Detoxification enzymes: comparative analysis across hemipteran species

Mariano Volonté
- Laboratorio de Neurobiología de Insectos (LNI), Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CENEXA, CONICET, La Plata, Buenos Aires, Argent

Lucila Traverso
- Laboratorio de Neurobiología de Insectos (LNI), Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CENEXA, CONICET, La Plata, Buenos Aires, Argent

Jose Manuel Latorre Estivalis
Laboratorio de Insectos Sociales, Instituto de Fisiología, Biología Molecular y Neurociencias, Universidad de Buenos Aires - CONICET, Ciudad Autónoma de Buenos Aires, Argentina.

Francisca Cunha Almeida
Grupo de Investigación en Filogeografía y Filogenias Moleculares, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Autónoma

Sheila Ons (sheila.ons@presi.unlp.edu.ar)
- Laboratorio de Neurobiología de Insectos (LNI), Centro Regional de Estudios Genómicos, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CENEXA, CONICET, La Plata, Buenos Aires, Argent

---

Research Article

Keywords: GLUTATHIONE TRANSFERASES, CYTOCHROMES P450, CARBOXYL CHOLINESTERASES, INSECTS, PENTATOMID, TRIATOMINE, HETEROPTERA

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1665017/v1

Abstract

Background:

Hemiptera is one of the most speciose Orders of insects, and the most speciose considering hemimetabolous. Through their evolutive history, hemipterans with different feeding habits have adapted to deal with different chemical challenges. Three major gene families are involved in xenobiotic detoxification in insects: the cytochromes P450 (CYPs), carboxyl/cholinesterases (CCEs), and glutathione transferases (GSTs). Here we perform a comparative analysis on the complement of these gene superfamilies across five hemipteran species; four heteropterans (the plant feeders pentatomids *Nezara viridula* and *Halyomorpha halys*; the hematophagous *Cimex lectularius*, *Cimicidae*, and *Rhodnius prolixus, Reduviidae*), and one plant feeder Auchenorryncha (*Nilaparvata lugens*).

Results:

Our results point to an expansion of several enzyme families associated to xenobiotic detoxification in heteropterans respect to other species, and present a dynamic evolution pattern including CYP3 clan, hormone and pheromone processing class in the CCE superfamily and Sigma class in GST superfamily. Other detoxification-related families are reduced in the Hemiptera species analyzed here: reduction or even absence of class Epsilon and reduced class Delta in GST superfamily; absence of mitochondrial family CYP12; absence of CYP9 family in CYP3 clan; and reduction or even absence of some dietary/detoxification groups of CCEs. Interestingly, the most polyphagous species analyzed here (*H. halys*), is also the one that presents the larger repertoire of detoxification enzymes. Gene-cluster analysis suggests that this could be due to gene duplication events.

Conclusions:

The evolutionary analysis performed here reveals characteristics that are both common and particular for heteropterans. The composition and organization of detox-related gene families could shed light on evolutionary forces that shaped their divergence. These families are both important for the detoxification of diet products and to confer tolerance or resistance to synthetic insecticides. Furthermore, we present here the first comprehensive analysis of detoxificant gene superfamilies in *N. viridula*, an understudied species in spite of its economic relevance as a crop pest. The information obtained is of interest for basic insect science, as much as for the control of harmful species and the management of insecticide resistance.

Background

Hemiptera is one of the most speciose orders of insects, and the most speciose considering hemimetabolous; it comprises more than 50,000 species. The suborder Heteroptera (true bugs) includes

species adapted to a diversity of ecological niches and lifestyles. Phylogenomic analysis and fossil records suggest that the origin of the suborder Heteroptera is coincident with a shift from herbivory to predation [1], whereas in Pentatomorpha and Cimicomorpha, two lineages within Heteroptera, a shift back to herbivory occurred. Within Cimicomorpha, hematophagy emerged independently in the subfamily Triatominae (kissing bugs) and the family Cimicidae (bed bugs). Dissimilar diets present different challenges for these phylogenetically related heteropterans. Through their evolutive history, hemipterans with different feeding habits have adapted their detoxificant physiology to deal with different chemicals. Whereas phytophagous insects are exposed to toxins, repellents or anti-digestive compounds from plants [2], hematophagy led to the ingestion of toxic amounts of heme, iron and amino acids in every meal [3].

Three major gene families are involved in xenobiotic detoxification in insects: the cytochromes P450 (CYPs), carboxyl/cholinesterases (CCEs), and glutathione transferases (GSTs) [4]. It has been largely demonstrated that the ability of herbivorous insects to detoxify plant allelochemicals affects their capability to utilize different plants as a potential host [5]. In fact, a positive correlation between the number of detoxifying genes in a particular species and the complexity of its food sources has been reported [5]. On the other hand, a reduced complement of detoxifying enzymes in species such as the honey bee *Apis mellifera* [6] or the human head lice *Pediculus humanus* [7] were related with their specialized diet. Likewise, generalist hymenopterans seem to present a higher number of GSTs, CCEs and CYPs when compared to specialist species of the same Order [8].

Even though several exceptions exist to the correlation between a broad diet and a large number of detoxification-related enzymes, it is generally accepted that expansions or contractions of these gene families throughout evolution could be related to functional adaptations to the environment [8]. The evolution of gene families is determined by a dynamic process of gene duplication and loss (the gene birth-and death model), which is mediated by mutation genetic drift, natural selection, and chromosomal rearrangements [9]. Paralogue gene proliferations in genomes (or “gene blooms”) are originated by gene duplications, causing significant expansions in particular subfamilies [10]. Newly duplicated genes can be fixed if it is adaptive or lost through pseudogenefication. Hence, the comparative analysis of insect detoxicant complement could indicate those groups of genes involved in the adaptation of the species to particular ecological niches.

In a previous work, we found that triatomines, who are obligated blood-feeders in their complete life cycle, present reductions (or even absence) of some families of GSTs, CCEs and CYPs, whereas others are expanded [4]. Up to our knowledge, a comprehensive analysis comparing xenobiotic-detoxification related genes in hemipterans with different feeding habits has not been reported to date; hence, some of the particularities found in triatomines could be common to Heteroptera or to Hemiptera. Even though a previous report compared the CYP, CCE and GST superfamilies in 160 insect species [11] providing interesting conclusions regarding adaptations to different diets, it did not include any species of the order Hemiptera. Hence, the information on this group, relevant both in terms of number of species and its economical/sanitary impact, is still scarce. Furthermore, an evolutionary analysis identifying differences in CYP, GST and CCE complements between sap and blood sucking heteropterans could reveal molecular

adaptations to either hematophagy or plant-feeding. A comparative analysis of gene expansions/reductions and selective pressure between blood and sap sucking species could reveal common characteristics within the heteropterans, and also differences that could shed light on the evolutive forces exerted on different genes and gene families.

