Recent Selection on the Y-to-Dot Translocation in *Drosophila pseudoobscura*

Amanda M. Larracuente*[1,2] and Andrew G. Clark[1]
[1]Molecular Biology and Genetics, Cornell University
[2]Department of Biology, University of Rochester
*Corresponding author: E-mail: alarracu@bio.rochester.edu.
Associate editor: Naoko Takezaki

Abstract

The *Drosophila pseudoobscura* dot chromosome acquired genes from the ancestral *Drosophila* Y chromosome in a Y-to-dot translocation event that occurred between 12.7 and 20.8 Ma. The formerly Y-linked genes mostly retained their testis-specific expression but shrank drastically in size, mostly through intron reduction, since becoming part of the dot chromosome in this species. We investigated the impact of this translocation on the evolution of the both the Y-to-dot translocated region and the original segments of the dot chromosome in *D. pseudoobscura*. Our survey of polymorphism and divergence across the chromosome reveals a reduction in variation, a deletion polymorphism segregating at high frequency, and a shift in the frequency spectra, all consistent with a history of recent selective sweeps in the Y-to-dot translocated region but not on the rest of the dot chromosome. We do find evidence for recombination primarily as gene conversion on the dot chromosome; however, predicted recombination events are restricted to the part of the dot chromosome outside the translocation. It therefore appears that recombination has resulted in a degree of decoupling between the ancestral Y region and the conserved region of the dot chromosome.

Key words: dot chromosome, positive selection, Y chromosome.

Introduction

Drosophila dot chromosomes are exceptional in many features, including their small size, low gene density, and large heterochromatin proportion. Named for its small size, the dot chromosome is nonrecombining via classical crossing over (Bridges 1935; Ashburner 1989); 80% of the dot chromosome is constitutively heterochromatic. The remainder of the chromosome contains a conserved complement of genes with euchromatic gene density (Sun et al. 2000), and their regulation appears uniquely controlled by chromosome-specific and heterochromatin-associated proteins (Larsson et al. 2001; Larsson and Meller 2006; Riddle and Elgin 2006). Recent evidence shows convincingly that the dot chromosome was once an ancient X chromosome (Vicoso and Bachtrog 2013). Thus, it is not surprising that the dot chromosome more closely resembles an X chromosome than an ordinary autosome in many ways (Larsson and Meller 2006; Riddle and Elgin 2006; Vicoso and Bachtrog 2013).

In *Drosophila pseudoobscura*, the dot chromosome is even more exceptional, as it has acquired a major translocation from the ancestral *Drosophila* Y chromosome (Carvalho and Clark 2005; Larracuente et al. 2010). The current Y chromosome of *D. pseudoobscura* is not homologous to the *D. melanogaster* Y chromosome (Carvalho and Clark 2005) and instead may be a degenerated neo-Y chromosome that originated in an X-autosome fusion event in an ancestor of *D. pseudoobscura* (the X fused with an arm of an autosome referred to as the "Muller D"; Carvalho and Clark 2005; Larracuente et al. 2010). The translocation of the Y chromosome to the dot chromosome occurred between 12.7 and

20.8 Ma: both the X-A fusion and Y-A translocation are found in *D. affinis* and *D. azteca* (clade split from *D. pseudoobscura* ancestor between 12.7 and 14.9 Ma; Gao et al. 2007) but are not found in the *obscura* group species *D. bifasciata* and *D. guanche* (Carvalho and Clark 2005; the *D. bifasciata* and *D. pseudoobscura* lineages diverged between 17.8 and 20.8 Ma; Gao et al. 2007).

The Y-to-dot translocation was accompanied by a drastic reduction in gene size: *Drosophila* Y-linked genes can reach megabases in size (Gatti and Pimpinelli 1983; Kurek et al. 2000), but in *D. pseudoobscura* these genes shrank at least 10-fold (to the kilobase range) after translocating to the dot chromosome (Carvalho and Clark 2005; Larracuente et al. 2010). Most Y-to-dot translocated genes retained their testis-specific expression after moving to the dot chromosome (Carvalho and Clark 2005). Whereas the large introns of the Y chromosome may best be explained by the reduced efficacy of natural selection owing to the lack of meiotic recombination on the Y chromosome (reviewed in Hill and Robertson 1966; Felsenstein 1974; Charlesworth and Charlesworth 2000); once dot-linked, the formerly Y-linked genes may enjoy the benefits of a larger effective population size (N_e), and perhaps even recombination (via gene conversion).

