Genetica (2012) 140:297–305
DOI 10.1007/s10709-012-9682-5

Recent duplications drive rapid diversification of trypsin genes in 12 *Drosophila*

Luolan Li · Shabana Memon · Yuanchu Fan · Sihai Yang · Shengjun Tan

Received: 22 May 2012 / Accepted: 7 September 2012 / Published online: 18 September 2012

Abstract Trypsin participates in many fundamental biological processes, the most notably in digesting food. The 12 species of *Drosophila* provide a great opportunity to analyze the duplication pattern of trypsins and their association with dietary changes. Here, we find that the trypsin family expands dramatically after speciation. The duplication events are strongly related to the host preferences, with significantly more copy numbers in species breeding on rotting fruits. Temporal analysis of the duplication events indicates that the occurrences of these events are not simultaneous, but rather correlate to the ecological change or host shift. Furthermore, we find that the specialists and generalists have different adaptive selections, which is revealed by dynamic duplication and/or deletion and relatively high *Ka/Ks* values on the duplicated events in specialists. Our findings suggest that the duplication of trypsin genes has played an important role in the adaption of *Drosophila* to the diverse ecosystems.

Electronic supplementary material The online version of this article (doi:10.1007/s10709-012-9682-5) contains supplementary material, which is available to authorized users.

L. Li · S. Memon · S. Yang (✉) · S. Tan (✉)
State Key Laboratory of Pharmaceutical Biotechnology,
Department of Biology, School of Life Sciences, Nanjing
University, 22 Hankou Rd., Nanjing 210093, China
e-mail: sihaiyang@nju.edu.cn

S. Tan
e-mail: clamp131@gmail.com

Y. Fan
First Hospital, Peking University, Beijing 100871, China

S. Tan
Key Laboratory of Zoological Systematics and Evolution,
Institute of Zoology, Chinese Academy of Sciences,
Beijing 100101, China

Keywords Trypsin · Duplication · Deletion · Food resources · Adaptive evolution

Introduction

Trypsin is one of the largest families of serine proteases. It emerged early in evolution and is ubiquitous in both vertebrates and invertebrates. This family takes part in many fundamental biological processes, these include the synthesis of melanin (Tang et al. 2006) and antimicrobial peptide (Levashina et al. 1999), hemolymph coagulation (Iwanaga et al. 1998), and the activation of rapid immune response to antigens (Gorman and Paskewitz 2001). Among these functions, the most prominent role is played in the digestion of food by recognizing and hydrolyzing peptide bonds (Rawlings and Barrett 1994). It becomes the most abundant protease in the invertebrate digestive system (Muhlia-Almazan et al. 2008).

The evolutionary process of ecological adaptation, in which selective forces may play important roles, is crucial for the survival of insects. In *Drosophila*, olfactory and gustatory receptors are responsible for identifying and locating food. Studies of these two families revealed a strong correlation between the expansion and contraction of these gene families and adaptation to host specialization for species (McBride 2007). Therefore, it can also be expected that other functional related protein (such as digestive proteases), could exhibit similar patterns of molecular evolution. Indeed, a study in pepsins found a number of gene duplication and loss events, and each round of gene duplication is characterized by adaptive evolution (Carginale et al. 2004).

How the trypsin family has evolved has attracted tremendous interest. Early researches suggested that the trypsin family evolved both in complexity and size in different organisms (Patthy 1999; Zdobnov et al. 2002). Comparisons of the trypsin family in fruit fly and mosquitoes uncovered dramatic changes in gene family size, indicating that the expansion of trypsin genes contributed to the adaptation of mosquitoes to their hematophagous trait (Wu et al. 2009). Similar studies in a leaf-eating monkey revealed that the expansion of trypsin genes have contributed to an adaptation to the specific diet (Zhang et al. 2002). These studies indicate the vital role of trypsin family in adaptive evolution for different species. However, because of lacking whole-genome sequencing data, there has been no comprehensive investigation of changes in the copy number variation of the trypsin family in closely related species.

The release of 12 worldwide *Drosophila* genomes has dramatically facilitated investigations into whether an adaptive evolutionary role exists for this protein family. The *Drosophila* genus captures a range of evolutionary distances, from closely related sibling species pairs, such as *D. pseudoobscura* and *D. persimilis*, to more distantly related species, *D. yakuba* and *D. grimshawi*. Additionally, there are the cosmopolitan species, *D. melanogaster* and *D. simulans,* as well as species with highly restricted geographic ranges for example *D. sechellia* (Singh et al. 2009). There are also three specialist species: *D. erecta* (Lachaise et al. 1988), *D. mojavensis* (Pfeiler and Markow 2001) and *D. sechellia* (R'Kha et al. 1991). More importantly, they have diverse host preferences corresponding to different yeast communities (Singh et al. 2009).