Here we perform a comparative analysis on the complement of GST, CCE, and CYP gene superfamilies across five hemipteran species; four heteropterans (*Nezara viridula*, *Halyomorpha halys*, *Cimex lectularius* and *Rhodnius prolixus*), and one Auchenorryncha (*Nilaparvata lugens*). The stink bug *N. viridula* (Hemiptera: Pentatomidae) is an important crop pest, which generates a severe economic impact, particularly in soybean [12]. In spite of its economical relevance from North to South America, the genome of *N. viridula* has not been published to date, although a high-quality and complete transcriptome is available [12]. *Halyomorpha halys* (Hemiptera: Pentatomidae), also phytophagous, has exceptionally high levels of polyphagy, which makes it an efficient invasive plague that expanded from Asia to North and South America, Europe and Australia [13]. *Nilaparvata lugens* (Hemiptera: Delphacidae) is a planthopper which is a monophagous herbivore of rice, constituting a serious threat for production [14]. The bed bug *C. lectularius* (Hemiptera: Cimicidae) and the kissing bug *R. prolixus* (Hemiptera: Reduviidae) are obligate hematophagous during their entire life cycle, and are species of sanitary relevance as human ectoparasites. Furthermore, *R. prolixus* is a vector of *Trypanosoma cruzi*, the causative agent of Chagas disease, a neglected life-threatening disease which affects millions of people all over the world (https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis). The comparative analysis presented herein revealed particularities in the detoxificant-related complement of Heteroptera with respect to other insect species, and point to gene families that could maintain a fast evolution rate.

Results And Discussion

CYP, CCE and GST repertoires were compared across four heteropterans (*N. viridula*, *H. halys*, *R. prolixus* and *C. lectularius*) and one Auchenorryncha (*N. lugens*). The latter was included in the analysis as a hemipteran non-heteroptera outgroup. Considering the sum of members of all the gene superfamilies under analysis, 151 detox genes in *N. viridula*, 116 in *C. lectularius*, 194 in *R. prolixus,* 238 in *H. halys* and 96 in *N. lugens* were identified (Table 1). CYP superfamily was the largest in terms of number of genes in all the species, followed by CCE, being GST the least represented (Table 1). This distribution is not observed in other insect species; *Drosophila melanogaster*, for example, possesses more GST than CCE genes [4]. When considering each superfamily separately, the *H. halys* CYPs (134), CCEs (74) and GSTs (30) were the most abundant. By comparison, the *N. lugens* repertoire was reduced in every family (48 in CYPs, 36 in CCEs and 12 in GSTs) (Table 1). The CCE and CYP superfamilies of *R. prolixus* have 119 and 61 members, respectively, followed by *N. viridula* with 82 and 64, respectively (Table 1). The GST superfamily resulted slightly larger (15) in *N. viridula* than in *C. lectularius* and *R. prolixus*, counting 14 members in each species (Table 1).

When the sum of CYPs, CCEs and GSTs was compared to the complete gene sets in the genomes, significant expansions in *C. lectularius, H. halys* and *R. prolixus* were detected compared to *N. lugens* (Supplementary information 1). Within Heteroptera, both *H. halys* and *R. prolixus* presented significant expansions with respect to *C. lectularius* (Supplementary information 1). This was supported by different numbers of duplications and losses (Figure 1), with *H. halys* and *R. prolixus* being the species with a higher number of duplication events detected in the gene families studied.

Cytochromes P450

CYPs are involved in the degradation of xenobiotics, from both the diet and the environment, being an important factor in the response of insects to chemical stress [15]. They also participate in key metabolic processes, such as the degradation of pheromones or the biosynthesis of signaling molecules. CYP genes are divided into 4 large clans (Mitochondrial, CYP2, CYP3 and CYP4 clans), which in turn are subdivided into families and subfamilies [15].

CYP complement in the *N. lugens* genome was significantly smaller when compared with the CYP complement of other heteropteran genomes (Supplementary information 1). Also, the CYP superfamily was significantly expanded in both *R. prolixus* and *H. halys* when compared to *C. lectularius* (Supplementary information 1). Phylogenetic analysis allowed the classification of CYP superfamily members into clans and families (Figure 2), which presented different evolutionary dynamics (Table 2).

Mitochondrial Clan

The number of mitochondrial clan genes was similar among species: 8 genes in *N. lugens*, *R. prolixus* and *H. Halys*, 7 in *C. lectularius*, and 6 in *N. viridula* (Table 1). These numbers reflect stability in the evolution of the clan, with few events of gene duplications and losses, although the number of losses was slightly higher (Table 2). Many genes from this clan have essential roles in the life cycle of insects, such as the *Halloween* genes that are involved in the synthesis of ecdysteroids [15]. *Halloween* genes belong to families 315 (or *sad*), 302 (or *disembodied*; *dib*), and 314 or (*shadow*; *shd*). All of them were represented with 1 gene for each family in the hemipteran databases, with the exception of 2 CYP302 genes in *H. halys* (Figure 2A). The family CYP394 from the mitochondrial clan had 1 representative in both *R. prolixus* and *C. lectularius*, whereas CYP3221 had one representative in each one of the pentatomids. *N. lugens* had no representatives of these families, but it had 1 gene classified as CYP419. Interestingly, CYP404 was represented by one gene in *N. lugens*, *C. lectularius* and *R. prolixus*, but was absent in the pentatomids. The largest mitochondrial CYP family in the databases analyzed here was CYP301, with 3 genes detected in *R. prolixus*, *H. halys* and *N. lugens*, and 2 in *C. lectularius* and *N. viridula* (Figure 2A). Coincident to previous observations for triatomines, none of the sequences identified in the hemipterans analyzed herein were clustered with the CYP12 family, related to insecticide resistance [10, 15].

CYP2 Clan

Like the mitochondrial clan, CYP2 contains several *Halloween* genes (CYP306 and CYP307 families; named *phantom* and *spook* respectively) [15](Figure 2B). Eight CYP2 genes were detected in *N. lugens*, 7 in *R. prolixus,* 6 in *C. lectularius* and *H. halys*, and 4 in *N. viridula* (Table1, Figure 2B). The evolutionary dynamic of this clan was similar to that of the Mitochondrial clan: few events, with gains slightly more prevalent than losses (Table 2). A single member of the families CYP305 and CYP307 were detected in each species database, with the exception of 2 CYP307 genes in *N. lugens*. For CYP15, one member was detected for each database. Two genes were classified in the CYP306 family in *R. prolixus* whereas one member was found in each of the other species under analysis. CYP304, previously reported in *D. melanogaster* and other insects, was represented by 1 gene in *N. lugens*, but not in the other species analyzed, suggesting a loss of this family in Heteroptera. Finally, only one sequence belonging to each of the CYP18 and CYP303 families were detected in the 4 genomes analyzed, but not in the *N. viridula* transcriptome (Figure 2B). In a previous work, CYP303 and CYP18 were not detected in the transcriptomes of *Triatoma dimidiata* and *Triatoma infestans* [4], which could suggest that these families are low or conditionally expressed in heteropterans, hindering their identification even in complete transcriptomic databases.