In regions of low recombination like the dot chromosome, selection at one site will interfere with selection at linked sites (Hill–Robertson Interference; Hill and Robertson 1966; Felsenstein 1974). Because recombination is hypothesized to be low or nonexistent on the *D. pseudoobscura* dot chromosome (Larracuente et al. 2010), as it is in *D. melanogaster*

and *D. simulans* (Ashburner 1989; Jensen et al. 2002; Wang et al. 2002; Wang et al. 2004; Arguello et al. 2010), the Y-to-dot translocation would be expected to affect levels of neutral variability across the entire dot chromosome. Patterns of nucleotide variation on the dot chromosomes of *melanogaster* subgroup species are highly variable (Wang et al. 2002; Wang et al. 2004; Arguello et al. 2010) and are consistent with background selection (Arguello et al. 2010; see Wang et al 2004 for a possible case of balancing selection in *D. melanogaster*). In contrast, one dot-linked locus surveyed in *D. pseudoobscura* has low levels of diversity and a skewed frequency spectrum in *D. pseudoobscura*, consistent with the effects of recent selection: nucleotide polymorphism and divergence at the *eyeless* locus revealed recent introgression between *D. pseudoobscura* and *D. persimilis* and the possibility of a trans-species selective sweep (Machado and Hey 2003).

To determine whether adaptive evolution has shaped the Y-to-dot translocation and has had an impact on dot chromosome evolution in *D. pseudoobscura*, we studied patterns of polymorphism and divergence across the dot chromosome. Our results are consistent with a history of recurrent selective sweeps in the Y-to-dot region, possibly favoring reductions in size of the translocated region over time. We do not detect the signature of recurrent selective sweeps outside of the Y-to-dot region, however, suggesting that recombination has decoupled the evolution of the former Y and conserved regions of the dot chromosome.

Results

Reduced Diversity on the Dot

We surveyed polymorphism and divergence at 20 dot-linked loci spanning the dot chromosome in 64 lines of *D. pseudoobscura* from nine different geographic locations. Eleven regions are in the Y-to-dot translocation and nine are on the rest of the dot chromosome (table 1).

Average overall θ_w (population mutation rate per nucleotide site) and π (pairwise measure of diversity per nucleotide site) for the dot chromosome are 0.00096 and 0.00038, respectively, which is significantly lower than variation reported on the autosomes ($P_\theta = 5.14 \times 10^{-5}$, $P_\pi = 6.0 \times 10^{-11}$, Mann–Whitney U test, MWU) and X chromosome of *D. pseudoobscura* ($P_\theta = 4.54 \times 10^{-8}$, $P_\pi = 2.48 \times 10^{-9}$, MWU). Several loci had no variation, both in the Y-to-dot translocated region (*kl-3_1*, *kl-3_3*, *Ppr-Y_2*; table 1) and the part of the dot not associated with the translocation (conserved dot; GA10734; table 1). Interestingly, the most variable

Table 1. Summary Statistics for the 20 Loci Surveyed from the Dot Chromosome.

