Rapid Functional Divergence of a Newly Evolved Polyubiquitin Gene in *Drosophila* and Its Role in the Trade-off between Male Fecundity and Lifespan

Zubing Zhan,[1,2] Yun Ding,[1,2] Ruopin Zhao,[1] Yue Zhang,[1] Haijing Yu,[1,3] Qi Zhou,‡,[1] Shuang Yang,[1] Hui Xiang,*,[1] and Wen Wang*,[1]

[1]CAS-Max Planck Junior Research Group, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

[2]Graduate University of the Chinese Academy Sciences, Beijing, China

[3]Laboratory for Conservation and Utilization of Bio-resources and Human Genetics Center of Yunnan University, Kunming, China

‡Present address: Department of Integrative Biology, University of California, Berkeley

***Corresponding author:** E-mail: wwang@mail.kiz.ac.cn; xiangh@mail.kiz.ac.cn.

Associate editor: Naoko Takezaki

Abstract

The cost of reproduction is a pivotal trade-off with various biological processes during the evolution of organisms. However, the genes and molecular mechanisms underlying the evolution of balancing reproductive capacity and its cost are still largely unknown. Here, we present a comprehensive study on the evolution, expression, and biological functions of a newly evolved pair of X-linked polyubiquitin tandemly duplicated genes, *CG32744* and *CG11700*, of which the duplication event occurred in *Drosophila melanogaster* lineage after the split from *D. simulans* clade. We found that *CG32744* retains conserved polyubiquitin-coding sequences across *Drosophila* species and is ubiquitously expressed, whereas *CG11700* has accumulated numerous amino acid changes and shows a male-specific expression pattern. Null mutants of *CG11700* have a higher male fecundity but shorter lifespan, whereas its overexpression decreases male fecundity. In contrast, the null mutants of the peptide-conserved *CG32744* do not exhibit such phenotypes. These results suggest that *CG11700* might have experienced neofunctionalization and evolved important functions in the trade-off between male fecundity and lifespan and that *CG32744* likely has retained the ancestral function.

Key words: duplicated polyubiquitin gene, *CG11700*, functional divergence, reproductive cost.

Introduction

Gene duplication and subsequent functional divergence have long been recognized as the most important mechanism for the origin of evolutionary novelties (Ohno 1970; Gu et al. 2002; Zhang 2003; Zhou et al. 2008). Functional divergence of duplicated genes that are preserved in genomes can lead to two alternative evolutionary fates: 1) subfunctionalization, in which duplicated genes, respectively, partition regulatory patterns or functions of the ancestral gene and 2) neofunctionalization, in which one copy acquires a novel function (Force et al. 1999; Lynch and Conery 2000; Long et al. 2003; Zhang 2003). So far, a number of new *Drosophila* duplicated genes that evolved novel functions have been reported, such as *jingwei* (Long and Langley 1993; Zhang et al. 2004), *Sdic* (Nurminsky, Nurminskaya, Benevolenskaya, et al. 1998; Nurminsky, Nurminskaya, De Aguiar, et al. 1998), *sphinx* (Wang et al. 2002; Dai et al. 2008), *ms(3)K81* (*K81*) (Loppin et al. 2005), and *nsr* (Ding et al. 2010). Interestingly, many of these new duplicated genes exhibit male-specific or even testis-specific expression, suggesting that male-specific expression might be a frequent fate of newly evolved genes (Long et al. 2003; Kaessmann et al. 2009). In addition, some of these duplicated genes with male-biased expression pattern are tandemly clustered on the X chromosome, such as *Sdic* (Nurminsky, Nurminskaya, Benevolenskaya, et al. 1998; Nurminsky, Nurminskaya, De Aguiar, et al. 1998) and *Tektins* (Dorus et al. 2008), which is consistent with the finding that there is an excess of young X-linked genes with male-biased expression pattern in *Drosophila* (Zhang et al. 2010). This observation does not readily follow the predictions of the existing models for chromosomal distribution of genes with male-biased expression, such as meiotic sex chromosome inactivation (MSCI) and sexual antagonism driving germ line X inactivation (SAXI) models (Wu and Xu 2003; Kaiser and Bachtrog 2010), and it indicates that new genes with male-biased expression are likely to locate on X chromosome initially but might escape from X to autosomes later (Zhang et al. 2010).

Reproductive cost is a pivotal trade-off with various biological processes during the evolution of organisms (Harshman and Zera 2007). Due to its fundamental importance, reproductive cost has been of interest to researchers from a variety of biological fields (Zera and Harshman 2001; Harshman and Zera 2007). Given that organisms have a limited pool of resources (Paukku and Kotiaho 2005), investment to reproduction usually reduces longevity, growth, future fecundity, etc. The reproductive cost for females can be attributed to several factors, including tolerating

harassment and courtship by males, copulation, harmful effects of male seminal products, producing eggs, pregnancy, giving birth, maternal care, and, as is most likely, some combination of these factors (Kotiaho and Simmons 2003). In males, reproductive cost may result from sperm production, courtship, copulation, parental care, etc. (Paukku and Kotiaho 2005). However, compared with females, reproductive cost in males has been much less explored. More importantly, the genes and mechanisms underlying the evolution of a balance between the reproductive capability and its cost are still largely unknown.

In this study, we investigated the evolution, expression, and biological functions of newly evolved pair of X-linked polyubiquitin tandemly duplicated genes, *CG11700* and *CG32744*. The duplication event occurred after the split of *Drosophila melanogaster* from the *D. simulans* clade. Both *CG32744* and *CG11700* have ubiquitin-encoding repeats. Despite that *CG11700* was annotated as a pseudogene in the recent FlyBase (*CR11700*: dmel_r5.2) due to a lack of expressed sequence tags (ESTs), the result of our gene-specific reverse-transcription polymerase chain reaction (RT-PCR) revealed that it is transcribed and has a potential open reading frame (ORF) of 906 bp. Ubiquitins are encoded by a multigene family. The peptide sequences are highly conserved in the ubiquitin gene family. They play important roles in both cellular stress responses and protein degradation in eukaryotes (Hershko and Ciechanover 1998; Kerscher et al. 2006). Interestingly, although peptide sequence is usually highly conserved in the ubiquitin gene family, *CG11700* accumulated numerous amino acid replacements in four ubiquitin-coding repeats, suggesting that it might have diverged to evolve new functions. Our functional investigations showed that *CG11700* is involved in the trade-off between male fecundity and lifespan in *D. melanogaster*.

Materials and Methods

Synteny Analysis of *CG32744* and *CG11700*

We performed BLAST (Altschul et al. 1997) and synteny analyses (gene order) to reveal the duplication process using genomic sequence data from five species downloaded from FlyBase (ftp://ftp.flybase.net/genomes/) (*D. melanogaster* [Dmel: R5.16], *D. sechellia* [Dsch: 1.3], *D. simulans* [Dsim: R1.3], *D. yakuba* [Dyak: R1.3], and *D. erecta* [Dere: R1.3]). The annotated orthologs (homologs) for *CG32744* and *CG11700* in other species are *GM12582* in *D. sechellia*, *GE16472* in *D. yakuba*, and *GG17684* in *D. erecta*. The orthologous gene in *D. simulans* is not intact due to a number of sequencing gaps.

