RESEARCH ARTICLE   **Open Access**

Functional requirements driving the gene duplication in 12 Drosophila species

Yan Zhong[†], Yanxiao Jia[†], Yang Gao, Dacheng Tian, Sihai Yang[*] and Xiaohui Zhang[*]

Abstract

Background: Gene duplication supplies the raw materials for novel gene functions and many gene families arisen from duplication experience adaptive evolution. Most studies of young duplicates have focused on mammals, especially humans, whereas reports describing their genome-wide evolutionary patterns across the closely related *Drosophila* species are rare. The sequenced 12 *Drosophila* genomes provide the opportunity to address this issue.

Results: In our study, 3,647 young duplicate gene families were identified across the 12 *Drosophila* species and three types of expansions, species-specific, lineage-specific and complex expansions, were detected in these gene families. Our data showed that the species-specific young duplicate genes predominated (86.6%) over the other two types. Interestingly, many independent species-specific expansions in the same gene family have been observed in many species, even including 11 or 12 *Drosophila* species. Our data also showed that the functional bias observed in these young duplicate genes was mainly related to responses to environmental stimuli and biotic stresses.

Conclusions: This study reveals the evolutionary patterns of young duplicates across 12 *Drosophila* species on a genomic scale. Our results suggest that convergent evolution acts on young duplicate genes after the species differentiation and adaptive evolution may play an important role in duplicate genes for adaption to ecological factors and environmental changes in *Drosophila*.

Keywords: Young duplication, Environmental factor, Convergent evolution, Adaptive evolution

Background

Gene duplication is one of the dominant driving forces in adaptive evolution of genome and genetic systems [1]. Duplicate genes are considered to be the raw materials and the primary mechanism for generation of novel gene functions [2]. At least 15% of genes in human genome and 8% to 20% of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cervisiae* genomes are thought to arise from gene duplications [3].

Young duplicate genes will ultimately suffer one of three long-term fates: (i) one copy may lose gene function by nonfunctionalization/pseudogenization; (ii) one copy may evolve a new beneficial function by means of neofunctionalization and the other retain the old; or (iii) both duplicated copies may stably be maintained with daughter copy partitioning of ancestral gene function by

subfunctionalization [3-7]. Many models have been proposed that pseudogenization could be the most common fate of duplicated genes [8-10]. In addition, evidences for adaptive evolution of pseudogenes have been reported in many organisms, such as pseudogenes in 80 *Arabidopsis* accessions [11] and the *rcsA* gene in *Yersinia pestis* [12]. Similarly, the preservation of duplicated genes might be a by-product of neutral evolution [1,9,13], or might be adaptive substitutions during or after fixation of duplicates [1], indicating that selection for neofunctionalization is the mechanism to keep them [14,15].

Previous studies conducted in many organisms have widely reported that duplicate genes undergo adaptive evolution. At the genome-wide level, the signatures of adaptive natural selection of young gene duplicates are found with high frequency in the human, macaque, mouse and rat genomes [16]. Furthermore, gene duplicates from *Drosophila pseudoobscura* neo-X chromosome [17] and a group of digestive protease encoding genes that are associated with recent, lineage-specific

* Correspondence: sihaiyang@nju.edu.cn; xiaohuizhang@nju.edu.cn
[†]Equal contributors
State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, 22 Hankou Rd., Nanjing 210093, China

duplications in *Drosophila Arizonae* [18] are detected under adaptive evolution. Researches have been focused almost either on recent duplication events occurring in humans or other mammals involved in human diseases [19,20] or on the duplication and adaptive evolution of single gene families, such as *chalcone synthase* genes and *MADS-box* genes in plants [21,22], fatty acid biosynthesis genes in bacteria [23], or *Toll-like receptor* genes in *Drosophila* [24]. Although gene gain and loss is estimated with a *Drosophila*-wide perspective [25], a systematic investigation of the genetic character and evolutionary pattern of young duplicate genes across the closely related *Drosophila* species has not been reported.

Sequencing of the genomes of the 12 worldwide *Drosophila* species (*Drosophila* 12 Genomes Consortium 2007) [26] provides the opportunity to reveal the evolutionary genetics of recent duplications. These species capture a range of evolutionary distances: closely related sister-species, such as *D. simulans* and *D. sechellia* or *D. pseudoobscura* and *D. persimilis*; distantly related species classified into different subgenera, such as *Sophophora* and *Drosophila*. There are also cosmopolitan species such as *D. melanogaster* and *D. simulans* or highly restricted species such as *D. sechellia*, distributed in some specific geographic ranges [27]. Additionally, the diverse host preferences provide a way to connect recent duplications with ecological and environmental factors. In this work, we conducted a genome-wide investigation of young duplicate genes across 12 *Drosophila* species to uncover their evolutionary patterns.

Results

Young duplicate genes in 12 *Drosophila* genomes

Across the 12 *Drosophila* genomes, a total of 22,488 gene families were detected, including 3,647 young duplicate gene families (see Methods; Table 1), suggesting that approximately 16.2% of the total gene families included young duplicates. In these young duplicate gene families, three types were defined based on their expansion patterns: species-specific expansions, lineage-specific expansions and complex expansions. The species-specific young duplicate gene families clearly predominated (3159/3647 = 86.6%) over the other two types of expansions. On the other hand, uneven distribution of the species-specific young duplicate genes among different species, ranging from 54 to 794, was also observed. For example, *D. melanogaster* had the least family number (54), while the highest three values were found in *D. willistoni* (318), *D. yakuba* (569) and *D. grimshawi* (794), respectively (Table 1). This uneven distribution of the young duplicate genes was also found in the lineage-specific expansions and the complex expansions. For example, 114 duplicate gene families were detected in lineages of *D. persimilis* and *D. pseudobsura*, which is approximately 11.9- and 4.9-fold greater than that in lineages of *D. erecta* and *D. yakuba* or *D. sechellia* and *D. simulans*, respectively. Also as expected, if there are more species (e.g. > 3) in a group of lineage-specific and complex expansions, fewer duplicate gene families were detected.

Table 1 Number of young duplicate gene families for three types of expansion

Species	Species-specific expansions	Lineage-specific expansions						Complex expansions													Total
		2	2	2	3	5	6	2	3	4	5	6	6	7	8	9	10	11	11	12	
D. simulans	139	+[a]	-	-	+	+	+	-	-	+	+	-	+	+	+	+	+	+	+	+	
D. sechellia	289	+	-	-	+	+	+	-	-	+	+	-	-	-	+	+	-	+	+	+	
D. melanogaster	54	-[b]	-	-	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	
D. yakuba	569	-	+	-	-	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	
D. erecta	89	-	+	-	-	+	+	-	-	-	+	+	+	+	+	+	+	+	+	+	
D. ananassae	150	-	-	-	-	-	+	-	-	-	+	-	-	-	+	+ …[c]	+	+	+	+	
D. pseudobsura	247	-	-	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	+	
D. persimilis	232	-	-	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
D. willistoni	318	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+	+	
D. virilis	118	-	-	-	-	-	-	+	-	-	+	-	+	-	-	-	+	+	+	+	
D. mojavensis	160	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	+	
D. grimshawi	794	-	-	-	-	-	-	+	+	-	-	+	-	-	-	-	-	+	+	+	
		29	12	143	11	5	1	21	177	5	44	3	9	1	6	1 1 1 1 3	2	3	1 2	1 2 3	
Total	**3159**			**201**							**287**										**3647**

[a] Corresponding species is involved in the relevant expansion events.
[b] Corresponding species is not involved in the relevant expansion events.
[c] Not all expansion events are showed because of space limitation.

