BMC Genomic Data

RESEARCH Open Access

Genome-wide phylogenetic analysis and expansion of gene families involved in detoxification in *Smittia aterrima* (Meigen) and *Smittia pratorum* (Goetghebuer) (Diptera, Chironomidae)

Bin Mao[1], Yue Zheng[1], Yunli Xiao[1], Kaixia Yang[1], Jingru Shangguan[1], Mi Shen[1], Hao Sun[1], Xiangliang Fang[1] and Yue Fu[1*]

Abstract

Smittia aterrima (Meigen, 1818) and *Smittia pratorum* (Goetghebuer, 1927) are important indicator insects for aquatic environments, showing extensive tolerance to the environment. However, the genome-wide phylogenetic relationships and characteristics of the detoxification mechanisms in *S. aterrima* and *S. pratorum* remain unclear. Based on the genomes of the two species obtained in our preliminary studies and nine genomes from the NCBI database, we found that chironomids diverged from other mosquitoes approximately 200 million years ago (MYA), and *S. aterrima* and *S. pratorum* diverged about 30 MYA according to phylogenetic analysis. Gene family evolution analysis showed significant expansion of 43 and 15 gene families in *S. aterrima* and *S. pratorum*, respectively, particularly those related to detoxification pathways. Positive selection analysis reveals that genes under positive selection are crucial for promoting environmental adaptation. Additionally, the detoxification-associated gene families including Cytochrome P450 (CYP), Glutathione S-transferases (GST), ATP-binding cassette (ABC), carboxylesterase (CCE), and UDP-glucuronosyltransferase (UGT) were annotated. Our analysis results show that these five detoxification gene families have significantly expanded in the chironomid genomes. This study highlights the genome evolution of chironomids and their responses to mechanisms of tolerance to environmental challenges.

Keywords *Smittia aterrima*, *Smittia pratorum*, Detoxification, Environmental adaptation, Phylogenetic analysis

*Correspondence:
Yue Fu
fuyue20190125@163.com
[1]Hubei Key Laboratory of Economic Forest Germplasm Improvement and Resources Comprehensive Utilization, Hubei Collaborative Innovation Center for the Characteristic Resources Exploitation of Dabie Mountains, Hubei Zhongke Research Institute of Industrial Technology, College of Biology and Agricultural Resources, Huanggang Normal University, Huanggang City 438000, Hubei, China

Introduction

The non-biting chironomid midges (Diptera: Chironomidae) represent the most diverse group of freshwater insects, with over 6,000 species known globally. with some species also thrive in terrestrial and marine environments [1]. The larvae of Chironomidae are highly diverse and abundant, inhabiting the water's bottom and serving as crucial indicators of water quality and eutrophication among benthic organisms. These larvae exhibit a highly polyphagous diet, consuming a variety of organic matter including detritus, algae, microorganisms, and macrophytes [2]. Their feeding behavior significantly contributes to the release of nitrogen and phosphorus from the substrate, thereby accelerating the mineralization of organic matter in the water. Consequently, Chironomidae larvae play an essential role in maintaining water ecological balance and are widely used in biomonitoring studies [3]. Their polyphagous nature and resilience in extreme habitats make them valuable model organisms for studying genetic and physiological mechanisms of adaptation and evolution.

Chironomidae exhibits a wide diversity of species that have uniquely adapted to thrive in extreme abiotic conditions and possess an exceptional capacity to digest substantial quantities of organic matter. Additionally, they are skilled at adapting to toxic compounds from environmental pollutants and have developed resistance to xenobiotics [4, 5]. A primary mechanism for establishing this resistance involves the use of detoxification enzymes [6]. Detoxification gene families found in various insect species illustrate a notable physiological adaptation to their complex food sources and environments [7, 8]. Numerous cytochrome P450s (P450s), glutathione-S-transferases (GSTs), and choline/carboxylesterases (CCEs) have been identified in biting mosquitoes, including Culex pipiens quinquefasciatus, with evidence of an expansion in detoxification enzymes linked to insecticide resistance [9]. Similarly, in Anopheles sinensis, 93 P450s, 31 GSTs, and 50 CCEs have been characterized, indicating an expansion of the CYP9J subfamily relative to Anopheles gambiae [10]. We have already obtained high-quality chromosome-level genomes of *S. aterrima* and *S. pratorum* [11]. Both species represent the group's successful adaptation to extreme environments and possess the smallest genomes yet documented within Chironomidae. This finding further supports the hypothesis [12] that a smaller genome size may be a factor contributing to adaptation in harsh environments.

To better understand the environmental adaptation and pollution resistance mechanisms of *S. aterrima* and *S. pratorum*, this study examines the crucial role played by major families of detoxification enzymes—including ATP-binding cassette (ABC), CCE, P450s (CYP), GSTs, and UDP-glucuronosyltransferase (UGT)—in their survival strategies. Our research focuses on elucidating the tolerance and detoxification mechanisms of these two species through the analysis and comparison of their detoxification-related gene families. Additionally, this study underscores the importance of investigating genome evolution and responses to environmental challenges to enhance our understanding of how organisms adapt to changes such as climate change or habitat destruction.

Methods

Phylogenetic analysis and divergence time estimation

Protein-coding genes from high-quality genomes of 11 representative Diptera species—Culicidae (*Anopheles arabiensis*, *Culex quinquefasciatus*), Sciaridae (*Bradysia coprophila*), Drosophilidae (*Drosophila melanogaster*), Stratiomyidae (*Hermetia illucens*), Calliphoridae (*Lucilia cuprina*), Muscidae (*Musca domestica*), and Chironomidae (*Belgica antarctica*, *Polypedilum vanderplanki*, *S. aterrima*, *S. pratorum*)—were obtained from the NCBI Genomes database for gene family orthology inference. These 11 species were clustered using OrthoFinder v2.5.2 [13], with Diamond employed as the protein sequence alignment tool to ultimately derive the orthologous gene families for each species.

The phylogeny of these species was constructed from single-copy gene protein sequences obtained using OrthoFinder. First, MAFFT v7.394 [14] was utilized to align the homologous regions with the L-INS-i strategy. Next, BMGE v1.12 [15] filtered out regions with unreliable homology, using the default parameters set to strict mode ('-m BLOSUM90 -h 0.4'). The resulting alignments were combined into a supermatrix using FASconCAT-G v1.04 [16]. Phylogenetic tree construction was performed with IQ-TREE v2.1.3 [17], employing '--symtest-remove-bad --symtest-pval 0.10' to filter genes not conforming to the SRH hypothesis prior to tree building. Protein substitution models were automatically selected by IQ-TREE, with the model type restricted to LG and using a heuristic partition search strategy ('-m MFP -mset LG -msub nuclear -rclusterf 10'). Node support values were assessed using Ultrafast Bootstrap [18] and the SH-aLRT algorithm ('-B 1000 --alrt 1000') [19].

Divergence times were estimated using MCMCTree in PAML v4.9j [20]. Key parameters for MCMCTree were set as follows: clock=2, BDparas=1 1 0.1, kappa_gamma=6 2, alpha_gamma=1 1, rgene_gamma=2 20 1, sigma2_gamma=1 10 1. Fossil calibration evidence was sourced from the Paleobiology Database (PBDB, https://paleobiodb.org/navigator/). Four fossil calibration points were utilized: Trichoptera+Lepidoptera+Diptera (root node): < 323.2 million years ago (Mya), Chironomidae: 201.3–252.17 Mya, Culicidae: 93.5–145 Mya, and Sciaroidea: 208.5–252.17 Mya.

Gene family evolution analyses

CAFÉ v4.2.1 [21] was used to detect expansions or contractions of gene families across the phylogenetic tree, utilizing a single birth-death parameter lambda (λ) and a significance level of 0.01. The analysis also concurrently evaluated an error correction model for the input data. Additionally, functional enrichment analysis (GO and KEGG) was performed using ClusterProfiler v3.10.1 [22] with default parameters ($p=0.01$, FDR=0.05), based on significantly expanded gene families.

