The Genomic Basis of Evolutionary Novelties in a Leafhopper

Zheng Li [ORCID],*,[1] Yiyuan Li,[1,2] Allen Z. Xue,[1] Vy Dang,[3] V. Renee Holmes,[4] J. Spencer Johnston,[4] Jeffrey E. Barrick,[3] and Nancy A. Moran[1]

[1]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA
[2]State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo, Zhejiang, China
[3]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA
[4]Department of Entomology, Texas A&M University, College Station, TX, USA

*Corresponding author: E-mail: liz7@arizona.edu.
Associate editor: John Parsch

Abstract

Evolutionary innovations generate phenotypic and species diversity. Elucidating the genomic processes underlying such innovations is central to understanding biodiversity. In this study, we addressed the genomic basis of evolutionary novelties in the glassy-winged sharpshooter (*Homalodisca vitripennis*, GWSS), an agricultural pest. Prominent evolutionary innovations in leafhoppers include brochosomes, proteinaceous structures that are excreted and used to coat the body, and obligate symbiotic associations with two bacterial types that reside within cytoplasm of distinctive cell types. Using PacBio long-read sequencing and Dovetail Omni-C technology, we generated a chromosome-level genome assembly for the GWSS and then validated the assembly using flow cytometry and karyotyping. Additional transcriptomic and proteomic data were used to identify novel genes that underlie brochosome production. We found that brochosome-associated genes include novel gene families that have diversified through tandem duplications. We also identified the locations of genes involved in interactions with bacterial symbionts. Ancestors of the GWSS acquired bacterial genes through horizontal gene transfer (HGT), and these genes appear to contribute to symbiont support. Using a phylogenomics approach, we inferred HGT sources and timing. We found that some HGT events date to the common ancestor of the hemipteran suborder Auchenorrhyncha, representing some of the oldest known examples of HGT in animals. Overall, we show that evolutionary novelties in leafhoppers are generated by the combination of acquiring novel genes, produced both de novo and through tandem duplication, acquiring new symbiotic associations that enable use of novel diets and niches, and recruiting foreign genes to support symbionts and enhance herbivory.

Key words: Cicadellidae, Auchenorrhyncha, genome, horizontal gene transfer, brochosome, evolutionary novelty.

Introduction

Evolutionary innovations generate phenotypic novelties, help organisms expand into new ecological niches, and often result in lineage diversification (Hunter 1998; Wagner and Lynch 2010; Rabosky 2017). Classic examples of evolutionary novelties include the wings of insects (Nicholson et al. 2014) and the flowers of angiosperms (Endress 2011). Multiple hypotheses have been forwarded for how evolutionary novelties originate and evolve (Wagner 2011). At the gene level, two major molecular mechanisms have been proposed. The first is based on modifications of existing genes. Multiple studies have shown that changes in the regulation of gene expression or in existing coding gene sequences can generate phenotypic novelties (Carroll 2008; Blount et al. 2012). The second category involves the acquisition of novel genes. New genes can be acquired from foreign organisms through horizontal gene transfer (HGT, or lateral gene transfer) (Moran and Jarvik 2010;

Acuña et al. 2012; Wybouw et al. 2016; Husnik and McCutcheon 2018; McKenna et al. 2019; Xia et al. 2021), and recent genomic studies have shown that HGT is surprisingly common in some groups, including insects (Li et al. 2022). Novel genes can also arise from de novo gene birth from noncoding sequences (Long et al. 2013; Jin et al. 2021) and from gene or genome duplication followed by the divergence of paralogous genes (Ohno 1970; Lynch and Force 2000; Birchler and Yang 2022). In eukaryotes, gene duplication is the most common mechanism for creating new genes. Beyond changes to the genome itself, the acquisition of heritable, obligate symbionts is another process that confers novel phenotypes and functional capabilities (Perreau and Moran 2022).

The origin of evolutionary novelties can be complex and may involve different mechanisms at multiple levels of biological organization. In this study, we use a leafhopper called the glassy-winged sharpshooter (*Homalodisca*

Open Access
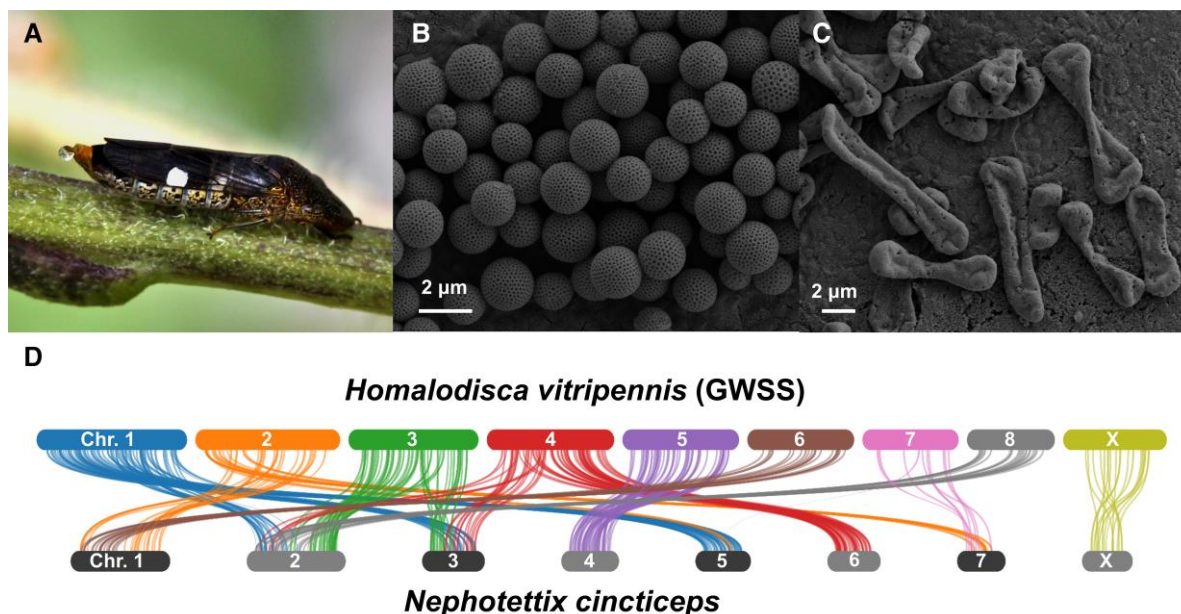
FIG. 1. (A) *Homalodisca vitripennis* female (GWSS); (B) scanning electron microscopy (SEM) of GWSS integumental brochosomes; (C) SEM of GWSS egg brochosomes; (D) syntenies between two leafhoppers, GWSS versus *Nephotettix cincticeps* (rice green leafhopper). Bars represent chromosomes. The length of the bars is proportional to the length of the chromosome-level scaffolds in the assemblies.

vitripennis, GWSS) (fig. 1A) as a study system to understand the genomic basis of several distinct types of evolutionary novelties. Leafhoppers (Insecta: Hemiptera: Cicadellidae) constitute the second largest hemipteran family with more than 20,000 species (Wahlberg et al. 2006).

One prominent evolutionary innovation in leafhoppers is the production of proteinaceous nanostructures called brochosomes (fig. 1B, C). Brochosomes are secreted into the hindgut lumen by cells in a portion of the Malpighian tubules and then excreted. Leafhoppers then spread brochosomes using distinctive setae on their hind legs to coat their bodies and, in some species, egg masses (Tulloch and Shapiro 1954; Day and Briggs 1958; Rakitov 1999; Rakitov et al. 2018). Brochosomes are hydrophobic and prevent fouling of leafhopper surfaces by their sugary exudates, and it has also been hypothesized that they are involved in camouflage and preventing microbial infections (Rakitov 2004; Rakitov and Gorb 2013a, 2013b; Yang et al. 2017). Molecular phylogenies support a single origin of brochosomes in the ancestor of leafhoppers. Four novel gene families are involved in brochosome production (Rakitov et al. 2018). The largest of these families was reported to contain 28 paralogs within a single genome, suggesting gene duplication during the evolution of these novel leafhopper nanostructures.

