

Recurrent Events of Positive Selection in Independent *Drosophila* Lineages at the Spermatogenesis Gene *roughex*

Ana Llopart¹ and Josep M. Comeron

Department of Biology, University of Iowa, Iowa City, Iowa 52242

Manuscript received December 20, 2007

Accepted for publication April 8, 2008

ABSTRACT

Our understanding of the role of positive selection in the evolution of genes with male-biased expression can be hindered by two observations. First, male-biased genes tend to be overrepresented among lineage-specific genes. Second, novel genes are prone to experience bursts of adaptive evolution shortly after their formation. A thorough study of the forces acting on male-biased genes therefore would benefit from phylogenywide analyses that could distinguish evolutionary trends associated with gene formation and later events, while at the same time tackling the interesting question of whether adaptive evolution is indeed idiosyncratic. Here we investigate the *roughex* (*rux*) gene, a dose-dependent regulator of *Drosophila* spermatogenesis with a C-terminal domain responsible for nuclear localization that shows a distinct amino acid sequence in the *melanogaster* subgroup. We collected polymorphism and divergence data in eight populations of six *Drosophila* species, for a total of 99 *rux* sequences, to study rates and patterns of evolution at this male-biased gene. Our results from two phylogeny-based methods (PAML and HyPhy) as well as from population genetics analyses (McDonald–Kreitman-based tests) indicate that amino acid replacements have contributed disproportionately to divergence, consistent with adaptive evolution at the Rux protein. Analyses based on extant variation show also the signature of recent selective sweeps in several of the populations surveyed. Most important, we detect the significant and consistent signature of positive selection in several independent *Drosophila* lineages, which evidences recurrent and concurrent events of adaptive evolution after *rux* formation.

GENOMEWIDE analyses of gene expression in a variety of organisms including *Drosophila* have revealed differences in messenger RNA abundances between the two sexes for a substantial fraction of genes (JIN *et al.* 2001; ARBEITMAN *et al.* 2002; PARISI *et al.* 2003; RANZ *et al.* 2003; GIBSON *et al.* 2004; REINKE *et al.* 2004; RINN *et al.* 2004). On the other hand, several molecular evolutionary studies have now described that genes with sex-biased expression show fast rates of protein evolution (NUZHIDIN *et al.* 2004; ZHANG *et al.* 2004; ZHANG and PARSCH 2005). This observation is consistent with earlier findings showing that reproductive proteins evolve more quickly than proteins with non-sex-related functions, granted that much of the difference in expression profiles between sexes occurs in the germ line (THOMAS and SINGH 1992; CIVETTA and SINGH 1995; VACQUIER 1998; BEGUN *et al.* 2000; WYCKOFF *et al.* 2000; BIRKHEAD and PIZZARI 2002; SWANSON and VACQUIER 2002; CASTILLO-DAVIS *et al.* 2004). Additionally, there are several examples of genes where these high rates of amino acid replacements result from adaptive protein evolution likely triggered

by sexual selection and/or sexual conflict (AGUADÉ *et al.* 1992; CIVETTA and SINGH 1998; AGUADÉ 1999; BIRKHEAD and PIZZARI 2002; SWANSON and VACQUIER 2002; BEGUN and LINDFORS 2005; ELLEGREN and PARSCH 2007). Although there is increasing evidence for adaptive evolution in female reproductive genes (SWANSON *et al.* 2001, 2003; TURNER and HOEKSTRA 2006; KELLEHER *et al.* 2007; LAWNICZAK and BEGUN 2007), most reported cases pertain to the male class (TORGERSON *et al.* 2002; SWANSON *et al.* 2003; GOOD and NACHMAN 2005; NIELSEN *et al.* 2005; CIVETTA *et al.* 2006; PROSCHEL *et al.* 2006; BAUER DUMONT *et al.* 2007).

Our understanding of how natural selection works on male-biased genes in *Drosophila* can be confounded, however, by two features. First, male-biased genes have higher effective birth and extinction rates, and this appears to be widespread among several *Drosophila* genomes (BEGUN and LINDFORS 2005; WAGSTAFF and BEGUN 2005a,b; BEGUN *et al.* 2006, 2007; LEVINE *et al.* 2006; HAERTY *et al.* 2007; ZHANG *et al.* 2007). In agreement with this idea, a large fraction of novel genes show sex-specific expression, and testis-specific expression in particular (LEVINE *et al.* 2006). Second, novel genes, evolved through gene duplication (LONG *et al.* 2003) or by “*de novo*” recruitment of noncoding DNA (LEVINE *et al.* 2006), are prone to bursts of adaptive evolution

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU476916–EU477049.

¹Corresponding author: Department of Biology, University of Iowa, 215 Biology Bldg. (BB), Iowa City, IA 52242. E-mail: ana-llopart@uiowa.edu

shortly after their birth (LONG and LANGLEY 1993; BETRAN and LONG 2003; JONES and BEGUN 2005; CHEN *et al.* 2007). Thus, a thorough study of selective trends acting on male-biased genes would benefit from phylogenywide analyses that could take into account evolutionary trends occurring shortly after gene formation and later events.

Here we analyze the molecular evolution of the *roughex* (*rux*) gene, a dose-dependent regulator of spermatogenesis (GONCZY *et al.* 1994), in several independent lineages of *Drosophila*. *rux* is indeed a male-biased gene, with transcript levels in males significantly higher than in females (PARISI *et al.* 2003; RANZ *et al.* 2003; RIFKIN *et al.* 2003; GIBSON *et al.* 2004), and earlier evolutionary analyses of the Rux protein by AVEDISOV *et al.* (2001) reported rapid rates of amino acid replacements attributed to low overall selective constraints. Interestingly, our analysis of the recently available 12 *Drosophila* genomes (CLARK *et al.* 2007) reveals that the Rux domain responsible for proper intracellular (nuclear) localization is present only in the *melanogaster* subgroup. This suggests that the *rux* function (*sensu Drosophila melanogaster*) likely evolved 8–15 million years ago (LACHAISE *et al.* 1988). In all, *rux* is an ideal candidate for investigating whether adaptive evolution in male-biased proteins occurs repeatedly in independent lineages or if, in contrast, adaptation is idiosyncratic in nature (LEVINE and BEGUN 2007).

To obtain a comprehensive picture of the evolutionary trends in *rux* we analyzed polymorphism data in eight populations from six species of the *melanogaster* subgroup (for a total of 99 chromosomes). Our study reveals the signature of recurrent events of adaptive evolution, with a significant excess of amino acid replacements fixed between species. More interestingly, positive (Darwinian) selection in the Rux protein has occurred independently in several lineages and these concurrent events are detected as bursts of amino acid replacements after *rux* formation as well as patterns of polymorphism consistent with more contemporary selective sweeps in different species.

MATERIALS AND METHODS

Drosophila stocks and population surveys: To obtain polymorphism and divergence data we surveyed eight populations in six *Drosophila* species: one population each of *D. santomea*, *D. melanogaster*, *D. mauritiana*, and *D. sechellia* and two populations each (one island and one mainland) of *D. simulans* and *D. yakuba*. The Tucson *Drosophila* Stock Center made available *D. simulans* flies from the Democratic Republic of Congo, and we used also ethanol-preserved males of *D. simulans* collected in the island of São Tomé. T. F. Mackay kindly provided isofemale lines of *D. yakuba* from Abidjan, Ivory Coast, and the “island” populations of *D. yakuba* and *D. santomea* were collected in São Tomé (LLOPART *et al.* 2005). Highly inbred lines of *D. melanogaster* from Raleigh, North Carolina, were kindly supplied by T. F. Mackay and are described in FRY *et al.* (1998). *D. mauritiana* and *D. sechellia*

lines were generously provided by J. A. Coyne. One additional stock of *D. mauritiana* (Tucson *Drosophila* Stock Center no. 14021-0241.07) was also used to investigate the exon–intron structure of the *rux* gene.