In order to understand the contribution of the trypsin family in adaptation to dietary changes, we analyzed the evolutionary processes of trypsins with respect to host preference and geographic range in 12 species of *Drosophila*. We identified ~200 copies of the trypsin genes in each species and constructed a phylogenetic tree to detect gene duplication and deletion events. We also performed statistical tests on the relationship between duplication events and host preference or geographic range. Our findings suggested a strong correlation between the expansion of the trypsin family and a host shift from decaying trees to rotting fruits. At the same time, our results provided insight into the ecological forces behind its adaptive evolution.

Materials and methods

Identification of the trypsin family genes

To obtain the trypsin family genes, the coding sequences (CDSs) of 12 *Drosophila* species were downloaded from FlyBase databases (http://flybase.org/, Drosophila 12 Genomes Consortium 2007), and the trypsin domain (PF00089) from Pfam website (http://pfam.janelia.org/). Using the consensus sequence of trypsin domain as a query, a TBLASTN search was then performed against all the CDSs of 12 *Drosophila* in Bioedit v7.0.5 (Hall 1999). The threshold expectation value was set to 10, and the other numerical options were left at default values. The obtained candidate sequences were further surveyed to confirm whether they encoded trypsin motif using the Pfam database v24.0 (Punta et al. 2011).

Construction of phylogenetic tree and determination of duplication or deletion events

Multiple sequence alignment of the predicted trypsin genes was performed using the ClustalW program in MEGA 4 (Kumar et al. 2007) with default parameters. A phylogenetic tree based on nucleotide sequences was constructed using the Neighbor-Joining (NJ) method. NJ analysis was done using p-distance (or Poisson Correction methods) and pairwise deletion of gaps. Support for each node was tested with 1,000 bootstrap replicates.

The phylogenetic tree was first divided into several clades based on one criterion: a clade should contain at least one species of both the *Sophophora* and *Drosophila* subgenera (Hahn et al. 2007). Therefore each clade could be assumed a copy present in the most recent common ancestor (MRCA). The duplication or deletion events were then counted manually in each clade on the tree. If there was no orthologous trypsin gene for a certain species in a given clade, we inferred that there was a deletion event in this species (Fig. S1, clades labeled with '−'). The duplication events were counted as the paralog number minus one in a given clade for each species. In order to get a conserved and accurate estimate, we only counted the events within the branches with a bootstrap values of more than 50 (Fig. S1, clades labeled with '+'). The duplication events could be further classified into two groups: species-specific (copies individually duplicated in one specific species) and group-specific duplication (copies duplicated in several species).

In order to group the *drosophila* species with respect to the pattern of duplications or deletions, we created two phylogenetic trees based on the discrete morphology (parsimony) method in the program PARS of the PHYLIP package v3.6 (written by J. Felsenstein; available at: http://evolution.genetics.washington.edu/phylip.html). During the construction of duplication tree, we defined two states for species in each clade, "1" if having at least one duplication event, and "0" if none. The deletion tree was constructed in a similar way.

Detection of the age of duplication events and selective forces

To understand the process of duplication in trypsin family, the non-synonymous rate (*Ka*) and the synonymous rate (*Ks*) were computed pairwise for each duplication event in each clade based on the multiple sequence alignments in

MEGA. The *Ks* values were used to roughly represent the age of the duplication events. The selective pressures were estimated by the ratios of non-synonymous (*Ka*) to synonymous (*Ks*) nucleotide substitutions rates, also known as *Ka/Ks*, on the CDSs of the duplicated genes. A positive selection was indicated if $Ka/Ks > 1$, while purifying selection if $Ka/Ks \ll 1$. To confirm the positive selection, we also calculated the *Ka/Ks* values by "codeml" as implemented in PAML (Yang 1997).

Results

Evolutionary pattern of duplication and/or deletion events

Using the amino acid sequence of trypsin domain as a query, the TBLASTN search was performed in 12 whole-genome sequenced *Drosophila* genomes. In total, 2,598 trypsin copies were identified and used to construct the combined phylogenetic tree (Fig. S1). The copy numbers of trypsins in the 12 *Drosophila* genomes varied from 167 in *D. mojavensis* to 258 in *D. yakuba* (Fig. 1), averaging on 217. The phylogenetic tree was divided into 154 clades (Fig. S1). Within each clade, the trypsin genes were assumed to have evolved from the same ancestor. Therefore, the most recent common ancestor (MRCA) of *Drosophila* which lived ~60 million years ago might have 154 trypsin genes, and after differentiation, this family in all 12 species has undergone or is undergoing frequent gene duplication and/or deletion. The duplication or deletion events in each species were then counted for each clade (see "Materials and Methods" for more details). Within the 154 MRCA copies, 80 were duplicated in at least one species and 31 in 6 species or more. The deletion events were less frequent but still affected 45 copies.

