Bulletin of Entomological Research

cambridge.org/ber

Comparative genomics reveals evolutionary drivers of the dietary shift in Hemiptera

Guangyao Wu[1], Chunyan Wu[1], Youssef Dewer[2], Peiyao Li[1], Baojun Hao[3], Liansheng Zang[1] and Fengqi Li[1]

[1]National Key Laboratory of Green Pesticide, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Center for R&D of Fine Chemicals of Guizhou University, Guiyang 550025, China; [2]Phytotoxicity Research Department, Central Agricultural Pesticide Laboratory, Agricultural Research Center, Dokki 12618, Giza, Egypt and [3]School of Life and Health Science, Kaili University, Guizhou 556000, China

Research Paper

Cite this article: Wu G, Wu C, Dewer Y, Li P, Hao B, Zang L, Li F (2024). Comparative genomics reveals evolutionary drivers of the dietary shift in Hemiptera. *Bulletin of Entomological Research* **114**, 41–48. https://doi.org/10.1017/S0007485323000597

Received: 29 August 2023
Revised: 6 November 2023
Accepted: 9 November 2023
First published online: 15 December 2023

Keywords:
adaptive evolution; comparative genomics; Hemipteran; *Oncopeltus fasciatus*; *Riptortus pedestris*

Corresponding author:
Fengqi Li; Email: fqli@gzu.edu.cn

Abstract

Hemiptera insects exhibit a close relationship to plants and demonstrate a diverse range of dietary preferences, encompassing phytophagy as the predominant feeding habit while a minority engages in carnivorous or haematophagous behaviour. To counteract the challenges posed by phytophagous insects, plants have developed an array of toxic compounds, causing significant evolutionary selection pressure on these insects. In this study, we employed a comparative genomics approach to analyse the expansion and contraction of gene families specific to phytophagous insect lineages, along with their adaptive evolutionary traits, utilising representative species from the Hemiptera order. Our investigation revealed substantial expansions of gene families within the phytophagous lineages, especially in the Pentatomomorpha branch represented by *Oncopeltus fasciatus* and *Riptortus pedestris*. Notably, these expansions of gene families encoding enzymes are potentially involved in hemipteran-plant interactions. Moreover, the adaptive evolutionary analysis of these lineages revealed a higher prevalence of adaptively evolved genes in the Pentatomomorpha branch. The observed branch-specific gene expansions and adaptive evolution likely contribute significantly to the diversification of species within Hemiptera. These results help enhance our understanding of the genomic characteristics of the evolution of different feeding habits in hemipteran insects.

Introduction

Hemiptera encompasses several crucial agricultural pests such as aphids and whiteflies, as well as significant sanitary pests including *Triatoma rubrofasciata*, *Rhodnius prolixus*, and *Cimex lectularius* (Rosenfeld *et al.*, 2016; Liu *et al.*, 2019; Huang *et al.*, 2021; Rijal *et al.*, 2021). Therefore, the study of Hemiptera holds great significance. There are 97,000 to 103,590 known species representing over 10% of all known insect species (Johnson *et al.*, 2018). The distinguishing characteristic that differentiates Hemiptera from other insects is the variation in mouthparts, which has evolved from chewing mouthparts in their ancestors to piercing-sucking mouthparts to enhance adaptability (Li *et al.*, 2017). These specialised piercing-sucking mouthparts have undergone continuous modification and adaptation, enabling different Hemiptera species to feed on plant phloem, blood, or other fluids (Li *et al.*, 2017).

Hemipterans exhibit a remarkable biodiversity, spanning a wide range of life histories and habitats, encompassing both terrestrial and aquatic ecosystems (Li *et al.*, 2017). The diversity of environments drives their exploration of various food sources such as plants, arthropods, fungi, and vertebrate blood (Rider, 1996). Based on previous phylogenetic study of Hemipteroid, there is a temporal correlation between the evolution of true bugs in Hemipteroids and a shift in feeding habits, implying a possible transition from phytophagy to predation (Sweet, 1979; Johnson *et al.*, 2018). The successful evolution of dietary shifts in Hemiptera insects can be attributed to the interaction between trophic niche and genome-level changes (Weirauch *et al.*, 2019). The environment and prevailing conditions are primary factors that influence genetic variation within insect populations, which leads to the divergence of different species. Among various environmental factors, the trophic niche, which denotes the principal source of nourishment, plays a pivotal role in moulding phenotypic and genomic alterations in insects (Johnson *et al.*, 2018).

Studies investigating dietary variations in Hemiptera revealed that within this insect order, *Acyrthosiphon pisum*, whitefly, *Laodelphax striatellus*, *Oncopeltus fasciatus*, and *Riptortus pedestris* are classified as phytophagous insects (Li *et al.*, 2017). On the other hand, *T. rubrofasciata*, *R. prolixus*, *C. lectularius*, and *Gerris buenoi* are carnivorous insects (Rosenfeld *et al.*, 2016; Armisén *et al.*, 2018; Liu *et al.*, 2019; Huang *et al.*, 2021; Rijal *et al.*, 2021). Except *G. buenoi*, the habitats of these insects are primarily semi-aquatic (Armisén *et al.*, 2018).

Plants employ diverse defence strategies against phytophagous insects, they have developed various mechanisms of insect resistance specifically targeting Hemiptera to protect themselves (Santamaria *et al.*, 2013). Consequently, this dynamic interplay exerts selective pressure on phytophagous Hemiptera species.

