Correlated Evolution among Six Gene Families in *Drosophila* Revealed by Parallel Change of Gene Numbers

Dong-Dong Wu[1], David M. Irwin[1,2,3], and Ya-Ping Zhang*[,1,4]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario, Canada

[3]Banting and Best Diabetes Centre, University of Toronto, Ontario, Canada

[4]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China

*Corresponding author: E-mail: zhangyp@mail.kiz.ac.cn.

Accepted: 8 April 2011

Abstract

Proteins involved in a pathway are likely to evolve in a correlated fashion, and coevolving gene families tend to undergo complementary gains and losses. Accordingly, gene copy numbers (i.e., repertoire size) tend to show parallel changes during the evolution of coevolving gene families. To test and verify this hypothesis, here we describe positive correlations among the repertoire sizes of six gene families, that is, trypsin-like serine protease, odorant-binding protein, odorant receptor, gustatory receptor, cytochrome P450, and glutathione S-transferase after excluding the possibility of phylogenetic constraint and random drift. The observed correlations are indicative of parallel changes in the repertoire sizes of the six gene families that are due to similar demands for the quantity of these different genes in different lineages of *Drosophila*. In conclusion, we propose that the correlated evolution among these six gene families in *Drosophila* is a signature of a parallel response to ecological adaptation.

Key words: correlated evolution, Drosophila, gene family.

Correlated evolution is commonly observed when traits are functionally related. For example, comparative genomic and phylogenetic studies, for example, phylogenetic profiling (Pellegrini et al. 1999), have indicated that correlated evolution is commonly observed among different proteins that function in a pathway (Pazos and Valencia 2008). Functionally related genes often show similar responses to evolutionary pressures, functional specificities, and phylogenetic tree topologies (Fryxell 1996; Pazos and Valencia 2008).

Coevolving gene families having related functions tend to undergo complementary gains and losses (Fryxell 1996). Accordingly, gene copy numbers (i.e., repertoire size) may show parallel changes during the evolution of coevolved gene families. To test this hypothesis, here, we illustrate changes in the repertoire sizes of six gene families: odorant-binding protein (*OBP*), odorant receptor (*OR*), gustatory receptor (*GR*), trypsin-like serine proteases (*Tryp_SPc*), cytochrome P450 (*CYP450*), and glutathione S-transferase (*GST*) in 12 Drosophila genomes (Clark et al. 2007) and analyze the pattern of changes during the evolution of Drosophila. The principal reason for choosing these six gene families are that many proteins in these families have related functions. Many Tryp_SPc have direct roles in the digestion of food (Rawlings and Barrett 1994; Wu et al. 2009). CYP450 and GSTs are two classes of the major enzymes responsible for the detoxification of toxic compounds contained in or produced from food (Tijet et al. 2001; Ranson et al. 2002; Low et al. 2007; Chung et al. 2009). The above three gene families have related functions in the digestion and processing of food. ORs, GRs, and OBPs play roles in chemosensory perception, a process important in the finding and identification of good (edible) food and in the avoidance of poisonous food (Nei et al. 2008). Chemosensory information during digestion also plays an important role in the regulation of various aspects of gastrointestinal functions, such as the secretory activity of gastrointestinal glands, resorptive activity, motility and blood supply of the intestinal tract, and satiation (Hofer et al. 1999). Chemical stimulants in the intestinal lumen can stimulate neural afferent pathways, especially the intestinal vagal sensory afferent fibers and increase the release of gastrointestinal hormones from enteroendocrine cells in the intestinal epithelium (Hofer et al. 1999). The sizes of these six gene families are also very large in Drosophila thus can evolve dynamically yielding changes

in repertoire size that can easily be detected. Therefore, we hypothesize that the sizes of these gene families may evolve in a correlated fashion.