CYP3 Clan

CYP3 and CYP4 clans are always bigger than the Mitochondrial and CYP2 clans in the insect genomes [15]. These clans have proliferated as a result of gene duplication events, which allowed their diversification and neo-functionalization [10]. Similar results were observed in the hemipteran species analyzed herein. The CYP3 clan has a very dynamic evolution and a significant expansion trend with 2.5 folds more gene duplications than losses (Table 2). Within the CYP3 clan, there are multiple xenobiotic metabolizing families involved in phytochemical detoxification and insecticide resistance, mainly represented by CYP6 and CYP9 families [15].

The number of sequences encoding CYP3 enzymes was 55 in *R. prolixus*, 31 in *C. lectularius*, 78 in *H. halys* and 41 in *N. viridula* (Table 1; Figure 2C); in *N. lugens* the CYP3 complement was significantly reduced (8 genes) with respect to their counterpart in heteropterans (Table 1; Figure 2C). Accordingly, many duplications were observed in all phylogenetic branches leading to Heteroptera species. In particular, *R. prolixus* had 30 duplications compared to only 11 in *C. lectularius*, which explains most of the difference between these two hematophagous species. A large number of duplications was observed in the ancestor of *N. viridula* and *H. halys*, and again in the lineage of *H. halys* (Figure 1, Table 2). The discrepancy between the two pentatomids could be related to differences in the broad of their diet, although the much lower number of genes and duplications, and more gene losses detected in *N. viridula* could be due to the low expression of some of these genes, impairing their detection in the assembled transcriptome. Genomic information of this species will be important to refine present results.

CYP6 is the most abundant CYP3 family in the herbivorous species analyzed (35.9% of the CYP3 are CYP6 in *H. halys*, 39% in *N. viridula* and 62.5% in *N. lugens*), and the second larger in the hematophagous

(22.6% sequences in *C. lectularius* and 14.5% in *R. prolixus*). Within this family, subfamily LV seems to be exclusive for pentatomids (18 in *H. halys* and 11 in *N. viridula*). The second larger family in pentatomids, and the more abundant in both *R. prolixus* and *C. lectularius*, is CYP395 (23.1 % of the sequences in *H. halys*, 12.2% in *N. viridula*, 25.8% in *C. lectularius* and 25.4% in *R. prolixus*). However, none of the *N. lugens* CYP sequences was classified within this family. Subfamilies CYP395-S and CYP395-R are only represented in pentatomids, whereas subfamilies CYP395-C to CYP395-F contain only *R. prolixus* sequences. Families CYP3225 to CYP3231 are exclusive for the pentatomids. CYP3225 family is larger in *H. halys* (11 sequences) when compared to *N. viridula* (3 sequences). On the contrary, subfamily CYP3226 contains 2 genes from *H. halys* and 5 from *N. viridula*. CYP3227 contains 6 *H. halys* and 4 *N. viridula* sequences, whereas CYP3228 to CYP3231 are represented by one gene in both *H. halys* and *N. viridula*.

Families CYP3084 to CYP3089 and CYP3091 are absent in pentatomids. The latter is represented by 6 sequences in *R. prolixus* and 1 in *C. lectularius* and *N. lugens.* Families exclusive for *R. prolixus* are CYP3084, CYP3085, CYP3086, CYP3088 and CYP3096. Family CYP3090, which was previously described only in triatomines [4], seems to be ubiquitous in Heteroptera with 1 sequence in *H. halys*, *N. viridula, R. prolixus*; and *C. lectularius*. Finally, CYP3092 family, described for triatomines for the first time in previous works [4, 16], was also detected in pentatomids (4 sequences in *R. prolixus*, 7 in *H. halys*, 4 in *N. viridula*), even though subfamily CYP3092A is exclusive for *R. prolixus*, whereas subfamilies CYP3092D and CYP3092E were only detected in the pentatomids.

We observed that CYP9 family (belonging to CYP3 Clan) was absent in all the hemipteran databases analyzed here. In a previous work, we reported the absence of CYP9 family in triatomine species [4]. Furthermore, a recent work analyzing the genome of the predaceous hemipteran *Orius laevigatus* (*Anthocoridae*) also reported the lack of CYP9 family [17]. Globally, results point to the lack of CYP9 class as a common observation in hemipteran genomes. Given that CYP9 is a large family in insect genomes belonging to different Orders, and that it was associated with insecticide resistance and xenobiotic detoxification [10], the lack of this family in Hemiptera could be a relevant finding for evolutive and applied entomology. On the contrary, CYP395 family is one of the largest groups in heteropteran databases. Given the absence of the CYP9 family, CYP395 could have a role in detoxification in heteropterans.

CYP4 Clan

CYP4 is the second most numerous CYP clan in the genomes of insects from different orders [15]. Many genes belonging to this clan have been involved in detoxification [18]. The analysis revealed 49, 42, 31, 12 and 24 CYP4 genes, in *R. prolixus*, *H. halys*, *N. viridula, C. lectularius* and *N. lugens*, respectively. CYP4 subfamily was significantly larger respect to the total CYP complement in *N. lugens* compared to *H. halys* and *C. lectularius* (Supplementary information 1). This was due to the small number of CYP3 genes detected in the planthopper. A significant reduction was also detected for CYP4 in *C. lectularius* compared to both *N. viridula* and *R. prolixus* (Table 1, Figure 2D, Supplementary information 1). This result reflects

an expansion of the clan in *R. prolixus* due to 37 gene duplication events (Table 2). Our methodology cannot rule out that the expansion occurred in the ancestor of the hematophagous species with a posterior massive gene loss in *C. lectularius*. However, previous data comparing CYP4 complement in different triatomine species, including *R. prolixus*, supports the first hypothesis [4]. A large number of duplications (23) was also detected in the ancestor of the pentatomids (Table 2). Again, the losses detected in *N. viridula* should be confirmed with genomic information.

CYP4 genes were classified in CYP4, CYP3222, CYP3223, CYP3224 and CYP3093 families, being 3222 to 3224 exclusive for pentatomids (Figure 2D). Besides, CYP3093 family is exclusive to *R. prolixus.* CYP3093 is expanded in *R. prolixus* (71.4% of the CYP4 clan), due to a gene bloom (Figure 2D). It was proposed that expansions of CYP genes occur in response to environmental stimuli, leading to the potential to develop insecticide resistance [10]. Functional information on CYP40393 does not exist for *R. prolixus* to date, with the exception of bioinformatic molecular docking models. For several CYP43093 members, these models proposed a favorable interaction between the pyrethroid deltamethrin and the active site, suggesting a possible role in insecticide metabolism [19]. In *T. infestans* CYP40393 are highly expressed in tegument, which is the first barrier to toxics; members of this family are overexpressed in resistant *T. infestants* populations [19]. Altogether, the evidence allows us to hypothesize that CYP43093 bloom confers to *R. prolixus* the potential to acquire resistance to chemical insecticides. The bloom of CYP3093 observed in *R. prolixus* is not shared by other triatomine species [4], nor by other heteropterans (present results), suggesting that it may be a recent event in the evolution of this species.