Detoxification of plant secondary metabolites is a crucial adaptation strategy in the evolutionary dynamics of phytophagous insects, enabling them to neutralise or mitigate the effects of these compounds (Seppey *et al.*, 2019). Several gene families play pivotal roles in this process, including cytochrome P450 monooxygenases, carboxylesterases (CEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs), and ATP-binding cassette (ABC) transporters (Voelckel 2014). These genes are instrumental in the structural adjustment, modification, and transport of toxic compounds. In addition to producing compounds that are resistant to insects, plants also produce protein inhibitors targeting specific mechanisms of insect resistance. Notable gene families associated with insect resistance mechanisms include endopeptidases, such as cysteine (CYSs) and serine (SERs) proteases, along with more specialised enzymes like glycoside hydrolases (GHs) (Pauchet *et al.*, 2010; McKenna *et al.*, 2016). Furthermore, chemosensory systems, including the olfactory and gustatory sensors, play a pivotal role in the adaptation of phytophagous insects to their host plants (Goldman-Huertas *et al.*, 2015).

During lineage divergence, insects undergo genomic changes that contribute to adaptive development. Identifying these genes and their associations with phenotypic variations in different insect species is crucial for understanding the process of speciation (Hurst, 2009). Genomic changes such as gene duplication events lead to the expansion of gene families, resulting in the emergence of new gene copies with similar or identical functions (Kondrashov, 2012). Additionally, other genomic changes, such as point mutations in genes, alter their functions. Branch-specific gene expansion refers to the expansion of gene families in specific lineages and encompasses various mechanisms, including but not limited to adaptive evolution (Innan and Kondrashov, 2010). Despite these mechanisms are not exclusively related to adaptation, the production of different gene copies through duplication provides opportunities for natural selection to operate.

Materials and methods

Data sources and quality assessment

This study encompassed the genomes of 13 Hemiptera species carefully selected to represent five suborders of Hemiptera, ensuring a well-balanced sampling across the suborders of Cimicomorpha, Pentatomomorpha, Gerromorpha, Fulgoromorpha, and Sternorrhyncha. Sequence data of *R. prolixus*, *C. lectularius*, and *A. pisum* (CDS and protein sequences) was obtained from Ensembl Metazoa. Additionally, sequence data of *O. fasciatus* and *R. pedestris* (CDS and pep sequences) were acquired from InsectBase. Sequence of *G. buenoi* was obtained from the I5 K-pilot project, while *T. rubrofasciata*, *L. striatellus*, *Sogatella furcifera*, and *Trialeurodes vaporariorum* were obtained from the I5 K-pilot and GigaDB databases. Sequence data for the whitefly Mediterranean cryptic species (MED) and whitefly Middle East-Asia Minor 1 cryptic species (MEAM1) was obtained from the Whitefly Genome Database (http://www.whiteflygenomics.org/) and sequence data for *Diaphorina citric* was obtained from NCBI.

For phylogenetic and divergence time analyses, we introduced the fruit fly *Drosophila melanogaster* as an outgroup. The genomic data of *D.melanogaster* were obtained from Ensembl Metazoa. Detailed information about the genomic data sources pertinent to this study can be found in the list provided in supplementary Table S1. The quality assessment of all employed genomic datasets was conducted using BUSCO (version 5.2.2) (Waterhouse *et al.*, 2018) with the arthropoda_odb10.2020-09-10 dataset and the model tran.

Phylogenetic analysis of species

We employed the genomes of the 13 Hemiptera species and one outgroup mentioned earlier, resulting in a total of 14 species, to identify direct homologues. To limit redundancy arising from selective splice variants, we retained gene models encoding the longest protein sequence for each locus. Direct homologous gene analysis was conducted using DIAMOND in conjunction with OrthoFinder (version 2.5.4) (Emms and Kelly, 2019). Single-copy genes were identified based on the OrthoFinder results and subsequently utilised for downstream analyses, including the construction of phylogenetic trees, estimation of divergence times, and investigation of gene family contraction and expansion.

For the phylogenetic tree construction, a total of 22,723 protein sequences from the single-copy gene families were utilised. MUSCLE (version 3.8.95) (Edgar, 2004) was used to generate multiple sequence alignments with default parameters for the protein sequences within each single-copy family. The resulting alignments were combined into a super alignment matrix, which was then employed for phylogenetic tree reconstruction using the PROTGAMMAJTT model in the RAxML software (Stamatakis, 2014).

Functional annotation

The protein Pfam family was identified by InterProScan (Jones *et al.*, 2014). The uniref50 database was annotated using Blastp with an *e*-value threshold set at 1e-20. GO term annotation was performed using eggnog (Suzek *et al.*, 2015). Orthologous groups (OGs) were preliminarily selected based on the annotation in the uniref50 database. Subsequently, candidate OGs were identified by matching the uniref50 annotation and the Pfam and GO term annotation results.

Analysis of contraction and expansion of gene families

Based on the orthogroups (gene families) information derived from the identified orthologous genes using OrthoFinder (Emms and Kelly, 2019), the expansion and contraction of orthologous gene families were assessed using CAFE v4.2.1 (De Bie *et al.*, 2006). This software utilises birth and death processes to model the gene gain and loss events throughout the phylogenetic history process.