Interestingly, three young duplicate gene families were detected respectively in complex expansions occurring in 11 and 12 species (Table 1). Although the six families were classified as the complex expansion type, species-specific duplication events also were found in all of these families (Additional file 1: Figure S1), especially independent duplications after the species differentiation in many species. For example, 15, 16 and 91 species-specific duplicate clades across all of the 12 species were detected in family 2,419, 7,827 and 8,177, respectively. In the other 3 families, 8, 11 and 145 species-specific duplicate clades were distributed in 8, 11 and 11 species, respectively. In addition, some lineage-specific duplications were also found in these families. All these results suggested that these duplicate gene families were likely to have been shaped by convergent evolution due to independent expansions in many species after the species differentiation.

Distribution of young duplicate genes on chromosomes

To explore the distribution of the young duplicate genes on the chromosomes, stochastic simulations were implemented using the observed gene numbers with 10,000 times random repeats. The chromosomal distribution was significantly non-random ($P < 0.05$); for example, chromosome 2 (2L & 2R), 3 (3L & 3R) and X contained a mass of young duplicate genes (Additional file 2: Figure S2). Figure 1 shows graphs representing these simulation results. Furthermore, the windows in Figure 1 with the observed number larger than the upper level of the confidence intervals correspond to hotspot regions on the chromosomes for the young duplications.

As shown in Figure 1, hotspot regions were found in all three types of young duplications, especially in species-specific expansions (Figure 1A), e.g. on chromosome 2 in *D. grimshawi* and on chromosome 3 in *D. yakuba*. In contrast, few hotspot regions were found in species such as in *D. ananassae* and *D. melanogaster*. Interestingly, some duplication hotspot regions were shared by more than one species in the species-specific expansions (marked by dash lines in Figure 1A), also suggesting convergent evolution of these genes among different species. However, none of shared hotspot region were detected in lineage-specific duplications, although the two species had similar trend lines which were generated by the observations and simulation numbers (Figure 1B). In complex expansions, few hotspot regions were detected along the chromosomes (Figure 1C).

Functional preference of young duplicate genes

To further reveal the genetic characteristics of the young duplications, the domains of the duplicates were detected using Pfam searches. Subsequently, the protein domains were counted in each species. For species-specific expansions, a total of 1,277 different domains were found in 12 species, averaging 106 protein domains in each species (Additional file 3: Table S1). Interestingly, approximately 84% of protein domains occurred only once or twice, suggesting that most domains were unique. However, the frequency of some protein domains, such as DUF1676 in *D. willistoni*, annexin and dynein_IC2 in *D. melanogaster*, inositol_P and PAP2 in *D. yakuba*, were high in one species but low (0 or 1) across the other 11 species, suggesting that these species-specific duplicate proteins might be driven by adaptive evolution in each species. Furthermore, some protein domains occurred in a lineage-specific manner, although they were detected in the species-specific expansion events. For example, the expansion of domains Gb3_synth and Gly_transf_sug shared by *D. mojavensis* and *D. virilis*, were greatly expanded only in these two species. A similar situation was also observed in the alpha-amylase domain, which occurred in two closely related *Drosophila* species, *D. sechellia* and *D. simulans*. Although different types and numbers of protein domains were examined in each species, we still found that approximately 4% of the domains appeared simultaneously in ≥ 6 species. Prominent examples of these protein domains were trypsin, p450 and WD40, which were detected in 12, 11 and 11 species, respectively (Additional file 3: Table S1). These proteins are all important in response to environmental stimuli [28,29]. To investigate whether the high-frequency domains also occupied in large numbers in each species or vice versa, we examined the occurrence-frequency of the top 20 domains in 12 genomes. Interestingly, these high-frequency domains also had a large number of copies in the related species (Figure 2A), suggesting that these high-frequency duplicated proteins play an important role in the evolution of these species.

An identical approach was also used for the gene families of lineage-specific expansions and complex expansions. Our results showed that most gene families contained limited protein domains, although the number of the same domain was always different. However, some protein domains were still undergoing rapid expansion independently in many species, e.g. the six shared duplicate gene families in complex expansions occurring in 11 and 12 species (Table 1; Figure 2B and C). Furthermore, Ank, EGF, Peptidase_M17 and Peptidase_M17_N, which were all conserved and widespread domains in organisms for survival, exhibited high frequencies in 11 species (Figure 2B). In the shared expansion events of 12 species (Figure 2C), 12 of the top 20 protein domains such as histones, HSP70, Lys, co-occurred in 12 species. Numerous previous studies have shown that these protein domains are closely related to stress responses and pathogens in the environment, for example, histones are involved in stress responses [30], HSP70 protects cells
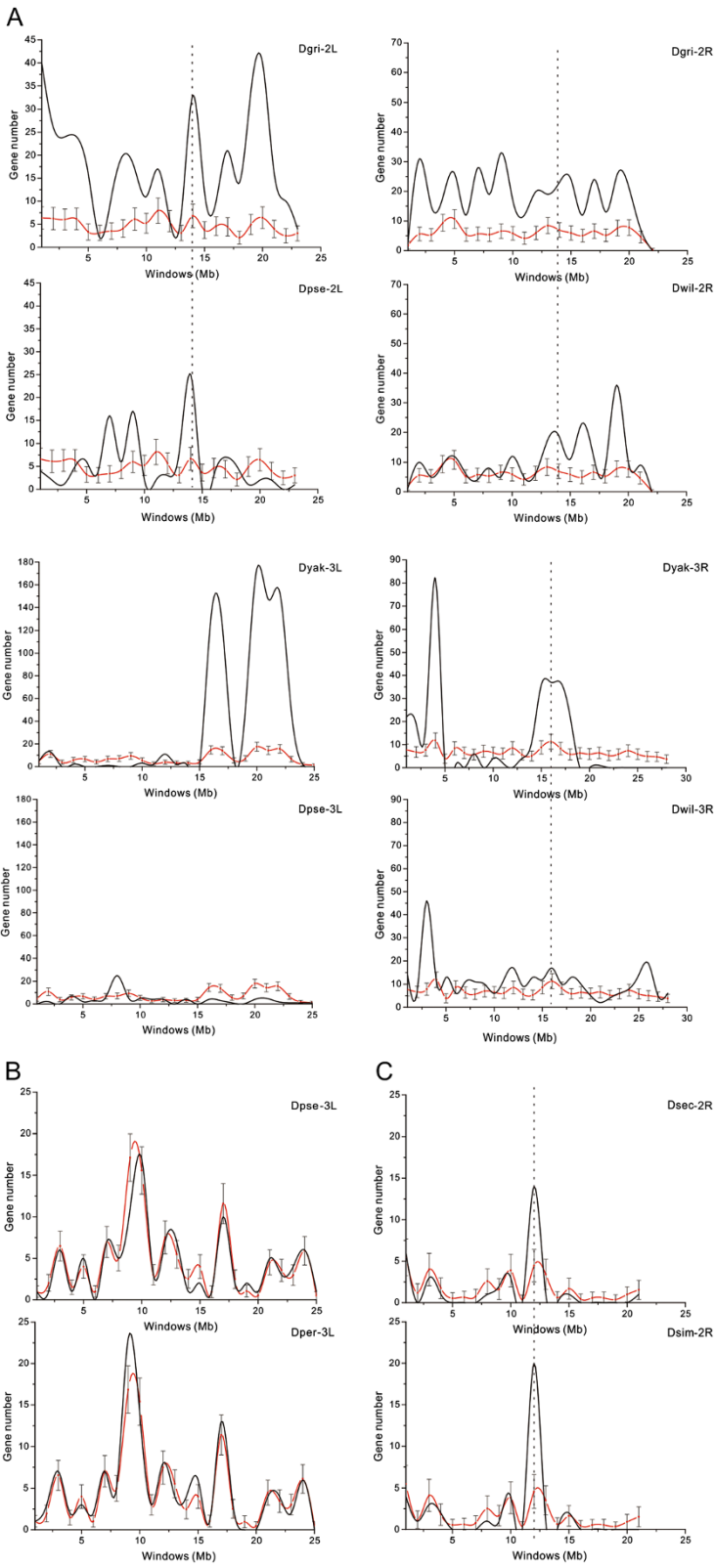
Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 Distribution of young duplicate genes on chromosomes. (A) Species-specific duplicates on chromosome 2 and 3. **(B)** Lineage-specific duplicates of *D. pseudoobscura-D. persimilis* pair on chromosome 3. **(C)** Duplicates of complex expansions occurring in 11 species on chromosome 2. Black lines represent observations, while black bands and red lines (red dots) show confidence intervals and average numbers of genes in corresponding windows. Dashed lines indicate the shared hotspot regions between the species. Dgri: *D. grimshawi*, Dpse: *D. pseudoobscura*, Dper: *D. persimilis*, Dsec: *D. sechellia*, Dsim: *D. simulans*, Dwil: *D. willistoni*, Dyak: *D. yakuba*; 2L: chromosome 2L, 2R: chromosome 2R, 3L: chromosome 3L, 3R: chromosome 3R.

