Molecular Biology and Evolution

Molecular Evolution of a Y Chromosome to Autosome Gene Duplication in *Drosophila*

Kelly A. Dyer,*[,1] Brooke E. White,[1] Michael J. Bray,[1] Daniel G. Piqué,[1] and Andrea J. Betancourt*†[,2]

[1]Department of Genetics, University of Georgia

[2]Institute of Evolutionary Biology, University of Edinburgh, Ashworth Labs, Edinburgh, United Kingdom

†Present address: Institute for Population Genetics, University of Veterinary Medicine Vienna, Vienna 1210, Austria

***Corresponding author:** kdyer@uga.edu, andrea.betancourt@vetmeduni.ac.at.

Associate editor: Jody Hey

Abstract

In contrast to the rest of the genome, the Y chromosome is restricted to males and lacks recombination. As a result, Y chromosomes are unable to respond efficiently to selection, and newly formed Y chromosomes degenerate until few genes remain. The rapid loss of genes from newly formed Y chromosomes has been well studied, but gene loss from highly degenerate Y chromosomes has only recently received attention. Here, we identify and characterize a Y to autosome duplication of the male fertility gene *kl-5* that occurred during the evolution of the testacea group species of *Drosophila*. The duplication was likely DNA based, as other Y-linked genes remain on the Y chromosome, the locations of introns are conserved, and expression analyses suggest that regulatory elements remain linked. Genetic mapping reveals that the autosomal copy of *kl-5* resides on the dot chromosome, a tiny autosome with strongly suppressed recombination. Molecular evolutionary analyses show that autosomal copies of *kl-5* have reduced polymorphism and little recombination. Importantly, the rate of protein evolution of *kl-5* has increased significantly in lineages where it is on the dot versus Y linked. Further analyses suggest this pattern is a consequence of relaxed purifying selection, rather than adaptive evolution. Thus, although the initial fixation of the *kl-5* duplication may have been advantageous, slightly deleterious mutations have accumulated in the dot-linked copies of *kl-5* faster than in the Y-linked copies. Because the dot chromosome contains seven times more genes than the Y and is exposed to selection in both males and females, these results suggest that the dot suffers the deleterious effects of genetic linkage to more selective targets compared with the Y chromosome. Thus, a highly degenerate Y chromosome may not be the worst environment in the genome, as is generally thought, but may in fact be protected from the accumulation of deleterious mutations relative to other nonrecombining regions that contain more genes.

Key words: sex chromosome, population genetics, molecular evolution, genetic linkage.

Research article

Introduction

Sex chromosomes have evolved many times in a diverse assemblage of organisms (Bull 1983), and the study of their evolution has provided insight into how population genetic processes affect large-scale patterns of genome evolution (Charlesworth et al. 2005; Ming and Moore 2007). Sex chromosomes are thought to usually evolve from a pair of homologous autosomes, frequently resulting in the cessation of recombination on all or part of one of the pair (reviewed in Charlesworth B and Charlesworth D 2000). The reduction in recombination has important evolutionary consequences. In order for any single variant to be fixed on a nonrecombining chromosome, all linked variants on the entire chromosome must be fixed as well, leading to the fixation of linked weakly deleterious variants and the loss of genetic variation. Furthermore, on nonrecombining chromosomes, genetic drift may be sufficiently strong to prevent purging of deleterious mutations, possibly resulting in their fixation (Muller 1964; Charlesworth 1978; Gordo and Charlesworth 2000). Finally, the accumulation of deleterious mutations may hamper the fixation of beneficial mutations, eventually resulting in selection favoring an end to the expression of the maladapted genes (Peck 1994; Orr and Kim 1998). The result of these processes is the eventual degeneration of the chromosome: genes become nonfunctional and repetitive DNA accumulates (Steinemann and Steinemann 1998; Bachtrog et al. 2008; Marais et al. 2008).

Snapshots of the early stages of sex chromosome evolution have been documented in many species and appear to fit the scenario outlined above very well. The study of highly degenerate sex chromosomes, however, poses several technical challenges, due to the high proportion of heterochromatin, and has only recently been tackled with a high degree of success (Carvalho et al. 2000, 2001; Koerich et al. 2008). (For simplicity, we will refer to the sex chromosomes as XX and XY, with the Y chromosome taken to be the degenerate chromosome.) Surprisingly, recent work shows that degenerate sex chromosomes can gain genes: genes that function in male fertility and reproduction are recruited to Y chromosomes via duplication and/or translocation events (Skaletsky et al. 2003; Koerich et al.

2008). Through time, the continual process of gene loss and gain can cause formerly homologous X and Y chromosomes to differ substantially in their gene content and gene order and related Y chromosomes to differ from one another (reviewed in Graves 2006, Koerich et al. 2008, and Hughes et al. 2010). Thus, the long-term evolution of Y chromosomes appears to be a dynamic process of gene gain and loss, and not just gene loss.

Here, we focus on the *Drosophila* Y chromosome, where the lack of recombination has resulted in a highly degenerate chromosome: in *D. melanogaster*, the heterochromatic Y comprises 40% of the haploid DNA content but carries only 13 known protein-coding genes (Carvalho et al. 2001; Koerich et al. 2008; Arguello et al. 2010). Further consequences of linkage can be seen in the remaining Y-linked genes themselves, which often have huge introns (>1 MB in length) loaded with repetitive DNA and low levels of polymorphism (Kurek et al. 2000; Reugels et al. 2000). Recent genomic analyses have underscored the dynamic nature of Y chromosome evolution in *Drosophila*. In *D. pseudoobscura*, the ancestral *Drosophila* Y has been entirely replaced: several of the ancestral Y genes have relocated to an autosome and the current Y is likely descended from a former autosome whose homolog became fused to the X chromosome (Carvalho and Clark 2005; Larracuente et al. 2010). An analysis using data from the 12 *Drosophila* genomes project shows that the process of Y replacement can also happen in a piecemeal fashion (Koerich et al. 2008). Testing 12 *D. melanogaster* Y-linked genes for Y linkage in each of the 11 other *Drosophila* species with sequenced genomes, Koerich et al. (2008) found that although some appear to have been Y linked throughout the 60 My history of the genus (e.g., the male fertility factors *kl-2* and *kl-3*), others have recently been acquired by the melanogaster lineage (e.g., *FDY*) or lost in other lineages (e.g., *PRY* in the lineage leading to *D. mojavensis*). The gene *kl-5*, an essential male fertility factor, is thought to have been acquired at least twice by the Y chromosome, once in the lineage leading to the melanogaster group, and independently in the common ancestor of the *Drosophila* subgenus. Overall, the degenerate *Drosophila* Y now appears to be mostly acquiring genes; Koerich et al. (2008) estimated that the rate of gene gain has been more than ten times higher than the rate of loss since the origin of the genus.

We have identified a third translocation of *kl-5*, from the Y chromosome to an autosome in the testacea group of *Drosophila*. In *D. melanogaster*, *kl-5* encodes a dynein heavy chain expressed during spermatogenesis and is essential for male fertility. When a gene moves from the nonrecombining male-limited Y chromosome, the direction of selection may change (as it can now experience selection in females), and the efficacy of selection may be drastically improved by recombination. Consistent with this idea, other duplications to an autosome have shown indirect evidence of an increase in the efficacy of selection, as evidenced by a reduction in intron size (Carvalho and Clark 2005). Here, we characterize the testacea group *kl-5* duplication and ask if patterns of selection and constraint acting on the protein sequence of *kl-5* have changed along with its genomic location.

Materials and Methods

Fly Strains and Rearing

Fly stocks of species used in this study were derived from wild-caught strains or obtained from the *Drosophila* Species Stock Center. We used species from the *Drosophila* subgenus, including a variety of species from the quinaria group and representative species from the immigrans and tripunctata groups. The primary focus of this study is the testacea species group, also in the *Drosophila* subgenus. The testacea group has a holarctic distribution and is comprised of four mushroom-feeding species: *D. putrida* is found in eastern and southern North America (Patterson and Stone 1952), *D. neotestacea* in North America, *D. testacea* in Europe and continental Asia, and *D. orientacea* in Japan and eastern Asia (Grimaldi et al. 1992; Chen et al. 1998). These species are morphocryptic but reproductively isolated: only *D. orientacea* and *D. testacea* produce fertile F1 female offspring when crossed in the laboratory; all other combinations will not mate or do not produce offspring (Grimaldi et al. 1992; Dyer K, unpublished data). The chromosomal structure of the testacea group is similar to *D. melanogaster*: flies have two metacentric autosomes, X and Y chromosomes, and a small dot chromosome. The X chromosome is syntenic with that of *D. melanogaster* (Patterson and Stone 1952; Dyer K, unpublished data). Synteny of the large autosomes to each Müller's element has not been established. Based on large-scale gene conservation of the dot chromosome in the *Drosophila* genus (Riddle and Elgin 2006; Riddle et al. 2009; Leung et al. 2010), we presume that the small dot chromosome in the testacea group is syntenic with the dot of other *Drosophila* species. Finally, males of *D. neotestacea* that lack a Y chromosome (XO) are sterile (James and Jaenike 1990; Dyer K, unpublished data).