Adaptive evolutionary analysis

Furthermore, we conducted adaptive evolutionary selection analyses on the 13 Hemiptera species selected in this study. These analyses aimed to examine the differential occurrence of positive Darwinian selection on specific branches of the phylogeny for individual homologous genes. For this purpose, single-copy

https://doi.org/10.1017/S0007485323000597 Published online by Cambridge University Press

orthologous genes generated by OrthoFinder were employed for selective pressure analysis. Multiple sequence comparisons were performed using MACSE (Ranwez *et al.*, 2018), and subsequent analyses utilised the branching site model of the codeml program within the PAMLpackage (https://github.com/Hua-CM/BatchPAML) (Yang, 2007).

To identify positively selected genes, we employed Bonferroni correction and Benjamini-Hochberg false discovery rate (FDR) control to select genes with FDR-corrected *p*-values below 0.05. We performed comparative sequence recombination checks using the Pairwise Homoplasy Index (PHI) (Bruen *et al.*, 2006), Neighbour Similarity Score (NSS), and Maximum Chi-Square tests implemented in the PhiPack program to exclude false-positive selection events resulting from sequence recombination. Recombination events were considered present when the *p*-value (*q*-value) for PHI was less than 0.05 and when at least one other test supported the occurrence of recombination. In this study, we systematically examined the branches of Pentatomomorpha, Fulgoromorpha, Sternorrhyncha, and Cimicomorpha for adaptive evolution using the aforementioned process. Enrichment analysis was conducted to identify genes associated with adaptive evolution, following the procedure described earlier. We utilised the web-based tool agriGO (systemsbiology.cau.edu.cn/agriGOv2) (Tian *et al.*, 2017) and the FlyBase database to assess the enrichment of genes in specific Gene Ontology (GO) terms (Thurmond *et al.*, 2019). Furthermore, the KOBAS (Xie *et al.*, 2011) and BlastKOALA software (Kanehisa *et al.*, 2016) were employed to statistically evaluate the enrichment of genes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto, 2000).

Results

Representative sampling of Hemiptera insects

To investigate the adaptation of Hemiptera to host plants, we conducted a sampling of 13 Hemiptera species, representing various suborders. Specifically, our sampling included five species from the Sternorrhyncha suborder, three species from the Cimicomorpha suborder, two species from the Pentatomomorpha suborder, two species from the Fulgoromorpha suborder, and one species from the Gerromorpha suborder. These species inhabited different habitats, with *G. buenoi* being semi-aquatic and the remaining species terrestrial (Table S1). The feeding habits of these species varied and consisted of three categories: phytophagous, predaceous, and haematophagous. Notably, within the Cimicomorpha suborder alone, all three feeding habits were represented, among which *T. rubrofasciata* and *R. prolixus* were predaceous, *C. lectularius* being haematophagous. *O. fasciatus* and *R. pedestris* were phytophagous (Table S1). Therefore, our sample encompassed different Hemiptera suborders, diverse habitats, and a range of feeding habits, including a substantial number of phytophagous species (Table S1). We assessed the integrity of the genome sequences obtained from the sampled species using BUSCO analysis, which revealed sequence integrity ranging from 74.8% to 99.9% across the sampled Hemiptera species and outgroups. With these sequences, we identified 22,723 directly homologous genes and constructed a time-calibrated species phylogeny using 221 single-copy gene sequences. The sequences from these orthologous groups (OGs) were further subjected to functional annotation, particularly focusing on gene families involved in phytophagous insect-plant interactions. Among these candidate OGs, we identified 72 candidates from eight gene families, including 11 UDP-glycosyl transferases (UGTs), 61 cytochrome P450s (P450s), 13 carboxylesterases (CEs), 1 glutathione S-transferase (GST), 4 serine proteases (SERs), 2 cysteineproteases (CYSs), 21 ATP-binding cassette transporters (ABCs), and 3 glycoside hydrolases (GHs) (Table 1).

Phytophagous insects show more frequent gene expansion

According to the CAFE analysis, the predicted λ (gain) value for gene expansions was 0.0020 gain/gene/million years, and the μ (loss) value for gene contractions was 0.0036 loss/gen e/million years (De Bie *et al.*, 2006). Out of the total 22,723 orthologous groups (OGs) examined, the root of the phytophagous insect suborder exhibited a higher frequency of genes. In the root of the Sternorrhyncha suborder, 21 genes underwent expansion, while 5 genes were contracted. Similarly, within the Fulgoromorpha suborder, 64 genes experienced expansion, while 18 genes underwent contraction. In the Cimicomorpha suborder, which includes two predaceous and one haematophagous insect species, all three species are non-phytophagous, and the root exhibited expansion of 3 genes and contraction of 39 genes. Conversely, in the Pentatomomorpha suborder, which comprises two phytophagous insect species, the root showed an expansion of 82 genes and a contraction of 9 genes. The number of gene contractions and expansions in each branch is depicted in fig. 1. Significance analysis using a *p*-value threshold of <0.01 revealed that 11.11% of the 2525 OGs identified by CAFE exhibited significant contractions or expansions.