Positive Correlations among the Repertoire Sizes of These Gene Families

To conduct the analysis, we determined the sizes of the *Tryp_SPc* and *CYP450* genes repertoires by using Blast to search the genomes of the 12 *Drosophila* species followed by gene prediction and refinement (see Materials and Methods in the supplementary materials, Supplementary Material online). In addition, we used the gene repertoire sizes that were determined for the *GST*, *OR*, *OBP*, and *GR* gene families from previous studies (Low et al. 2007; Vieira et al. 2007; Gardiner et al. 2008; Nei et al. 2008). These analyses showed that *Drosophila* genomes contain on average approximately 240, 50, 60, 65, 35, and 90 members for the *Tryp_SPc*, *OBP*, *GR*, *OR*, *GST*, and *CYP450* gene families, respectively (supplementary table S1, Supplementary Material online) and that these gene numbers were variable (fig. 1). When we focused on only intact (i.e., potentially functional) genes, positive correlations in number of genes in the six gene families were found in the 12 *Drosophila* species (table 1, fig. 1), that is, a species which had a gene family with a large size tended to have larger sizes for all of its other gene families. This observation suggests that parallel changes in the quantity and, thus potentially the demand for, products of each of these different gene families occurred during the evolution of *Drosophila*. When gene family size was considered in a pairwise manner, two pairs (*Tryp_SPc-CYP450* and *Tryp_SPc-OR*) failed to show a clear significant correlation ($P = 0.186$ and $P = 0.177$), four pairs showed a correlation that almost showed statistical significance (*Tryp_SPc-GR* [$P = 0.053$], *OBP-CYP450* [$P = 0.071$], *GST-CYP450* [$P = 0.064$], *OBP-OR* [$P = 0.083$]), whereas the remaining nine pairs showed statistically significant positive correlations in gene family size, that is, *OBP-Tryp_SPc* ($P = 0.001$), *GR-OBP* ($P = 0.025$), *GR-OR* ($P = 4.62 \times 10^{-5}$), *GST-Tryp_SPc* ($P = 0.006$), *GST-OBP* ($P = 0.008$), *GST-GR* ($P = 0.036$), *GST-OR* ($P = 0.025$), *CYP450-GR* ($P = 0.002$), *CYP450-OR* ($P = 0.001$) (table 1, fig. 1). After using a false discovery rate controlling procedure for multiple testing, eight pairs still show significant correlations: *Tryp_SPc-OBP*, *Tryp_SPc-GST*, *OBP-GR*, *OBP-GST*, *GR-OR*, *CYP450-GR*, *OR-GST*, and *CYP450-OR*. Using a Bonferroni correction for multiple testing, a more conservative method in which the *P* values are multiplied by the number of comparisons, four pairs continued to show significant correlation: *Tryp_SPc-OBP*, *GR-OR*, *CYP450-GR*, and *CYP450-OR*.

An association of traits across species could suggest a common evolutionary force. However, due to phylogenetic constraints, closely related species should be more similar to each other than to more distantly related species. Therefore, we evaluated the contribution of phylogenetic inertia to the evolution of the sizes of the six gene families using a series of phylogenetic comparative methods. First, we used Moran's autocorrelation index I (Gittleman and Kot 1990). The size of only the *Tryp_SPc* gene family showed evidence of phylogenetic autocorrelation ($P = 0.033$) (supplementary table S2, Supplementary Material online). The phylogenetic dependency of the *Tryp_SPc* gene family was also supported by the phylogenetic eigenvector regression method ($P = 0.012$) (Diniz-Fi et al. 1998). We also employed four complementary tests from orthogram to diagnose the phylogenetic dependency (Ollier et al. 2006), and the statistics generated from these tests support only a slight role for phylogenetic history and suggest evolutionary independence among these gene families (supplementary table S3, Supplementary Material online). Furthermore, to correct for any bias introduced by phylogenetic inertia, we conducted a phylogenetic-independent contrast analysis to deduce the values of the "contrasts," which are statistically independent, for the six gene families, and found that 8 of the 15 pairs of gene families retained significant correlation in their gene family sizes, that is, *OBP-Tryp_SPc* ($P = 0.031$), *GR-Tryp_SPc* ($P = 0.028$), *GR-OR* ($P = 1.47 \times 10^{-4}$), *Tryp_SPc-GST* ($P = 0.049$), *GST-GR* ($P = 0.014$), *GST-OR* ($P = 0.042$), *CYP450-GR* ($P = 0.025$), *CYP450-OR* ($P = 0.016$) (table 2).

