

Research Article

Evolution of the Heme Peroxidases of Culicidae (Diptera)

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, 715 Sumter Street, Coker Life Sciences Building, Columbia, SC 29208, USA

Correspondence should be addressed to Austin L. Hughes, austin@biol.sc.edu

Received 17 August 2012; Accepted 17 October 2012

Academic Editor: Zainulabeuddin Syed

Copyright © 2012 Austin L. Hughes. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Phylogenetic analysis of heme peroxidases (HPXs) of Culicidae and other insects revealed six highly conserved ancient HPX lineages, each of which originated by gene duplication prior to the most recent common ancestor (MRCA) of Hemimetabola and Holmetabola. In addition, culicid HPX7 and HPX12 arose by gene duplication after the MRCA of Culicidae and Drosophilidae, while HPX2 orthologs were not found in any other order analyzed except Diptera. Within Diptera, HPX2, HPX7, and HPX12 were relatively poorly conserved at the amino acid level in comparison to the six ancient lineages. The genome of *Anopheles gambiae* included genes encoding five proteins (HPX10, HPX11, HPX13, HXP14, and HPX15) without orthologs in other genomes analyzed. Overall, gene expression patterns did not seem to reflect phylogenetic relationships, but genes that evolved rapidly at the amino acid sequence level tended to have divergent expression patterns as well. The uniquely high level of duplication of HPXs in *A. gambiae* may have played a role in coevolution with malaria parasites.

1. Introduction

The production of nitric oxide (NO) is an important immune defense mechanism against cellular microorganisms in insects and other invertebrates [1]. Nitric oxide synthase (NOS), encoded by a single gene in insect genomes sequenced to date [2], is the major enzyme involved in NO production, but the full pathway of NO production is only beginning to be understood. In the mosquito *Anopheles gambiae* (Diptera: Culicidae) when infected by malaria parasites (Apicomplexa: *Plasmodium*), a heme peroxidase (HPX2) and NADPH oxidase (NOX5) were found to play a crucial role in potentiating NO in antiparasite defense [3].

NOX5 is represented by a single ortholog in insects, but HPX2 is a member of a multigene heme peroxidase family in *A. gambiae* and other insects [3–5]. Other mosquito heme peroxidases (HPXs) of known function include those expressed in the salivary glands of female *A. gambiae* and *A. albimanus* [6, 7] and one involved in the catalysis of protein-crosslinking in the chorion of *Aedes aegypti* eggs [8]. In another member of Diptera, the fruit fly *Drosophila melanogaster*, heme peroxidases have been implicated in chorion assembly and other developmental processes [9, 10].

For nearly 5000 orthologous genes, Waterhouse and colleagues [4] compared the amino acid sequence distance between *D. melanogaster* and *A. gambiae* with that between *D. melanogaster* and *Ae. aegypti*. On average genes with known immune function showed a greater level of amino acid sequence divergence than other genes, but certain immune-related proteins were well conserved [4]. HPXs included in these analyses likewise included a number of more conserved and less conserved proteins.

Here I use a phylogenetic analysis in order to establish orthologous sets of HPX genes in Culicidae. Unlike previously published analyses of insect HPXs [4, 5], the present phylogenetic analysis includes HPXs from Hemimetabola, making it possible to time gene duplication events in the HPX family relative to the time of the most recent common ancestor (TMRCA) of Hemimetabola and Holometabola (including Diptera). Using sets of orthologs established by the phylogenetic analysis, I analyze the patterns of amino acid sequence conservation of HPXs within and between two families of Diptera: (1) Culicidae (mosquitoes), represented by three species with completely sequenced genomes (*A. gambiae*, *Ae. aegypti*, and *Culex quinquefasciatus*) and (2) Drosophilidae (fruit flies), represented by *D. melanogaster*

and *D. grimshawi*. In addition, using data from the MozAtlas *A. gambiae* gene expression database [11], I compare across-tissue expression patterns with phylogenetic relationships and with patterns of amino acid sequence conservation.