from stress [31], and Lys (lysozyme) acts as a bacterio-lytic enzyme by hydrolyzing bacterial cell walls [32], suggesting that these shared duplications play an important role in adaption to ecological factors and environmental changes in *Drosophila*.

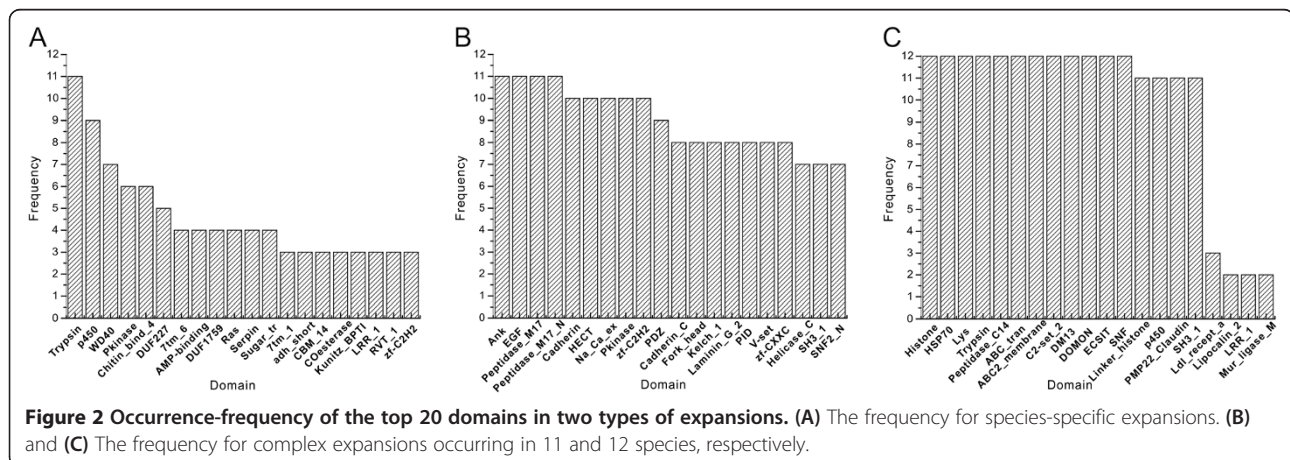## Nonsynonymous and synonymous substitution between paralogs and orthologs

The ratio of nonsynonymous to synonymous nucleotide substitution (*Ka/Ks*) is considered as an important parameter indicating the strength of functional constraints. The smaller the *Ka/Ks* ratio is, the stronger the functional constraints are. The 12 *Drosophila* whole-genome data offer us unprecedented opportunity to explore the different selection pressure between paralogs and orthologs. Therefore, we examined *Ka/Ks* ratios for paralogs and orthologs in each duplicate gene family.

The average *Ka/Ks* between paralog gene pairs or ortholog gene pairs in these young duplicate gene families were 0.626 and 0.445, respectively, which was significantly (*P* < 0.05) larger than the genome-wide *Ka/Ks* (0.218) between ortholog pairs, suggesting relaxation of the functional constraints in the young duplicate gene families. Figure 3 shows that most of the gene pairs (91.2%), including paralogs and orthologs, had *Ka/Ks* ratios less than 1, demonstrating that most young duplicate genes were under purifying selection. However, there were still 229 and 82 gene pairs with *Ka/Ks* ratios greater than 1 for paralogs and orthologs, respectively, indicating that some young duplicate genes are driven by positive selection. However, in the gene pairs with *Ka/Ks* values exceeding 1, many values were just slightly greater than 1 and only few pairs were detected to have *Ka/Ks* ratios significantly greater than 1.

Based on the strengths of boxes and whisker lines in species-specific expansions in Figure 3, it was clear that *Ka/Ks* between paralogs had a broader dispersed distribution, larger median and quartile values than orthologs, indicating that paralogs had higher *Ka/Ks* than orthologs. Similar results were also obtained in lineage-specific expansions (Figure 3D), with the exception of *D. sechellia* vs. *D. simulans* and *D. persimilis* vs. *D. pseudoobscura*. To further ensure that the *Ka/Ks* of paralogs were significantly greater than those of orthologs, we conducted paired *t*-tests. Apart from four pairs, the other *Ka/Ks* ratios of paralogs and orthologs exhibited highly significant (*P* < 0.01) or significant (*P* < 0.05) differences (Additional file 4: Table S2). All the results showed that paralogs had significantly higher *Ka/Ks* than orthologs and indicated that paralogs are subject to weaker functional constraints and faster evolutionary processes than orthologs.

Linear analog was also performed between the mean *Ka/Ks* of paralogs and orhtologs (Additional file 5: Figure S3). In the same family, the dot above the trend lines (slope = 1) indicated that paralogs have higher evolutionary rates than orthologs. Interestingly, it was clear that some dots lay far away the trend lines. Detection of the protein domains of these dots (Additional file 4: Table S2) showed that most of the domains detected in the genes of upper dots, such as Coesterase [33,34], Turandot [35] and MIP [36] were
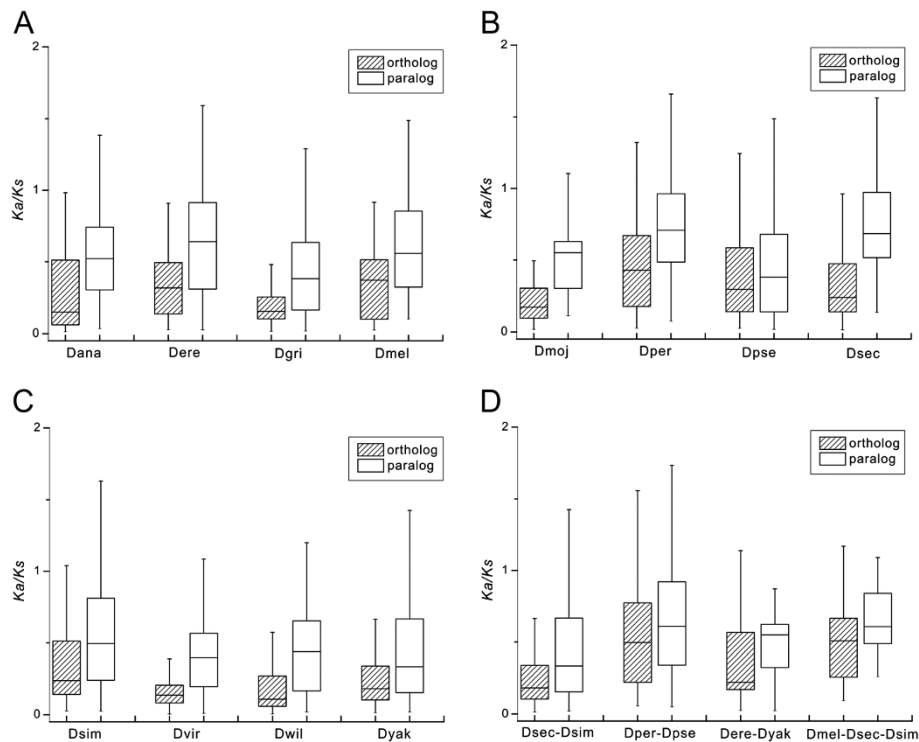