Most leafhoppers feed on plant phloem or xylem sap, in which essential amino acids are scarce. These nutrients are supplied by two bacterial symbionts, *Sulcia* and *Nasuia*, which are among the oldest insect symbionts, acquired over 200 million years ago in a common ancestor of the hemipteran suborder Auchenorrhyncha (Moran et al. 2005; Bennett and Moran 2013; Cao and Dietrich

2022). These symbionts are hosted in specialized organs called bacteriomes. Interestingly, the sharpshooters (subfamily Cicadellinae) acquired a new bacterial symbiont, *Baumannia*, which replaced *Nasuia* (Moran et al. 2003; Takiya et al. 2006). This acquisition coincided with a shift from phloem-feeding to xylem-feeding and resulted in the generation of a novel cell type that hosts *Baumannia* (Moran 2007).

During the long-term evolution of intracellular bacterial symbionts, such as *Sulcia* and *Nasuia*, the symbiont genomes shrink over time due to relaxed selection and genetic drift causing the loss of non-essential genes (McCutcheon and Moran 2011; Bennett and Moran 2013; Bennett and Moran 2015). In several cases, sap-feeding insects have acquired foreign genes from other microbes through HGT; some of these acquired genes support interactions with symbionts (Mao et al. 2018; Van Leuven et al. 2019; Mao and Bennett 2020). Gene acquisitions by HGT can also facilitate other evolutionary novelties in these insects. For example, aphids acquired genes for carotenoid biosynthesis from fungi (Moran and Jarvik 2010; Nováková and Moran 2012). In whiteflies, genes transferred from plants enable the detoxification of plant compounds (Xia et al. 2021). In leafhoppers and cicadas, multiple HGT events have been documented for genes that appear to support interactions with symbionts or to enhance herbivory (Husnik and McCutcheon 2018; Mao et al. 2018; Van Leuven et al. 2019; Mao and Bennett 2020). Thus, HGT appears to be a central mechanism by which sap-feeding insects have colonized new ecological niches and diversified.

In this study, we applied PacBio long-read sequencing and the Dovetail Omni-C technique to generate a chromosome-level genome assembly for the GWSS, a sap-

feeding insect that is one of the most-studied leafhopper species because it is an agricultural pest. We evaluated the quality and completeness of the assembly using karyotyping and flow cytometry. By combining this assembly with transcriptomic and proteomic data, we investigated the genomic basis of three evolutionary novelties: 1) novel proteins underlying the production of brochosomes; 2) interactions with a new symbiont, *Baumannia*; and 3) foreign gene acquisitions by HGT. We also inferred the phylogenetic placement of these HGT events. Our study reveals that a combination of different genetic mechanisms contributed to the evolutionary novelties that fueled leafhopper diversification.

Table 1. Summary of the Chromosome-level Genome Assembly of the Glassy-Winged Sharpshooter.

Category	Length (bp)	No. of genes	GC%
All	**2,304,988,238**	**21,980**	**33.14**
Chromosome 1	251,804,026	2,742	33.00
Chromosome 2	238,757,426	2,759	32.90
Chromosome 3	211,258,354	2,093	32.90
Chromosome 4	206,277,168	2,541	33.39
Chromosome 5	186,394,357	2,120	33.04
Chromosome 6	168,834,166	1,447	32.89
Chromosome 7	148,585,955	1,233	32.76
Chromosome 8	132,882,420	1,265	33.22
Chromosome X	162,104,374	1,456	33.53
Unplaced contigs	598,089,992	4,324	33.25
	(373–1,015,107)		

Results

Assembly and Annotation of the GWSS Genome

We assembled the GWSS genome from 237.9 Gb of PacBio long reads and 53.8 Gb of Omni-C reads. After genome polishing, our assembly had an overall length of 2.3 Gb with a scaffold N50 of 168.8 Mb. Nine scaffolds were larger than 132.9 Mb and suggested a haploid chromosome count of $n = 9$. We found no published karyotype information for GWSS, so we performed karyotyping on males to estimate the chromosome number and to identify the X chromosome. We found a haploid chromosome number of $n = 9$ in the GWSS, which is consistent with the nine large scaffolds from our genome assembly (supplementary fig. S1, Supplementary Material online; Table 1). The total length of the nine chromosome-level scaffolds is 1.7 Gb, which is 74.1% of the total genome assembly (Table 1; supplementary table S1, Supplementary Material online).

Most leafhoppers have an XO sex determination system (Tree of Sex Consortium 2014). To identify the X chromosome in the GWSS genome, we mapped male and female Illumina reads to our nine chromosome-level scaffolds. Given the XO sex determination system, we expect the male versus female sequencing depth of the X chromosome to be approximately half what it is for the autosomes. In our genome assembly, eight scaffolds had similar normalized read depths between sexes. By contrast, the seventh largest scaffold (162.1 Mb) had about half of the normalized sequencing read depth ratio between sexes when compared to other chromosomes (supplementary fig. S2, Supplementary Material online). Therefore, we inferred that the seventh largest scaffold to be the X chromosome. We subsequently named the other eight autosomes as chromosomes 1–8 based on their sizes, in order from longest to shortest (Table 1).

To further evaluate our assembly, we estimated genome size using flow cytometry. This provided an estimate of 1.89 Gb and 1.99 Gb for male and female genome sizes, respectively (supplementary table S2, Supplementary Material online). The X chromosome size estimated by flow cytometry (199 ± 40 Mb) is consistent with the inferred X chromosome size (162.1 Mb), further supporting its assignment as

the sex chromosome. The DNA content estimated by flow cytometry is also consistent with the total length of the nine chromosome-level scaffolds, supporting the overall accuracy of our GWSS genome assembly.

We used Benchmarking Universal Single-Copy Orthologs (BUSCO) to evaluate the completeness of the genome assembly (Simão et al. 2015). Querying the single copy orthologs of Hemiptera resulted in a BUSCO score for the GWSS assembly of 93.7% (92.7% single and duplicated, 1.0% fragmented, 6.2% missing). This score is similar to that of the other recent GWSS genome assembly (Ettinger et al. 2021). The BUSCO score for the nine chromosome-level scaffolds alone was 90.8% (86.6% single and duplicated, 4.2% fragmented, 9.2% missing) (supplementary table S3, Supplementary Material online).

The NCBI RefSeq annotation pipeline was used to annotate the genome (O'Leary et al. 2016). We used WindowMasker (Morgulis et al. 2006) to mask repetitive elements, which constituted 39.4% of the genome. We then aligned 42 GWSS transcriptomes containing 3,410,135,668 reads onto the repeat-masked genome. We predicted a total of 21,980 annotated genes with 19,904 protein-coding genes and 611 pseudogenes. 17,656 annotated genes were found on the nine chromosome-level scaffolds, with 1,456 on the X chromosome and 16,200 on the autosomes (Table 1; supplementary table S1, Supplementary Material online).