Analysis of Expression of *CG32744* and *CG11700* by RT-PCR and Online Resource Data

Total RNA was extracted with Trizol reagent (Invitrogen Life Technologies, Carlsbad CA) from different developmental stages (0- to 24-h-old embryos, 1- to 3-day-old larvae, 1- to 3-day-old pupae, and 1- to 5-day-old adult males and females) of *D. melanogaster* and *D. yakuba*, from testes (with accessory glands) dissected from adult males and from gonadectomized adult males of *D. melanogaster*. Total RNA was treated by RNase-free DNase I (MBI Fermentas, Vilnius, Lithuania) and reverse transcribed to synthesize cDNA using Superscript III reverse transcriptase (Invitrogen Life Technologies, Carlsbad CA). Using primers designed in the first (forward primer) and second annotated exon (reverse primer), which cover the 5′ Untranslated Regions (5′ -UTR) and the coding sequence of the target genes, we conducted RT-PCR to confirm the gene structure of *CG11700* and *CG32744* in *D. melanogaster*. We also used these primers to analyze their expression pattern as well as the single-copy ortholog (*GE16472*) in *D. yakuba*. All PCRs were performed using rTaq polymerase (Takara, Dalian, China) with gene-specific primers. *Gapdh2* in *D. melanogaster* and *rp49* in *D. yakuba* were set as the internal control. In *D. melanogaster*, the forward (F) and reverse (R) primers for each gene were CG11700RT-F and CG11700RT-R, CG32744RT-F and CG32744RT-R, Gapdh2RT-F and Gapdh2RT-R. In *D. yakuba*, the primers for each gene were GE16472RT-F and GE16472RT-R, rp49RT-F and rp49RT-R. The sequences of all primers used in the study were listed in supplementary table S1 (Supplementary Material online). In both *D. melanogaster* and *D. yakuba*, PCR using genomic DNA as template was set as a positive control reaction.

The expression patterns of both *CG11700* and *CG32744* were further analyzed by searching the modENCODE Temporal Expression Data from FlyBase (http://flybase.org/).

Analysis of Molecular Evolution

The coding regions in *CG32744*, *CG11700* and the orthologs in other species were aligned using Clustal X 2.0 (Larkin et al. 2007) and were further manually adjusted to correct frame. Because *GE16472* has only three ubiquitin-coding repeats in its open reading frame, only the alignment for these three repeats was performed for further analysis. Their phylogenetic relationship was constructed using MEGA 4.0 (Tamura et al. 2007). The branch lengths of figure 1*E* were estimated by d_S using PAML version 4.4c (Yang 1997, 2007). The numbers of synonymous and nonsynonymous substitutions as well as the ratio of nonsynonymous substitutions over synonymous substitutions (d_N/d_S, ω) on each branch were analyzed using the codeml program of PAML version 4.4c under free ratio model (model = 1, NS sites = 0) (Yang 1997, 2007), and the branch lengths were estimated by the estimated d_S value. Several models were used for comparisons in PAML analysis. First, we assumed the same ω for all branches and estimated the ω values under one ratio model (model A). Second, the ω values for all branches were calculated under the free ratio model (model B), which assumes that each branch has a different ω value. Then, the ω value of each gene's branch was fixed as 1, and the ω of the other branches were calculated under the two-ratio model (model C-G). We performed likelihood ratio test (statistics $2\Delta\ell > 0$ is tested assuming that it follows chi-square