DNA preparation and sequencing: We extracted genomic DNA from single males using the Puregene DNA isolation kit for paraffin-embedded tissue (Gentra Systems, Minneapolis) with minor modifications. PCR reactions were performed using primers designed to either *D. melanogaster* or *D. yakuba* sequences from GenBank. To control for possible demography that may mimic the fingerprints of selection in island populations we used multilocus data from LLOPART *et al.* (2005) in the cases of *D. yakuba* and *D. santomea*, and for the population of *D. simulans* collected in São Tomé we amplified the putatively neutral gene *Xdh*. After cleanup of the PCR fragments using the Wizard MagneSil system (Promega, Madison, WI) both strands were sequenced directly on a 3730 DNA analyzer using the Big Dye 3.1 chemistry (Applied Biosystems, Foster City, CA). To edit the sequences and assemble the contigs we used the software Sequencher 4.1 (Gene Codes, Ann Arbor, MI). Haplotypes in the *Xdh* region were determined experimentally by performing cycle-sequencing reactions using allele-specific primers, whereas *rux* haplotypes were inferred directly because it is an X-linked gene. Alignment of the amino acid sequences was obtained using ClustalX (THOMPSON *et al.* 1997). We surveyed exon 2 (~1000 bp) of the *rux* gene. One additional region encompassing exon 1 (22 bp), intron 1 (~100 bp), and 5′-noncoding sequence (~600 bp) was also investigated in the *D. melanogaster*, *D. simulans*, *D. mauritiana*, and *D. sechellia* populations. In the 5′-noncoding region of *D. sechellia* we discovered the remnants of an unknown non-LTR retrotransposon (sequence AF237761 in GenBank), which was excluded from the study. We restricted our analyses to the portion of the 5′-noncoding sequence that could be aligned among all species. The *D. leissieri*, *D. erecta*, and *D. orena rux* sequences used here were published elsewhere (AVEDISOV *et al.* 2001), and we also used previously reported sequences of the *rux* gene in the *D. yakuba* and *D. santomea* populations from São Tomé (LLOPART *et al.* 2005).

To investigate the exon–intron structure of the *rux* gene in *D. mauritiana* we extracted total RNA from 1-day-old adult females and males using the RNAeasy kit (QIAGEN, Valencia, CA), and first-strand cDNA was obtained using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA). RT-PCR reactions were performed using primers designed to two exons. Cleanup of PCR products and sequencing were carried out as described above. Sequences newly reported in this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU476916–EU477049.

Inference about the Rux nuclear domain: Earlier studies using Southern blot analysis and PCR with degenerated primers failed to detect *rux* homologous sequences in several species outside the *melanogaster* subgroup (AVEDISOV *et al.* 2001). To explore the possibility that *rux* may be a novel gene emerged in the *melanogaster* subgroup, we searched all available *Drosophila* genomes using both BLASTP and TBLASTN programs, using default parameters (<http://flybase.net/blast/>). While BLASTP is used for finding similar amino acid sequences in annotated protein databases, TBLASTN allows you to compare a protein sequence to the six-frame translations of a nucleotide database. Our searches were performed using the complete amino acid sequence, the C-terminal domain (amino acids 188–335 in AVEDISOV *et al.* 2001), and the non-C portion of the *D. melanogaster* Rux (supplemental Figure S1).

Inference about selection: To detect the effects of positive selection using divergence data (one sequence per species analyzed) we first applied the maximum-likelihood method

(NIELSEN and YANG 1998; YANG *et al.* 2000) implemented in the program codeml of PAML (version 3.15) (YANG 1997). This method is based on the comparison of several site-specific models with different assumptions regarding the ratio of nonsynonymous to synonymous substitution rates ($\omega = d_N/d_S$) used as an estimate of the selective pressure across sites. We compared models M1 (nearly neutral) *vs.* M2a (positive selection) and M7 (beta) *vs.* M8 (beta and ω), using equilibrium codon frequencies calculated from the average nucleotide frequencies at third codon positions (CodonFreq = 2). Model M1 assumes two classes of sites, conserved sites with ω estimated from the data ($\omega < 1$) and neutral sites ($\omega = 1$), while M2a considers an extra third class of sites under positive selection ($\omega > 1$) (WONG *et al.* 2004; YANG *et al.* 2005). In M7 and M8 ω varies across sites following a beta distribution, with M8 having the additional discrete class of sites with $\omega > 1$.

We identified positively selected sites at which nonsynonymous substitutions occur at a higher rate than synonymous ones with the Bayes Empirical Bayes (BEB) calculation of posterior probabilities for site classes (YANG *et al.* 2005), implemented in codeml (PAML, models M2a and M8). To account for potential heterogeneity across sites in d_S we also applied two recently developed methods (KOSAKOVSKY POND and FROST 2005) executed by the software HyPhy (POND and FROST 2005) and available in the web-based interface DataMonkey (<http://www.datamonkey.org/>; POND and FROST 2005). Both methods are based on maximum-likelihood estimates, but one assumes that substitution rates across sites can vary according to a gamma distribution and infers the rate at which individual sites evolve (random-effects models, REL) while the other estimates the ratio of nonsynonymous to synonymous substitutions on a site-by-site basis (fixed-effects models, FEL).

McDonald and Kreitman (MK) tests (MCDONALD and KREITMAN 1991) as well as basic calculations of polymorphism statistics were performed using the program DnaSP 4.10 (ROZAS *et al.* 2003). To estimate the selection coefficients (γ) associated with nonlethal amino acid replacement mutations we used the Markov chain Monte Carlo (MCMC) method implemented in the web-based program MKPRF (<http://cbsuapps.tc.cornell.edu/mkprf.aspx>) (BUSTAMANTE *et al.* 2002, 2003; BARRIER *et al.* 2003). Posterior means and probabilities of $\gamma \leq 0$ were estimated after 10,000 iterations (1000 burn-in iterations).

To contrast whether polymorphism patterns were compatible with a standard neutral model we applied the *DH* test (ZENG *et al.* 2006), which is based on the combined study of *D* (TAJIMA 1989b) and *H* (FAY and WU 2000) statistics. The *H* statistic was calculated for *D. yakuba* and *D. santomea* using *D. teissieri* as outgroup and for *D. simulans*, *D. mauritiana*, and *D. sechellia* using *D. melanogaster* as outgroup, and for the *D. melanogaster* polymorphisms we used the *D. simulans* sequence. Statistical significance of the *DH* test was assessed using neutral coalescent simulations produced by the program ms (<http://home.uchicago.edu/~rhudson1/>; HUDSON 2002). Probabilities were estimated by generating 10,000 replicate samples with the same number of mutations as that observed in the data. Simulations were performed under the assumption of no recombination and with likely levels of intragenic recombination estimated using the maximum-likelihood method implemented in the program MaxHap (<http://home.uchicago.edu/~rhudson1/>; HUDSON 2001).

To determine whether polymorphism patterns at the *rux* gene could be explained by demographic events in island populations, we simulated different population dynamics using the program ms (<http://home.uchicago.edu/~rhudson1/>; HUDSON 2002). We modeled an ancestral panmictic population that undergoes a size reduction (*i.e.*, island colonization), experiences reduced constant size for some time period, and

finally expands by growing exponentially. The population growth is specified by the equation $N_2 = N_1 \exp^{-\alpha t}$, where N_2 is the current population size, N_1 is the size prior to the expansion, α is the growth parameter, and t is the time at which the growth began. Simulations to evaluate *DH* values in the *D. yakuba* and *D. simulans* island populations consider variables to be both the time at which the size reduction occurred and the period of constant reduced size. In contrast, the modeling of *D. santomea* assumes that the island colonization occurred ~0.45 million years ago, an estimate of the time of the *D. yakuba*–*D. santomea* split (LLOPART *et al.* 2002). A total of 10,000 replicates were used to estimate *P*-values.

RESULTS

Protein divergence in *Rux*: Rates of evolution at nonsynonymous (d_N) and synonymous (d_S) sites in the *rux* gene were first estimated by maximum likelihood using PAML. Consistent with an overall pattern of functional constraints on protein evolution in the nine species of the *melanogaster* subgroup, the average d_N/d_S ratio (ω) across sites and lineages is < 1 ($\omega = 0.47$; Table 1). Likelihood-ratio tests comparing models that assume no positive selection (M1 and M7) to models that allow for a fraction of sites evolving under positive selection (M2a and M8) show, however, that the observed pattern of divergence at *rux* is best explained by the inclusion of a class of sites under positive selective pressure ($P \leq 0.0001$). These comparisons suggest also that 5–7% of codons are evolving by Darwinian selection with estimates of $\omega \sim 4$. The BEB method identifies 17 codon sites as positively selected, with 3 sites having a very high ($P > 0.95$) posterior probability (Table 1).