A hundred percent of the CYP4 sequences of *C. lectularius* (12) belong to the CYP4 family, which is also represented in *R. prolixus* (14 genes, 28.5% of the clan), *H. halys* (30 genes, 71.4%), *N. viridula* (20 genes, 64.5%) and *N. lugens* (23 genes, 95.8%). It has been suggested a role of this family in insecticide tolerance and resistance in triatomine species, given their higher expression in *T. infestans* resistant to pyrethroids [4, 20]. Moreover, RNAi mediated gene-silencing of CYP4 family members led to an increased susceptibility to deltamethrin in both *T. infestans* [21] and *R. prolixus* [22].

Carboxyl/Cholinesterases

Database searches and manual gene curation (Supplementary information 4) revealed 36 genes encoding CCEs in *N. lugens*, 74 in *H. halys*, 46 in *C. lectularius* and 54 in *N. viridula*, whereas *R. prolixus* genome encodes 61 genes (Table 1). It is a dynamic detoxification gene family with 150 duplications and 54 gene losses detected (Table 2, Figure 1). These numbers suggest an expansion in the superfamily compared with *D. melanogaster* (Supplementary information 1), which was more pronounced in some branches of the phylogeny. Interestingly, very large expansions occurred in the ancestors of pentatomids (51 duplications) and individually in *R. prolixus* (41) and in *C. lectularius* (25). Overall, this dynamic was mostly branch specific, suggesting that it could be the result of adaptation to different environmental niches. The number of gene losses is certainly overestimated due to lack of genomic data in *N. viridula*.

The CCE superfamily includes proteins with different functions which were classified into three functional classes [23, 24]: neuro/developmental (ND) class which, with the exception of acetylcholinesterases (Ach), lacks the catalytic activity; dietary/detoxification (DD) class, and hormone and pheromone processing (HPP) esterases. In a previous work, we showed that triatomine species lack DD class, and that most of the CCEs in these species were classified as HPP [4]. More recently, Bailey et al [17] also reported the lack of DD in the predaceous hemipteran *O. laevigatus.*

We have classified CCE sequences of *N. lugens*, *N. viridula* and *C. lectularius* based on their phylogenetic relationships to CCEs of *H. halys* and *R. prolixus* previously reported [4, 13] and the analysis of characteristic conserved residues in their sequences (Supplementary information 2; Figure 3). Interestingly, DD class was only represented in *N. lugens* (4 genes). The lack of DD class CCEs in blood feeding and predaceous Hemiptera has been proposed as a consequence of their diet; given that these species are not exposed to plant secondary metabolites they would not need this class of CCEs [4, 17]. However, our results revealed that DD CCEs are also absent in pentatomids, suggesting its lose during the evolution of Heteroptera.

The sequences considered as HPP were defined here taking into account phylogenetic relationships and the presence of β-esterases in the groups. In agreement to previous reports on Heteroptera [4, 17] we observed an important diversification in the HPP. Forty HPP class CCEs were detected in *R. prolixus*, 30 in *C. lectularius*, 56 in *H. halys*, 46 in *N. viridula* and 16 in *N. lugens*. These numbers are considerably bigger than the reported for species of other Orders. The HPP clades that contain heteropteran sequences harbor large expansions, especially in pentatomids. Three CCE sequences of *H. halys*, 2 of *N. viridula*, 2 of *N. lugens* and 21 of *C. lectularius* could not be classified as β-esterases by sequence identity analysis, but they are part of the hormone and pheromone processing class according to the phylogeny (Figure 3).

The ND class encodes neuroligin, gliotactin, glutactin, and neurotactin proteins, which are not catalytic. It also encodes Ach, which is involved in neurotransmission, and is the target site of organophosphate and carbamate insecticides [25]. ND class presented smaller species-specific expansions in the species analyzed as compared to the HPP class. In the ND class, the gene complement was significantly expanded in *R. prolixus* in respect to *N. viridula* and *H. halys* (Supplementary information 1). *D. melanogaster* and other dipteran genomes possess 1 Ach encoding-gene. A single gene encoding Ach was also found in *N. lugens* genome, while *R. prolixus*, *H. halys,* and *N. viridula* have 2 each. Remarkably, the *C. lectularius* genome encodes 5 Ach paralogue genes (Figure 3). All the species analyzed possess 1 gene classified as “putative neuroreceptor” (Gliotactin or clade K) within the neurodevelopmental class. *N. lugens*, *H. halys*, *R. prolixus* and *C. lectularius* have 1 gene in the neurotactin group and 1 in the “uncharacterized” or I group, whereas *N. viridula* transcriptome do not have any representatives in these groups. *H. halys*, *C. lectularius* and *N. viridula* have 1 gene in the glutactin group while *R. prolixus* has 2. The most numerous neurodevelopmental class is the one formed by neuroligins: 2 in *N. viridula*, 5 in *C. lectularius*, 7 in *H. halys*, *10 in N. lugens* and a significant expansion with 13 representatives in *R. prolixus* (Supplementary information 1).

Altogether, our results and previous reports suggest a particular configuration of CCEs complement in Heteroptera, with a lack of DD and expansions in HPP class. In the absence of DD class, catalytic β-esterases could have a role in detoxification in Heteroptera. In this sense, the expansions observed in the HPP class may functionally counteract the absence of DD in the ability of these species to cope with toxic xenobiotics.

Glutathione Transferases

GST enzymes play a fundamental role in the detoxification of endogenous and xenobiotic compounds, but they also participate in hormone biosynthesis, intracellular transport and protection against oxidative stress [26]. These enzymes can metabolize compounds by reductive dehydrochlorination or by conjugation reactions with reduced glutathione, generating soluble metabolites that are easier to eliminate. GSTs are classified in microsomal, mitochondrial and cytosolic, depending on their location in the cell; mitochondrial GSTs have not been found in insects. While 7 types of cytosolic GSTs have been recognized in mammals, insects have 4 of these families known as Omega, Sigma, Theta and Zeta. Delta and Epsilon classes, associated with insecticide resistance in Diptera, were only reported in insects to date [26].

Fourteen GSTs were found in both *R. prolixus* and *C. lectularius* genomes, 15 in *N. viridula* (15) and 30 in *H. halys* (Table 1). Twelve GST encoding genes were identified in the *N. lugens* genome, in agreement with a previous report [14]. Although this family was less dynamic in terms of gene birth and death as compared to the other detox families, it also presented an expansion trend (Figure 1). The largest number of duplications were observed in the ancestor of the pentatomids and in *H. halys*, suggesting an adaptive role of this family in phytophagous heteropterans. The absence of genomic data for *N. viridula* may be hiding duplications in this species and leading to overestimation of the number of gene losses. The expansion of the GST family observed in the *H. halys* genome is significant when compared with both *N. lugens* and *R. prolixus* (Supplementary information 1).