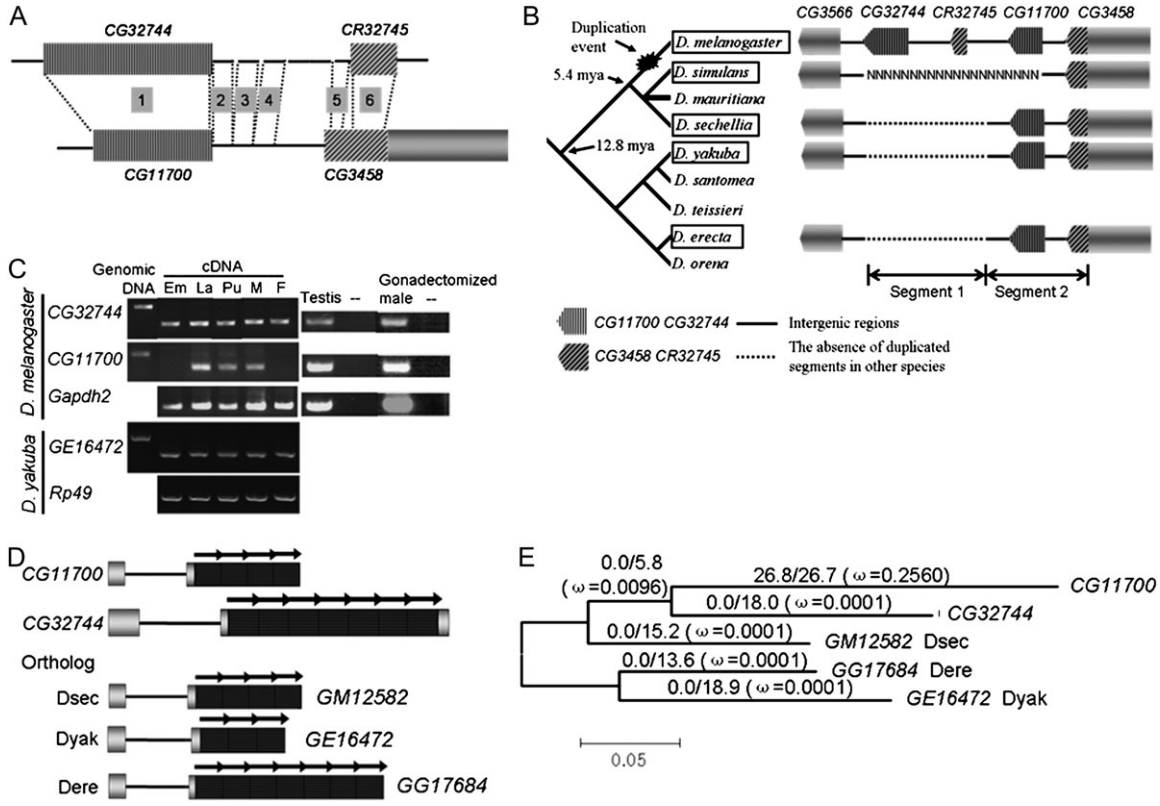

Fig. 1. Evolutionary analysis of *CG32744* and *CG11700* in the *Drosophila melanogaster* subgroup. (*A*) Genomic organization and homologous regions of the two paralogous regions containing *CG32744* and *CG11700* in *D. melanogaster*. Six sequential homologous regions, including *CG32744* and *CG11700* (block 1); three intergenic regions (blocks 2, 3, 4); upstream of *CR32745* and the 5′ tip of *CG3458* (block 5); and *CR32745* and part of *CG3458* (block 6) show sequence identities of 92%, 92%, 75%, 92%, 93%, and 97%, respectively. This suggests that these homologous regions arose through a segmental duplication event. (*B*) The phylogenetic distribution of duplicates *CG32744* and *CG11700* in the *D. melanogaster* subgroup. The divergence time among species (Zhou et al. 2008) and the time of duplication events are labeled in the phylogenetic tree. The duplicated segments are indicated at the bottom. The lengths of segment 1 and 2 are 4,074 and 4,992 bp. The dotted lines indicate the absence of one segment in the outgroup species. *N* indicates a missing sequence in *D. simulans*. (*C*) RT-PCR results. Em, La, Pu, M, and F denote embryo, larvae, pupae, adult males, and adult females, respectively. In both *D. melanogaster* and *D. yakuba*, PCR using genomic DNA as templates was set as positive control reaction for primers of *CG32744*, *CG11700*, and the ortholog in *D. yakuba*. "—" indicates the RT-PCR-negative control for testes and gonadectomized male bodies with the testes and accessory glands removed, in which everything is the same as the cDNA samples, except that reverse transcriptase has been omitted. (*D*) The gene structures of *CG11700*, *CG32744*, and the orthologs in *D. sechellia* (*GM12582*), *D. yakuba* (*GE16472*), and *D. erecta* (*GG17684*). The gray and black boxes represent UTRs and coding regions, respectively. The lines between boxes indicate introns. Each arrow above the coding region denotes a ubiquitin-coding repeat. (*E*) Substitution patterns of *CG11700* in the *D. melanogaster* subgroup. Numbers of synonymous and nonsynonymous substitutions, as well as ω values (d_N/d_S), are labeled on each branch.

distribution, where $\Delta\ell$ is the difference of the logarithm of the likelihood values calculated under the two models) to compare the fit of these models with the data.

Population Genetic Analysis on *CG11700*

To obtain polymorphism data for *CG11700*, the entire coding region and the first intron of *CG11700* from nine *D. melanogaster* strains (high gr500 CA, North N4, EC158, ZS28, ZS(6), ZS(29), ZS(11), ZW56, and Canton S) were amplified using ExTaq polymerase (Takara, Dalian, China), which is a high fidelity DNA polymerase. The PCR and sequencing primers are CG11700pop-F and R (supplementary table S1, Supplementary Material online). The sequences have been deposited into GenBank (accession nos. JQ080076-JQ080084). The sequences were assembled using Seqman 5.0 in DNASTAR (DNASTAR, Inc.) and were then manually checked. The coding regions of these sequencing data were used for further analysis. Including the sequence from the fly reference genome, homologous sequences were aligned with Clustal X 2.0 (Larkin et al. 2007). DnaSP v5 package (Rozas et al. 2003) was used to calculate the numbers of synonymous and nonsynonymous sites, the numbers of synonymous and nonsynonymous substitutions, and average nucleotide pairwise synonymous and nonsynonymous differences.

Additionally, we performed lineage-specific McDonald and Kreitman test (McDonald and Kreitman 1991; Presgraves 2007; Meisel et al. 2010) to investigate whether the *CG11700* gene in *D. melanogaster* was subject to positive selection or not. Two genes (*GM12582* in *D. sechellia* and *GG17684* in *D. erecta*) were used as outgroup genes to infer the ancestral sequence of *CG11700* and *CG32744* in

D. melanogaster. Because the sequence of *GE16472* in *D. yakuba* has a big deletion compared with *CG11700*, we excluded it when we were inferring the ancestral sequence.

Examination of Amino Acid Substitution and Remodeling of 3D Structure of *CG32744* and *CG11700*

The amino acid sequences of ubiquitin-coding units in *CG32744* and *CG11700* in *D. melanogaster*, as well as those in the ortholog of other four species of the *D. melanogaster* subgroup, were aligned by ClustalW2.0 (Larkin et al. 2007). The amino acid substitutions were counted and mapped to the secondary structure motifs of ubiquitin (http://stud.chem.uni.wroc.pl/users/lucek/JAREMKO/ubiquitin.htm), which was downloaded from Protein Data Bank (http://www.pdb.org).

In order to find structural effects of amino acid replacement, we remodeled the 3D structure for four ubiquitin peptide units in *CG11700* and one in *CG32744* (because seven ubiquitin-coding repeats in *CG32744* encode the same peptide), employing the Protein Structure Prediction server with default parameters (http://ps2.life.nctu.edu.tw/) (Chen et al. 2006).

Overexpression of *CG11700* and Western Blotting Assay

We amplified the coding region of *CG11700* using the primers CG11700CDS-F and R (supplementary table S1, Supplementary Material online) with *Eco*R I and *Xba* I sites added in the primers, respectively. The amplified product of *CG11700* was then cloned into pUAST with *Eco*R I and *Xba* I sites, and myc tag was annealed to the 5′ end of *CG11700* in frame. The resultant plasmid pUAST-myc-CG11700 was transformed into *w1118* flies according to standard protocols of germ line transformation (Rubin and Spradling 1982). Females of transgenic flies (*w/w; UAS-myc-CG11700/UAS-myc-CG11700;+/+*) and *w1118* (*w/w; +/+;+/+*) were crossed to males carrying *tubulin-GAL4* (Bloomington Stock Number: 5138. *yw/Y; +/+;P{w[+mC]=tubP-GAL4}LL7/TM3, Sb*). Correspondingly, their offspring were the overexpression line (*w/w; UAS-myc-CG11700/+; tubulin-GAL4/+*) and the wild-type control (*w/w; +/+; tubulin-GAL4/+*), respectively.

For western blotting analysis of myc-tagged CG11700 protein, the cell lysates were extracted from the male individuals of the *CG11700* overexpression line, denatured, and subjected to immunoprecipitation with anti-myc antibodies, followed by immunoblotting with the secondary antibody.

Generation of Mutants of *CG11700* and *CG32744*

The fly stocks used in this study were maintained under standard culture conditions. The mutant flies carrying *P{w[+mC] y[+mDint2]=EPgy2}* elements (Mata et al. 2000) around *CG11700* (Bloomington Stock Number: 16008, *y[1] w[67c23] P{w[+mC] y[+mDint2]=EPgy2}EY09407*) and *CG32744* (Bloomington Stock Number: 22365, *y[1] w[67c23] P{w[+mC] y[+mDint2]=EPgy2}EY20141*) were obtained from Bloomington *Drosophila* Stock Center. *y[1] w[67c23] P{EPgy2}EY09407* and *y[1] w[67c23] P{EPgy2}EY20141* were integrated in the 3′ downstream region of *CG11700* and the 5′ upstream region of *CG32744*, respectively.