Positive selection analysis

To estimate the rate of adaptive evolution for each gene, we conducted dN/dS analysis on all single-copy genes (1:1 orthologues) of 11 Diptera species. These single-copy genes were identified using the same OrthoFinder parameters as previously outlined. The analysis was performed using the CodeML program in the PAML v4.9e package with the branch-site model [20] to estimate the dN/dS ratio for each homologous cluster. The branch-site model was applied to detect positive selection within gene families, focusing on the common ancestor of *S. aterrima* and *S. pratorum* as the foreground clade and the other nine species as the background branches. Protein sequences of each single-copy gene were aligned using MAFFT v7.310 [14]. For individual gene analysis, we pruned the relevant species from the phylogenetic tree obtained from GoTree v0.4.2 [23], based on the IQ-TREE generated earlier. The branch-site model A (model=2, NSsites=2, fix_omega=0, omega=1.5) was compared against the null model (model=2, NSsites=2, fix_omega=1, omega=1) using the maximum likelihood phylogenetic tree as input. The likelihood ratio test (LRT) was used to evaluate the log-likelihood difference $2\Delta l = 2(l1 - l0)$ between the null (log-likelihood value l0) and alternative (log-likelihood value l1) models, following a χ^2 distribution. Genes were identified as positively selected if they had a *p*-value below 0.05 after FDR correction. Additionally, we used the aBSREL model of Hyphy v2.5.36 [24] to analyze gene evolution, considering the most recent common ancestor of *S. aterrima* and *S. pratorum* as the foreground branches and the other nine species as the background branches. Gene families with a *p*-value less than 0.05 were considered to be under positive selection.

Analysis of the detoxification-related gene families

To investigate the detoxification metabolism of *S. aterrima* and *S. pratorum*, we utilized BITACORA v1.3 [25] to annotate five gene families: Cytochrome P450 (CYP), Glutathione S-transferases (GST), ATP-binding cassette (ABC), carboxylesterase (CCE), and UDP-glucuronosyltransferase (UGT). BITACORA employed BLASTP and TBLASTN to align the protein-coding genes and the genome assembly results annotated by MAKER, respectively. HMMER v3.3 [26, 27] was then used to identify the protein domains within these gene families. Reference sequences for the GST, ABC, CCE, and UGT gene families were obtained from the NCBI GenBank database for *Culex quinquefasciatus*, *D. melanogaster*, *Culex pipiens pallens*, and *Anopheles gambiae*. Reference sequences for the CYP gene family were sourced from Dermauw et al. [28]. HMM files for CYP (PF00067), GST (PF14497 and PF02798), ABC (PF00005), CCE (PF06203), and UGT (PF00201) were downloaded from the Pfam database. To identify detoxification metabolism-related gene families, an e-value of 1e-5 and 10 was set for BLAST and HMMER searches, respectively. TBLASTN was used to search for gene families CCE, ABC, UGT, CYP, and GST, setting maximum intron lengths at 5,000 bp for CCE, ABC, and UGT, 15,000 bp for CYP, and 1,000 bp for GST. The resulting genome assembly data were aligned to predict new genes. Predicted genes were functionally validated against the non-redundant protein sequence database (nr) via online BLASTP. Gene families were classified, and phylogenetic trees were constructed to verify any erroneous sequences. To refine gene sequences, positive gene, exon, and intron boundaries were manually inspected using IGV, aided by transcriptome alignment to BAM, BRAKER de novo predictions, and MAKER predictions. Sequence evolution analysis was conducted on specific gene families, including CCE, ABC, UGT, CYP, and GST. Proteins were aligned using MAFFT v7.450 with the L-INS-i strategy, followed by trimming the alignments with trimAl v1.4.1 using the "gappyout" heuristic method to remove ambiguous homology sites. Gene trees were then constructed using IQ-TREE v2.0.7 with automatic model selection and 1,000 ultrafast bootstrap replicates. Finally, the phylogenetic tree was visualized using EvolView v3 [29].

Results and discussion

Gene orthology and comparative analysis with other genomes

Comparative genomic analyses were conducted among *S. aterrima*, *S. pratorum*, and nine other Diptera insects representing six families. In total, 147,267 genes (93.5%) were clustered into 16,283 gene families using OrthoFinder (Table 1). Amongst them, 3,232 were single-copy orthologs (families), and 2,522 were multi-copy orthologs present in all 11 insect genomes. Specifically for *S. aterrima* and *S. pratorum*, there were 7,247 gene families containing 201 species-specific genes and 7,041 gene families containing 68 species-specific genes, respectively (Table 2).

Gene family evolution, expansion, and contraction

Species phylogenetic trees were constructed using 3,232 single-copy gene protein sequences obtained by the

Table 1 Statistics of the gene families of *S. Aterrima*, *S. pratorum*, and nine other Diptera insects

Statistics	Value
Number of species	11
Number of genes	157,538
Number of genes in orthogroups	147,267
Number of unassigned genes	10,271
Percentage of genes in orthogroups	93.5
Number of orthogroups	16,283
Number of species-specific orthogroups	3,341
Number of genes in species-specific orthogroups	14,285
Percentage of genes in species-specific orthogroups	9.1
Mean orthogroup size	9
Number of orthogroups with all species present	5,754
Number of single-copy orthogroups	3,232

above-described Orthofinder. After removing 691 genes with the symtest test in IQ-TREE, the remaining 2,837 genes (1,829,410 amino acid sites) were used to construct the tree, achieving full support (100/100 for UFB/SH-aLRT) at all nodes. The phylogenetic tree indicated that *S. aterrima*, *S. pratorum*, and two additional chironomids formed a cluster, with *S. aterrima*, *S. pratorum*, and *Belgica antarctica* (Orthocladiinae) constituting a sister lineage to *Polypedilum vanderplanki* (Chironominae), consistent with previous findings [30, 31]. Approximately 200 million years ago, chironomids diverged from other mosquitoes, with the split between Orthocladiinae and Chironominae occurring around 100 million years ago, and the divergence of *S. aterrima* and *S. pratorum* taking place about 30 million years ago (Fig. 1a).

Regarding *S. aterrima*, the analysis of gene family evolution revealed that 532 gene families underwent expansion, while 388 gene families experienced contraction (Fig. 1a). Among these, 43 orthologous groups showed significant expansion ($p<0.05$), including those related to trypsin, cytochrome P450, lectin C-type domain, glucose dehydrogenase, odorant receptors, and trehalose receptors. Conversely, twelve orthologous groups were observed to undergo significant contraction (Table S1, Fig. 1b). For *S. pratorum*, the analysis of gene family evolution indicated that 257 gene families expanded, while 841 contracted (Fig. 1a). Notably, fifteen orthologous groups, including those associated with lipase 3, SET and MYND domain-containing proteins, UDP-glucosyltransferase 2, and cytochrome P450, underwent significant expansion. Similarly, thirteen orthologous groups were significantly contracted (Table S2, Fig. 1c).

Gene family GO/KEGG enrichment analyses

The GO enrichment analyses conducted on *S. aterrima* and *S. pratorum* indicate that the expanded gene families in these species are predominantly associated with metabolic processes. *S. aterrima* exhibits enrichment in steroid, hormone, xenobiotic, and terpenoid metabolism, whereas *S. pratorum* shows enrichment in functions related to the regulation of glycosylation proteins, hematopoietic stem cells, cholesterol, and fatty acid metabolism (Table S3, S4, Fig. 2a and b). These findings suggest that both species have developed mechanisms to adapt to extreme environments, with steroids playing a key role in driving various metabolic processes and contributing to physiological adaptations to local abiotic conditions [32, 33]. Additionally, mechanisms such as xenobiotic, terpenoid, cholesterol, and fatty acid metabolism enable detoxification against terpenoids and pesticides, as well as resistance to multiple adverse environments [34–37]. KEGG pathway enrichment analyses further underscore the importance of detoxification metabolism, hormones, sterols, and secondary metabolite synthesis in *S. aterrima* and *S. pratorum* (Table S5, S6, Fig. 2c and d). These metabolic processes play critical biological roles in diverse organisms. For instance, detoxification metabolism aids in the removal of harmful substances from cells, thereby promoting cellular health and survival. Hormones are essential for regulating a wide range of physiological