Comparative Genomics of Leafhoppers and Planthoppers

To compare the genome content and organization of leafhoppers (Hemiptera: Auchenorrhyncha: Cicadellidae) and planthoppers (Hemiptera: Auchenorrhyncha: Fulgoroidea), we used the three available chromosome-level genome assemblies, including the GWSS, *Nephotettix cincticeps* (rice green leafhopper), and *Nilaparvata lugens* (brown planthopper). Based on gene cluster assessment, 6,989 gene families are shared by these three species. The two leafhoppers share 2,337 gene clusters that are not found in *Ni. lugens* (supplementary fig. S3, Supplementary Material online). The GWSS and *Ne. cincticeps* shared 519 gene families
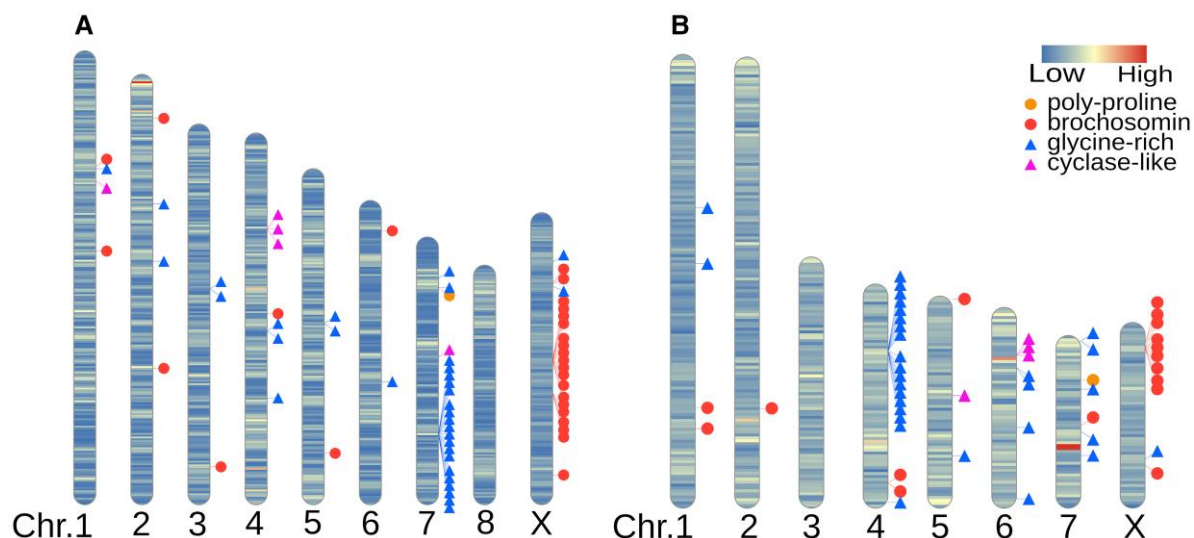
Fɪɢ. 2. Ideogram with chromosomal locations of brochosome related genes in two leafhopper genomes. The intensity of the shading on each chromosome represents gene density. (A) GWSS genome. (B) *Nephotettix cincticeps* (rice green leafhopper) genome.

that are absent in *Ni. lugens*. We also compared synteny between these three species. We observed little synteny between GWSS or *Ne. cincticeps* when aligned to *Ni. lugens* (supplementary fig. S4, Supplementary Material online). We found 113 syntenic blocks between the two leafhoppers (fig. 1D). The X chromosome exhibited similar conservation of gene content and arrangement between species to what was found for the autosomes.

Genomic Distribution of Brochosome-Related Genes

Brochosomes are produced by specialized glandular segments of the Malpighian tubules in leafhoppers (Rakitov et al. 2018). A previous study using transcriptomics and proteomics identified four major gene families related to brochosome production; these families are referred to as brochosomins, glycine-rich, poly-proline, and cyclase-like proteins (Rakitov et al. 2018). However, the molecular and genetic basis of brochosome production is still not fully understood. We used this list of brochosome-related genes to locate homologous sequences in the GWSS genome. Overall, we identified 68 brochosome-related genes from these four major gene families.

We next analyzed the locations of each brochosome-related gene in the GWSS genome (fig. 2; supplementary table S4, Supplementary Material online). We found 28 genes in the brochosomin gene family. Of these, 21 are located in two major clusters on the X chromosome (fig. 2; supplementary table S4, Supplementary Material online). We also observed 33 genes in the glycine-rich family, distributed across multiple chromosomes, with a cluster of 21 genes on chromosome 7. We found one poly-proline gene and five cyclase-like genes on autosomes.
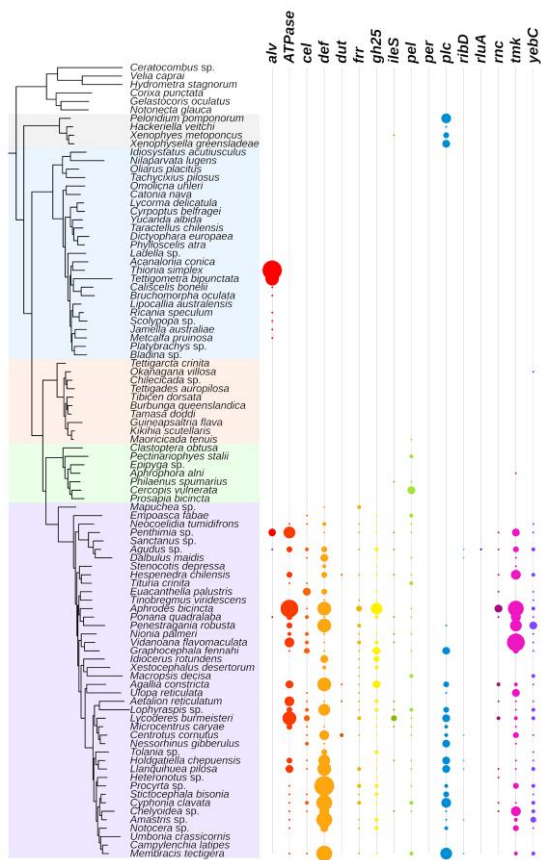
To compare repertoires of brochosome-related genes between leafhopper species, we used the reciprocal best blast hits of the GWSS brochosome-related genes to locate

homologous sequences in the *Ne. cincticeps* genome. We found 53 brochosome-related genes from the four gene families in the rice green leafhopper genome (fig. 2; supplementary table S4, Supplementary Material online). We found only 17 genes in the brochosomin gene family for *Ne. cincticeps*, suggesting that the GWSS has undergone more duplications of these loci. Nine brochosomin genes form a cluster on the likely X chromosome based on its homology to the X chromosome of the GWSS. We also found a gene cluster of 17 genes in the glycine-rich family on chromosome four. As for the GWSS, we observed one poly-proline gene and five cyclase-like genes on autosomes of *Nephotettix*.

Expression Levels of Brochosome-Related Genes

To examine the expression of brochosome-related genes, we sequenced transcriptomes from the GWSS Malpighian tubules and control samples that consisted of pooled tissues from the gut and other internal organs from the same individuals, with three biological replicates of each type of sample. We mapped transcriptome reads to the genome and performed differential expression analyses. A total of 1,190 genes are significantly upregulated in the Malpighian tubules (supplementary table S5, Supplementary Material online). Given that Malpighian tubules have other biological functions, unrelated to brochosome production, we focused on the expression of known brochosome-related genes (supplementary table S6, Supplementary Material online). Of the 68 identified brochosome-related genes, transcripts were detected for 53. Of these, 37 are significantly upregulated in the Malpighian tubules, and two are more highly expressed in other tissues (fig. 3; supplementary table S6, Supplementary Material online). The 15 genes for which transcripts were not detected might be expressed at different times or have functions unrelated to brochosome

Fig. 3. Differential expression of 68 brochosome-related genes in Malpighian tubules versus other internal organs (n Malpighian tubules = 3, other internal organs = 3). Horizontal dashed line: cut-off for the adjusted P-value < 0.05. Vertical dashed lines: Log_2 fold change > |1|.

production. Potentially, these genes have been silenced by the accumulation of mutations, but we did not detect signs of pseudogene formation, such as frameshifts or truncation.

To examine the expression of brochosome-related genes at the protein level, we used liquid chromatography-tandem mass spectrometry (LC-MS/MS) to compare the proteins found in the Malpighian tubules with those in the remaining tissues from the GWSS abdomen. Across all proteomics samples, we detected 3,537 proteins, including 22 of the 68 brochosome-related proteins. Of these 22 proteins, 13 were significantly higher in the Malpighian tubules; none were higher in other tissues (supplementary tables S6 and S7, Supplementary Material online). Twelve of these were also upregulated in the transcriptome analyses. To determine which of these proteins are also found in brochosomes, we performed LC-MS/MS proteomics on three samples containing brochosomes and other biomolecules isolated from the insect integument. We identified 15 brochosome-related proteins and five of these are among the top 150 most abundant proteins in terms of the peptide-spectrum matches (PSMs) in these samples. Twelve of these 15 proteins found in brochosomes were found significantly upregulated in the Malpighian tubule transcriptomes. Nine of these 15 proteins were also found in the Malpighian tubule proteomics samples and were significantly upregulated in the Malpighian tubule transcriptomes (fig. 3; supplementary tables S6 and S8, Supplementary Material online).

