Recurrent Tandem Gene Duplication Gave Rise to Functionally Divergent Genes in Drosophila

Chuanzhu Fan,†1 Ying Chen,‡2 and Manyuan Long*†
*Department of Ecology and Evolution, The University of Chicago; and †The Committee on Genetics, The University of Chicago

Tandem gene duplication is one of the major gene duplication mechanisms in eukaryotes, as illustrated by the prevalence of gene family clusters. Tandem duplicated paralogs usually share the same regulatory element, and as a consequence, they are likely to perform similar biological functions. Here, we provide an example of a newly evolved tandem duplicate acquiring novel functions, which were driven by positive selection. CG32708, CG32706, and CG6999 are 3 clustered genes residing in the X chromosome of Drosophila melanogaster. CG6999 and CG32708 have been examined for their molecular population genetic properties (Thornton and Long 2005). We further investigated the evolutionary forces acting on these genes with greater sample sizes and a broader approach that incorporate between-species divergence, using more variety of statistical methods. We explored the possible functional implications by characterizing the tissue-specific and developmental expression patterns of these genes. Sequence comparison of species within D. melanogaster subgroup reveals that this 3-gene cluster was created by 2 rounds of tandem gene duplication in the last 5 Myr. Based on phylogenetic analysis, CG32708 is clearly the parental copy that is shared by all species. CG32706 appears to have originated in the ancestor of Drosophila simulans and D. melanogaster about 5 Mya, and CG6999 is the newest duplicate that is unique to D. melanogaster. All 3 genes have different expression profiles, and CG6999 has in addition acquired a novel transcript. Biased polymorphism frequency spectrum, linkage disequilibrium, nucleotide substitution, and McDonald–Kreitman analyses suggested that the evolution of CG6999 and CG32706 were driven by positive Darwinian selection.

Introduction

It is well recognized that gene duplication is prevalent in eukaryotes. Genomic analyses of model organisms have shown that over one-third of all protein-coding genes belong to multigene families (Rubin et al. 2000; Kent et al. 2003). The mechanisms of gene duplication can be classified based on its scale (e.g., whole-genome duplication, segmental duplication, and tandem gene duplication) or whether it is RNA mediated (retroposition and transposition). Comparative genomic analysis between closely related species has revealed that tandem duplication is one of major mechanisms creating new genes, particularly genes clustered into a gene family, which have been documented in many organisms (e.g., Anderson and Roth 1977; Stark 1993; Brown et al. 1998; Eichler and Sankoff 2003; Leister 2004; Cardoso et al. 2006; Ponce and Hartl 2006; Shoja and Zhang 2006; Tuskan et al. 2006; Hazkani-Covo and Graur 2007). It is believed that tandem gene duplication could arise by unequal crossing over, which results from homologous recombination between paralogous sequences or nonhomologous recombination by replication-dependent chromosome breakages (Arguello et al. 2007).

A newly duplicated gene must overcome substantial hurdles before fixation. Once fixed, duplicated genes may face different fates: 1 of the 2 copies could lose its function and become pseudogenized due to the accumulation of degenerative mutations or both copies can maintain the same function. It is also possible that the 2 copies can accumulate different mutations leading to the duplicated genes taking on different roles that had previously been performed by the original gene, a process known as subfunctionalization. The most remarkable fate of gene duplication is neofunctionalization, whereby the new copy evolves a novel function driven and maintained by selection, whereas the old copy still retains the original function.

The location of duplicated copy can be adjacent to the original (tandem) or somewhere else in the genome (dispersed), for example, the duplicate generated by RNA-mediated retrotransposition. Separated from their regulatory elements, the dispersed duplicated copies will likely evolve novel functions by recruiting new regulatory elements (e.g., Wang et al. 2002). In contrast, the tandemly duplicated gene would tend to maintain a similar function to their parental copy due to their sharing the same regulatory elements and this has been demonstrated in many examples (e.g., Li et al. 2006; Ponce and Hartl 2006; Arisue et al. 2007). Given the apparent importance of tandem gene duplication for gene expansion in the eukaryotes, it is of great interest to know whether the tandem gene duplication can also generate novel functions. It has been recognized that gene duplication followed by divergence is one of the most important mechanism for generating new genes with novel functions, and such genetic novelty could involve in increasing organismic complexity, speciation, and adaptation processes (Long et al. 2004; Roth et al. 2007).