In this study, we used wild-caught *D. putrida* (from Rochester, NY), *D. testacea* (from Munich, Germany), *D. neotestacea* (from Rochester, New York; Gatlinburg, Tennessee; Deary, Idaho; and Vancouver, British Columbia), and *D. orientacea* (from Sapporo, Japan). As there is no signature of population structure among *D. neotestacea* populations at the loci we surveyed (results not shown), data from all populations were combined in these analyses. In addition to wild-caught flies, we used three laboratory stocks: two stocks containing autosomal eye-color mutations (*dark* and *orange*) isolated from *D. testacea* collected in Munich, Germany, and an introgression stock containing an autosomal copy of *kl-5* from *D. orientacea* in a *D. testacea* background. To construct this introgression stock, we crossed *D. testacea* females with *D. orientacea* males and then backcrossed the hybrid F1 females to *D. testacea* males. We then backcrossed this stock to *D. testacea* for seven generations, using females heterozygous for the *D. orientacea* and *D. testacea* alleles for each cross. To make the *D. orientacea* allele homozygous, we performed two final crosses. We first sib-mated heterozygous males and females from the seventh backcross generation offspring, and then used the resulting offspring that were homozygous for

the *D. orientacea kl-5* allele to construct the final introgression stock. For the preceding crosses, flies were genotyped for the *kl-5* allele using a restriction fragment length difference that distinguishes the *D. orientacea* versus *D. testacea kl-5* alleles. As this genotyping can only be done after the crosses have been performed, in all cases we performed additional crosses but retained only those using flies with the appropriate genotype. All crosses were single-pair crosses using virgin females. All crosses and laboratory stocks were reared on Instant *Drosophila* Medium (Carolina Biological) supplemented with a piece of fresh *Agaricus bisporus* mushroom and housed at 20 °C on a 12-h light–dark cycle with 50% relative humidity.

Molecular Methods

Degenerate primers for *kl-5* were based on the consensus coding sequence of *kl-5* from *D. melanogaster* (GenBank accession NM_001015499) and *D. hydei* (GenBank accession AF031494). For our initial test of *kl-5* Y linkage and for phylogenetic analyses, we used degenerate primers that amplify 1,192 bp of exon 12 of *kl-5* in *D. melanogaster* (primers kl5–F8 and kl5–R6 in supplementary table S1, Supplementary Material online). We also designed primers for *kl-5* to amplify across introns, the locations of which are conserved between *D. melanogaster* and *D. hydei* (Kurek et al. 2000). Degenerate primers for all other loci were previously published or designed using the genome sequences from *D. melanogaster*, *D. pseudoobscura*, *D. virilis*, and *D. mojavensis* (Clark et al. 2007). All primers were designed using Oligo6 or Primer3, and are listed in supplementary table S1 (Supplementary Material online). Sequences from this study have been deposited in GenBank (HQ685127–HQ685755).

Genomic DNA was extracted from single flies using the Qiagen Puregene Tissue Kit, and polymerase chain reaction (PCR) was conducted using standard protocols. All PCRs for cloning used Invitrogen High Fidelity Taq, 25 cycles of PCR to reduce PCR recombination, and were cloned using a TOPO TA Invitrogen cloning kit. From each individual, at least five inserts were screened to infer haplotype in order to minimize the impact of PCR error. For all DNA sequencing, amplicons were purified using Exosap (USB), and DNA was sequenced in both directions using Big Dye v3.1 chemistry (Applied Biosystems) and run out on an ABI 3730 at core facilities at the University of Edinburgh or the University of Georgia. Chromatograms were analyzed using Sequencher (Gene Codes) and sequences were aligned manually in Se-Al (Rambaut 1996). For genotyping offspring of crosses, flies were scored visually when eye-color mutants were used, and by sequencing alleles in both parents and offspring when molecular markers were used. For some crosses, we also used a restriction fragment length polymorphism to distinguish between *kl-5 D. orientacea* and *D. testacea* alleles.

To analyze expression of *kl-5*, we used reverse transcriptase PCR (RT-PCR) (Invitrogen SuperScript III One-Step kit) with primers that span two introns in *D. melanogaster* and *D. hydei* (primers kl5-F44 and kl5-R46 in supplementary table S1, Supplementary Material online). Total RNA was isolated separately from heads and reproductive tracts of virgin males and females using Qiagen's RNeasy kit, including a DNAse step, and quantified using a Nanodrop spectrophotometer. A reaction using genomic DNA was included as a positive control. As an additional control, we used RT-PCR to amplify the autosomal *tpi* gene from the same RNA samples. PCR products were sequenced in both directions to confirm *kl-5* expression and to detect which alleles were expressed.

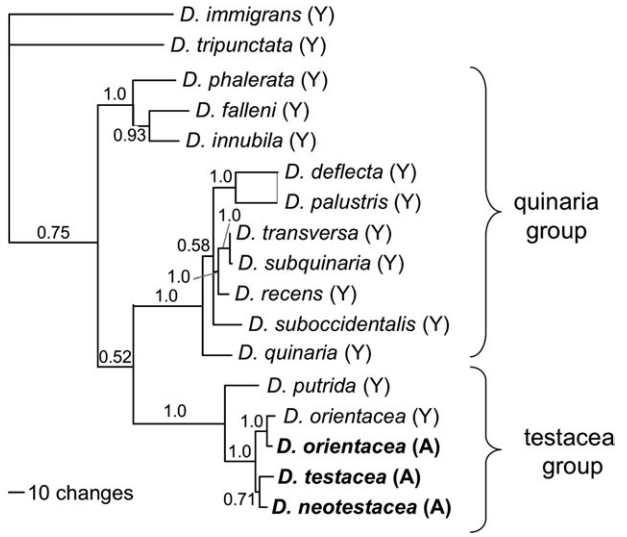

Fig. 1. Phylogeny of *kl-5* based on Bayesian phylogenetic analysis. Support value for each node is shown if that partition was found in >50% of the sampled trees. The Y or A after each species name indicates the genomic location of *kl-5*, either Y linked or autosomal. See text for details on model parameters.

Molecular Evolutionary Methods and Analyses

We sequenced a fragment of *kl-5* from all species shown in figure 1 in order to confirm that we amplified the correct locus and to infer the phylogeny of this locus. We cloned fragments from species where we identified more than one allele at *kl-5* (apparent by "double peaks" on the chromatograms). Phylogenetic analyses of *kl-5* were conducted using MrBayes v3.1 (Ronquist and Huelsenbeck 2003). Based on the results of hierarchical likelihood ratio tests in MrModelTest (Nylander 2004), we used a general time reversible $+\Gamma$ model of DNA evolution (Lanave et al. 1984; Tavaré 1986; Yang 1994). We ran the program for 5 million iterations, collecting samples every 1,000 iterations. The first 500,000 samples were excluded as burn-in, and the most probable tree was constructed from the final 4.5 million samples. We conducted four independent runs to ensure that the chains converged to the same tree.

We assayed the level of polymorphism at *kl-5* in each of the four testacea group species (*D. testacea*, *D. orientacea*, *D. neotestacea*, and *D. putrida*). Ten to 25 individual wild-caught flies of each species were sequenced for approximately 1800 bp of *kl-5* coding sequence (primers kl5-F8 and kl5-R5 in supplementary table S1, Supplementary Material online), which includes the fragment used for

the Y survey and corresponds to part of exon 12 in *D. melanogaster*. To test for the presence and amount of recombination, we inferred phase directly by cloning and/or sequencing F1 offspring of wild-caught females and males.

To compare the ratio of diversity and divergence in *kl-5* with other genes in the genome, we sequenced several additional loci in wild-caught individuals from each of the four testacea group species, as well as from a single individual from as many of the other species surveyed as possible. For *D. neotestacea*, we used flies from Rochester, New York and Gatlinburg, Tennessee, a subset of the populations used for *kl-5* diversity in this species. The loci included *tpi*, which is located on an autosome; *bt*, *eIF4G*, *gw*, and *plexA*, located on the dot chromosome; *mof*, *pgd*, *RpL36*, and *skpA*, which are X linked; and *kl-2* and *kl-3*, which are Y linked. Analyses of diversity and deviations of the site frequency spectrum (Tajima 1989) from neutrality were completed in DnaSP (Librado and Rozas 2009), and Hudson-Kreitman-Aguade (HKA) tests (Hudson et al. 1987) were performed using MLHKA (Wright and Charlesworth 2004). For the HKA analyses, we only used *kl-5* sequences from females.

To compare rates of evolution in lineages of *kl-5* that are autosomal versus Y linked, we used PAML (Yang 2007) to estimate the standardized rate of protein evolution (d_N/d_S, or ω) under different evolutionary models, and compared models using likelihood ratio tests. To examine sequence data for evidence of positive selection, we used polarized and unpolarized McDonald–Kreitman (MK) tests (McDonald and Kreitman 1991; Holloway et al. 2007). Finally, we used maximum likelihood methods (Welch 2006; Betancourt et al. 2009) to estimate α, the estimated proportion of replacement substitutions attributable to positive selection. For the sample of genome-wide loci, we estimated the confidence intervals for α using bootstrap samples; confidence intervals for *kl-5* were calculated numerically, by evaluating the likelihood for different values of α. For loci or classes of loci with no polymorphism (e.g., loci on the dot in some species), it is not possible to obtain α estimates.

Functional Assay of Male Fertility

We measured male fertility using three separate assays. In the first, a single male and single female were placed in a vial, transferred to fresh food after 4 days, and then removed after an additional 4 days. All offspring were counted. In the second, we used a sperm exhaustion assay to detect more subtle differences in male fertility. In this assay, we placed a single male with five virgin females for 24 h, and then transferred the male to a new vial with five virgin females for 24 h. The females from each vial were kept as a group and transferred to fresh food every second day for 14 days, and all their offspring were counted. In the first two assays, males were 6 days old and females were 7–10 days old when initially used. All females were pure *D. testacea*, and males were either pure *D. testacea* or *D. testacea* homozygous for *kl-5* introgressed from *D. orientacea*. Flies were reared at a low density, collected on light CO_2 as virgins, and placed together in vials by gentle air aspiration. In the third assay, we dissected the testes of sexually mature males from the introgression stock described above that were isolated from females for 1 week, and scored motility of mature sperm from the seminal vesicles. Males were classified as having many, few, or no motile sperm (as in Orr 1992).