Symbiosis-related Genes in the GWSS

The GWSS has two types of bacteriomes for hosting endosymbionts. The red bacteriome contains a relatively recently (~30 Mya) acquired symbiotic bacterium, *Baumannia*. The yellow bacteriome hosts *Baumannia* and *Sulcia*, an older (~250 Mya) symbiont that is found in most leafhoppers as well as most other

Auchenorrhyncha. A previous study identified host genes that are upregulated in these two bacteriomes using transcriptome-based assemblies as the reference (Mao and Bennett 2020). To re-evaluate their differential expression analyses based on the new reference genome, we mapped this RNAseq data from this study to the GWSS genome. In the red bacteriome, 1,504 host genes are upregulated, and in the yellow bacteriome, 2,063 genes are upregulated. Overall, 521 host genes are upregulated in both bacteriomes (supplementary table S9, Supplementary Material online), which is significantly more overlap between the two sets of genes than expected by chance (hypergeometric test, $P < 10^{-273}$). We also mapped the genes upregulated in each type of bacteriome to our assembled chromosomes. The overlapping high gene density regions show that some chromosomal locations of upregulated genes are shared between the two bacteriomes (supplementary fig. S5, Supplementary Material online).

Evolutionary History of HGT Genes

Previous studies have revealed genes arising from HGT in leafhoppers, with at least fourteen such genes in the GWSS (Mao et al. 2018; Mao and Bennett 2020). To study the evolutionary history of these HGT genes, we first used the GWSS genome to identify their chromosomal locations. We observed that the majority of HGT genes are located on autosomes, with six on chromosome 1 and two on chromosome 2. One HGT gene, the *ATPase* gene, is found on the X chromosome (supplementary fig. S6, Supplementary Material online). We also confirmed that there are multiple paralogs of the horizontally acquired *cel*, *gh25*, and *plc* genes. The paralogs of *cel* and *gh25* were found on the same chromosomes (supplementary fig. S6 and table S10, Supplementary Material online).

To estimate the phylogenetic placement of the HGT sources, we assembled a dataset with 102 transcriptomes and 25 genomes across the Auchenorrhyncha (supplementary table S11, Supplementary Material online). We identified sequences that were homologous to the fourteen HGT genes in this dataset and constructed gene family phylogenies for these genes (fig. 4; supplementary fig. S7, Supplementary Material online). Based on gene presence and absence in the context of the gene tree topologies, our results support multiple independent HGT events from different bacterial sources to leafhoppers. Strikingly, the *pel* genes were present in Cercopoidea, Cicadoidea, and Membracoidea, and the gene tree topology supports a single origin through HGT in the common ancestor of these three superfamilies, which span the Auchenorrhyncha suborder. We also found that the *alv* gene was likely horizontally transferred to the common ancestor of Auchenorrhyncha (fig. 4; supplementary fig. S7 and table S10, Supplementary Material online). Similarly, *cel*, *def*, *gh25*, and the other four genes were possibly transferred to the common ancestor of Membracoidea. In contrast to these HGT genes with deep origins, the *per* and *rluA* genes were only found in one or two species of leafhoppers, suggesting relatively recent HGT events (fig. 4; supplementary fig. S7 and table S10,

Fig. 4. Phylogeny of Auchenorrhyncha and HGT gene copy numbers in transcriptomes. The phylogeny is adopted from Skinner *et al.* 2020, the shading correspond to different major lineages (Top to bottom: Outgroups; Coleorrhyncha; Fulgoridae; Cicadoidea; Cercopoidea; Membracoidea). The size of each circle represents the gene copy number found in a given transcriptome.

Supplementary Material online). We found orthologs of the *plc* gene in the Coleorrhyncha and Auchenorrhyncha, but the phylogeny of the gene tree suggests two independent transfers of *plc* in these two suborders (fig. 4; supplementary fig. S7 and table S10, Supplementary Material online).

Discussion

Genome Assembly of the GWSS

Leafhoppers are a highly diverse insect lineage and include some major agricultural pest species. The GWSS itself is a significant invasive pest species, primarily as a vector of bacterial pathogens, including *Xylella fastidiosa* in grapes and other woody hosts. To date, the only other chromosome-level genome assembly for a leafhopper is that for *Neophotettix cincticeps* in the subfamily Deltocephalinae (Yan et al. 2021). Our GWSS genome represents the subfamily Cicadellinae, which diverged from Deltocephalinae over 125 million years ago (Cao et al. 2022). Our genome N50 is higher than a recent assembly for GWSS that was based on PacBio and Nanopore reads

(Ettinger et al. 2021). Our total assembly length is larger but consistent with the other assembly. Karyotyping confirmed that the chromosome number corresponds to the number of large scaffolds in our genome assembly. Furthermore, the combined length of the chromosome-level scaffolds is close to the genome size estimated from flow cytometry. Thus, the many small scaffolds likely represent a combination of contaminants plus sequences on homologous chromosomes that did not assemble due to heterozygosity in our sample. Overall, the agreement between the assembly and empirical approaches suggests we have a high-quality and accurate genome.

Comparative Genomics of Auchenorrhyncha

We identified and compared syntenic regions between the two leafhopper genomes. The patterns imply that chromosome fission and/or fusion occurred during the divergence of these species. Genomes of leafhoppers and planthoppers show little synteny, a result that reflects the deep divergence between these two groups. In the comparison between leafhoppers, autosomes and the X chromosomes showed similar levels of synteny, in contrast to comparisons between aphids, in which X chromosome synteny is greater (Li et al. 2020; Mathers et al. 2021). In this regard, leafhoppers resemble other hemipteran insects such as kissing bugs and planthoppers, in which conservation of synteny is similar for the X and the autosomes (Mathers et al. 2021). Thus, the selective conservation of the X appears to be unique to aphids and likely reflects their unusual reproductive biology (Li et al. 2020; Mathers et al. 2021).

Brochosome and Brochosome-Related Genes

Brochosomes are novel proteinaceous structures found only in leafhoppers (Tulloch and Shapiro 1954; Day and Briggs 1958; Rakitov 1999; Rakitov et al. 2018). Four major gene families have been hypothesized to underlie brochosome production (Rakitov et al. 2018). A previous study based on transcriptomes revealed multiple copies of the brochosomin, glycine-rich, and cyclase-like gene families per genome (Rakitov et al. 2018). We found that many of these genes were organized into gene clusters. In GWSS and *Nephotettix*, there are two and one major gene clusters of brochosomin genes, respectively. In GWSS, the glycine-rich genes form a cluster on chromosome 7. Similarly, glycine-rich genes also form a cluster on chromosome 4 of *Nephotettix*. No synteny is found between chromosome 7 of the GWSS and chromosome 4 of *Nephotettix*. This observation suggests that glycine-rich genes expanded independently by tandem duplication in these two leafhopper lineages. In both genomes, we observed that cyclase-like genes have tandem duplicates on an autosome. Overall, we found that tandem duplication plays an important role in the expansion and diversification of these novel gene families. Unlike these cases, we confirmed that the remaining brochosome-associated gene family, encoding the poly-proline protein, exists

as a single copy in both the GWSS and *Nephotettix* genomes.

The two brochosomin gene clusters of the GWSS are situated on the X chromosome, and the single brochosomin gene cluster of *Nephotettix* is on what is probably the X chromosome, based on its homology to the GWSS X. Interestingly, it has been shown that some sharpshooters, including GWSS, produce two kinds of brochosomes (Rakitov 2004). These include the spherical integumental brochosomes that most leafhoppers excrete and anoint onto their integument, and the so-called egg brochosomes, which are larger and more elongated (Rakitov 2000). Egg brochosomes, produced only by females, are collected as a deposit on the wings and then used to cover the egg masses following oviposition into leaf tissue (Rakitov 2004). *Nephotettix* belongs to the Deltocephalinae, which only produces integumental brochosomes (Rakitov 2004). Potentially, the differences in brochosomin gene numbers and brochosomin gene clusters on the X chromosome are related to the female-specific production of egg brochosomes in GWSS.

In addition to genomic analyses, we used transcriptome analyses and shotgun proteomics to identify candidate genes that are important for brochosome production. Transcriptomics and shotgun proteomics of the Malpighian tubules identified 37 and 13 genes from brochosome-related gene families that are highly upregulated, respectively. We detected 15 of these proteins in proteomic analyses of partially purified brochosomes. This evidence from both Malpighian tubules and brochosomes themselves supports the importance of these genes for brochosome production. By cross-referencing genomes, transcriptomes, and proteomics, our study provides a comprehensive candidate list of brochosome-related genes of the GWSS.