2. Methods

2.1. Sequences and Alignment. The phylogenetic analysis was based on 81 heme peroxidase (HPX) sequences from 10 insect species with completely or nearly completely sequenced genomes, belonging to five orders: (1) the pea aphid *Acyrtosiphon pisum* (order Hemiptera); (2) the body louse *Pediculus humanus* (order Phthiraptera); (3) the red flour beetle *Tribolium confusum* (order Coleoptera); (4) the jewel wasp *Nasonia vitripennis* (order Hymenoptera); (5) the honeybee *Apis mellifera* (Hymenoptera); (6) the yellow fever mosquito *Aedes aegypti* (Diptera); (7) the African malaria mosquito *Anopheles gambiae* (Diptera); (8) the Southern house mosquito *Culex quinquefasciatus* (Diptera); (9) the fruit fly *Drosophila melanogaster* (Diptera); (10) the Hawaiian fruit fly *Drosophila grimshawi* (Diptera). Of the five orders represented, two (Hemiptera and Phthiraptera) belong to Hemimetabola, while the rest belong to Holometabola. Preliminary analyses including species from the order Lepidoptera and additional species from Hymenoptera produced results similar to those reported (not shown). However, since the main focus of the present analyses was on HPXs of Diptera, only the above 11 species were used in the final analysis for ease of presentation. Nomenclature of individual proteins and their putative orthologs was based on names provided by VectorBase (<http://www.vectorbase.org/>) for *A. gambiae*, where those were available.

Amino acid sequences were aligned by the CLUSTAL W algorithm in the MEGA 5 program [12]. In all comparisons among aligned sequences, amino acid sites at which the alignment postulated a gap in any sequence were excluded from all comparisons. Following Waterhouse and colleagues [4], the DBLOX protein of *A. gambiae*, which has an internally duplicated structure, was divided into N-terminal and C-terminal segments, each of which was aligned separately with the other HPXs. The same approach was applied to 8 additional sequences related to DBLOX, which were found to share the same internally duplicated structure. The phylogenetic analysis was thus applied to 90 aligned sequences, including both N-terminal and C-terminal segments of the 9 internally duplicated sequences.

Phylogenetic analyses were conducted using the MEGA program, version 5.05 [12], by the maximum likelihood (ML) method. The Model test function in MEGA was used to choose models by the Bayes Information Criterion (BIC). The reliability of branching patterns was tested by bootstrapping (1000 samples). Relative rate tests were conducted by Tajima's [13] method in MEGA.

Expression data for *A. gambiae* were downloaded from the MozAtlas database [11]. Expression data for the following tissue types were included separately for adult males and adult (blood-fed) females: (1) the head, including the brain;

(2) the salivary gland; (2) the midgut; (3) the Malpighian tubules; (4) the thorax excluding the gut, Malpighian tubules, and the gonads. In addition, there were data for the following sex-specific tissues: (1) ovary in females; (2) testis in males; (3) accessory glands in males. For each of these tissues, the mean expression value from MozAtlas was standardized with gene, and the genes were clustered by the McQuitty algorithm using the Manhattan distance in Minitab version 15.0 (<http://www.minitab.com/>).

3. Results

3.1. Phylogenetic Analyses. A phylogenetic tree of insect HPXs was constructed by the ML method using the WAG+G+I model at 301 aligned amino acid sites (Figure 1). In the tree there were seven major clusters each defining a lineage which included one or more sequences from Diptera and one or more sequences from Hemimetabola; these seven clusters are here designated HPX1, HPX4, HPX5, HPX6, DBLOX-N, DBLOX-C, and CG42331 (Figure 1). The first four lineages were named according to the nomenclature used in VectorBase for the *Anopheles gambiae* sequence. The DBLOX-N and DBLOX-C clusters were named, respectively, for the N-terminal and C-terminal portions of *A. gambiae* DBLOX (Figure 1). The CG42331 cluster was named for the *Drosophila melanogaster* sequenced included in that cluster (Figure 1). Each of the seven clusters was supported by an internal branch that received at least 86% bootstrap support; and five of the seven clusters received 99% percent or greater bootstrap support (Figure 1).

