 could be mapped to our differentially methylated promoters. Among these mapped genes, the relationship between DNA methylation at promoter sites and gene expression is complex and unclear (Figure 8). Ten genes show significant hyper-methylation in *Nasuia* bacteriocytes, but only one, RNA pseudouridine synthase (*rluA*), exhibits hypo-methylation in *Sulcia.* Despite the hyper-methylation in *Nasuia* bacteriocytes, *rluA* is a horizontally transferred gene in ALF that is more highly expressed in *Nasuia* bacteriocytes, with low expression in *Sulcia* bacteriocytes. Additionally, the *PANK* gene (pantothenate kinase) exhibits significant hypo-methylation in both *Sulcia* and *Nasuia* bacteriocytes and is highly expressed in both bacteriocyte types. Taken together, these results indicate that methylation status alone does not dictate gene expression levels and that other regulatory factors may be in play. Furthermore, it implies that the influence of methylation on gene expression may be context-dependent, varying across different genes and cell types.

The lack of a consistent correlation between DNA methylation and gene expression in our data suggests an intricate interplay of regulatory factors. Two possible scenarios arise from this observation. First, DNA methylation may be playing a less traditional or secondary role in the regulation of these genes. Alternatively, there may be other regulatory mechanisms at play (as described above). The symbiotic relationship between host and symbiont reflects a long evolutionary process that has spanned millions of years. Insect hosts have acquired novel genes and functions to support persistently degrading symbiont

genomes (McCutcheon et al., 2019). The acquisition and assignment of these genes to the host-symbiont interface is complex and likely idiosyncratic. As such, the regulation of these genes may not follow one consistent pattern, and multiple different mechanisms evolve to manage the localized expression of individual genes among bacteriocytes. Finally, it's possible that *Nasuia* and *Sulcia* bacteriocytes have evolved unique, species-specific ways of managing gene expression via methylation, in order to maintain their respective symbiotic relationships. Understanding how symbiont support genes are regulated may require a specific investigation of the origin and evolutionary history of individual genes.

4.5 CONCLUSION

Our use of fluorescence-based cell sorting for isolating specific symbiont-containing bacteriocyte cell populations demonstrates a successful application. It opens new possibilities for downstream analyses that more precisely address questions such as how bacteriocytes function and evolve. Currently, little is known about the role of epigenetic control in *Macrosteles quadrilineatus* and related insects in the Auchenorrhyncha. Using genome-wide DNA methylation sequencing approaches in *Macrosteles quadrilineatus*, we show differential methylation patterns between *Sulcia* and *Nasuia* bacteriocytes compared to non-symbiotic host tissues and cells. Despite the complexity of methylation patterns, our results show that this biological process plays a role in regulating the expression of host genes involved in maintaining complex symbioses. Methylation is likely, but one evolved mechanism among other regulatory processes that have entwined roles in maintaining symbioses (Ball et al., 2009; Cedar and Bergman, 2009; Luco et al., 2010). Our findings lay a foundation for further exploration into the biology of these symbiotic relationships. We anticipate that this study will serve as a useful reference for future investigations into insect symbiosis, particularly in exploring the role of DNA methylation and other regulatory mechanisms in these complex systems.

4.6 FIGURES

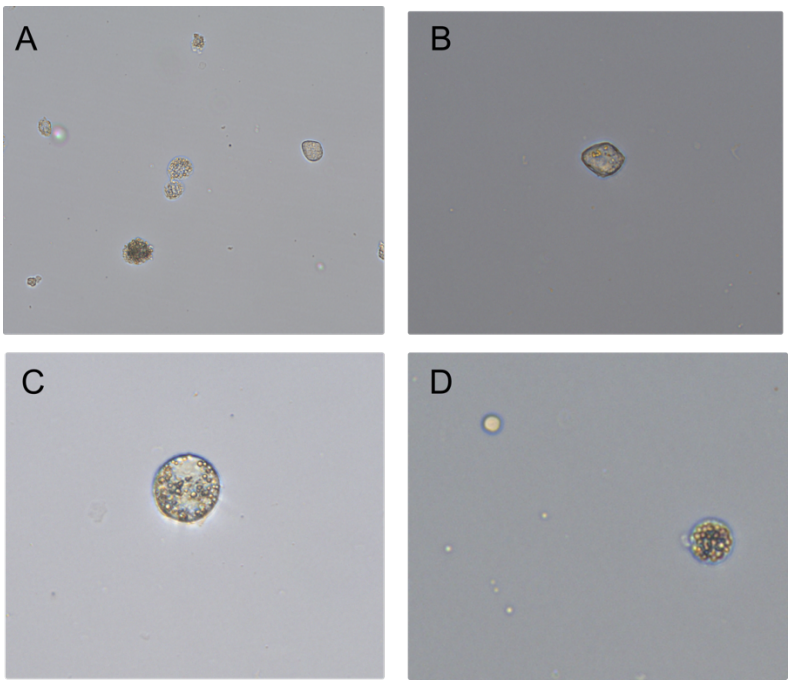


Figure 4.1: Cells from 1% trypsin digestion at room temperature.
A) Undigested bacteriome cells at 20x. B) Bacteriome cells at 10 minutes at 40x. C) Bacteriome cells at 20 minutes at 40x. D) Bacteriome cells at 30 minutes at 40x.

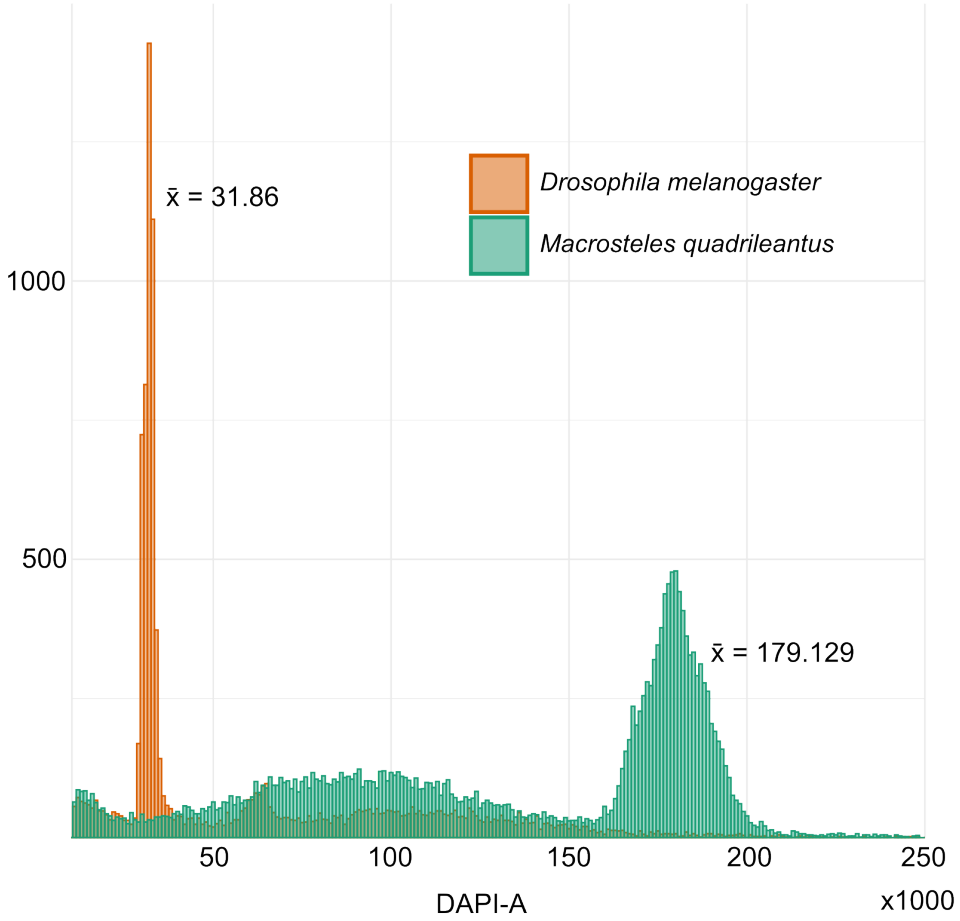


Figure 4.2: DAPI fluorescence of ALF head cells.
D. melanogaster was used as a control marker to identify ALF specific host signal. Mean fluorescence of each species is denoted by x̄.

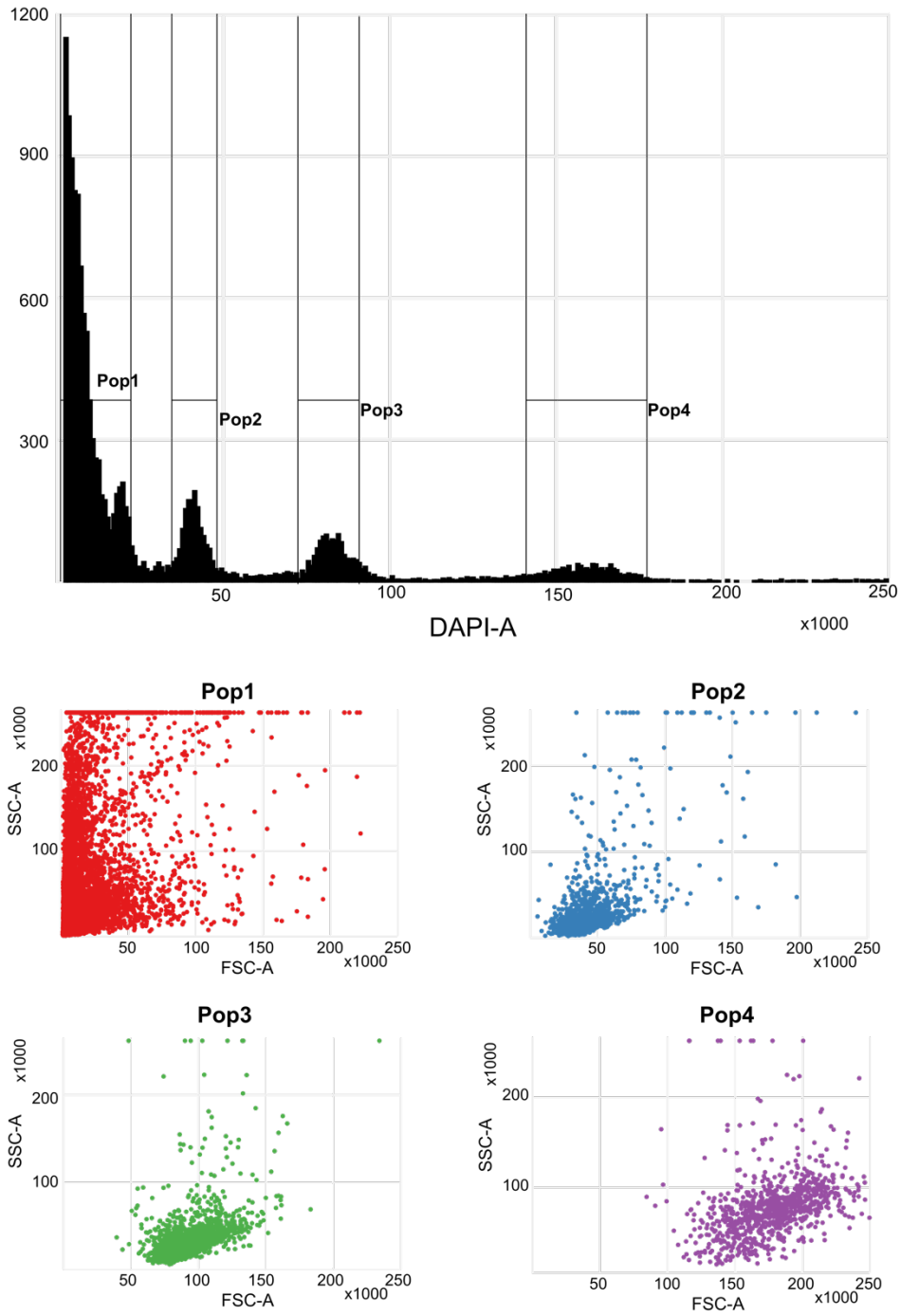


Figure 4.3: Sorted populations from *Macrosteles quadrilineatus* bacteriocytes. Populations were sorted by their DAPI fluorescence. Pop1 is predicted to be cell debris. Pop2 are host cells fluorescence as determine by control study. Pop3 and Pop4 are predicted to be bacteriocytes due to increased fluorescence possibly indicating increased ploidy level. Forward scatter (FSC-A) and side scatter (SSC-A) are included for each population.