Previous studies have found tremendous morphological diversity in brochosomes (Rakitov 2004; Rakitov et al. 2018). For example, the diameters of the regular spherical brochosomes produced by most species vary from 200 to 700 nm but may be as large as 5 μm (Rakitov 2004). It remains to be understood what genes and proteins are responsible for this variation and whether it has any significance in terms of insect fitness. Potentially, gene duplication has contributed to this morphological diversity. Future experiments could address the biological consequences of silencing brochosome-related genes. CRISPR/Cas9 genome editing was recently demonstrated in GWSS (de Souza Pacheco et al. 2022), and this approach might be applied to test how disrupting genes we identified impacts brochosome biogenesis and morphology. Our high quality GWSS genome and the comprehensive list of brochosome-related genes will serve as a foundation for understanding the biology of brochosomes.

Symbiosis-Related Genes in GWSS

Sharpshooters, including the GWSS, acquired a novel symbiont, *Baumannia* (Moran et al. 2003; Takiya et al. 2006).

The red bacteriome of the GWSS represents a novel organ that supports this new symbiosis, and it has a distinctive pattern of gene expression (Mao and Bennett 2020). By mapping transcriptome reads of both bacteriome types to our genome assembly, we evaluated symbiosis-related genes potentially used to support each endosymbiont. Consistent with the previous study, we found more upregulated genes in the more ancient yellow bacteriome, which hosts both *Baumannia* and *Sulcia*, potentially linked to the fact that the *Sulcia* genome is more highly reduced. The new chromosome-level genome assembly allowed us to identify the locations of symbiosis-related genes, as was done in previous analyses for aphids and a psyllid, in suborder Sternorrhyncha (Li et al. 2020). We found that symbiosis-related genes occur on all GWSS chromosomes. We did not observe an enrichment of these genes on the autosomes relative to the X chromosome, as is observed in aphids (Li et al. 2020). In the GWSS genome, a similar overall profile of chromosomal regions supports each bacteriome. In part, this reflects the significant overlap in the sets of GWSS genes that are upregulated in each type of bacteriome. This pattern is consistent with previous observations (Mao and Bennett 2020).

HGT in Auchenorrhyncha

During their evolution, insects have occasionally acquired genes from bacteria, fungi, and plants (Moran and Jarvik 2010; Acuña et al. 2012; Wybouw et al. 2016; Husnik and McCutcheon 2018; McKenna et al. 2019; Xia et al. 2021, Li et al. 2022). Previous studies found multiple HGT genes in draft genomes of the GWSS and other leafhoppers (Mao et al. 2018; Mao and Bennett 2020). We confirmed the presence of these genes in our genome assembly for GWSS and identified their chromosomal locations. We found that the majority of the HGT genes are on the autosomes of the GWSS; only the ATPase gene is located on the X chromosome. Paralogs of *cel* and *gh25* are found as tight clusters on the same chromosomes. These observations suggest that these two HGT genes have been duplicated by tandem duplications.

Previous studies have shown that some HGT genes are found in lineages of leafhoppers and cicadas, suggesting a common ancestry of these genes in these hemipteran insects (Mao et al. 2018; Van Leuven et al. 2019; Mao and Bennett 2020). However, the phylogenetic placement of the HGT events remained unresolved. To better place these HGT sources, we used a phylogenomic approach to study HGT gene presence and absence and to infer HGT gene tree topologies across Auchenorrhyncha. We found that some HGT genes likely have deep origins. For example, the *alv* gene was likely transferred to the common ancestor of the Auchenorrhyncha, which places the time of this event at around 250–300 Mya (Johnson et al. 2018; Cao and Dietrich 2022), and possibly lost in the Cicadoidea (cicadas), Cercopoidea (spittlebugs), and Membracidae (treehoppers). Similarly, the *pel* genes are shared by the Cicadoidea, Cercopoidea, and

Membracoidea. Based on molecular dating of the species phylogeny (Johnson et al. 2018), *pel* was acquired by the common ancestor of these major lineages around 250 Mya. These HGT genes are some of the oldest known examples of HGT in animals (Husnik and McCutcheon 2018). In contrast, *per* and *rluA* are only found within the Deltocephalinae, suggesting relatively recent HGT events.

Interestingly, we found the *plc* gene in the genomes of moss bugs (Coleorrhyncha) and leafhoppers. The gene tree topology supports independent acquisitions of *plc* in these two lineages. Our gene tree supports an origin from *Providencia* in the Membracoidea (Mao et al. 2018; Mao and Bennett 2020), whereas moss bugs possibly acquired *plc* from a different bacterium. The *plc* genes have functions related to lipid metabolism, and they are not highly expressed in the bacteriomes of GWSS (Mao et al. 2018; Mao and Bennett 2020). The function and expression pattern of the *plc* genes in Coleorrhyncha is unknown. We identified different numbers of paralogs of HGT genes among closely related species. Although some of the variation may be due to the incomplete nature of transcriptomes, some is likely due to different levels of gene duplication following the HGT event. Overall, our study shows repeated HGT-based acquisitions of foreign genes during the evolution of Auchenorrhyncha.

Conclusion

A broad question in evolution is how novel phenotypic features originate, and central to that question is the problem of how novelty can arise from existing genomic repertoires. Leafhoppers exhibit some striking new features, including brochosome production and highly specialized relationships with intracellular symbionts. Our analysis of a high-quality leafhopper genome gives insight into the genomic basis of these features. We show that both acquisition of foreign genes and gene duplications contributed. Some questions remain that will require sequencing of additional insect species to resolve. For example, some of the brochosome-related genes belong to novel gene families with unknown origins.

Materials and Methods

Sample Preparation for Genome Sequencing

Multiple *H. vitripennis* individuals were collected from crape myrtles (*Lagerstroemia indica*) near the campus of the University of Texas at Austin. For a high-quality genome assembly, a total of 0.6 g of male and female adult individuals were frozen and shipped to Dovetail Genomics (Santa Cruz, CA, USA). Three male individuals were used for DNA extraction and Pacbio long-read sequencing. Other male and female individuals were pooled for Dovetail Omni-C library preparation. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30× sequence coverage.

Assembly of the *H. vitripennis* Genome

To assemble the *H. vitripennis* genome, 237.9 Gb of PacBio CLR reads were used as an input to WTDBG2 v2.5 (Ruan and Li 2020), minimum read length 20000, and minimum alignment length 8192. Additionally, realignment was enabled with the "-R" option, and read type was set with the option "-x sq." Blast results of the WTDBG2 output assembly against the NCBI nt database were used as input for blobtools v1.1.1 (Laetsch and Blaxter 2017), and scaffolds identified as possible contamination were removed from the assembly. Finally, purge_dups v1.2.3 (Guan et al. 2020) was used to remove haplotypic duplications.

The de novo genome assembly and Dovetail Omni-C library reads were used as input data for genome scaffolding using the HiRise assembler version v2.1.6–072ca03871cc (Putnam et al. 2016). Dovetail Omni-C library sequences were aligned to the draft input assembly using bwa (Li and Durbin 2009). The separations of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for the genomic distance between read pairs, and the model was used to identify and break putative misjoins, score prospective joins, and make joins above a threshold. The HiRise scaffolds were then polished by Nextpolish (Hu et al. 2020) using the PacBio long reads and Omni-C library short reads used in the genome assembly and scaffolding.

To evaluate the completeness of our genome assembly, BUSCO version 3.0.2 (Simão et al. 2015) was used on the chromosome-level assembly using the single-copy orthologous gene set for Hemiptera from OrthoDB version 9 (Zdobnov et al. 2017).

Genome Annotation

The NCBI Eukaryotic Genome Annotation Pipeline was used for genome annotation (O'Leary et al. 2016). Repeat families found in the genome assemblies of *H. vitripennis* were identified and masked using WindowMasker (Morgulis et al. 2006). Over 20,000 transcripts of GWSS and high-quality proteins of GWSS and other closely related insects were retrieved from Entrez, aligned to the genome by Splign (Kapustin et al. 2008), minimap2 (Li 2018), or ProSplign (https://www.ncbi.nlm.nih.gov/sutils/static/prosplign/prosplign.html). Additionally, 3,410,135,668 reads from 42 GWSS RNA-Seq datasets were also aligned to the repeat-masked genome. Protein, transcript, and RNA-Seq read alignments were passed to Gnomon for gene prediction. Gnomon predictions selected for the final annotation set were assigned to models based on known and curated RefSeq and models based on Gnomon predictions. The overall quality of the annotations was assessed using BUSCO v4 (Seppey et al. 2019). The detailed annotation pipeline can be found at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/.