Results

Y Linkage and Phylogenetic History of *kl-5*

We tested for Y linkage of *kl-5* in each of 16 species by amplifying part of the gene from four males and four females (fig. 1). In most species tested, *kl-5* amplified only in males, demonstrating that it is located on the Y chromosome. The exception was the testacea group: although in *D. putrida*, *kl-5* amplified only in males, in *D. neotestacea*, *D. orientacea*, and *D. testacea*, it amplified in both males and females. As *D. putrida* is an outgroup to the other three testacea group species (Perlman et al. 2003), this result suggests the possibility of a non–Y-linked duplication of *kl-5* in the common ancestor of *D. testacea*, *D. orientacea*, and *D. neotestacea*.

We sequenced the amplified fragment to confirm that the amplification in females was due to a recent duplication of *kl-5*, and not artefactual amplification of a related gene. An ancient duplicate would show high levels of protein divergence from *kl-5* gene of other *Drosophila*, but the fragment sequenced from the testacea group species is highly similar to that of other *Drosophila*, regardless of Y linkage (supplementary figure S1, Supplementary Material online). Phylogenetic analysis confirms that the testacea group *kl-5* is not an ancient duplicate (fig. 1): the testacea group *kl-5* is not basal to the other 13 species, as would be expected if it were not *kl-5*, but a related gene. Instead, the phylogenetic analysis places the testacea group in relation to other *Drosophila* in a position similar to that in a phylogeny based on mitochondrial DNA (Perlman et al. 2003).

Genomic Location of *kl-5*

We further characterized the genomic location of *kl-5*, asking if it shows patterns consistent with non–Y-linked inheritance and whether the Y-linked copy is retained in the genome. Sequencing of wild-caught testacea group flies showed that both males and females were frequently heterozygous, suggesting that the putatively new copy of *kl-5* is autosomal, rather than X-linked (summary of data in table 1; alleles shown in supplementary table S2, Supplementary Material online). For all three species where *kl-5* appears to have a non–Y-linked copy, we tested for a residual Y-linked copy, using different approaches for each species. In *D. neotestacea*, we used genetic crosses, crossing two *D. neotestacea* stocks fixed for different *kl-5* alleles. All three *kl-5* genotypes (both homozygotes and the heterozygote) occurred among both male and female F2 offspring, as is expected for an autosomal locus. Note that this result is not expected for either an X-linked locus

Table 1. *kl-5* polymorphism in the testacea group.

			All Sites					Synonymous					Replacement					
Species	M/F	N	*L*	*S*	π	θ	*D*	*L*	*S*	π	θ	*D*	*L*	*S*	π	θ	D	π_A/π_S
putrida	M	20	2048	5	0.033	0.069	−1.59*	456.7	5	0.149	0.309	−1.59*	1586.3	0	0	0	NA	0.00
orientacea	F	14	2048	11	0.112	0.169	−1.32	451.8	6	0.269	0.417	−1.27	1591.2	5	0.068	0.099	−1.09	0.25
orientacea	M (Y)	12	2048	0	0	0	NA	451.0	0	0	0	NA	1593.0	0	0	0	NA	0.00
testacea	M + F	41	2038	10	0.082	0.115	−0.86	453.9	8	0.246	0.412	−1.15	1583.1	2	0.035	0.03	−0.33	0.14
testacea	F	17	2038	7	0.089	0.102	−0.44	453.8	6	0.282	0.391	−0.94	1583.2	1	0.033	0.019	1.53	0.12
neotestacea	M + F	50	2048	2	0.034	0.022	0.99	455.4	1	0.112	0.049	1.73	1587.6	1	0.012	0.014	−0.24	0.10
neotestacea	F	24	2048	2	0.029	0.026	0.27	455.5	1	0.114	0.059	1.57	1587.5	1	0.005	0.017	−1.16	0.05

NOTE.—Shown are number of alleles (*N*), polymorphism measured by π and θ, each $\times 10^{-2}$, length in base pairs (*L*), segregating sites (*S*), and Tajima's *D* (*D*). Samples sorted by both males (M) and females (F) or by one sex only. M (Y) in *D. orientacea* is for the Y-linked copy only. Significance indicated by an asterisk at the $P = 0.05$ level, based on coalescent simulations. NA indicates that Tajima's D cannot be calculated due to a lack of segregating polymorphism.

or for a combination of autosomal and Y-linked loci. Consistent with the lack of Y copy in *D. neotestacea*, no segregating variant found in two or more individuals was exclusive to males ($S = 2$).

In *D. testacea*, the lack of male-specific variants also suggests the absence of a Y-linked copy of *kl-5* (again, no segregating polymorphism found in two or more individuals was exclusive to males, with $S = 4$). We eliminated the possibility of a Y-linked copy that was very similar in sequence to the autosomal copy by using an introgression stock, which has the autosomal copy of *kl-5* from *D. orientacea* in a largely *D. testacea* background. A Y-linked *D. testacea* allele would be easily distinguished from the introgressed *D. orientacea* allele by sequencing, and should occur in all males from the introgression stock. However, sequences of males from this stock show no evidence of a *D. testacea kl-5* allele. In addition, the pattern of segregation of *kl-5* is inconsistent with a Y-linked *D. testacea* copy, as sibling matings between flies heterozygous for the *D. orientacea* and *D. testacea* alleles at *kl-5* (after seven generations of backcrossing, see Materials and Methods) produce some male offspring with only *D. orientacea kl-5* alleles ($n = 2$ crosses, with 30/119 and 20/107 male offspring homozygous for *D. orientacea* alleles, close to the expected 1:3 ratios).

In *D. orientacea*, in contrast, both autosomal and Y-linked copies of *kl-5* appear to be present: 11 of 19 segregating sites are present only in males (supplementary table S2, Supplementary Material online), suggesting that we have amplified both Y-linked and non–Y-linked copies of *kl-5* in this species. We confirmed the presence of this Y-linked copy in two ways. First, we showed that these male-specific variants (identified in the population sample) occurred only in the sons, and not in the daughters, of wild-caught females (among six wild-caught females tested, 9/9 sons and 0/12 daughters had the 11 male-specific variants). Second, we were able to show that these variants were entirely linked, as would be expected if they distinguished a Y linked from an autosomal copy of *kl-5*. We did this by directly sequencing each haplotype, using primers specific to the putative Y-linked and autosomal copies. These data also show that the putatively Y-linked copy of *kl-5* in *D. orientacea* does not harbor any polymorphic sites. To summarize the results for all four testacea group species: we find that *D. neotestacea* and *D. testacea kl-5* only have autosomal copies of *kl-5*, whereas *D. orientacea kl-5* has both autosomal and Y-linked copies and in *D. putrida kl-5* is Y linked.

Having established that there is an autosomal copy of *kl-5*, we attempted further to resolve its location. Because the *D. orientacea* autosomal copy is potentially independent of that in the other species (see below), we first establish whether or not it is in the same location as in *D. testacea* using a segregation test. To do this, we backcrossed fertile F1 *D. testacea–D. orientacea* hybrid females to each of the parental species. We constructed the backcross so that we could distinguish three *kl-5* alleles by sequencing: the *D. orientacea* and *D. testacea* alleles present in the hybrid female, and a third allele from the male parent, which was homozygous for a different *D. orientacea* or *D. testacea* allele than that in the hybrid female. Genotyping of the backcross progeny shows that the *D. orientacea* and *D. testacea kl-5* alleles cosegregate. No individuals carrying all three alleles were observed among 43 progeny, indicating that *kl-5* is on the same linkage group in these two species (two locus model is rejected with $\chi^2 = 14.55$, $P = 0.0009$).

We tested *kl-5* for linkage to several available markers, including two autosomal eye-color mutations (*dark* and *orange*; both unmapped) and three loci with sequence variants (*tpi*, on Müller's element E, or 3R in *D. melanogaster*, and *bt* and *gw*, both on Müller's element F, or the dot chromosome, in *D. melanogaster* and in all sequenced *Drosophila* species for which their locations are known; Clark et al. 2007). As there was little within-species polymorphism at dot chromosome loci available for mapping, for these loci we instead followed the segregation of between species differences in backcross progeny between *D. orientacea* and *D. testacea*. The results show that *kl-5* is completely linked to the dot loci: variants at *kl-5* show perfect association with variants at *bt* and *gw* ($n = 14$; uncorrected cM = 0). In contrast, *kl-5* shows no association with any of the other three tested markers (*tpi*: $n = 15$ in the same backcross progeny to establish dot linkage, uncorrected cM = 80; *orange*: $n = 48$; uncorrected cM = 37.5, *dark*: $n = 52$; uncorrected cM = 34.6).