Adaptive evolution is more widespread in the phytophagous branch Pentatomomorpha

We specifically examined the phytophagous insect branches of Pentatomomorpha, Fulgoromorpha, and Sternorrhyncha for adaptive evolution, as well as the non-phytophagous insect branch of Cimicomorpha. The results respectively a total of 18, 3, 1, and 3 genes undergoing adaptive evolution in the branches of Pentatomomorpha, Fulgoromorpha, Sternorrhyncha, and Cimicomorpha (FDR < 0.05) (Table 2). Among these genes, one member of the ABC gene family, ACYPI010103, exhibited adaptive evolution in the Pentatomomorpha branch with *FDR* value 7.6046E-08 (Table 2). The GO functional enrichment analysis of these adaptively evolved genes revealed that these genes were enriched in 28 different GO terms such as skeletal muscle myosin thick filament assembly, myosin filament assembly, and striated muscle myosin thick filament assembly (Table S2). In terms of KEGG functional enrichment, the adaptively evolved genes were found to be enriched in GTP-binding proteins (Table S3).

In contrast, no significantly enriched GO or KEGG terms were observed for the adaptively evolved genes in the Fulgoromorpha and Sternorrhyncha branches ($P < 0.05$). For the non-phytophagous Cimicomorpha branch, the three adaptively evolved genes were found to be enriched in a single GO term: brown fat cell differentiation (Table S2). However, no significantly enriched pathways were detected when analysing KEGG enrichment (Table S3).

Discussion

In this study, we employed comparative genomics to investigate the genomic characteristics, including gene family contraction, expansion and adaptive evolution, across different feeding

https://doi.org/10.1017/S0007485323000597 Published online by Cambridge University Press

Table 1. Candidate gene categories and key terms/identifiers are used to select from the complete annotation sequences using InterProScan

Gene family categor	InterProScan (Pfam or InterPro identifiers) or Gene Ontology	UnifRef keyword	Number of OGs
UDP-glycosyltransferases (UGTs)	PF00201	name: UDP-glucuronosyltransferase'	11
Cytochrome P450 oxidases (P450s)	PF00067	name: 'Cytochrome P450'	61
Carboxylesterases (CEs)	PF02230, PF00135	name: 'carboxylesterase' OR name: 'carboxylic ester hydrolase'	14
Glutathione S-transferases (GSTs)	PF00043, PF02798	name: 'Glutathione S-transferase'	1
Serine proteases (SERs)	PF00450, PF12146, PF05577,GO:0008236	name: 'Serine protease'	4
Cysteine proteases (CYSs)	PF00112	name: 'cysteine protease' OR name:'Papain'	2
ABC transporters (ABCs)	PF00005, PF00664	name: 'ABC'	21
Glycoside hydrolases (GHs)	PF00232, PF00704, PF03659, PF07971	name: 'Glycoside hydrolase'	3
Total			72

InterProScan Scan-derived categories for Inclusion in OGs are considered for inclusion in candidate (OGs), OGs are required to have at least one sequence matching both UniRef and InterProScan entries. Additionally, supplementary Gene Ontology terms are taken into account.

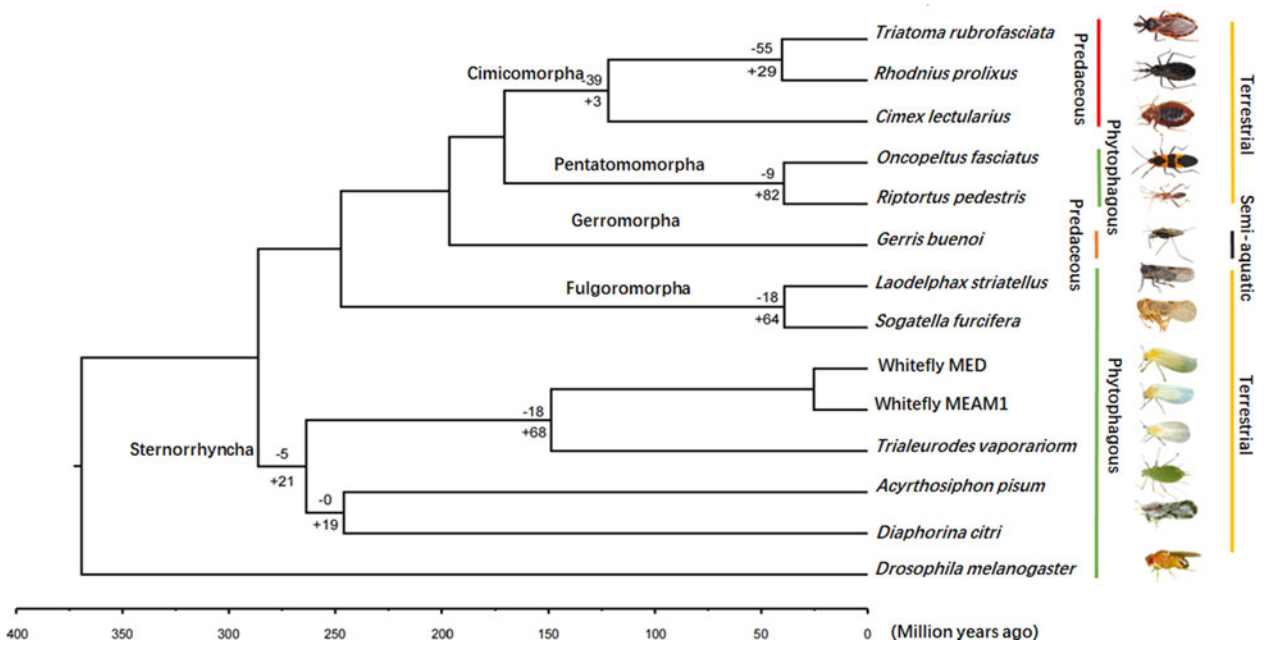

Figure 1. The research quantitatively analyses phylogenetics and gene family expansions using genomic datasets. The species tree is based on 221 single-copy orthologous genes, with strong bootstrap support (99% or 100%). Data are from sequenced genomes. (OGs) expand (+) and contract (-) at each subsequence root. The study included herbivorous, semi-aquatic, and blood-feeding insects.