Chromosome Number Confirmation by Karyotyping

Three male adult individuals of *H. vitripennis* were collected on the campus of the University of Texas at

Austin. For karyotyping, the insects were injected with 50 μL of 2% colchicine solution in the abdomen and left at room temperature overnight. They were then dissected in 1X PBS solution to separate and remove both testicular follicles from the upper abdomen. Testicular follicles were each transferred to 100 μL of 0.075 M sodium citrate solution for 10 min and subsequently fixed in modified Carnoy's solution (3:1 absolute ethanol:glacial acetic acid) for 1 hour. Finally, they were each added to 100 μL of 50% acetic acid and the tissue was homogenized by blowing air into the solution with a micropipette. We then spotted 10 μL of the acetic acid solution containing testicular follicle cells onto slides and allowed it to air-dry at room temperature. All samples were stained with 15 μL Giemsa stain (5%) for 30 min, then rinsed completely and mounted in deionized water. The slides were viewed under a Nikon Eclipse te2000-u inverted fluorescence microscope, and cells with clear chromosome segregation were recorded and photographed with a Nikon DS-Ri2 Microscope Camera.

Genome Size Estimation by Flow Cytometry

Genome size was estimated as described in Johnston et al. 2019. In brief, a *H. vitripennis* head with unknown genome size and a *Drosophila virilis* female head standard (1C = 328 Mbp) were placed together into 1 mL of ice-cold Galbraith buffer in a 2 mL Dounce tissue grinder. Nuclei were released from both tissues by grounding with 15 gentle strokes using an A (loose) pestle, then filtered through nylon mesh into a 1.5 mL microfuge tube, and stained with 25 μl of propidium iodide (1 mg/ml) for 2 hours in the dark at 4°C. The mean red PI fluorescence from 2C sample and standard nuclei was quantified using a CytoFlex flow cytometer (Beckman Coulter). Haploid (1C) DNA quantity was calculated as (2C sample mean fluorescence/2C standard mean fluorescence) × 328 Mbp. The X chromosome genome size was estimated as the genome size difference between XX females and XO males.

Assignment of the X Chromosome and Autosomes

The X chromosome was assigned following the method previously used in the pea aphid and psyllid genomes (Li et al. 2019; Li et al. 2020). We mapped whole genome sequencing reads from male and female individuals back to our chromosome-level genome assembly. The male and female sequencing reads were obtained from the i5K insects genome project (Thomas et al. 2020) through GenBank (BioProject: PRJNA168119, Accession: SRX326930, SRX326929, SRX326928, SRX326927) and were cleaned with Trimmomatic version 0.38 (Bolger et al. 2014). The clean reads were mapped to the chromosome-level assembly using Bowtie2 version 2.3.4.3 (Langmead and Salzberg 2012) with default parameters. The resulting SAM files were converted to BAM files, sorted, and indexed using SAMtools version 1.9 (Danecek et al. 2021). We estimated the sequencing depth based on 10-kb sliding windows with 2-kb steps, and the

sequencing depth of each window was estimated using Mosdepth version 0.2.3 (Pedersen and Quinlan 2018). We normalized the overall sequencing depths among male individuals and female individuals based on methods used in Li et al. 2020. The overall sequencing depth distribution was plotted using a violin plot in ggplot2 version 3.2.1 (Wickham 2016). The X chromosome assigned to the chromosome had about half the ratio of sequencing depth between males and females compared to the others.

Comparative Genomics and Genome Synteny Analyses

We compared the genomes of the GWSS, *Nephotettix cincticeps* (rice green leafhopper) (Yan et al. 2021), and *Nilaparvata lugens* (brown planthopper) (Ye et al. 2021). The protein sequences of each genome were downloaded. We used OrthoVenn2 (Xu et al. 2019) with e-value = 1e−5 and inflation value = 1.5 to cluster orthologous groups and to create a Venn diagram of these three genomes. We used MCScanX (Wang et al. 2012) to evaluate the whole genome synteny between these species. All parameters were used as defaults in MCScanX. SynVisio (https://github.com/kiranbandi/synvisio) was used to display syntenies between genomes.

Genomic Distribution of Brochosome-related Genes

We used the list of brochosome-related genes as blastp queries to locate homologous sequences in the GWSS. We manually curated homologs and annotated them on our genome assembly using the "protein2genome" mode of exonerate version 2.2.0 (Slater and Birney 2005). Similarly, we used reciprocal best blast hits of the GWSS brochosome-related genes to locate homologous sequences in the *Ne. cincticeps* genome. These homologous sequences were used as queries in the second round of blastp with the annotated proteins of *Ne. cincticeps* genome as the database. The location of each brochosome-related gene was shown in an ideogram produced by the R package Rideogram (Hao et al. 2020). The gene density was calculated with a sliding window size of 1 Mb.

RNAseq Differential Expression Analyses of Brochosome-Related Genes

To perform differential expression analyses of brochosome-related genes, we generated transcriptomes for Malpighian tubules and the rest of the organs of the abdomen of *H. vitripennis* with three biological replicates. For each paired set of transcriptomes, we used four frozen adult males for dissection and RNA extraction. The insects were dissected in cold 1X PBS solution to remove all organs from the abdomen. All the Malpighian tubules were separated from the rest of the organs of the abdomen and pooled into a 1.7 mL microcentrifuge tube containing 200 μL PBS. The remaining organs were also pooled in 200 μL PBS. The RNA was then extracted from both samples using the Quick-RNA MiniPrep kit following the manufacturer's instructions (Zymo Research, Irvine, CA). Total RNA in

samples was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), and RNA integrity was checked with a 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Sequencing libraries were quantified using a Qubit and validated using a TapeStation as well as by quantitative PCR (Applied Biosystems, Carlsbad, CA, USA). The sequencing libraries were multiplexed and clustered onto a flow cell and loaded on an Illumina HiSeq 4000 using a 2×150 paired-end (PE) configuration.

Differential gene expression analyses were performed to identify brochosome-related genes that are significantly up-regulated in the Malpighian tubules. We used a list of candidate brochosome-related genes from a previous study that was based on proteomics and transcriptomics of the leaf-hopper *Graphocephala fennahi* (Rakitov et al. 2018). The protein sequences of these candidate genes were used as queries in blastp against all proteins inferred from our genome and from the other two GWSS genomes in GenBank. We used the same approach from the previous section for sequence read cleaning. The clean reads were mapped to the chromosome-level assembly using HISAT2 version 2.1.0 (Kim et al. 2019) with -k 3. We used featureCounts (Liao et al. 2014) to estimate the number of reads mapped to the exons of each candidate gene (–type exon). Counts of the genes were normalized, and we identified differentially expressed genes using DESeq2 version 1.20.0 in R (Love et al. 2014; R Core Team 2014). The Malpighian tubules were treated as one condition and the rest of the organs of the abdomen sample was used as the other condition. The Wald significance test was used to identify differentially expressed genes. EnhancedVolcano (https://github.com/kevinblighe/EnhancedVolcano) was used to make the volcano plot.