Table 2 Distribution of interspecific gene families

Species	1:1:1	*N: N:N*	Chironomidae	Species-specific	Others	Unassigned
Anopheles arabiensis	3,232	3,461	0	803	4,566	526
Bradysia coprophila	3,232	4,042	0	3,064	4,600	1,608
Culex quinquefasciatus	3,232	3,854	0	1,890	5,304	801
Drosophila melanogaster	3,232	3,954	0	689	4,660	1,420
Hermetia illucens	3,232	3,816	0	1,438	4,631	874
Lucilia cuprina	3,232	3,817	0	1,992	5,417	1,078
Musca domestica	3,232	4,256	0	785	5,807	808
Belgica antarctica	3,232	4,054	435	555	3,881	1,353
Polypedilum vanderplanki	3,232	4,437	431	2,800	5,685	1,278
Smittia aterrima	3,232	4,015	403	201	4,185	294
Smittia pratorum	3,232	3,809	382	68	3,528	231

Note: 1:1:1: Shared single-copy orthologs; N:N: N: Shared multi-copy orthologs; Chironomidae: Unique genes of Chironomidae; Species-specific: Species-specific genes; Others: Other orthologs; Unassigned: orthologs which cannot be assigned into any orthogroups

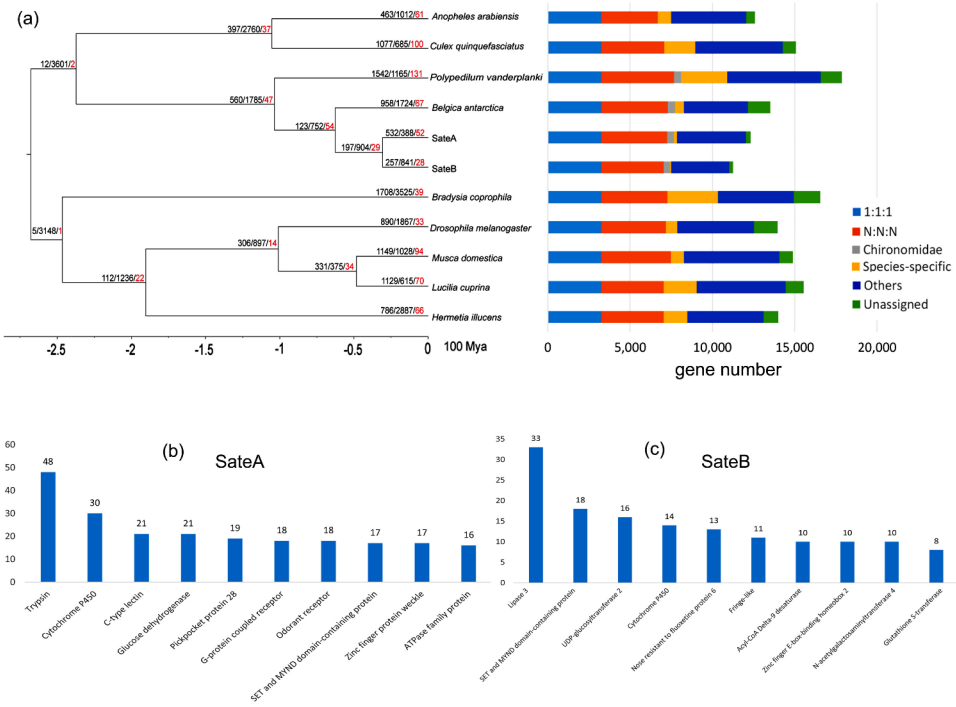

Fig. 1 Phylogeny, dating and gene family evolution of *S. aterrima* (Sate A), *S. pratorum* (Sate B) with other insect species. (**a**) the dated tree and orthologue statistics. Branch length represents divergence time (unit, 100 Mya); numbers on the branches represent gene families expanding/contracting/rapidly evolving. "1:1:1" represents shared single-copy orthologues, "N: N:N" represents multicopy orthologues shared by all species, "Chironomidae" represents orthologues unique to Chironomidae, "Species-specific" represents Species-specific genes, "others" represents unclassified orthologues and "unassigned" represents orthologues that cannot be assigned to any orthogroups. (**b**, **c**) Significantly expanded gene families (orthologue numbers ≥ 10) in SateA(**b**) and SateB (**c**)

processes, including growth, development, metabolism, and reproduction [38, 39]. Sterols are vital for maintaining cell membrane integrity, signaling pathways, and the biosynthesis of other molecules. Secondary metabolites are involved in ecological interactions and evolutionary adaptations [40]. Overall, these pathways may pave the way for adaptation to various environmental problems.

Positively selected genes in *S. aterrima* and *S. pratorum*

To gain deeper insights into the evolutionary process of expanded gene families in *S. aterrima* and *S. pratorum*, we identified 1,726 gene families that underwent positive selection using codeml analysis, while 354 gene families showed signs of positive selection based on Hyphy analysis. Of these, 172 gene families were identified as undergoing positive selection by both methods. Enrichment cluster analysis was then conducted on these positively selected gene families in both species, revealing that GO enrichment was primarily centered around nucleotide metabolism and synthesis, as well as the biosynthesis of pyridine-containing compounds known for their antimicrobial and antiviral activities [41] (Table S7, Fig. 3a). Similarly, KEGG pathway enrichment highlighted RNA transport, glycolysis, sugar synthesis, the citric acid cycle, and other metabolic pathways (Table S8, Fig. 3b). These pathways are closely intertwined with gene expression, regulation, energy production, and substance metabolism. Such mechanisms allow *S. aterrima* and *S. pratorum* to degrade and assimilate exogenous compounds, releasing their chemical energy and converting it into forms usable by cells and tissues. This metabolic flexibility enables them to withstand harsh environmental conditions, such as drought, by utilizing stored energy reserves and other metabolic substrates. Consequently, these positively selected genes are essential for facilitating environmental adaptation and provide significant insights into the underlying evolutionary processes.

Analysis of detoxification-related gene families in *S. aterrima* and *S. pratorum*

To better understand the environmental adaptation and pollution resistance mechanisms of *S. aterrima* and *S. pratorum*, this study examines the critical role played by major families of detoxification enzymes, such as ATP-binding cassette (ABC), carboxylesterase (CCE), Cytochrome P450 (CYP), Glutathione S-transferases (GST), and UDP-glucuronosyltransferase (UGT), in their survival strategies. Additionally, the chironomid species *Polypedilum vanderplanki* was selected for analysis due to its larval form's exceptional ability to withstand