To determine whether the observed positive correlation in gene family size could simply be caused by random changes in the sizes of the gene families, we conducted a genome-wide analysis of the correlation of the sizes of each of the six gene families with the sizes of other gene families found in *Drosophila* genomes. Hahn et al. (2007) had previously described the gene families that exist in the 12 near complete *Drosophila* genomes, and for our analysis, we used 149 of these gene families that have one or more gene in each of the genomes and five or more genes in at least one species. We then computed the Pearson correlation coefficients between the sizes of each pair of gene families in the 12 *Drosophila* and used this as an empirical data set. When we examine the correlation coefficients between our six gene families (*Tryp_SPc*, *OBP*, *OR*, *GR*, *CYP450*, and *GST*), we found that 6 of 15 pairwise coefficients (*OBP-Tryp_SPc*, *GR-OR*, *Tryp_SPc-GST*, *GST-OBP*, *CYP450-GR*, and *CYP450-OR*) were higher than the 95th percentile rank value of the empirical data (which is 0.682), significantly more than that expected by random from the empirical data ($\chi^2 = 38.23$, $P = 6.30 \times 10^{-10}$, degrees of freedom = 1). However, the gene families in Hahn et al. (2007) were assembled by a modified reciprocal BlastP method using the annotated protein sequences of *Drosophila*, which would result in the loss of many genes especially in large gene families. To further address the issue as to whether random evolutionary process could produce the correlations of repertoire