We also investigated the signature of positive selection based on the d_N/d_S ratio by applying the FEL and REL methods (implemented in the HyPhy package), which incorporate not only nonsynonymous but also synonymous rate variation among codon sites explicitly (KOSAKOVSKY POND and FROST 2005; POND and FROST 2005). Results from FEL and REL show good agreement with those obtained using PAML (Table 1). In the nine species of the *melanogaster* subgroup the REL method identifies seven sites with very strong evidence for positive selection (posterior probabilities > 0.95) and the FEL approach detects three positively selected sites ($P < 0.05$). Notably, several codon sites are identified by all three methods (BEB, FEL, and REL) as positively selected.

To investigate whether the results suggesting positive selection based on divergence data are due to the effect of a single lineage we analyzed separately the two main clades of the *melanogaster* subgroup. The first clade (*D. yakuba*–*D. erecta*) includes *D. yakuba*, *D. santomea*, *D. teissieri*, *D. orena*, and *D. erecta*, while the second clade (*D. simulans*–*D. melanogaster*) includes *D. simulans*, *D. mauritiana*, *D. sechellia*, and *D. melanogaster* (Figure 1). In both clades an extra class of nonsynonymous sites with ω significantly > 1 is required to best explain the

TABLE 1
Results of positive selection in the *rux* gene from PAML and HyPhy

	<i>n</i>	<i>L</i>	<i>L_s</i>	<i>d_N/d_S</i>	<i>P</i>		Selection parameters (M8)	PSS		
					M1 vs. M2a	M7 vs. M8		M8	FEL	REL
<i>D. melanogaster</i> subgroup	9	1032	1.55	0.47	0.0001	<0.00001	$p_1 = 0.071$ $\omega_1 = 3.92$	17 (3)	15 (3)	52 (7)
<i>D. yakuba</i> – <i>D. erecta</i> clade	5	930	0.91	0.44	0.039	0.036	$p_1 = 0.065$ $\omega_1 = 3.53$	33 (3)	17 (1)	41 (2)
<i>D. simulans</i> – <i>D. melanogaster</i> clade	4	996	0.30	0.69	0.001	0.001	$p_1 = 0.051$ $\omega_1 = 19.53$	15 (2)	4 (0)	11 (2)

n, number of taxa analyzed; *L*, sequence length in base pairs; *L_s*, tree length in substitutions per codon under a PAML model with one ω for all sites (M0); *d_N/d_S*, ratio average across all sites and lineages under PAML model M0; *P*, probability value of the likelihood-ratio tests between the listed models of PAML. Selection parameters (M8), parameter estimates under model M8 of PAML: p_1 is proportion of sites in the class with “ $\omega > 1$ ” and ω_1 is the estimate of ω for that class. PSS (M8), number of positively selected sites estimated using the BEB method (PAML) under M8 with posterior probabilities $P > 0.50$ ($P > 0.95$ in parentheses); PSS (FEL), number of positively selected sites estimated using the FEL analysis (HyPhy) with $P < 0.20$ ($P < 0.05$ in parentheses) [$P < 0.20$ is consistent with type I error rates of <5% (KOSAKOVSKY POND and FROST 2005)]; PSS (REL), number of positively selected sites estimated using the REL analysis (HyPhy) with posterior probabilities $P > 0.50$ ($P > 0.95$ in parentheses).

data (Table 1). The BEB method identifies 33 and 15 codon sites positively selected in the *D. yakuba*–*D. erecta* and the *D. simulans*–*D. melanogaster* clades, respectively. The methods implemented in HyPhy (FEL and REL) also detect codon sites with $\omega > 1$ in both clades.

Remarkably, the distribution of codon sites under positive selection across the gene differs between the two clades of the *melanogaster* subgroup (Figure 2). In the *D. yakuba*–*D. erecta* clade PAML identifies sites possibly under positive selection to be evenly distributed

along the protein sequence, with sites with the highest posterior probabilities located in the N-terminal two-thirds of the protein (amino acids 1–217 in AVEDISOV *et al.* 2001). In contrast, in the *D. simulans*–*D. melanogaster* clade the signature of positive selection is detected only in the C-terminal domain (amino acids 188–335 in AVEDISOV *et al.* 2001) of the protein. An equivalent distribution of sites with ω significantly greater than one is obtained by FEL and REL methods, with positive selection in the *D. simulans*–*D. melanogaster* clade detected only in the C-terminal third of the protein. Thus, results of both PAML and HyPhy show evidence for adaptive evolution in the *melanogaster* subgroup as well as in the two main clades within the subgroup with spatial differences across the Rux protein. (supplemental Table S1 lists all codon sites identified as positively selected using PAML and HyPhy methods in the different taxa.)

Rux polymorphism: We surveyed patterns of nucleotide polymorphism at the *rux* gene in six species, with a total of 99 chromosomes from eight populations. Table 2 lists summary statistics of sequence polymorphism. In agreement with previous reports, *D. sechellia* shows extremely reduced levels of polymorphism, a pattern that is suggestive of a severely reduced effective population size (KLIMAN *et al.* 2000). Another overall feature that stands out from our survey is the high number of nonsynonymous polymorphisms observed in the *rux* gene of *D. mauritiana* (18 replacement and 11 synonymous polymorphisms in a sample of 6 chromosomes). Furthermore, all these polymorphisms are located in the C-terminal half of the protein and most (10) are clustered in an ~80-bp-long region of exon 2. This region corresponds to a six-amino-acid repeat unit that has expanded in *D. simulans*, *D. mauritiana*, and *D. sechellia* but is not present in either *D. melanogaster* or

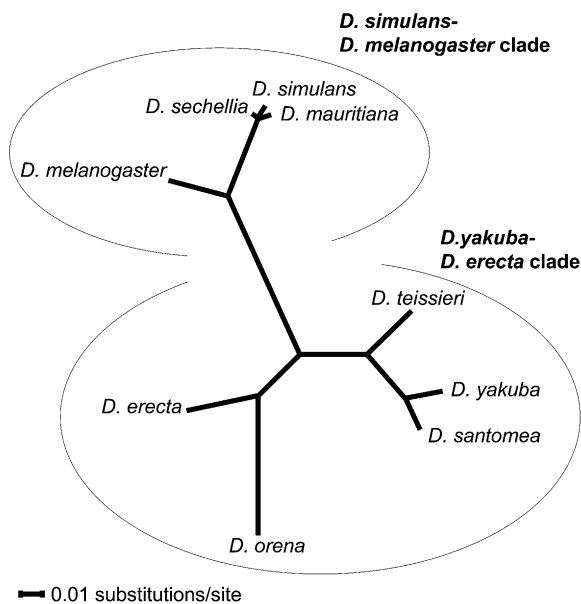


FIGURE 1.—Phylogenetic relationship among the nine species of the *D. melanogaster* subgroup inferred from *rux* gene genealogy. The proposed phylogeny is in complete agreement with that obtained from the 12 Drosophila Genomes Project (CLARK *et al.* 2007).