Due to the frequent duplications and/or deletions, we explored whether there was a relationship between duplication or deletion of trypsins and some factors, for instance the evolutionary phylogeny of *Drosophila*. For the deletion events, a phylogenetic tree was constructed based on the presence or absence of deletion events in 154 clades for each species. Interestingly, there was no relationship observed between deletion events and their species phylogeny (Fig. 2a). This indicates deletion events may occur randomly in the evolutionary history of *Drosophila* species. On the contrary, a significant consistency was observed between duplication events and their species phylogeny (Fig. 2b). For example, the closely related species, such as *D. pseudoobscura* and *D. persimilis* or *D. yakuba* and *D. erecta*, had similar duplicated copies. However, *D. willistoni*, which was the first one splitting from the ancestor, had a much different pattern with the other *Sophophora* subgenus members. On this tree, *D. grimshawi* had the most diverse duplication pattern, and this is likely due to its Hawaiian endemic.

Duplication events associated with food resources of *Drosophila*

On the combined tree of trypsin genes, we could also distinguish how many times the duplication events occurred in each species. 56 duplication events were detected in each species on average, ranging from 30 in *D. mojavensis* to 80 in *D. simulans* (Fig. 1). It was postulated that the duplication of trypsin genes could be associated with the

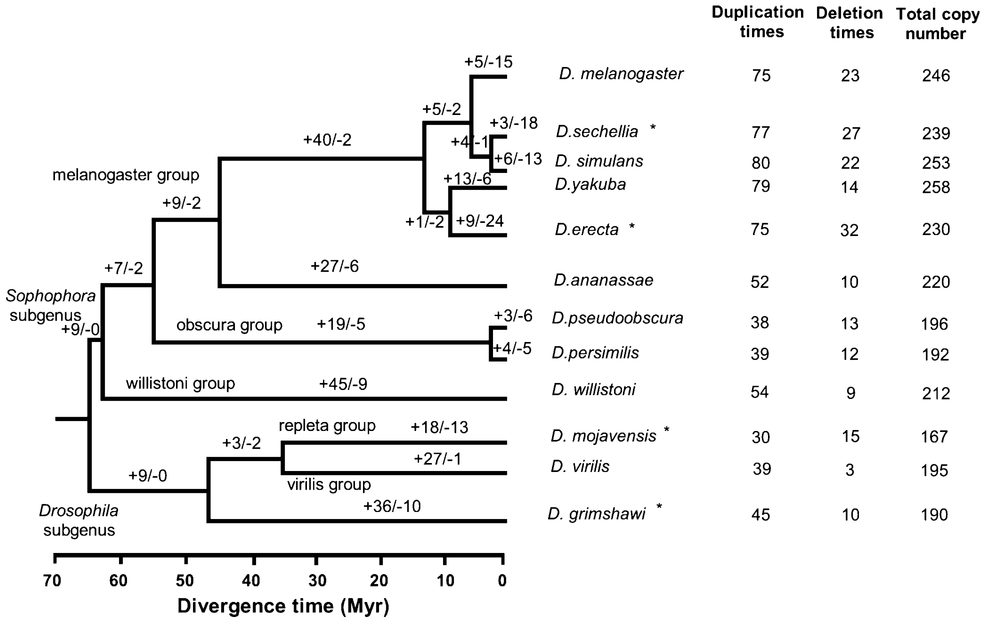

Fig. 1 The duplication and deletion times occurred in each species. On each branch of the tree the number of duplication/deletion events is given. The specialists are labeled with '*': *D. sechellia* is endemic to the Seychelles Islands and breeds in *M. citrifolia*, *D. erecta* is an African Drosophilid that specializes on *P. candelabrum*, *D. mojavensis* is a cactophilic, and *D. grimshawi* is a Hawaiian endemic