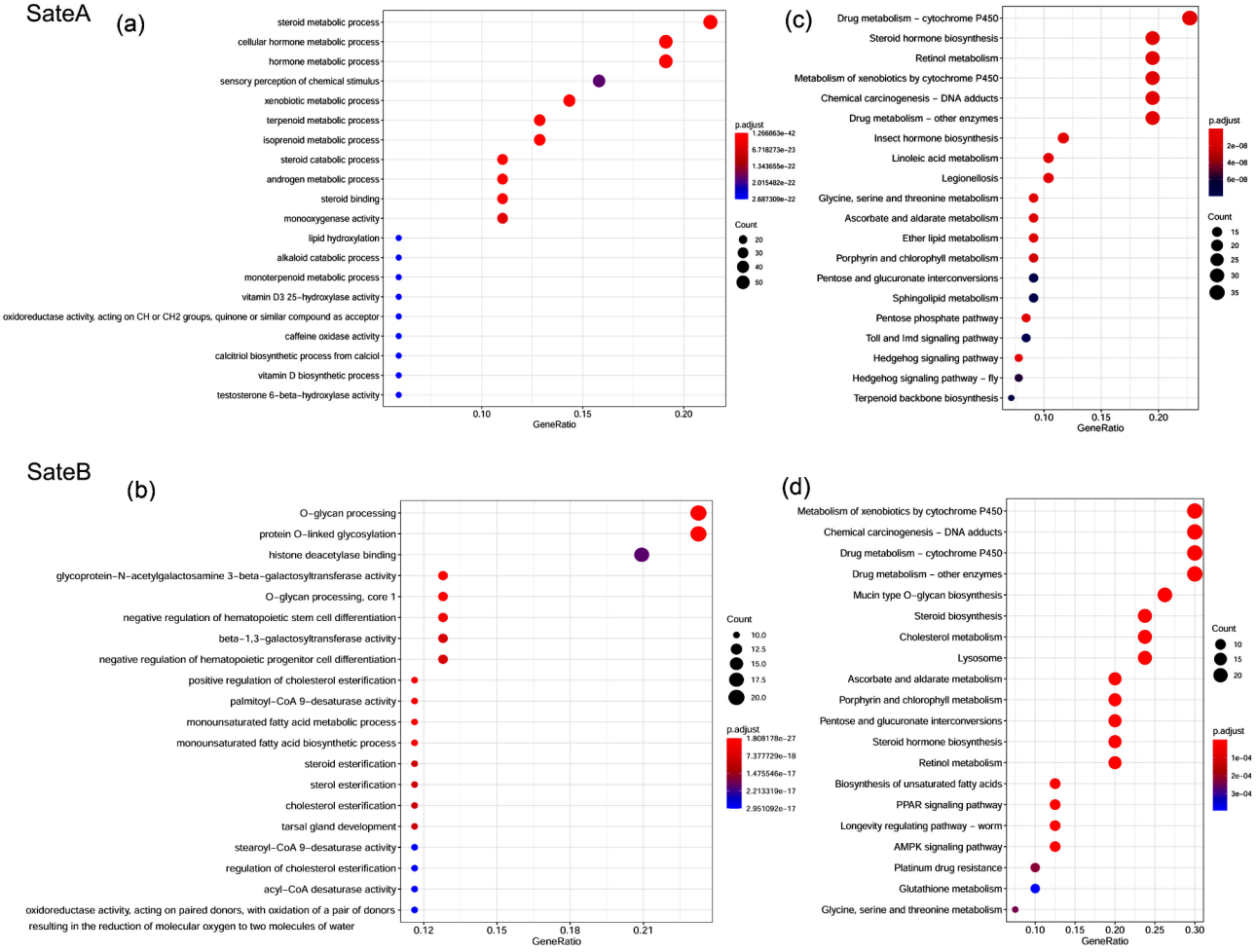

Fig. 2 (**a**, **b**) GO enrichment analyses of expanded gene families of *S. aterrima* (SateA, **a**) and *S. pratorum* (SateB, **b**). The vertical axis represents the path name and the horizontal axis represents gene ratio. The size of the dot indicates the number of differentially expressed genes in the pathway and the colour of the point corresponds to different q-value ranges. (**c**, **d**) KEGG enrichment analyses of expanded gene families of *S. aterrima* (SateA, **c**) and *S. pratorum* (SateB, **d**)

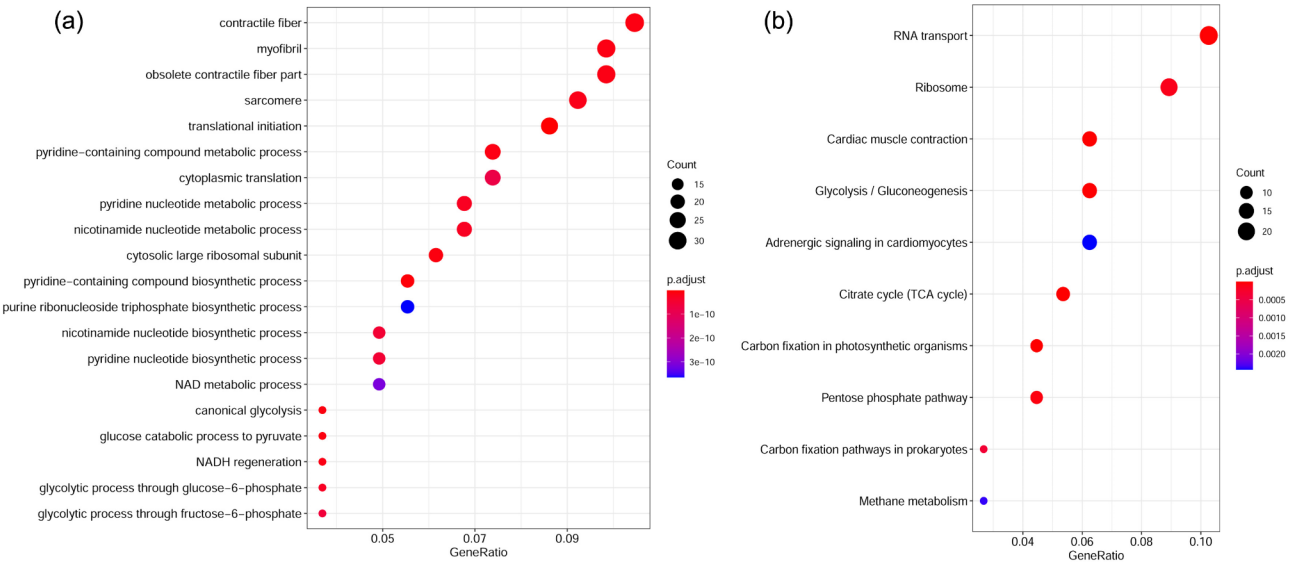

Fig. 3 (**a**, **b**) GO (**a**) and KEGG (**b**) enrichment analyses of positively selected gene families of *S. aterrima* and *S. pratorum*

near-complete water loss [42], and its similar life habits to *S. aterrima* and *S. pratorum*. Furthermore, the well-studied model insect *Drosophila melanogaster* from the Drosophilidae family [43] was included in the analysis, and the corresponding numbers of gene families for each species are presented in Table 3. Our analysis revealed expansions of five gene sub-families within the chironomid species genomes.

The results presented in this study provide novel insights into ABC proteins in chironomids, revealing their expansion and clustering into multiple gene clusters (Fig. 4). It was discovered that *S. aterrima*, *S. pratorum*, and *P. vanderplanki* possess a total of 63, 67, and 73 ABC transporter genes, respectively. By analyzing their phylogenetic relationships (as shown in Figure S1) and domain organization, the ABC gene family was assigned into eight subfamilies: ABCA, ABCB, ABCC, ABCD, ABCE, ABCF, ABCG, and ABCH. Notably, the ABCC, ABCG, and ABCH subfamilies contained the highest number of members, while the ABCD, ABCE, and ABCF subfamilies were relatively sparse across all species. To further elucidate the expansion scenario, we compared the ABC proteins of *S. aterrima* and *S. pratorum* with those of *P. vanderplanki* and *D. melanogaster* (Table 3). Consistent with previous studies [44, 45], high conservation of ABC transporters was observed in proteins involved in fundamental cellular processes. Among these are the mitochondrial half transporters of the ABCB subfamily, which are involved in iron metabolism and the transport of Fe/S protein precursors. Members of the ABCD, ABCE, and ABCF subfamilies are implicated in the transport of very long-chain fatty acids, initiation of gene transcription, and protein translation, respectively. The analysis revealed a particularly high number of gene duplications in the ABCH subfamilies, which contain 9, 9, and 19 members in *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, respectively. While the function of ABCH proteins remains unclear, this subfamily has been found to expand significantly in aquatic arthropods [44], suggesting a potential role in adaptation to water environments. Additionally, the ABCC and ABCG subfamilies showed expansion trends in chironomid species. The ABCC subfamily has multiple functions, including acting as a chloride/anion channel in epithelial cells (cystic fibrosis transmembrane conductance regulator), serving as targets for sulfonylurea antidiabetic drugs (sulfonylurea receptors), and including multidrug resistance-associated proteins (MRPs) that transport a variety of substrates such as drugs, endogenous compounds, glutathione and glucuronyl conjugates, and cyclic nucleotides [44, 46–48]. Furthermore, the ABCG subfamily has been extensively studied, with the white protein in *Drosophila* functioning as a transporter of eye pigment precursors (guanine and tryptophan) in the eye cells of the fly [49]. Additionally, the ABCG subfamily is known to contribute to the detoxification of xenobiotics [37, 50–53].