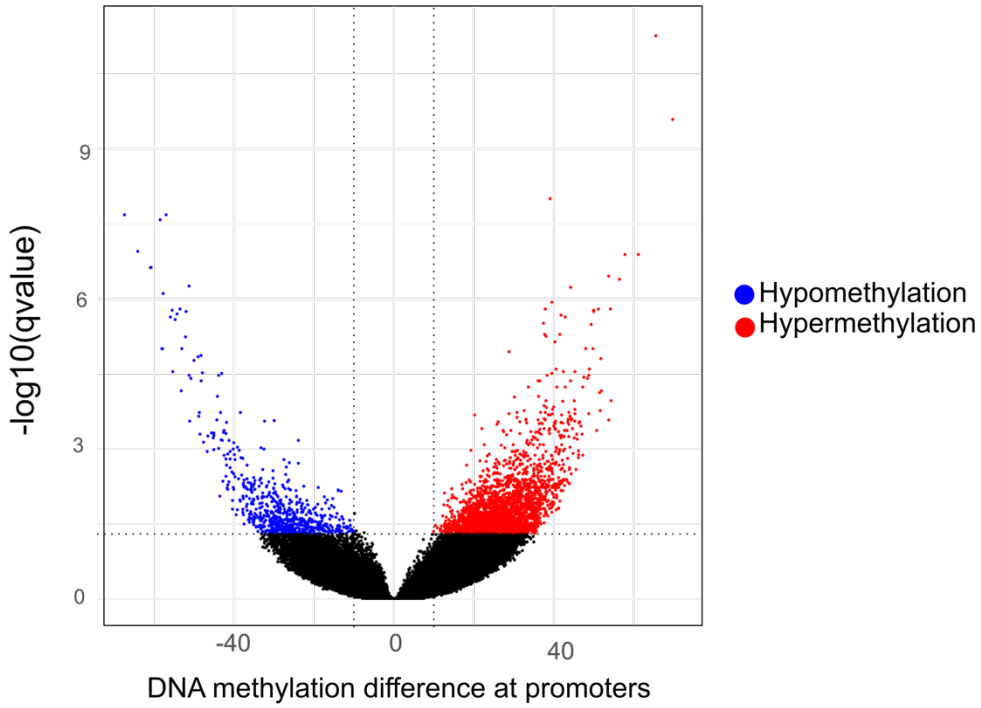


Figure 4.4: Differential methylation patterns between bacteriocyte cell types and non-symbiotic host cell types.
Hyper-methylated and hypo-methylated CpG sites in promoter sequences between symbiotic and non-symbiotic host cells. Significant methylation difference was identified by sites that exhibit >10% difference and a Benjamini-Hochberg adjust *p*-value < 0.05 (qvalue).

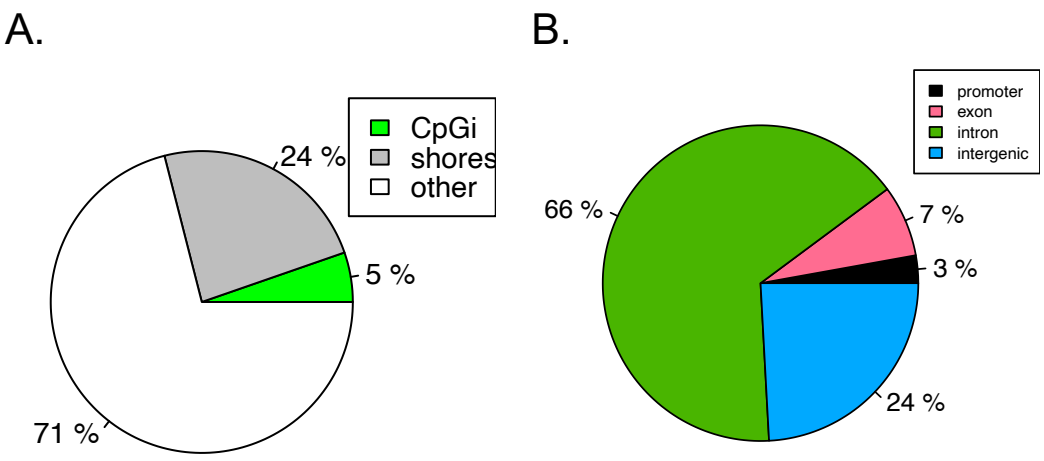


Figure 4.5: Global methylation patterns between bacteriocyte cell types and non-symbiotic host cell types.
A) Percentages of CpG sites differentially methylated within CpG islands and shores. B) Percentages of CpG sites differentially methylated within genomic regions.

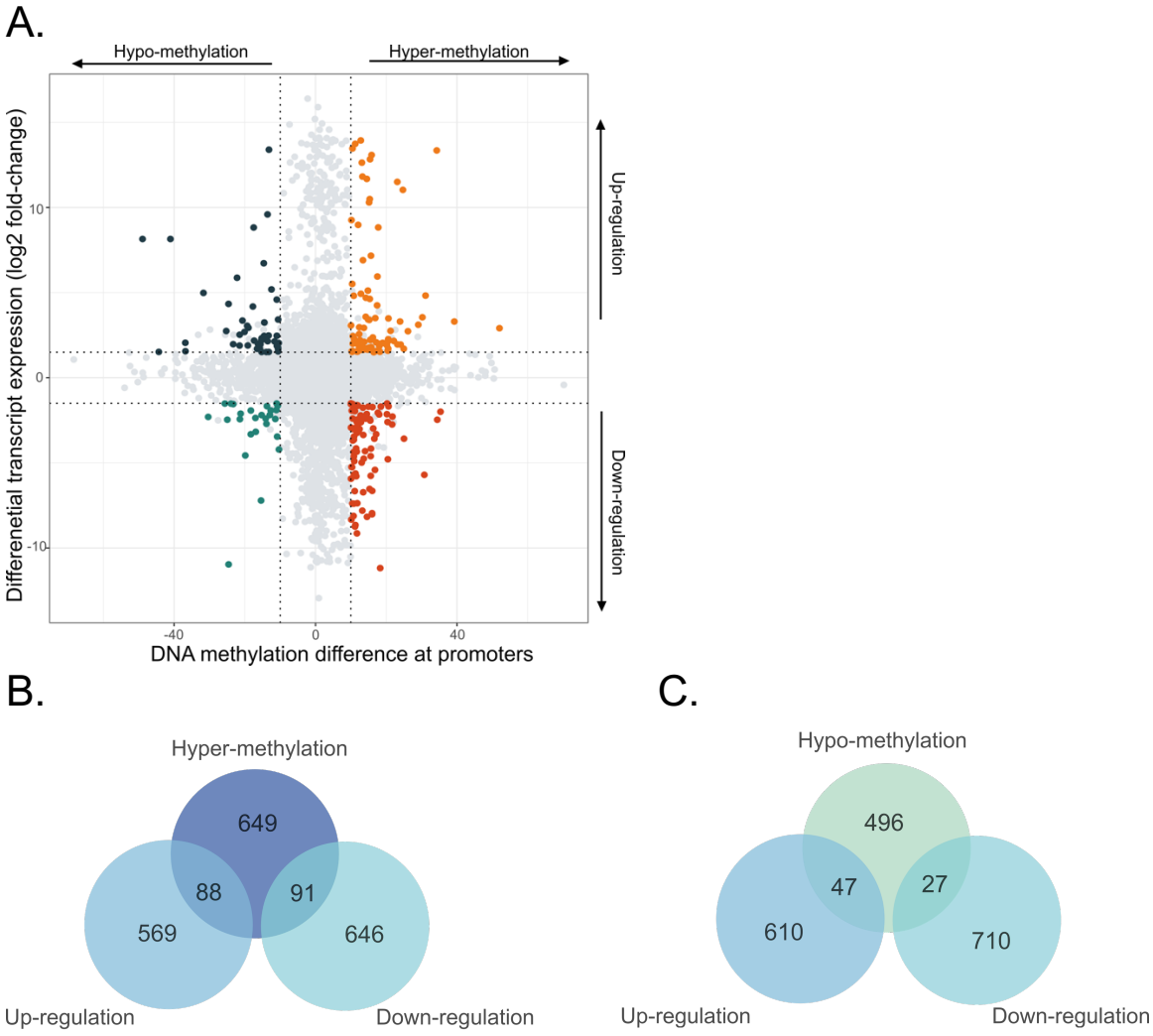


Figure 4.6: Differential methylation patterns between bacteriocyte cell types and non-symbiotic host cell types correlated with gene expression patterns.
A) DNA methylation difference at promoter sequences correlated with gene expression from DESeq2. Genes that exhibit significant methylation and significant gene expression are indicated by colored circles. Each quadrant is separated by their relationship between hypo-methylation, hyper-methylation, up-regulation and down-regulation. B) Number of genes significantly hyper-methylated that exhibit patterns of up-regulation or down-regulation. C) Number of genes significantly hypo-methylated that exhibit patterns of up-regulation or down-regulation.

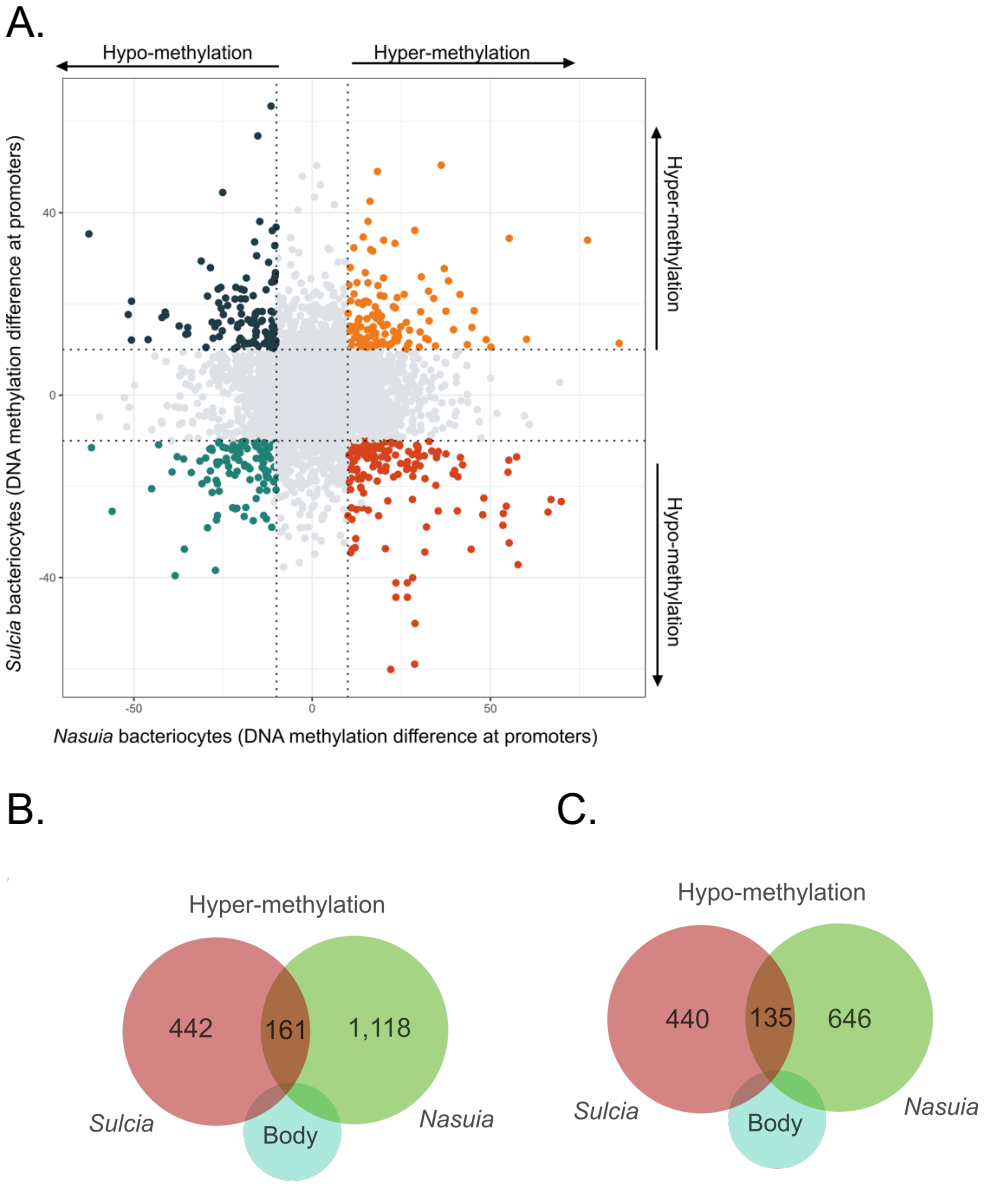
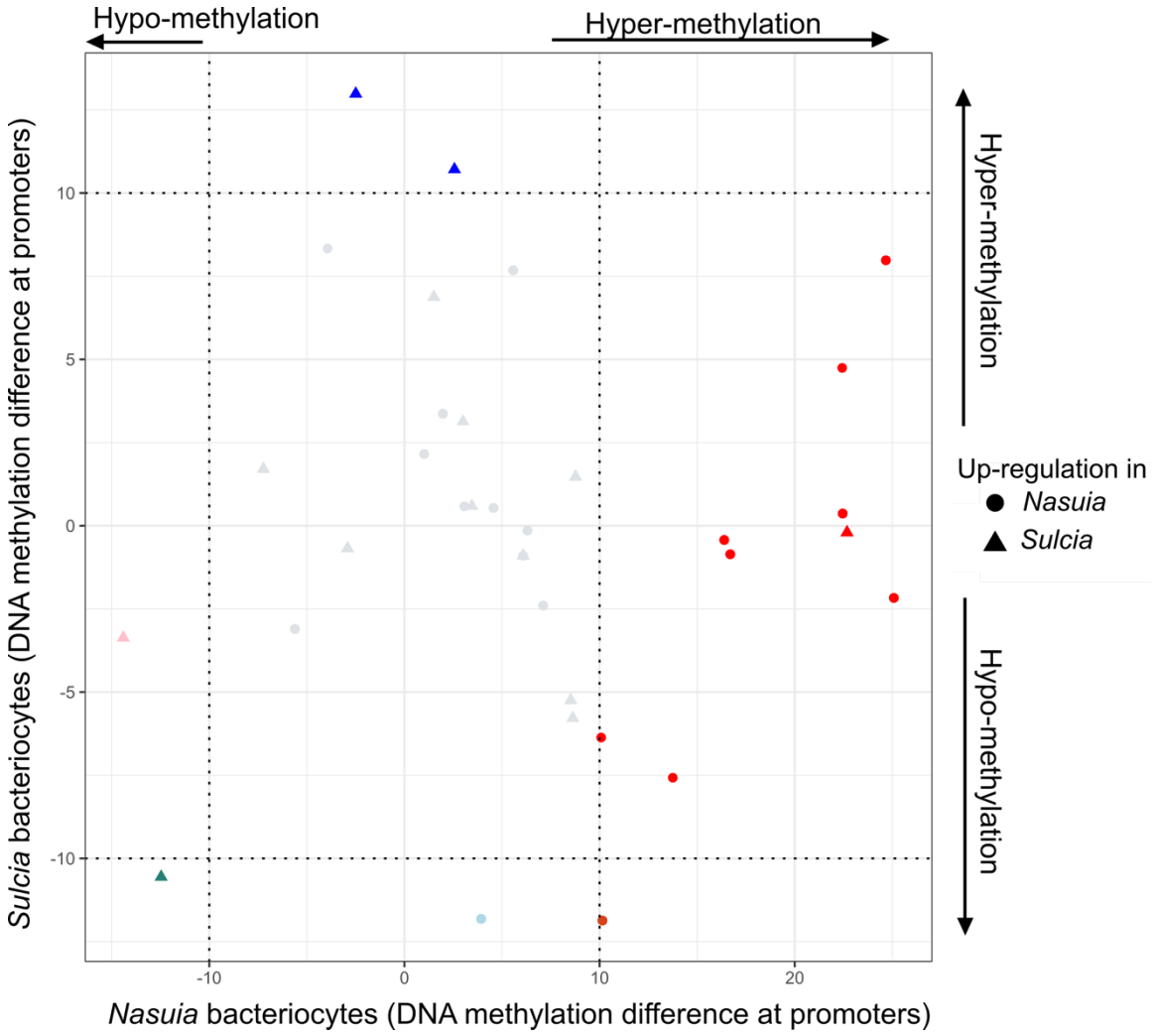