Microsomal GSTs have not been reported to play a role in the detoxification of xenobiotics; even though they differ structurally from cytosolic GSTs, they catalyze similar reactions [26]. We found 3 sequences in *H. halys* and *C. lectularius*, 2 in *N. lugens*, while *R. prolixus* and *N. viridula* databases encode for only 1 microsomal GSTs.

The hemipteran species analyzed herein present a low number of Delta (2 in *H. halys*, *N. lugens* and *C. lectularius*; 1 in *R. prolixus,* 0 in *N. viridula*) and Epsilon GSTs (2 in *N. lugens* and 0 in the heteropterans; Table 1) (Figure 4, Supplementary Figure 2). With the only exception of *N. lugens* and *Trialeurodes vaporariorum* [17], the absence of Epsilon GSTs is a common finding in hemipteran genomes. Conversely, other GST groups are larger in these species compared to other Orders [4]. Particularly, Sigma class was the largest group, especially in pentatomids (Table 1). Fifty percent of the total GSTs of *R. prolixus* (7 out of 14), 60% in *N. viridula* (9 out of 15), 42.8% in *C. lectularius* (6 out of 14), 63.3% in *H. halys* (19 out of 30), and 25% in *N. lugens* (3 out of 12) belong to Sigma class (Table 1). These results suggest that

Sigma class GSTs could be relevant for detoxification in Hemiptera, compensating the absence or reduction in other GST groups.

The Omega class has a cysteine residue in their active site, which allows the catalysis of thiol transferase and reduction reactions that are not catalyzed by the other GST classes [27]. Among the hemipterans studied herein, *H. halys* possess 3 genes in this class, *N. viridula* has 2, whereas *N. lugens*, *R. prolixus* and *C. lectularius* have 1 representative each (Figure 4, Table 1).

The Theta class has been proposed to be a contributor to the detoxification of xenobiotics, due to the protective action against oxidants and dehalogenating activity [6]. *R. prolixus* has 3 genes in this class, whereas *C. lectularius*, *H. halys* and *N. viridula* have 2 and the *N. lugens* genome encodes 1. Finally, 1 gene belonging to the Zeta class was detected in each of the hemipteran databases analyzed here.

Genomic clusters and evolution of detoxifying gene superfamilies

Paralogue genes that are originated by relatively recent duplications are usually organized in clusters in the genomes. Furthermore, duplicated genes that are in adjacent genomic regions could be regulated in a coordinated way. For this reason, it is possible that evolutive forces may be acting to maintain genes in a cluster organization even a long time after the duplication events that generated the cluster [28, 29].

A positive linear correlation was observed between the total number of detoxifying genes (CYP, CCE and GST families) and the detoxifying genes organized in genomic clusters (R=0.96) (Figure 5A). This tendency is maintained when each superfamily is considered separately (R=0.99 for GSTs; R=0.99 for CYPs; R=0.81 for CCEs). The detoxifying genes forming clusters represented 25% in *N. lugens*, 55.2 % in *C. lectularius*, 62.4% in *R. prolixus* and in 57.14% *H. halys* (Figure 5B). These results are in accordance with the large estimated number of gene duplications (Figure 1).

Considering the total of CYP members identified in the four species under analysis with available genomic sequence 58.1% form clusters; these percentages are 55.7 for CCEs and 26.8 for GSTs (Figure 5C). All the families within the CYP superfamily showed a tendency to form clusters, with the exception of the mitochondrial clan. CYP2 contains 48% of the genes in clusters, CYP3 67% and CYP4 66% (Figure 5D). The large percentage of genes in genomic clusters is in accordance with the large number of species specific, probably recent, duplications observed in the CYP3 and CYP4 gene families.

Most CYP3 and CYP4 genes were clustered in the heteropterans (Figure 6A); whereas most of the genes belonging to CYP2 were clustered in the hematophagous species analyzed. HPP class of CCE has 68% of genes located in clusters. This tendency is strong in Heteroptera, being very evident in *C. lectularius* (95% of clustered genes).

In the case of the GSTs, the highest proportion of clustered genes (close to 50%) is given by the Sigma and Omega classes (Figures 5D and 6C).

Our phylogenetic analysis of gene gains and losses revealed some interesting patterns. The large number of both gene duplications and losses suggests that the detox gene families have a very dynamic evolution, with high turnover rates resulting in many species and lineage-specific genes. This evolutionary pattern resembles that of chemosensory gene families, which similarly to the detox gene families, have a large adaptive potential [29]. Interestingly, it has been recently suggested that some chemosensory genes are involved in detoxification in insects, probably by sequestering toxic molecules [30].

Among the detoxification superfamilies, the CYP superfamily was the most dynamic, with a total of 299 events, the majority of which were duplications, indicating that this family is expanding. The CYP3 and CYP4 clans are the most dynamic within the superfamily (Figure 1, Table 2). The second more dynamic superfamily was CCEs, with 204 events. The GST superfamily had less events in comparison, which is in accordance with having less genes as compared to the other superfamilies. Overall, all families had more duplications than losses, which suggests a general expansion trend.

Comparisons across species showed an expansion of detox genes in the lineage leading to the phytophagous heteropterans (Figure 1). This is especially true for superfamilies CCE and CYP. The GST superfamily had fewer events overall, but appears to be more dynamic in the phytophagous species as compared to the hematophagous species. Besides, a large expansion was observed in CYP in *H. halys*, which also had many duplications in the GST family. These results are in accordance with phytophagy being a newly acquired diet in the group from the predator common ancestor of all the heteropterans [1]. A phytophagous diet represents an important adaptive challenge, as a result of the arms race between plants trying to avoid herbivory and insects trying to overcome the toxicity of chemical compounds produced by plants. The broad complement of detoxificative enzymes resulting from both ancestral and species-specific duplication in *H. halys* are certainly instrumental for its feeding on different plant species, facilitating its success as an invasive species. *N. viridula* had a very different pattern, most likely as a consequence of having only transcriptomic data available for this analysis. Due to the incompleteness of transcriptomes, gene numbers and duplications are probably underestimated, while gene losses are likely overestimated in this species. However, *H. halys* is a polyphagous pentatomid, when compared to *N. viridula*, suggesting that the observation of a broader complement of detoxificant enzymes could be not a consequence of the lack of genomic information in the latter, but a reflect of different feeding habits between the species.

In contrast with the phytophagous species, significant expansions were not observed in the ancestors of the hematophagous, which in turn had many species-specific expansions. The GST superfamily, however, seems to be shrinking among the blood-feeding heteropterans. Hematophagy evolved independently in *C. lectularius* and *R. prolixus* as their most recent common ancestor was likely a predator[1]. Hence, it makes sense that these two species had independent evolution of their detox gene complements in response to the diet change, with many species-specific duplications and losses. *R. prolixus*, however, had many more duplications as compared to *C. lectularius*, which accounts for most of the large difference in the total number of detox genes between these two species. In *C. lectularius*, besides the GSTs, the CYPs also seem to be in a reduction trend.