Table 3 Comparison of detoxification-related gene families among the *S. aterrima* (SateA), *S. pratorum* (SateB), *Polypedilum Vanderplanki* (Pvan) and *Drosophila melanogaster* (Dmel)

Family	Clan	SateA	SateB	Pvan	Dmel
ABC		**63**	**67**	**73**	**56**
	ABCA	8	11	19	10
	ABCB	6	5	6	8
	ABCC	15	17	16	14
	ABCD	2	2	2	2
	ABCE	1	1	1	1
	ABCF	3	3	3	3
	ABCG	19	19	17	15
	ABCH	9	9	19	3
CCE		**46**	**42**	**46**	**35**
	Dietary/Detoxification	26	22	23	13
	Hormone/Semiochemical	9	9	11	8
	Neuro-developmental	11	11	12	14
CYP		**156**	**120**	**127**	**87**
	Mitochondrial	9	10	13	12
	CYP2	12	10	7	7
	CYP3	90	70	66	36
	CYP4	45	30	41	32
GST		**52**	**54**	**61**	**38**
	Microsomal	6	7	6	3
	Delta	17	26	22	11
	Epsilon	9	7	7	14
	Omega	1	1	3	3
	Sigma	10	5	14	1
	Zeta	7	4	7	4
	Theta	1	2	1	2
	unclassified	1	2	1	0
UGT		**48**	**50**	**39**	**35**

The CCE family is primarily composed of catalytic proteins responsible for the hydrolysis of various carboxylic esters, although there are instances of non-catalytic members as well [54, 55]. These proteins also participate in the modification of insect juvenile hormones. In the genomes of *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, 46, 42, and 46 genes encoding CCEs were identified, respectively—significantly more than the number found in the *D. melanogaster* genome (Table 3). Moreover, the CCE gene family is divided into three distinct clades representing dietary/detoxification, hormone/semiochemical, and neuro/developmental functions. Phylogenetic analysis (Figure S2) revealed clear orthologous relationships for these genes among the three chironomid species. The dietary/detoxification class is notably expanded in chironomids, with *S. aterrima*, *S. pratorum*, and *P. vanderplanki* possessing 26, 22, and 23 genes, respectively, compared to only 13 in *D. melanogaster*. This expansion suggests a potential role in the detoxification of dietary or allelochemical compounds.

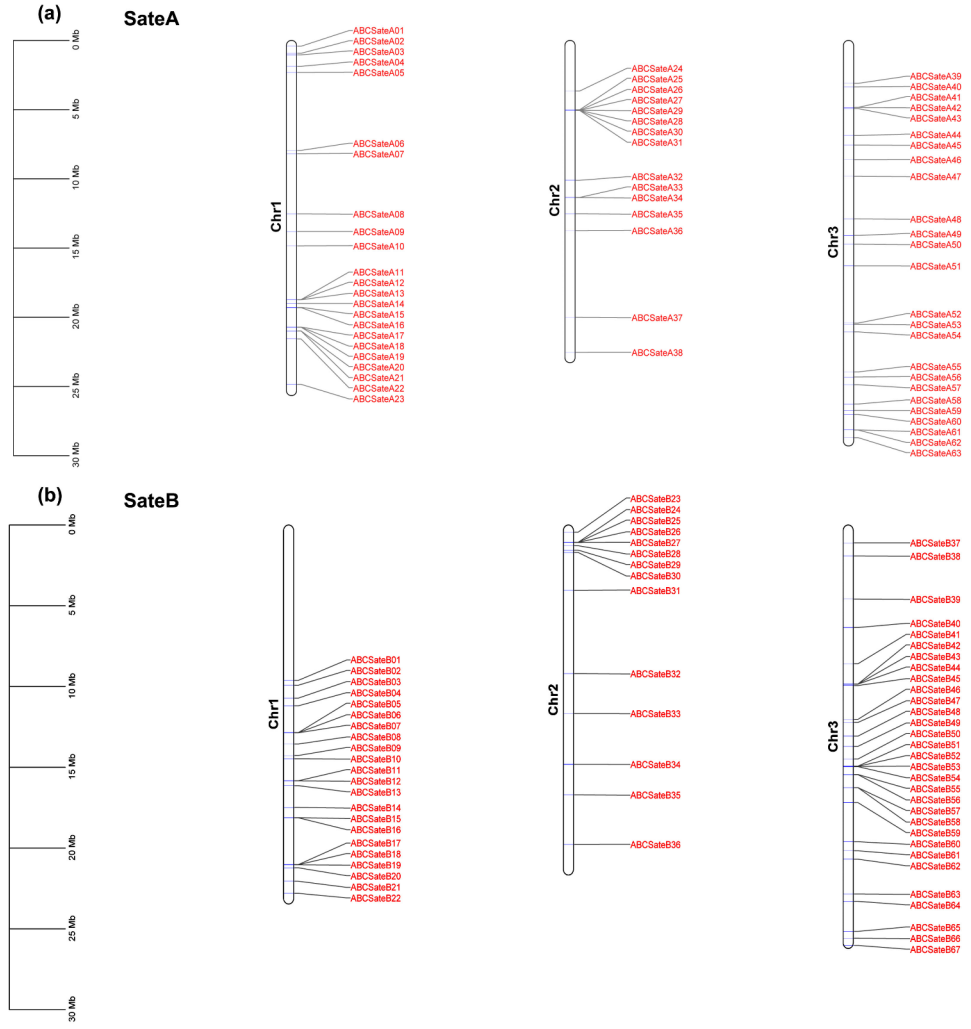

Fig. 4 The genomic positions of ABC genes. The predicted 63 (**a**) and 67 (**b**) ABC genes were mapped to the genome of *S. aterrima* (SateA) and *S. pratorum* (SateB)

The hormone/semiochemical class primarily consists of secreted enzymes and is associated with neonicotinoid resistance, reproduction, odor processing, and possible gut or salivary detoxification. Finally, the neuro/developmental class is conserved among insects and plays a role in the hydrolysis of the neurotransmitter acetylcholine [55–57].

The CYP gene family, widely distributed in eukaryotic genomes, can be categorized into four clans: Mitochondrial P450, CYP2, CYP3, and CYP4. These clans play a pivotal role in metabolic processes involving natural products, xenobiotics, and endogenous compounds [58], particularly in insects, which utilize a suite of mitochondrial P450 paralogues to adapt to environmental pressures [8, 59]. In the complete genomic sequences of *S. aterrima* and *S. pratorum*, we identified a total of 156 and 120 CYP genes, respectively. These genes can be divided into the four main clans, with 9 and 10 genes in the mitochondrial P450 clan, 12 and 10 genes in the CYP2 clan, 90 and 70 genes in the CYP3 clan, and 45 and 30 genes in the CYP4 clan. Compared to *D. melanogaster*, chironomids exhibit a significant expansion in the CYP gene family, particularly in the CYP2 clan, where *S. aterrima* and *S. pratorum* possess 12 and 10 genes, respectively. The CYP2 gene family is crucial for the metabolism of drugs and environmental chemicals and is expressed in both vertebrates and invertebrates. These enzymes are essential for detoxification and influence a range of physiological and toxicological processes [8, 60, 61]. The CYP3 clan genes show substantial expansion in all three chironomid species, with 90, 70, and 66 genes found in *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, respectively, compared to the 36 genes found in *D. melanogaster*. These genes, which are prevalent among insect P450 genes, have been shown to be associated with xenobiotic metabolism and insecticide resistance when induced by phenobarbital, pesticides, or natural products [8]. Meanwhile, the CYP4 clan genes also underwent significant expansion, with

45, 30, and 41 genes in *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, respectively. Although insect CYP4 genes are numerous in insect genomes, certain CYP4 genes can be induced by xenobiotics as metabolizers, while others are linked to odorant or pheromone metabolism [8]. (Table 3; Fig. 5).