Figure 4.7: Differential methylation patterns between each bacteriocyte. A) DNA methylation difference at promoter sequences between *Nasuia* bacteriocytes and *Sulcia* bacteriocytes. Genes that exhibit significant methylation and significant gene expression are indicated by colored circles. Each quadrant is separated by their relationship between hypo-methylation, hyper-methylation, up-regulation and down-regulation. B) Number of genes significantly hyper-methylated that are shared by *Nasuia* bacteriocytes and *Sulcia* bacteriocytes (denoted as *Nasuia* or *Sulcia*). C) Number of genes significantly hypo-methylated that are shared by *Nasuia* bacteriocytes and *Sulcia* bacteriocytes (denoted as *Nasuia* or *Sulcia*).

A.


Hypo-methylation
Hyper-methylation
Sulcia bacteriocytes (DNA methylation difference at promoters)
Nasuia bacteriocytes (DNA methylation difference at promoters)
10
5
0
-5
-10
-10
0
10
20
Hyper-methylation
Hypo-methylation
Up-regulation in
Nasuia
Sulcia

B.

Gene Product	Nasuia qvalue	Nasuia methylation difference	Sulcia qvalue	Sulcia methylation difference	Body mean expression	Nasuia mean expression	Sulcia mean expression
translation initiation factor IF-2, mt	1.28E-03	5.572	1.27E-10	7.677	0.039	45.156	1.305
translation factor GUF1 homolog, mt	1.89E-03	6.085	5.56E-01	-0.902	0.353	25.485	368.611
bifunctional glutamate/proline--tRNA ligase	1.85E-01	-7.221	7.37E-01	1.701	0.059	2.068	12.878
DNA polymerase alpha	2.56E-05	10.083	2.53E-04	-6.367	4.616	26.434	17.347
DNA-directed RNA polymerase II subunit RPB1	5.14E-06	8.625	5.40E-05	-5.782	0.088	0.252	2.875
mt basic amino acids transporter	3.78E-01	1.512	8.20E-10	6.875	0.229	184.150	864.858
RNA pseudouridine synthase	7.83E-03	10.148	1.06E-04	-11.868	0.000	72.439	3.858
Peptide deformylase	4.68E-25	16.699	4.54E-01	-0.857	0.000	23.036	1.531
elongation factor Ts, mt	1.55E-07	22.459	9.29E-01	0.367	0.598	1819.534	84.695
Peptide deformylase	5.33E-12	6.304	8.70E-01	-0.141	0.000	3.982	0.230
ribonuclease H1	3.96E-01	1.970	1.82E-02	3.365	1.023	372.074	19.174
B(0,+)-type amino acid transporter	3.15E-01	-3.934	2.69E-03	8.333	1.629	13.509	3.984
DNA-directed RNA polymerase III subunit RPC10	7.44E-03	3.067	5.45E-01	0.584	20.330	208.597	54.039
ranslation factor GUF1 homolog, mt	5.43E-01	1.015	5.83E-02	2.154	10.930	713.614	287.020
genome maintenance exonuclease, mt	5.25E-03	7.112	2.46E-01	-2.402	1.131	2072.985	183.967
Isoleucine-tRNA ligase	5.55E-06	8.775	3.45E-01	1.470	0.073	17.403	294.909
3'-phosphoadenosine 5'-phosphosulfate synthase	2.52E-01	3.931	1.53E-03	-11.818	32.472	7491.896	455.106
phenylalanine--tRNA ligase, mt	4.62E-75	22.434	1.82E-06	4.742	25.422	199.851	58.913
3'-5' exonuclease	5.15E-32	25.081	1.92E-01	-2.169	0.112	630.243	32.276
aspartate--tRNA ligase	1.64E-01	-2.916	7.25E-01	-0.684	0.472	6.926	119.279
threonine--tRNA ligas	1.92E-27	16.389	7.63E-01	-0.429	0.084	188.760	11.169
NADP-dependent malic enzyme	1.24E-03	8.522	7.02E-03	-5.247	557.879	128.516	1730.025
cysteine--tRNA ligase, mt	1.14E-07	-14.404	3.94E-02	-3.365	12.737	140.157	150.837
B(0,+)-type amino acid transporter	1.57E-15	22.682	9.37E-01	-0.202	0.236	8.402	119.361
DNA ligase 3	1.03E-02	3.007	8.92E-05	3.132	5.063	13.089	59.424
pantothenate kinase	2.27E-06	-12.466	1.92E-09	-10.556	32.580	142.026	268.975
methionine--tRNA ligase	3.59E-01	2.564	7.76E-08	10.710	15.455	107.797	256.645
tryptophan--tRNA ligase, mt	1.02E-10	24.672	1.45E-02	7.978	15.746	70.014	13.043
arginine--tRNA ligase	1.29E-02	3.460	5.65E-01	0.593	29.270	48.512	126.216
mutS protein homolog 4	4.52E-04	-5.608	5.35E-03	-3.103	0.521	27.515	5.656
DNA polymerase theta	9.65E-06	13.752	9.31E-03	-7.575	0.377	1.878	1.144
DNA topoisomerase 2	0.0041644	4.563	6.53E-01	0.534	4.047	163.162	15.920

Figure 4.8: Differential methylation patterns in putative symbiosis support genes between each bacteriocyte.
Putative symbiont genes identified in Mao et al. (2018) mapped their differentially methylated promoter sequences. A) Genes that are significantly hypo-methylated or hyper-methylated in either *Sulcia* or *Nasuia* bacteriocytes are colored. Genes that show no significance in methylation differences in both bacteriocytes are in grey. Each sample is labeled by the bacteriocyte that exhibited the highest up-regulation according to mean gene expression from Mao et al. (2018). B) Table showing numerical values for each putative symbiont support genes, including their adjusted *p*-value (qvalue), their methylation difference and their mean gene expression.

References

Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res *38*, W7–W13. https://doi.org/10.1093/nar/gkq291.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology *13*, 1–9. .

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Álvarez-Carretero, S., Kapli, P., and Yang, Z. (2023). Beginner's Guide on the Use of PAML to Detect Positive Selection. Molecular Biology and Evolution *40*, msad041. https://doi.org/10.1093/molbev/msad041.

Alves, J.M.P., Serrano, M.G., Maia da Silva, F., Voegtly, L.J., Matveyev, A.V., Teixeira, M.M.G., Camargo, E.P., and Buck, G.A. (2013). Genome Evolution and Phylogenomic Analysis of Candidatus Kinetoplastibacterium, the Betaproteobacterial Endosymbionts of Strigomonas and Angomonas. Genome Biol Evol *5*, 338–350. https://doi.org/10.1093/gbe/evt012.

Aminetzach, Y.T., Macpherson, J.M., and Petrov, D.A. (2005). Pesticide Resistance via Transposition-Mediated Adaptive Gene Truncation in Drosophila. Science *309*, 764–767. https://doi.org/10.1126/science.1112699.

Andrade, C., Oliveira, B., Guatelli, S., Martinez, P., Simões, B., Bispo, C., Ferrario, C., Bonasoro, F., Rino, J., and Sugni, M. (2021). Characterization of coelomic fluid cell types in the starfish Marthasterias glacialis using a flow cytometry/imaging combined approach. Frontiers in Immunology *12*, 807. .

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Anisimova, M., Bielawski, J.P., and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Molecular Biology and Evolution *18*, 1585–1592. .

Ankrah, N.Y.D., Wilkes, R.A., Zhang, F.Q., Zhu, D., Kaweesi, T., Aristilde, L., and Douglas, A.E. (2020). Syntrophic splitting of central carbon metabolism in host

cells bearing functionally different symbiotic bacteria. ISME J *14*, 1982–1993. https://doi.org/10.1038/s41396-020-0661-z.

Arab, D.A., and Lo, N. (2021). Evolutionary Rates are Correlated Between Buchnera Endosymbionts and the Mitochondrial Genomes of Their Aphid Hosts. J Mol Evol *89*, 238–248. https://doi.org/10.1007/s00239-021-10001-9.

Arab, D.A., Bourguignon, T., Wang, Z., Ho, S.Y.W., and Lo, N. (2020). Evolutionary rates are correlated between cockroach symbionts and mitochondrial genomes. Biol Lett *16*, 20190702. https://doi.org/10.1098/rsbl.2019.0702.

Bai, Y., Yang, C., Halitschke, R., Paetz, C., Kessler, D., Burkard, K., Gaquerel, E., Baldwin, I.T., and Li, D. (2022). Natural history–guided omics reveals plant defensive chemistry against leafhopper pests. Science *375*, eabm2948. .

Ball, M.P., Li, J.B., Gao, Y., and Lee, J.H. (2009). Cytosine methylation may interact with other epigenetic features, such as histone modifications. Nature Biotech *27*, 361–368. .

Bandi, Venkat, and Gutwin, Carl (2020). Interactive Exploration of Genomic Conservation. (Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine), p. 10 pages, 5090.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Banttari, E.E., and Zeyen, R.J. (1979). Interactions of mycoplasmalike organisms and viruses in dually infected leafhoppers, planthoppers and plants. In Leafhopper Vectors and Plant Disease Agents, (Elsevier), pp. 327–347.

Bao, Y.-Y., Qin, X., Yu, B., Chen, L.-B., Wang, Z.-C., and Zhang, C.-X. (2014). Genomic insights into the serine protease gene family and expression profile analysis in the planthopper, Nilaparvata lugens. BMC Genomics *15*, 507. https://doi.org/10.1186/1471-2164-15-507.

Barlogie, B., Raber, M.N., Schumann, J., Johnson, T.S., Drewinko, B., Swartzendruber, D.E., Göhde, W., Andreeff, M., and Freireich, E.J. (1983). Flow Cytometry in Clinical Cancer Research1. Cancer Research *43*, 3982–3997. .

Baum, J.A., and Roberts, J.K. (2014). Progress towards RNAi-mediated insect pest management. In Advances in Insect Physiology, (Elsevier), pp. 249–295.

Bauman, J.G.J., and Bentvelzen, P. (1988). Flow cytometric detection of ribosomal RNA in suspended cells by fluorescent in situ hybridization. Cytometry: The Journal of the International Society for Analytical Cytology *9*, 517–524. .

Baumann, P. (2005). Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. Annu Rev Microbiol *59*, 155–189. https://doi.org/10.1146/annurev.micro.59.030804.121041.

Beanland, L., Madden, L.V., Hoy, C.W., Miller, S.A., and Nault, L.R. (2005). Temporal Distribution of Aster Leafhopper Sex Ratios and Spatial Pattern of Aster Yellows Phytoplasma Disease in Lettuce. Annals of the Entomological Society of America *98*, 756–762. https://doi.org/10.1603/0013-8746(2005)098[0756:TDOALS]2.0.CO;2.

Becker, A., Schlöder, P., Steele, J.E., and Wegener, G. (1996). The regulation of trehalose metabolism in insects. Experientia *52*, 433–439. .

Bell-Roberts, L., Douglas, A.E., and Werner, G.D.A. (2019). Match and mismatch between dietary switches and microbial partners in plant sap-feeding insects. Proceedings of the Royal Society B: Biological Sciences *286*, 20190065. https://doi.org/10.1098/rspb.2019.0065.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Statist Soc B *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bennett, G. (2020). Evolving integrated multipartite symbioses between plant-sap feeding insects (Hemiptera) and their endosymbionts. In Cellular Dialogues in the Holobiont, (CRC Press), pp. 173–200.