Each of the seven major clusters included, in addition to sequences from the three mosquito species, apparent orthologs from the two *Drosophila* species (Figure 1). Furthermore, each of the seven clusters included at least one sequence from Hemimetabola (*Acyrtosiphon pisum* or *Pediculus humanus*); this topology supported the hypothesis that each of these seven major groups of HPX orthologs arose by gene duplications that occurred prior to the most recent common ancestor (MRCA) of Hemimetabola and Holometabola. Since each of the two portions of DBLOX formed a cluster including sequences from Hemimetabola (Figure 1), the phylogenetic analysis supported the hypothesis that the internal duplication of DBLOX occurred prior to the MRCA of Hemimetabola and Holometabola.

Outside the seven major clusters, *A. gambiae* HPX sequences showed very different patterns of relatedness. These patterns were examined further by an additional phylogenetic analysis of the 32 sequences corresponding to the subtree of the original phylogeny (Figure 2) that included the HPX1 cluster and the other sequences that fell outside the seven major clusters (Figure 2). The phylogeny was constructed on the basis of the WAG+G+I model at 402 aligned sites; and the topology of the phylogenetic tree (Figure 2) was broadly similar to that of the corresponding portion of the original tree (Figure 1).

In both trees, *A. gambiae* HPX2 formed a cluster with apparent orthologs from the other mosquito species and from *Drosophila* (Figures 1 and 2). However, there were