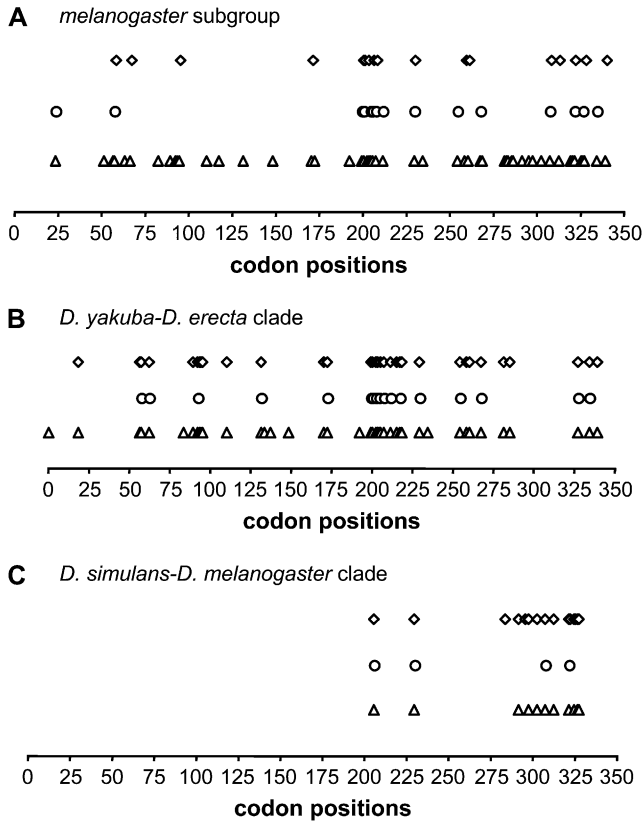


FIGURE 2.—Distribution of positively selected sites along exon 2 of the *rux* gene identified by PAML (BEB) (diamonds, posterior probabilities >0.50), FEL (circles, posterior probabilities >0.50), and REL (triangles, $P < 0.20$) in species of (A) the *melanogaster* subgroup, (B) the *D. yakuba*–*D. erecta* clade, and (C) the *D. simulans*–*D. melanogaster* clade.

the *D. yakuba*–*D. erecta* clade. To investigate the possibility that the abundance of replacement polymorphisms is due to a change in the exon–intron structure of the *rux* gene in *D. mauritiana* we obtained the sequence of the cDNA. Sequence analyses indicate that the exon–intron boundaries are conserved between *D. mauritiana*

and *D. melanogaster* and that the hypervariable 80-bp-long region is present in the *D. mauritiana* cDNA (supplemental Figure S2). Thus the region where most amino acid replacement polymorphisms cluster does not represent a new intron evolved in *D. mauritiana*, but a portion of the coding sequence with severely reduced selective constraints in this species.

Inference about selection using patterns of divergence and polymorphism: To determine whether patterns of variation in the *rux* gene are consistent with a neutral model we applied the MK test that compares patterns of divergence and polymorphism (MCDONALD and KREITMAN 1991). This test contrasts the ratio of polymorphic to fixed differences at amino acid replacement and synonymous or silent sites and is based on the prediction of the neutral theory that divergence between species correlates with variation within species (KIMURA 1983). We used two different approaches to assess the statistical significance of the difference in the ratios: the 2×2 goodness-of-fit test proposed originally (MCDONALD and KREITMAN 1991) and a maximum-likelihood method based on the Poisson random field framework (MKPRF) (SAWYER and HARTL 1992; BUSTAMANTE *et al.* 2003). This latter approach also allows us to estimate directly the direction and strength of selection (γ ; $\gamma = 2N_e s$) acting on amino acid replacement mutations in the *rux* gene. The results of the MK and MKPRF tests as well as estimates of γ on nonlethal amino acid replacements are shown in Table 3.

The comparison between *D. yakuba* and *D. santomea* indicates that amino acid replacements have contributed disproportionately to divergence between these sibling species ($G = 12.49$, $P = 0.0004$), consistent with adaptive evolution in the Rux protein. MK-based tests between very closely related species, however, should be interpreted carefully because some of the observed fixed differences may be due to ancestral polymorphism (CHARLESWORTH *et al.* 2005). To minimize these confounding effects we performed MK tests using the more

TABLE 2
Polymorphism in the *rux* gene

Species	Population	<i>n</i>	ST	SS	SR	θ_s	π_s	<i>L</i> (bp)
<i>D. yakuba</i>	Ivory Coast	15	27	19	8	0.021	0.016	973
<i>D. yakuba</i>	São Tomé	16	28	21	7	0.024	0.015	940
<i>D. santomea</i>	São Tomé	16	21	18	3	0.020	0.013	940
<i>D. melanogaster</i>	Raleigh, NC	11	13	12	1	0.004	0.005	1686
<i>D. simulans</i>	Congo	15	18	12	6	0.004	0.005	1724
<i>D. simulans</i>	São Tomé	14	20	14	6	0.005	0.004	1724
<i>D. mauritiana</i>	Mauritius	6	29	11	18	0.005	0.005	1740
<i>D. sechellia</i>	Seychelles	6	1	0	1	0.000	0.000	1697

n, number of chromosomes sampled; ST, total number of segregating mutations in the sample; SS, number of silent (synonymous and noncoding) polymorphisms; SR, number of replacement polymorphisms; θ_s , estimate of silent heterozygosity per site (WATTERSON 1975); π_s , average number of silent differences per site; *L*, number of sites analyzed excluding gaps.

TABLE 3
Adaptive evolution in the Rux protein

Species	FS	SS	FR	SR	n_1	n_2	N.I.	$P(G)$	γ	$P(\gamma \leq 0)$
<i>D. yakuba</i> – <i>D. santomea</i>	6	45	14	15	31	16	0.14	0.0004	NA	NA
<i>D. teissieri</i> – <i>D. yakuba</i>	31	30	33	12	1	31	0.38	0.019	2.56	0.0098
<i>D. teissieri</i> – <i>D. santomea</i>	33	18	31	3	1	16	0.18	0.004	6.39	0.0008
<i>D. simulans</i> – <i>D. melanogaster</i>	86	33	45	7	29	11	0.41	0.035	2.24	0.011
<i>D. simulans</i> – <i>D. mauritiana</i>	15	30	5	19	29	6	1.90	0.28	3.61	0.74
<i>D. simulans</i> – <i>D. sechellia</i>	9	21	3	6	29	6	0.86	0.86	38.74	0.16

FS, number of silent differences fixed between species; SS, number of silent polymorphisms; FR, number of nonsynonymous differences fixed between species; SR, number of nonsynonymous polymorphisms; n_1 and n_2 , number of chromosomes sampled in each species; N.I., neutrality index. $P(G)$, P -values for the 2×2 G -tests (MK test). γ , posterior mean selection intensity associated with nonlethal amino acid replacements estimated by the MKPRF framework, with $\gamma = 2N_e s$, where N_e indicates the diploid effective population size and s the selection coefficient. $P(\gamma \leq 0)$, P -values based on the maximum-likelihood method MKPRF after 10,000 iterations; NA, values could not be estimated because of lack of informative sites.

distant *D. teissieri* and also detected the signature of adaptive evolution in both *D. teissieri*–*D. yakuba* ($G = 5.52$, $P = 0.019$) and *D. teissieri*–*D. santomea* ($G = 8.30$, $P = 0.004$) comparisons. Congruently, estimates of γ are positive (+2.56 and +6.39) and significantly >0 ($P = 0.0098$ and $P = 0.0008$) for the *D. teissieri*–*D. yakuba* and *D. teissieri*–*D. santomea* comparisons, respectively. In the *D. simulans*–*D. melanogaster* clade there is also a minor but significant excess of amino acid replacements fixed between these species and the estimate of γ is also >0 ($\gamma = +2.24$, $P = 0.011$). When we remove singletons to reduce the confounding effects of weakly deleterious mutations that will not contribute to long-term evolutionary trends, the excess of amino acid replacements between species remains significant in both comparisons ($G = 8.01$, $P = 0.0045$ for *D. yakuba*–*D. santomea* and $G = 6.51$, $P = 0.012$ for *D. simulans*–*D. melanogaster*). We find no evidence for recent selection at the Rux protein in the *D. simulans* lineage after the split from its closer relatives *D. mauritiana* and *D. sechellia*.

Finally, we performed lineage-specific MK tests to determine whether the excess of adaptive nonsynonymous substitutions is associated with a single lineage or results from independent events in different lineages. Using parsimonious criteria we assigned fixed differences to either the *D. yakuba* or the *D. santomea* lineages, using the *D. teissieri* sequence as the outgroup, and *D. erecta* was used to polarize fixed differences in the lineage leading to the ancestor of *D. yakuba* and *D. santomea* after the split from *D. teissieri*. (The *D. orena* sequence was not used as the outgroup because it shows extreme GC bias at some loci; see TAKANO-SHIMIZU 1999.) In the comparison between *D. melanogaster* and *D. simulans*, the *D. yakuba* sequence was used as the outgroup. Lineage-specific MK tests indicate a rapid accumulation of amino acid replacements in three lineages of the *melanogaster* subgroup (*D. yakuba*, *D. santomea*, and *D. melanogaster* lineages), as well as in the lineage leading to the *D. yakuba* and *D. santomea* pair after the

split from *D. teissieri* (Figure 3). While the high number of amino acid replacements in the comparison *D. yakuba*–*D. santomea* is due to an excess of fixations in each lineage after the split, in the *D. melanogaster*–*D. simulans* comparison only the *D. melanogaster* lineage shows evidence of adaptive evolution (Figure 3). Following SMITH and EYRE-WALKER (2002) we estimated the number of adaptive amino acid substitutions in the *D. yakuba*, *D. santomea*, and *D. melanogaster* lineages to be 4.84, 5.50, and 12.92, respectively.