Fragment	Region	Alignment[a] Length	n[b]	S[c]	θ_w All (Silent[d])	π All (Silent[d,e])	D_{Taj}[e]
Y-to-dot translocation							
kl-3_1	Y-to-dot	583	53	0	0 (0)	0 (0*[#†])	NA
kl-3_2	Y-to-dot	684	35	3	0.00166 (0.00438)	0.00041 (0.00107*[#†])	−1.484*
kl-3_3	Y-to-dot	384	63	0	0 (0)	0 (0*[#†])	NA
ARY_1	Y-to-dot	289	54	1	0.00076 (0.00339)	0.00037 (0.00165*[#†])	−0.675
kl-2_1	Y-to-dot	626	46	1	0.00036 (0.0017)	0.00014 (0.00063*[#†])	−0.860
kl-2_2	Y-to-dot	736	58	1	0.00029 (0)	0.00005 (0*[#†])	−1.088*
ORY_1	Y-to-dot	610	56	2	0.00085 (0.00064)	0.00039 (0.00011*[#†])	−0.934
ORY_2	Y-to-dot	546	63	1	0.00039 (0.00126)	0.00037 (0.00119*[#†])	−0.0687
ORY_3	Y-to-dot	458	61	7	0.00334 (0.00446)	0.00058 (0.00078*[#†])	−2.104*[#]
Ppr-Y_1	Y-to-dot	505	61	5	0.00265 (0.00242)	0.00041 (0.00037*[#†])	−1.971*[#]
Ppr-Y_2	Y-to-dot	477	61	0	0 (0)	0 (0*[#†])	NA
Concatenated[f]	Y-to-dot	5898	64	21	(0.0019)	(0.00045)	−2.332
Conserved region of the dot							
GA27948	Conserved dot	578	61	2	0.00074 (0.00093)	0.00042 (0.00014*[#†])	−0.749
GA10714	Conserved dot	450	61	2	0.00095 (0.00459)	0.00015 (0.00070*[#†])	−1.442*
GA10734	Conserved dot	498	56	0	0 (0)	0 (0*[#†])	NA
GA14409	Conserved dot	540	50	4	0.00178 (0.00238)	0.00032 (0.00043*[#†])	−1.862*
GA14323	Conserved dot	521	50	1	0.00043 (0.00075)	0.00008 (0.00013*[#†])	−1.103
GA15199	Conserved dot	475	60	1	0.00046 (0)	0.00014 (0*[#†])	−0.891
GA15170	Conserved dot	518	61	3	0.00135 (0.00358)	0.00145 (0.00155*[#†])	0.153
GA13377	Conserved dot	569	61	2	0.00075 (0.0013)	0.00012 (0.0002*[#†])	−1.442*
EY	Conserved dot	546	58	6	0.00239 (0.00242)	0.00221 (0.00224*[#†])	−0.192

[a]Alignment length is the region considered for the polymorphism analysis.

[b]*n* is the sample size for each locus.

[c]*S* is the number of segregating sites.

[d]The silent sites considered are both noncoding and synonymous sites.

[e]The significance of π and Tajima's D are indicated with a * for a significant value ($P < 0.05$) under the standard neutral model with constant population size and [#] for a significant value under the model of population expansion described in the Results. † indicates that this value has a FDR of <5%.

[f]Summary statistics for the concatenated Y-to-dot loci (excluding *kl-3_2*) are reported. θ_h (measure of diversity per silent site that depends on high frequency variants) estimated from the concatenated data set for selective sweep simulations was 1.65×10^{-5}.

fragment we surveyed in the Y-to-dot translocated region (ORY_3 $\theta_w = 0.00446$; table 1) had 97 and 69 bp deletion polymorphisms segregating in our sample at frequencies of 35.6% and 3.39%, respectively. The most variable of the dot chromosome loci outside the translocation (conserved dot) was ey ($\theta_w = 0.00239$), which is nearly an order of magnitude higher than previous estimates at this locus (Machado and Hey 2003). The Y-to-dot translocation does not significantly differ from the rest of the dot chromosome in levels of variation ($P_\theta = 0.8300$, $P_\pi = 0.8173$, MWU; table 2).