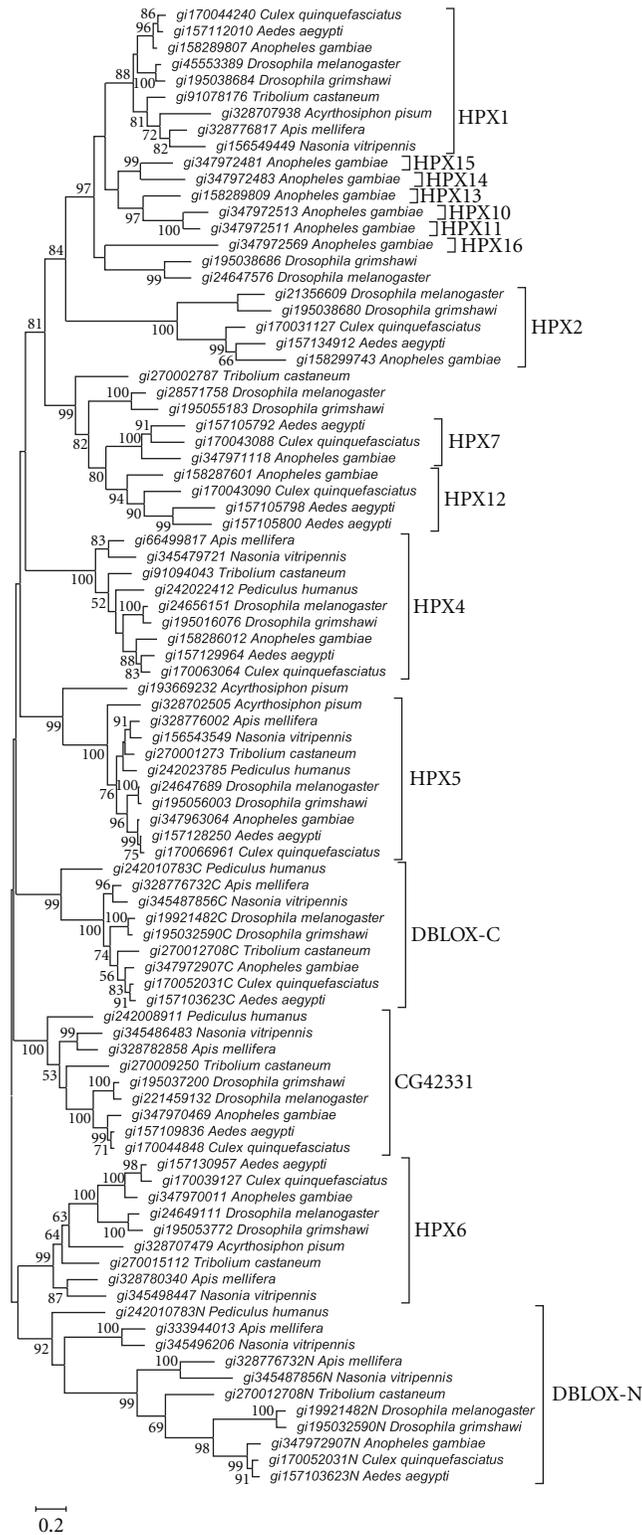


FIGURE 1: Maximum likelihood tree of insect heme peroxidases, based on the WAG+G+I model at 301 aligned amino acid sites. Numbers on the branches represent the percentage of 1000 bootstrap pseudosamples supporting the branch; only values $\geq 50\%$ are shown.

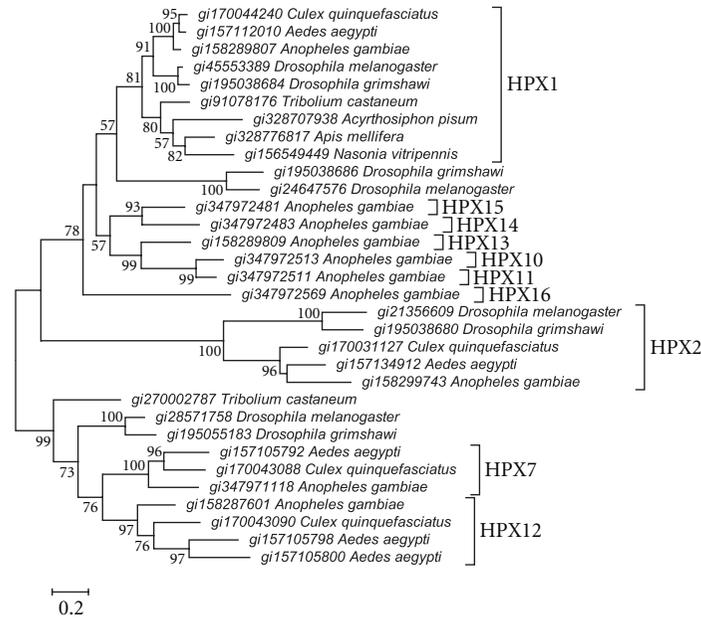


FIGURE 2: Maximum likelihood tree of select insect heme peroxidases, based on the WAG+G+I model at 402 aligned amino acid sites. Numbers on the branches represent the percentage of 1000 bootstrap pseudosamples supporting the branch; only values $\geq 50\%$ are shown.

no apparent orthologs of HPX2 from outside of Diptera (Figures 1 and 2). In both trees, there was a cluster including *A. gambiae* HPX10, HPX11, HPX13, HPX14, and HPX15, with no apparent orthologs from any of the other genomes analyzed (Figures 1 and 2). Bootstrap support for the latter cluster was weak in both trees (Figures 1 and 2). However, there was high bootstrap support for the clustering of *A. gambiae* HPX10, HPX11, and HPX13 (99%; Figure 2); and for clustering of *A. gambiae* HPX10 and HPX11 (99%, Figure 2). There was moderately high support for clustering of *A. gambiae* HPX4 and HPX15 (93%, Figure 2).