Here, we provide an example of tandemly duplicated genes acquiring novel transcription patterns, which could potentially lead to novel biological functions. CG32708, CG32706, and CG6999 are 3-gene cluster on the X chromosome of Drosophila melanogaster. By investigating their homologous counterparts in the D. melanogaster subgroup species, we have found that this 3-gene cluster was created by 2 rounds of tandem gene duplication in the last 5 Myr. Though the newly duplicated copies (CG32706 and CG6999) have diverged biological functions from their parental copy (CG32708), they share great similarity both in their DNA and protein sequences with only few substitutions in D. melanogaster. All the 3 genes have different
expression patterns, which can potentially lead to diverged biological functions. Particularly, we have found that the newest duplicate, CG6999, has a novel transcript with shorter sequence compared with its major transcript. We further observed that the homologous copy of CG32706 in species Drosophila simulans, Drosophila mauritiana, and Drosophila sechellia has undergone extensive sequence divergence compared with D. melanogaster CG32706. Sequence divergence and population genetic tests strongly suggested that CG6999 and CG32706 evolved under positive selection.

Materials and Methods

Stocks, Sampling, and DNA Extraction

We used isofemale stocks of D. melanogaster (Oregon-R), D. mauritiana, D. sechellia, Drosophila yakuba, Drosophila teissieri, Drosophila santomea, and Drosophila erecta, which have been kept in our laboratory for over 50 generations. We sequenced a collection of Drosophila lines to generate the polymorphism for the 3 genes in D. melanogaster and D. simulans. The 26 isofemale D. melanogaster strains sampled include 10 from North America (NA), 7 from Zimbabwe (ZS), 5 from Taiwan, and 4 from Israel (FS). The polymorphism of D. simulans was generated from 22 population samples, of which 6 are from Africa, 6 from NA, 5 from France, 2 from FS, 2 from South America, and 1 from Southern Pacific Cook Island. To avoid potential problem with population structure within McDonald–Kreitman (MK) test (McDonald and Kreitman 1991), we restricted our analysis to the 6 D. simulans African ancestral strains to test the evolution of CG32706 using a MK test.

Genomic DNA of D. melanogaster, D. simulans, D. mauritiana, D. sechellia, D. yakuba, D. teissieri, D. santomea, and D. erecta was extracted using Puregene DNA isolation kits (Genta Systems, Minneapolis, MN) from 25–30 flies (for microarray hybridization, Southern blotting, and genomic DNA polymerase chain reaction [PCR] amplification) or single male fly (for the D. melanogaster and D. simulans population genetic analysis). We did not observe any potential heterozygous sites in the sequence traces, so the sequence from each individual was considered to be a haplotype.

Duplication Identification, DNA Amplification, and Sequencing

We first identified the potential duplicated candidates in the D. melanogaster subgroup species using microarray-based comparative genomic hybridization (CGH) methods. Genomic DNA was digested using DNeasi, and 3′ termini of the fragmentation products were labeled with biotin-dideoxyuridine triphosphate (ddUTP). The target DNA fragments (~100–150 bp) were hybridized onto The GeneChip Drosophila melanogaster Genome Array following the standard Affymetrix protocol (Affymetrix, Santa Clara, CA). The ratio of pairwise comparisons for each probe was calculated using hybridization intensity among 8 species, and the median value of intensity fold-change in all probes for each feature was taken as threshold for gene duplication criterion. The detailed methodology for the duplication identification was described in Fan and Long (2007).

PCR were performed in the standard thermal cycler using Invitrogen Taq polymerase following the manufacturer’s protocol, with annealing temperature adjusted based on the length of fragments with 1 kb/min. The double-stranded PCR products were purified using a Qiagen PCR purification kit or Qiagen miniprep Gel purification system. Purified PCR products were sequenced using Applied Biosystems 3730XL 96-capillary automated DNA sequencer. The entire fragments of the blocks were sequenced using the sequence walk procedure. Sequences were edited and assembled. ClustalX was used to align sequences for further analyses (Thompson et al. 1997). Manual adjustments were made where necessary.

Expression Analysis

Retrotranscription (RT)–PCR was used to analyze the expression profile in different developmental stages and tissues. Total RNA was extracted from D. melanogaster adult virgin females, males, 2-hour-old eggs, second- and third-instar larvae, and pupae using a Qiagen total RNA extraction kit. We examined tissue differential expression pattern by RT-PCR with RNA extractions from the head, body without ovary/testis, accessions gland, and ovary/testis. Testis and ovary were obtained by dissecting mature male and female flies in saline solution, and removed testis and ovary were preserved in RNA later solution. Total RNA was extracted from flies or tissues following the Qiagen protocol.