The GST gene family is responsible for regulating essential traits related to toxin metabolism in insects, comprising two main types: microsomal and cytosolic [7, 62–65]. The results provide new insights into the presence of GST proteins in chironomids, highlighting the identification of a total of 52, 54, and 61 GST genes in *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, respectively. These numbers represent a notable increase compared to the 38 GST genes present in *D. melanogaster*. Phylogenetic analysis (Fig. 6) indicates that the GST gene family can be classified into seven subfamilies: Microsomal, Delta, Epsilon, Omega, Sigma, Zeta, and Theta. Of these,

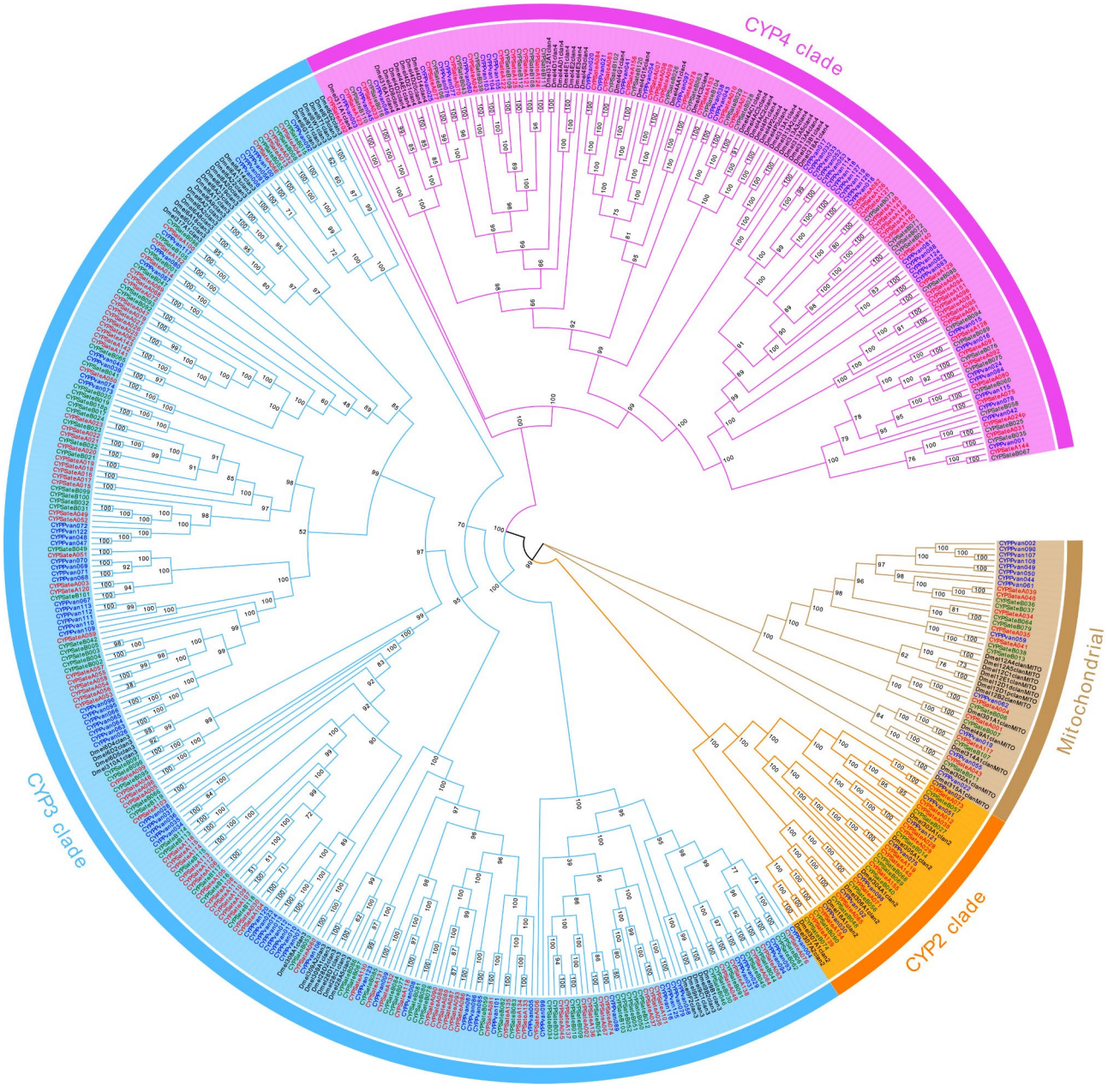

Fig. 5 Phylogenetic tree based on the CYP gene family of *Smittia aterrima* (SateA in red), *Smittia pratorum* (SateB in green), *Polypedilum vanderplanki* (Pvan in blue), and *Drosophila melanogaster* (Dmel in black). Colored lines correspond to branches with species using the same colors as for the CYP names. Numbers on the tree correspond to bootstrap values. CYP clans were colored brown for the mitochondrial clan, orange for clan2, blue for clan3 and pink for clan4

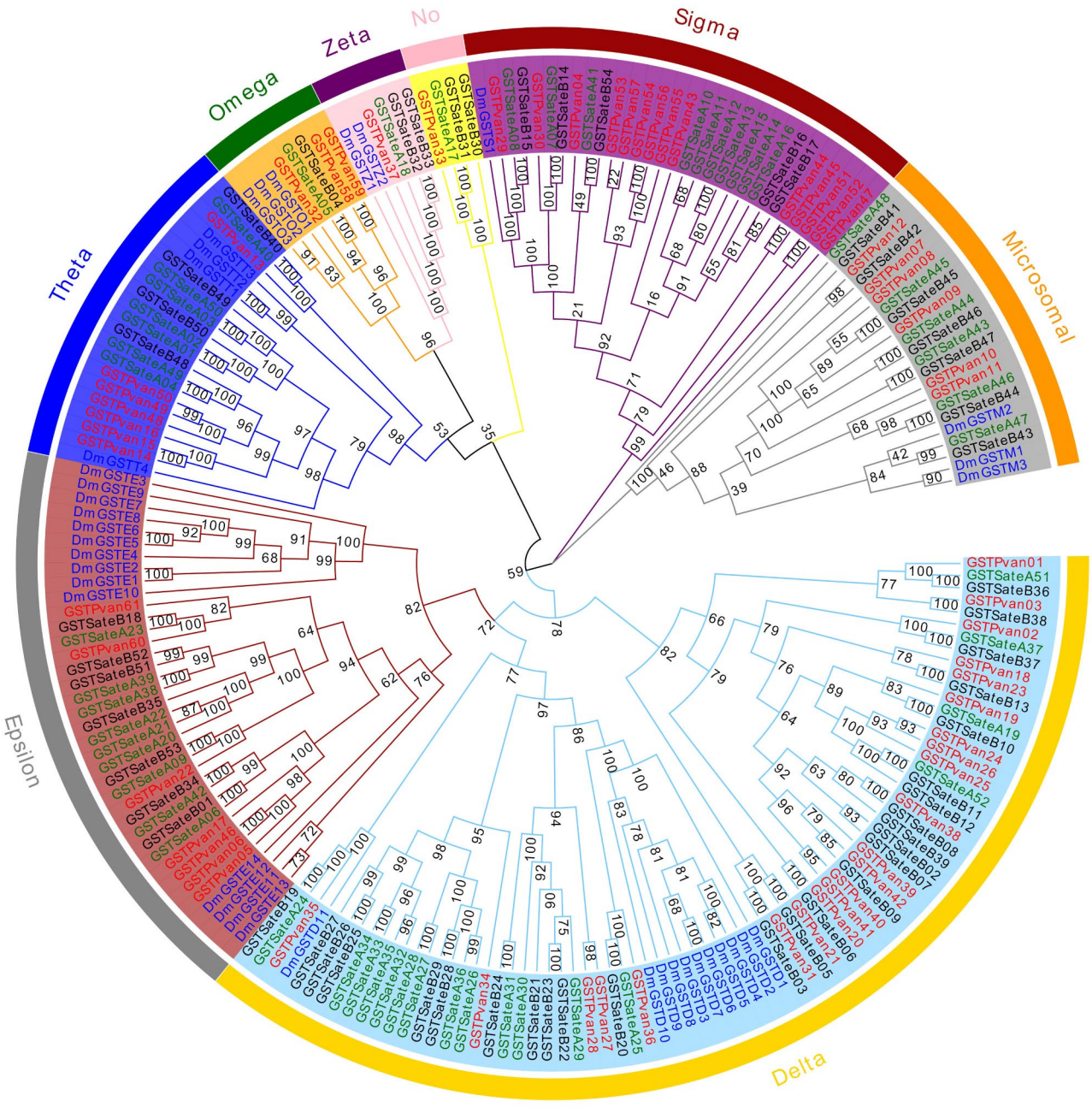

Fig. 6 Phylogenetic tree based on the GST gene family of *Smittia aterrima* (SateA, green name), *Smittia pratorum* (SateB, black name), *Polypedilum vanderplanki* (Pvan, red name), and *Drosophila melanogaster* (Dm, blue name). Bootstrap values are indicated at the nodes. The No branch is an unclassified gene