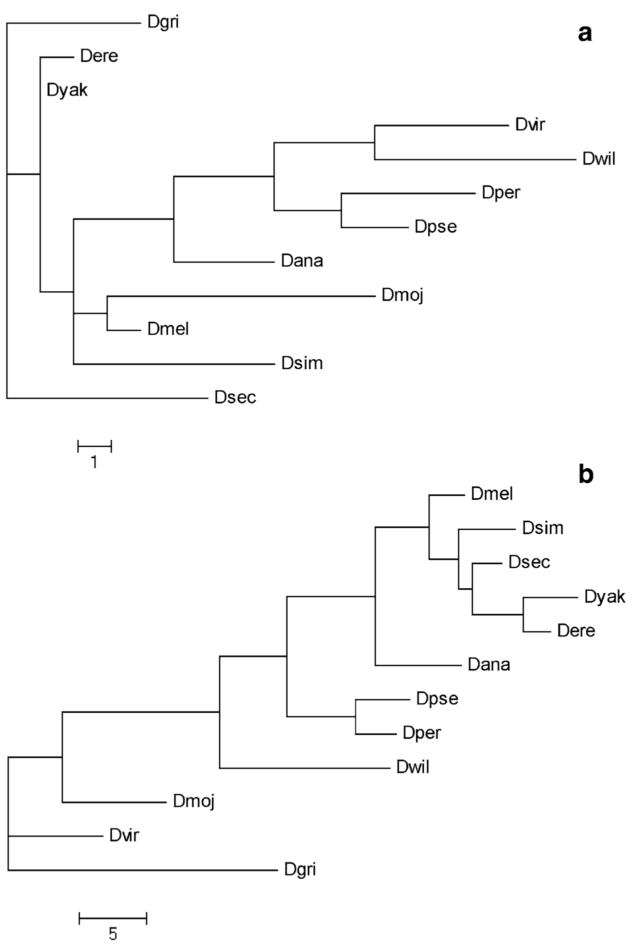

Fig. 2 Grouping of fruitflies according to the absent (**a**) or duplicated (**b**) pattern of trypsin family genes. The trees were created based on the discrete morphology method using the programs PARS of the PHYLIP package v3.6

adaptation to different food resources because the function of trypsin genes was related to the digestion of food. There are 2 major kinds of resources for the 12 *Drosophila* to breed on: rotting fruits and/or decaying trees (Singh et al. 2009). Notably, there were significantly more duplications in the species which breed on rotting fruits than those on decaying trees ($P < 0.01$, Fig. 1). There are two specialist species, *D. sechellia* and *D. erecta*, which evolved to specialize only on specific fruits, *Morinda citrifolia* and *Pandanus candelabrum* respectively (Fig. 1). However, their duplication events had no difference to the generalist species breeding on rotting fruits in our research ($P = 0.92$, *t* test).

In the *Sophophora* subgenus, *D. pseudoobscura* and *D. persimilis* can breed on both decaying trees and rotting fruits, and have less duplication events than those only breeding on rotting fruits. In the *Melanogaster* group, five species showed a distinctly excessive number of duplication events (averaging 77), half of which were accumulated in their ancestor after splitting with *D. ananassae* (Fig. 1). Within the species breeding on decaying trees, the Cactophilic *D. mojavensis* had less duplication times though not significantly ($P = 0.45$, *t* test). In contrast to this pattern, we observed significantly more duplications in the African species in comparison to their America or Asia relatives ($P < 0.01$, *t* test). These results suggest that food resources could be important incentive factor for the expansion of trypsin genes.

Temporal evolutionary patterns of the duplication events

The expansion of trypsin genes is a gradually accumulated process, and the duplication events could exhibit different patterns across the evolutionary history of *Drosophila*. The synonymous rate (*Ks*) of the emerging time of duplication events was calculated and used to estimate the relative age of gene duplications.

The results revealed that the average *Ks* value of the species-specific was 0.30, much smaller than 1.45 of the group-specific duplications, which arise in the ancestor of at least two species (Table S2). Then the number of duplication events for each species was counted in each *Ks* range (from 0 to 2 with an interval of 0.1, Table S1). These events were distributed unevenly among different *Ks* range. In $Ks < 0.1$ group, *D. grimshawi* had significantly excessive duplication events ($P < 0.01$, *t* test; Fig. 3a), while *D. sechellia* had none. *D. grimshawi* is an endemic living in Hawaii. Half of the duplication events had very small *Ks* values, indicating these novel events could be consequences of adapting to the recently changed ecological niche. In $Ks > 1.0$ group, 5 of 6 species in the *Melanogaster* subgroup (except *D. ananassae*) had more duplication events than the others ($P < 0.01$, *t* test; Fig. 3b). More than 77 % duplications arise in the ancestor of these five species after it split from *D. ananassae* (Fig. 1). This indicates these events could be ancient and retained in descendants.

To evaluate the different patterns of duplication events at different time among species, we compared three species from different groups: *D. ananassae*, *D. willistoni* and *D. grimshawi*. All of these species diverged at least 42 million years ago. They accumulated a large number of species-specific duplication events. Obviously, these species displayed different duplication peaks during their evolutionary history: *D. grimshawi* had excessive duplications recently, which was associated to the adaptation to Hawaii environment; *D. willistoni* had a stable rise of duplication events when $Ks < 0.7$; while *D. ananassae* had the most constant duplication rate (Fig. 3c).