Bennett, G.M., and Chong, R.A. (2017). 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: Genes, Genomes, Genetics *7*, 3073–3082. .

Bennett, G.M., and Mao, M. (2018). Comparative genomics of a quadripartite symbiosis in a planthopper host reveals the origins and rearranged nutritional responsibilities of anciently diverged bacterial lineages. Environ Microbiol *20*, 4461–4472. https://doi.org/10.1111/1462-2920.14367.

Bennett, G.M., and Moran, N.A. (2013). Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. Genome Biol Evol *5*, 1675–1688. https://doi.org/10.1093/gbe/evt118.

Bennett, G.M., and Moran, N.A. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci USA *112*, 10169–10176. https://doi.org/10.1073/pnas.1421388112.

Bennett, G.M., and O'Grady, P.M. (2012). Host–plants shape insect diversity: Phylogeny, origin, and species diversity of native Hawaiian leafhoppers (Cicadellidae: Nesophrosyne). Mol Phylogenet Evol *65*, 705–717. https://doi.org/10.1016/j.ympev.2012.07.024.

Bennett, G.M., and O'Grady, P.M. (2013). Historical biogeography and ecological opportunity in the adaptive radiation of native Hawaiian leafhoppers (Cicadellidae: Nesophrosyne). Journal of Biogeography *40*, 1512–1523. .

Bennett, G.M., McCutcheon, J.P., MacDonald, B.R., Romanovicz, D., and Moran, N.A. (2014). Differential Genome Evolution Between Companion Symbionts in an Insect-Bacterial Symbiosis. MBio *5*. https://doi.org/10.1128/mBio.01697-14.

Bennett, G.M., Abbà, S., Kube, M., and Marzachì, C. (2016a). Complete Genome Sequences of the Obligate Symbionts " *Candidatus* Sulcia muelleri" and " *Ca.* Nasuia deltocephalinicola" from the Pestiferous Leafhopper *Macrosteles quadripunctulatus* (Hemiptera: Cicadellidae). Genome Announc *4*. https://doi.org/10.1128/genomeA.01604-15.

Bennett, G.M., McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2016b). Lineage-Specific Patterns of Genome Deterioration in Obligate Symbionts of Sharpshooter Leafhoppers. Genome Biol Evol *8*, 296–301. https://doi.org/10.1093/gbe/evv159.

Bennett, G.M., Abbà, S., Kube, M., and Marzachì, C. (2016c). Complete genome sequences of the obligate symbionts "Candidatus Sulcia muelleri" and "Ca. Nasuia deltocephalinicola" from the pestiferous leafhopper Macrosteles quadripunctulatus (Hemiptera: Cicadellidae). Genome Announc. *4*, e01604-15. .

Benton, R., Vannice, K.S., and Vosshall, L.B. (2007). An essential role for a CD36-related receptor in pheromone detection in Drosophila. Nature *450*, 289–293. https://doi.org/10.1038/nature06328.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P.F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution *69*, 313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Bewick, A.J., Vogel, K.J., Moore, A.J., and Schmitz, R.J. (2016). Evolution of DNA Methylation across Insects. Mol Biol Evol msw264. https://doi.org/10.1093/molbev/msw264.

Bhagwat, N., Dulmage, K., Pletcher Jr, C.H., Wang, L., DeMuth, W., Sen, M., Balli, D., Yee, S.S., Sa, S., and Tong, F. (2018). An integrated flow cytometry-based platform for isolation and molecular characterization of circulating tumor single cells and clusters. Scientific Reports *8*, 5035. .

Bingham, K.N., Lee, M.D., and Rawlings, J.S. (2015). The use of flow cytometry to assess the state of chromatin in T cells. JoVE (Journal of Visualized Experiments) e53533. .

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99. .

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

Bonasio, R., Li, Q., Lian, J., Mutti, N.S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., and Ding, Y. (2012). Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Current Biology *22*, 1755–1764. .

Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D.H., and Keeling, P.J. (2017). Parallel genome reduction in symbionts descended from closely related free-living bacteria. Nat Ecol Evol *1*, 1160–1167. https://doi.org/10.1038/s41559-017-0237-0.

Bourguignon, T., Kinjo, Y., Villa-Martín, P., Coleman, N.V., Tang, Q., Arab, D.A., Wang, Z., Tokuda, G., Hongoh, Y., Ohkuma, M., et al. (2020). Increased Mutation Rate Is Linked to Genome Reduction in Prokaryotes. Curr Biol *30*, 3848-3855.e4. https://doi.org/10.1016/j.cub.2020.07.034.

Braendle, C., Miura, T., Bickel, R., Shingleton, A.W., Kambhampati, S., and Stern, D.L. (2003). Developmental Origin and Evolution of Bacteriocytes in the Aphid–Buchnera Symbiosis. PLOS Biology *1*, e21. https://doi.org/10.1371/journal.pbio.0000021.

Brambila, J., and Hodges, G.S. (2008). Bugs (Hemiptera). In Encyclopedia of Entomology, J.L. Capinera, ed. (Dordrecht: Springer Netherlands), pp. 591–611.

Brevik, K., Bueno, E.M., McKay, S., Schoville, S.D., and Chen, Y.H. (2021). Insecticide exposure affects intergenerational patterns of DNA methylation in the Colorado potato beetle, Leptinotarsa decemlineata. Evolutionary Applications *14*, 746–757. https://doi.org/10.1111/eva.13153.

Brodbeck, B.V., Mizell, R.F., French, W.J., Andersen, P.C., and Aldrich, J.H. (1990). Amino acids as determinants of host preference for the xylem feeding

leafhopper, Homalodisca coagulata (Homoptera: Cicadellidae). Oecologia *83*, 338–345. https://doi.org/10.1007/BF00317557.

Brodbeck, B.V., Andersen, P.C., Oden, S., and Mizell, R.F. (2014). Preference-performance linkage of the xylem feeding leafhopper, Homalodisca vitripennis (Hemiptera Cicadellidae). Environmental Entomology *36*, 1512–1522. .

Brown, M.A., Wallace, C.S., Anamelechi, C.C., Clermont, E., Reichert, W.M., and Truskey, G.A. (2007). The use of mild trypsinization conditions in the detachment of endothelial cells to promote subsequent endothelialization on synthetic surfaces. Biomaterials *28*, 3928–3935. .

Bublitz, D.C., Chadwick, G.L., Magyar, J.S., Sandoz, K.M., Brooks, D.M., Mesnage, S., Ladinsky, M.S., Garber, A.I., Bjorkman, P.J., Orphan, V.J., et al. (2019). Peptidoglycan Production by an Insect-Bacterial Mosaic. Cell *179*, 703-712.e7. https://doi.org/10.1016/j.cell.2019.08.054.

Buchner, P. (1965). Endosymbiosis of animals with plant microorganisms.

Cameron, S.L. (2014). Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. Annu Rev Entomol *59*, 95–117. https://doi.org/10.1146/annurev-ento-011613-162007.

Campbell, M.A., Van Leuven, J.T., Meister, R.C., Carey, K.M., Simon, C., and McCutcheon, J.P. (2015a). Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont *Hodgkinia*. Proc Natl Acad Sci USA *112*, 10192–10199. https://doi.org/10.1073/pnas.1421386112.

Campbell, M.A., Van Leuven, J.T., Meister, R.C., Carey, K.M., Simon, C., and McCutcheon, J.P. (2015b). Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. Proceedings of the National Academy of Sciences *112*, 10192–10199. .

Campbell, M.A., Łukasik, P., Simon, C., and McCutcheon, J.P. (2017). Idiosyncratic Genome Degradation in a Bacterial Endosymbiont of Periodical Cicadas. Curr Biol *27*, 3568-3575.e3. https://doi.org/10.1016/j.cub.2017.10.008.

Campbell, M.A., Łukasik, P., Meyer, M.C., Buckner, M., Simon, C., Veloso, C., Michalik, A., and McCutcheon, J.P. (2018). Changes in Endosymbiont Complexity Drive Host-Level Compensatory Adaptations in Cicadas. MBio *9*, 10.1128/mbio.02104-18. https://doi.org/10.1128/mbio.02104-18.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

Cao, Y., and Dietrich, C.H. (2022). Phylogenomics of flavobacterial insect nutritional endosymbionts with implications for Auchenorrhyncha phylogeny. Cladistics *38*, 38–58. https://doi.org/10.1111/cla.12474.

Cao, Y., Dietrich, C.H., Zahniser, J.N., and Dmitriev, D.A. (2022). Dense sampling of taxa and characters improves phylogenetic resolution among deltocephaline leafhoppers (Hemiptera: Cicadellidae: Deltocephalinae). Systematic Entomology *47*, 430–444. https://doi.org/10.1111/syen.12540.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics *10*, 295–304. .

Chang, H.-H., Cho, S.-T., Canale, M.C., Mugford, S.T., Lopes, J.R.S., Hogenhout, S.A., and Kuo, C.-H. (2015). Complete Genome Sequence of “Candidatus Sulcia muelleri” ML, an Obligate Nutritional Symbiont of Maize Leafhopper (*Dalbulus maidis*). Genome Announc *3*. https://doi.org/10.1128/genomeA.01483-14.

Chaperon, D.-N. (2006). Construction and complementation of in-frame deletions of the essential Escherichia coli thymidylate kinase gene. Applied and Environmental Microbiology *72*, 1288–1294. .

Charles, H., Balmand, S., Lamelas, A., Cottret, L., Pérez-Brocal, V., Burdin, B., Latorre, A., Febvay, G., Colella, S., Calevro, F., et al. (2011). A Genomic Reappraisal of Symbiotic Function in the Aphid/Buchnera Symbiosis: Reduced Transporter Sets and Variable Membrane Organisations. PLoS ONE *6*, e29096. https://doi.org/10.1371/journal.pone.0029096.

Chasen, E.M., Undersander, D.J., and Cullen, E.M. (2015). Revisiting the Economic Injury Level and Economic Threshold Model for Potato Leafhopper (Hemiptera: Cicadellidae) in Alfalfa. Journal of Economic Entomology *108*, 1748–1756. https://doi.org/10.1093/jee/tov120.

Chong, R.A., and Moran, N.A. (2016). Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. Proc Natl Acad Sci USA *113*, 13114–13119. https://doi.org/10.1073/pnas.1610749113.

Chong, R.A., Park, H., and Moran, N.A. (2019). Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. Mol Biol Evol *36*, 1481–1489. https://doi.org/10.1093/molbev/msz082.

Clark, M.A., Moran, N.A., and Baumann, P. (1999). Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol Biol Evol *16*, 1586–1598. https://doi.org/10.1093/oxfordjournals.molbev.a026071.

Comas-Riu, J., and Rius, N. (2009). Flow cytometry applications in the food industry. Journal of Industrial Microbiology and Biotechnology *36*, 999–1011. https://doi.org/10.1007/s10295-009-0608-x.

ffrench-Constant, R.H., Williamson, M.S., Davies, T.G.E., and Bass, C. (2016). Ion channels as insecticide targets. Journal of Neurogenetics *30*, 163–177. https://doi.org/10.1080/01677063.2016.1229781.

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evolutionary Biology *10*, 1–21. .

Cui, F., Li, M.-X., Chang, H.-J., Mao, Y., Zhang, H.-Y., Lu, L.-X., Yan, S.-G., Lang, M.-L., Liu, L., and Qiao, C.-L. (2015). Carboxylesterase-mediated insecticide resistance: Quantitative increase induces broader metabolic resistance than qualitative change. Pesticide Biochemistry and Physiology *121*, 88–96. https://doi.org/10.1016/j.pestbp.2014.12.016.

Cui, Y., Xie, Q., Hua, J., Dang, K., Zhou, J., Liu, X., Wang, G., Yu, X., and Bu, W. (2013). Phylogenomics of Hemiptera (Insecta: Paraneoptera) based on mitochondrial genomes. Syst Ento *38*, 233–245. https://doi.org/10.1111/j.1365-3113.2012.00660.x.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat Methods *9*, 772–772. https://doi.org/10.1038/nmeth.2109.

Darzynkiewicz, Z., Juan, G., and Bedner, E. (1999). Determining cell cycle stages by flow cytometry. Current Protocols in Cell Biology *1*, 8.4. 1-8.4. 18. .

Degnan, P.H., Lazarus, A.B., Brock, C.D., and Wernegreen, J.J. (2004). Host–Symbiont Stability and Fast Evolutionary Rates in an Ant–Bacterium Association: Cospeciation of Camponotus Species and Their Endosymbionts, Candidatus Blochmannia. Syst Biol *53*, 95–110. https://doi.org/10.1080/10635150490264842.

Degnan, P.H., Lazarus, A.B., and Wernegreen, J.J. (2005). Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. *15*, 1023–1033. https://doi.org/10.1101/gr.3771305.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics *23*, 673–679. https://doi.org/10.1093/bioinformatics/btm009.

DeSalle, R., Gregory, T.R., and Johnston, J.S. (2005). Preparation of samples for comparative studies of arthropod chromosomes: visualization, in situ hybridization, and genome size estimation. In Methods in Enzymology, (Elsevier), pp. 460–488.

Després, L., David, J.-P., and Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. Trends in Ecology & Evolution *22*, 298–307. https://doi.org/10.1016/j.tree.2007.02.010.