Excision of the *P{EPgy2}EY09407* and *P{EPgy2}EY20141* insertions were carried out by crossing females flies carrying the *P{EPgy2}* insertions with *yw; Sp/CyO; Δ2-3,sb/TM6B* males carrying Δ2-3transposase (Robertson et al. 1988). Recovered white-eyed progenies were backcrossed to isogenic *y[1] w[67c23] P{EPgy2}* lines to establish stocks: 1) null mutants, in which *CG32744* or *CG11700* was deleted by *P*-element imprecise excision, and 2) wild-type controls, in which *CG32744* or *CG11700* was recovered by *P*-element precise excision. The deletion in the null mutant lines was validated by PCR amplification of *CG32744* and *CG11700* using primers covering the 5′-UTR and the coding sequence of the target genes, including CG11700scF and CG11700scR, CG32744scF and CG32744scR (supplementary table S1, Supplementary Material online). Wild-type control lines were confirmed by PCR amplification and sequencing of the targeted region. Two independent null mutant and control lines of each gene were isolated. All of these flies were homozygous or hemizygous viable.

Fecundity and Lifespan Assays

We separated sexes on a plate under light CO_2 anesthesia and performed two assays, as below. For fecundity assays, ten 2- to 3-day-old virgin flies of the two null mutant lines ($CG11700^{-}$/Y; +/+; +/+, $CG32744^{-}$/Y; +/+; +/+), the wild-type control for null mutants ($CG11700^{+}$/Y; +/+; +/+, $CG32744^{+}$/Y; +/+; +/+), the overexpression line (*w/Y; UAS-myc-CG11700/+; tubulin-GAL4/+*), and the wild-type control for the overexpression line (*w/Y; +/+; tubulin-GAL4/+*) were set up separately to mate with three 3- to 4-day-old *w1118* virgin females and were transferred to fresh vials every 2–4 days. Six replicates for each line were set. The enclosed offspring were collected and counted until the male died after the initial mating. For lifespan test for *CG11700*, ten male flies in each vial of the two lines of the null mutants ($CG11700^{-}$/Y; +/+; +/+) and the wild-type control for null mutants ($CG11700^{+}$/Y; +/+; +/+) were set up to mate with ten *w1118* females in a vial. Two days after mating, all 20 flies in the vial were transferred to another vial with fresh food. During the experimental time course, the flies were transferred to new vials with fresh food every 3–4 days in order to keep uniform living conditions for them. In order to get enough male flies to calculate the survival rate, we pooled all individuals from two independent lines (130 for the two mutant lines and 60 for the wild-type control lines, respectively). The corresponding survival rate (%) was calculated at each time point with an interval of 3–4 days.

Statistical Analysis

Nested analysis of variance (ANOVA) was used to determine whether the effect of the gene state (null mutant

or wild type) on the number of offspring was significant in fecundity assays; and paired *t*-test was used to compare the overall average survival rate of mutants and the wild-type control flies during the experimental time course.

Results and Discussion

CG32744 and *CG11700* Were Derived from Segmental Tandem Duplication and Both of Them Encode Putative Ubiquitins in *D. melanogaster*

The gene pair *CG32744* and *CG11700* was first reported in our systematic identification of new genes in the *D. melanogaster* subgroup (Zhou et al. 2008). *CG32744* and *CG11700* lie in the chromosome bands 5E4 and 5E5 in *D. melanogaster*, respectively. We took a further step to check our previous observation by comparing paralogous regions as well as the orthologous regions from the genomes of five closely related species of the *D. melanogaster* subgroup (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, and *D. erecta*). Our syntenic analysis and BLAST results show not only the similarity between gene regions (*CG32744* and *CG11700* [92%] as well as *CR32745* and the tip of *CG3458* [97%]) but also the similarity between intergenic regions (75–93%) and another partial gene duplication region (*CR32745* and the tip of *CG3458*, 97%) (fig. 1*A* and supplementary data S1, Supplementary Material online). Based on this observation and the tandem arrangement of the duplicated segments (fig. 1*A* and *B*), we propose that the duplication of *CG32744* and *CG11700* should have occurred by a single event of tandem segmental duplication. The BLAST hits of segments 1 and 2 have the same genomic locations in the genomes of other species (supplementary table S2, Supplementary Material online). Therefore, there is only one segment in the outgroup species, including *D. sechellia*, *D. yakuba*, and *D. erecta* (because there are sequencing gaps in the orthologous region in *D. simulans*, we excluded this species from following analyses). When comparing the sequence similarity of these regions among species, we found that *CG32744* has a higher sequence identity to the orthologous segment in the outgroup species than *CG11700* (supplementary table S2, Supplementary Material online). In addition, *CG32744*, but not *CG11700*, has an identical expression pattern to the orthologous *D. yakuba* gene as revealed by our RT-PCR analysis (fig. 1*C*). All these observations suggest that *CG32744* more likely retains the features of the ancestral gene.

CG32744 has a transcript of 1,913 bp with an ORF of 1,605 bp, encoding a putative polypeptide of 534 amino acids. *CG11700* was previously annotated as a pseudogene (*CR11700*: FB2007_01, dmel_r5.2) due to a lack of ESTs, but our RT-PCR result reveals that *CG11700* is actually transcribed and that the introns are spliced (fig. 1*C*). Therefore, *CG11700* likely is an active gene, and we reannotated it as *CG11700* instead of *CR11700*. *CG11700* has a potential ORF of 906 bp encoding a polypeptide of 301 amino acids.

In order to test whether or not *CG32744* and *CG11700* are functional, we calculated the d_N/d_S value of the coding region of *CG32744*, *CG11700*, and the orthologs among species in the *D. melanogaster* subgroup. The d_N/d_S value is significantly less than 1 in all comparisons (supplementary table S3, Supplementary Material online), suggesting that *CG11700* is possibly evolving under functional constraint. In addition, we employed PAML (Yang 1997, 2007) to calculate the numbers of synonymous and nonsynonymous substitutions as well as d_N/d_S (ω) on each branch under free ratio model (model = 1, NS sites = 0) and found that the ω values for all the branches are lower than 1 (fig. 1*E*, supplementary table S4, Supplementary Material online), suggesting functional constraint on the genes across the phylogeny. However, it should be noted that the ω value for the branch leading to *CG11700* (0.256) are much higher than the other ω values ($\leq$0.0001) (supplementary table S3, Supplementary Material online and fig. 1*E*), suggesting that relaxation of functional constraint or positive selection specifically operated on *CG11700*.