Population Genetic Analysis

Basic population genetic analyses were performed in DnaSP (Rozas et al. 2003). The sequence diversity was quantified as nucleotide diversity (π) (Nei 1987) and Watterson’s θ (1975). Tests of deviation from neutrality were conducted using tests from Tajima (1989), Fu & Li (1993), and Fay & Wu (2000), and significance was assessed using coalescent simulations. The neutral coalescent process was simulated using 2,000 replicates with the number of segregating sites set to that observed in the data. However, these approaches, based on the polymorphic spectrum, are of limited utility in testing for neutrality in young genes because a reduced level of diversity and skew toward rare are expected (Thornton 2007). Therefore, we also used MK test as implemented in DnaSP. In the MK test for CG32706, we compared the polymorphism generated from the D. simulans lines and fixed mutations between D. simulans and D. mauritiana. In addition, we also investigated linkage disequilibrium (LD).

Phylogenetic Analysis and Sequence Divergence Calculation

The phylogenetic analysis for both DNA and protein sequence was performed using the Neighbor-Joining and
maximum likelihood methods implemented in PAUP* 4.0b10 (Swofford 2002), with 10,000 bootstrap replicates to assess support. We further calculated the number and rate of nonsynonymous and synonymous substitution for 3 genes in *D. melanogaster* using codon model (Codeml) implemented in PAML (Yang 1997, 2007) under a model in which all branches were allowed their own $K_a/K_s$ value. To generate the data using free ratio for each branch, we aligned the coding sequences of CG32708, CG32706, and CG6999. Because we have determined CG32708 as the parental copy of the other 2 genes, the tree for Codeml analysis used CG32708 as outgroup. The numbers of synonymous and nonsynonymous substitution along each branch were calculated under a model in which the $K_a/K_s$ ratio ($\omega$) was free to vary along each branch. (Goldman and Yang 1994; Yang 1997).

We calculated the $K_a/K_s$ ratio using maximum likelihood algorithm using Perl script incorporated PAML for CG32706 homologues between species *D. simulans, D. mauritiana*, and *D. sechellia*. The significance of $K_a/K_s$ that deviated from neutrality ($\omega = 1$) was tested using likelihood ratio test (Yang 1998).

**Results**

**Tandem Duplication in *D. simulans, D. mauritiana*, and *D. melanogaster***

Our initial microarray CGH suggested that there are multiple homologous copies of CG6999 in the species of *D. melanogaster*. By blasting the candidate sequences that we identified from CGH against the genomic sequence of *D. melanogaster*, we found 3 copies, CG32708, CG32706, and CG6999, closely adjacent with only 500 bp separating the protein-coding sequences on the X chromosome near 8C5 (fig. 1). To characterize the gene content and structure in all *D. melanogaster* subgroup species, we designed a pair of primers that were located in the flanking sequences of CG32708 and CG6999 (fig. 1) to amplify and obtain the homologous sequences in all *D. melanogaster* subgroup species. The PCR and sequencing results indicated that a single homologue is present in *D. yakuba, D. erecta, D. santomea, and D. teissieri*, and 2 homologous copies in *D. simulans* and *D. mauritiana* (fig. 1). Phylogenetic analysis clearly showed that 2 duplication events occurred in the last 5 Myr. The first duplication event occurred before the divergence of ancestor of *D. melanogaster* and *D. simulans* approximately 5 Mya, and a more recent duplication happened in the branch of *D. melanogaster* in the last 1–2 Myr (fig. 2).

**Expression Analysis by RT-PCR**

We performed RT-PCR to investigate the pattern of gene expression across tissues and developmental stages in *D. melanogaster* for all 3 genes. Overall, the expression levels were different among the 3 genes, with CG6999 having the highest expression and CG32706, the lowest expression (fig. 3). Interestingly, we found 2 transcripts of CG6999 in both female and male flies (fig. 3a). Transcript “B” is a novel shorter transcript that has a 5′ splicing site located in the first exon of transcript “A.” To dissect the differential expressions of the novel transcript of CG6999, we conduct RT-PCR to examine the expression profile using different tissues and found that only reproductive organs (testis and ovary) show the expression of the novel CG6999 transcript “B” (fig. 3c and d). The differential expression profiles of the 3 genes also appear to be consistent across different developmental stages. CG32708 and CG32706 tend to have lower expression in second-instar larva than in third-instar larva and pupa. CG6999, however, has an equal expression level in both transcripts (A and B) (fig. 3b).