Amino acid substitutions have a strong effect on gene expression or change of function and can reflect the

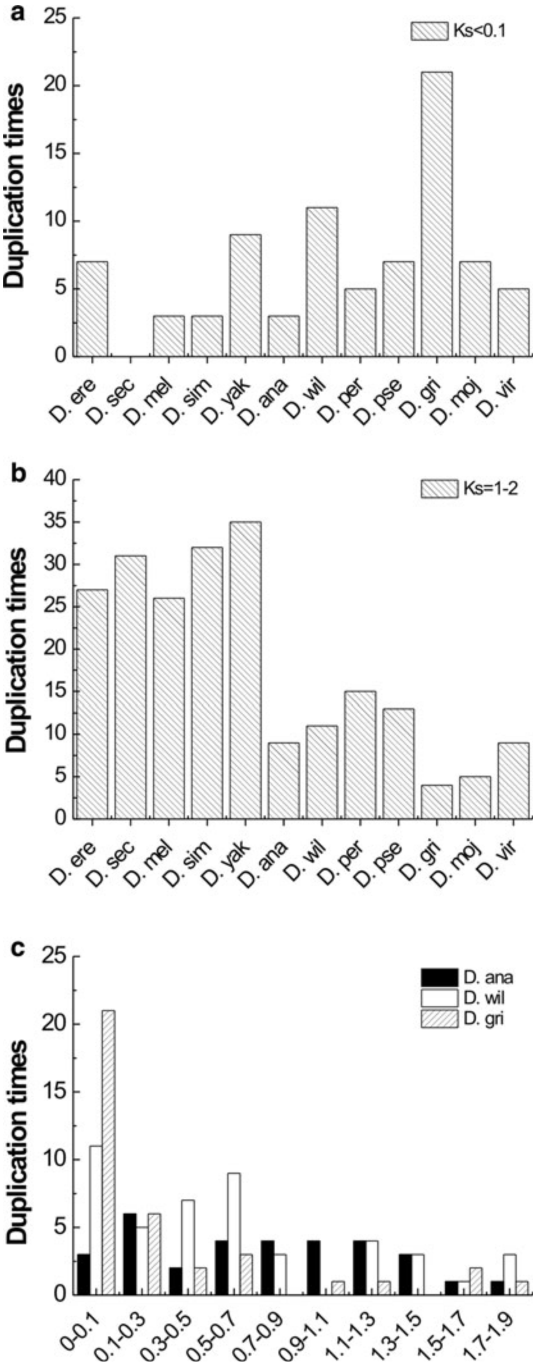

Fig. 3 The duplication times during different species (**a**, **b**) and *Ks* ranges (**c**)

selective forces imposing on evolving genes. We found that there was a strong linear relationship between *Ka* and *Ks* values for each species (Table 1, $P < 0.0001$), and the average value of the linear regression slopes was 0.29, much smaller than 1. This result indicates that paralogs stably diverged from each other at a low rate. The *Melanogaster* group had a lower slope value (0.25) while the obscura group had a relatively high value (0.35). This suggests the evolutionary rates of the duplicated genes were different between species. The specialists did not exhibit significantly higher *Ka/Ks* on the long time-scale. For example, 0.307 in *D. erecta* and 0.278 in *D. yakuba*, or 0.331 in *D. mojavensis* and 0.325 in *D. virilis*. When comparing the group-specific and species-specific duplications, however we found the latter were all higher than the former. This suggests that the trypsin genes of specialists were fast evolving and under relaxed selection constraint after functionally fixation.

Selective forces on duplicated trypsin genes

There is compelling evidence that directional selection has played an important role in speciation (Rieseberg et al. 2002). To detect the selective pressure, we examined the ratio of non-synonymous to synonymous substitutions (*Ka/Ks*) for each duplication event. Interestingly, the *Ka/Ks* values had significant differences between group-specific and species-specific duplications ($P < 0.05$, *t* test). The *Ka/Ks* values for group-specific duplications were smaller (averaging 0.35), indicating these duplicated copies were functional and experienced purifying selection. The *Ka/Ks* for species-specific duplications were elevated (averaging 0.41), reflecting a comparative relaxation of purifying selection. Moreover, we detected 11 species-specific duplication events under positive selection ($Ka/Ks > 1$). For example, in *D. erecta*, we found three *D. erecta* -specific duplication events with $Ka/Ks > 1$ (3.66, 2.01 and 1.19) in $Ks < 0.1$ range (0.022, 0.047 and 0.067, respectively), as well as another one in $Ks = 0.1$–0.2 group ($Ka/Ks = 1.42$ in $Ks = 0.15$; Table S2). The *Ka/Ks* values calculated by PAML also supported the existence of these positive selections with $Ka/Ks > 1$. In Hawaii specialist *D. grimshawi*, four duplication events were found under positive selection in $Ks < 0.1$ group (Table S2) and two of these *Ka/Ks* values were validated >1 by PAML. In total, among the 11 events under positive selection, 9 were in the specialist species (Table S2). For *D. ananassae*, the only 2 pair of species-specific duplicated paralogs showed tremendous divergence, and consequently did not have a calculable *Ka* and *Ks* values. This may be the result of a strong diversified selection.