Figure 2 Occurrence-frequency of the top 20 domains in two types of expansions. (A) The frequency for species-specific expansions. **(B)** and **(C)** The frequency for complex expansions occurring in 11 and 12 species, respectively.

Figure 3 Box plot comparing the mean *Ka/Ks* ratios of paralog gene pairs and ortholog gene pairs. (A), **(B)** and **(C)** *Ka/Ks* ratios of species-specific duplicates. **(D)** *Ka/Ks* ratios of lineage-specific duplicates. The top bar is maximum observation, the lower bar is minimum observation, the top of box is third quartile, the bottom of box is lower or first quartile, the middle bar in box is median value.

involved in stress responses. These results also suggested that the young duplicates result from adaption to the environment both in species-specific and lineage-specific expansions.

Evolutionary analysis of young duplicate genes across 12 *Drosophila* species

To detect the timing of recent duplication in each species, the *Ks* values were calculated. We adopted the common assumption that *Drosophila* species experienced about 10 generations/year and that the single nucleotide mutation rate was 5.8×10^{-9} mutations per generation [37]. Furthermore, only *Ks* values lower than 1.0 were kept to avoid the saturation of nucleotide substitutions.

On the whole, the young duplication events occurred over a short span of time (0.082-5.282 MYA). The duplication time of species-specific expansions fell in a range from 1.238 MYA (*D. simulans*) to 3.573 MYA (*D. melanogaster*) (standard deviation, 0.712) (Table 2), indicating that all the species-specific expansions occurred within a short time. Moreover, most of our estimates of duplication time were less than the species divergence time reported by Tamura [38]. However, the divergence time of several closely related species previously reported, including *D. simulans* vs. *D. sechellia* (< 0.93 ± 0.49 MYA)

and *D. pseudoobscura* vs. *D. persimilis* (0.85 ± 0.29 MYA) was slightly lower than their respective family duplication times (1.238, 2.313 and 1.327, 1.573 MYA). Similarly, higher duplication times in the four species were also found in lineage-specific expansions and complex expansions. Moreover, the lowest standard deviations of the duplication time were detected between these lineage species in lineage-specific expansions, which suggested that closely related species duplicated in close periods, especially the species pair *D. persimilis* vs. *D. pseudoobscura* (2.341 and 2.401 MYA). In the six lineage species, there was a group of duplication times with more compact distribution and smaller values than those in species-specific expansions, which indicated that the expansion occurred over a more concentrated and closer period in lineage-specific expansions than in species-specific expansions. It was clear that less species and closer relationships caused such results. Finally, although the highest standard deviations were found in complex expansions, especially in 11 species with a broader range for duplication time (0.765–5.282 MYA) and a larger standard deviation (1.686) than those in others, their distributions of duplication time were still in relative compact period. This demonstrated that these duplicated genes in complex expansions might appear at relatively scattered time compared with

Table 2 Duplication time for species-specific expansions, lineage-specific expansions and complex expansions

Species	Species-specific expansions	Lineage-specific expansions						Complex expansions		
D. simulans	1.238	1.561	-	-	1.492	1.916	0.981	2.763	1.531	1.655
D. sechellia	2.313	2.249	-	-	1.281	1.575	0.084	2.576	0.765	0.525
D. melanogaster	3.573	-	-	-	1.022	0.769	0.998	0.674	1.169	3.794
D. yakuba	1.613	-	3.526	-	-	0.641	0.564	4.843	1.644	2.549
D. erecta	1.874	-	2.775	-	-	0.539	0.082	5.131	1.287	0.760
D. ananassae	2.572	-	-	-	-	-	0.582	3.372	5.282	1.244
D. pseudoobscura	1.327	-	-	2.401	-	-	-	4.108	-	3.790
D. persimilis	1.573	-	-	2.341	-	-	-	3.268	1.017	4.490
D. willistoni	2.085	-	-	-	-	-	-	2.419	2.560	1.473
D. virilis	2.513	-	-	-	-	-	-	3.509	5.153	1.003
D. mojavensis	3.388	-	-	-	-	-	-	-	1.332	1.023
D. grimshawi	2.160	-	-	-	-	-	-	2.162	4.755	0.449
Standard deviation	0.712	0.344	0.375	0.030	0.192	0.553	0.371	1.207	1.686	1.349

duplicates in the other two types. Therefore, we might infer that *Drosophila* species have consistently duplicated to adapt to environmental changes.

Discussion

Convergent evolution of young duplicate genes across the 12 *Drosophila* species

Convergent evolution plays an important role in biological adaptation, by which distantly related organisms independently evolve similar structures or functions in order to adapt to similar environments or ecological niches [39], such as, the specialized oxygen transport function of oxygen transport hemoglobins in jawed and jawless vertebrates [40] and the similar substrate of apolipoprotein (a) in humans and hedgehogs [41]. Although there are many other theories could explain the evolutionary process of young duplicates, such as genomic drift proposed by Nei [42,43], convergent evolution might be more convincible for two evidences detected in our study.

In our study, the phylogenetic trees (Additional file 1: Figure S1) and the chromosomal distributions (Figure 1) of young duplicate genes also provide evidence of convergent evolution. Six young duplicate families were found in complex expansions occurring in 11 or 12 species with many species-specific duplication clades across these 11 or 12 species. Interestingly, the phenomenon that the independent duplicates with similar function preference are under convergent evolution has also been previously reported both in animals and plants. For example, histone proteins are highly alkaline proteins in eukaryotic genomes which package DNA into nucleosomes [44] and independent convergent evolution has produced striking similarities between plant and animal histones [45]. Another example of similar genetic charac-

teristics shared by distant species is the digestion function of lysozymes (Lys domain) in animals. Lysozymes are usually present in tears, saliva and other bodily fluids and have independently been recruited to the stomach and play important roles in enzyme functions across vertebrates [46]. Furthermore, some duplication hotspot regions were shared by more than one species across their chromosomes in species-specific expansions. Interestingly, conserved duplication hotspots have also been previously detected between *D. melanogaster* and *D. simulans* [47]. Similar function preference and identical hotspot regions arising from independent duplications suggest that the young duplicate genes have undergone convergent evolution which appears to have played an important role in the independent evolution of adaptive traits in 12 *Drosophila* species.