**Sequence Divergence of the 3 Genes**

We estimated the $K_a/K_s$ ratio of CG32708 and CG32706 across species. The average $K_a/K_s$ ratio for CG32708 in all *D. melanogaster* subgroup species is equal...
to 0.23, which indicates that CG32708 are under strong functional constraints. However, the sequence of CG32706 is highly diverged between the clade of *D. simulans*–*D. mauritiana*–*D. sechellia* and the clade of *D. melanogaster*, and there are extensive deletions in the 5′ and 3′ ends of the sequences (supplementary fig. 1, Supplementary Material online), which is even higher than the sequence divergence in the intron regions (data not shown). Because CG6999 is a novel gene in the species of *D. melanogaster*, we calculated its $K_a/K_s$ values against *D. melanogaster* CG32706. Overall, 7 nonsynonymous substitutions, 1 synonymous substitution, and a 6-base insertion are observed (fig. 4). The ratio of $K_a/K_s$ (1.6) along the CG6999 lineage indicates that CG6999 has undergone accelerated divergence after the gene duplication event, and because the ratio is largely greater than 1, it seems likely that this was driven by positive selection. The distribution of the substitutions and insertions are primarily located at near 5′ or 3′ end of the gene. This biased distribution of substitutions is also seen in *D. simulans* CG32706.

**Positive Selection of CG6999 in *D. melanogaster***

To further investigate whether adaptive evolution had affected these 3 genes, we collected polymorphism data from 26 *D. melanogaster* lines. Because the local population structure can lead to a departure from neutrality under certain tests, we tested for gene flow and population subdivision using $F_{st}$. The $F_{st}$ values clearly show the high gene flow and low population subdivision among 4 local populations (table 1). Among the 3 genes, only CG6999 shows a significant bias in the site-frequency spectrum away from neutral expectations (table 2). The negative values of these

---

**Table 1**

<table>
<thead>
<tr>
<th>Pop1</th>
<th>Pop2</th>
<th>3-Gene Cluster</th>
<th>CG32708</th>
<th>CG32706</th>
<th>CG6999</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>ZS</td>
<td>0.45109</td>
<td>0.16667</td>
<td>0.46336</td>
<td>0.15781</td>
</tr>
<tr>
<td>NA</td>
<td>FS</td>
<td>0.06476</td>
<td>0.07407</td>
<td>0.0331</td>
<td>0.05761</td>
</tr>
<tr>
<td>NA</td>
<td>TWN</td>
<td>-0.05289</td>
<td>0.04762</td>
<td>-0.10624</td>
<td>-0.06702</td>
</tr>
<tr>
<td>ZS</td>
<td>FS</td>
<td>0.44062</td>
<td>0</td>
<td>0.608</td>
<td>0</td>
</tr>
<tr>
<td>ZS</td>
<td>TWN</td>
<td>0.49308</td>
<td>0</td>
<td>0.5858</td>
<td>0</td>
</tr>
<tr>
<td>FS</td>
<td>TWN</td>
<td>0.03468</td>
<td>0</td>
<td>-0.06061</td>
<td>0</td>
</tr>
</tbody>
</table>

**NOTE.**—TWN, Taiwan.
tests suggest that either positive selection or demographic process (e.g., older bottleneck, population expansion, recently fixed duplication, and hidden population structure) drove the evolution of CG6999.

We further investigated the above possibilities by performing LD analysis. An LD test covering the entire of the 3 genes regions was conducted, and the significant associations were estimated using homologous sequences of Drosophila simulans as outgroup and that of CG6999 was estimated using Drosophila melanogaster CG32706 sequence as outgroup. N, population size; L, gene length (bp); S, the number of segregation sites.

### Table 2
Levels of Polymorphism in Drosophila melanogaster and Neutrality Tests on the Site-Frequency Spectrum

<table>
<thead>
<tr>
<th>Summary Statistic</th>
<th>CG32708</th>
<th>CG32706</th>
<th>CG6999</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>L</td>
<td>798</td>
<td>766</td>
<td>786</td>
</tr>
<tr>
<td>S</td>
<td>4</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>H</td>
<td>0.00064</td>
<td>0.0043</td>
<td>0.0017</td>
</tr>
<tr>
<td>Θ</td>
<td>0.00131</td>
<td>0.0038</td>
<td>0.0053</td>
</tr>
</tbody>
</table>

Tajima’s D

-1.36, *P* = 0.08

0.47, *P* = 0.73

-2.30, *P* = 0.001

Fu & Li’s *D*

-0.90, *P* = 0.32

-0.48, *P* = 0.37

-3.64, *P* = 0.001

Fay & Wu’s *H*

0.39, *P* = 0.55

-2.043, *P* = 0.019

-10.38, *P* = 0.001

* The significance as *P* < 0.01.