Culicid HPX7 and HPX12 clustered together with sequences from the two *Drosophila* species and one sequence from *Tribolium castaneum*; the latter cluster received 99% bootstrap support in both phylogenetic analyses (Figures 1 and 2). Because the *Drosophila* sequence fell outside the cluster of culicid HPX7 and HPX12 in both phylogenetic analyses (Figures 1 and 2), the topology is consistent with the hypothesis that the gene duplication giving rise to culicid HPX7 and HPX12 occurred after the MRCA of Culicidae and Drosophilidae.

3.2. Sequence Conservation. In order to obtain evidence regarding conservation of amino acid sequences in different HPXs, aligned amino acid sets were compared in 10 HPX orthologous groups represented in both Culicidae and Drosophilidae (Table 1). Comparisons were made among the three mosquito species, between the two *Drosophila* species, and between the two families (Table 1). HPX5 showed the strongest conservation in all comparisons (Table 1). HPX2 was much less conserved; in fact, of the 9 other orthologs, 6 were significantly more conserved than HPX2 in all comparisons (Table 1). On the other hand, HPX12, unique

to Culicidae (Figure 1), was significantly less conserved than HPX2 in comparisons within Culicidae (Table 1). HPX7 was significantly less conserved than HPX2 in the comparison between *Ae. aegypti* and *C. quinquefasciatus*, and HPX7 was less conserved than HPX2, though not significantly so, in the comparison among the three culicid species (Table 1). CG42331 did not differ significantly from HPX2 in the level of conservation within Culicidae, but CG42331 was significantly more conserved than HPX2 within Drosophilidae and in both families (Table 1).

For each of the orthologous groups listed in Table 1, relative rate tests were used to compare the rates of amino acid evolution in *A. gambiae* sequences with that in *Ae. aegypti* orthologs, using *D. melanogaster* orthologs as an outgroup. In no case was there a significant rate difference between *A. gambiae* and *Ae. aegypti* (not shown). Similar analyses likewise showed no significant differences in the rate of amino acid sequence evolution between *A. gambiae* and *C. quinquefasciatus*.

3.3. Expression Pattern. When the 15 *A. gambiae* genes were clustered on the basis of adult male and female expression pattern (Figure 3), clustering showed little relationship with phylogenetic relationships (Figures 1 and 2). For example, the closely related pair of genes HPX14 and HPX15 did not show very similar expression patterns, nor did the closely related pair of genes HPX10 and HPX11 (Figure 3). Likewise, the closely related pair HPX2 and HPX7 did not show similar expression patterns (Figure 3). A cluster of genes very similar in terms of expression pattern (HPX1, DBLOX, HPX14, HPX, and HPX10; Figure 3) shared low levels of expression across all tissues analyzed. By contrast, CG42331 had the most divergent expression pattern (Figure 3), with

TABLE 1: Percentages of amino acid residues in heme peroxidases that are conserved among sets of dipteran taxa.

Protein	No. aligned residues	Set of taxa compared			
		<i>Ae. aegypti</i> and <i>C. quinquefasciatus</i>	<i>Ae. aegypti</i> , <i>C. quinquefasciatus</i> , and <i>A. gambiae</i>	<i>D. melanogaster</i> and <i>D. grimshawi</i>	Culicidae and Drosophilidae
HPX1	672	87.5%***	79.8%***	85.3%***	56.3%***
HPX2	609	68.1%	50.6%	61.9%	24.8%
HPX4	868	85.5%***	74.3%***	88.1%***	54.0%***
HPX5	746	93.4%***	91.4%***	96.0%***	78.8%***
HPX6	800	82.0%*	64.9%***	78.8%***	41.4%***
HPX7	654	58.3%**	46.0%	67.1% ¹	18.7% ¹
HPX12	654	41.4%***	35.5%***		
CG42331	1132	74.9%	54.8%	72.1%***	34.0%**
DBLOX-N	585	86.8%***	72.5%***	83.8%***	39.5%***
DBLOX-C	775	92.5%***	86.5%***	92.3%***	66.3%***

Chi-square tests of the hypothesis that the proportion conserved is equal to that for HPX2 in the same set of taxa (Bonferroni-corrected): * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

¹Homologous to both HPX7 and HPX12 (see Figure 1).