We further investigated the polymorphism data of *CG11700* in *D. melanogaster* to test if *CG11700* is under functional constraints (table 1). Using the DnaSP v5 package (Rozas et al. 2003), we found that there are 15 synonymous polymorphic sites and 4 nonsynonymous polymorphic sites in the population data and that the average synonymous difference per synonymous site (πS) (0.02898) is significantly higher than the average nonsynonymous difference per nonsynonymous site (πN) (0.00321) ($P < 0.01$) in the *D. melanogaster* population. These results suggest that *CG11700* may be under functional constraint in the population of *D. melanogaster*. Additionally, to examine whether or not *CG11700* experienced positive selection after it diverged from *CG32744*, we further performed lineage-specific McDonald and Kreitman test (McDonald and Kreitman 1991; Presgraves 2007; Meisel et al. 2010). We inferred the ancestral sequence of *CG11700* and *CG32744* using *Dsch_GM12582* and *Dere_GG17684* as outgroup sequences (because the *GE16472* sequence of *D. yakuba* has a big deletion compared with *CG11700*, we excluded it when we were inferring the ancestral sequence). Compared with the ancestral sequence, 13 fixed synonymous and 20 fixed nonsynonymous substitutions happened in the lineage of *D. melanogaster*'s *CG11700* (tables 1 and 2). Based on our population data for *CG11700*, there are 15 synonymous and 4 nonsynonymous polymorphic sites in the *D. melanogaster* (table 1). Thus, the lineage-specific McDonald and Kreitman test indicates that the *CG11700* of *D. melanogaster* might have experienced positive selection after it diverged from *CG32744* ($P < 0.01$ by both *G*-test and Fisher's exact test) (table 2). Alternatively, it is possible that the fixation of the high number of nonsynonymous substitutions happened to *CG11700* was due to relaxation of functional constraint because the polymorphism in *D. melanogastor* is apparently under negative selection as mentioned earlier.

Our RT-PCR results clearly show that *CG32744* is widely expressed in all developmental stages and both sexes, whereas *CG11700* is only expressed in males in *D. melanogaster* (fig. 1*C*), which is consistent with the

Table 1. Fixed and Polymorphic Sites in the Coding Region of *Drosophila melanogaster CG11700* Gene.

	■ΔΔ■○○○●●○Δ●●●●●●○ΔΔ○○○●●ΔΔ●■●●Δ●●Δ●■○ΔΔΔ●●○ΔΔ○Δ○
	● ●
	6347789111111111222233333333333344555555555556666
	0202100001234590234001333446788231222334447903334
	4787077492240098256585527709138470296440 98
CG11700	GCCGAAAAATGTGCAATGACAAAAACGTGCAAATTGAATGCGCACTTCA
High gr500 CA	.....C.....CA....T....
North N4	.....C.....C...
EC158	.....C.....C...
Canton S	.....C.....C...
ZS28	TA.A.....T.....CA.....
ZS(6)	..TA.....T.....CA.....
ZS(29)	.....A.....A.....CA.....
ZS(11)	...A.....T.....G.....C.T....C.T.
ZW56	T.....GT.....C.C.....
Ancestral sequence	....GGCGTCTGCGGCCTG.GGCGT.TGCGG.CC.C.G...AGG.CC.G

NOTE.—The inferred ancestral sequence of *CG11700* and *CG32744* was constructed using *Dsch_GM12582* and *Dere_GG17684* as outgroup sequences. Δ Synonymous polymorphic changes within *D. melanogaster*; ■ Nonsynonymous polymorphic changes within *D. melanogaster*; ○ Synonymous fixed differences between the *CG11700* and the ancestral sequence; ● Nonsynonymous fixed differences between the CG11700 and the ancestral sequence. Position_117 and Position_345 in *CG11700* not only show fixed differences between the *D. melanogaster* and the ancestral sequence but also have evolved synonymous polymorphism within *D. melanogaster*.

modENCODE Temporal Expression Data from FlyBase (http://flybase.org) (supplementary figs. S1 and S2, Supplementary Material online). Although specific antibodies for these two proteins are not available, our in vivo tagged protein experiment confirms the expression of CG11700 protein (see below). The tissue-specific expression pattern further indicates the functionality of *CG11700*. The evolution of the male-specific expression pattern for *CG11700* might be due to changes of its upstream regulatory elements because the sequence alignments of the upstream regions of *CG32744*, *CG11700*, and the orthologs in outgroup species show that there are many *CG11700*-specific mutations in *D. melanogaster* (supplementary fig. S3, Supplementary Material online).

Previous studies suggested that X-linked genes tend to move to autosomes and avoid the X inactivation during spermatogenesis (Betrán et al. 2002; Emerson et al. 2004; Zhang et al. 2010). Similar to the reported duplicated genes *Sdic* (Nurminsky, Nurminskaya, Benevolenskaya, et al. 1998; Nurminsky, Nurminskaya, De Aguiar, et al. 1998) and *Tektins* (Dorus et al. 2008), which show male-biased expression pattern, *CG11700* and *CG32744* are X-linked young duplicates and one of them (*CG11700*) obtained male-biased expression pattern, which is consistent with the two-step model (Zhang et al. 2010) rather than MSCI or SAXI models (Wu and Xu 2003; Kaiser and Bachtrog 2010).

CG11700 Accumulated Numerous Amino Acid Substitutions

Usually ubiquitin peptides are well conserved, manifested by the fact that 72 of the 76 amino acids are invariable among fungi, plants, and animals (Sharp and Li 1987). The ORFs of *CG11700* and *CG32744* comprised four and seven ubiquitin coding units, respectively. The ORFs of

Table 2. Result of Lineage-Specific McDonald and Kreitman Test for *CG11700*.

	Synonymous Substitutions	Nonsynonymous Substitutions
Observed polymorphic sites within *Drosophila melanogaster*	**15**	**4**
Inferred fixed sites along the *CG11700* lineage	**13**	**20**

NOTE.—$P = 0.009007$ by Fisher's exact test; $P = 0.00475$ by *G*-test.