branches of the Hemiptera suborder. We analysed whitefly MED cryptic species, whitefly MEAM cryptic species, and *T. vaporariorum* from the polyphagous Sternorrhyncha suborder, as well as *A. pisum* from the monophagous suborder. In addition, *O. fasciatus* and *R. pedestris*, belonging to the Pentatomomorpha suborder, were also monophagous. *R. pedestris,* commonly known as the bean bug, primarily feeds on leguminous plants, particularly *G. max* which is commonly referred to as soybean (Huang *et al.*, 2021). On the other hand, *O. fasciatus,* also known as the large milkweed bug, is a specialised seed feeder and has been observed to in 13 Hemiptera species (Panfilio *et al.*, 2019). Among these, three are predaceous: *T. rubrofasciata* and *R. prolixus* from the Cimicomorpha suborder, and *G. buenoi* from the Gerromorpha suborder (Armisén *et al.*, 2018; Liu *et al.*, 2019; Huang *et al.*, 2021). There is also a haematophagous species, *C. lectularius*, which belongs to the Cimicomorpha suborder. The remaining nine species were phytophagous, consisting of *L. striatellus* and *S. furcifera* from the monophagous Fulgoromorpha suborder, which generally undergo nymphal development on a limited number of host species within the Asclepias genus. To test the hypothesis that gene expansion and adaptive evolution contribute to dietary changes in Hemiptera, we analysed gene family expansions and adaptive evolutionary events in the 13 Hemiptera species. It is widely recognised that phytophagous

https://doi.org/10.1017/S0007485323000597 Published online by Cambridge University Press

Table 2. Result of evolutionary selection pressure analysis on Pentatomomorpha, Fulgoromorpha, Sternorrhyncha, and Cimicomorpha branches

Branch	Number	OGs	Representative gene in *Acyrthosiphon pisum*	FDR	Gene description	pfam or Gene ontology
Cimicomorpha	3	OG0000408	ACYPI010047	0.007281	Histone deacetylase 3	
		OG0005435	ACYPI003705	0.007281	ADP-ribosylation factor-like 4A	
		OG0002769	ACYPI002812	0.013461	Kinesin heavy chain	
Pentatomomorpha	18	OG0003750	ACYPI010103	7.6046E-08	ATP-binding cassette sub-family F member 3	PF00005
		OG0004057	ACYPI002371	7.6046E-08	Ubiquitin-conjugating enzyme E2 W-like	
		OG0004544	ACYPI009001	1.9975E-07	26S proteasome regulatory subunit 4	
		OG0006564	ACYPI001001	2.17375E-05	Paramyosin, long form	GO:0008236
		OG0001069	ACYPI000969	0.000125772	Tropomyosin-1	
		OG0005013	ACYPI000662	0.000220743	Zinc finger protein 706	
		OG0000323	ACYPI007135	0.002435271	Serine/threonine-protein phosphatase 4 catalytic subunit	
		OG0004111	ACYPI56655	0.006950125	Acetyl-CoA acetyltransferase, mitochondrial	
		OG0000039	ACYPI000267	0.009426111	Alpha-tocopherol transfer protein-like	
		OG0003650	ACYPI009782	0.0118158	Polymerase (RNA) II (DNA directed) polypeptide H	
		OG0002897	ACYPI006711	0.012963455	Elongation factor 1-alpha	
		OG0003325	ACYPI002823	0.01975175	GTP-binding protein SAR1b	
		OG0003607	ACYPI002584	0.023669923	V-type proton ATPase catalytic subunit A	
		OG0005435	ACYPI003705	0.032497143	ADP-ribosylation factor-like 4A	
		OG0003432	ACYPI000895	0.032555333	GTP-binding protein Di-Ras2	
		OG0000720	ACYPI007027	0.03542625	Fructose-bisphosphate aldolase	
		OG0004573	ACYPI003710	0.043599412	Calcineurin B-like	GO:0008236
		OG0005795	ACYPI007002	0.049585	Dodo-like	
Fulgoromorpha	3	OG0005206	ACYPI005806	0.0280575	Enolase	
		OG0005272	ACYPI007179	0.0280575	Ribosomal protein L27a	
		OG0002355	ACYPI067654	0.03838	SNW domain-containing protein 1	
Sternorrhyncha	1	OG0004177	ACYPI002837	0.01041	pre-mRNA-processing factor 6	
Gerromorpha	8	OG0001916	ACYPI003057	0	Actin-interacting protein 1	
		OG0004259	ACYPI006557	6.17375E-10	Exportin-1	
		OG0005417	ACYPI52009	2.1875E-07	Obg-like ATPase 1	
		OG0003650	ACYPI009782	3.99875E-07	Polymerase (RNA) II (DNA directed) polypeptide H	
		OG0001603	ACYPI003934	0.000005315	Phosphatidylinositol transfer protein alpha isoform-like	
		OG0004522	ACYPI006441	0.00000805	Proteasome subunit alpha type 6-like	
		OG0002949	ACYPI009856	0.00001	40S ribosomal protein S16-like	
		OG0002538	ACYPI010127	0.002031875	Ribosomal protein S3A	