The small dot chromosome is conserved throughout most *Drosophila* species (Riddle and Elgin 2006; Riddle et al. 2009; Leung et al. 2010), and present in the testacea

group (Patterson and Stone 1952). Consistent with the general conservation of gene content within Müller's elements in *Drosophila*, most of the dot-linked gene content is conserved between *D. melanogaster* and *D. virilis*, an evolutionary distance similar to that of *D. melanogaster* and the testacea group (of 69 genes dot linked in *D. melanogaster*, 66 remain dot linked in *D. virilis*, including *bt* and *gw*) (Leung et al. 2010). Nevertheless, it is possible that the two markers we used are no longer dot linked in the testacea group. We therefore further confirmed linkage of *kl-5* to the dot chromosome using four dot-linked loci—*bt*, *gw*, *eIF4G*, and *plexA*, all of which again appear to be dot linked throughout the genus (Clark et al. 2007)—and flies from the introgression stock. Specifically, we crossed *D. testacea* males to females from the sixth backcross during the construction of the introgression stock. The females used here were heterozygous for autosomal copies of *D. orientacea* and *D. testacea kl-5*, but primarily *D. testacea* in genome composition (with 2% of loci unlinked to *kl-5* expected to originate from *D. orientacea*). Among 32 genotyped offspring in crosses to *D. testacea*, there was a perfect association between *kl-5* and all four of the dot-linked loci (14 individuals heterozygous for *D. orientacea* and *D. testacea* alleles at all five loci, 18 homozygous at all loci, 2% linkage rejected with $\chi^2 = 35.12$, $P < 1 \times 10^{-5}$ from simulation).

Phylogenetic Reconstruction of *kl-5* Duplication and Loss

We used the phylogeny recovered from the *kl-5* sequences to ask about the history of duplication and loss of *kl-5* in these species. The simplest scenario—a single Y to autosome translocation of *kl-5* in the ancestor of all three species, followed by a single loss of the Y-linked copy in the ancestor of *D. neotestacea* and *D. testacea*—is incompatible with the observed phylogeny. If this reconstruction were correct, we would expect the Y-linked *D. orientacea kl-5* to be basal to all of the autosomal copies; instead, the autosomal and Y-linked copies in *D. orientacea* cluster together (fig. 1 and supplementary figure S2, Supplementary Material online). Note that scenarios involving incomplete lineage sorting, although they can give rise to incongruities between gene and species trees, would also be expected to result in a basal *D. orientacea* Y-linked copy—that is, as long as there was a single origin of the autosomal copies *kl-5*, these copies should form a monophyletic group. It is unlikely that this incompatibility is due to a misinferred phylogeny as the clustering of these two copies has strong support (posterior probability = 1.00), and as the *D. orientacea* Y and autosomal copies share a total of 15 private synapomorphies.

The phylogeny in figure 1 is consistent with two parsimonious reconstructions, both involving two translocations within the testacea group (supplementary figure S3, Supplementary Material online). First, *kl-5* may have translocated from the Y to an autosome twice: once in the common ancestor of *D. testacea* and *D. neotestacea* (followed by loss of the Y copy), and again in the *D. orientacea* lineage (without the loss of the Y copy). Second, *kl-5* may have moved once from the Y to an autosome in the common ancestor of all three testacea group species (followed by the loss of the Y-linked copy), and then moved back to the Y in *D. orientacea* lineage. A third possibility, also consistent with the phylogenetic reconstruction, is a single Y to autosomal movement of *kl-5* followed by subsequent homogenization of the autosomal and Y-linked copies in *D. orientacea* via nonhomologous recombination.

Mechanism of Duplication

We asked whether the testacea group species show a wholesale relocation of Y-linked genes by testing two additional genes (*kl-2* and *kl-3*) for Y linkage in the testacea group. Both genes are Y linked in *D. melanogaster* and 11 other *Drosophila* species from throughout the genus (Koerich et al. 2008). In all four testacea group species, we were able to amplify one fragment of *kl-2* and three independent fragments of *kl-3* in males, but not in females, indicating that these genes are still Y linked in this group. In addition, we were able to amplify *kl-3* in a broad range of related species, all of which showed Y linkage. Thus, it appears that the Y to dot duplication in the testacea group is limited to *kl-5* or a portion of the Y chromosome and does not encompass the entire Y.

One possible mechanism for small-scale translocations of this nature is retroposition via an mRNA intermediate. If *kl-5* has been retrotransposed, we expect that it will either lack introns or have introns in different locations than in species where the gene is Y linked. Intron locations in Y-linked copies of *kl-5* are conserved between two highly divergent *Drosophila* species, *D. melanogaster* and *D. hydei* (Kurek et al. 2000). We used several degenerate primer sets that span four putative introns of *kl-5* (introns two through five in *D. melanogaster*) and attempted to amplify across each of them in all four testacea group species. All four introns are present in the same locations as *D. melanogaster* in *D. orientacea*, *D. testacea*, and *D. neotestacea* (supplementary table S3, Supplementary Material online). Locations of two of these four introns were also conserved in *D. putrida* (we were unable to amplify the relevant product for the others). This suggests a DNA-based mechanism of duplication, as might occur, for example, via a DNA transposon or a segmental duplication. In *D. orientacea* males, the amplicon that spans introns two through four contains a male-specific variant; thus, for these introns, the Y-linked and autosomal copies are identical in length and location.

Genes on the Y chromosome, where *kl-5* is normally found, necessarily have male-limited expression; and, in species where it has been previously studied, *kl-5* function is specific to the male reproductive tract. We asked if the autosomal copy of *kl-5* has retained a testes-specific expression pattern in the testacea group by amplifying *kl-5* transcripts from different tissues using RT-PCR (fig. 2). The results show that, as expected, *kl-5* expression in *D. putrida* is limited to the male reproductive tract. In flies where *kl-5* is dot linked, *kl-5* is weakly expressed in females and in the

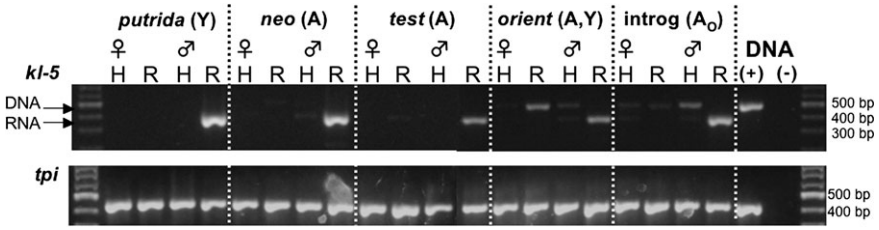

FIG. 2. RT-PCR results of *kl-5* expression from virgin adult flies. Each of the four testacea group species, *Drosophila putrida* (putrida), *D. neotestacea* (neo), *D. testacea* (test), *D. orientacea* (orient), and the introgression stock (Introg) is included, and separated by male and female head (H) or reproductive tract (R). Also indicated is the genomic location of *kl-5*, Y linked (Y) and/or autosomal (A). The *kl-5* fragment amplified is across two introns, and is 489 bp for DNA versus 383 bp for RNA. RT-PCR on the autosomal gene *tpi* is shown as a positive control.

male head, and highly expressed in the male reproductive organs. This expression pattern is consistent with the bulk of *kl-5* expression occurring during spermatogenesis, and indicates that the autosomal duplication of *kl-5* likely included its *cis*-regulatory elements (again implicating a DNA-based mechanism of duplication).

We sequenced the RT-PCR products to check for previously unidentified copies of *kl-5*, and to ask whether both autosomal and Y-linked copies are expressed in *D. orientacea*. As expected, sequencing of the RT-PCR products of *D. testacea* and *D. neotestacea* did not reveal any additional nucleotide variants, which would indicate a Y-linked copy of *kl-5* not previously identified. We also found that the introgression stock expresses only the dot-linked copy of *D. orientacea* allele of *kl-5*, as expected. In *D. orientacea*, the coding sequence of the RT-PCR fragment used for the expression analysis does not contain any differences between the Y- and dot-linked copies. We therefore sequenced a different RT-PCR product (amplified with primers F8 and R6 in supplementary table S1, Supplementary Material online) and found that both the Y-linked and autosomal alleles are present, as evidenced by double peaks in the sequencing chromatogram corresponding to differences between the duplicates.

Molecular Evolutionary Analyses

We compared polymorphism, recombination, and divergence between autosomal and Y-linked copies of *kl-5*, and between *kl-5* and other loci in the genome. First, we examined DNA sequence polymorphism at *kl-5* in all four testacea group species; summary statistics of polymorphism and divergence are shown in table 1 and supplementary table S4 (Supplementary Material online). Across approximately 1800 bp of coding sequence, the level of silent polymorphism (θ_S) of the dot chromosome copy ranges from $\theta_S = 0.0006$ (in *D. neotestacea* females) to $\theta_S = 0.004$ (in *D. orientacea* and *D. testacea* females). Y-linked copies of *kl-5* exhibit no polymorphism in *D. orientacea* and very low polymorphism in *D. putrida* ($\theta_S = 0.003$).

Like dot and Y-linked loci in other *Drosophila* species (Zurovcova and Eanes 1999; Wang et al. 2002; Wang et al. 2004; Betancourt et al. 2009; Arguello et al. 2010), the level of polymorphism at *kl-5* appears to be lower than is typical for related *Drosophila* species (Dyer and Jaenike 2004; Dyer et al. 2007). However, it is formally possible that polymorphism in these species is depressed genome-wide and not unusually low at *kl-5*. We therefore used HKA tests (Hudson et al. 1987) to statistically compare the level of variation at *kl-5* to that of nine other loci. Specifically, we surveyed polymorphism in *D. neotestacea*, *D. testacea*, and *D. orientacea* at one autosomal locus (*tpi*), four X-linked loci (*mof*, *RpL36*, *skpA*, and *pgd*), four dot-linked loci (*bt*, *eIF4G*, *gw*, and *plexA*), and two Y-linked loci (*kl-2* and *kl-3*) (supplementary table S5, Supplementary Material online), and estimated divergence at these loci relative to *D. putrida*. (In the HKA tests, we used appropriate scaling factors to adjust for the expected reduction in N_e for X- and Y-linked loci due to inheritance patterns alone.) Consistently across all three surveyed species, when *kl-5* is compared with Y-linked or dot loci, there is little difference among the loci. However, *kl-5* alone and *kl-5* and the Y and dot loci together show a severe and significant reduction in polymorphism compared with X-linked and autosomal loci (fig. 3 and supplementary table S6, Supplementary Material online).