Adaptive evolution supported by functional bias of young duplicates

It is well-known that duplicate genes face three possible fates: pseudogenization, subfunctionalization and neofunctionalization. Pseudogenization is considered as the most common fate of duplicate genes [8-10], but more evidence support the models of subfunctionalization or neofunctionalization, as the mechanisms for the preservation of duplicate genes under adaptive selection [6,15,48,49]. Many previous studies have shown that the duplicated genes could adapt to various conditions, in particular, genes encoding membrane or secreted proteins which are always involved in ecological stimuli or stress. For example, adaptive gene duplications have been found in response to biotic stress [50], antibiotics [51,52], weedicides [53] or pesticides [54,55], drugs or toxins [56], extreme temperatures [57,58], nutrient limitation [59,60] and symbiosis between host and parasite [61].

In this study, it was shown that the protein domains of trypsin, p450, WD40 and Pkinase in species-specific expansions, Ank, EGF, histone, HSP70 and Lys in complex expansions occurred with high frequency across the 12 Drosophila species. Interestingly, these young duplicates were also involved in different aspects of interactions with the environment. Trypsin is one of the largest families of secreted serine proteases found in the digestive system of vertebrates and invertebrates. Although it participates in many basic physiological processes [62-64], it is predominantly involved in diet and digestion. The high frequency of trypsin across 12 Drosophila species indicated that consistent and independent duplication for adaptation to specific dietary habits was due to their diverse ecosystems [27,29]. For example, investigations of trypsin family conducted in various genomes, such as fruit fly [65], mosquitoes [66] and leaf-eating monkey [67], have all indicated that adaption occurs in response to specific diets. In particular, researches into the rapid diversification of trypsin genes in 12 Drosophila species have provided insights into the ecological forces driving the adaptive evolution by comparing the relationship between duplications and host preference shifts [65].

Another protein domain shared between 11 Drosophila species detected in this study was cytochrome p450 (CYP). P450 comprise a superfamily of enzymes that occurs with a high degree of diversity in all organisms [68]. Among the various biological functions of p450, we focused on the oxidation of xenobiotic compounds, which facilitates their excretion from the organism [69,70]. Abuse of insecticides has forced adaptive evolution in Drosophila over an extremely short period. A single p450 gene, Cyp6g1, is sufficient and necessary for DDT resistance [28] and its cross-resistance to a wide range of other insecticides has also been identified in Drosophila [71,72]. Furthermore, functional divergence and positive selection detected in mammalian CYP genes, provide insights into the adaptive selection of CYPs in response to high diversity of xenobiotics [73].

Other expanded domains were also identified with roles in adaption to various ecological factors, especially stress. For example, some SAPK (stress-active protein kinases, Pkinase) mediate cellular responses to toxins and physical stress [74] and TAK1 (transforming growth factor-β-activate kinase, Pkinase) is a key regulator in response to diverse stimuli in adaptive immunity [75], ankyrin proteins (Ank) play a role in stress responses and disease resistance both in animals [76] and plants [77], histones [30] and HSP70 proteins protect cells from stress [31], and Lys proteins act as bacteriolytic enzymes by hydrolyzing cell bacterial walls [32].

These observations indicate that shared young duplications reflect adaptive evolution of the Drosophila species to global ecological pressures.

Adaptive evolution contributes to specific functional preference

In this study, although most paralogs and orthologs of these young duplicate gene families had Ka/Ks ratios lower than 1, some Ka/Ks ratios greater than 1 were also found both in species-specific and lineage-specific expansions (Figure 3), demonstrating that they were under adaptive selection. Furthermore, paralog gene pairs had higher Ka/Ks ratios than ortholog gene pairs across 12 Drosophila species. It can be concluded that the paralogs have higher frequency of adaptive evolution than the orthologs [48]. Previous research has indicated that many genes families in Drosophila are driven by adaptive selection, such as, elastase/chymotrypsin, trypsin and astacin, which are all involved in digestive processes in D. arizonae [26], two immunity-related gene families, Toll-like receptors and lysozyme in D. melanogaster [24,78], and metallothionein genes involved in metal tolerance [79,80]. Moreover, positive selection is a major force driving the evolution of male-specific recent duplicates on neo-X chromosome in D. pseudoobscura [17] and segmentally duplicated seminal fluid genes in D. melanogaster [81].

Functional analysis of those gene families in which paralogs had higher Ka/Ks ratios than orthologs with ratios exceeding 1 (Figure 3) showed that adaptive evolution leads to species-specific and lineage-specific functional preference for each Drosophila (Additional file 3: Table S1). Examples include the PDZ-domain containing gene family in D. sechellia, stress-inducible humoral factor Turandot (Turandot domain) in D. yakuba and the Pam16 family (Pam16 domain) in the D. pseudoobscura and D. persimilis pair. Combining the functional preference with the host preference of corresponding Drosophila species [27], it seems reasonable to infer that Drosophila evolve for adaption to a given environment [82].

Interestingly, we found that D. sechellia only distribute in the Seychelle Islands in the Indian Ocean and prefer inhabiting Morinda citrifolia, the fruit of which contain nutrients such as alkaloids. Alkaloids are widely known that play an important role in inhibiting tumors by reducing microtubules disruption during mitosis [83]. Coincidently, PDZ proteins are involved in the interaction between syntrophin-associated serine/threonine kinase and microtubule-associated serine/threonine kinases and are recognized as tumor suppressors [84,85]. Therefore, we speculated that duplication of PDZ in D. sechellia is closely associated with adaption to this unique habitat.

The D. yakuba species in Africa exhibited propagation of Turandot proteins which are adaptively resistant to high temperature, dehydration and UV irradiation [86,87]. In contrast, the Pam16 proteins of D. pseudoobscura and D. persimilis play important regulatory roles in recruiting heat shock protein partners and responding to cold harde-

ning [88-90], indicating superior adaptation of these species to their specific habitats situated in the regions of higher latitude in the Northern Hemisphere compared with other species.

Furthermore, other evidence of adaptive evolution in a single species or lineage species pairs was also detected in domain analysis (Additional file 3: Table S1). For example, annexin and dynein_IC2 in *D. melanogaster*, which are sperm-specific proteins (annexin X and cytoplasmic dynein intermediate chain) and absent in other species of the *melanogaster* species subgroup, support the hypothesis that male reproductive functions are driven by selective sweep and rapid molecular evolution [91]. Another example is alpha-amylase (Amy), a digestive enzyme of the two closely related species of the 12 *Drosophila*, *D. sechellia* and *D. simulans*, for which genetic variation of duplicated amylase genes has been reported revealing adaptive evolution in *Drosophila* [92].

Consequently, adaptive evolution of *Drosophila* species leads to young duplicate genes exhibiting specific function preference in response to ecological factors and environmental changes.

Conclusions

We identified 3,647 young duplicate gene families across 12 *Drosophila* species and detected three types of expansions in these gene families: species-specific, lineage-specific and complex expansions. We found that the species-specific young duplicate genes predominated (86.6%) over the other two types. Furthermore, we observed that, in the same gene family, independent species-specific expansions occurred in many species, even including 11 or 12 *Drosophila* species, suggesting that these young duplicate genes were under convergent evolution after the *Drosophila* species differentiation. We also found that the functional preference of the young duplicate genes was mainly related to responses to environmental stimuli and biotic stresses, suggesting that adaptive evolution may play an important role in duplicate genes for adaption to ecological factors in *Drosophila* species. This work may help us to better understand the evolutionary patterns of young duplicate genes across 12 *Drosophila* species.