insects require specific genes for the detoxification of plant secondary metabolites and digestion of plant tissues on their host plants, to achieve the purpose of adapting to different environments comparing to predaceous and haematophagous insect (Heidel-Fischer and Vogel, 2015; Simon *et al.*, 2015). Both gene family expansion and adaptive evolution are important molecular mechanisms underlying organismal adaptation to varying environments (Innan and Kondrashov, 2010; Kondrashov, 2012).

https://doi.org/10.1017/S0007485323000597 Published online by Cambridge University Press

Previous studies on beetles have demonstrated the significance of gene family expansion and adaptive evolution in the dietary shift to phytophagy (Seppey *et al.*, 2019). In our study, we observed gene family expansion in the phytophagous insect branch. However, regarding adaptive evolution, a consistent association between adaptive evolution and gene family expansion was only detected in the Pentatomomorpha suborder. Therefore, adaptive evolution may only play a minor role in the intrinsic mechanism of gene family expansion in Hemiptera. Our analysis identified the phytophagous trophic niche as a driving force behind gene family expansion in Hemiptera. However, the expansion of these genes did not lead to functional divergence (Innan and Kondrashov, 2010; Kondrashov, 2012).

There is a shared anatomical feature of a specialised piercing and sucking mouthpart within the Hemiptera insects. However, these insects have undergone diversification to exploit a wide range of food sources, including seeds and plant tissues (phytophagy), and even vertebrate blood (haematophagy) (Li *et al.*, 2017). As a result of this diversification, numerous hemipterans have become significant agricultural pests or vectors of human diseases. Consequently, extensive efforts have been devoted to genome sequencing in these species, aiming to unravel their genomic composition and elucidate their complex biological characteristics (Li *et al.*, 2017; Johnson *et al.*, 2018).

When conducting comparative analyses of datasets from different species, it is crucial to ensure that all analysed species have comparable gene contents. To mitigate the risk of false-positive gene identification, we implemented an additional filtering step based on the method proposed by Seppey *et al.* (2019). This process combined InterProScan results using Pfam or InterPro identifiers, along with Gene Ontology information. Additionally, we integrated the relevant keywords from the carefully selected UniRef dataset for further filtration (Table 1). Although this stringent filtering approach might have potentially excluded some candidate orthogroups (OGs) from our analysis, it was crucial in ensuring the accuracy and validity of the OGs used in this study.

In addition to the gene expansion in Pentatomomorpha, the positive selection analyses also identified a higher number of genes in the branch that underwent positive selection for adaptive evolution. A noteworthy gene in this context is the ABC transporter, which is known to be involved in the detoxification of plant secondary metabolites (Yazaki, 2006). This finding suggests that selective pressure has played a role in the expansione expansion of the gene family associated with detoxification enzymes (Yazaki, 2006; Calla, 2021).

Emerging research indicates that ABC transporters are also involved in insect detoxification of pesticides, Bt and plant secondary metabolites (Yazaki, 2006; Heckel, 2012; Merzendorfer, 2014). In addition to the ABC transporter, several other adaptively evolved genes in the Pentatomomorpha branch are associated with muscle development, including skeletal muscle myosin thick filament assembly, myosin filament assembly, and muscle cell development, and others (Table 3). The adaptive evolution of these genes provides a mechanistic foundation for the enhanced locomotor capacity of Pentatomomorpha. Several mechanisms contribute to gene family expansion, with adaptive evolution being one of the driving factors, but it may not be the predominant factor (Innan and Kondrashov, 2010; Seppey *et al.*, 2019). Although a significant number of gene family expansions and adaptive evolutions were observed in the Pentatomomorpha branch, we did not find the same level of

Table 3. Candidate OGs with CAFE *p*-value <0.01 were identified

	Pentatomomorpha		Fulgoromorpha		Cimicomorpha		Sternorrhyncha	
Gene family category	Expansion	Contract	Expansion	Contract	Expansion	Contract	Expansion	Contract
UDP-glycosyltransferases (UGTs)	0	0	0	0	0	1	0	0
Cytochrome P450 oxidases (P450s)	1	0	0	1	0	0	0	0
Carboxylesterases (CEs)	1	0	0	0	0	1	0	0
Glutathione S-transferases (GSTs)	0	0	0	0	0	0	0	0
Serine proteases (SERs)	0	0	0	0	0	0	0	0
Cysteine proteases (CYSs)	2	0	0	1	0	0	0	0
ABC transporters (ABCs)	0	0	0	0	0	0	0	0
Glycoside hydrolases (GHs)	1	0	0	0	0	0	0	0

OGs were categorised as Cytochrome P450 (P450), Carboxylesterase (CE), Glutathione S-Transferase (GST), and Cysteine Protease (CYS) based on OrthoDB v8 functional annotations.

https://doi.org/10.1017/S0007485323000597 Published online by Cambridge University Press

cooperativity of gene family expansions and adaptive evolutions in Hemiptera as observed in Coleoptera (Seppey *et al.*, 2019). Therefore, for gene family expansion, neutral or purifying selection forces may play a more prominent role.