Dettman, J.R., Sztepanacz, J.L., and Kassen, R. (2016). The properties of spontaneous mutations in the opportunistic pathogen Pseudomonas aeruginosa. BMC Genomics *17*, 27. https://doi.org/10.1186/s12864-015-2244-3.

Dietrich, C.H. (2005). KEYS TO THE FAMILIES OF CICADOMORPHA AND SUBFAMILIES AND TRIBES OF CICADELLIDAE (HEMIPTERA: AUCHENORRHYNCHA). Flen *88*, 502–517. https://doi.org/10.1653/0015-4040(2005)88[502:KTTFOC]2.0.CO;2.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dombrovsky, A., Arthaud, L., Ledger, T.N., Tares, S., and Robichon, A. (2009). Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction. Genome Res *19*, 2052–2063. https://doi.org/10.1101/gr.091611.109.

Douglas, A.E. (1989). Mycetocyte Symbiosis in Insects. Biological Reviews *64*, 409–434. https://doi.org/10.1111/j.1469-185X.1989.tb00682.x.

Douglas, A.E. (2009). The microbial dimension in insect nutritional ecology. Func Eco *23*, 38–47. https://doi.org/10.1111/j.1365-2435.2008.01442.x.

Douglas, A.E. (2011). Lessons from Studying Insect Symbioses. Cell Host Microbe *10*, 359–367. https://doi.org/10.1016/j.chom.2011.09.001.

Douglas, A.E. (2016). How multi-partner endosymbioses function. Nat Rev Microbiol *14*, 731–743. https://doi.org/10.1038/nrmicro.2016.151.

Douglas, A.E. (2017). The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. Curr Opin Insect Sci *23*, 65–69. https://doi.org/10.1016/j.cois.2017.07.012.

Dowton, M., Cameron, S.L., Austin, A.D., and Whiting, M.F. (2009). Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera–a lineage with both rapidly and slowly evolving mitochondrial genomes. Molecular Phylogenetics and Evolution *52*, 512–519. .

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol *7*, 214. https://doi.org/10.1186/1471-2148-7-214.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., and Markowitz, S. (2011). Geneious Pro v5. 5.6.

Dumas, P., Morin, M.D., Boquel, S., Moffat, C.E., and Morin, P.J. (2019). Expression status of heat shock proteins in response to cold, heat, or insecticide exposure in the Colorado potato beetle Leptinotarsa decemlineata. Cell Stress and Chaperones *24*, 539–547. https://doi.org/10.1007/s12192-019-00983-3.

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 Biology *5*, e96. https://doi.org/10.1371/journal.pbio.0050096.

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

Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. Current Opinion in Genetics & Development *4*, 255–259. .

Ehrlich, M., and Lacey, M. (2013). DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. Epigenomics *5*, 553–568. .

Elbein, A.D., Pan, Y.T., Pastuszak, I., and Carroll, D. (2003). New insights on trehalose: a multifunctional molecule. Glycobiology *13*, 17R-27R. https://doi.org/10.1093/glycob/cwg047.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology *20*, 1–14. .

Engsontia, P., Sangket, U., Chotigeat, W., and Satasook, C. (2014). Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. J Mol Evol *79*, 21–39. https://doi.org/10.1007/s00239-014-9633-0.

Fares, M.A., Moya, A., and Barrio, E. (2004). GroEL and the maintenance of bacterial endosymbiosis. Trends in Genetics *20*, 413–416. https://doi.org/10.1016/j.tig.2004.07.001.

Feder, M.E., and Hofmann, G.E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annual Review of Physiology *61*, 243–282. .

Feldhaar, H. (2011). Bacterial symbionts as mediators of ecologically important traits of insect hosts. Ecological Entomology *36*, 533–543. https://doi.org/10.1111/j.1365-2311.2011.01318.x.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., and Halpern, M.E. (2010). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences *107*, 8689–8694. .

Feron, R., and Waterhouse, R.M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. Gigascience *11*, giac006. https://doi.org/10.1093/gigascience/giac006.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends in Genetics *5*, 103–107. .

Fisher, R.A. (1992). Statistical methods for research workers. In Breakthroughs in Statistics, (Springer), pp. 66–70.

Fletcher, M., Löcker, H., Mitchell, A., and Gopurenko, D. (2017). A revision of the genus *Orosius* Distant (Hemiptera: Cicadellidae) based on male genitalia and DNA barcoding: Revision of *Orosius*. Austral Entomology *56*, 198–217. https://doi.org/10.1111/aen.12224.

Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G.A., and Maleszka, R. (2009). Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. BMC Genomics *10*, 1–11. .

Francis, F., Vanhaelen, N., and Haubruge, E. (2005). Glutathione S-transferases in the adaptation to plant secondary metabolites in the Myzus persicae aphid. Archives of Insect Biochemistry and Physiology *58*, 166–174. https://doi.org/10.1002/arch.20049.

Fronk, D.C., and Sachs, J.L. (2022). Symbiotic organs: the nexus of host–microbe evolution. Trends in Ecology & Evolution *37*, 599–610. https://doi.org/10.1016/j.tree.2022.02.014.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science *220*, 1049–1051. https://doi.org/10.1126/science.220.4601.1049.

Glastad, K.M., Hunt, B.G., and Goodisman, M.A. (2014). Evolutionary insights into DNA methylation in insects. Current Opinion in Insect Science *1*, 25–30. .

González, J., Lenkov, K., Lipatov, M., Macpherson, J.M., and Petrov, D.A. (2008). High Rate of Recent Transposable Element–Induced Adaptation in

Drosophila melanogaster. PLOS Biology *6*, e251. https://doi.org/10.1371/journal.pbio.0060251.

González, J., Karasov, T.L., Messer, P.W., and Petrov, D.A. (2010). Genome-Wide Patterns of Adaptation to Temperate Environments Associated with Transposable Elements in Drosophila. PLOS Genetics *6*, e1000905. https://doi.org/10.1371/journal.pgen.1000905.

Gossett, J.M., Porter, M.L., Vasquez, Y.M., Bennett, G.M., and Chong, R.A. (2023). Genomic comparisons reveal selection pressure and functional variation between nutritional endosymbionts of cave-adapted and epigean Hawaiian planthoppers. Genome Biology and Evolution *15*, evad031. .

Gray, M.W., Burger, G., and Lang, B.F. (2001). The origin and early evolution of mitochondria. Genome Biology *2*, 1–5. .

Greenway, G. (2022). Estimating Economic Impacts of Beet Leafhopper1 on Potato and Carrot Crops in the Columbia Basin Region of Oregon and Washington. Swen *47*, 89–96. https://doi.org/10.3958/059.047.0107.

Grimaldi, D., Engel, M.S., Engel, M.S., and Engel, M.S. (2005). Evolution of the Insects (Cambridge University Press).

Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics *36*, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

Günthardt, M.S., and Wanner, H. (1981). The feeding behaviour of two leafhoppers on Vicia faba. Ecological Entomology *6*, 17–22. https://doi.org/10.1111/j.1365-2311.1981.tb00968.x.

Gupta, A., and Nair, S. (2020). Dynamics of Insect–Microbiome Interaction Influence Host and Microbial Symbiont. Frontiers in Microbiology *11*. .

Gupta, A., and Nair, S. (2022). Heritable Epigenomic Modifications Influence Stress Resilience and Rapid Adaptations in the Brown Planthopper (Nilaparvata lugens). IJMS *23*, 8728. https://doi.org/10.3390/ijms23158728.

Halkka, O. (1960). Chromosomal Evolution in the Cicadellidae. Hereditas *46*, 581–591. https://doi.org/10.1111/j.1601-5223.1960.tb03102.x.

Han, M.V., Thomas, G.W., Lugo-Martinez, J., and Hahn, M.W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular Biology and Evolution *30*, 1987–1997. .

Hand, S.C. (1987). Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. The Biological Bulletin *173*, 260–276. .

Hansen, A.K., and Moran, N.A. (2011a). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci USA *108*, 2849–2854. https://doi.org/10.1073/pnas.1013465108.

Hansen, A.K., and Moran, N.A. (2011b). Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. Proceedings of the National Academy of Sciences *108*, 2849–2854. .

Hare, E.E., and Johnston, J.S. (2011). Genome size determination using flow cytometry of propidium iodide-stained nuclei. Molecular Methods for Evolutionary Genetics 3–12. .

Heidel-Fischer, H.M., and Vogel, H. (2015). Molecular mechanisms of insect adaptation to plant secondary compounds. Current Opinion in Insect Science *8*, 8–14. https://doi.org/10.1016/j.cois.2015.02.004.

Hembry, D.H., Bennett, G., Bess, E., Cooper, I., Jordan, S., Liebherr, J., Magnacca, K.N., Percy, D.M., Polhemus, D.A., Rubinoff, D., et al. (2021). Insect Radiations on Islands: Biogeographic Pattern and Evolutionary Process in Hawaiian Insects. The Quarterly Review of Biology *96*, 247–296. https://doi.org/10.1086/717787.

Hendry, T.A., de Wet, J.R., Dougan, K.E., and Dunlap, P.V. (2016). Genome Evolution in the Obligate but Environmentally Active Luminous Symbionts of Flashlight Fish. Genome Biol Evol *8*, 2203–2213. https://doi.org/10.1093/gbe/evw161.

Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., and Ezekowitz, R.A.B. (1999). Phylogenetic Perspectives in Innate Immunity. Science *284*, 1313–1318. https://doi.org/10.1126/science.284.5418.1313.

Hogenhout, S.A., Ammar, E.-D., Whitfield, A.E., and Redinbaugh, M.G. (2008a). Insect Vector Interactions with Persistently Transmitted Viruses. Annu. Rev. Phytopathol. *46*, 327–359. https://doi.org/10.1146/annurev.phyto.022508.092135.

Hogenhout, S.A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H.N., and Namba, S. (2008b). Phytoplasmas: bacteria that manipulate plants and insects. Molecular Plant Pathology *9*, 403–423. https://doi.org/10.1111/j.1364-3703.2008.00472.x.

Hogenhout, S.A., Van der Hoorn, R.A., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. Molecular Plant-Microbe Interactions *22*, 115–122. .

Hu, Q.-L., Ye, Y.-X., Zhuo, J.-C., Huang, H.-J., Li, J.-M., and Zhang, C.-X. (2022). Chromosome-level Assembly, Dosage Compensation and Sex-biased Gene Expression in the Small Brown Planthopper, Laodelphax striatellus. Genome Biology and Evolution *14*, evac160. https://doi.org/10.1093/gbe/evac160.

Hua, J., Smith, D.R., Borza, T., and Lee, R.W. (2012). Similar Relative Mutation Rates in the Three Genetic Compartments of Mesostigma and Chlamydomonas. Protist *163*, 105–115. https://doi.org/10.1016/j.protis.2011.04.003.

Huang, H.-L., Hsing, H.-W., Lai, T.-C., Chen, Y.-W., Lee, T.-R., Chan, H.-T., Lyu, P.-C., Wu, C.-L., Lu, Y.-C., and Lin, S.-T. (2010). Trypsin-induced proteome alteration during cell subculture in mammalian cells. Journal of Biomedical Science *17*, 1–10. .

Huelsenbeck, J.P., Larget, B., Miller, R.E., and Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. Syst Biol *51*, 673–688. https://doi.org/10.1080/10635150290102366.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

Husnik, F., and McCutcheon, J.P. (2016). Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. Proc Natl Acad Sci USA *113*, E5416–E5424. https://doi.org/10.1073/pnas.1603910113.

Husnik, F., and McCutcheon, J.P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol *16*, 67–79. https://doi.org/10.1038/nrmicro.2017.137.

Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., et al. (2013). Horizontal Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested Mealybug Symbiosis. Cell *153*, 1567–1578. https://doi.org/10.1016/j.cell.2013.05.040.

Ibrahim, S.F., and van den Engh, G. (2007). Flow cytometry and cell sorting. Cell Separation: Fundamentals, Analytical and Preparative Methods 19–39. .

Jain, R.G., Robinson, K.E., Asgari, S., and Mitter, N. (2021). Current scenario of RNAi-based hemipteran control. Pest Management Science *77*, 2188–2196. .

Johanson, K.O., and McHenry, C.S. (1980). Purification and characterization of the beta subunit of the DNA polymerase III holoenzyme of Escherichia coli. J Bio Chem *255*, 10984–10990. https://doi.org/10.1016/S0021-9258(19)70404-9.

Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., Allen, J.M., Petersen, M., Donath, A., Walden, K.K.O., et al. (2018). Phylogenomics and the evolution of hemipteroid insects. Proceedings of the National Academy of Sciences *115*, 12775–12780. https://doi.org/10.1073/pnas.1815820115.

Kanamori, Y., Saito, A., Hagiwara-Komoda, Y., Tanaka, D., Mitsumasu, K., Kikuta, S., Watanabe, M., Cornette, R., Kikawada, T., and Okuda, T. (2010). The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues. Insect Biochemistry and Molecular Biology *40*, 30–37. https://doi.org/10.1016/j.ibmb.2009.12.006.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

Kawai, A. (1997). Prospect for integrated pest management in tea cultivation in Japan. Japan Agricultural Research Quarterly *31*, 213–218. .

Keller, T.E., Han, P., and Yi, S.V. (2016). Evolutionary transition of promoter and gene body DNA methylation across invertebrate–vertebrate boundary. Molecular Biology and Evolution *33*, 1019–1028. .

Khaliq, A.M., Javed, M., Sohail, M., and Sagheer, M. (2014). Environmental effects on insects and their population dynamics. Journal of Entomology and Zoology Studies *2*, 1–7. .