In the 12 *Drosophila*, there are four specialists: *D. sechellia* and *D. erecta*, which only breed on certain plants; *D. mojavensis*, a cactophilic; *D. grimshawi*, a Hawaiian endemic. No similar evolutionary patterns were observed in these duplication events, while all of them had significantly more deletion events compared with generalists. For instance, *D. mojavensis* is a cactophilic species native to the Southwestern United States and has undergone the most dramatic gene duplication and deletion events (18 and 13, Fig. 1), significantly different to *D. virilis*

Table 1 The linear relationship between *Ka* and *Ks* values for duplicated trypsin genes in the 12 *Drosophila* genomes

Species	Slope	R^2	*P* value	*Ka/Ks*	Host preference	Range
D.ere	0.307	0.611	1.17E − 12	0.498	Rotting fruits (*P. candelabrum*)	Africa
D.sec	0.198	0.254	2.81E − 05	0.338	Rotting fruits (*M. citrifolia*)	Africa (island)
D.yak	0.278	0.607	4.44E − 16	0.346	Rotting fruits	Africa
D.wil	0.333	0.699	2.56E − 14	0.354	Rotting fruits	America
D.ana	0.229	0.596	2.06E − 09	0.316	Rotting fruits	Asia & Pacific
D.mel	0.244	0.531	7.70E − 12	0.302	Rotting fruits	Cosmopolitan
D.sim	0.276	0.448	1.97E − 10	0.339	Rotting fruits	Cosmopolitan
D.gri	0.227	0.843	1.11E − 16	0.432	Decaying trees (Hawaiian)	America (island)
D.moj	0.331	0.659	1.58E − 06	0.441	Decaying trees (cacti)	America
D.vir	0.325	0.889	3.53E − 13	0.336	Decaying trees	Holarctic
D.per	0.354	0.652	2.77E − 09	0.43	Both	America
D.pse	0.344	0.709	9.30E − 10	0.401	Both	America

($P < 0.01$). The same phenomenon was observed between *D. grimshawi* and *D. virilis* ($P < 0.05$), as well as between *D. erecta* and *D. yakuba* ($P < 0.01$). *D. sechellia* also had more deletion events than *D. simulans*, though the difference was not statistically significant (18 vs 13, Fig. 1). The specialists all had special environments and narrow food resources; therefore presumably there was no need for them to preserve many copies of trypsin genes. The purifying selection could be the major force for the gene deletion events.

Discussion

Duplication of trypsin genes associated with food sources contributes to the adaptive evolution of *Drosophila*

Trypsin genes have two main functions, hydrolysis of food protein and activation of zymogenes of digestive proteases (Amino et al. 2001). Their ability to digest protein must have emerged early in evolution, since the trypsin coding genes are also found in eubacterial genomes (Rypniewski et al. 1994). Both the size and the complexity of trypsin family are significantly different between organisms (Zdobnov et al. 2002), presumably due to the functional specialization forced by selective pressures (Wu et al. 2009). In the 12 *Drosophila* species, we detected approximately 200 copies of trypsin genes in each genome, changing from 167 to 258, and varied ~1.5-fold. Through a phylogenetic analysis, we inferred 154 trypsin copies in their MRCA. After the species split, this multi-gene family underwent frequent expansions and contractions in all lineages at different historic stages.

Moreover, the duplication pattern strongly correlated to the species phylogeny (Fig. 2b). The closely related species, such as *D. pseudoobscura* and *D. persimilis* or *D. yakuba* and *D. erecta*, have similar duplicated copies. *D. grimshawi*, which differentiated early and is a Hawaiian endemic, had the most special duplicated copies to others. On the other hand, there was evidence that food resources played an important role in speciation of *Drosophila*. Reproductive isolation in *D. pseudoobscura* using different food types, starch and maltose demonstrated that the isolation was due solely to the process of adaptation to the novel regimes (Dodd 1989). As the function of desisting food for trypsin proteins, our findings verified that duplications of trypsin genes could take part in *Drosophila* speciation.

The adaptation to abiotic and biotic environment was important in shaping the ecological diversity of the species. The role of chemoreception genes in the evolution of host specialization has been well studied (McBride 2007). Aside of the olfactory and gustatory organs, digestive system is also essential in response to specific ecological conditions. Correspondingly, the duplication events in trypsin family were widely reported in various species (Wu et al. 2009; Kelleher et al. 2007; Baptista et al. 1998), and expanded trypsin genes were found in adaptation to new food resources; for instance, new diet of blood meals in mosquito (Wu et al. 2009), and leaves in a leaf-eating monkey (Zhang et al. 2002). In support, we observed that duplication events were significantly more frequent in species breeding on rotting fruits than those on decaying trees ($P < 0.01$), suggesting more nutrients in rotting fruits.