Concluding Remarks

Here we present a comparative analysis of detoxification gene superfamilies among Heteroptera species with different feeding specializations. The availability of transcriptomic data allowed us to include *N. viridula*, an important crop pest that is still understudied. Nevertheless, a complete genomic sequence of the species is essential to close gaps in the understanding of its detoxification gene complement and in the evolution of these genes in Heteroptera. Our results indicate a reduction in Heteroptera of several enzyme families associated to xenobiotic detoxification: absence of the class Epsilon and a reduced class Delta in the GST superfamily; absence of the mitochondrial family CYP12; absence of the CYP9 family in CYP3 clan; and the absence of the DD class of CCEs. Conversely, other detoxification-related families are expanded in the group , and have had a dynamic evolution: CYP3 clan, HPP in the CCE superfamily and Sigma class in GST superfamily. These characteristics were previously proposed as particularities of the Triatomine subfamily, or even to non-herbivorous heteropterans [4, 17]. Our results demonstrate that these features could be extended to the suborder Heteroptera.

Our comparative analysis suggests that diet may be an important driver of detoxificant gene family evolution in Heteroptera. Nevertheless, a more dense species sampling would be necessary to check this hypothesis. For instance, it would be interesting to include more species belonging to Cimicomorpha, both phytophagous and carnivorous, to separate diet from phylogenetic constraints in the evolution of the detox gene superfamilies in this group. In addition , the inclusion of non-pentatomid phytophagous species would allow checking whether switching to an herbivorous diet is associated with an increase in the detoxification gene repertoire. On the other hand, it would be interesting to study the particular genes involved in the parallel evolution of blood compound detoxification in *C. lectularius* and *R. prolixus*.

Methods

Gene identification and sequence analysis

BLASTp searches (with expected threshold < 0.0001) were performed using the PFAM domains PFAM domains PF02798 and PF00043 (GSTs), PF00135 (CCEs) and PF00067 (CYPs) as queries in the predicted protein datasets of *C. lectularius* (NCBI accession number GCF_000648675.2), *H. halys* (i5k number OGS1.2), *R. prolixus* (NCBI accession number GCA_000181055.3) and *N. lugens* (NCBI accession number GCA_014356525.1).

PFAM domains mentioned above were used as queries to perform tBLASTn searches [31] on a *N. viridula* complete transcriptome [12]. All the positive hits against the non-redundant transcriptome were manually curated and analyzed using homologous sequences as reference to reconstruct transcripts whenever evident errors in sequencing (such as frameshifts or missassemblies) were detected. Transcripts encoding proteins shorter than 100 amino acids were discarded.

Positive hits for all the searches were re-analyzed in two sequential steps: an InterProScan search [32] using the Gene3d, PfamA, and SuperFamily applications, and BLASTp searches against the non-redundant database on the NCBI. Those hits not belonging to superfamilies of interest were discarded. Sequences encoding CYP450s, CCEs and GSTs from the *N. viridula* transcriptome are presented in Supplementary information 3. Accession numbers of all sequences used for the analysis from *N. lugens*, *C. lectularius*, *R. prolixus* and *H. halys* are presented in Supplementary information 4.

Phylogenetic analysis

Protein sequence alignments for the target protein families were generated with MAFFT [33] using the G-INS-i strategy and the following settings: option "leave gappy regions" active; Unaligned level = 0.1; Offset value = 0.12; Maxiterate = 1000. The alignments were trimmed using trimAl v1.2 [34] with default parameters except for the gap threshold (-gt) that was fixed at 0.3. The trimmed alignments were used to build a phylogenetic tree for each family based on the approximately maximum likelihood approach with IQ-tree v 1.6.12 [35] using -B 1000 and -alrt 1000 settings to combine ModelFinder option, tree searching based on 1,000 replicates and estimation of branch support using the approximate Likelihood Ratio Test based on the Shimodaira-Hasegawa (aLRT-SH) and the ultrafast bootstrap or UFBoot procedures [36]. The best-fit amino acid substitution model estimated by IQ-tree for each protein family and chosen according to Bayesian Information Criterion were: WAG+F+R8 for CCEs, LG+I+G4 for CYP2, Q.yeast+R7 for CYP3, Q.pfam+F+R6 for CYP4, LG+F+R4 in Mitochondrial Clan and LG+R4 for GSTs. The phylogenetic trees were displayed and edited using FigTree (http://tree.bio.ed.ac.uk/software/figtree) and iTol online tool (https://itol.embl.de).

N. viridula CYP and GST sequences were named according to their phylogenetic relationships to annotated proteins in the species with sequenced genomes [13, 16]. Once phylogenetic trees were generated, sequence identity analyses were performed using CD-HIT with an identity threshold of 30% to confirm the proposed annotation. In the case of the CCEs sequences were aligned to a reference sequence from *D. melanogaster* (CG17907) using Clustal Omega [37]. The presence of conserved regions and specific residues along the sequences was verified to confirm classification (Supplementary information 2).

Duplication analysis and enrichment statistical analysis

Chi-Squared Tests were performed in order to identify significant expansions in gene families. Two by two contingency tables (Supplementary information 1) were built between two species and:

-The number of sequences encoding either CYPs, CCEs or GSTs *vs.* the rest of genes of a particular genome.

-The number of sequences encoding a particular detoxificant superfamily (CYPs, CCEs or GSTs) *vs.* the sum of genes belonging to the other two superfamilies under study.

-The number of sequences encoding enzymes from a particular group belonging to CYPs, CCEs or GSTs *vs.* the rest of sequences of the same superfamily.

Gene tree *vs.* species tree reconciliation method was used. This method compares gene trees with the species tree and identifies gene gains and losses that could explain the differences between them. In this way, we manually inspected the phylogenetic trees obtained for each family following [38]. This method may underestimate the number of events, as it does not take into account genes that may have been gained and were posteriorly lost (we did not analyze pseudogenes). However, because each orthogroup is analyzed separately, the results are more accurate than when automated methods are applied [38]. The objective was to have an overall idea of the gene birth-and-death dynamics of each family and detect differences between species and clans/classes. It is important to bear in mind that the estimates of gene gain and losses are dependent on the annotations; faulty annotations will lead to an overestimate of losses.

Cluster analysis

General feature format (GFF) files were downloaded for *R. prolixus* (GCA_000181055.3_Rhodnius_prolixus-3.0.3.gff) from NCBI; *H. halys* (halhal_OGSv1.2.gff) from i5K; *C. lectularius* (GCF_000648675.2_Clec_2.1_genomic.gff) from NCBI; and *N. lugens* (GCF_014356525.1.gff) from NCBI. Target gene locations (coordinates and scaffold) for the four species were obtained from each GFF file using a list of IDs and the Seqtk tool. Two genes were considered into a genomic cluster when they belong to the same gene family and were separated by less than 100,000 bp [39].

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations

Consent for publication

Not applicable

Availability of data and materials

Database used for this study are deposited under the following accession numbers: *C. lectularius:* NCBI accession number GCF_000648675.2; *H. halys:* i5k number OGS1.2; *R. prolixus:* NCBI accession number GCA_000181055.3; *N. lugens*: NCBI accession number GCA_014356525.1. The assembled transcriptomic dataset for *N. viridula* is available at the NCBI-TSA GGPJ00000000.