Inferences about selection based on the frequency of polymorphic mutations: To study whether patterns of intraspecific variation in the *rux* gene are consistent with a neutral model or, conversely, are suggestive of recent adaptive events we applied the DH test (ZENG *et al.* 2006). This test is based on the combined detection of both low- and high-frequency polymorphisms summarized by Tajima's D (TAJIMA 1989b) and Fay and Wu's H (FAY and WU 2000) statistics, respectively. An excess of low-

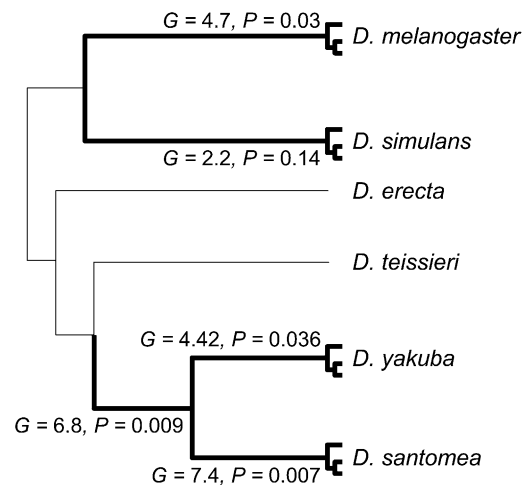


FIGURE 3.—Results of lineage-specific MK tests in the *rux* locus [goodness-of-fit (G) values and probabilities (P)]. Branch lengths do not represent real genetic distances.

TABLE 4
Neutrality tests based on polymorphism data of the *rux* gene

Species	Population	ST	<i>D</i>	<i>H</i>	$P(DH)_{r=0}$	$P(DH)_{r=\rho}$	ρ
<i>D. yakuba</i>	Ivory Coast	27	-1.35	-3.41	0.045	<0.00001	194.8
<i>D. yakuba</i>	São Tomé	28	-1.62	-4.87	0.028	<0.00001	189.8
<i>D. santomea</i>	São Tomé	21	-1.42	-9.13	0.012	<0.00001	420.3
<i>D. melanogaster</i>	Raleigh, NC	13	0.11	1.29	0.34	0.40	9.9
<i>D. simulans</i>	Congo	18	1.47	0.01	0.31	0.37	4.4
<i>D. simulans</i>	São Tomé	20	-1.31	-8.22	0.017	0.008	3.4
<i>D. mauritiana</i>	Mauritius	29	0.91	1.07	0.36	0.57	24.8
<i>D. sechellia</i>	Seychelles	1	-0.93	NA	NA	NA	NA

ST, total number of segregating mutations in the sample; *D*, Tajima's *D* statistic; *H*, Fay and Wu's *H* statistic. $P(DH)_{r=0}$ and $P(DH)_{r=\rho}$, *P*-values for the *DH* tests of neutrality (see text) based on coalescent simulations without recombination ($r = 0$) and with estimated levels of intragenic recombination ($r = \rho$), respectively; ρ , estimates of intragenic recombination; NA, values could not be estimated because of lack of informative sites.

frequency polymorphisms is expected after the removal of variation associated with a selective sweep (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989) or the elimination of deleterious mutations in small populations (CHARLESWORTH *et al.* 1993). When recombination occurs, however, a selective sweep is also expected to generate a combination of both high- and low-frequency mutations segregating in the population (FAY and WU 2000), the simultaneous detection of which (*i.e.*, the *DH* test) is an effective way to identify the signature of positive selection using within-species variation (ZENG *et al.* 2007). Table 4 summarizes the results of *DH* tests for the *rux* gene under the conservative assumption of no recombination and under likely values of intragenic recombination estimated from the data (TAJIMA 1989b; WALL 1999; HUDSON 2001; ZENG *et al.* 2006, 2007). In all, three populations of the *D. yakuba*-*D. erecta* clade (Ivory Coast and São Tomé populations of *D. yakuba* and the São Tomé population of *D. santomea*) as well as the *D. simulans* population from São Tomé show polymorphisms in the *rux* gene that are skewed significantly toward low and high frequency of derived mutations.

Although our results from the *DH* tests are consistent with a recent selective sweep in the *rux* region, demographic events, such as a recent population bottleneck or an expansion, can also skew the frequency spectrum, mimicking the effects of positive selection (TAJIMA 1989a). This possibility is particularly likely for *D. santomea* and the São Tomé population of *D. yakuba*, where there is evidence for nonequilibrium based on multilocus analyses (LLOPART *et al.* 2005; BACHTROG *et al.* 2006). Despite the fact that many of these nonselective scenarios are expected to have a smaller influence on *H* and *DH* than on other population statistics (PRZEWSKI 2002; ZENG *et al.* 2006, 2007), we investigated whether demography alone could explain *DH* estimates as well as the number of polymorphisms observed at the *rux* locus in island populations. Using a number of combinations of changes in population sizes, we conducted coalescent

simulations where a population undergoes an instantaneous reduction in size (*i.e.*, invasion of an island) followed by a period of time for local adaptation and final expansion (see MATERIALS AND METHODS and supplemental Table S2 for a detailed description of the specific parameters used in the simulated demography). These simulations, however, are not intended to be exhaustive: we have inevitably missed more complex population dynamics. They serve nonetheless to assess how extreme cases of demography can possibly influence neutrality tests based on the frequency spectrum of polymorphisms in island populations. The results indicate that none of the simple demographic scenarios considered entirely explains the observed number of polymorphisms and the frequency spectrum in the *rux* locus (Table 5 and supplemental Table S2). While population dynamics that incorporate a very recent and brief population size reduction for *D. yakuba* and *D. simulans* are the most compatible with the number of polymorphisms observed today, they fail to produce significant *DH* tests. Likewise results in the *D. santomea* population show that none of the modeled demographic histories produces *DH*s that are equal or more extreme than those observed in the *rux* region.

Despite the fact that none of the simulated demographic scenarios can fully explain the polymorphism patterns in the *rux* locus, we cannot formally rule out more complex neutral population dynamics (*e.g.*, bottleneck followed by expansion combined with species introgression) having an effect on the frequency spectrum of polymorphisms in island populations. This appears to be the case in *D. santomea* for which the study of 29 loci showed 7 loci with more negative Tajima's *D* than *rux* (LLOPART *et al.* 2005). In contrast, none of the 29 loci analyzed in the island population of *D. yakuba* showed a more extreme *D*-value than that of *rux*, which strongly supports the notion of positive selection. As for the *D. simulans* population from São Tomé, there is no multilocus analysis yet to test mutation-drift equilibrium;

TABLE 5

Neutrality tests based on polymorphism data of the *rux* gene under different simulated demographic scenarios in island populations

Species	T (years)	d (years)	$P(S)_4$	$P(DH)_4$	$P(DH S)_4$	$P(S)_5$	$P(DH)_5$	$P(DH S)_5$
<i>D. yakuba</i>	500	4	0.06	0.05	0.01	0.11	0.04	0.01
<i>D. simulans</i>	500	4	0.03	0.02	0.00	0.07	0.02	0.006
<i>D. santomea</i>	450000	400	<0.0001	NA	<0.0001	0.002	NA	<0.0001