Although the lack of polymorphism at *kl-5* is likely due to the low recombination environment of the dot chromosome, in principle the reduction could be due to a recent selective sweep or purging of neutral variation caused by the fixation of a nearby positively selected allele. There is no strong evidence of such a sweep, however, seen in the frequency spectrum estimated from the polymorphism data. That is, a complete sweep would result in an excess of rare variants at *kl-5* indicated by a negative Tajima's *D* (Tajima 1989), but *D* is not significantly negative in any of the species (table 1). An incomplete sweep—one currently in the process of reaching a high frequency—would be signified by an excess of high frequency–derived segregating sites, as indicated by a negative Fay and Wu's *H* (Fay and Wu 2000). Here, the results are more ambiguous; two species show some evidence of an ongoing sweep at *kl-5* (*D. orientacea*: $H = -1.88$, $P = 0.044$; *D. testacea*: $H = -1.88$, $P = 0.040$), but the third does not (*D. neotestacea*; $H = 0.29$, $P = 0.446$; *D. putrida* was used as the outgroup to calculate the *H* statistic in all cases). However, because a variety of demographic scenarios can also produce a negative *H*, this signal alone is generally not considered to be sufficient evidence for positive selection (Wakeley and Aliacar 2001; Przeworski 2002; Jensen et al. 2005).

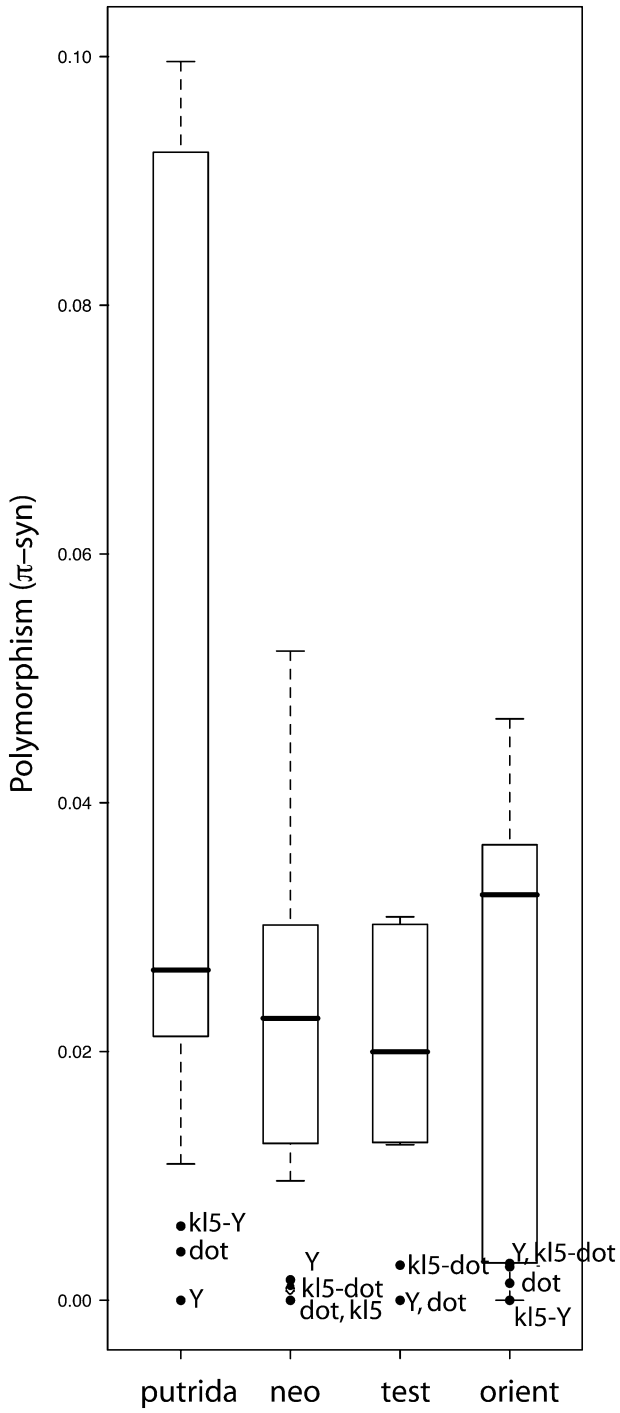

Fig. 3. Comparison of within-species polymorphism in the testacea group. Boxes show values for X-linked and autosomal loci, with boxes encompassing the middle 50% of values and whiskers extending to extreme values. Polymorphism values for X-linked loci have been multiplied by 4/3 and for Y-linked loci by 4 to correct for difference in population size due to inheritance patterns. The dots show estimates averaged across loci for Y-linked and dot loci and the estimates for the dot-linked copy of *kl-5* separately.

In any case, the low polymorphism probably largely reflects the lack of recombination at this locus: the autosomal copy of *kl-5* shows no evidence of intralocus recombination in any of the species ($R_m = 0$), as might be expected given the lack of crossing over on the dot chromosome seen in other species (Wang et al. 2002, 2004; Riddle and Elgin 2006; Betancourt et al. 2009; Arguello et al. 2010).

We next asked if the *kl-5* duplication is associated with long-term changes in the rate of protein evolution, using the codeml program within the PAML package (Yang 2007) and *kl-5* sequences from the species shown in figure 1 (table 2 and supplementary figure S4, Supplementary Material online). First, we found that autosomal copies of *kl-5* in the testacea group species show a rate of protein evolution approximately 5-fold higher than that of Y-linked copies of *kl-5* (model with uniform ω: ln $L = -4604.45$, $\omega = 0.033$; model with separate ω for testacea group excluding *D. putrida*: ln $L = -4595.250$, $\omega_Y = 0.028$, $\omega_{testacea} = 0.151$; $\Lambda = 18.4$, $P < 0.0001$). There was no evidence that the higher rate of protein evolution was solely due to the duplication events, as estimating an additional ω for this branch did not significantly improve the fit of the model and only marginally affected the ω estimates (Model 4 ln $L = -4595.197$; $\omega_Y = 0.028$, $\omega_{testacea} = 0.144$, $\omega_{duplication} = 0.198$; $\Lambda = 0.1$, $P = 0.94$). The elevation in the rate of evolution is also not due to any (undetected) recombination between the two *D. orientacea* copies, as excluding these sequences does not qualitatively change the results (model with uniform ω: ln $L = -4844.19$, $\omega = 0.033$, ln $L = -4595.197$; model with separate ω for testacea group excluding *D. putrida*: ln $L = -4840.21$ $\omega_Y = 0.031$, $\omega_{testacea} = 0.133$; $\Lambda = 7.9$, $P = 0.005$). Visual inspection of the protein alignment (supplementary figure S1, Supplementary Material online) indicates that 13 amino acid changes (of 397 total amino acid sites) occurred in the testacea group lineages, which suggests that the increased ω is not due to just one or a few amino acid changes.

A high rate of protein evolution might be the result of either adaptive evolution or relaxed purifying selection. We attempted to distinguish between these forces using MK tests and estimates of α (the proportion of excess amino acid substitutions over that expected due to drift alone; McDonald and Kreitman 1991; Smith and Eyre-Walker 2002). These tests use the ratio of polymorphism to divergence at synonymous sites to that at nonsynonymous sites: Under neutrality, we expect these ratios to be equal, whereas recurrent adaptive evolution results in greater than expected nonsynonymous divergence. We conducted both unpolarized (using the total number of substitutions between two species) and polarized (using an outgroup to assign substitutions between species to one of the two lineages) MK tests for polymorphisms within *D. orientacea*, *D. testacea*, and *D. neotestacea*, and divergence to *D. putrida*. All of these tests were nonsignificant ($P > 0.15$; supplementary table S7, Supplementary Material online). Due to low polymorphism levels, estimates of α at *kl-5* loci are highly unreliable (or, in cases where there is no polymorphism, unobtainable), with broad confidence intervals ($\alpha_{testacea} = 0.19$ [−4.3 to 0.96], $\alpha_{orientacea} = -2.70$ [−12.8 to 0.04], $\alpha_{neotestacea} = -3.85$ [−28.95 to 0.82]). The confidence intervals for these α estimates overlap with those for α estimates from the set of genome-wide loci. However, α

Table 2. PAML molecular evolutionary analysis.