Duplications of trypsin genes and the radiation patterns of *Drosophila*

The early *Drosophila* experienced habitats of a relatively "poorer" source of nutrients like decaying plants (Starmer 1981), and breeding on fruits was an adaptive way to radiate. *Ks* is widely used as a reliable parameter to predict the emerging time of duplication events because it is unaffected by selection and constantly increases over time (Chung et al. 2006; Gu et al. 2002; Li et al. 2005; Liao and Zhang 2006). Our data show that out of the 12 *Drosophila*, five *melanogaster* species had significant excess of trypsin duplications compared to the others ($P < 0.01$, *t* test; Fig. 3b) in $Ks > 1.0$ group ($Ks = 1.13$ on average). This suggests that the age of duplications was quite old in these five *melanogaster*. These five had an Afro tropical origin and they colonized the rest of the world only recently (Lachaise et al. 1988). The demographic changes were thought fixed by numerous beneficial mutations, as revealed by signatures of directional selection in drosophila genomes (Stephan and Li 2007). Consistently, we found more than 77 % duplication events happened in their ancestor, suggesting that duplications of trypsin genes might be one of the steps for the ancestor of *Drosophila* out of Africa.

On the contrary, *D. willistoni* and *D. ananassae* revealed no apparent high rate of duplications in the early stage. However, they both had a significant excess of recent duplications (Fig. 3c). This indicates that they had a recent adaption to the new niches. The duplication events in *D. grimshawi* were even younger since there existed a significant excess when $Ks < 0.1$ (Fig. 3a), which was consistent with the recent formation of Hawaiian rainforests sustaining *Drosophila* (DeSalle 1992). *D. sechellia* was also an island endemic but had no recent duplications ($Ks < 0.1$). The different duplication patterns between them may be due to the different spectrums of food resources. *D. sechellia* is a specialist on the fruit of *P. candelabrum* (Lachaise et al. 1988), while *D. grimshawi* could breed on seven families of endemic Hawaiian plants (Magnacca and O'Grady 2006). The less duplication events in *D. sechellia* may have resulted from an adaptation to the narrow food resources. Although we could not infer the exact time of the duplication events, the trypsin genes were duplicated at different periods among species. For the relationship between duplications of trypsin genes and adaption to the ecologies, our results could be additional evidences for predicting the biogeographic radiation patterns of *Drosophila*.

Adaptive selection explains differences between specialists and generalists

There are two major distinct features in four specialists (*D. sechellia, D. erecta, D. mojavensis* and *D. grimshawi*) compared to the generalists: dynamic duplication and/or deletion occurrences and relatively high *Ka/Ks* values on the duplicated events. The four specialists all had more than 10 deletion events after splitting from their sister-species. *D. sechellia* and *D. erecta* have evolved to specialize exclusively on *M. citrifolia* and *P. candelabrum* respectively, and both of them had a high number of gene deletion but little duplication (Fig. 1). All the trypsin genes were duplicated/deleted dramatically in the specialists than in their sister generalist lineages, except for those in *D. sechellia* ($P = 0.25$), which is most likely due to their short divergence time (0.93 million years). Similar observation was also reported in the analysis of olfactory and gustatory gene families (McBride et al. 2007; McBride 2007). Both the demography and ecology could be responsible for the patterns observed in the evolution of chemoreceptor genes (Gardiner et al. 2008) in *Drosophila*. The narrower food resources could be the major reason for contraction of these gene families. *D. sechellia* and *D. erecta* preferentially lost bitter-taste receptors, and this is the critical step in the evolution of their host preference (Matsuo et al. 2007). In the lifecycle of *Drosophila*, food resources possessed an important role. Consequently, the related gene families, for instance those responsible for detecting the chemical stimuli and digesting food, exhibited strong evidences of adaptive evolution related to shift of host. Combined with our results, all three gene families, olfactory, gustatory and trypsin, had a convergent evolution, especially for the specialists, demonstrating the important role played by food resources in species-specific adaptation and specialization.

The *Ka/Ks* value was also used to detect the selective forces on the duplicated trypsin genes. When plotting *Ka* against *Ks* values of all duplication events in 12 species (Table 1), we found that the slopes of the linear regressions were all smaller than 1. The slopes represent the average selective pressure on trypsin genes in each of the 12 species. Therefore over the long-time scale of evolution, the duplicated trypsin genes were generally under a constant purifying selection. However, the small value of slopes does not necessarily mean that the duplicated trypsin genes in specialists were under the same selective pressure as the generalists. It is also possible that a different selective pressure on newly duplicated genes may be masked by those copies duplicated long ago and have lost their role of adaptation to new environment.

Indeed, we found the average *Ka/Ks* value of species-specific duplication was higher than that of group-specific (0.41 vs 0.35; Table S2), while the average *Ks* value of the species-specific was much smaller (0.30 vs 1.45). This indicates that the overall level of selection pressure was changed during different stages of speciation and within different groups of duplications. Besides, evidence of positive selection was found in duplication events with $Ks < 0.1$ in some species. In total, 11 species-specific duplication events were detected under positive selection, among which 9 were in specialist or endemic. The positive selection detected in the newly-born duplication genes, especially among the specialists, suggested that the duplicated trypsins help adapting to the new environments. After their function stabilized, the role of these genes switched from adapting to the altered ecosystem to maintaining their present function. Accordingly their selective pressure thus became negative.