Results of DH tests obtained under different demographic scenarios with the conservative assumption of no recombination. We simulated a panmictic population with an ancestral size of N_0 that experiences a severe instantaneous size reduction to $N_1 = N_0 \times 0.001$, followed by a period of constant size, followed by an exponential growth to a final size of N_2 . We studied two alternative scenarios, $N_2 = 10 \times N_1$ and $N_2 = 100 \times N_1$. The mutation parameter for the *rux* locus θ_0 was estimated from polymorphism data of the *D. yakuba* ($\theta_0 = 8.75$) and *D. simulans* ($\theta_0 = 5.17$) mainland populations. In the case of *D. santomea* the mutation parameter was estimated on the basis of the predictions of the isolation model (WAKELEY and HEY 1997) for the ancestor of *D. yakuba* and *D. santomea* ($\theta_0 = 25.38$; Table 5 in LLOPART *et al.* 2005). T , time ago at which the severe reduction in size occurred; d , time during which the population size remained constant and reduced. $P(S)_4$ and $P(S)_5$ indicate the fraction of the simulated replicates with the number of segregating sites equal to or greater than that observed in the island samples assuming that N_2 is either 10,000 or 100,000, respectively. $P(DH)_4$ and $P(DH)_5$ indicate P -values for the DH tests, and $P(DH | S)_4$ and $P(DH | S)_5$ indicate P -values for the DH tests conditioning on the observed number of segregating sites assuming that N_2 is either 10,000 or 100,000, respectively. NA, demographic scenarios producing three or fewer polymorphisms on average.

meanwhile we collected polymorphism data for the putatively neutral locus *Xdh*. The analysis of the frequency spectrum at *Xdh* shows no significant departure from neutrality ($\theta_S = 0.064$; $D = 0.27$, $H = -12.35$, $P = 0.10$ without recombination and $P = 0.073$ with recombination), which again would be consistent with positive selection on *rux*. Finally, we considered possible demographic effects for the *D. yakuba* population collected in Ivory Coast, the only mainland population for which there is a significant excess of low- and high-frequency derived polymorphisms. A study of 21 genes in this population of *D. yakuba* indicates a minor, if any, effect of demography on the frequency spectrum of *rux* (Ivory Coast population of *D. yakuba* mean $D = -0.19$; A. WILLIFORD, A. LLOPART and J. M. COMERON, unpublished data). This again supports the idea of recent positive selection acting on the *rux* gene in *D. yakuba*.

DISCUSSION

Here we combine several phylogeny-based methods with molecular population genetics analyses to show that variation between and within species at the *rux* gene in *Drosophila* is best explained by recurrent events of positive selection in independent lineages. In particular we detect a significant excess of amino acid replacements fixed between species that is observed when comparing to synonymous divergence (d_N/d_S approach using both PAML and HyPhy) as well as when contrasting levels of polymorphism and divergence for both replacement and silent mutations (MK and MKPRF approaches). This excess, we suggest, is the result of adaptive fixations of selectively advantageous amino acid replacements in the Rux protein and has occurred concurrently in different lineages.

Evidence for adaptive protein evolution in different lineages has been previously reported in genes responsible for reproductive isolation among *Drosophila* species, and in these cases positive selection is restricted to ancient events (BARBASH *et al.* 2004; PRESGRAVES and STEPHAN 2007). *rux* evolution is distinctive because our analyses of extant variation at the *rux* gene reveal also an excess of both high- and low-frequency polymorphisms (DH test) in four of eight populations surveyed, a signature of recent events of positive selection in recombining regions that cannot be explained by simple demographic scenarios alone. Although these departures in the frequency spectrum of polymorphisms are consistent with adaptive amino acid fixations, divergent gene expression cannot be ruled out as an additional target of selection (CHEN *et al.* 2007). Indeed male-biased genes are prone to show patterns of expression that are divergent between species (MEIKLEJOHN *et al.* 2003; MICHALAK and NOOR 2003; RANZ *et al.* 2003; NUZHIDIN *et al.* 2004). In agreement with the possibility of selection at gene expression divergence, *rux* mRNA abundance differs significantly between males of *D. melanogaster* and *D. simulans* (PARISI *et al.* 2003; RANZ *et al.* 2003) and also between males of the very closely related *D. yakuba* and *D. santomea* (A. LLOPART, R. RATNAPPAN and J. M. COMERON, unpublished data).

A possible caveat to the use of the standard method to detect positive selection based on the d_N/d_S ratio (based on PAML software) is that a portion of synonymous mutations in *Drosophila* are not neutral (AKASHI 1995; AKASHI and SCHAEFFER 1997; KLIMAN 1999; McVEAN and VIEIRA 2001; COMERON and GUTHRIE 2005; LLOPART *et al.* 2008). Therefore one could argue that in genes with strong selection on synonymous mutations (high codon bias) events of positive selection acting on amino acid

replacements might be overestimated due to reduced rate of evolution at synonymous sites (reduced d_S and hence high d_N/d_S). Several lines of evidence suggest that our assessment of positive selection for *rux* based on maximum-likelihood estimates of d_N/d_S is not likely to be caused by selection on synonymous mutations. First, *rux* does not show a strong bias in synonymous codons, with 57% of *D. melanogaster* genes showing greater bias based on the frequency of preferred codons. Second, both PAML and HyPhy reveal that different clades exhibit different regions of the *rux* gene with codon sites under positive selection. Finally, it is worth noting that divergence estimates by PAML in *Drosophila* genes do not generate a negative relationship between d_S and the degree of codon bias (DUNN *et al.* 2001). This latter feature is likely a consequence of the methodology used by PAML to estimate the number of synonymous sites (BIERNE and EYRE-WALKER 2003) and, for our purposes, would also suggest that estimates of $\omega > 1$ in the *rux* gene are not likely to be caused by reduced d_S associated with increased selection on synonymous sites.

MK-based methods to assess adaptive evolution on amino acid changes also assume neutrality of synonymous sites. Hence one could also contend that weak selection on synonymous sites would decrease the ratio of divergence to polymorphism for synonymous mutations and create the appearance of increased divergence for amino acid replacements under the MK framework. False positives, however, would appear only if replacement mutations were more neutral than synonymous mutations, an unlikely event based on the observation that both divergence and polymorphism levels are reduced at replacement sites (see also ANDOLFATTO 2005; SHAPIRO *et al.* 2007). More important, segregating mutations are also under stronger selective constraints when they are replacements than when they are synonymous based on an excess of low-frequency variants in the former class of mutations (SHAPIRO *et al.* 2007). Thus granted that weak selection on synonymous mutations might bias precise estimates of selection (γ) based on the MKPRF framework, it is not likely that it could create the appearance of an excess of amino acid replacements between species in the absence of positive selection at the protein level. Another possible bias in the estimates of γ may arise from equilibrium departures at synonymous sites, perhaps associated with changes in effective population size (AKASHI 1996; McVEAN and VIEIRA 2001; EYRE-WALKER 2002; COMERON and GUTHRIE 2005; KERN and BEGUN 2005; AKASHI *et al.* 2006). Although we cannot formally rule out this possible bias, we argue that, in the absence of adaptive evolution, it is unlikely that the bias would result in the detection of different positively selected sites in different clades.

Indeed, our results indicate also that amino acid replacements possibly subjected to positive selection are distributed in a different manner along the Rux protein

in the two clades of the *melanogaster* subgroup. This is consistent with different portions of the protein having different selective histories in independent lineages. We observe that the N-terminal domain (amino acids 1–217 in AVEDISOV *et al.* 2001) evolves under positive selection at amino acid divergence solely in the *D. yakuba*–*D. erecta* clade. Our BLASTP and TBLASTN searches indicate that this domain is present in every genome of the 12 *Drosophila* species that have been sequenced (supplemental Figure S1). This domain is responsible for the physical interaction with the product of the gene *CycA* (AVEDISOV *et al.* 2000), and future studies will contrast whether its fast evolution is also accompanied by rapid coevolution of *CycA*, as seen in the interactors of the *Drosophila* hybrid inviability gene *Nup96* (PRE-SGRAVES and STEPHAN 2007). For now, preliminary analyses based on d_N/d_S estimates in the *melanogaster* subgroup show no evidence of rapid protein evolution for *CycA*, with an overall $\omega < 0.1$ and no increased probability for models that allow for positively selected sites (M1 *vs.* M2a and M7 *vs.* M8; $P > 0.20$ in both cases).