Locus	Genome Location	*N*	Model	ω_1	ω_2	ω_3	Ln *L*	Model Comparison	df	LR	*P* value
kl-5	dot, Y	16	1	0.033			−4604.452				
			2	0.028	0.151		−4595.250	1 vs. 2	1	18.40	0.000
			3	0.031	0.189		−4602.337	1 vs. 3	1	2.11	0.040
			4	0.028	0.144	0.198	−4595.197	1 vs. 4	2	18.50	0.000
								2 vs. 4	1	0.10	0.940
kl-3	Y	15*	1	0.015			−4766.801				
			2	0.011	0.093		−4751.924	1 vs. 2	1	29.77	0.000
bt	dot	12*	1	0.054			−2562.544				
			2	0.049	0.127		−2560.570	1 vs. 2	1	3.95	0.047
eIF4G	dot	14*	1	0.030			−2899.242				
			2	0.030	0.001		−2898.686	1 vs. 2	1	1.112	0.291
plexA	dot	14	1	0.030			−3053.318				
			2	0.029	0.051		−3053.115	1 vs. 2	1	0.406	0.524
tpi	A	16	1	0.037			−1248.009				
			2	0.033	0.306		−1245.216	1 vs. 2	1	5.59	0.018
pgd	X	14	1	0.032			−2714.142				
			2	0.031	0.042		−2714.037	1 vs. 2	1	0.21	0.647
mof	X	15	1	0.016			−2377.556				
			2	0.017	0.001		−2376.529	1 vs. 2	1	2.05	0.155
RpL36	X	14	1	0.011			−570.425				
			2	0.012	0.001		−570.342	1 vs. 2	1	0.08	0.773
skpA	X	15	1	0.003			−1340.023				
			2	0.003	0.000		−1339.821	1 vs. 2	1	0.40	0.526

NOTE.—Included are the number of species (*N*) used in analysis, in which for loci with an asterisk *Drosophila virilis* was used as the outgroup, as these loci did not amplify in *D. immigrans* or *D. tripunctata*; d_N/d_S value, as indicated by ω; likelihood of the model (ln *L*); and likelihood ratio (LR) statistic. Models are as follows: model 1 has one ω for entire tree (ω_1); model 2 has one ω for the nontestacea group species plus the linage to *D. putrida* (ω_1) and a second for the testacea group (ω_2), excluding the lineage to *D. putrida*; model 3 has one ω for the lineage leading to the common ancestor of *D. neotestacea*, *D. orientacea*, and *D. testacea* (ω_2), and another for the remaining lineages (ω_1); model 4 has three ω values: one for the nontestacea group species (ω_1), a second for *D. neotestacea*, *D. orientacea*, and *D. testacea* (ω_2), and a third for the common ancestor of these three species (ω_3). These models are depicted visually in supplementary figure S4 (Supplementary Material online).

values for *kl-5* tend to be both lower than estimates for other loci and negative (supplementary figure S5, Supplementary Material online). Negative estimates of α in other low recombination regions indicate an excess of protein polymorphism relative to divergence (Sheldahl et al. 2003; Presgraves 2005; Betancourt et al. 2009), suggesting that purifying selection is too weak to prevent the segregation of deleterious protein variants at *kl-5*. Consistent with this interpretation, the ratio of nonysnonymous to synonymous polymorphism within species, π_A/π_S, is significantly higher than that of five unlinked loci in species where *kl-5* is on the dot. (table 1 and supplementary figure S5, Supplementary Material online; *D. orientacea*: π_A/π_S-*kl-5* = 0.25, π_A/π_S-unlinked = 0–0.04 [95% confidence intervals; $Z = 5$, $P < 0.0001$]; *D. testacea*: π_A/π_S-*kl-5* = 0.12, π_A/π_S unlinked = 0–0.06 [$Z = 2.3$, $P < 0.018$]; *D. neotestacea*: π_A/π_S-*kl-5* = 0.05, π_A/π_S unlinked = 0–0.03 [$Z = 1.97$, $P < 0.048$]). In contrast, in *D. putrida* where *kl-5* is Y linked, there are only synonymous segregating sites, and the π_A/π_S ratio is not different than that of the unlinked loci (π_A/π_S-*kl-5* = 0, π_A/π_S unlinked = 0−0.03; $Z = 0$, $P = 1.0$).

Finally, we attempted to distinguish whether the high rate of protein evolution of testacea group *kl-5* is specific to this gene or due to some more general characteristic of the testacea group or of the locus. To accomplish this, we collected sequence data from one autosomal (*tpi*), four X-linked (*mof*, *pgd*, *RpL36*, and *skpA*), and three dot-linked loci (*bt*, *eIF4G*, *plexA*) from as many of the 16 species in figure 1 as possible. (The other genes used in this study—*kl-2* and *gw*—did not amplify consistently outside the testacea group.) For all sequenced genes, we repeated the same PAML analyses as before, comparing a model with one ω estimate for the whole tree with a model with a separate ω estimated for the testacea group excluding the lineage to *D. putrida*. Note that the genealogies for these genes are not all consistent with the phylogeny in figure 1; however, the testacea group is always 100% monophyletic, and we use the conservative approach of applying the genealogy from each gene rather than forcing the same tree for all genes.

The results of the PAML analysis show that high ω estimates are not a clade-wide characteristic of the testacea group, such as might be the case if these species tend to have low effective population sizes (table 2 and fig. 4). Of the five X-linked and nondot autosomal loci, only one (*tpi*) showed elevated ω for the group, and the increase was small. In addition, the elevated ω estimate for testacea group *kl-5* does not appear to be a general characteristic of dot chromosome loci in these species, as only one of three dot loci shows this pattern (*bt*; table 2). However, high rates of protein evolution may be characteristic of testacea group sperm proteins—*kl-3* does show higher testacea group ω estimates—though we were only able to test this one other gene for this pattern.

Functional Assay of Male Fertility

Because *D. orientacea* retains a Y-linked copy of *kl-5*, it may be that the dot copy is nonfunctional or substantially impaired in function. The presence of fertile males in the introgression stock, which lack a Y-linked allele and rely solely

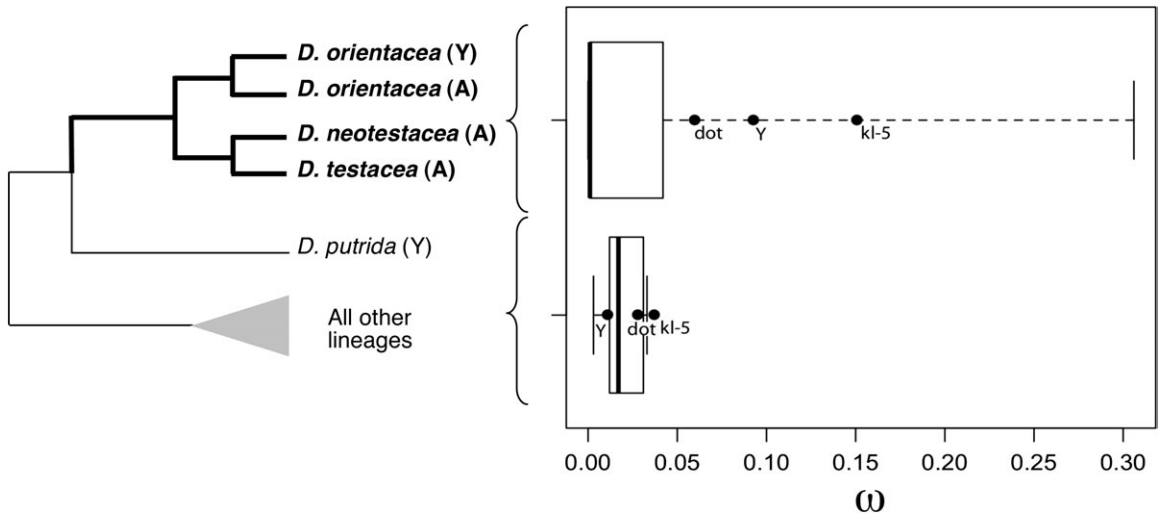

FIG. 4. Rates of protein evolution for testacea group lineages versus other lineages. The phylogeny shows the tree inferred for *kl-5*, with Y or autosomal linkage indicated after the species name. The box plot shows PAML estimated rates of protein evolution (ω, or d_N/d_S), with two rates estimated: 1) ω for the testacea group lineages (shown in bold), and 2) ω for all other lineages (i.e., ω_1 and ω_2 of model 2 in table 2). The boxes indicate the ω values for unlinked X and autosomal loci (with boxes enclosing the middle 50% of values and whiskers extending to the extreme values), the mean ω values for the dot and Y-linked loci, and the ω value for *kl-5*.

on a *D. orientacea* autosomal copy of the locus, shows that the gene is functional. We asked if the *D. orientacea* dot copy shows subtler defects in function by comparing the fertility of the introgressed males with that of pure species *D. testacea* males. In the first assay, we compared the fertility of a single male via mating with a single female, and found no difference in the number of offspring produced regardless of which species the dot chromosome was from (Kruskal–Wallace $\chi^2 = 0.23$, $P = 0.63$; fig. 5). Next, we conducted a sperm exhaustion assay, which may be more sensitive to subtle differences in fertility. Introgression males with a dot chromosome from *D. orientacea* produced on average more offspring than pure species *D. testacea* males, a result that is marginally significant (Kruskal–Wallace $\chi^2 = 3.81$, $P = 0.051$; fig. 5). Finally, we dissected mature sperm from 30 *D. testacea* males with an introgressed *D. orientacea* dot chromosome, and all 30 males had "many" motile sperm. In sum, these results indicate that there are no hybrid incompatibilities between loci residing on the *D. orientacea* dot and loci elsewhere in the *D. testacea* genome; furthermore, the preservation of high levels of fertility shows that the *D. orientacea* dot copy has not suffered the fixation of severely deleterious mutations.

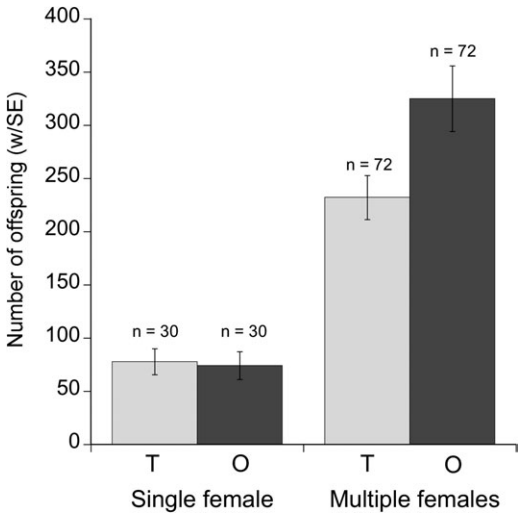

FIG. 5. Male fertility of pure species *Drosophila testacea* males versus *D. testacea* with the dot chromosome introgressed from *D. orientacea*. In the first assay, a single male was paired with a single female for 8 days, and in the second assay a single male was paired with ten females over 2 days; see text for details. The male type is indicted by the T (pure *D. testacea*) or O (introgression male). In no case were the introgression males less fertile than the pure species males (see text for details).