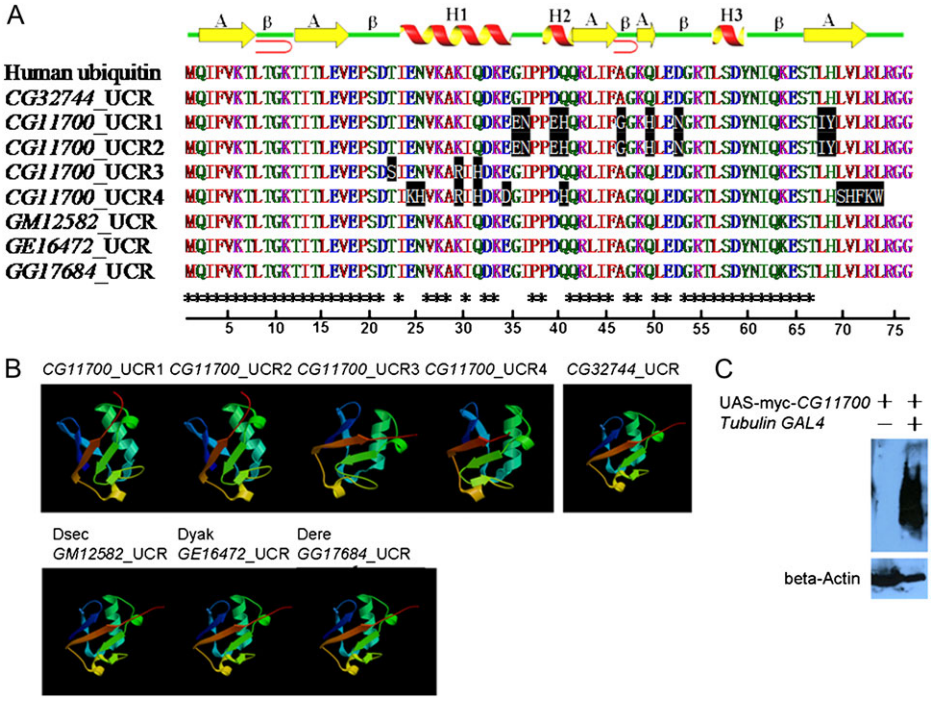

FIG. 2. Peptide sequence alignment (*A*), 3D remodeling (*B*), and western blotting analysis (*C*) of CG11700 protein. (*A*) Amino acid sequence alignment and the secondary structure prediction of the ubiquitin-coding repeats in *CG32744*, *CG11700*, and the orthologs (*GM12582* in *Drosophila sechellia*, *GE16472* in *D. yakuba*, and *GG17684* in *D. erecta*, respectively). Wiring diagram of secondary structures for ubiquitin is presented above the sequence alignment. Yellow arrows: β-strands (*A*); red helices: α-helices; β above bold lines: β-turns (β); red curves below bold lines: β-hairpins. UCR denotes ubiquitin-coding repeats. *CG11700*_UCR1/2/3/4 denotes its four ubiquitin-coding repeats, in which the amino acid replacements are marked with black blocks. (*B*) 3D structure prediction of ubiquitin domains for CG32744, CG11700, and the outgroup orthologs. Dsec, Dyak, and Dere denote *D. sechellia*, *D. yakuba*, and *D. erecta*, respectively. Obvious differences in the 3D structures of the four repeat units of *D. melanogaster CG11700* suggest that the mutations may have functional effects. (*C*) Western blotting analysis of CG11700 protein. The male flies (*w*/*Y*; *UAS-myc-CG11700*/+; *tubulin-GAL4*/+ and *w*/*Y*; +/+; *tubulin-GAL4*/+) were used for the isolation of protein extracts and performed western blotting analysis with anti-myc antibody.

the orthologs in *D. sechellia*, *D. yakuba*, and *D. erecta* comprised four, three, and seven ubiquitin-coding repeats, respectively (fig. 1*D*). We aligned the putative peptide sequence of each ubiquitin-coding repeat of *CG32744* and *CG11700* in *D. melanogaster* to those of the orthologs in other four species (supplementary data S3, Supplementary Material online). Surprisingly, we found that the *D. melanogaster* CG11700 putative protein has a total of 32 amino acid replacements in its four ubiquitin-coding repeats (9, 9, 3, and 11, respectively) (fig. 2*A*). In contrast, the ubiquitin-coding repeats in *CG32744* and in the orthologs of *D. yakuba* and *D. erecta* encode completely identical peptide sequences (supplementary fig. S4, Supplementary Material online), and there is only one amino acid substitution between the *CG32744* and the ortholog of *D. sechellia* (R74L) (supplementary fig. S4, Supplementary Material online). Moreover, the ubiquitin-coding repeats of *CG32744* are also completely identical to those of other ubiquitin genes in *D. melanogaster*, such as Ubi-p63E (*CG11624*), *RpL40* (*CG2960*), and *RpS27A* (*CG5271*) (supplementary fig. S5, Supplementary Material online), and to those of human ubiquitin gene (fig. 1*D*).

Considering that previous studies have shown concerted evolution of ubiquitin genes (Sharp and Li 1987; Nei et al. 2000), we employed GENECONV (Sawyer 1989) to detect gene conversion events among ubiquitin-coding repeats within or between *CG11700*, *CG32744*, and the orthologs in other species. As shown in supplementary table S5 (Supplementary Material online), gene conversion events may have ever occurred within *CG11700* and between *CG11700* and *CG32744* in *D. melanogaster*. Nevertheless, many amino acid replacements in *CG11700* clearly show that *CG11700* has diverged dramatically from the ancestral polyubiquitin gene and might have evolved novel functions, whereas *CG32744* possibly keeps the functions of the ancestral gene.

To evaluate the influence of amino acid replacements on the protein structure of *CG11700*, we superimposed their locations onto the structure-resolved human ubiquitin (Vijay-Kumar et al. 1987; Jackson 2006). The human ubiquitin is composed of one β-sheet, five β-strands, three α-helices, six β-turns, one β-bugle, and two β-hairpins (http://stud.chem.uni.wroc.pl/users/lucek/JAREMKO/ubiquitin.htm) (fig. 2*A*). Among the 32 amino acid replacements, 12, 4, 9, and 2 are located in the α-helix, β-turn, β-strand, and β-hairpin, respectively (fig. 2*A*, supplementary table S6, Supplementary Material online); the remaining 7 are not in any secondary structure motifs. The results of 3D-structure remodeling revealed different structures between the four ubiquitin peptide units of *CG11700* and the other ubiquitin peptides

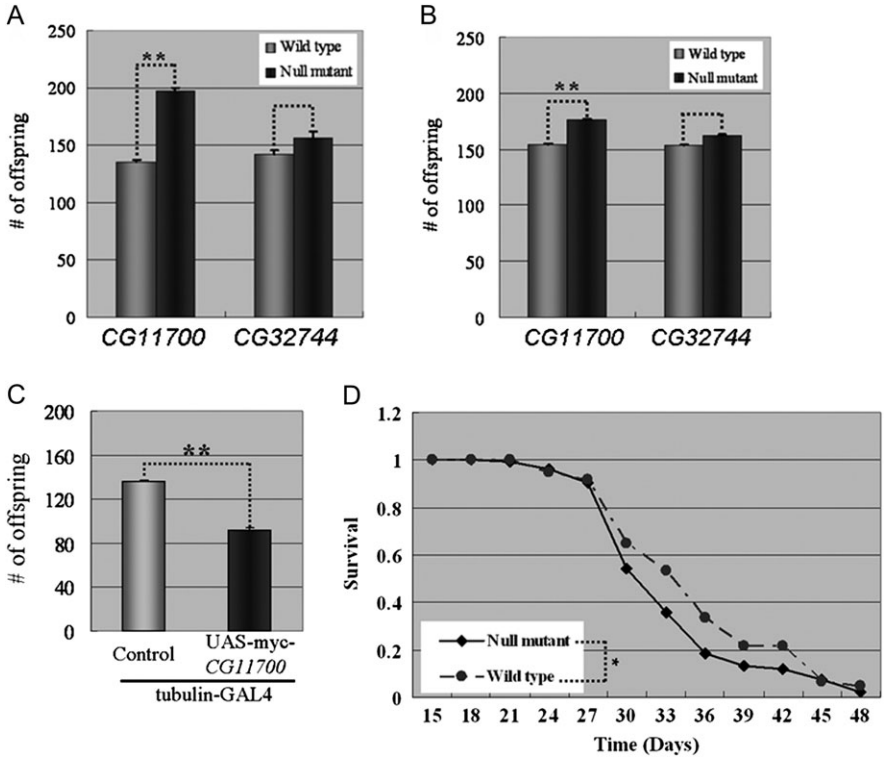

FIG. 3. The male fecundity and lifespan assays. (*A*) and (*B*) The male fecundity of the two null mutants and two wild-type controls of both *CG32744* and *CG11700*, indicated by the average number of offspring, as shown by the *y* axis. Arrow bars show standard errors. (*C*) The male fecundity of overexpression line and wild-type control of *CG11700*, indicated by the number of offspring, as shown by the *y* axis. Arrow bars show standard errors. (*D*) The lifespan of *CG11700*-null males and the wild-type controls, indicated by the survival rate (as shown by the *y* axis) calculated at each time point (as shown by the *x* axis), showing that the lifespan of *CG11700*-null mutants is significantly shorter than that of wild-type controls. Levels of statistical significance are shown in the figure (*$P < 0.05$, **$P < 0.01$).