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

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012). Symbiont-mediated insecticide resistance. Proc. Natl. Acad. Sci. U.S.A. *109*, 8618–8622. https://doi.org/10.1073/pnas.1200231109.

Kim, D., Minhas, B.F., Li-Byarlay, H., and Hansen, A.K. (2018). Key Transport and Ammonia Recycling Genes Involved in Aphid Symbiosis Respond to Host-Plant Specialization. G3 Genes|Genomes|Genetics *8*, 2433–2443. https://doi.org/10.1534/g3.118.200297.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology *37*, 907–915. .

Kinjo, Y., Lo, N., Martín, P.V., Tokuda, G., Pigolotti, S., and Bourguignon, T. (2021). Enhanced Mutation Rate, Relaxed Selection, and the “Domino Effect” are associated with Gene Loss in *Blattabacterium* , A Cockroach Endosymbiont. Molecular Biology and Evolution *38*, 3820–3831. https://doi.org/10.1093/molbev/msab159.

Kiryutin, B., Souvorov, A., and Tatusova, T. (2007). ProSplign–protein to genomic alignment tool. In Proc. 11th Annual International Conference in Research in Computational Molecular Biology, San Francisco, USA, p.

Kobiałka, M., Michalik, A., Walczak, M., Junkiert, Ł., and Szklarzewicz, T. (2015). Symbiotic microorganisms of the leafhopper Deltocephalus pulicaris (Fallén, 1806) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae): molecular characterization, ultrastructure and transovarial transmission. Polish Journal of Entomology *84*, 289–304. https://doi.org/10.1515/pjen-2015-0025.

Kobiałka, M., Michalik, A., Walczak, M., and Szklarzewicz, T. (2018). Dual “Bacterial-Fungal” Symbiosis in Deltocephalinae Leafhoppers (Insecta, Hemiptera, Cicadomorpha: Cicadellidae). Microb Ecol *75*, 771–782. https://doi.org/10.1007/s00248-017-1075-y.

Koga, R., and Moran, N.A. (2014a). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J *8*, 1237–1246. https://doi.org/10.1038/ismej.2013.235.

Koga, R., and Moran, N.A. (2014b). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. The Isme Journal *8*, 1237. .

Koga, R., Meng, X.-Y., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. Proc Natl Acad Sci USA *109*, E1230–E1237. https://doi.org/10.1073/pnas.1119212109.

Koga, R., Bennett, G.M., Cryan, J.R., and Moran, N.A. (2013a). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage:

Replacement of symbionts in sap-feeding insects. Environ Microbiol *15*, 2073–2081. https://doi.org/10.1111/1462-2920.12121.

Koga, R., Bennett, G.M., Cryan, J.R., and Moran, N.A. (2013b). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environmental Microbiology *15*, 2073–2081. .

Krishan, A. (1975). Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. The Journal of Cell Biology *66*, 188–193. .

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution *35*, 1547. .

Kunkel, L.O. (1926). Studies on aster yellows. American Journal of Botany 646–705. .

Kupper, M., Gupta, S.K., Feldhaar, H., and Gross, R. (2014). Versatile roles of the chaperonin GroEL in microorganism-insect interactions. FEMS Microbiol Lett *353*, 1–10. https://doi.org/10.1111/1574-6968.12390.

Kwak, Y., Argandona, J.A., Degnan, P.H., and Hansen, A.K. (2023). Chromosomal-level assembly of Bactericera cockerelli reveals rampant gene family expansions impacting genome structure, function and insect-microbe-plant-interactions. Molecular Ecology Resources *23*, 233–252. https://doi.org/10.1111/1755-0998.13693.

Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Res *6*, 1287. https://doi.org/10.12688/f1000research.12232.1.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research *35*, 3100–3108. .

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution *34*, 772–773. .

Langmead, B., and Salzberg, S.L. (2012a). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357. .

Langmead, B., and Salzberg, S.L. (2012b). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Leulier, F., and Lemaitre, B. (2008). Toll-like receptors — taking an evolutionary approach. Nat Rev Genet *9*, 165–178. https://doi.org/10.1038/nrg2303.

Li, H., Leavengood, J.M., Chapman, E.G., Burkhardt, D., Song, F., Jiang, P., Liu, J., Zhou, X., and Cai, W. (2017). Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. Proc R. Soc B *284*, 20171223. https://doi.org/10.1098/rspb.2017.1223.

Li, Y., Zhang, B., and Moran, N.A. (2020). The Aphid X Chromosome Is a Dangerous Place for Functionally Important Genes: Diverse Evolution of Hemipteran Genomes Based on Chromosome-Level Assemblies. Molecular Biology and Evolution *37*, 2357–2368. https://doi.org/10.1093/molbev/msaa095.

Li, Z., Li, Y., Xue, A.Z., Dang, V., Holmes, V.R., Johnston, J.S., Barrick, J.E., and Moran, N.A. (2022). The Genomic Basis of Evolutionary Novelties in a Leafhopper. Molecular Biology and Evolution *39*, msac184. https://doi.org/10.1093/molbev/msac184.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. .

Liu, K., Dong, Y., Huang, Y., Rasgon, J.L., and Agre, P. (2013). Impact of trehalose transporter knockdown on Anopheles gambiae stress adaptation and susceptibility to Plasmodium falciparum infection. Proceedings of the National Academy of Sciences *110*, 17504–17509. https://doi.org/10.1073/pnas.1316709110.

Long, H., Miller, S.F., Williams, E., and Lynch, M. (2018). Specificity of the DNA Mismatch Repair System (MMR) and Mutagenesis Bias in Bacteria. Mol Biol Evol *35*, 2414–2421. https://doi.org/10.1093/molbev/msy134.

LourençO, J.M., Glémin, S., Chiari, Y., and Galtier, N. (2013). The determinants of the molecular substitution process in turtles. J Evol Biol *26*, 38–50. https://doi.org/10.1111/jeb.12031.

Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biol *15*, 10–1186. .

Lowdon, R.F., Jang, H.S., and Wang, T. (2016). Evolution of epigenetic regulation in vertebrate genomes. Trends in Genetics *32*, 269–283. .

Lowe, T.M., and Chan, P.P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res *44*, W54–W57. https://doi.org/10.1093/nar/gkw413.

Lu, K., Song, Y., and Zeng, R. (2021). The role of cytochrome P450-mediated detoxification in insect adaptation to xenobiotics. Current Opinion in Insect Science *43*, 103–107. .

Luan, J.-B., Chen, W., Hasegawa, D.K., Simmons, A.M., Wintermantel, W.M., Ling, K.-S., Fei, Z., Liu, S.-S., and Douglas, A.E. (2015). Metabolic Coevolution in the Bacterial Symbiosis of Whiteflies and Related Plant Sap-Feeding Insects. Genome Biol Evol *7*, 2635–2647. https://doi.org/10.1093/gbe/evv170.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996–1000. .

Łukasik, P., Nazario, K., Van Leuven, J.T., Campbell, M.A., Meyer, M., Michalik, A., Pessacq, P., Simon, C., Veloso, C., and McCutcheon, J.P. (2018). Multiple origins of interdependent endosymbiotic complexes in a genus of cicadas. Proc Natl Acad Sci USA *115*, E226–E235. https://doi.org/10.1073/pnas.1712321115.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biology *8*, e1000506. .

Ma, W., Xu, L., Hua, H., Chen, M., Guo, M., He, K., Zhao, J., and Li, F. (2021). Chromosomal-level genomes of three rice planthoppers provide new insights into sex chromosome evolution. Molecular Ecology Resources *21*, 226–237. https://doi.org/10.1111/1755-0998.13242.

Madeira, F., Pearce, M., Tivey, A.R., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research *50*, W276–W279. .

Malinverni, J.C., Werner, J., Kim, S., Sklar, J.G., Kahne, D., Misra, R., and Silhavy, T.J. (2006). YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. Molecular Microbiology *61*, 151–164. .

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.

Mao, M., and Bennett, G.M. (2020). Symbiont replacements reset the co-evolutionary relationship between insects and their heritable bacteria. ISME J *14*, 1384–1395. https://doi.org/10.1038/s41396-020-0616-4.

Mao, M., Yang, X., Poff, K., and Bennett, G. (2017a). Comparative Genomics of the Dual-Obligate Symbionts from the Treehopper, Entylia carinata (Hemiptera: Membracidae), Provide Insight into the Origins and Evolution of an Ancient Symbiosis. Genome Biol Evol *9*, 1803–1815. https://doi.org/10.1093/gbe/evx134.

Mao, M., Yang, X., Poff, K., and Bennett, G. (2017b). Comparative genomics of the dual-obligate symbionts from the treehopper, Entylia carinata (Hemiptera: membracidae), provide insight into the origins and evolution of an ancient symbiosis. Genome Biology and Evolution *9*, 1803–1815. .

Mao, M., Yang, X., and Bennett, G.M. (2018a). Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. Proc Natl Acad Sci USA *115*, E11691–E11700. https://doi.org/10.1073/pnas.1811932115.

Mao, M., Yang, X., and Bennett, G.M. (2018b). Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. Proceedings of the National Academy of Sciences *115*, E11691–E11700. .

Martin, A.P. (1999). Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol Biol Evol *16*, 996–1002. https://doi.org/10.1093/oxfordjournals.molbev.a026189.

Mathers, T.C., Wouters, R.H.M., Mugford, S.T., Swarbreck, D., van Oosterhout, C., and Hogenhout, S.A. (2021). Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome. Molecular Biology and Evolution *38*, 856–875. https://doi.org/10.1093/molbev/msaa246.

Matsuura, Y., Moriyama, M., Łukasik, P., Vanderpool, D., Tanahashi, M., Meng, X.-Y., McCutcheon, J.P., and Fukatsu, T. (2018). Recurrent symbiont recruitment from fungal parasites in cicadas. Proc Natl Acad Sci USA *115*, E5970–E5979. https://doi.org/10.1073/pnas.1803245115.

McCutcheon, J.P., and Moran, N.A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci USA *104*, 19392–19397. https://doi.org/10.1073/pnas.0708855104.

McCutcheon, J.P., and Moran, N.A. (2010a). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biology and Evolution *2*, 708–718. .

McCutcheon, J.P., and Moran, N.A. (2010b). Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. Genome Biol Evol *2*, 708–718. https://doi.org/10.1093/gbe/evq055.

McCutcheon, J.P., and Moran, N.A. (2012a). Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol *10*, 13–26. https://doi.org/10.1038/nrmicro2670.

McCutcheon, J.P., and Moran, N.A. (2012b). Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology *10*, 13. .

McCutcheon, J.P., and von Dohlen, C.D. (2011). An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs. Curr Biol *21*, 1366–1372. https://doi.org/10.1016/j.cub.2011.06.051.

McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009a). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA *106*, 15394–15399. https://doi.org/10.1073/pnas.0906424106.

McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009b). Origin of an Alternative Genetic Code in the Extremely Small and GC–Rich Genome of a Bacterial Symbiont. PLoS Genet *5*, e1000565. https://doi.org/10.1371/journal.pgen.1000565.

McCutcheon, J.P., Boyd, B.M., and Dale, C. (2019). The Life of an Insect Endosymbiont from the Cradle to the Grave. Curr Biol *29*, R485–R495. https://doi.org/10.1016/j.cub.2019.03.032.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., and Gilbert, S.F. (2013). Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences *110*, 3229–3236. .

McKight, P.E., and Najab, J. (2010). Kruskal-wallis test. The Corsini Encyclopedia of Psychology 1–1. .

McKinnon, K.M. (2018). Flow cytometry: an overview. Current Protocols in Immunology *120*, 5.1. 1-5.1. 11. .

Metcalfe, C.J., Li, J., Giorgi, D., Doležel, J., Piperidis, N., and Aitken, K.S. (2019). Flow cytometric characterisation of the complex polyploid genome of Saccharum officinarum and modern sugarcane cultivars. Sci Rep *9*, 19362. https://doi.org/10.1038/s41598-019-55652-3.

Michalik, A., Castillo Franco, D., Kobiałka, M., Szklarzewicz, T., Stroiński, A., and Łukasik, P. (2021). Alternative Transmission Patterns in Independently Acquired

Nutritional Cosymbionts of Dictyopharidae Planthoppers. MBio *12*, e0122821. https://doi.org/10.1128/mBio.01228-21.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution *37*, 1530–1534. .

Mira, A., and Moran, N.A. (2002). Estimating Population Size and Transmission Bottlenecks in Maternally Transmitted Endosymbiotic Bacteria. Microb Ecol *44*, 137–143. https://doi.org/10.1007/s00248-002-0012-9.

Monnin, D., Jackson, R., Kiers, E.T., Bunker, M., Ellers, J., and Henry, L.M. (2020). Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis. Curr Biol *30*, 1949-1957.e6. https://doi.org/10.1016/j.cub.2020.03.011.

Montella, I.R., Schama, R., and Valle, D. (2012). The classification of esterases: an important gene family involved in insecticide resistance - A review. Mem. Inst. Oswaldo Cruz *107*, 437–449. https://doi.org/10.1590/S0074-02762012000400001.

Moran, N.A. (1996a). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA *93*, 2873–2878. https://doi.org/10.1073/pnas.93.7.2873.

Moran, N.A. (1996b). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences *93*, 2873–2878. .

Moran, N.A. (2007). Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci USA *104*, 8627–8633. https://doi.org/10.1073/pnas.0611659104.