Shotgun Proteomics (LC-MS/MS) of Malpighian Tubules

Multiple *H. vitripennis* individuals were collected from wild sunflowers (*Helianthus annuus*) near the entrance of the McKinney Roughs Nature Park at Cedar Creek, TX, and from crape myrtles (*Lagerstroemia indica*) near the campus of the University of Texas at Austin. To dissect the Malpighian tubules, female *H. vitripennis* individuals were frozen at −20°C for 5 min, then dissected in 1X PBS solution to extract the organs from the abdomen. Malpighian tubules were separated from the rest of the organs. The Malpighian tubules from four individuals were pooled into 1.7 mL microcentrifuge tubes containing 200 μL 1X PBS. For comparison, the remaining organs of the abdomen from two individuals (to achieve roughly equivalent protein concentration as the Malpighian tubules) were also pooled (Malpighian tubules: $n = 4$; gut: $n = 4$). All samples were then centrifuged at 4,000 x g for 10 min at 4°C to pellet the tissue. Following the removal of the supernatant, the pellet was homogenized with a pestle. Two different lysis buffers were added to the samples. For six samples (Malpighian tubules: $n = 3$; gut: $n = 3$), 300 μL of immunoprecipitation (IP) lysis buffer (25 mM Tris, 1 mM

ethylenediaminetetraacetic acid [EDTA], 5% glycerol, 0.15 M NaCl, 1% Triton X-100) was added. For two samples (Malpighian tubules: $n = 1$; gut: $n = 1$), 300 μL of sodium dodecyl sulfate (SDS) lysis buffer (0.05% SDS, 5 mM EDTA pH 8.0, 1 mM phenylmethylsulfonyl fluoride [PMSF], 10 mM dithiothreitol [DTT], 0.8 mg DNase I) was added. All samples were then centrifuged at 12,000 x g for 20 min at 4°C. The supernatants were transferred to clean microcentrifuge tubes and stored at −20°C until needed. Protein concentration was quantified with the Quick Start Bovine Serum Albumin Bradford assay according to the manufacturer's protocol (Bio-Rad Laboratories, Inc., Hercules, CA) and using an Eppendorf Biophotometer 6131.

In preparation for protein electrophoresis, Malpighian tubule samples were diluted in water to attain appropriate protein concentrations, giving a total of 15 μL volume per sample. 3 μL 6X SDS loading dye containing reducing agent (0.6 M DTT, 0.35 M Tris pH 6.8, 30% v/v glycerol, 10% w/v SDS, 0.012% w/v bromophenol blue) was added to each sample. Samples were then heated at 100°C for 10 min, allowed to cool to room temperature, and centrifuged for 30 s. The samples were run in 12% polyacrylamide gel for 20 min at 70 V and a constant 30 mA. The gel was then stained with Coomassie G-250 for 30 min and destained with 10% ethanol and 5% acetic acid solution overnight. Sample lanes were cut from the gel and stored in the same destain solution at 4°C until trypsin digestion.

The LC-MS/MS shotgun proteomics was performed by the Proteomics Core at UT Austin. Raw LC-MS/MS spectra were processed using Proteome Discoverer (v2.3) (Orsburn 2021). We used the Percolator node in Proteome Discoverer to assign unique peptide spectral matches (PSMs) at false discovery rate (FDR) <5% to the composite form of the GWSS reference proteome which comprises 31,235 proteins. In order to identify proteins statistically significantly associated with each bait, we calculated both a $\log_2$ fold-change and a Z-score for each protein based on the observed PSMs in the bait (expt) versus control (ctrl) pulldown. The fold-change was computed for each protein according to the methods reported in a recent study (Lee et al. 2020).

Shotgun Proteomics (LC-MS/MS) of Purified Brochosomes

To obtain purified brochosomes of *H. vitripennis*, five female individuals were stored at −80°C and soaked in acetone for 25 min. The solution was then sonicated using a Bransonic 2800 bath sonicator for 3 min to suspend the brochosome particles in the solution. The brochosome solution was then filtered using a glass fiber syringe filter with a pore size of 1 μm to remove impurities and leafhopper fragments. To further concentrate the brochosome solution, the filtered solution was centrifuged using an Eppendorf 5804R centrifuge at 16,000 x g for 25 min. After centrifugation, the supernatant was removed and the concentrated product was stored in ambient conditions.

In preparation for protein electrophoresis, the dried pellet samples were resuspended in 18 µL of Laemmli's buffer (50 mM Tris-HCl pH 8.8, 2% SDS, 10% glycerol, 0.01% bromophenol blue, 100 mM DTT) (Rakitov et al. 2018). They were then heated at 100°C for 10 min, allowed to cool to room temperature, and centrifuged for 30 s. The samples were then run on a protein gel and further analyzed in the same manner as the Malpighian tubule samples.

Identifying Symbiosis-Related Genes With Differential Expression Analyses

Using transcriptome data from a previous study (Mao and Bennett 2020), we compared transcriptomes of three tissue types: red bacteriome, yellow bacteriome, and the remainder of the body, with three biological replicates for each tissue type. These data were downloaded through the AWS links on Genbank (BioProject: PRJNA342859, Accession: SRR10060917, SRR10060918, SRR10060919). We used Trimmomatic version 0.38 (Bolger et al. 2014) to remove low-quality reads. Filtered reads were mapped to our chromosome-level assembly using HISAT2 version 2.1.0 (Kim et al. 2019) with -k 3. We used featureCounts (Liao et al. 2014) to estimate the number of reads mapped to the exons of each gene (–type exon). Counts of the genes were normalized, and we identified differentially expressed genes using DESeq2 version 1.20.0 in R (Love et al. 2014; R Core Team 2014). To take into account biases due to variation in sequencing depth and RNA composition, the median of ratios method implemented in DESeq2 was used for normalization. Given that there were three tissue types, two differential analyses were performed. Each bacteriome type was treated as one condition and the body as the other condition. The Wald significance test was used to identify differentially expressed genes. Symbiosis-related genes were identified as genes that were >2-fold upregulated in the two bacteriomes compared to the body. The statistical significance of overlapping upregulated host genes between two bacteriomes was tested with a hypergeometric test in package "stats" version 3.6.1 in R (R Core Team 2014).

Evolutionary History of HGT Genes

Previous studies identified multiple bacterial genes that were horizontally transferred from bacteria into leafhoppers using the draft genomes of *H. vitripennis* and *Macrosteles quadrilineatus* (Mao et al. 2018; Mao and Bennett 2020). To understand the evolutionary sources of HGT genes previously identified in these two leafhoppers, we used their translated sequences of HGT genes as blastp queries. We also assembled the translated protein sequences of 25 hemipteran genomes and 102 transcriptomes as our blast database. A blastp search with an E-value of 1e −5, a bit-score of 100, and percent identity of 40% as thresholds were used to find homologous protein sequences of queries from the blast database. All protein sequences from the blastp hits were clustered into gene families using OrthoFinder version 2.3.12 (Emms

and Kelly 2019) with defaults. The bacterial protein sequences used as outgroups in previous studies (Mao et al. 2018; Mao and Bennett 2020) were also downloaded from Genbank. For each HGT gene family plus outgroup sequences, homologous protein sequences were aligned using MUSCLE v3.8.31 (Edgar 2004) with default parameters. Sequences with low overlap in the protein alignments were removed by using -resoverlap 0.75 -seqoverlap 60 in trimAl v1.2 (Capella-Gutiérrez et al. 2009). The "-gappyout" option was used to remove columns with many gaps from the protein sequence alignments. We then manually inspected each protein alignment and removed any ambiguous sequences. Finally, gene trees were built with IQ-TREE multicore version 1.6.1 with 1000 bootstrap replicates using models selected by MFP ModelFinder (Nguyen et al. 2015).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

Homalodisca vitripennis genome assembly and sequencing data are available on NCBI under the BioProject ID PRJNA776266. RefSeq genome annotation is available at: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/021/130/785/GCF_021130785.1_UT_GWSS_2.1/; Annotation information is available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Homalodisca_vitripennis/100/; Genome information is available at: https://www.ncbi.nlm.nih.gov/genome/? term=txid197043[orgn]

References

Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ, *et al.* 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci USA* **109**:4197–4202.

Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol* **5**:1675–1688.

Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci USA* **112**:10169–10176.

Birchler JA, Yang H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* **00**:1–9.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**:513–518.

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

Cao Y, Dietrich CH. 2022. Phylogenomics of flavobacterial insect nutritional endosymbionts with implications for Auchenorrhyncha phylogeny. *Cladistics* **38**:38–58.

Cao Y, Dietrich CH, Zahniser JN, Dmitriev DA. 2022. Dense sampling of taxa and characters improves phylogenetic resolution among deltocephaline leafhoppers (Hemiptera: Cicadellidae: Deltocephalinae). *Syst Entomol* **47**:430–444.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**:25–36.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**:giab008.