(fig. 2*B*), indicating that the mutations might have important functional impacts on *CG11700*.

In order to investigate whether the mutations in the ubiquitin repeat units of *CG11700* have changed its function as a ubiquitin, we generated the *CG11700* overexpression lines (*w/Y; UAS-myc-CG11700/+; tubulin-GAL4/+*) and corresponding wild-type control lines (*w/Y; +/+; tubulin-GAL4/+*) and extracted the protein from the males, since *CG11700* only expresses in males. Protein lysate was subjected to immunoprecipitation with anti-myc antibodies, followed by immunoblotting with anti-myc antibody and secondary antibody (fig. 2*C*). It is well known that ubiquitins, which are usually linked together to form polyubiquitin chains via the lysine48 (K48) residue of each ubiquitin, form (Ub)n-ubiquitinylated substrates through binding to various proteins in vivo, and thus, a smear-like blot banding pattern usually is a characteristic of a ubiquitin protein (Hershko and Ciechanover 1998; Kerscher et al. 2006). Interestingly, our western blotting result for *D. melanogaster* CG11700 showed a smear-like blot banding pattern (fig. 2*C*), which is consistent with the cases in mammalian cells (Moriyoshi et al. 2004; Adhikary et al. 2005; Azakir and Angers 2009; Yin et al. 2010), indicating that *CG11700* may still encode polyubiquitin peptides and bind to protein substrates in males.

Proteasome subunits have been reported to be involved in ubiquitin-mediated proteolysis and usually experienced rampant gene duplication in *Drosophila* (Belote and Zhong 2009). Similar to other new genes, these newly evolved proteasome subunit genes also exhibit testis-specific expression (Belote and Zhong 2009). The subunits of ubiquitin–proteasome system could interact with each other to accomplish the protein degradation processes. If one subunit has mutations, some of other subunits may also accumulate mutations to operate in coordination. Nevertheless, there has been no new gene duplication of proteasome subunits in *D. melanogaster* (Belote and Zhong 2009). We also carried out an amino acid sequence analysis on proteasome subunits in *Drosophlia* species and did not find any significant change in *D. melanogaster* (data not shown). Therefore, we did not observe an obvious coevolution of proteasome subunits with the appearance of *CG11700* in *D. melanogaster* at the level of copy number variation or amino acid changes.

CG11700 Reduces Male Fecundity but Extends Lifespan

It has been reported that ubiquitin can function in reproductive processes, such as gametogenesis and fertilization (Baarends et al. 1999; Bebington et al. 2001; Sakai et al. 2004). Whether *CG11700*, which is expressed only in adult males (fig. 1*C*), can perform special reproductive functions, remains unclear. We generated *CG11700-* and *CG32744-*

null mutants by *P*-element excision to analyze their in vivo functions. The deletion of *CG11700* and *CG32744* was confirmed by PCR (supplementary fig. S6, Supplementary Material online). Although we did not observe any obvious morphological change in the testes of the null mutant flies of *CG11700*, both the independent excision lines of *CG11700* show a consistent phenotype; that is, the mutant males can produce significantly more progeny than the wild-type control ($P < 0.01$ by *t*-test; $P = 0.025$ by nested ANOVA [$F_{1,20} = 5.90$]) (fig. 3*A* and *B*). As we could not determine the exact region of the deletion using genomic PCR to exclude the possibility that the phenotypes might be caused by the disruption of nearby genes in the *P*-element mobilization, we carried out an overexpression experiment of *CG11700*. As expected, the male individuals of the overexpression line produced significantly fewer offspring ($P = 0.000184$ by *t*-test; see fig. 3*C*), further suggesting that the phenotype observed in the null mutants is indeed caused specifically by the *CG11700* gene. In contrast, no significant fecundity change was observed in the male null mutants of *CG32744* ($P = 0.67$ by the nested ANOVA [$F_{1,20} = 0.18$]) (fig. 3*A* and *B*). These results suggest that *CG11700* has evolved a novel function in the suppression of male fecundity in *D. melanogaster*. However, as a dramatically diverged ubiquitin-encoding gene, how *CG11700* acts to decrease the male fecundity at the molecular level remains to be elucidated. Given that a number of proteasome subunits and ubiquitin-activating enzyme (Uba1) are present in the sperm proteome (Wasbrough et al. 2010), it is possible that *CG11700* might be involved in ubiquitin-mediated proteolysis or transfer of testis proteins, which might result in changes in efficiency in either fertilization or mediation of female characteristics. These issues await more investigation in future functional studies.

Given the unexpected observation that *CG11700* null mutants can significantly increase male fecundity, which invokes its possible roles in the trade-off between reproduction and other physiological processes, especially that involving the lifespan, we tried to measure lifespan of the null mutant and wild-type control flies of *CG11700* using male flies kept on standard food to test the reproductive cost. Comparison of the survival curves during the whole experimental time course showed a significant lower survival rate of *CG11700* null mutant males than wild-type males ($P = 0.01738$ by paired *t*-test) (fig. 3*D*), indicating the existence of reproductive cost. Considering that *CG11700* encodes a ubiquitin protein (fig. 2*C*), we suspect that the loss of function of *CG11700* might cause abnormal accumulation of some proteins, which is likely to influence the lifespan. Further comparison of the proteomes between the null mutants and the wild-type controls will thus be quite worthwhile.

In summary, the polyubiquitin gene *CG11700* has accumulated many nonsynonymous substitutions and apparently evolved a novel function involving in repressing fecundity and extending lifespan after the duplication that occurred in the *D. melanogastor* lineage, whereas the paralogous gene *CG32744* has retained the ancestral function. The neofunctionalization of *CG11700* might have contributed to the adaptive evolution of *D. melanogaster* by balancing the cost of reproduction. Further study on its molecular function will lend more pieces of evidence to demonstrate how it fulfills such an important role. The significant result of lineage-specific McDonald–Kreitman test indicates positive selection if we assume the polymorphism in the present population is neutral. However, because the polymorphism in *D. melanogastor* is under negative selection, it is possible that the significant result of the lineage-specific McDonald–Kreitman test is due to relaxation of functional constraint. Therefore, it is likely that positive selection might have operated on *CG11700* considering that it has acquired a new function, but relaxation of functional constraint might also have occurred some time.