* The significance as *P* < 0.05.

Positive Selection of CG32706 Orthologs in D. simulans, D. sechellia, and D. mauritiana

We calculated the *Ks/Ka* ratio of CG32706 in the species D. simulans, D. sechellia, and D. mauritiana (table 4). The average *Ks/Ka* is equal to 2.067. The *Ks/Ka* ratio between D. sechellia and D. mauritiana is significantly greater than 1 (3.406, *P* = 0.05), suggesting positive selection (table 4). The result of the MK test revealed a significant excess of replacement substitutions between species, indicating strong positive selection acting on CG32706 after the species diverged within the D. simulans clade (table 5). Moreover, the polymorphism analysis revealed that CG32706 is not a pseudogene, as shown by the excess of synonymous substitution in D. simulans polymorphism spectrum (table 6).

### Discussion
Positive Selection Drive the Evolution of CG32706 and CG6999

The sequence and phylogenetic analyses from the D. melanogaster subgroup species clearly suggest the 3-gene cluster in D. melanogaster is a product of 2 rounds of gene duplication, with CG6999 originating 1–2 Mya and CG32706 derived from 5 Mya. Thornton and Long (2005) previously generated sequence polymorphism data for CG6999 and CG32708 (synonymyms to CG6997 in Thornton and Long 2005) from 10 ZS D. melanogaster lines. They compared the parologs of CG32708 and CG6999 in D. melanogaster using population genetics and MK analyses and found no evidence for selection for new protein functions after gene duplication. In this study, we further pursued this question by using a combination of polymorphism and divergence analyses using comparative sequences from all D. melanogaster subgroup species. Several complement lines of evidence suggest that the evolution of CG6999 was likely to have been driven by positive Darwinian selection. First, the significant skew toward rare alleles in the site-frequency spectrum suggests an excess of rare allele in the D. melanogaster population particularly in the gene CG6999. Second, the LD test indicates that CG6999 has a remarkably high number of the significantly associated sites, consistent with the notion that CG6999 is linked to a site that is under selection within or immediately outside the gene region. Third, the excess of nonsynonymous substitutions occurs after gene duplication between CG32706 and CG6999 in D. melanogaster (fig. 4).

### Table 3
The Number of Significant Associations between Pairs of SNPs

<table>
<thead>
<tr>
<th>Region (bp)</th>
<th>CG32708</th>
<th>CG32706</th>
<th>CG6999</th>
<th>CG32708–CG32706</th>
<th>CG32708–CG6999</th>
<th>CG32706–CG6999</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>1,106</td>
<td>1,227</td>
<td>1,236</td>
<td>2,333</td>
<td>3,569</td>
<td>2,463</td>
<td>3,569</td>
</tr>
<tr>
<td>Significant associations</td>
<td>2</td>
<td>11</td>
<td>19</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>56</td>
</tr>
</tbody>
</table>

* Pairwise comparisons show a significant association after applying the Bonferroni procedure.
We further noticed that the sequences of CG32706 in *D. simulans* and *D. mauritiana* are highly diverged with a significant Kₐ/Kₛ ratio deviating from neutrality, which apparently shows the accelerated rate of evolution after it diverged from its ancestor CG32708. Moreover, the significant Fay & Wu's test for CG32706 indicated that selection drove the excess of high-frequency alleles in the *D. melanogaster* population, and the MK test of CG32706 orthologues in *D. simulans* and *D. mauritiana* strongly suggests positive selection after species divergence in the *D. simulans* sister group.