Neutral or purifying selection is likely the primary mode of occurrence for gene family expansions in Hemiptera. It can be thought that genes associated with nutrition and detoxification might result from gene duplication rather than the emergence of novel functions (Panfilio *et al.*, 2019; Volonté *et al.*, 2022).

Conclusions

In this study, we analysed branch-specific gene family expansions and adaptive evolutions by comparing genomic data from representative species of Hemiptera. Our study revealed a higher occurrence rate of gene family expansions in the phytophagous branch, particularly within the Pentatomomorpha branch. Furthermore, through further analysis of adaptive evolutions within each branch, we identified a greater abundance of genes that underwent adaptive evolutions in the Pentatomomorpha branch.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0007485323000597.

Acknowledgements. This research was funded by National Key Research and Development Plan (grant number 2023YFE0104800); the Program of Talent Introduction in Guizhou University(2023-28), and Lure Research and Application Technology of Agricultural Sucking Pests in Ningxia of China, Grant/Award Number: NGSB-2021-10-02. Thanks for the computing support of the State Key Laboratory of Public Big Data, Guizhou University.

Competing interests. None.

Reference

Armisén D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn S-J, Poelchau MF, Chao H, Dinh H, Doddapaneni HV, Dugan S, Gibbs RA, Hughes DST, Han Y, Lee SL, Murali SC, Muzny DM, Qu J, Worley KC, Munoz-Torres M, Abouheif E, Bonneton F, Chen T, Chiang L-M, Childers CP, Cridge AG, Crumière AJJ, Decaras A, Didion EM, Duncan EJ, Elpidina EN, Favé M-J, Finet C, Jacobs CGC, Cheatle Jarvela AM, Jennings EC, Jones JW, Lesoway MP, Lovegrove MR, Martynov A, Oppert B, Lillico-Ouachour A, Rajakumar A, Refki PN, Rosendale AJ, Santos ME, Toubiana W, van der Zee M, Vargas Jentzsch IM, Lowman AV, Viala S, Richards S and Khila A (2018) The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC Genomics* **19, 832**. https://doi.org/10.1186/s12864-018-5163-2.

Bruen TC, Philippe H and Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681.

Calla B (2021) Signatures of selection and evolutionary relevance of cytochrome P450s in plant-insect interactions. *Current Opinion in Insect Science* **43**, 92–96.

De Bie T, Cristianini N, Demuth JP and Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics (Oxford, England)* **22**, 1269–1271.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Emms DM and Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238.

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG and Whiteman NK (2015) Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3026–3031.

Heckel DG (2012) Learning the ABCs of Bt: ABC transporters and insect resistance to *Bacillus* thuringiensis provide clues to a crucial step in toxin mode of action. *Pesticide Biochemistry and Physiology* **104**, 103–110.

Heidel-Fischer HM and Vogel H (2015) Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in insect Science* **8**, 8–14.

Huang HJ, Ye YX, Ye ZX, Yan XT, Wang X, Wei ZY, Chen JP, Li JM, Sun ZT and Zhang CX (2021) Chromosome-level genome assembly of the bean bug *Riptortus pedestris*. *Molecular Ecology Resources* **21**, 2423–2436.

Hurst LD (2009) Fundamental concepts in genetics: genetics and the understanding of selection. *Nature Reviews Genetics* **10**, 83–93.

Innan H and Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* **11**, 97–108.

Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KKO, Kozlov AM, Podsiadlowski L, Mayer C, Meusemann K, Vasilikopoulos A, Waterhouse RM, Cameron SL, Weirauch C, Swanson DR, Percy DM, Hardy NB, Terry I, Liu S, Zhou X, Misof B, Robertson HM and Yoshizawa K (2018) Phylogenomics and the evolution of hemipteroid insects. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 12775–12780.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R and Hunter S (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* **30**, 1236–1240.

Kanehisa M and Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27–30.

Kanehisa M, Sato Y and Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology* **428**, 726–731.

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings. Biological Sciences* **279**, 5048–5057.

Li H, Leavengood JM Jr., Chapman EG, 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. *Proceedings. Biological Sciences* **284, 20171223**. https://doi.org/10.1098/rspb.2017.1223.

Liu Q, Guo Y, Zhang Y, Hu W, Li Y, Zhu D, Zhou Z, Wu J, Chen N and Zhou XN (2019) A chromosomal-level genome assembly for the insect vector for Chagas disease, *Triatoma rubrofasciata*. *Gigascience* **8**, giz089. https://doi.org/10.1093/gigascience/giz089.

McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn SJ, Arsala D, Benoit JB, Blackmon H, Bledsoe T, Bowsher JH, Busch A, Calla B, Chao H, Childers AK, Childers C, Clarke DJ, Cohen L, Demuth JP, Dinh H, Doddapaneni H, Dolan A, Duan JJ, Dugan S, Friedrich M, Glastad KM, Goodisman MA, Haddad S, Han Y, Hughes DS, Ioannidis P, Johnston JS, Jones JW, Kuhn LA, Lance DR, Lee CY, Lee SL, Lin H, Lynch JA, Moczek AP, Murali SC, Muzny DM, Nelson DR, Palli SR, Panfilio KA, Pers D, Poelchau MF, Quan H, Qu J, Ray AM, Rinehart JP, Robertson HM, Roehrdanz R, Rosendale AJ, Shin S, Silva C, Torson AS, Jentzsch IM, Werren JH, Worley KC, Yocum G, Zdobnov EM, Gibbs RA and Richards S (2016) Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology* **17**, 227.