FIGURE 3: Hierarchical clustering of gene expression of 15 *Anopheles gambiae* HPXs across 11 tissue types. The scale on the left represents expression pattern similarity score.

particularly high levels in both male and female thorax. Next most divergent were HPX2 (with high levels in male thorax and testis and in the head in both sexes) and HPX6 (with high levels in the head in both sexes).

4. Discussion

Phylogenetic analysis of the heme peroxidases (HPXs) of Culicidae showed support for the existence of six separate lineages that originated by gene duplication events

that occurred prior to the MRCA of Hemimetabola and Holometabola, which is believed to have occurred in the late Carboniferous period, 318–300 million years ago [14]: HPX1, HPX4, HPX5, HPX6, CG42331, and DBLOX. Likewise the results supported the hypothesis that the internal duplication of DBLOX occurred before the MRCA of Hemimetabola and Holometabola.

In addition to these ancient HPX lineages, culicid genomes include certain members of the HPX family that were apparently duplicated more recently than the six ancient lineages. Culicid HPX7 and HPX12 clustered with a *Tribolium castaneum* sequence, implying an origin before the MRCA of Diptera and Coleoptera. The duplication giving rise to HPX7 and HPX12 appeared to have occurred independently in the culicid lineage after the MRCA of Culicidae and Drosophilidae. However, since HPX7 and HPX12 were found in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*, the duplication occurred prior to the MRCA of those three genera.

Neither of the two available genomes of Hemimetabola was found to contain orthologs of HPX7, HPX12, HPX2, HPX10, HPX11, HPX13, HPX14, HPX15, or HPX16. The absence of these genes from Hemimetabola suggests the possibility that these genes have arisen in the Holometabola after the MRCA of Hemimetabola and Holometabola. However, it may be that future analysis of additional genomes of Hemimetabola will discover orthologs of these genes. Therefore the conclusion that they are unique to Holometabola remains tentative.

HPX2, which is known to be involved in resistance to malaria parasite infection in *A. gambiae* [3], did not have orthologs in any of the other orders included in the present analysis. In addition, the genome of *A. gambiae* included six HPXs (HPX10, HPX11, HPX13, HPX14, HPX15, and HPX16) with no reported orthologs in any other species. Thus, extensive HPX duplication appears to be a unique trait of *A. gambiae* among the mosquito genomes analyzed.

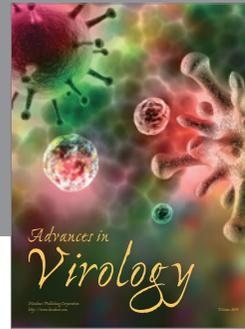
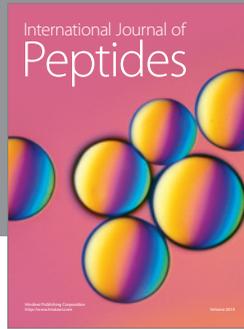
Members of the six ancient families generally showed a greater level of amino acid sequence conservation in Diptera than did sets of orthologs not found outside Holometabola (HPX2, HPX7, and HPX12). Of the latter, HPX2 was somewhat more conserved in Culicidae than was either HPX7 or HPX12. Because of the presence of relatively poorly conserved HPXs and the unique paralogs, the HPX family of Culicidae and of *A. gambiae* in particular seem to have undergone a degree of rapid evolution unusual for insect HPXs. By contrast, CG42331 was the one ancient lineage that was relatively nonconserved within Culicidae.

Overall, gene expression patterns did not reflect phylogenetic relationships, with closely related gene pairs often showing marked differences in expression pattern. On the other hand, there was some association between lack of amino acid sequence conservation and divergent patterns of gene expression (Table 1 and Figure 3). For example, CG42331, HPX2, and HPX12 represented relatively unconserved genes whose patterns of gene expression were dissimilar to those of other HPX genes, as indicated by the fact that these three genes clustered apart from the other genes in the cluster analysis of gene expression patterns (Figure 3).