Day MF, Briggs M. 1958. The origin and structure of brochosomes. *J Ultrastruct Res* **2**:239–244.

de Souza Pacheco I, Doss A-LA, Vindiola BG, Brown DJ, Ettinger CL, Stajich JE, Redak RA, Walling LL, Atkinson PW. 2022. Efficient CRISPR/Cas9-mediated genome modification of the glassy-winged sharpshooter *Homalodisca vitripennis* (Germar). *Sci Rep* **12**:6428.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf* **5**:113.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**:238.

Endress PK. 2011. Evolutionary diversification of the flowers in angiosperms. *Am J Bot* **98**:370–396.

Ettinger CL, Byrne FJ, Collin MA, Carter-House D, Walling LL, Atkinson PW, Redak RA, Stajich JE. 2021. Improved draft reference genome for the glassy-winged sharpshooter (*Homalodisca vitripennis*), a vector for Pierce's disease. *G3 (Bethesda)* **11**:jkab255.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**:2896–2898.

Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**:e251.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**:2253–2255.

Hunter JP. 1998. Key innovations and the ecology of macroevolution. *Trends Ecol Evol* **13**:31–36.

Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol* **16**:67–79.

Jin G, Ma P-F, Wu X, Gu L, Long M, Zhang C, Li D-Z. 2021. New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos. *Mol Biol Evol* **38**:5752–5768.

Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KKO, *et al.* 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc Natl Acad Sci USA* **115**:12775–12780.

Johnston JS, Bernardini A, Hjelmen CE. 2019. Genome size estimation and quantitative cytogenetics in insects. *Methods Mol Biol* **1858**:15–26.

Kapustin Y, Souvorov A, Tatusova T, Lipman D. 2008. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* **3**:20.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**:907–915.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res.* **6**:1287.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359.

Lee C, Cox RM, Papoulas O, Horani A, Drew K, Devitt CC, Brody SL, Marcotte EM, Wallingford JB. 2020. Functional partitioning of a liquid-like organelle during assembly of axonemal dyneins. *eLife* **9**:e58662.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.

Li Y, Liu Z, Liu C, Shi Z, Pang L, Chen C, Chen Y, Pan R, Zhou W, Chen X-X, *et al.* 2022. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**:1–13.

Li Y, Park H, Smith TE, Moran NA. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol Biol Evol* **36**:2143–2156.

Li Y, Zhang B, Moran NA. 2020. The aphid X chromosome is a dangerous place for functionally important genes: diverse evolution of hemipteran genomes based on chromosome-level assemblies. *Mol Biol Evol* **37**:2357–2368.

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

Long M, VanKuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: little did we know. *Annu Rev Genet* **47**:307–333.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**:459–473.

Mao M, Bennett GM. 2020. Symbiont replacements reset the co-evolutionary relationship between insects and their heritable bacteria. *ISME J* **14**:1384–1395.

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

Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA. 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol Biol Evol* **38**:856–875.

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

McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H, *et al.* 2019. The evolution and genomic basis of beetle diversity. *Proc Natl Acad Sci USA* **116**:24729–24737.

Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA* **104**(Suppl 1):8627–8633.

Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. 2003. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. *Environ Microbiol* **5**:116–126.

Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**:624–627.

Moran NA, Tran P, Gerardo NM. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl Environ Microbiol* **71**:8802–8810.

Morgulis A, Gertz EM, Schäffer AA, Agarwala R. 2006. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* **22**:134–141.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274.

Nicholson DB, Ross AJ, Mayhew PJ. 2014. Fossil evidence for key innovations in the evolution of insect diversity. *Proc R Soc Lond B Biol Sci* **281**:20141823.

Nováková E, Moran NA. 2012. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Mol Biol Evol* **29**:313–323.

Ohno S. 1970. *Evolution by gene duplication*. Berlin: Springer.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**:D733–D745.

Orsburn BC. 2021. Proteome Discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* **9**:15.

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**:867–868.

Perreau J, Moran NA. 2022. Genetic innovations in animal–microbe symbioses. *Nat Rev Genet* **23**:23–39.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* **26**:342–350.

R Core Team. 2014. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rabosky DL. 2017. Phylogenetic tests for evolutionary innovation: the problematic link between key innovations and exceptional diversification. *Philos Trans R Soc Lond B Biol Sci* **372**:20160417.

Rakitov RA. 1999. Secretory products of the Malpighian tubules of Cicadellidae (Hemiptera, Membracoidea): an ultrastructural study. *Int J Insect Morphol Embryol* **28**:179–193.

Rakitov RA. 2000. Secretion of brochosomes during the ontogenesis of a leafhopper, *Oncometopia orbona* (F.) (Insecta, Homoptera, Cicadellidae). *Tissue Cell* **32**:28–39.

Rakitov RA. 2004. Powdering of egg nests with brochosomes and related sexual dimorphism in leafhoppers (Hemiptera: Cicadellidae). *Zool J Linn Soc* **140**:353–381.

Rakitov R, Gorb SN. 2013a. Brochosomal coats turn leafhopper (Insecta, Hemiptera, Cicadellidae) integument to superhydrophobic state. *Proc Biol Sci* **280**:20122391.

Rakitov R, Gorb SN. 2013b. Brochosomes protect leafhoppers (Insecta, Hemiptera, Cicadellidae) from sticky exudates. *J R Soc Interface* **10**:20130445.

Rakitov R, Moysa AA, Kopylov AT, Moshkovskii SA, Peters RS, Meusemann K, Misof B, Dietrich CH, Johnson KP, Podsiadlowski L, et al. 2018. Brochosomins and other novel proteins from brochosomes of leafhoppers (Insecta, Hemiptera, Cicadellidae). *Insect Biochem Mol Biol* **94**:10–17.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**:155–158.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* **1962**:227–245.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.

Skinner RK, Dietrich CH, Walden KKO, Gordon E, Sweet AD, Podsiadlowski L, Petersen M, Simon C, Takiya DM, Johnson KP. 2020. Phylogenomics of Auchenorrhyncha (Insecta: Hemiptera) using transcriptomes: examining controversial relationships via degeneracy coding and interrogation of gene conflict. *Syst Entomol.* **45**:85–113.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf* **6**:31.

Takiya DM, Tran PL, Dietrich CH, Moran NA. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. *Mol Ecol* **15**:4175–4191.

Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, et al. 2020. Gene content evolution in the arthropods. *Genome Biol* **21**:15.

Tree of Sex Consortium. 2014. Tree of Sex: a database of sexual systems. *Sci Data* **1**:140015.

Tulloch GS, Shapiro JE. 1954. Brochosomes and leafhoppers. *Science* **120**:232–232.

Van Leuven JT, Mao M, Xing DD, Bennett GM, McCutcheon JP. 2019. Cicada endosymbionts have tRNAs that are correctly processed despite having genomes that do not encode all of the tRNA processing machinery. *mBio* **10**:e01950–18.

Wagner A. 2011. The molecular origins of evolutionary innovations. *Trends Genet* **27**:397–410.

Wagner GP, Lynch VJ. 2010. Evolutionary novelties. *Curr Biol* **20**:R48–R52.

Wahlberg N, Grimaldi D, Engel MS. 2006. Evolution of the insects. *Syst Biol* **55**:692–693.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-H, 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.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Cham: Springer International Publishing AG Switzerland.

Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biol Evol* **8**:1785–1801.

Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, Yang X, Yang F, Wu Q, Xie W, et al. 2021. Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell* **184**:3588.

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, 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.

Yan B, Yu X, Dai R, Li Z, Yang M. 2021. Chromosome-level genome assembly of *Nephotettix cincticeps* (Uhler, 1896) (Hemiptera: Cicadellidae: Deltocephalinae). *Genome Biol Evol* **13**:evab236.

Yang S, Sun N, Stogin BB, Wang J, Huang Y, Wong T-S. 2017. Ultra-antireflective synthetic brochosomes. *Nat Commun* **8**:1285.

Ye Y-X, Zhang H-H, Li D-T, Zhuo J-C, Shen Y, Hu Q-L, Zhang C-X. 2021. Chromosome-level assembly of the brown planthopper genome with a characterized Y chromosome. *Mol Ecol Resour* **21**:1287–1298.

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res* **45**:D744–D749.