Merzendorfer H (2014) Chapter One - ABC transporters and their role in protecting insects from pesticides and their metabolites. In Cohen E (ed.), *Advances in Insect Physiology*. New York: Academic Press, pp. 1–72.

Panfilio KA, Vargas Jentzsch IM, Benoit JB, Erezyilmaz D, Suzuki Y, Colella S, Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, Weirauch MT, Hughes DST, Murali SC, Werren JH, Jacobs CGC, Duncan EJ, Armisén D, Vreede BMI, Baa-Puyoulet P, Berger CS, Chang CC, Chao H, Chen MM, Chen YT, Childers CP, Chipman AD, Cridge AG, Crumière AJJ, Dearden PK, Didion EM, Dinh H,

Doddapaneni HV, Dolan A, Dugan S, Extavour CG, Febvay G, Friedrich M, Ginzburg N, Han Y, Heger P, Holmes CJ, Horn T, Hsiao YM, Jennings EC, Johnston JS, Jones TE, Jones JW, Khila A, Koelzer S, Kovacova V, Leask M, Lee SL, Lee CY, Lovegrove MR, Lu HL, Lu Y, Moore PJ, Munoz-Torres MC, Muzny DM, Palli SR, Parisot N, Pick L, Porter ML, Qu J, Refki PN, Richter R, Rivera-Pomar R, Rosendale AJ, Roth S, Sachs L, Santos ME, Seibert J, Sghaier E, Shukla JN, Stancliffe RJ, Tidswell O, Traverso L, van der Zee M, Viala S, Worley KC, Zdobnov EM, Gibbs RA and Richards S (2019) Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biology* **20**, 64.

Pauchet Y, Wilkinson P, Chauhan R and Ffrench-Constant RH (2010) Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PLoS One* **5**, e15635.

Ranwez V, Douzery EJP, Cambon C, Chantret N and Delsuc F (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution* **35**, 2582–2584.

Rider DA (1996) The true bugs of the world (Hemiptera: Heteroptera). classification and natural history. *Annals of the Entomological Society of America* **89**, 609–610.

Rijal JP, Joyce AL and Gyawaly S (2021) Biology, ecology, and management of Hemipteran pests in almond orchards in the United States. *Journal of Integrated Pest Management* **12, 1-14**. https://doi.org/10.1093/jipm/pmab018.

Rosenfeld JA, Reeves D, Brugler MR, Narechania A, Simon S, Durrett R, Foox J, Shianna K, Schatz MC, Gandara J, Afshinnekoo E, Lam ET, Hastie AR, Chan S, Cao H, Saghbini M, Kentsis A, Planet PJ, Kholodovych V, Tessler M, Baker R, DeSalle R, Sorkin LN, Kolokotronis S-O, Siddall ME, Amato G and Mason CE (2016) Genome assembly and geospatial phylogenomics of the bed bug *Cimex lectularius*. *Nature Communications* **7, 10164**. https://doi.org/10.1038/ncomms10164.

Santamaria ME, Martínez M, Cambra I, Grbic V and Diaz I (2013) Understanding plant defence responses against herbivore attacks: an essential first step towards the development of sustainable resistance against pests. *Transgenic Research* **22**, 697–708.

Seppey M, Ioannidis P, Emerson BC, Pitteloud C, Robinson-Rechavi M, Roux J, Escalona HE, McKenna DD, Misof B, Shin S, Zhou X, Waterhouse RM and Alvarez N (2019) Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. *Genome Biology* **20**, 98.

Simon JC, d'Alençon E, Guy E, Jacquin-Joly E, Jaquiéry J, Nouhaud P, Peccoud J, Sugio A and Streiff R (2015) Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics* **14**, 413–423.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* **30**, 1312–1313.

Suzek BE, Wang Y, Huang H, McGarvey PB and Wu CH (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics (Oxford, England)* **31**, 926–932.

Sweet MH (1979) On the original feeding habits of the Hemiptera (Insecta). *Annals of the Entomological Society of America* **72**, 575–579.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC and Calvi BR (2019) FlyBase 2.0: the next generation. *Nucleic Acids Research* **47**, D759–D765.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W and Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* **45**, W122–W129.

Voelckel CJG (2014) Insect-plant interactions. *Annual Plant Reviews* **47**.

Volonté M, Traverso L, Estivalis JML, Almeida FC and Ons S (2022) Comparative analysis of detoxification-related gene superfamilies across five Hemipteran species. *BMC Genomics* **23**, 757.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV and Zdobnov EM (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**, 543–548.

Weirauch C, Schuh RT, Cassis G and Wheeler WC (2019) Revisiting habitat and lifestyle transitions in Heteroptera (Insecta: Hemiptera): insights from a combined morphological and molecular phylogeny. *Cladistics* **35**, 67–105.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY and Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* **39**, W316–W322.

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

Yazaki K (2006) ABC transporters involved in the transport of plant secondary metabolites. *FEBS Letters* **580**, 1183–1191.