The dot chromosome has significantly less diversity than the autosomes overall ($P_\theta = 1.061 \times 10^{-6}$, $P_\pi = 2.524 \times 10^{-8}$, MWU; table 2), including when comparing genes just in the Y-to-dot translocation ($P_\theta = 1.043 \times 10^{-5}$, $P_\pi = 3.4 \times 10^{-6}$, MWU; table 2) and the non-Y parts of the dot ($P_\theta = 3.127 \times 10^{-5}$, $P_\pi = 1.9 \times 10^{-5}$, MWU; table 2). The dot also has significantly less diversity than the X chromosome ($P_\theta = 4.844 \times 10^{-5}$, $P_\pi = 3.1 \times 10^{-9}$, table 2). These patterns are mimicked at silent sites: the Y-to-dot translocation and the rest of the dot have similar levels of silent variation ($P_\theta = 0.7289$, $P_\pi = 0.8173$, MWU; table 2 and fig. 1). Silent site diversity on the dot chromosome is significantly less than the autosomes ($P_\theta = 1.061 \times 10^{-7}$, $P_\pi = 2.524 \times 10^{-8}$, MWU; table 2 and fig. 1) and the X chromosome ($P_\theta = 4.84 \times 10^{-5}$, $P_\pi = 3.1 \times 10^{-9}$, MWU; table 2 and fig. 1).

The level of silent polymorphism surveyed on the dot chromosome is lower than expected based on reported levels of silent autosomal polymorphism under a neutral coalescent model of constant population size ($P = 0.0002$). Silent site divergence between $D.\ pseudoobscura$ and $D.\ miranda$ is not significantly different between the dot chromosome and the autosomes ($P = 0.5485$, MWU) or between the dot chromosome and the X chromosome ($P = 0.8694$, MWU; supplementary table S1, Supplementary Material online, and fig. 1), indicating that the low of polymorphism on the dot chromosome is not likely due to a lower neutral mutation rate.

Evidence for Gene Conversion

The $D.\ melanogaster$ dot chromosome does not undergo meiotic crossover events (Ashburner 1989). However, because population genetic analyses in $D.\ melanogaster$ and $D.\ simulans$ have documented historical recombination events,

several authors have concluded that the dot chromosome shows evidence for very low levels of recombination by crossing over and gene conversion (Jensen et al. 2002; Wang et al. 2002; Wang et al. 2004; Arguello et al. 2010). To determine whether the $D.\ pseudoobscura$ dot chromosome has experienced recombination, we calculated the minimum number of recombination events as R_m (Hudson and Kaplan 1985) and R_h (Myers and Griffiths 2003). We find evidence for low levels of recombination across the $D.\ pseudoobscura$ dot chromosome: the R_m density is 0.0031/kb/chromosome (2 events per 10 kb in 64 chromosomes) with a lower bound of 1.918×10^{-5} (2 events per 1,630 kb in 64 chromosomes) and the R_h density is 0.0194 (7 events per 10 kb in 64 chromosomes) with a lower bound of 6.71×10^{-5} (7 events per 1,630 kb in 64 chromosomes). While the R_m density in this sample is lower than the dot chromosomes of $D.\ simulans$ (0.0451 R_m/kb/chromosome, Wang et al. 2004, and a lower bound of 0.024 R_m/kb/chromosome, Arguello et al. 2010) and $D.\ melanogaster$ (0.005 R_m/kb/chromosome, Wang et al. 2002, and a lower bound of 0.010 R_m/kb/chromosome, Arguello et al. 2010), our sample size differs from that of other studies, therefore it is difficult to directly compare these estimates of recombination with other regions of the genome and other species.

Consistent with low rates of recombination, we find a high level of linkage disequilibrium (LD): 57.6% pairs showing very high LD (449/780 pairs with $0.9 < |D'| < 1$). Despite this high LD, only 16% (125/780) of pairwise comparisons between SNPs rejected the null hypothesis of independence. Consistent with recombination events breaking down LD over long distances, the product moment correlation between distance between SNPs and r^2 (−0.1186) is statistically significant ($P = 0.0009$; fig. 2). This break down of LD over long distances is driven by the conserved region of the dot chromosome: the correlation between distance between SNPs (−0.2658) and r^2 is only significantly negative for the conserved region of the dot ($P = 0.0018$; fig. 2) and not for the Y-to-dot translocation ($P = 0.3066$; fig. 2). These relationships hold under a nonlinear model where LD decays exponentially with distance across the dot chromosome ($P = 0.0082$). Again, this pattern is driven by the conserved region of the dot ($P = 0.00853$) and not the Y-to-dot translocation ($P = 0.351$).