On the other hand, the C-terminal domain (amino acids 188–335 in AVEDISOV *et al.* 2001) shows evidence of positively selected sites in both clades of the *melanogaster* subgroup. Whole-genome cDNA microarray analyses show that some regions of the 3' end of the annotated *rux* gene are indeed transcribed in species outside the *melanogaster* subgroup (ZHANG *et al.* 2007; supplemental Figure S1). Nevertheless our BLASTP and TBLASTN searches indicate that this domain has no detectable amino acid similarity (supplemental Figure S1). This is consistent with the C-terminal domain being the result of extremely fast divergence rather than “*de novo*” formation. Functional analyses in *D. melanogaster* indicate, however, that the C-terminal domain is necessary for proper intracellular localization to the nucleus (AVEDISOV *et al.* 2000). Thus we suggest that this highly diverged domain, which comprises a third of the Rux protein (148 of 335 amino acids in *D. melanogaster*), may have promoted the evolution of a distinct function exclusive to the *melanogaster* subgroup. It is tempting to propose that *rux*, as a male-biased gene with mutants showing male, but not female, sterility, evolves under positive selection because of its role in spermatogenesis and that this is somehow achieved by postcopulatory sexual selection and/or sexual conflict (BIRKHEAD and PIZZARI 2002; SWANSON and VACQUIER 2002; ELLEGREN and PARSCH 2007). It is also possible, however, that adaptive evolution acting at the level of amino acid divergence may just be the result of its distinct protein domain. In all, adaptive evolution in the “ancestral” (*i.e.*, N-terminal) domain of the Rux protein appears to be quite idiosyncratic (LEVINE and BEGUN 2007), while the C-terminal domain shows evidence of positive selection in independent *Drosophila* lineages within the *melanogaster* subgroup. This raises the interesting possibility

that male-biased genes of either recent origin or ancient descent might be under different selective regimes.

We thank the editor and two anonymous reviewers for valuable comments and suggestions. We are also grateful to R. R. Hudson for making available ms and MaxHap programs. Our thanks also go to S. L. Kosakovsky Pond, S. Frost, and A. Poon for the Datamonkey webserver and to the Computational Biology Service Unit (Cornell University) for the MKPRF web server. Part of this work was carried out using computational resources of the Carver Center for Comparative Genomics (University of Iowa). This work was supported by the Roy J. Carver Charitable Trust grant 05-2258 and the National Science Foundation grant DEB-03-44209 to J.M.C. and by a Carver Scientific Research Initiative Grants Funding Initiative Award to A.L.

LITERATURE CITED

- AGUADÉ, M., 1999 Positive selection drives the evolution of the *Acp29AB* accessory gland protein in *Drosophila*. *Genetics* **152**: 543–551.
- AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1992 Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila*. *Genetics* **132**: 755–770.
- AKASHI, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.
- AKASHI, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**: 1297–1307.
- AKASHI, H., and S. W. SCHAEFFER, 1997 Natural selection and the frequency distributions of “silent” DNA polymorphism in *Drosophila*. *Genetics* **146**: 295–307.
- AKASHI, H., W. Y. KO, S. PIAO, A. JOHN, P. GOEL *et al.*, 2006 Molecular evolution in the *Drosophila melanogaster* species subgroup: frequent parameter fluctuations on the timescale of molecular divergence. *Genetics* **172**: 1711–1726.
- ANDOLFATTO, P., 2005 Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- ARBEITMAN, M. N., E. E. FURLONG, F. IMAM, E. JOHNSON, B. H. NULL *et al.*, 2002 Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**: 2270–2275.
- AVEDISOV, S. N., I. KRASNOSELSKAYA, M. MORTIN and B. J. THOMAS, 2000 *Roughex* mediates G(1) arrest through a physical association with cyclin A. *Mol. Cell. Biol.* **20**: 8220–8229.
- AVEDISOV, S. N., I. B. ROGOZIN, E. V. KOONIN and B. J. THOMAS, 2001 Rapid evolution of a cyclin A inhibitor gene, *roughex*, in *Drosophila*. *Mol. Biol. Evol.* **18**: 2110–2118.
- BACHTROG, D., K. THORNTON, A. CLARK and P. ANDOLFATTO, 2006 Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evol. Int. J. Org. Evol.* **60**: 292–302.
- BARBASH, D. A., P. AWADALLA and A. M. TARONE, 2004 Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biol.* **2**: e142.
- BARRIER, M., C. D. BUSTAMANTE, J. YU and M. D. PURUGGANAN, 2003 Selection on rapidly evolving proteins in the Arabidopsis genome. *Genetics* **163**: 723–733.
- BAUER DUMONT, V. L., H. A. FLORES, M. H. WRIGHT and C. F. AQUADRO, 2007 Recurrent positive selection at *bgn*, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein *bam*. *Mol. Biol. Evol.* **24**: 182–191.
- BEGUN, D. J., and H. A. LINDFORS, 2005 Rapid evolution of genomic *Acp* complement in the *melanogaster* subgroup of *Drosophila*. *Mol. Biol. Evol.* **22**: 2010–2021.
- BEGUN, D. J., P. WHITLEY, B. L. TODD, H. M. WALDRIP-DAIL and A. G. CLARK, 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879–1888.
- BEGUN, D. J., H. A. LINDFORS, M. E. THOMPSON and A. K. HOLLOWAY, 2006 Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags. *Genetics* **172**: 1675–1681.
- BEGUN, D. J., H. A. LINDFORS, A. D. KERN and C. D. JONES, 2007 Evidence for *de novo* evolution of testis-expressed genes in the *Drosophila yakuba/Drosophila erecta* clade. *Genetics* **176**: 1131–1137.
- BETRAN, E., and M. LONG, 2003 *Dntf-2r*: a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics* **164**: 977–988.
- BIERNE, N., and A. EYRE-WALKER, 2003 The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics* **165**: 1587–1597.
- BIRKHEAD, T. R., and T. PIZZARI, 2002 Postcopulatory sexual selection. *Nat. Rev. Genet.* **3**: 262–273.
- BUSTAMANTE, C. D., R. NIELSEN, S. A. SAWYER, K. M. OLSEN, M. D. PURUGGANAN *et al.*, 2002 The cost of inbreeding in Arabidopsis. *Nature* **416**: 531–534.
- BUSTAMANTE, C. D., R. NIELSEN and D. L. HARTL, 2003 Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among classes of mutations using DNA polymorphism data. *Theor. Popul. Biol.* **63**: 91–103.
- CASTILLO-DAVIS, C. I., F. A. KONDRASHOV, D. L. HARTL and R. J. KULATHINAL, 2004 The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. *Genome Res.* **14**: 802–811.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CHARLESWORTH, B., C. BARTOLOME and V. NOEL, 2005 The detection of shared and ancestral polymorphisms. *Genet. Res.* **86**: 149–157.
- CHEN, S. T., H. C. CHENG, D. A. BARBASH and H. P. YANG, 2007 Evolution of *hydra*, a recently evolved testis-expressed gene with nine alternative first exons in *Drosophila melanogaster*. *PLoS Genet.* **3**: e107.
- CIVETTA, A., and R. S. SINGH, 1995 High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J. Mol. Evol.* **41**: 1085–1095.
- CIVETTA, A., and R. S. SINGH, 1998 Sex-related genes, directional sexual selection, and speciation. *Mol. Biol. Evol.* **15**: 901–909.
- CIVETTA, A., S. A. RAJAKUMAR, B. BROUWERS and J. P. BACIK, 2006 Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. *Mol. Biol. Evol.* **23**: 655–662.
- CLARK, A. G., M. B. EISEN, D. R. SMITH, C. M. BERGMAN, B. OLIVER *et al.*, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.
- COMERON, J. M., and T. B. GUTHRIE, 2005 Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in *Drosophila*. *Mol. Biol. Evol.* **22**: 2519–2530.
- DUNN, K. A., J. P. BIELAWSKI and Z. YANG, 2001 Substitution rates in *Drosophila* nuclear genes: implications for translational selection. *Genetics* **157**: 295–305.
- ELLEGREN, H., and J. PARSCH, 2007 The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* **8**: 689–698.
- EYRE-WALKER, A., 2002 Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**: 2017–2024.
- FAY, J. C., and C. I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FRY, J. D., S. V. NUZHIDIN, E. G. PASYUKOVA and T. F. MACKAY, 1998 QTL mapping of genotype-environment interaction for fitness in *Drosophila melanogaster*. *Genet. Res.* **71**: 133–141.
- GIBSON, G., R. RILEY-BERGER, L. HARSHMAN, A. KOPP, S. VACHA *et al.*, 2004 Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics* **167**: 1791–1799.
- GONCZY, P., B. J. THOMAS and S. DiNARDO, 1994 *roughex* is a dose-dependent regulator of the second meiotic division during *Drosophila* spermatogenesis. *Cell* **77**: 1015–1025.
- GOOD, J. M., and M. W. NACHMAN, 2005 Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol. Biol. Evol.* **22**: 1044–1052.
- HAERTY, W., S. JAGADEESHAN, R. J. KULATHINAL, A. WONG, K. RAVI RAM *et al.*, 2007 Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**: 1321–1335.