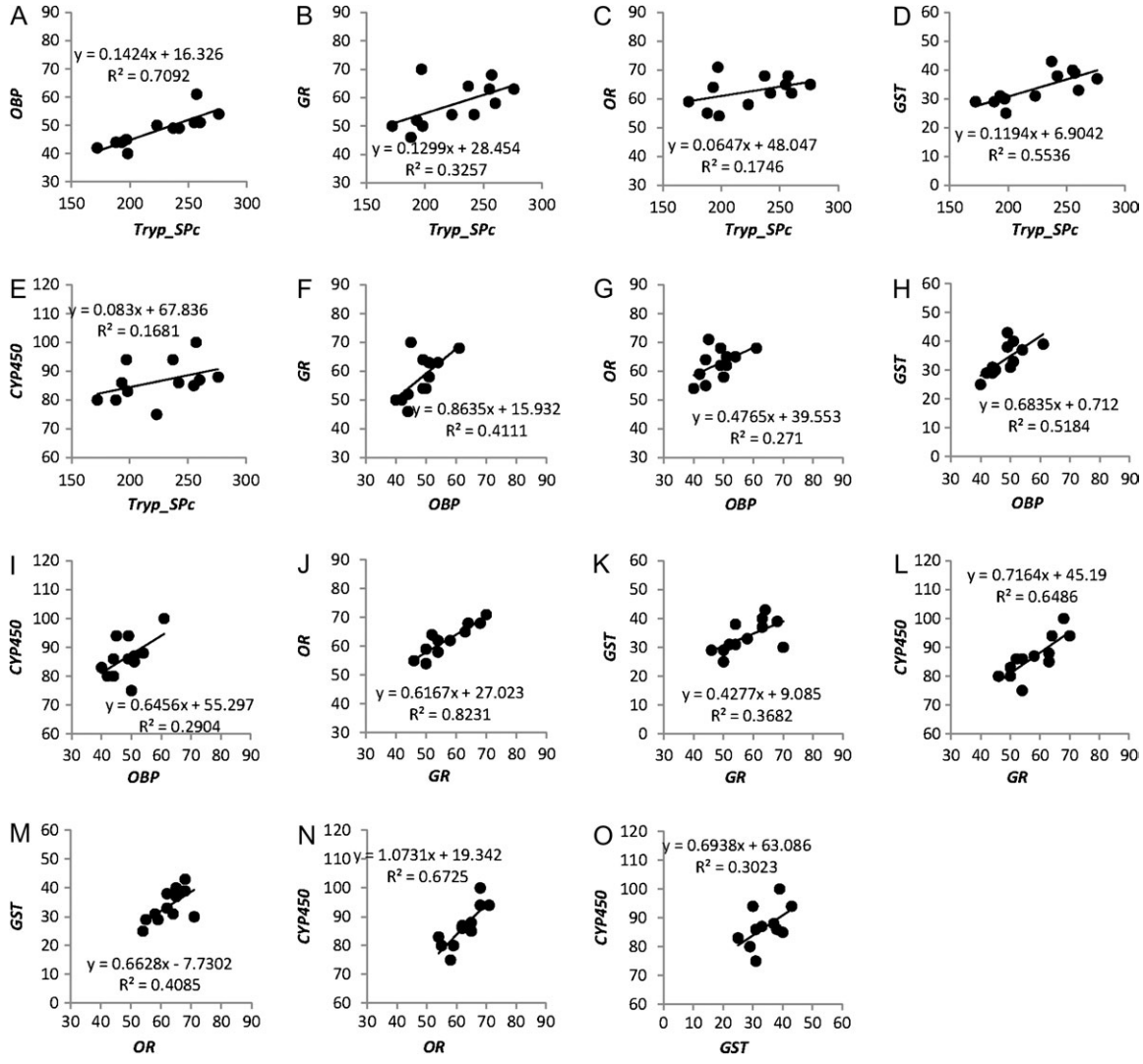

Fig. 1.—Correlation of gene family sizes for six gene families in 12 *Drosophila* species. (*A*–*O*) are linear-regression plots of intact gene numbers of each of 15 pairs of gene families.

sizes among the six gene families, we simulated the gene numbers in the 12 extant *Drosophila* species under a stochastic birth-and-death evolutionary process with 100 replications for each of the six gene families (Hahn et al. 2005; De Bie et al. 2006) (see Materials and Methods in the supplementary materials, Supplementary Material online). We found that six of the gene family pairs still showed positive correlation that were significantly higher than that from those generated by our simulations, that is, *CYP450-GR*, *CYP450-OR*, *OBP-Tryp_SPc*, *OBP-GST*, *Tryp_SPc-GST*, and *OR-GR*. These results further support our conclusion that the correlated changes in these six gene family sizes is not due to a random process.

Potentially correlated changes in gene family size could simply be due to parallel changes in genome sizes (i.e., all gene families in bigger genomes will likely be larger) or gene number content (i.e., genomes with a greater number of genes likely have larger gene families). When we considered these possibilities, we found that neither of them could explain the correlations that we observed for our six gene families (table 3), that is, we found that the sizes of our six candidate gene families do not correlate with either genome size or genome gene number. The failure of the sizes of these six gene families to be correlated to genome size or genome gene number suggests that the correlated evolution of the six gene families is not due to genomic causes.

Gene copy number variation is considered to be a pivotal factor underlying the complexity of functional traits (Demuth et al. 2006; Hahn et al. 2007). Comparative genomic analyses have demonstrated that large disparities in the number of genes involved in same functional processes occur among organisms (Hahn et al. 2007), suggesting that changes in gene numbers may explain differences in specific traits between

Table 1
Correlation among Intact Gene Repertoire Sizes of the Six Gene Families in 12 *Drosophila* Species

	Tryp_SPc	*OBP*	*GR*	*OR*	*GST*
OBP	0.842				
	(0.001)				
GR	*0.571*	0.641			
	(*0.053*)	(0.025)			
OR	0.418	*0.521*	0.907		
	(0.177)	(*0.083*)	(4.62×10^{-5})		
GST	0.744	0.720	0.607	0.639	
	(0.006)	(0.008)	(0.036)	(0.025)	
CYP450	0.410	*0.539*	0.805	0.820	*0.550*
	(0.186)	(*0.071*)	(0.002)	(0.001)	(*0.064*)

Note.—Correlation coefficients with their statistical significance (below in brackets) are shown for each pair of gene families. The shaded boxes indicate those with statistically significant (at 95% level) correlations. Values shown in italics are only marginally significant (with $0.05 < P < 0.1$).

species. In addition, functionally related gene families often demonstrate similar evolutionary pressures, functional specificities, because natural selection tends to retain functionally complementary gene gains and losses on these gene families (Fryxell 1996; Pazos and Valencia 2008). Our observation of a positive correlation in the sizes of the repertoire of six specific gene families (*Tryp_SPc*, *GR*, *OR*, *OBP*, *GST*, and *CYP450*) indicated that a similar and parallel demand for these families exists among *Drosophila*.