Although gene duplication is the major throughway of sub- and neo-functionalization for a species to adapt to the new environment, redundant genes are a great burden for an organism. Therefore, either gene deletion or fast-evolving diversification could be effective strategy to economize the cost. Considerably we found a greater number of gene deletion events and positive selection pressure on the newly duplicated genes in specialists were detected in support of this assumption. Our results demonstrated that specialists had a remarkably different evolutionary dynamics compared with their sister generalists, and both effective food resources and radiation incentives were highly associated with adaptation to the narrow ecologies.

Acknowledgments This work was supported by the National Natural Science Foundation of China (30970198, 0930008 and J1103512) and Qing Lan Project to S.Y. We thank Robin Burns for helpful discussions and suggestions on the manuscript, which significantly improved the writing.

Conflict of interest The authors declare that they have no competing interests.

References

Amino R, Tanaka AS, Schenkman S (2001) Triapsin, an unusual activatable serine protease from the saliva of the hematophagous vector of Chagas' disease Triatoma infestans (Hemiptera: Reduviidae). Insect Biochem Mol Biol 31(4–5):465–472

Baptista AM, Jonson PH, Hough E, Petersen SB (1998) The origin of trypsin: evidence for multiple gene duplications in trypsins. J Mol Evol 47(3):353–362

Carginale V, Trinchella F, Capasso C, Scudiero R, Riggio M, Parisi E (2004) Adaptive evolution and functional divergence of pepsin gene family. Gene 333:81–90

Chung WY, Albert R, Albert I, Nekrutenko A, Makova KD (2006) Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network. BMC Bioinf 7:46

DeSalle R (1992) The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. Mol Biol Evol 9(5):905–916

Dodd DMB (1989) Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. Evolution 43(6):1308–1311

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450(7167):203–218

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG (2008) Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Mol Ecol 17(7):1648–1657

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

Gu ZL, Nicolae D, Lu HHS, Li WH (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet 18(12):609–613

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

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Iwanaga S, Kawabata S, Muta T (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. J Biochem 123(1):1–15

Kelleher ES, Swanson WJ, Markow TA (2007) Gene duplication and adaptive evolution of digestive proteases in *Drosophila* arizonae female reproductive tracts. PLoS Genet 3(8):e148

Kumar S, Tamura K, Dudley J, Nei M (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the drosophila-melanogaster species subgroup. Evol Biol 22:159–225

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

Li WH, Yang J, Gu X (2005) Expression divergence between duplicate genes. Trends Genet 21(11):602–607

Liao BY, Zhang JZ (2006) Evolutionary conservation of expression profiles between human and mouse orthologous genes. Mol Biol Evol 23(3):530–540

Magnacca KN, O'Grady PM (2006) A subgroup structure for the Modified mouthparts species group of Hawaiian *Drosophila*. Proc Hawaiian Entomol Soc 38:87–101

Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y (2007) Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. PLoS Biol 5(5):e118

McBride CS (2007) Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc Natl Acad Sci USA 104(12):4996–5001

McBride CS, Arguello JR, O'Meara BC (2007) Five *Drosophila* genomes reveal non neutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177(3): 1395–1416

Muhlia-Almazan A, Sanchez-Paz A, Garcia-Carreno FL (2008) Invertebrate trypsins: a review. J Comp Physiol B 178(6): 655–672

Patthy L (1999) Genome evolution and the evolution of exon-shuffling—a review. Gene 238(1):103–114

Pfeiler E, Markow TA (2001) Ecology and population genetics of Sonoran Desert *Drosophila*. Mol Ecol 10(7):1787–1791

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2011) The Pfam protein families database. Nucleic Acids Res 40(D1):D290–D301

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

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. P Natl Acad Sci USA 99(19):12242–12245

R'Kha S, Capy P, David JR (1991) Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. Proc Natl Acad Sci USA 88(5):1835–1839

Rypniewski WR, Perrakis A, Vorgias CE, Wilson KS (1994) Evolutionary divergence and conservation of trypsin. Protein Eng 7(1):57–64

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

Starmer WT (1981) A comparison of *Drosophila* habitats according to the physiological attributes of the associated yeast communities. Evolution 35(1):38–52

Stephan W, Li H (2007) The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity (Edinb) 98(2):65–68

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

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

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13(5):555–556

Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM, Mueller HM, Dimopoulos G, Law JH, Wells MA, Birney E, Charlab R, Halpern AL, Kokoza E, Kraft CL, Lai Z, Lewis S, Louis C, Barillas-Mury C, Nusskern D, Rubin GM, Salzberg SL, Sutton GG, Topalis P, Wides R, Wincker P, Yandell M, Collins FH, Ribeiro J, Gelbart WM, Kafatos FC, Bork P (2002) Comparative genome and proteome analysis of Anopheles gambiae and *Drosophila melanogaster*. Science 298(5591):149–159

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