**Table 2.** Mean and Median Diversity Estimates θ_w and π for Silent Sites.

Location	n^a	θ_w Mean (Median)	π Mean (Median)	References
Autosomes[b]	56	0.0147 (0.0128)	0.0134 (0.0102)	Hamblin and Aquadro 1999; Schaeffer et al. 2001; Machado et al. 2002; Schaeffer et al. 2003; Haddrill et al. 2010
X	8	0.0228 (0.0179)	0.0154 (0.013)	Kovacevic and Schaeffer 2000; Machado et al. 2002; Haddrill et al. 2010
Dot	20	0.00171 (0.00128)	0.00056 (0.00029)	This study
Y-to-dot	11	0.00166 (0.00126)	0.00053 (0.00037)	This study
Conserved dot	9	0.00177 (0.0013)	0.00060 (0.00020)	This study

[a]n is the number of loci.
[b]Excludes the dot chromosome.
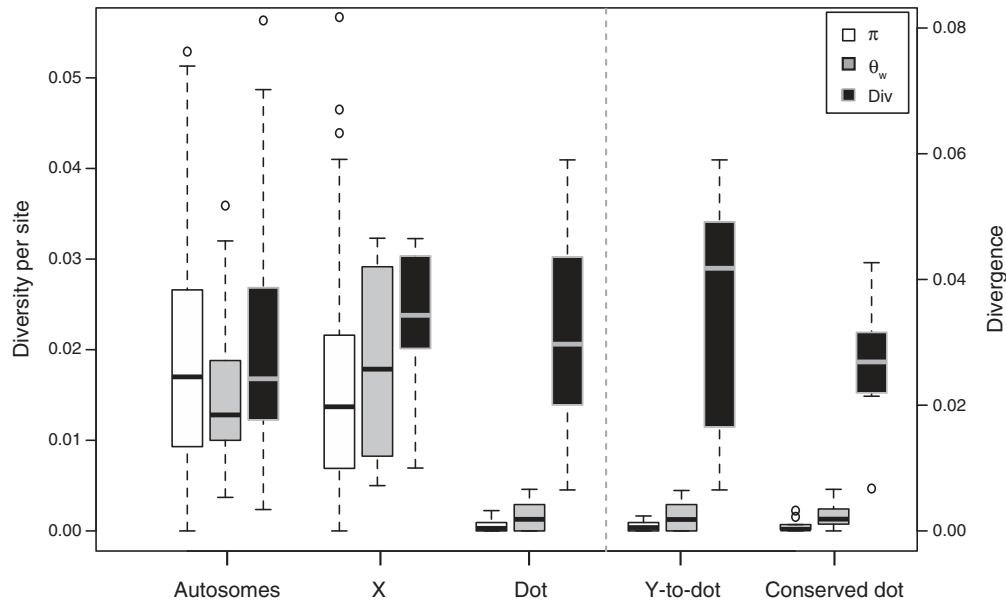
FIG. 1. Reduced silent site diversity on the *Drosophila pseudoobscura* dot chromosome. Shown are the medians and interquartile ranges of π, θ_w, and divergence between *D. pseudoobscura* and *D. miranda* for the autosomes, the X chromosome, and the dot chromosome (outliers represented by open circles). The dot is further partitioned into the Y-to-dot translocated region and the conserved region of the dot, not involved in the translocation.
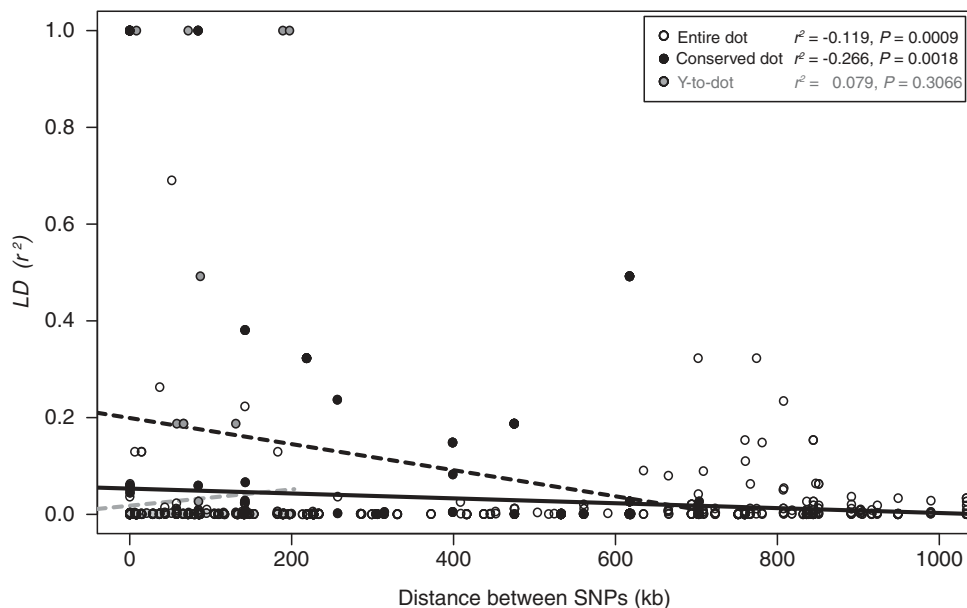
FIG. 2. The relationship between LD (measured as r^2) and distance between SNPs. The distance between SNPs and r^2 for the entire dot chromosome (open black circles, solid black line), just the conserved region of the dot chromosome (filled black circles, dotted black line), and just the Y-to-dot translocation (filled gray circles, dotted gray line). The negative Pearson's product moment coefficient ($r^2 = -0.119$) is statistically significant for the entire dot chromosome ($P = 0.0009$), indicating that this chromosome has experienced recombination. This pattern appears to be driven by the conserved region of the dot chromosome ($r^2 = -0.266$, $P = 0.0018$).