What Are the Potential Forces Driving the Correlated Evolution?

Coevolving proteins are subject to common evolutionary constraints and show higher level of similarity of evolutionary pattern than those of unrelated proteins. Here, the observed correlated evolution of gene families is suggestive that proteins in these gene families interact in a network or play related roles in the same pathway. *GR*, *OR*, *OBP*, *Tryp_SPc*, *GST*, and *CYP450* indeed do have related functions, linked by chemosensory perception and diet. We believe that this may be the common link and we do not know of any other (nondiet) common physiological process that links these proteins.

Table 2
Correlation of the Sizes of Intact Genes for the Six Gene Families in 12 *Drosophila* Species Analyzed by the Method of Phylogenetic Independent Contrasts

	Tryp_SPc	*OBP*	*GR*	*OR*	*GST*
OBP	0.648				
	0.031				
GR	0.659	0.439			
	0.028	0.177			
OR	0.448	0.195	0.902		
	0.167	0.566	1.47×10^{-4}		
GST	0.605	0.308	0.714	0.619	
	0.049	0.357	0.014	0.042	
CYP450	0.081	0.215	0.667	0.701	0.520
	0.812	0.525	0.025	0.016	0.101

Note.—Correlation coefficients with their statistical significance are shown for each pair of gene families. The shaded boxes indicate those with statistically significant (at the 95% level) correlations.

Table 3
Correlation of the Sizes of the Six Gene Families with Genome Size and Number of Protein Coding Sequences in 12 *Drosophila* Species

	R(Genome Size)	*P* Value
Tryp_SPc	−0.124	0.702
OBP	0.113	0.727
GR	0.292	0.357
OR	0.162	0.615
GST	0.041	0.900
CYP450	0.516	0.086
	R(Number of Proteins)	*P* Value
Tryp_SPc	−0.079	0.807
OBP	0.127	0.695
GR	−0.138	0.670
OR	−0.091	0.780
GST	0.055	0.865
CYP450	−0.248	0.438

Note.—Correlation coefficients for each gene family with genome size (top) and $\log_{10}$-transformed numbers of protein coding sequence (bottom). The significance of the coefficients is shown on the right (*P* value). The $\log_{10}$-transformed numbers of protein coding sequences was obtained from *Drosophila* 12 Genomes Consortium (Clark et al. 2007).

Chemosensory perception contributes profoundly to the fitness of an organism through processes such as smell and taste, which are involved in behaviors such as the finding and identifying food, choosing mates, facilitating communication, taking precaution against predators, and avoiding toxins (Nei et al. 2008). Peripheral chemosensory perception in insects is performed by several groups of multigene families including the olfactory and GRs. It has been demonstrated that *GR* display a pattern of evolution similar to that seen for the *OR* genes (Gardiner et al. 2008). OBPs were proposed to recognize odorants in the environment and shuttle them to underlying olfactory receptors (Pelosi 1994). Therefore, *GR*, *OR*, and *OBP* are joined together by their functions in the chemosensory perception. We proposed that chemosensory perception is one of the potential forces driving the correlated evolution of these three gene families.