Moran, N.A., and Bennett, G.M. (2014). The Tiniest Tiny Genomes. Annu Rev Microbiol *68*, 195–215. https://doi.org/10.1146/annurev-micro-091213-112901.

Moran, N.A., and Telang, A. (1998). Bacteriocyte-Associated Symbionts of Insects. BioScience *48*, 295–304. https://doi.org/10.2307/1313356.

Moran, N.A., Dale, C., Dunbar, H., Smith, W.A., and Ochman, H. (2003). Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. Environ Microbiol *5*, 116–126. https://doi.org/10.1046/j.1462-2920.2003.00391.x.

Moran, N.A., Tran, P., and Gerardo, N.M. (2005). Symbiosis and Insect Diversification: an Ancient Symbiont of Sap-Feeding Insects from the Bacterial Phylum Bacteroidetes. AEM *71*, 8802–8810. https://doi.org/10.1128/AEM.71.12.8802-8810.2005.

Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics *42*, 165–190. .

Moran, N.A., McLaughlin, H.J., and Sorek, R. (2009). The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. Science *323*, 379–382. https://doi.org/10.1126/science.1167140.

Morgulis, A., Gertz, E.M., Schaffer, A.A., and Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. Bioinformatics *22*, 134–141. https://doi.org/10.1093/bioinformatics/bti774.

Myers, K.N., Conn, D., and Brown, A.M.V. (2021). Essential Amino Acid Enrichment and Positive Selection Highlight Endosymbiont's Role in a Global Virus-Vectoring Pest. MSystems *6*, e01048-20. https://doi.org/10.1128/mSystems.01048-20.

Nakabachi, A., and Moran, N.A. (2022). Extreme Polyploidy of *Carsonella* , an Organelle-Like Bacterium with a Drastically Reduced Genome. Microbiol Spectr *10*, e00350-22. https://doi.org/10.1128/spectrum.00350-22.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., Ishikawa, H., Kudo, T., and Fukatsu, T. (2005a). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. Proc Natl Acad Sci USA *102*, 5477–5482. https://doi.org/10.1073/pnas.0409034102.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., Ishikawa, H., Kudo, T., and Fukatsu, T. (2005b). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. Proceedings of the National Academy of Sciences *102*, 5477–5482. .

Nakahara, Y., Shimura, S., Ueno, C., Kanamori, Y., Mita, K., Kiuchi, M., and Kamimura, M. (2009). Purification and characterization of silkworm hemocytes by flow cytometry. Developmental & Comparative Immunology *33*, 439–448. https://doi.org/10.1016/j.dci.2008.09.005.

Nauen, R., and Denholm, I. (2005). Resistance of insect pests to neonicotinoid insecticides: Current status and future prospects. Archives of Insect Biochemistry and Physiology *58*, 200–215. https://doi.org/10.1002/arch.20043.

NCBI (2017). The NCBI Eukaryotic Genome Annotation Pipeline (NCBI).

Nielsen, R., and Yang, Z. (1998). Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. Genetics *148*, 929–936. https://doi.org/10.1093/genetics/148.3.929.

Nielson, M.W. (1979). Taxonomic relationships of leafhopper vectors of plant pathogens. US Agric Res Ser Reprints.

Nikoh, N., McCutcheon, J.P., Kudo, T., Miyagishima, S., Moran, N.A., and Nakabachi, A. (2010). Bacterial Genes in the Aphid Genome: Absence of Functional Gene Transfer from Buchnera to Its Host. PLOS Genetics *6*, e1000827. https://doi.org/10.1371/journal.pgen.1000827.

Noda, H., Watanabe, K., Kawai, S., Yukuhiro, F., Miyoshi, T., Tomizawa, M., Koizumi, Y., Nikoh, N., and Fukatsu, T. (2012). Bacteriome-associated endosymbionts of the green rice leafhopper Nephotettix cincticeps (Hemiptera: Cicadellidae). Appl Entomol Zool *47*, 217–225. https://doi.org/10.1007/s13355-012-0110-1.

Nozaki, T., and Shigenobu, S. (2022). Ploidy dynamics in aphid host cells harboring bacterial symbionts. Scientific Reports *12*. https://doi.org/10.1038/s41598-022-12836-8.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., and Shukla, M. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research *42*, D206–D214. .

Padhi, A., Verghese, B., and Otta, S.K. (2009). Detecting the form of selection in the outer membrane protein C of Enterobacter aerogenes strains and Salmonella species. Microbiol Res *164*, 282–289. https://doi.org/10.1016/j.micres.2006.12.002.

Pal, A., and Vicoso, B. (2015). The X Chromosome of Hemipteran Insects: Conservation, Dosage Compensation and Sex-Biased Expression. Genome Biology and Evolution *7*, 3259–3268. https://doi.org/10.1093/gbe/evv215.

Pathak, M.D. (1968). Ecology of common insect pests of rice. Annual Review of Entomology *13*, 257–294. .

Patiño-Navarrete, R., Moya, A., Latorre, A., and Peretó, J. (2013). Comparative Genomics of Blattabacterium cuenoti: The Frozen Legacy of an Ancient Endosymbiont Genome. Genome Biol Evol *5*, 351–361. https://doi.org/10.1093/gbe/evt011.

Peng, L., Zhao, Y., Wang, H., Song, C., Shangguan, X., Ma, Y., Zhu, L., and He, G. (2017). Functional Study of Cytochrome P450 Enzymes from the Brown

Planthopper (Nilaparvata lugens Stål) to Analyze Its Adaptation to BPH-Resistant Rice. Frontiers in Physiology *8*. .

Perreau, J., Zhang, B., Maeda, G.P., Kirkpatrick, M., and Moran, N.A. (2021). Strong within-host selection in a maternally inherited obligate symbiont: *Buchnera* and aphids. Proc Natl Acad Sci USA *118*, e2102467118. https://doi.org/10.1073/pnas.2102467118.

Pers, D., and Hansen, A.K. (2021). The boom and bust of the aphid's essential amino acid metabolism across nymphal development. G3 Genes|Genomes|Genetics *11*, jkab115. https://doi.org/10.1093/g3journal/jkab115.

Price, D.R., Duncan, R.P., Shigenobu, S., and Wilson, A.C. (2011a). Genome expansion and differential expression of amino acid transporters at the aphid/Buchnera symbiotic interface. Molecular Biology and Evolution *28*, 3113–3126. .

Price, D.R.G., Tibbles, K., Shigenobu, S., Smertenko, A., Russell, C.W., Douglas, A.E., Fitches, E., Gatehouse, A.M.R., and Gatehouse, J.A. (2010). Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional characterization. Insect Molecular Biology *19*, 97–112. https://doi.org/10.1111/j.1365-2583.2009.00918.x.

Price, D.R.G., Duncan, R.P., Shigenobu, S., and Wilson, A.C.C. (2011b). Genome Expansion and Differential Expression of Amino Acid Transporters at the Aphid/Buchnera Symbiotic Interface. Mol Biol Evol *28*, 3113–3126. https://doi.org/10.1093/molbev/msr140.

Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. *26*, 342–350. https://doi.org/10.1101/gr.193474.115.

Qin, D., Zhang, L., Xiao, Q., Dietrich, C., and Matsumura, M. (2015). Clarification of the Identity of the Tea Green Leafhopper Based on Morphological Comparison between Chinese and Japanese Specimens. PLOS ONE *10*, e0139202. https://doi.org/10.1371/journal.pone.0139202.

Queffelec, J., Postma, A., Allison, J.D., and Slippers, B. (2022). Remnants of horizontal transfers of Wolbachia genes in a Wolbachia-free woodwasp. BMC Ecol Evol *22*, 36. https://doi.org/10.1186/s12862-022-01995-x.

Radyuk, S.N., Klichko, V.I., Spinola, B., Sohal, R.S., and Orr, W.C. (2001). The peroxiredoxin gene family in Drosophila melanogaster. Free Radical Biology and Medicine *31*, 1090–1100. .

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology *67*, 901. .

Rech, G.E., Bogaerts-Márquez, M., Barrón, M.G., Merenciano, M., Villanueva-Cañas, J.L., Horváth, V., Fiston-Lavier, A.-S., Luyten, I., Venkataram, S., Quesneville, H., et al. (2019). Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila. PLOS Genetics *15*, e1007900. https://doi.org/10.1371/journal.pgen.1007900.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 217–223. .

Ribeiro Lopes, M., Simonet, P., Duport, G., Gaget, K., Balmand, S., Sugio, A., Simon, J.-C., Parisot, N., and Calevro, F. (2021). Isolation of insect bacteriocytes as a platform for transcriptomic analyses. RNA Abundance Analysis: Methods and Protocols 185–198. .

Rieseberg, M., Kasper, C., Reardon, K.F., and Scheper, T. (2001). Flow cytometry in biotechnology. Appl Microbiol Biotechnol *56*, 350–360. https://doi.org/10.1007/s002530100673.

Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J.-C., Tagu, D., and Fukatsu, T. (2008). Large Gene Family Expansion and Variable Selective Pressures for Cathepsin B in Aphids. Molecular Biology and Evolution *25*, 5–17. https://doi.org/10.1093/molbev/msm222.

Roesch Goodman, K., Welter, S.C., and Roderick, G.K. (2012). GENETIC DIVERGENCE IS DECOUPLED FROM ECOLOGICAL DIVERSIFICATION IN THE HAWAIIAN NESOSYDNE PLANTHOPPERS: DIVERGENCE DECOUPLED FROM DIVERSIFICATION. Evolution *66*, 2798–2814. https://doi.org/10.1111/j.1558-5646.2012.01643.x.

Rolton, A., Delisle, L., Berry, J., Venter, L., Webb, S.C., Adams, S., and Hilton, Z. (2020). Flow cytometric characterization of hemocytes of the flat oyster, Ostrea chilensis. Fish & Shellfish Immunology *97*, 411–420. .

Rostant, W.G., Wedell, N., and Hosken, D.J. (2012). Chapter 2 - Transposable Elements and Insecticide Resistance. In Advances in Genetics, S.F. Goodwin, T. Friedmann, and J.C. Dunlap, eds. (Academic Press), pp. 169–201.

RStudio Team (2018). RStudio: Integrated Development Environment for R (Boston, MA: RStudio, Inc.).

Sabater-Muñoz, B., Toft, C., Alvarez-Ponce, D., and Fares, M.A. (2017). Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. ISME J *11*, 1291–1304. https://doi.org/10.1038/ismej.2017.18.

Santos-Garcia, D., Mestre-Rincon, N., Ouvrard, D., Zchori-Fein, E., and Morin, S. (2020). *Portiera* Gets Wild: Genome Instability Provides Insights into the Evolution of Both Whiteflies and Their Endosymbionts. Genome Biol Evol *12*, 2107–2124. https://doi.org/10.1093/gbe/evaa216.

Sarda, S., Zeng, J., Hunt, B.G., and Yi, S.V. (2012). The evolution of invertebrate gene body methylation. Molecular Biology and Evolution *29*, 1907–1916. .

Schauvliege, R., Janssens, S., and Beyaert, R. (2007). Pellino Proteins: Novel Players in TLR and IL-1R Signalling. Journal of Cellular and Molecular Medicine *11*, 453–461. https://doi.org/10.1111/j.1582-4934.2007.00040.x.

Schrader, L., Kim, J.W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun *5*, 5495. https://doi.org/10.1038/ncomms6495.

Schuler, M.A. (1996). The role of cytochrome P450 monooxygenases in plant-insect interactions. Plant Physiology *112*, 1411. .

Sharp, A.J., Spotswood, H.T., Robinson, D.O., Turner, B.M., and Jacobs, P.A. (2002). Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations. Human Molecular Genetics *11*, 3145–3156. https://doi.org/10.1093/hmg/11.25.3145.

Shcherbakov, D. (2002). The 270 million year history of Auchenorrhyncha (Homoptera) (Denisia).

Sheldon, F.H., Jones, C.E., and McCracken, K.G. (2000). Relative Patterns and Rates of Evolution in Heron Nuclear and Mitochondrial DNA. Mol Biol Evol *17*, 437–450. https://doi.org/10.1093/oxfordjournals.molbev.a026323.

Shih, H.-T., Su, C.-C., Chang, C.-J., Vargas, S., Dai, Z., and Chen, J. (2019). Draft Genome Sequence of “Candidatus Sulcia muelleri” Strain KPTW1 from Kolla paulula, a Vector of Xylella fastidiosa Causing Pierce’s Disease of Grapevine in Taiwan. Microbiology Resource Announcements *8*, 10.1128/mra.01347-18. https://doi.org/10.1128/mra.01347-18.

Silva, F.J., and Santos-Garcia, D. (2015). Slow and Fast Evolving Endosymbiont Lineages: Positive Correlation between the Rates of Synonymous and Non-Synonymous Substitution. Front Microbiol *6*. https://doi.org/10.3389/fmicb.2015.01279.

Silver, K.S., Du, Y., Nomura, Y., Oliveira, E.E., Salgado, V.L., Zhorov, B.S., and Dong, K. (2014). Voltage-Gated Sodium Channels as Insecticide Targets. Adv In Insect Phys *46*, 389–433. https://doi.org/10.1016/B978-0-12-417010-0.00005-7.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas, J., Heddi, A., and Calevro, F. (2016). Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. Scientific Reports *6*, 19967. .

Sinha, K., and Bhatnagar, R. (2010). GroEL provides protection against Bacillus anthracis infection in BALB/c mice. Molecular Immunology *48*, 264–271. .