We find both very long (between 43 and 130 kb; supplementary fig. S1, Supplementary Material online) and very short haplotype blocks across the dot chromosome (88–124 bp; supplementary fig. S1, Supplementary Material online), indicating that at least some of these recombination events were crossovers. All predicted recombination events occurred in the conserved part of the dot chromosome, outside of the translocation (supplementary fig. S1, Supplementary Material online).

To attribute recombination events to meiotic crossing over or gene conversion, we estimated the population level recombination parameter $\rho = 4Nr$ (where N is the effective population size and r is the rate of recombination per base pair per generation) and the relative rate of gene conversion, $f = g/\rho$ (where g is the probability of a gene conversion event per base pair). We used a composite likelihood approach to search over a grid of values of ρ and f, for several different conversion tract lengths. The maximum composite likelihood

for the different conversion tract lengths was at 150 bp, giving an estimate $\rho = 0.000092$ and $f = 170.2$ (table 3), indicating that there is support for a model where gene conversion events far outnumber crossover events. In all cases, the composite likelihood for the model assuming no gene conversion was lower than a model with both gene conversion and recombination for all tract lengths examined (table 3). Our estimates of ρ and f are similar to estimates on the dot chromosome of *D. simulans* (Wang et al. 2004) and slightly higher than estimates observed in *D. melanogaster* (Arguello et al. 2010). This is consistent with *D. pseudoobscura* having a larger effective population size (Schaeffer et al. 1987; Riley et al. 1989; Schaeffer and Miller 1992b) and overall higher rates of recombination (Hamblin and Aquadro 1999; Ortiz-Barrientos et al. 2006) than *D. melanogaster*. We estimate that the average ρ on the dot chromosome is three orders of magnitude lower than the X and autosomes; however, the estimates from the X and autosomes do not consider a model with gene conversion (Kovacevic and Schaeffer 2000; Machado et al. 2002).

Accounting for the Demographic History of *D. pseudoobscura*

For all but one of the loci sampled (GA15170), θ_w is greater than π and Tajima's D (D_{Taj}) is negative (table 1). Of the 16 loci where segregating sites were found, seven of these loci have a significantly negative D_{Taj}. Because we show no evidence for recombination within the Y-to-dot translocation and three loci have no variation, we concatenated the data set and treat the 11 loci as a single locus. This concatenated data set also has a significantly negative D_{Taj} ($D_{Taj} = -2.13$; $P = 0.001$) under a standard neutral model. Thus, it appears that the frequency spectra of loci on the dot chromosome are significantly skewed toward rare variants, with the greatest skew occurring in the Y-to-dot region.