Supplementary Material

Supplementary data S1–S3, figures S1–S6, and tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Drs P. Shi and HP. Li for their helpful comments and suggestions on the manuscript. Thanks are also given to HF. Jiang, D. Li, X. Li, and XY. Li for discussions during the design of the experiments. This work was supported by an NSFC key grant (No. 30930056) to W.W., State Key Laboratory of Genetic Resources and Evolution Project (GREKF08-02), and a grant of The National Natural Science Foundation of China (NSFC) (30871342) to H.Y.

References

Adhikary S, Marinoni F, Hock A, et al. (14 co-authors). 2005. The ubiquitin ligase HectH9 regulates transcriptional activation by Myc and is essential for tumor cell proliferation. *Cell* 123:409–421.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Azakir BA, Angers A. 2009. Reciprocal regulation of the ubiquitin ligase Itch and the epidermal growth factor receptor signaling. *Cell Signal.* 21:1326–1336.

Baarends WM, Roest HP, Grootegoed JA. 1999. The ubiquitin system in gametogenesis. *Mol Cell Endocrinol.* 151:5–16.

Bebington C, Doherty FJ, Fleming SD. 2001. The possible biological and reproductive functions of ubiquitin. *Hum Reprod Update.* 7:102–111.

Belote JM, Zhong L. 2009. Duplicated proteasome subunit genes in *Drosophila* and their roles in spermatogenesis. *Heredity* 103:23–31.

Betrán E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* 12(12):1854–1859.

Chen CC, Hwang JK, Yang JM. 2006. (PS)2: protein structure prediction server. *Nucleic Acids Res.* 34:W152–W157.

Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W, Long M. 2008. The evolution of courtship

behaviors through the origination of a new gene in *Drosophila*. *Proc Natl Acad Sci U S A.* 105:7478–7483.
Ding Y, Zhao L, Yang S, et al. (12 co-authors). 2010. A young *Drosophila* duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. *PLoS Genet.* 6:e1001255.
Dorus S, Freeman ZN, Parker ER, Heath BD, Karr TL. 2008. Recent origins of sperm genes in *Drosophila*. *Mol Biol Evol.* 25: 2157–2166.
Emerson JJ, Kaessmann H, Betrán E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. *Science* 303(5657): 537–540.
Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
Gu Z, Nicolae D, Lu HH, Li WH. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* 18:609–613.
Harshman LG, Zera AJ. 2007. The cost of reproduction: the devil in the details. *Trends Ecol Evol.* 22:80–86.
Hershko A, Ciechanover A. 1998. The ubiquitin system. *Annu Rev Biochem.* 67:425–479.
Jackson SE. 2006. Ubiquitin: a small protein folding paradigm. *Org Biomol Chem.* 4:1845–1853.
Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet.* 10:19–31.
Kaiser VB, Bachtrog D. 2010. Evolution of sex chromosomes in insects. *Annu Rev Genet.* 44:91–112.
Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol.* 22:159–180.
Kotiaho JS, Simmons LW. 2003. Longevity cost of reproduction for males but no longevity cost of mating or courtship for females in the male-dimorphic dung beetle Onthophagusbinodis. *J Insect Physiol.* 49:817–822.
Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nat Rev Genet.* 4:865–875.
Long M, Langley CH. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95.
Loppin B, Lepetit D, Dorus S, Couble P, Karr TL. 2005. Origin and neofunctionalization of a *Drosophila* paternal effect gene essential for zygote viability. *Curr Biol.* 15:87–93.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
Mata J, Curado S, Ephrussi A, Rørth P. 2000. Tribbles coordinates mitosis and morphogenesis in *Drosophila* by regulating string/CDC25 proteolysis. *Cell* 101:511–522.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351(6328):652–654.
Meisel RP, Hilldorfer BB, Koch JL, Lockton S, Schaeffer SW. 2010. Adaptive evolution of genes duplicated from the *Drosophila pseudoobscura* neo-X chromosome. *Mol Biol Evol.* 27(8):1963–1978.
Moriyoshi K, Iijima K, Fujii H, Ito H, Cho Y, Nakanishi S. 2004. Seven in absentia homolog 1A mediates ubiquitination and degradation of group 1 metabotropic glutamate receptors. *Proc Natl Acad Sci U S A.* 101:8614–8619.
Nei M, Rogozin IB, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci U S A.* 97(20):10866–10871.
Nurminsky DI, Nurminskaya MV, Benevolenskaya EV, Shevelyov YY, Hartl DL, Gvozdev VA. 1998. Cytoplasmic dynein intermediate-chain isoforms with different targeting properties created by tissue-specific alternative splicing. *Mol Cell Biol.* 18:6816–6825.
Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* 396:572–575.
Ohno S. 1970. Evolution by gene duplication. New York: Springer.
Paukku S, Kotiaho JS. 2005. Cost of reproduction in *Callosobruchusmaculatus*: effects of mating on male longevity and the effect of male mating status on female longevity. *J Insect Physiol.* 51:1220–1226.
Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays* 29(4):386–391.
Robertson HM, Preston CR, Phillis RW, Johnson-Schlitz DM, Benz WK, Engels WR. 1988. A stable genomic source of P element transposase in *Drosophila melanogaster*. *Genetics* 118:461–470.
Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
Rubin GM, Spradling AC. 1982. Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218:348–353.
Sakai N, Sawada MT, Sawada H. 2004. Non-traditional roles of ubiquitin-proteasome system in fertilization and gametogenesis. *Int J Biochem Cell Biol.* 36:776–784.
Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526–538.
Sharp PM, Li WH. 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *J Mol Evol.* 25:58–64.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
Vijay-Kumar S, Bugg CE, Cook WJ. 1987. Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol.* 194:531–544.
Wang W, Brunet FG, Nevo E, Long M. 2002. Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 99:4448–4453.
Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, Polpitiya A, Petritis K, Karr TL. 2010. The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J Proteomics.* 73:2171–2185.
Wu CI, Xu EY. 2003. Sexual antagonism and X inactivation—the SAXI hypothesis. *Trends Genet.* 19:243–247.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Yin L, Joshi S, Wu N, Tong X, Lazar MA. 2010. E3 ligases Arf-bp1 and Pam mediate lithium-stimulated degradation of the circadian heme receptor Rev-erb alpha. *Proc Natl Acad Sci U S A.* 107:11614–11619.
Zera AJ, Harshman LG. 2001. Physiology of life history trade-offs in animals. *Annu Rev Ecol Syst.* 32:95–126.
Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
Zhang J, Dean AM, Brunet F, Long M. 2004. Evolving protein functional diversity in new genes of *Drosophila*. *Proc Natl Acad Sci U S A.* 101:16246–16250.
Zhang YE, Vibranovski MD, Krinsky BH, Long M. 2010. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Res.* 20:1526–1533.
Zhou Q, Zhang G, Zhang Y, et al. 2008. On the origin of new genes in *Drosophila*. *Genome Res.* 18:1446–1455.