Discussion

In contrast to the long tradition of study of the evolution of new sex chromosomes, the ongoing evolution of degenerate Y chromosomes has only recently received attention (Skaletsky et al. 2003; Koerich et al. 2008; Hughes et al. 2009). Here, we have characterized a duplication of the gene *kl-5* from the Y chromosome to the dot chromosome in the common ancestor of three closely related testacea group *Drosophila* species. The initial translocation may have had a selective advantage, perhaps due to the elevated expression levels experienced by genes on the dot (Johansson et al. 2007) or to an overall increase of the number of gene copies and thus expression of *kl-5*. The translocation of a male fertility factor from a degenerate nonrecombining Y chromosome to any other region of the genome would also seem to enjoy a long-term evolutionary advantage; however, this Y-linked gene has merely moved to another region of reduced recombination.

It is notable that both the movement of *kl-5* characterized here and that of the whole ancestral *Drosophila* Y

chromosome in the obscura group of *Drosophila* (Carvalho and Clark 2005; Larracuente et al. 2010) involve translocations from the Y to the small dot chromosome. This apparent propensity of the dot and Y chromosome to exchange genetic material may have a mechanistic basis. That is, DNA-based duplications and translocations may be especially common in heterochromatic regions, such as those that make up large portions of the Y and dot chromosomes (Fiston-Lavier et al. 2007). Classical genetics suggests that the dot and Y chromosomes are near one another during meiosis as the X chromosome sometimes pairs with the dot as well as the Y (i.e., unpaired dot chromosomes cause high rates of nondisjunction of the X chromosome in *D. melanogaster*; Sandler and Novitski 1956). Recent work shows that the basis of this association may be physical: heterochromatic DNA threads form between the dot and X chromosome in females during achiasmatic meioses (Hughes et al. 2009), between the X and Y chromosomes in *Drosophila* males, where meiosis is always achiasmatic (Cooper 1964), and also between homologous pairs of dot chromosomes (Hartl et al. 2008). If interheterolog threads occasionally form between the dot and Y chromosome, breakage and subsequent DNA repair may result in the exchange of genetic material between these parts of the genome. Finally, in the melanogaster group, the dot and Y chromosome are known to share some repetitive satellite sequences (Lohe et al. 1993; Bayes and Malik 2009). If generally true of *Drosophila*, the process of genetic exchange between these chromosomes might be facilitated by these shared sequences.

Of the three testacea group species where *kl-5* now resides on an autosome, a Y-linked copy is retained only in *D. orientacea* (fig. 1). If the Y-to-autosome translocation occurred once in the ancestor of all three species, we would expect the Y-linked copy of *D. orientacea* to be basal to a monophyletic clade comprising all of the autosomal copies of *kl-5*. Instead, the phylogenetic analysis places the two *D. orientacea* copies together (fig. 1 and supplementary figure S2, Supplementary Material online). As discussed earlier, taking this analysis at face value, the phylogenetic tree is consistent with two parsimonious reconstructions, both involving two translocations (supplementary figure S3, Supplementary Material online). However, the phylogeny of *kl-5* is also consistent with a third explanation: a single translocation of *kl-5* from the Y to the dot in the common ancestor of *D. neotestacea*, *D. testacea*, and *D. orientacea*, and then subsequent genetic homogenization of the dot and Y copies of *kl-5* in *D. orientacea*. This scenario requires only one translocation off the Y and on to the dot chromosome, with ectopic recombination between the Y and the dot causing the two *D. orientacea* copies to cluster phylogenetically (fig. 1). Such ectopic recombination would have to occur between different chromosomes, which might be expected to occur rarely, but recent work shows that there is a low but appreciable rate of gene conversion between paralogs residing on different Müller's elements (Casola et al. 2010). In addition, the ectopic recombination would also have to occur in males (or in rare XXY females), and male recombination appears to be rare in *Drosophila*: only one subgroup, not closely related to the testacea group, is known to normally experience it (Matsuda et al. 1983). Any genetic exchange that does occur in males is insufficient to prevent the degeneration of neo-Y chromosomes (e.g., Bachtrog et al. 2008). However, some evidence suggests that a low rate of exchange is possible in males between X and Y chromosomes and among Y-linked duplicates (Williams et al. 1989; Kopp et al. 2006; Krsticevic et al. 2010), perhaps due to gene conversion during premeiotic cell divisions of spermatids (Hiraizumi et al. 1973). Moreover, only a very low rate of recombination—in fact, only a single gene conversion event since the split of *D. orientacea* from the common ancestor of the testacea group—is needed to explain the phylogenetic clustering of Y-linked and autosomal *kl-5* sequences. Further work may be able to disentangle the role of translocation versus gene conversion: If the genes flanking the *kl-5* translocation are syntenic between *D. testacea* and *D. orientacea*, this result would argue for a single translocation and subsequent homogenization, rather than multiple translocations. In any case, our current data add to previous evidence that the Y and dot chromosome exchange genetic material with each other more often than with other parts of the genome.

Strikingly, *kl-5* shows a 5-fold increase in the rate of protein evolution (ω) in the testacea group versus in other species ($\omega = 0.151$ in the testacea group versus $\omega = 0.028$ on the rest of the phylogeny; table 2 and fig. 4). As found for many duplicated genes, the high rate of protein substitution could be a simple consequence of relaxed purifying selection due to functional redundancy, if there were still a copy of *kl-5* present on the Y. But this is only likely in *D. orientacea*, which expresses both dot and Y-linked copies of *kl-5*, and does not explain the elevated rate of protein evolution throughout the testacea group. Substituting the dot chromosome from *D. orientacea* into the genome of *D. testacea* does not decrease male fertility, suggesting that any relaxed selection in the *D. orientacea* lineage has not been so substantial to allow the fixation of severely deleterious mutations. Furthermore, because *kl-5* is a highly conserved and essential male fertility factor, in no species is purifying selection likely to be entirely relaxed; rather, the high rate of protein substitution might reflect a gradual deterioration of *kl-5* function. Our data show that *kl-5* expression currently is not limited to the male reproductive tract, although the functional consequences of this, if any, are unknown.

Alternatively, the shift in the rate of protein evolution may be due to the new genomic environment. That is, if the dot is a better genomic environment than the Y, adaptive evolution may now be able to reverse previously fixed deleterious mutations or to fix new adaptive variants. It is not obvious, though, that the dot is a better environment, as a reduced effective population size and relaxed purifying selection appear to be the rule for dot-linked genes (Berry et al. 1991; Jensen et al. 2002; Wang et al. 2002, 2004; Betancourt et al. 2009; Haddrill et al. 2007;

Arguello et al. 2010; Leung et al. 2010). For the elevated protein substitution rate to be due to relaxed purifying selection, though, this requires that the dot would have an even lower effective size than the Y chromosome. The dot chromosome has a larger census population size as there is only one Y for every four copies of the dot. However, the effective size can be greatly reduced by linkage effects due to selection, and these linkage effects may well be stronger for the dot. For one, dot chromosome loci experience selection in both males and females, whereas Y-linked loci necessarily experience selection only in males. In addition, the dot may have more sites that are targets of purifying or positive selection, and the degenerative effects of linkage on the Y diminish as the number of selective targets declines (Bachtrog 2008; Kaiser and Charlesworth 2009). In contrast to the 13–20 genes on the *D. melanogaster* Y (Carvalho et al. 2009), the dot in *D. melanogaster* has approximately 80 genes, and we roughly estimate that it has 2.4- to 4.3-fold more nonsynonymous sites than the Y chromosome (dot: up to 195,000 sites/homolog; Y: up to 45,000–90,000 sites, given 13–20 genes and adjusting for up to 20% of sequence missing in the known Y-chromosome genes).

On balance, the data from this study suggest the dot is the worse environment. There is little suggestion that the rapid protein evolution at *kl-5* is due to positive selection—the pattern of genetic variation is inconsistent with a recent selective sweep, and MK tests of *kl-5* are not significant for any species (supplementary table S7, Supplementary Material online). Higher levels of amino acid polymorphisms are also consistent with weak purifying selection, and *kl-5* on the dot appears to suffer from these as well. Estimates of α are generally negative, and, equivalently, neutrality indices are larger than 1 (supplementary table S7, Supplementary Material online), though the confidence intervals are broad. Furthermore, in species where *kl-5* is dot linked, it suffers a higher relative level of amino acid polymorphism (π_A/π_S) compared with other loci in the genome. In comparison, in *D. putrida* where *kl-5* remains on the Y chromosome, π_A/π_S is similar to that of other loci.

In conclusion, we have shown that *kl-5* has moved from the Y chromosome to the tiny nonrecombining dot chromosome in the common ancestor of the three closely related testacea group species. The *kl-5* gene has thus moved at least three times, possibly four, during the evolutionary history of the *Drosophila* genus (Koerich et al. 2008; present study). The movement of genes may play an important role in causing inviability and sterility in species hybrids, particularly in the case of male fertility factors (Masly et al. 2006), and in genome evolution in general, and in the evolution of sex chromosomes in particular.