Competing interests

Authors have no competing interests

Funding

Funding was provided by Agencia Nacional de Promoción de la Innovación, el Desarrollo Tecnológico y la Innovación (PICTstartup 2018-0275 and PICT 2018-0862 to S.O, PICT 2019-02668 to J.M.L.E, PICT 2019-1533 to L.T.). S.O, J.M.L.E and F.C.A are researchers of Consejo Nacional de Investigación, Ciencia y Tecnología (CONICET). M.V and L.T are recipients of research fellowships from CONICET.

Authors' contributions

MV: Conceptualization, investigation, data curation, formal analysis, visualization, writing the original draft. LT: Conceptualization, investigation, data curation, formal analysis, writing review and editing. JMLE: Conceptualization, data curation, formal analysis, writing review and editing. FCA: Conceptualization, data curation, formal analysis, resources, writing review and editing. SO: Conceptualization, investigation, formal analysis, writing the original draft, funding acquisition, project administration.

Acknowledgements

Not applicable.

References

1. Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, et al. Phylogenomics and the evolution of hemipteroid insects. Proc Natl Acad Sci U S A. 2018;115:12775–80.
2. Wu J, Baldwin IT. New insights into plant responses to the attack from insect herbivores. Annu Rev Genet. 2010;44:1–24.
3. Sterkel M, Oliveira JHM, Bottino-Rojas V, Paiva-Silva GO, Oliveira PL. The Dose Makes the Poison: Nutritional Overload Determines the Life Traits of Blood-Feeding Arthropods. Trends Parasitol. 2017;33:633–44.
4. Traverso L, Lavore A, Sierra I, Palacio V, Martinez-Barnetche J, Latorre-Estivalis JM, et al. Comparative and functional triatomine genomics reveals reductions and expansions in insecticide resistance-related gene families. PLoS Negl Trop Dis. 2017;11:e0005313.
5. Rane R V., Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Curr Opin Insect Sci. 2019;31:131–8.
6. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler M a., Berenbaum MR, et al. A deficit of detoxification enzymes: Pesticide sensitivity and environmental response in the honeybee. Insect Mol Biol. 2006;15:615–36.
7. Lee SH, Kang JS, Min JS, Yoon KS, Strycharz JP, Johnson R, et al. metabolism. 2011;19:599–615.
8. Rane R V., Walsh TK, Pearce SL, Jermiin LS, Gordon KH, Richards S, et al. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? Curr Opin Insect Sci. 2016;13:70–6.

9. Nei M, Rooney AP. Concerted and Birth-and-Death Evolution of Multigene Families. Annu Rev Genet. 2005;39:121–52.
10. Feyereisen R. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochim Biophys Acta - Proteins Proteomics. 2011;1814:19–28.
11. Rane R V., Pearce SL, Li F, Coppin C, Schiffer M, Shirriffs J, et al. Genomic changes associated with adaptation to arid environments in cactophilic Drosophila species. BMC Genomics. 2019;20:1–22.
12. Lavore A, Perez-Gianmarco L, Esponda-Behrens N, Palacio V, Catalano MI, Rivera-Pomar R, et al. Nezara viridula (Hemiptera: Pentatomidae) transcriptomic analysis and neuropeptidomics. Sci Rep. 2018;8.
13. Sparks ME, Bansal R, Benoit JB, Blackburn MB, Chao H, Chen M, et al. Brown marmorated stink bug, Halyomorpha halys (Stål), genome: Putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. BMC Genomics. 2020;21:1–26.
14. Zhou WW, Liang QM, Xu Y, Gurr GM, Bao YY, Zhou XP, et al. Genomic Insights into the Glutathione S-Transferase Gene Family of Two Rice Planthoppers, Nilaparvata lugens (Stål) and Sogatella furcifera (Horváth) (Hemiptera: Delphacidae). PLoS One. 2013;8.
15. Feyereisen R. Evolution of insect P450. Biochem Soc Trans. 2006;34 Pt 6:1252–5.
16. Schama R, Pedrini N, Juárez MP, Nelson DR, Torres AQ, Valle D, et al. Rhodnius prolixus supergene families of enzymes potentially associated with insecticide resistance. Insect Biochem Mol Biol. 2015; August. https://doi.org/10.1016/j.ibmb.2015.06.005.
17. Bailey E, Field L, Rawlings C, King R, Mohareb F, Pak KH, et al. A scaffold-level genome assembly of a minute pirate bug, Orius laevigatus (Hemiptera: Anthocoridae), and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests. BMC Genomics. 2022;23:1–21.
18. Dulbecco AB, Caldero GM. Cytochrome P450 Genes of the CYP4 Clan and Pyrethroid Resistance in Chagas Disease Vectors. 2022;3 February:1–8.
19. Dulbecco AB, Moriconi DE, Calderón-Fernández GM, Lynn S, McCarthy A, Roca-Acevedo G, et al. Integument CYP genes of the largest genome-wide cytochrome P450 expansions in triatomines participate in detoxification in deltamethrin-resistant Triatoma infestans. Sci Rep. 2018;8:2–13.
20. Grosso CG, Blariza MJ, Mougabure-Cueto G, Picollo MI, García BA. Identification of three cytochrome P450 genes in the Chagas' disease vector Triatoma infestans: Expression analysis in deltamethrin susceptible and resistant populations. Elsevier B.V.; 2016.
21. Dulbecco AB, Moriconi DE, Lynn S, McCarthy A, Juárez MP, Girotti JR, et al. Deciphering the role of Rhodnius prolixus CYP4G genes in straight and methyl-branched hydrocarbon formation and in desiccation tolerance. Insect Mol Biol. 2020;29:431–43.
22. Paim RMM, Pessoa GCD, Nascimento BWL, Nascimento AMD, Pinheiro LC, Koerich LB, et al. Effect of salivary CYP4EM1 and CYP4EM2 gene silencing on the life span of Chagas disease vector Rhodnius prolixus (Hemiptera, Reduviidae) exposed to sublethal dose of deltamethrin. Insect Mol Biol. 2022;31:49–59.