Sloan, D.B., and Moran, N.A. (2012). Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Molecular Biology and Evolution *29*, 3781–3792. .

Sloan, D.B., Alverson, A.J., Wu, M., Palmer, J.D., and Taylor, D.R. (2012). Recent Acceleration of Plastid Sequence and Structural Evolution Coincides with Extreme Mitochondrial Divergence in the Angiosperm Genus Silene. Genome Biol Evol *4*, 294–306. https://doi.org/10.1093/gbe/evs006.

Sloan, D.B., Nakabachi, A., Richards, S., Qu, J., Murali, S.C., Gibbs, R.A., and Moran, N.A. (2014a). Parallel Histories of Horizontal Gene Transfer Facilitated Extreme Reduction of Endosymbiont Genomes in Sap-Feeding Insects. Mol Biol Evol *31*, 857–871. https://doi.org/10.1093/molbev/msu004.

Sloan, D.B., Nakabachi, A., Richards, S., Qu, J., Murali, S.C., Gibbs, R.A., and Moran, N.A. (2014b). Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Molecular Biology and Evolution *31*, 857–871. .

Smadja, C., Shi, P., Butlin, R.K., and Robertson, H.M. (2009). Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, Acyrthosiphon pisum. Molecular Biology and Evolution *26*, 2073–2086. https://doi.org/10.1093/molbev/msp116.

Smith, D.R., and Lee, R.W. (2010). Low Nucleotide Diversity for the Expanded Organelle and Nuclear Genomes of Volvox carteri Supports the Mutational-Hazard Hypothesis. Mol Biol Evol *27*, 2244–2256. https://doi.org/10.1093/molbev/msq110.

Sogin, E.M., Leisch, N., and Dubilier, N. (2020). Chemosynthetic symbioses. Curr Biol *30*, R1137–R1142. https://doi.org/10.1016/j.cub.2020.07.050.

Song, N., Liang, A.-P., and Bu, C.-P. (2012). A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. PloS One *7*, e48778. .

Spencer, J.L., and Hughson, S.A. (2023). Chapter six - Insect resistance to crop rotation. In Insect Resistance Management (Third Edition), D.W. Onstad, and L.M. Knolhoff, eds. (Academic Press), pp. 191–244.

Stapley, J., Santure, A.W., and Dennis, S.R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Molecular Ecology *24*, 2241–2252. https://doi.org/10.1111/mec.13089.

Stark, G.R., and Wahl, G.M. (1984). Gene amplification. Annual Review of Biochemistry *53*, 447–491. .

Stever, H., Eiben, J., and Bennett, G.M. (2021). Hawaiian *Nysius* Insects Rely on an Obligate Symbiont with a Reduced Genome That Retains a Discrete Nutritional Profile to Match Their Plant Seed Diet. Genome Biol Evol *13*, evab189. https://doi.org/10.1093/gbe/evab189.

Stewart, G.J., Wang, Y., and Niewiarowski, S. (1995). Methylcellulose protects the ability of anchorage-dependent cells to adhere following isolation and holding in suspension. Biotechniques *19*, 598–604. .

Sudakaran, S., Retz, F., Kikuchi, Y., Kost, C., and Kaltenpoth, M. (2015). Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. ISME J *9*, 2587–2604. https://doi.org/10.1038/ismej.2015.75.

Sudakaran, S., Kost, C., and Kaltenpoth, M. (2017). Symbiont acquisition and replacement as a source of ecological innovation. Trends in Microbiology *25*, 375–390. .

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res *34*, W609–W612. https://doi.org/10.1093/nar/gkl315.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics *9*, 465–476. .

Tabashnik, B.E. (1990). Implications of gene amplification for evolution and management of insecticide resistance. Journal of Economic Entomology *83*, 1170–1176. .

Takiya, D.M., Tran, P.L., Dietrich, C.H., and Moran, N.A. (2006a). Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Mol Ecol *15*, 4175–4191. https://doi.org/10.1111/j.1365-294X.2006.03071.x.

Takiya, D.M., Tran, P.L., Dietrich, C.H., and Moran, N.A. (2006b). Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Molecular Ecology *15*, 4175–4191. .

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res *28*, 33–36. https://doi.org/10.1093/nar/28.1.33.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., and Nikolskaya, A.N. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics *4*, 41. .

Tellis, M.B., Kotkar, H.M., and Joshi, R.S. (2023). Regulation of trehalose metabolism in insects: from genes to the metabolite window. Glycobiology *33*, 262–273. https://doi.org/10.1093/glycob/cwad011.

Toenshoff, E.R., Gruber, D., and Horn, M. (2012). Co-evolution and symbiont replacement shaped the symbiosis between adelgids (Hemiptera: Adelgidae) and their bacterial symbionts. Environmental Microbiology *14*, 1284–1295. https://doi.org/10.1111/j.1462-2920.2012.02712.x.

Trenbath, B.R. (1993). Intercropping for the management of pests and diseases. Field Crops Research *34*, 381–405. https://doi.org/10.1016/0378-4290(93)90123-5.

Tsai, J.H. (1979). Vector transmission of mycoplasmal agents of plant diseases. In The Mycoplasmas, (Elsevier), pp. 265–307.

Turner, W.F., and Pollard, H.N. (1959). Life histories and behavior of five insect vectors of phony peach disease (US Department of Agriculture).

US EPA, O. (2014). DDT - A Brief History and Status.

Van Leuven, J.T., Meister, R.C., Simon, C., and McCutcheon, J.P. (2014). Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. Cell *158*, 1270–1280. .

Van Leuven, J.T., Mao, M., Xing, D.D., Bennett, G.M., and McCutcheon, J.P. (2019). Cicada Endosymbionts Have tRNAs That Are Correctly Processed

Despite Having Genomes That Do Not Encode All of the tRNA Processing Machinery. MBio *10*. https://doi.org/10.1128/mBio.01950-18.

Vasquez, Y.M., and Bennett, G.M. (2022). A complex interplay of evolutionary forces continues to shape ancient co-occurring symbiont genomes. IScience *25*, 104786. https://doi.org/10.1016/j.isci.2022.104786.

Walberg, M., Gaustad, P., and Steen, H.B. (1996). Rapid flow cytometric assessment of mecillinam and ampicillin bacterial susceptibility. Journal of Antimicrobial Chemotherapy *37*, 1063–1075. .

Wallis, R.L. (1962). Spring migration of the six-spotted leafhopper in the western Great Plains. Journal of Economic Entomology *55*, 871–874. .

Waneka, G., Vasquez, Y.M., Bennett, G.M., and Sloan, D.B. (2021). Mutational Pressure Drives Differential Genome Conservation in Two Bacterial Endosymbionts of Sap-Feeding Insects. Genome Biology and Evolution *13*, evaa254. .

Wang, Q., Li, J., Dang, C., Chang, X., Fang, Q., Stanley, D., and Ye, G. (2018). Rice dwarf virus infection alters green rice leafhopper host preference and feeding behavior. PLOS ONE *13*, e0203364. https://doi.org/10.1371/journal.pone.0203364.

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.-H., Colbourne, J.K., Clark, A.G., and Werren, J.H. (2013). Function and evolution of DNA methylation in Nasonia vitripennis. PLoS Genetics *9*, e1003872. .

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res *40*, e49. https://doi.org/10.1093/nar/gkr1293.

Warren, D.L., Geneva, A.J., and Lanfear, R. (2017). RWTY (R We There Yet): an R package for examining convergence of Bayesian phylogenetic analyses. Molecular Biology and Evolution *34*, 1016–1020. .

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics *39*, 457–466. .

Weglarz, K.M., Havill, N.P., Burke, G.R., and von Dohlen, C.D. (2018). Partnering With a Pest: Genomes of Hemlock Woolly Adelgid Symbionts Reveal Atypical Nutritional Provisioning Patterns in Dual-Obligate Bacteria. Genome Biol Evol *10*, 1607–1621. https://doi.org/10.1093/gbe/evy114.

Weintraub, P.G., and Beanland, L. (2006). Insect Vectors of Phytoplasmas. Annual Review of Entomology *51*, 91–111. https://doi.org/10.1146/annurev.ento.51.110104.151039.

Wernegreen, J.J. (2002). Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet *3*, 850–861. https://doi.org/10.1038/nrg931.

Wernegreen, J.J. (2015). Endosymbiont evolution: predictions from theory and surprises from genomes: Endosymbiont genome evolution. Ann. N.Y. Acad. Sci. *1360*, 16–35. https://doi.org/10.1111/nyas.12740.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biom. Bull., 1, 80–83.

Wilson, A.C.C., and Duncan, R.P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. Proc Natl Acad Sci USA *112*, 10255–10261. https://doi.org/10.1073/pnas.1423305112.

Wilson, M.R., and Weintraub, P.G. (2007). An introduction to Auchenorrhyncha phytoplasma vectors. Bulletin of Insectology *60*, 177. .

Wing, J.B., Kitagawa, Y., Locci, M., Hume, H., Tay, C., Morita, T., Kidani, Y., Matsuda, K., Inoue, T., and Kurosaki, T. (2017). A distinct subpopulation of CD25− T-follicular regulatory cells localizes in the germinal centers. Proceedings of the National Academy of Sciences *114*, E6400–E6409. .

Wong, W.S.W., Yang, Z., Goldman, N., and Nielsen, R. (2004). Accuracy and Power of Statistical Methods for Detecting Adaptive Evolution in Protein Coding Sequences and for Identifying Positively Selected Sites. Genetics *168*, 1041–1051. https://doi.org/10.1534/genetics.104.031153.

Woolfit, M., and Bromham, L. (2003). Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol *20*, 1545–1555. https://doi.org/10.1093/molbev/msg167.

Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D., McCutcheon, J.P., and McDonald, B.R. (2010a). One bacterial cell, one complete genome. PloS One *5*, e10314. .

Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D., McCutcheon, J.P., McDonald, B.R., et al. (2010b). One Bacterial Cell, One Complete Genome. PLoS ONE *5*, e10314. https://doi.org/10.1371/journal.pone.0010314.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., et al. (2006). Metabolic

Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. PLoS Biol *4*, e188. https://doi.org/10.1371/journal.pbio.0040188.

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., and Kahne, D. (2005). Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in Escherichia coli. Cell *121*, 235–245. https://doi.org/10.1016/j.cell.2005.02.015.

Wyatt, G.R., and Kalf, G.F. (1957). The chemistry of insect hemolymph: II. Trehalose and other carbohydrates. The Journal of General Physiology *40*, 833–847. .

Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., and Dong, Y. (2010). Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. Nature Biotechnology *28*, 516–520. .

Xu, H.-S., Roberts, N., Singleton, F.L., Attwell, R.W., Grimes, D.J., and Colwell, R.R. (1982). Survival and viability of nonculturable Escherichia coli and Vibrio cholerae in the estuarine and marine environment. Microbial Ecology *8*, 313–323. .

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q., Coleman-Derr, D., Xia, Q., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res *47*, W52–W58. https://doi.org/10.1093/nar/gkz333.

Yan, B., Yu, X., Dai, R., Li, Z., and Yang, M. (2021). Chromosome-Level Genome Assembly of *Nephotettix cincticeps* (Uhler, 1896) (Hemiptera: Cicadellidae: Deltocephalinae). Genome Biology and Evolution *13*, evab236. https://doi.org/10.1093/gbe/evab236.

Yang, Z. (2007a). PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution *24*, 1586–1591. .

Yang, Z. (2007b). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang, Z., and Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol Evol *15*, 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A.-M.K. (2000). Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. Genetics *155*, 431–449. https://doi.org/10.1093/genetics/155.1.431.

Yang, Z., Wong, W.S.W., and Nielsen, R. (2005). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. Molecular Biology and Evolution *22*, 1107–1118. https://doi.org/10.1093/molbev/msi097.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: A Journal of Integrative Biology *16*, 284–287. .

Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res *45*, D744–D749. https://doi.org/10.1093/nar/gkw1119.

Zhang, X. (2008). The epigenetic landscape of plants. Science *320*, 489–492. .

Zhang, H., Lang, Z., and Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. Nature Reviews Molecular Cell Biology *19*, 489–506. .

Zhang, R., Zhang, X.-F., Chi, Y., Xu, Y., Chen, H., Guo, Z., and Wei, T. (2022). Nucleoprotein of a Rice Rhabdovirus Serves as the Effector to Attenuate Hemolymph Melanization and Facilitate Viral Persistent Propagation in its Leafhopper Vector. Front Immunol *13*, 904244. https://doi.org/10.3389/fimmu.2022.904244.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., and Jacobsen, S.E. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell *126*, 1189–1201. .

Zhao, Q., Shi, L., He, W., Li, J., You, S., Chen, S., Lin, J., Wang, Y., Zhang, L., Yang, G., et al. (2022). Genomic Variation in the Tea Leafhopper Reveals the Basis of Adaptive Evolution. Genomics Proteomics Bioinformatics S1672-0229(22)00098-5. https://doi.org/10.1016/j.gpb.2022.05.011.

Zhu, D.-T., Wang, X.-R., Ban, F.-X., Zou, C., Liu, S.-S., and Wang, X.-W. (2017). Methods for the extraction of endosymbionts from the whitefly Bemisia tabaci. JoVE (Journal of Visualized Experiments) e55809. .

Zimmerman, E.C. (1948). Insects of Hawaii.

eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale | Molecular Biology and Evolution | Oxford Academic.