Methods

Identification of young duplicate genes

The 12 *Drosophila* genome sequences and annotations were downloaded from the Flybase Datebase [ftp://ftp.flybase.net/genomes/] [93], and the exact version for each species is shown in Table 1 (Additional file 4). An all-versus-all Blastn search with E-value (1.0e-40) was processed across all nucleotide coding sequences (CDSs) in 12 *Drosophila* species, then coverage of > 60% was used to divide the genes into gene families. Subsequently, Clustalw2.0 [94] was used for the pairwise alignment of the nucleotide

sequences of genes in one family and the nucleotide diversity (π) was estimated with the Jukes and Cantor correction by MEGA v5.0 [95]. Young duplicate gene families were defined based on the following two conditions: (1) the number of family members was larger than the number of corresponding species in each family; (2) the highest identity of the paralogs in each family exceeded 90%.

To further analyze young duplicate gene families, three types of expansions were defined: species-specific expansions, lineage-specific expansions and complex expansions. Here, the species-specific expansions were denoted as young duplications occurring only in one species, while other species comprised ≤ 1 member or > 2 members but with less than 90% identity between paralogs. We also defined the latter two types as the young duplications of a gene family in species with a close (lineage-specific expansion) or distant (complex expansion) genetic relationship based on the species tree of the 12 *Drosophila* species [27]. Based on this principle, species-specific expansions were easily identified corresponding to each species. Furthermore, we defined the following six lineage-specific expansions: *D. sechellia-D. simulans*, *D. yakuba-D. erecta*, *D. pseudoobscura-D. persimilis*, *D. melanogaster-D. sechellia-D. simulans*, *D. melanogaster-D. sechellia-D. simulans-D. yakuba-D. erecta* and *D. melanogaster-D. sechellia-D. simulans-D. yakuba-D. erecta-D. ananassae*, and 14 complex expansions (Table 1 & Additional file 6: Table S4).

Sequence alignment and phylogenetic analysis

The amino acid sequences of the duplicate genes in each family were first aligned using the MUSCLE program with default options and then manually corrected using MEGA v5.0 [95]. The alignments were used to construct phylogenetic trees with 1,000 replicates using MEGA v5.0 based on the neighbor-joining (NJ) method [96]. NJ analysis was conducted using pairwise deletion of gaps and kimura-2 model (family3093, family9588 and family7827) or p-distance (family8177). Additionally, for the two families (family21 and family2419) with numerous members, NJ trees were constructed with default parameters and 1,000 bootstrap replicates in Clustalw2.0 [94].

Physical location and structural domains of the young duplicate genes

The hotspot regions for duplication events were examined by identifying the exact physical positions of young duplicate genes across the chromosomes. Detailed genome annotation information was available for the genes of *D. melanogaster* only. Thus, we performed Blast analysis using the CDSs of young duplicate genes in the other 11 genomes against the genome sequences of *D. melanogaster* to gain the position information.

We utilized the position information to process stochastic simulations with 10,000 random repeats by PERL

script. Each chromosome was divided into a number of windows (1Mb/1 Window). We incorporated all genes of corresponding species into the given window of each chromosome for the species-specific duplicate gene families, and an identical approach was taken for the families of lineage-specific expansions and complex expansions. All young duplicate genes identified in this study were further examined for structural domains using the Pfam database [Pfam database 26.0, http://pfam.sanger.ac.uk/] [97] with E-value 1.0.

Calculation of nonsynonymous to synonymous ratio and estimation of duplication time of paralogs

CDSs in each young duplicate gene family were aligned according to alignments of protein sequences in Clustalw2.0 [94]. Subsequently, the nonsynonymous substitutions (Ka), synonymous substitutions (Ks) and ratio of nonsynonymous to synonymous substitutions (Ka/Ks) were calculated for paralog and ortholog pairs in each duplicate gene family using MEGA v5.0 [95].

The mean Ks values were calculated for paralog pairs in each species-specific duplicated gene family and then used to estimate the timing of duplications. The calculations were performed using a single nucleotide mutation rate of 5.8×10^{-9} mutations per generation and it was assumed that *Drosophila* species experienced approximately 10 generations/year [37].

Additional files

Additional file 1: Figure S1. Phylogenetic trees of the six gene families of complex expansions occurring in 11 and 12 species. Dana: *D. ananassae*, Dere: *D. erecta*, Dgri: *D. grimshawi*, Dmoj: *D. mojavensis*, Dper: *D. persimilis*, Dpse: *D. pseudoobscura*, Dsec: *D. sechellia*, Dsim: *D. simulans*, Dvir: *D. virilis*, Dwil: *D. willistoni*, Dyak: *D. yakuba*. Genes without abbreviative species name are from *D. melanogaster*.

Additional file 2: Figure S2. Distribution of young duplicate genes on chromosomes. **(A)**-**(E)** Species-specific duplicates on chromosome 2L, 2R, 3L, 3R and X, respectively. (F) Lineage-specific duplicates of *D. pseudoobscura-D. persimilis* pair. **(G)**-**(K)** Duplicates of complex expansions occurring in 11 species on chromosome 2L, 2R, 3L, 3R and X, respectively. Black lines represent observations, while black bands and red lines (red dots) show confidence intervals and average numbers of genes in corresponding windows. Dmel: *D. melanogaster*. 2L: chromosome 2L, 2R: chromosome 2R, 3L: chromosome 3L, 3R: chromosome 3R.

Additional file 3: Table S1. Number of protein domains for young duplicate genes in species-specific expansions.

Additional file 4: Table S2. Structural domains detected in families corresponding to the dots excessively deviating from trend lines. **Table S3.** Versions of genome sequences and annotations for each species.

Additional file 5: Figure S3. Mean *Ka/Ks* ratios of paralog gene pairs vs. ortholog gene pairs. **(A)** *Ka/Ks* ratios of species-specific duplicates. **(B)** *Ka/Ks* ratios of lineage-specific duplicates. Black lines means the trend line of black dots and red lines represents trend lines with slope = 1.

Additional file 6: Table S4. Number of young duplicate gene families for three types of expansion.

Abbreviations
CDS: Nucleotide coding sequences; MYA: Million years ago.