Functional Divergence, Speciation, and Novel Transcript of Tandemly Duplicated Genes

CG32708, CG32706, and CG6999 are believed to play a role in the RNA-binding and alternative-splicing pathway (Park et al. 2004), though they act at different stages. CG32706 is an RNA-binding protein, which interacts selectively with premessenger RNA or messenger RNA (mRNA). CG6999 plays a role in the regulation of alternative nuclear mRNA splicing via the spliceosome (Dimova et al. 2003; Park et al. 2004). The cellular and biological function of CG32708 remains unclear. Based on the homologous sequences of its paralog in *D. melanogaster*, we believe it also plays a role in the RNA-splicing pathway. However, our evidence from both the expression data and the evolutionary analysis of sequences demonstrated that the new genes are likely to have replaced the major functions of their parental copies with the expansion of molecular and biological functionality. Interestingly, we have found a nontandem duplicated homologous copy, CG10993, in this gene family in *D. melanogaster*. CG10993 is located in 12C5 of X chromosome. To determine the relationship and function of CG10993 with other 3 homologous members, we performed a phylogenetic analysis and further calculated Kₐ and Kₛ values. The phylogenetic tree shows CG10993 resides in the basal lineage and has a closer relationship with *D. yakuba* CG32708 (fig. 5). Therefore, CG10993 is likely to be the ancestor of the 3 tandem genes diverged over 10 Mya. The Kₐ/Kₛ (0.3363/0.5614 = 0.599) test between CG10993 and CG32708 suggested that the 2 duplicated genes were subject to purifying selection.

It has been observed that X-linked genes evolved rapidly after gene duplication and the X chromosome appears to be a fertile ground for gene duplication and evolution. For example, studies from *Drosophila* and mammals found that the significant excess of retroposed genes originated from X chromosome, likely under adaptive evolution (Betran et al. 2002; Emerson et al. 2004). A previous genome analysis has predicted an excess of new genes on the X that are tandem duplicates, based on the fact that Kₛ seems to be low on the X between duplicates compared with the autosomes (Thorton and Long 2002). A recent investigation further demonstrated that the *Drosophila* X chromosome not only shows rapid origination and evolution of retroposed genes but also imposes noncoding RNA genes under positive selection through DNA-level duplication (Levine et al. 2006). Notably, our 3 gene cluster resides closely (less 1 cM distance) between CG32712 and CG32690, the 2 fast-evolved noncoding RNA genes found by Levine et al. (2006). Such coincidence may indicate a hot spot in the X chromosome for the novel gene origination under adaptive evolution. Furthermore, a whole-genome polymorphism and divergence analysis in *D. simulans* similarly found significantly less polymorphism and faster divergence on the X chromosome, indicating that X-linked genes are influenced by adaptive evolution (Begun et al. 2007).

It is well known that transcription is driven by regulatory elements, the interaction between the sequence-specific binding transcriptional factors (the trans-elements) and their DNA recognition sites (the cis-elements). In addition to the cis–trans interaction, the spatial and temporal expression of genes coding transcription factor genes can also regulate expression patterns, in which transcription

![Figure 5](https://example.com/fig5.png)

**Fig. 5.**—Maximum likelihood tree of *Drosophila melanogaster* CG6999 gene family. *Drosophila yakuba* CG32708 serves as outgroup. Bootstrap supports shown above the branches.

---

### Table 4

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Kₛ</th>
<th>Kₐ</th>
<th>Kₛ/Kₐ</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dsim/Dsec</td>
<td>0.063</td>
<td>0.072</td>
<td>1.143</td>
<td>P = 0.81</td>
</tr>
<tr>
<td>Dsim/Dmau</td>
<td>0.040</td>
<td>0.097</td>
<td>2.425</td>
<td>P = 0.42</td>
</tr>
<tr>
<td>Dmau/Dsec</td>
<td>0.032</td>
<td>0.109</td>
<td>3.406</td>
<td>P = 0.05*</td>
</tr>
<tr>
<td>Average</td>
<td>0.045</td>
<td>0.093</td>
<td>2.067</td>
<td>P = 0.19</td>
</tr>
</tbody>
</table>

* The significance as P < 0.05.

### Table 5

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Divergence</th>
<th>Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsynonymous</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Synonymous</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Fisher exact test</td>
<td>P = 0.015*</td>
<td></td>
</tr>
</tbody>
</table>

* The significance as P < 0.05.
Supplementary Materials

Supplementary figures 1 and 2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank a number of people: Chung-I Wu, Jerry Coyne, Peter Andolfatto, and Eviatar Nevo for providing fly strains; Xinning Li for performing the microarray hybridization; members in the Long laboratory for the valuable discussions and inputs; Adam Eyre-Walker and Stuart Wigby for critical reading the manuscript; and 3 anonymous reviewers for their critical comments and valuable suggestions. This work is supported by National Institutes of Health and National Science Foundation grants to M.L.

Literature Cited


Marcy Uyenoyama, Associate Editor

Accepted April 7, 2008