the Microsomal subfamily represents the microsomal type, while the other subfamilies belong to the cytosolic type. Further analysis of gene duplicates within these subfamilies reveals that members within the Microsomal, Delta, and Sigma subfamilies have significantly expanded in chironomids, whereas there was no expansion within the Epsilon, Omega, and Theta subfamilies when compared to *Drosophila* (Table 3). The microsomal class is considered crucial for protecting cells against oxidative damage and xenobiotics. Through our analysis, three microsomal GST genes were detected in the *D. melanogaster* genome, while *S. aterrima*, *S. pratorum*, and *P. vanderplanki* had 6, 7, and 6 microsomal GST genes, respectively. Among the genomes currently available, dipterans are characterized by a significant expansion of cytosolic GSTs. For example, *D. melanogaster* was found to have 35 cytosolic GSTs, while the three analyzed chironomids had 45, 45, and 54, respectively, compared to other insects with fewer than 20 GSTs [56, 66, 67]. The majority of the GST expansions are found in the Delta

and Sigma subfamilies. *S. aterrima*, *S. pratorum*, and *P. vanderplanki* exhibited 17, 26, and 22 genes in the Delta subfamily, respectively, compared to only 11 in *D. melanogaster*. This suggests that these enzymes play a significant role in the adaptation to their specific environments. The Sigma subfamily contains 10, 5, and 14 genes in *S. aterrima*, *S. pratorum*, and *P. vanderplanki*, respectively, with only 1 in *D. melanogaster*. These GSTs are critical in protecting against oxidative stress [68, 69].

UDP-glycosyltransferases (UGTs) are essential metabolic enzymes found widely distributed in all free-living organisms. They are responsible for facilitating the conjugation of a sugar donated by a UDP-glycoside to typically lipophilic molecules. This catalytic function aids in the formation of more hydrophilic compounds, thereby optimizing the degradation and excretion processes [70, 71]. In chironomids, UGT genes exhibit a significant expansion compared to *Drosophila melanogaster*. Specifically, *S. aterrima*, *S. pratorum*, and *P. vanderplanki* contain 48, 50, and 39 UGT genes, respectively, while *D. melanogaster* has 35 (Table 3). This expansion underscores the importance of these enzymes in safeguarding cellular systems against harmful xenobiotics and promoting efficient regulation of antibiotics.

Conclusions

This study elucidates the evolutionary history and genomic adaptations of *S. aterrima* and *S. pratorum*, emphasizing the importance of detoxification pathways in their environmental tolerance. The significant expansion of these detoxification gene families in *S. aterrima* and *S. pratorum* not only illuminates their genomic evolution but also highlights the molecular basis for their environmental resilience. These findings provide valuable insights into the genetic adaptations that enable these species to thrive in challenging aquatic environments and underscore their potential utility as bioindicators for pollution monitoring.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-024-01289-9.

Supplementary Figure 1: Phylogenetic tree based on the ABC gene family of *S. aterrima* (SateA), *S. pratorum* (SateB), *P. vanderplanki* (Pvan), and *D. melanogaster* (Dm, green name). Bootstrap values are indicated at the nodes

Supplementary Figure 2: Phylogenetic tree based on the CCE gene family of *S. aterrima* (SateA), *S. pratorum* (SateB), *P. vanderplanki* (Pvan), and *D. melanogaster* (CG, green name). Bootstrap values are indicated at the nodes

Supplementary Tables: **Supplementary Table S1**: Twenty most significantly expanded gene families of *S. aterrima*. **Supplementary Table S2**: Ten most significantly expanded gene families of *S. pratorum*. **Supplementary Table S3**: GO enrichment of expanded gene families in the *S. aterrima* genome. **Supplementary Table S4**: GO enrichment of expanded gene families in the *S. pratorum* genome. **Supplementary Table S5**: KEGG enrichment of expanded gene families in the *S. aterrima* genome. **Supplementary Table S6**: KEGG enrichment of expanded gene families in the *S. pratorum* genome. **Supplementary Table S7**: GO enrichment of gene families experiencing positive selection in the *S. aterrima* and *S. pratorum* genomes. **Supplementary Table S8**: KEGG enrichment of gene families experiencing positive selection in the *S. aterrima* and *S. pratorum* genomes

Acknowledgements
We acknowledge financial support from the Open Access Publication Fund of Huanggang Normal University (Hubei, China).

Author contributions
Yue Fu and Bin Mao conceived the study and wrote the manuscript, Yue Zheng and Xiangliang Fang collected the samples. Bin Mao, Mi Shen, Kaixia Yang, Jinru Shangguan, and Hao Sun performed bioinformatics analysis. Yunli Xiao put forward suggestions for modification. All authors read and approved the final manuscript.

Funding
This study was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 32070483, 31460572), Natural Science Foundation of Hubei Province (Grant No. 2020CFB757), Scientific Research Starting Foundation for Ph.D. of Huanggang Normal University (Grant No. 2020010, 2042024378), Scientific Research Project of Education Department of Hubei Province (Grant No. Q20232901), Huanggang Normal University Major Project Incubation Plan-Funded Program (Grant No. 204202315204).

Data availability
The data does not involve sequencing and are listed in the article.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 October 2024 / Accepted: 10 December 2024
Published online: 18 December 2024

References

1. McLachlan A, Armitage PD, Cranston PS, Pinder LCV. The Chironomidae: the biology and ecology of non-biting midges. J Anim Ecol. 1995;64:667.
2. Hirabayashi K, Wotton RS. Organic matter processing by chironomid larvae (Diptera: Chironomidae). Hydrobiologia. 1998;382:151–9.
3. Sæther OA. Chironomid communities as water quality indicators. Ecography. 1979;2:65–74.
4. Pinder LCV. Biology of Freshwater Chironomidae. Annu Rev Entomol. 1986;31:1–23.
5. Henriques-Oliveira AL, Nessimian JL, Dorvillé LF. Feeding habits of chironomid larvae (Insecta: Diptera) from a stream in the Floresta Da Tijuca, Rio De Janeiro, Brazil. Braz J Biol. 2003;63(2):269–81.
6. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R. Evolution of supergene families associated with insecticide resistance. Science. 2002;298(5591):179–81.
7. Ranson H, Hemingway J. Glutathione Transferases. In: *Comprehensive Molecular Insect Science.* Edited by Gilbert LI. Amsterdam: Elsevier; 2005: 383–402.
8. Feyereisen R. Evolution of insect P450. Biochem Soc Trans. 2006;34(Pt 6):1252–5.
9. Reddy BN, Rao BP, Prasad G, Raghavendra K. Identification and classification of detoxification enzymes from *Culex quinquefasciatus* (Diptera: Culicidae). Bioinformation. 2012;8(9):430–6.
10. Zhou D, Liu X, Sun Y, Ma L, Shen B, Zhu C. Genomic Analysis of Detoxification Supergene Families in the Mosquito *Anopheles sinensis*. PLoS ONE. 2015;10(11):e0143387.