- HUDSON, R. R., 2001 Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- JIN, W., R. M. RILEY, R. D. WOLFINGER, K. P. WHITE, G. PASSADOR-GURGEL *et al.*, 2001 The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* **29**: 389–395.
- JONES, C. D., and D. J. BEGUN, 2005 Parallel evolution of chimeric fusion genes. *Proc. Natl. Acad. Sci. USA* **102**: 11373–11378.
- KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- KELLEHER, E. S., W. J. SWANSON and T. A. MARKOW, 2007 Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. *PLoS Genet.* **3**: e148.
- KERN, A. D., and D. J. BEGUN, 2005 Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. *Mol. Biol. Evol.* **22**: 51–62.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- KLIMAN, R. M., 1999 Recent selection on synonymous codon usage in *Drosophila*. *J. Mol. Evol.* **49**: 343–351.
- KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**: 1913–1931.
- KOSAKOVSKY POND, S. L., and S. D. FROST, 2005 Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**: 1208–1222.
- LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol.* **22**: 159–225.
- LAWNICZAK, M. K., and D. J. BEGUN, 2007 Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. *Mol. Biol. Evol.* **24**: 1944–1951.
- LEVINE, M. T., and D. J. BEGUN, 2007 Comparative population genetics of the immunity gene, *Relish*: Is adaptive evolution idiosyncratic? *PLoS ONE* **2**: e442.
- LEVINE, M. T., C. D. JONES, A. D. KERN, H. A. LINDFORS and D. J. BEGUN, 2006 Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. USA* **103**: 9935–9939.
- LLOPART, A., S. ELWYN, D. LACHAISE and J. A. COYNE, 2002 Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evol. Int. J. Org. Evol.* **56**: 2262–2277.
- LLOPART, A., D. LACHAISE and J. A. COYNE, 2005 Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* **171**: 197–210.
- LLOPART, A., A. MABILLÉ, J. R. PETERS-HALL, J. M. COMERON and R. M. KLIMAN, 2008 A new test for selection applied to codon usage in *Drosophila simulans* and *D. mauritiana*. *J. Mol. Evol.* **66**: 224–231.
- LONG, M., and C. H. LANGLEY, 1993 Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**: 91–95.
- LONG, M., E. BETRAN, K. THORNTON and W. WANG, 2003 The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* **4**: 865–875.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MCVEAN, G. A., and J. VIEIRA, 2001 Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157**: 245–257.
- MEIKLEJOHN, C. D., J. PARSCH, J. M. RANZ and D. L. HARTL, 2003 Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **100**: 9894–9899.
- MICHALAK, P., and M. A. NOOR, 2003 Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol.* **20**: 1070–1076.
- NIELSEN, R., and Z. YANG, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 *envelope* gene. *Genetics* **148**: 929–936.
- NIELSEN, R., C. BUSTAMANTE, A. G. CLARK, S. GLANOWSKI, T. B. SACKTON *et al.*, 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**: e170.
- NUZHIDIN, S. V., M. L. WAYNE, K. L. HARMON and L. M. MCINTYRE, 2004 Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* **21**: 1308–1317.
- PARISI, M., R. NUTTALL, D. NAIMAN, G. BOUFFARD, J. MALLEY *et al.*, 2003 Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**: 697–700.
- POND, S. L., and S. D. FROST, 2005 Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.
- PRESGRAVES, D. C., and W. STEPHAN, 2007 Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, *Nup96*. *Mol. Biol. Evol.* **24**: 306–314.
- PROSCHEL, M., Z. ZHANG and J. PARSCH, 2006 Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* **174**: 893–900.
- PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.
- RANZ, J. M., C. I. CASTILLO-DAVIS, C. D. MEIKLEJOHN and D. L. HARTL, 2003 Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**: 1742–1745.
- REINKE, V., I. S. GIL, S. WARD and K. KAZMER, 2004 Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* **131**: 311–323.
- RIFKIN, S. A., J. KIM and K. P. WHITE, 2003 Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**: 138–144.
- RINN, J. L., J. S. ROZOWSKY, I. J. LAURENZI, P. H. PETERSEN, K. ZOU *et al.*, 2004 Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. *Dev. Cell* **6**: 791–800.
- ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- SAWYER, S. A., and D. L. HARTL, 1992 Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.
- SHAPIRO, J. A., W. HUANG, C. ZHANG, M. J. HUBISZ, J. LU *et al.*, 2007 Adaptive gene evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. USA* **104**: 2271–2276.
- SMITH, N. G., and A. EYRE-WALKER, 2002 Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.
- SWANSON, W. J., and V. D. VACQUIER, 2002 The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137–144.
- SWANSON, W. J., Z. YANG, M. F. WOLFNER and C. F. AQUADRO, 2001 Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* **98**: 2509–2514.
- SWANSON, W. J., R. NIELSEN and Q. YANG, 2003 Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18–20.
- TAJIMA, F., 1989a The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- TAJIMA, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAKANO-SHIMIZU, T., 1999 Local recombination and mutation effects on molecular evolution in *Drosophila*. *Genetics* **153**: 1285–1296.
- THOMAS, S., and R. S. SINGH, 1992 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VII. Varying rates of genic divergence as revealed by two-dimensional electrophoresis. *Mol. Biol. Evol.* **9**: 507–525.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- TORGERSON, D. G., R. J. KULATHINAL and R. S. SINGH, 2002 Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol. Biol. Evol.* **19**: 1973–1980.
- TURNER, L. M., and H. E. HOEKSTRA, 2006 Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol. Biol. Evol.* **23**: 1656–1669.
- VACQUIER, V., 1998 Evolution of gamete recognition proteins. *Science* **281**: 1996–1999.

- WAGSTAFF, B. J., and D. J. BEGUN, 2005a Comparative genomics of accessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. *Mol. Biol. Evol.* **22**: 818–832.
- WAGSTAFF, B. J., and D. J. BEGUN, 2005b Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* **171**: 1083–1101.
- WAKELEY, J., and J. HEY, 1997 Estimating ancestral population parameters. *Genetics* **145**: 847–855.
- WALL, J. D., 1999 Recombination and the power of statistical tests of neutrality. *Genet. Res. Camb.* **74**: 65–79.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WONG, W. S., Z. YANG, N. GOLDMAN and R. NIELSEN, 2004 Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**: 1041–1051.
- WYCKOFF, G. J., W. WANG and C. I. WU, 2000 Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**: 304–309.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YANG, Z., R. NIELSEN, N. GOLDMAN and A. M. PEDERSEN, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.
- YANG, Z., W. S. WONG and R. NIELSEN, 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**: 1107–1118.
- ZENG, K., Y. X. FU, S. SHI and C. I. WU, 2006 Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431–1439.
- ZENG, K., S. MANO, S. SHI and C. I. WU, 2007 Comparisons of site- and haplotype-frequency methods for detecting positive selection. *Mol. Biol. Evol.* **24**: 1562–1574.
- ZHANG, Y., D. STURGILL, M. PARISI, S. KUMAR and B. OLIVER, 2007 Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* **450**: 233–237.
- ZHANG, Z., and J. PARSCH, 2005 Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Mol. Biol. Evol.* **22**: 1945–1947.
- ZHANG, Z., T. M. HAMBUCH and J. PARSCH, 2004 Molecular evolution of sex-biased genes in *Drosophila*. *Mol. Biol. Evol.* **21**: 2130–2139.

Communicating editor: D. BEGUN