References
1. Ohno S: *Evolution by gene duplication*. London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag; 1970.
2. Long M, Langley CH: **Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila.** *Science* 1993, **260**(5104):91–95.
3. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151–1155.
4. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization.** *Genetics* 2000, **154**(1):459–473.
5. Lynch M, O'Hely M, Walsh B, Force A: **The probability of preservation of a newly arisen gene duplicate.** *Genetics* 2001, **159**(4):1789–1804.
6. Walsh JB: **How often do duplicated genes evolve new functions?** *Genetics* 1995, **139**(1):421–428.
7. Wagner A: **The fate of duplicated genes: loss or new function?** *Bioessays* 1998, **20**(10):785–788.
8. Hughes AL: **The evolution of functionally novel proteins after gene duplication.** *Proc Biol Sci* 1994, **256**(1346):119–124.
9. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerate mutations.** *Genetics* 1999, **151**(4):1531–1545.
10. Stoltzfus A: **On the possibility of constructive neutral evolution.** *J Mol Evol* 1999, **49**(2):169–181.
11. Wang L, Si W, Yao Y, Tian D, Araki H, Yang S: **Genome-Wide Survey of Pseudogenes in 80 Fully Re-sequenced Arabidopsis thaliana Accessions.** *PLoS ONE* 2012, **7**(12):e51769.
12. Sun YC, Hinnebusch BJ, Darby C: **Experimental evidence for negative selection in the evolution of a Yersinia pestis pseudogene.** *Proc Natl Acad Sci* 2008, **105**(23):8097–8101.
13. Dykhuizen D, Hartl DL: **Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background.** *Genetics* 1980, **96**(4):801–817.
14. Rodriguez-Trelles F, Tarrio R, Ayala FJ: **Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene.** *Proc Natl Acad Sci USA* 2003, **100**(23):13413–13417.
15. Thornton K, Long M: **Excess of amino acid substitutions relative to polymorphism between X-linked duplications in Drosophila melanogaster.** *Mol Biol Evol* 2005, **22**(2):273–284.
16. Kuwabara PE, Labouesse M: **The sterol-sensing domain: multiple families, a unique role?** *Trends Genet* 2002, **18**(4):193–201.
17. Meisel RP, Hilldorfer BB, Koch JL, Lockton S, Schaeffer SW: **Adaptive evolution of genes duplicated from the Drosophila pseudoobscura neo-X chromosome.** *Mol Biol Evol* 2010, **27**(8):1963–1978.
18. Kelleher ES, Swanson WJ, Markow TA: **Gene duplication and adaptive evolution of digestive proteases in Drosophila arizonae female reproductive tracts.** *PLoS Genet* 2007, **3**(8):e148.
19. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: **Recent segmental duplications in the human genome.** *Science* 2002, **297**(5583):1003–1007.
20. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, *et al*: **A genome-wide comparison of recent**

chimpanzee and human segmental duplications. *Nature* 2005, **437**(7055):88–93.

21. Yang J, Huang J, Gu H, Zhong Y, Yang Z: **Duplication and adaptive evolution of the chalcone synthase genes of Dendranthema (Asteraceae).** *Mol Biol Evol* 2002, **19**(10):1752–1759.

22. Hernández-Hernández T, Martínez-Castilla LP, Alvarez-Buylla ER: **Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein–protein interaction domains after major gene duplication events.** *Mol Biol Evol* 2007, **24**(2):465–481.

23. Kinsella RJ, Fitzpatrick DA, Creevey CJ, McInerney JO: **Fatty acid biosynthesis in Mycobacterium tuberculosis: lateral gene transfer, adaptive evolution, and gene duplication.** *Proc Natl Acad Sci* 2003, **100**(18):10320.

24. Medzhitov R, Preston-Hurlburt P, Janeway CA Jr: **A human homologue of the Drosophila Toll protein signals activation of adaptive immunity.** *Nature* 1997, **388**(6640):394–397.

25. Hahn MW, Han MV, Han SG: **Gene family evolution across 12 Drosophila genomes.** *PLoS Genet* 2007, **3**(11):e197.

26. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN: **Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures.** *Nature* 2007, **450**(7167):219–232.

27. Singh ND, Larracuente AM, Sackton TB, Clark AG: **Comparative genomics on the Drosophila phylogenetic tree.** *Annu Rev Ecol Evol Syst* 2009, **40**:459–480.

28. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, *et al*: **A single p450 allele associated with insecticide resistance in Drosophila.** *Science* 2002, **297**(5590):2253–2256.

29. Li L, Memon S, Fan Y, Yang S, Tan S: **Recent duplications drive rapid diversification of trypsin genes in 12 Drosophila.** *Genetica* 2012, **140**(7–9):297–305.

30. Camporeale G, Zempleni J, Eissenberg JC: **Susceptibility to heat stress and aberrant gene expression patterns in holocarboxylase synthetase-deficient Drosophila melanogaster are caused by decreased biotinylation of histones, not of carboxylases.** *J nutr* 2007, **137**(4):885–889.

31. Morano KA: **New tricks for an old dog: the evolving world of Hsp70.** *Ann N Y Acad Sci* 2007, **1113**:1–14.

32. Powning R, Davidson W: **Studies on insect bacteriolytic enzymes–II Some physical and enzymatic properties of lysozyme from haemolymph of Galleria mellonella.** *Comp Biochem Physiol B Comp Biochem* 1976, **55**(2):221–228.

33. Blackman R, Spence J, Field L, Javed N, Devine G, Devonshire A: **Inheritance of the amplified esterase genes responsible for insecticide resistance in Myzus persicae (Homoptera: Aphididae).** *Heredity* 1996, **77**(2):154–167.

34. Hemingway J, Hawkes NJ, McCarroll L, Ranson H: **The molecular basis of insecticide resistance in mosquitoes.** *Insect Biochem Mol Biol* 2004, **34**(7):653–665.

35. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B: **Drosophila intestinal response to bacterial infection: activation of host defense and stem cell proliferation.** *Cell Host Microbe* 2009, **5**(2):200–211.

36. Weig A, Deswarte C, Chrispeels MJ: **The major intrinsic protein family of Arabidopsis has 23 members that form three distinct groups with functional aquaporins in each group.** *Plant Physiol* 1997, **114**(4):1347–1357.

37. Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD: **Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila.** *Nature* 2007, **445**(7123):82–85.

38. Tamura K, Subramanian S, Kumar S: **Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks.** *Mol Biol Evol* 2004, **21**(1):36–44.

39. Doolittle RF: **Convergent evolution: the need to be explicit.** *Trends Biochem Sci* 1994, **19**(1):15.

40. Hoffmann FG, Opazo JC, Storz JF: **Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates.** *Proc Natl Acad Sci U S A* 2010, **107**(32):14274–14279.

41. Lawn RM, Schwartz K, Patthy L: **Convergent evolution of apolipoprotein(a) in primates and hedgehog.** *Proc Natl Acad Sci U S A* 1997, **94**(22):11992–11997.

42. Nozawa M, Kawahara Y, Nei M: **Genomic drift and copy number variation of sensory receptor genes in humans.** *Proc Natl Acad Sci* 2007, **104**(51):20421–20426.

43. Nozawa M, Nei M: **Evolutionary dynamics of olfactory receptor genes in Drosophila species.** *Proc Natl Acad Sci* 2007, **104**(17):7122–7127.

44. Youngson RM: *Collins Dictionary of Human Biology.* Glasgow: HarperCollins; 2006:352.

45. Fambrough DM, Bonner J: **On the Similarity of Plant and Animal Histones*.** *Biochemistry* 1966, **5**(8):2563–2570.

46. Arendt J, Reznick D: **Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?** *Trends Ecol Evol* 2008, **23**(1):26–32.

47. Cardoso-Moreira M, Emerson JJ, Clark AG, Long M: **Drosophila duplication hotspots are associated with late-replicating regions of the genome.** *PLoS genetics* 2011, **7**(11):e1002340.

48. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW: **Adaptive evolution of young gene duplicates in mammals.** *Genome Res* 2009, **19**(5):859–867.

49. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: **Selection in the evolution of gene duplications.** *Genome Biol* 2002, **3**(2):8.1–8.9.

50. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH: **Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli.** *Plant physiology* 2008, **148**(2):993–1003.

51. Koch AL: **Evolution of antibiotic resistance gene function.** *Microbiol Rev* 1981, **45**(2):355–378.

52. Martinez JL: **Antibiotics and antibiotic resistance genes in natural environments.** *Science* 2008, **321**(5887):365–367.

53. Widholm JM, Chinnala AR, Ryu JH, Song HS, Eggett T, Brotherton JE: **Glyphosate selection of gene amplification in suspension cultures of 3 plant species.** *Physiol Plantarum* 2001, **112**(4):540–545.