23. Montella IR, Schama R, Valle D. The classification of esterases: An important gene family involved in insecticide resistance - A review. Mem Inst Oswaldo Cruz. 2012;107:437–49.
24. Oakeshott JG, Claudianos C, Campbell P, Newcomb RD, Russell, R J. Biochemical genetics and genomics of insect esterases. Compr Mol Insect Sci. 2005;:309–81.
25. Casida JE, Durkin KA. Neuroactive Insecticides: Targets, Selectivity, Resistance, and Secondary Effects. Annu Rev Entomol. 2013;58:99–117.
26. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Molecular Biology. 2005;14:3–8.
27. Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, et al. Identification, characterization, and crystal structure of the omega class glutathione transferases. J Biol Chem. 2000;275:24798–806.
28. Vieira FG, Sánchez-Gracia A, Rozas J. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: Purifying selection and birth-and-death evolution. Genome Biol. 2007;8.
29. Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the arthropoda: Origin and evolutionary history of the chemosensory system. Genome Biol Evol. 2011;3:476–90.
30. Ingham VA, Anthousi A, Douris V, Harding NJ, Lycett G, Morris M, et al. A sensory appendage protein protects malaria vectors from pyrethroids. Nature. 2019; May. https://doi.org/10.1038/s41586-019-1864-1.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
32. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
33. Katoh K, Kuma K, Miyata T, Toh H. Improvement in the Katoh K, Kuma K, Miyata T TH. Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Inform. 2005;16:22–33.
34. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
35. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
36. Hordijk W, Gascuel O. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. Bioinformatics. 2005;21:4338–47.
37. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27:135–45.
38. Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family size evolution in drosophila chemosensory gene families: A comparative analysis with a critical appraisal of methods. Genome Biol Evol. 2014;6:1669–82.

39. Engsontia P, Sangket U, Chotigeat W, Satasook C. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: Implications for their adaptation and speciation. J Mol Evol. 2014;79:21–39.

Abbreviations

ABC transporters
ATP-binding cassette transporters.
Ach
Acethyl cholinesterase
aLRT-SH
approximate Likelihood Ratio Test based on the Shimodaira-Hasegawa.
BLAST
Basic Local Alignment Search Tool translated.
CCEs
Carboxyl/cholinesterases.
CD-HIT
Cluster Database at High Identity with Tolerance.
CYPs
The cytochromes P450.
DD
Dietary/Detoxification class.
GFF
General Feature Format.
GSTs
Glutathione Transferases.
HPP
Hormone and Pheromone Proccessing class.
IDs
Identification.
iTOL
Interactive Tree Of Life.
i5k
Sequencing Five Thousand Arthropod Genomes.
MAFFT
Multiple Alignment using Fast Fourier Transform.
NCBI
National Center for Biotechnology Information.
ND
Neurodevelopmental class.

trimAl
a tool for automated Alignment trimming.
UFBoot
the UltraFast bootstrap.

Tables

Tables 1-2 are available in the Supplementary Files section.

Figures

Figure 1

Species tree showing the number of gene gains (green) and losses (red) in each branch for GST, CCE and CYP superfamilies.

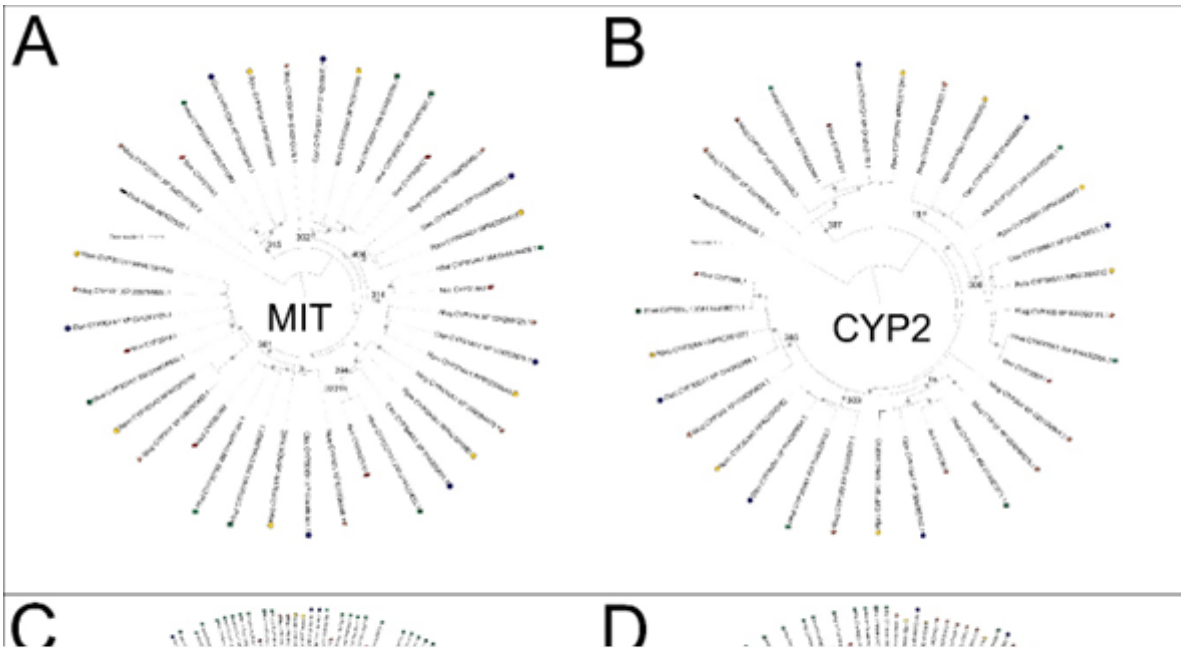

Figure 2

Phylogeny of CYP superfamily from *N. viridula* (Maroon: Nvir), *R. prolixus* (Yellow: Rpro), *H. halys* (Green: Hhal), *C. lectularius* (Blue: Clec) and *N. lugens* (Orange: Nlug) **A)** Mitochondrial Clan **B)** CYP2 Clan **C)** CYP3 Clan **D)** CYP4 Clan. A CYP gene from *Bemisia tabaci* was used as an outgroup (AEK21835.1 - NCBI).

Figure 3

Phylogeny of CCE superfamily from *N. viridula* (Maroon: Nvir), *R. prolixus* (Yellow: Rpro), *H. Halys* (Green: Hhal), *N. lugens* (Orange: Nlug) and *C. lectularius* (Blue: Clec). Cholinesterase 1 from *B. tabaci* was used

as an outgroup (XP_018913404.1 - NCBI).

Figure 4

Phylogeny of GST superfamily from *N. viridula* (Maroon: Nvir), *R. prolixus* (Yellow: Rpro), *H. Halys* (Green: Hhal), *N. lugens* (Orange: Nlug) and *C. lectularius* (Blue: Clec). A GST from *B. tabaci* was used as an outgroup (XP_018912034.1 - NCBI).

Figure 5

Cluster analysis across species and superfamilies. **A)** Total detoxification genes *vs*. detoxification genes in clusters for each species with available genome **B)** Total detoxification genes vs genes in clusters in Nlug, Clec, Rpro and Hhal for each superfamily under analysis. **C)** Sum of total genes vs sum of genes in clusters for each superfamily under analysis. **D)** Total genes vs genes in clusters for each clan/class under analysis.

Figure 6

Analysis across families with the majority of their genes in clusters on a species-by-species basis. **A)** CYP. **B)** CCE. **C)** GST.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.jpg
- SupplementaryFigure2.jpg
- SupplementaryInformation1chisquaretables.xlsx
- Supplementaryinformation2CCEs.xlsx
- Suplementaryinformation3NucleotidsequencesinN.viridula.txt
- Tables.docx