Food is a powerful driving force in the evolution of species. Many Tryp_SPc play important roles in the digestion of food (Rawlings and Barrett 1994). CYP450 and GSTs are two classes of the major enzymes responsible for the detoxification of toxic compounds contained in or produced from food (Tijet et al. 2001; Ranson et al. 2002; Low et al. 2007; Chung et al. 2009). After food is selected and

ingested, it must be digested to release nutrients, and here, adaptation of the *Tryp_SPc* family of proteases may have a role. Food is also known to contain, or can be metabolized into, toxic compounds, thus adaptation of the *GST* and *CYP450* families may occur to deal with novel diet-related toxins. Therefore, functions for food join *Tryp_SPc*, *GST* and *CYP450* together; and roles in diet may be a force driving the correlated evolution of these three gene families. In addition, the *ORs*, *OBP*, and *GR*s are involved in the sensing (finding and selection) of food, which may explain the correlation among the six gene families. However, we did not find strong evidence of correlation for some pairs, for example, *Tryp_SPc-OR*, *Tryp_SPc-CYP450*, which may be consequence of the fact that many genes in these families, especially Tryp_SPc, play roles in other unrelated pathways that are not related to food. We think that the correlations that we observed are attributed to those proteins within these families that have functions with food and chemosensory perception, such as food selection, finding, digestion, and detoxification and not due to those proteins within these families that have other functions. In addition, in contrast to the expectations of an adaptationist theory, a substantial portion of the chemosensory perception receptor gene repertoire appears to have been generated by genomic drift, a random process of gene duplication and deletion (Nozawa et al. 2007; Nei et al. 2008), which will also influence and confuse correlated evolution.

In conclusion, our observation of correlated changes in the sizes of gene families is better explained by adaptation driving correlated evolution of these gene families during the evolution of *Drosophila* because these gene family have correlated functions, such as for chemosensory perception and diet, with these results being consistent with a recent study proposing that ecological adaptation determines the functional mammalian olfactory subgenomes (Hayden et al. 2010).

Supplementary Material

Supplementary tables S1–S3 are available at *Genome Biology and Evolution online* (http://gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by grants from the National Basic Research Program of China (973 Program, 2007CB411600), the National Natural Science Foundation of China (30621092), and Bureau of Science and Technology of Yunnan Province.

Literature Cited

Chung H, et al. 2009. Characterization of Drosophila melanogaster cytochrome P450 genes. Proc Natl Acad Sci USA. 106:5731–5736.

Clark AG, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 450:203–218.

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

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS One. 1:e85.

Diniz-Fi JAF, de Sant'Ana CER, Bini LM. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution. 52:1247–1262.

Fryxell KJ. 1996. The coevolution of gene family trees. Trends Genet. 12:364–369.

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Mol Ecol. 17:1648–1657.

Gittleman JL, Kot M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Syst Zool. 39:227–241.

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res. 15:1153–1160.

Hahn MW, Han MV, Han SG, McVean G. 2007. Gene family evolution across 12 Drosophila genomes. PLoS Genet. 3:e197.

Hayden S, et al. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. Genome Res. 20:1–9.

Hofer D, Asan E, Drenckhahn D. 1999. Chemosensory perception in the gut. News Physiol Sci. 14:18–23.

Low WY, et al. 2007. Molecular evolution of glutathione S-transferases in the genus Drosophila. Genetics. 177:1363–1375.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 9:951–963.

Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. Proc Natl Acad Sci U S A. 104:20421–20426.

Ollier S, Couteron P, Chessel D. 2006. Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. Biometrics. 62:471–477.

Pazos F, Valencia A. 2008. Protein co-evolution, co-adaptation and interactions. EMBO J. 27:2648–2655.

Pellegrini M, et al. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A. 96:4285–4288.

Pelosi P. 1994. Odorant-binding proteins. Crit Rev Biochem Mol Biol. 29:199–228.

Ranson H, et al. 2002. Evolution of supergene families associated with insecticide resistance. Science. 298:179–181.

Rawlings ND, Barrett AJ. 1994. Families of serine peptidases. Methods Enzymol. 244:19–61.

Tijet N, Helvig C, Feyereisen R. 2001. The cytochrome P450 gene superfamily in Drosophila melanogaster: annotation, intron-exon organization and phylogeny. Gene. 262:189–198.

Vieira FG, Sánchez-Gracia A, Rozas J. 2007. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. Genome Biol. 8:R235.

Wu D-D, Wang G-D, Irwin DM, Zhang Y-P. 2009. A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito. Mol Biol Evol. 26: 2333–2341.

Associate editor: Takashi Gojobori