11. Fu Y, Fang X, Xiao Y, Mao B, Xu Z, Shen M, Wang X. Two chromosome-level genomes of Smittia aterrima and Smittia pratorum (Diptera, Chironomidae). Sci Data. 2024;11(1):165.
12. Cornette R, Gusev O, Nakahara Y, Shimura S, Kikawada T, Okuda T. Chironomid midges (Diptera, chironomidae) show extremely small genome sizes. Zoolog Sci. 2015;32(3):248–54.
13. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
14. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
15. Criscuolo A, Gribaldo S. BMGE (Block Mapping and gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210.
16. Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool. 2014;11(1):81.
17. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
18. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188–95.
19. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.
20. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
21. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013;30(8):1987–97.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
23. Lemoine F, Gascuel O. Gotree/Goalign: toolkit and go API to facilitate the development of phylogenetic workflows. NAR Genom Bioinform. 2021;3(3):lqab075.
24. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676–9.
25. Vizueta J, Sánchez-Gracia A, Rozas J. Bitacora: a comprehensive tool for the identification and annotation of gene families in genome assemblies. Mol Ecol Resour. 2020;20(5):1445–52.
26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.
27. Eddy SR. Accelerated Profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.
28. Dermauw W, Van Leeuwen T, Feyereisen R. Diversity and evolution of the P450 family in arthropods. Insect Biochem Mol Biol. 2020;127:103490.
29. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res. 2019;47(W1):W270–5.
30. Cranston PS, Hardy NB, Morse GE. A dated molecular phylogeny for the Chironomidae (Diptera). Syst Entomol. 2012;37(1):172–88.
31. Sæther OA. Phylogeny of the subfamilies of Chironomidae (Diptera). Syst Entomol 2000, 25.
32. Behmer ST, Nes WD. Insect sterol Nutrition and Physiology: A Global Overview. Adv Insect Physiol. 2003;31:1–72.
33. Svoboda JA, Thompson MJ, Robbins WE, Kaplanis JN. Insect steroid metabolism. Lipids. 1978;13(10):742–53.
34. Hartman JH, Widmayer SJ, Bergemann CM, King DE, Morton KS, Romersi RF, Jameson LE, Leung MCK, Andersen EC, Taubert S, et al. Xenobiotic metabolism and transport in Caenorhabditis elegans. J Toxicol Environ Health B Crit Rev. 2021;24(2):51–94.
35. Ikbal C, Ben HK, Ben HM. Insect growth regulator activity of Cestrum parqui saponins: an interaction with cholesterol metabolism. Commun Agric Appl Biol Sci. 2006;71(2 Pt B):489–96.
36. Souza ROO, Damasceno FS, Marsiccobetre S, Biran M, Murata G, Curi R, Bringaud F, Silber AM. Fatty acid oxidation participates in resistance to nutrient-depleted environments in the insect stages of Trypanosoma Cruzi. PLoS Pathog. 2021;17(4):e1009495.
37. Guo Z, Kang S, Zhu X, Xia J, Wu Q, Wang S, Xie W, Zhang Y. Down-regulation of a novel ABC transporter gene (Pxwhite) is associated with Cry1Ac resistance in the diamondback moth, Plutella xylostella (L). Insect Biochem Mol Biol. 2015;59:30–40.
38. Nijhout HF. Insect hormones. Princeton University Press; 1994.
39. Wu K, Li S, Wang J, Ni Y, Huang W, Liu Q, Ling E. Peptide hormones in the Insect Midgut. Front Physiol. 2020;11:191.
40. Prasad KVH. Insect–Plant Interactions and Role of Secondary Metabolites. In: *Insect Ecology: Concepts to Management*. Edited by Prasad KVH. Singapore: Springer Nature Singapore; 2022: 37–60.
41. Marinescu M, Popa CV. Pyridine compounds with antimicrobial and antiviral activities. Int J Mol Sci 2022, 23(10).
42. Shaikhutdinov N, Gusev O. Chironomid midges (Diptera) provide insights into genome evolution in extreme environments. Curr Opin Insect Sci. 2022;49:101–7.
43. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD. E: the genome sequence of Drosophila melanogaster. Science. 2000;287(5461):2185–95.
44. Sturm A, Cunningham P, Dean M. The ABC transporter gene family of Daphnia pulex. BMC Genomics. 2009;10:170.
45. Lu H, Xu Y, Cui F. Phylogenetic analysis of the ATP-binding cassette transporter family in three mosquito species. Pestic Biochem Physiol. 2016;132:118–24.
46. Kruh GD, Belinsky MG. The MRP family of drug efflux pumps. Oncogene. 2003;22(47):7537–52.
47. Moreau C, Prost AL, Dérand R, Vivaudou M. SUR, ABC proteins targeted by KATP channel openers. J Mol Cell Cardiol. 2005;38(6):951–63.
48. Meng X, Clews J, Kargas V, Wang X, Ford RC. The cystic fibrosis transmembrane conductance regulator (CFTR) and its stability. Cell Mol Life Sci. 2017;74(1):23–38.
49. Chen H, Rossier C, Lalioti MD, Lynn A, Chakravarti A, Perrin G, Antonarakis SE. Cloning of the cDNA for a human homologue of the Drosophila white gene and mapping to chromosome 21q22.3. Am J Hum Genet. 1996;59(1):66–75.
50. Kusuhara H, Sugiyama Y. ATP-binding cassette, subfamily G (ABCG family). Pflugers Arch. 2007;453(5):735–44.
51. Epis S, Porretta D, Mastrantonio V, Urbanelli S, Sassera D, De Marco L, Mereghetti V, Montagna M, Ricci I, Favia G, et al. Temporal dynamics of the ABC transporter response to insecticide treatment: insights from the malaria vector Anopheles Stephensi. Sci Rep. 2014;4:7435.
52. He C, Liang J, Liu S, Wang S, Wu Q, Xie W, Zhang Y. Changes in the expression of four ABC transporter genes in response to imidacloprid in Bemisia tabaci Q (Hemiptera: Aleyrodidae). Pestic Biochem Physiol. 2019;153:136–43.
53. Yang J, Chen S, Xu X, Lin G, Lin S, Bai J, Song Q, You M, Xie M. Novel-miR-310 mediated response mechanism to Cry1Ac protoxin in Plutella xylostella (L). Int J Biol Macromol. 2022;219:587–96.
54. Oakeshott JG, Devonshire AL, Claudianos C, Sutherland TD, Horne I, Campbell PM, Ollis DL, Russell RJ. Comparing the organophosphorus and carbamate insecticide resistance mutations in cholin- and carboxyl-esterases. Chem Biol Interact. 2005;157–158:269–75.
55. Oakeshott JG, Johnson RM, Berenbaum MR, Ranson H, Cristino AS, Claudianos C. Metabolic enzymes associated with xenobiotic and chemosensory responses in Nasonia Vitripennis. Insect Mol Biol. 2010;19(Suppl 1):147–63.
56. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyereisen R, Oakeshott JG. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol Biol. 2006;15(5):615–36.
57. Yu QY, Lu C, Li WL, Xiang ZH, Zhang Z. Annotation and expression of carboxylesterases in the silkworm, Bombyx mori. BMC Genomics. 2009;10:553.
58. Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol. 1999;29(9):757–77.
59. Lewis D. COMPACT: a structural approach to the modelling of cytochromes P450 and their interactions with xenobiotics. J Chem Technol Biot. 2001;76:237–44.
60. Waterman MR. Introduction: transcription and regulation of activities of cytochromes P450 metabolizing endogenous substrates. Faseb j. 1996;10(13):1455.
61. Kubota A, Stegeman JJ, Goldstone JV, Nelson DR, Kim EY, Tanabe S, Iwata H. Cytochrome P450 CYP2 genes in the common cormorant: evolutionary relationships with 130 diapsid CYP2 clan sequences and chemical effects on their expression. Comp Biochem Physiol C Toxicol Pharmacol. 2011;153(3):280–9.
62. Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. Crit Rev Biochem Mol Biol. 1995;30(6):445–600.

63. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Mol Biol. 2005;14(1):3–8.
64. Friedman R. Genomic organization of the glutathione S-transferase family in insects. Mol Phylogenet Evol. 2011;61(3):924–32.
65. Perry T, Batterham P, Daborn PJ. The biology of insecticidal activity and resistance. Insect Biochem Mol Biol. 2011;41(7):411–22.
66. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA. E: genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 2010;107(27):12168–73.
67. Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KH, Ponnala L, Macmil SL, Roe BA, Jander G. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol. 2010;19(Suppl 2):155–64.
68. Franciosa H, Bergé JB. Glutathione S-transferases in housefly (Musca domestica): location of GST-1 and GST-2 families. Insect Biochem Mol Biol. 1995;25(3):311–7.
69. Singh SP, Coronella JA, Benes H, Cochrane BJ, Zimniak P. Catalytic function of Drosophila melanogaster glutathione S-transferase DmGSTS1-1 (GST-2) in conjugation of lipid peroxidation end products. Eur J Biochem. 2001;268(10):2912–23.
70. Bock KW. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. Biochem Pharmacol. 2003;66(5):691–6.
71. Ahn SJ, Vogel H, Heckel DG. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. Insect Biochem Mol Biol. 2012;42(2):133–47.

Publisher's note