54. Guillemaud T, Raymond M, Tsagkarakou A, Bernard C, Rochard P, Pasteur N: **Quantitative variation and selection of esterase gene amplification in Culex pipiens.** *Heredity (Edinb)* 1999, **83**(Pt 1):87–99.

55. Tabashnik BE: **Implications of gene amplification for evolution and management of insecticide resistance.** *J Econ Entomol* 1990, **83**(4):1170–1176.

56. Gottesman MM, Hrycyna CA, Schoenlein PV, Germann UA, Pastan I: **Genetic analysis of the multidrug transporter.** *Annu Rev Genet* 1995, **29**:607–649.

57. Chen L, DeVries AL, Cheng CH: **Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish.** *Proc Natl Acad Sci U S A* 1997, **94**(8):3811–3816.

58. Riehle MM, Bennett AF, Long AD: **Genetic architecture of thermal adaptation in Escherichia coli.** *Proc Natl Acad Sci USA* 2001, **98**(2):525–530.

59. Sonti RV, Roth JR: **Role of gene duplications in the adaptation of Salmonella typhimurium to growth on limiting carbon sources.** *Genetics* 1989, **123**(1):19–28.

60. Brown CJ, Todd KM, Rosenzweig RF: **Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment.** *Mol Biol Evol* 1998, **15**(8):931–942.

61. Lai CY, Baumann L, Baumann P: **Amplification of trpEG: adaptation of Buchnera aphidicola to an endosymbiotic association with aphids.** *Proc Natl Acad Sci U S A* 1994, **91**(9):3819–3823.

62. Tang H, Kambris Z, Lemaitre B, Hashimoto C: **Two proteases defining a melanization cascade in the immune system of Drosophila.** *J Biol Chem* 2006, **281**(38):28097–28104.

63. Levashina EA, Langley E, Green C, Gubb D, Ashburner M, Hoffmann JA, Reichhart JM: **Constitutive activation of toll-mediated antifungal defense in serpin-deficient Drosophila.** *Science* 1999, **285**(5435):1917–1919.

64. Gorman MJ, Paskewitz SM: **Serine proteases as mediators of mosquito immune responses.** *Insect Biochem Molec* 2001, **31**(3):257–262.

65. Li L, Memon S, Fan Y, Yang S, Tan S: **Recent duplications drive rapid diversification of trypsin genes in 12 Drosophila.** *Genetica* 2012, **140**(7-9):297–305.

66. Wu DD, Wang GD, Irwin DM, Zhang YP: **A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito.** *Mol Biol Evol* 2009, **26**(10):2333–2341.

67. Zhang J, Zhang YP, Rosenberg HF: **Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey.** *Nat Genet* 2002, **30**(4):411–415.

68. Werck-Reichhart D, Feyereisen R: **Cytochromes P450: a success story.** *Genome Biol* 2000, **1**(6):3003.1.

69. Guengerich FP: **Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity.** *Chem Res Toxicol* 2001, **14**(6):611–650.

70. Omiecinski CJ, Remmel RP, Hosagrahara VP: **Concise review of the cytochrome P450s and their roles in toxicology.** *Toxicol Sci* 1999, **48**(2):151–156.

71. Wilson TG: **Cyromazine toxicity to Drosophila melanogaster (Diptera: Drosophilidae) and lack of cross-resistance in natural population strains.** *J Econ Entomol* 1997, **90**(5):1163–1169.

72. Wilson TG: **Genetic evidence that mutants of the methoprene-tolerant gene of Drosophila melanogaster are null mutants.** *Arch Insect Biochem* 1996, **32**(3–4):641–649.

73. da Fonseca RR, Antunes A, Melo A, Ramos MJ: **Structural divergence and adaptive evolution in mammalian cytochromes P450 2C.** *Gene* 2007, **387**(1–2):58–66.

74. Tibbles L, Woodgett J: **The stress-activated protein kinase pathways.** *Cell Mol Life Sci* 1999, **55**(10):1230–1254.

75. Sato S, Sanjo H, Takeda K, Ninomiya-Tsuji J, Yamamoto M, Kawai T, Matsumoto K, Takeuchi O, Akira S: **Essential function for the kinase TAK1 in innate and adaptive immune responses.** *Nat Immunol* 2005, **6**(11):1087–1095.

76. Miller MK, Bang ML, Witt CC, Labeit D, Trombitas C, Watanabe K, Granzier H, McElhinny AS, Gregorio CC, Labeit S: **The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules.** *J Mol Biol* 2003, **333**(5):951–964.

77. Yan J, Wang J, Zhang H: **An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism.** *Plant J* 2002, **29**(2):193–202.

78. Schlenke TA, Begun DJ: **Natural selection drives Drosophila immune system evolution.** *Genetics* 2003, **164**(4):1471–1480.

79. Maroni G, Wise J, Young J, Otto E: **Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster.** *Genetics* 1987, **117**(4):739–744.

80. Otto E, Young JE, Maroni G: **Structure and expression of a tandem duplication of the Drosophila metallothionein gene.** *Proc Natl Acad Sci USA* 1986, **83**(16):6025–6029.

81. Findlay GD, Yi X, Maccoss MJ, Swanson WJ: **Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating.** *PLoS Biol* 2008, **6**(7):e178.

82. Makino T, Kawata M: **Habitat variability correlates with duplicate content of Drosophila genomes.** *Mol Biol Evol* 2012, **29**(10):3169–3179.

83. Weaver BAA, Cleveland DW: **Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death.** *Cancer Cell* 2005, **8**(1):7–12.

84. Lumeng C, Phelps S, Crawford GE, Walden PD, Barald K, Chamberlain JS: **Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases.** *Nat Neurosci* 1999, **2**(7):611–617.

85. Valiente M, Andres-Pons A, Gomar B, Torres J, Gil A, Tapparel C, Antonarakis SE, Pulido R: **Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases.** *J Biol Chem* 2005, **280**(32):28936–28943.

86. Ekengren S, Hultmark D: **A family of Turandot-related genes in the humoral stress response of Drosophila.** *Biochem Bioph Res Co* 2001, **284**(4):998–1003.

87. Ekengren S, Tryselius Y, Dushay MS, Liu G, Steiner H, Hultmark D: **A humoral stress response in Drosophila.** *Curr Biol* 2001, **11**(9):714–718.

88. Frazier AE, Dudek J, Guiard B, Voos W, Li Y, Lind M, Meisinger C, Geissler A, Sickmann A, Meyer HE: **Pam16 has an essential role in the mitochondrial protein import motor.** *Nat Struct Mol Biol* 2004, **11**(3):226–233.

89. Kelley WL: **The J-domain family and the recruitment of chaperone power.** *Trends Biochem Sci* 1998, **23**(6):222–227.

90. Qin W, Neal SJ, Robertson RM, Westwood JT, Walker VK: **Cold hardening and transcriptional change in Drosophila melanogaster.** *Insect Mol Biol* 2005, **14**(6):607–613.

91. Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL: **Selective sweep of a newly evolved sperm-specific gene in Drosophila.** *Nature* 1998, **396**(6711):572–575.

92. Inomata N, Yamazaki T: **Adaptive evolution at the molecular level of the duplicatedAmy gene system inDrosophila.** *J Genet* 1996, **75**(1):125–137.

93. McQuilton P, St Pierre SE, Thurmond J: **FlyBase 101--the basics of navigating FlyBase.** *Nucleic acids res* 2012, **40**(Database issue):D706–D714.

94. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al*: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947–2948.

95. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**(10):2731–2739.

96. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**(4):406–425.

97. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**(D1):D290–D301.