It might be proposed that the evolutionary pattern of HPXs reflects coevolution of the insect host with parasitic microorganisms, particularly malaria parasites in the case of *A. gambiae*. However, there was no evidence of an increased rate of amino acid evolution of *A. gambiae* HPXs in comparison to orthologs in the other culicid species. Therefore coevolution with malaria parasites seems not to have enhanced the rate of amino acid replacement of the HPXs of *A. gambiae*. On the other hand, the extensive duplication of HPXs, unique in the present data to *A. gambiae*, may have played a role in coevolution with malaria parasites. An increased understanding of the functions of individual HPXs, including more detailed information on tissue expression, particularly in response to infection by malaria parasites, will provide evidence to test the latter hypothesis.

References

- [1] A. Rivero, "Nitric oxide: an antiparasitic molecule of invertebrates," *Trends in Parasitology*, vol. 22, no. 5, pp. 219–225, 2006.
- [2] N. M. Gerardo, B. Altincicek, C. Anselme et al., "Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*," *Genome Biology*, vol. 11, no. 2, article r21, 2010.
- [3] G. de A. Oliveira, J. Lieberman, and C. Barillas-Mury, "Epithelial nitration by a peroxidase/NOX5 system mediates mosquito antiplasmodial immunity," *Science*, vol. 335, pp. 856–859, 2012.
- [4] R. M. Waterhouse, E. V. Kriventseva, S. Meister et al., "Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes," *Science*, vol. 316, no. 5832, pp. 1738–1743, 2007.
- [5] G.-Q. Shi, Q.-Y. Yu, and Z. Zhang, "Annotation and evolution of the antioxidant genes in the silkworm, *Bombyx mori*," *Archives of Insect Biochemistry and Physiology*, vol. 79, no. 2, pp. 87–103, 2012.
- [6] I. M. B. Francischetti, J. G. Valenzuela, V. M. Pham, M. K. Garfield, and J. M. C. Ribeiro, "Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*," *Journal of Experimental Biology*, vol. 205, no. 16, pp. 2429–2451, 2002.
- [7] J. M. C. Ribeiro and J. G. Valenzuela, "Purification and cloning of the salivary peroxidase/catechol oxidase of the mosquito *Anopheles albimanus*," *Journal of Experimental Biology*, vol. 202, no. 7, pp. 809–816, 1999.
- [8] J. Li, S. R. Kim, and J. Li, "Molecular characterization of a novel peroxidase involved in *Aedes aegypti* chorion protein crosslinking," *Insect Biochemistry and Molecular Biology*, vol. 34, no. 11, pp. 1195–1203, 2004.
- [9] M. Fakhouri, M. Elalayli, D. Sherling et al., "Minor proteins and enzymes of the *Drosophila* eggshell matrix," *Developmental Biology*, vol. 293, no. 1, pp. 127–141, 2006.
- [10] R. E. Nelson, L. I. Fessler, Y. Takagi et al., "Peroxidasin: a novel enzyme-matrix protein of *Drosophila* development," *EMBO Journal*, vol. 13, no. 15, pp. 3438–3447, 1994.
- [11] D. A. Baker, T. Nolan, B. Fischer, A. Pinder, A. Crisanti, and S. Russell, "A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*," *BMC Genomics*, vol. 12, article 296, 2011.
- [12] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, pp. 2731–2739, 2011.
- [13] F. Tajima, "Simple methods for testing the molecular evolutionary clock hypothesis," *Genetics*, vol. 135, no. 2, pp. 599–607, 1993.
- [14] B. M. Wiegmann, M. D. Trautwein, J. W. Kim et al., "Single-copy nuclear genes resolve the phylogeny of the holometabolous insects," *BMC Biology*, vol. 7, article 34, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