Supplementary Material

Tables S1–S7 and figures S1–S5 are available at *Molecular Biology and Evolution*online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to J. Baines, J. Jaenike, and M. Kimura for assistance with fly collections, M. Gaertig and D. Smith for assistance in the laboratory, and B. and D. Charlesworth, D. Hall, and members of the Charlesworth lab group for useful discussion. This work was supported by Royal Society USA Research Fellowships to K.A.D. and A.J.B., a research scholarship from the Peach State Louis Stokes Alliances for Minority Participation program to D.G.P., and funds from the Ellison Medical Foundation and the University of Georgia Research Foundation to K.A.D.

References

Arguello J, Zhang Y, Kado T, Fan C, Zhao R, Innan H, Wang W, Long M. 2010. Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol Biol Evol.* 27:848–861.

Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* 179:1513–1525.

Bachtrog D, Hom E, Wong K, Maside X, De Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol.* 9:R30.

Bayes J, Malik H. 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* 326: 1538–1541.

Berry A, Ajioka J, Kreitman M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129:1111–1117.

Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by a lack of recombination. *Curr Biol.* 19:655–660.

Bull J. 1983. Evolution of sex determining mechanisms. Menlo Park (CA): Benjamin/Cummings. p. 316.

Carvalho A, Clark A. 2005. Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. *Science* 307: 108–110.

Carvalho A, Dobo B, Vibranovski M, Clark A. 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 98:13225–13230.

Carvalho A, Lazzaro B, Clark A. 2000. Y chromosomal fertility factors *kl-2* and *kl-3* of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proc Natl Acad Sci U S A.* 97:13239–13244.

Carvalho BA, Koerich L, Clark A. 2009. Origin and evolution of Y chromosomes: *Drosophila* tales. *Trends Genet.* 25:270–277.

Casola C, Cl G, Mw H. 2010. Nonallelic gene conversion in the genus *Drosophila*. *Genetics* 185:95–103.

Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A.* 75:5618–5622.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 355:1563–1572.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95: 118–128.

Chen H, Sidorenko VS, Watabe H. 1998. Sympatric distribution of two sibling species of *Drosophila, D. orientacea* and *D. testacea*. *Entomolog Sci.* 1:215–217.

Clark A, Eisen M, Smith D, Bergman C, Oliver B, Markow T, Kaufman T, Kellis M, Gelbart W, Iyer V. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.

Cooper KW. 1964. Meiotic conjunctive elements not involving chiasmata. *Genetics* 52:1248–1255.

Dyer KA, Jaenike J. 2004. Evolutionary stable infection by a male-killing endosymbiont in *Drosophila innubila*: molecular

evidence from the host and parasite genomes. *Genetics* 168: 1443–1455.
Dyer KA, Charlesworth B, Jaenike J. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc Natl Acad Sci U S A.* 104:1587–1592.
Fay J, Wu C. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.
Fiston-Lavier A, Anxolabehere D, Quesneville H. 2007. A model of segmental duplication formation in *Drosophila melanogaster*. *Genome Res* 17:1458–1470.
Gordo I, Charlesworth B. 2000. The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics* 154: 1379–1387.
Graves J. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124:901–914.
Grimaldi D, James AC, Jaenike J. 1992. Systematics and modes of reproductive isolation in the holarctic *Drosophila testacea* species group (Diptera: Drosophilidae). *Ann Entomolog Soc Am.* 85:671–685.
Haddrill P, Halligan D, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol.* 8:R18.
Hartl T, Sweeney S, Knepler P, Bosco G. 2008. Condensin II resolves chromosomal associations to enable anaphase I segregation in *Drosophila* male meiosis. *PLoS Genet.* 4:e1000228.
Hiraizumi Y, Slatko B, Langley C, Nill A. 1973. Recombination in *Drosophila melanogaster* male. *Genetics* 73:439–444.
Holloway A, Lawniczak M, Mezey J, Begun D, Jones C. 2007. Adaptive gene expression divergence inferred from population genomics. *PLoS Genet.* 3:2007–2013.
Hudson R, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
Hughes JF, Skaletsky H, Pyntikova T, et al. (17 co-authors). 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463:536–539.
Hughes S, Gilliland W, Cotitta J, Takeo S, Collins K, Hawley R. 2009. Heterochromatic threads connect oscillating chromosomes during prometaphase I in *Drosophila* oocytes. *PLoS Genet.* 5: e1000348.
James A, Jaenike J. 1990. "Sex Ratio" meiotic drive in Drosophila testacea. *Genetics* 126:651–656.
Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170:1401–1410.
Jensen M, Charlesworth B, Kreitman M. 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* 160:493–507.
Johansson A, Stenberg P, Bernhardsson C, Larsson J. 2007. Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophila melanogaster*. *EMBO J.* 26: 2307–2316.
Kaiser V, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* 25: 9–12.
Koerich L, Wang X, Clark A, Carvalho A. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature* 456: 949–951.
Kopp A, Frank AK, Barmina O. 2006. Interspecific divergence, intrachromosomal recombination, and phylogenetic utility of Y-chromosomal genes in *Drosophila*. *Mol Phylogenet Evol.* 38: 731–741.
Krsticevic F, Santos H, Januario S, Schrago C, Carvalho A. 2010. Functional copies of the Mst77F gene on the Y chromosome of *Drosophila melanogaster*. *Genetics* 184:295–307.
Kurek R, Reugels A, Lammermann U, Bünemann H. 2000. Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of *Drosophila* sp. *Genetica* 109:113–123.
Lanave C, Preparata G, Sacone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol.* 20:86–93.
Larracuente A, Noor M, Clark A. 2010. Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol.* 27:1612–1620.
Leung WD, Shaffer CD, Cordonnier J, Wong T, Itano MS, Slawson Tempel EE, et al. 2010. Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. *Genetics* 185: 1519–1534.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
Lohe A, Hilliker A, Roberts P. 1993. Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. *Genetics* 134:1149–1174.
Marais GAB, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Monéger F, Hobza R, Widmer A, Charlesworth D. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr Biol.* 18:545–549.
Masly J, Jones C, Noor M, Locke J, Orr H. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313: 1448–1450.
Matsuda M, Imai HT, Tobari YN. 1983. Cytogenetic analysis of recombination in males of*Drosophila ananassae*. *Chromosoma* 88:286–292.
McDonald J, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
Ming R, Moore P. 2007. Genomics of sex chromosomes. *Curr Opin Plant Biol.* 10:123–130.
Muller H. 1964. The relation of recombination to mutational advance. *Mutat Res.* 1:2–9.
Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre. Uppsala (Sweden): Uppsala University.
Orr HA. 1992. Mapping and characterization of a 'speciation gene' in *Drosophila*. *Genet Res.* 59:73–80.
Orr HA, Kim Y. 1998. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* 150:1693–1698.
Patterson JT, Stone WS. 1952. Evolution in the genus *Drosophila*. New York: The Macmillan Company. p. 618.
Peck J. 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* 137:597–606.
Perlman S, Spicer G, Shoemaker D, Jaenike J. 2003. Associations between mycophagous *Drosophila* and their *Howardula* nematode parasites: a worldwide phylogenetic shuffle. *Mol Ecol.* 12:237–249.
Presgraves D. 2005. Recombination enhances protein adaptation in*Drosophila melanogaster*. *Curr Biol.* 15:1651–1656.
Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189.
Rambaut A. 1996. Se-Al: sequence Alignment Editor [cited 2010 Dec]. Available from http://evolve.zoo.ox.ac.uk/
Reugels A, Kurek R, Lammermann U, Bunemann H. 2000. Mega-introns in the dynein gene DhDhc7(Y) on the heterochromatic Y chromosome give rise to the giant threads loops in primary spermatocytes of *Drosophila hydei*. *Genetics* 154: 759–769.
Riddle N, Elgin S. 2006. The dot chromosome of Drosophila: insights into chromatin states and their change over evolutionary time. *Chromosome Research* 14:405–416.
Riddle N, Shaffer C, Elgin S. 2009. A lot about a little dot—lessons learned from *Drosophila melanogaster* chromosome 4. *Biochem Cell Biol.* 87:229–241.

Ronquist F, Huelsenbeck J. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Sandler L, Novitski E. 1956. Evidence for genetic homology between chromosomes I and IV in *Drosophila melanogaster*, with a proposed explanation for the crowding effect in triploids. *Genetics* 41:189–193.

Sheldahl L, Weinreich D, Rand D. 2003. Recombination, dominance and selection on amino acid polymorphism in the *Drosophila* genome: contrasting patterns on the X and fourth chromosomes. *Genetics* 165:1195–1208.

Skaletsky H, Kuroda-Kawaguchi T, Minx P, Cordum H, Hillier L, Brown L, Repping S, Pyntikova T, Ali J, Bieri T. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.

Smith N, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.

Steinemann M, Steinemann S. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica* 102:409–420.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci.* 17:57–86.

Wakeley J, Aliacar N. 2001. Gene genealogies in a metapopulation. *Genetics* 159:893–905.

Wang W, Thornton K, Berry A, Long M. 2002. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* 295:134–137.

Wang W, Thornton K, Emerson J, Long M. 2004. Nucleotide variation and recombination along the fourth chromosome in *Drosophila simulans*. *Genetics* 166:1783–1794.

Welch J. 2006. Estimating the genomewide rate of adaptive protein evolution in Drosophila. *Genetics* 173:821–837.

Williams S, Kennison J, Robbins L, Strobeck C. 1989. Reciprocal recombination and the evolution of the ribosomal gene family of Drosophila melanogaster. *Genetics* 122:617–624.

Wright S, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol.* 39:306–314.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zurovcova M, Eanes WF. 1999. Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene Dhc-Yh3 of *Drosophila melanogaster* and *D simulans*. *Genetics* 153:1709–1715.