An excess of rare variants segregating in a population is consistent with several evolutionary scenarios. While selective sweeps (Tajima 1989) and background selection can both lead to reduced levels of neutral variability (Charlesworth et al. 1995) and a skew in the frequency spectrum toward rare alleles (Kaiser and Charlesworth 2009), so can nonselective forces such as changes in population size and a history of rapid population expansion. The fact that the skew in the frequency spectra was seen across nearly all loci sampled from the X chromosome and autosomes in *D. pseudoobscura* suggests that this species underwent a history of population expansion (Hamblin and Aquadro 1999; Kovacevic and Schaeffer 2000; Machado et al. 2002; Schaeffer 2002; Schaeffer et al. 2003; Haddrill et al. 2010). We therefore simulated data under a model of simple exponential expansion similar to a model inferred by Haddrill et al. (2010), where the population size at a given time t is $N^{-\lambda t}$. We assumed that the population expanded to ten times its original size and set $\lambda = 20$. The frequency spectra under this model of population expansion (e.g., mean and standard error under two demographic models: $D_{Taj\text{-}expansion} = -0.7969 \pm 0.6971$; $D_{Taj\text{-}constant} = -0.0439 \pm 0.9379$) are more consistent with the empirical autosomal data (supplementary fig. S2,

Table 3. Composite Likelihood Analysis of Recombination Rate (ρ) and the Ratio of Gene Conversions to Crossovers (f).

| Mean Tract Length[a]	ρ	f	ln L
0 (no gene conversion)	0.000139	0	−3795.031585
50	0.000092	470.588235	−3785.959551
100	0.000092	242.803504	−3785.767135
150	0.000092	170.212766	−3785.709478
200	0.000092	132.665832	−3785.742768
250	0.000092	111.389237	−3785.807385
300	0.000092	106.382979	−3785.907192
350	0.000090	100.125156	−3786.022639
400	0.000090	87.609512	−3786.136024
450	0.000090	78.848561	−3786.242966
500	0.000090	71.339174	−3786.344305
550	0.000090	67.584481	−3786.442007
600	0.000090	61.326658	−3786.536484

[a]The model with the highest likelihood is underlined.

Supplementary Material online). Of the 65 non-dot autosomal loci (or 64 with $S > 0$) analyzed here, only one and three reject the expansion model at false discovery rate (FDR) <5% for π and D_{Taj}, respectively (fig. 3). In contrast, all 11 of the Y-to-dot loci and all 9 of the conserved dot loci reject the expansion model at FDR <5% for π, whereas three of the eight individual Y-to-dot and one of the eight conserved dot loci (where $S > 0$) reject the expansion model at an FDR <5% for D_{Taj} (fig. 3). Overall, the Y-to-dot translocation concatenated data, where we find no evidence for recombination, reject the expansion model for both π and D_{Taj} ($P_\pi = 8 \times 10^{-4}$; $P_{D_{Taj}} < 1 \times 10^{-5}$). Therefore, the plausible demographic model involving population expansion (inferred from autosomal data; Haddrill et al. 2010) cannot explain the reduction in variation and skew in the frequency spectrum that we observe in the Y-to-dot translocation or the reduction in variation across the dot chromosome. This suggests that variability across the dot chromosome has been affected by more than just the demographic history of the species and implicates the action of natural selection.

Evidence for Positive Selection

Although the Y-to-dot translocation consistently rejects neutrality including under a model that accounts for the demographic history of the species, there is no evidence for recurrent positive selection in the protein-coding regions of any of the dot-linked loci surveyed here. McDonald–Kreitman tests (McDonald and Kreitman 1991) for the equality of the ratio of silent to replacement polymorphism and the ratio of silent to replacement divergence between species were not significant for any gene region (data not shown). There is therefore no evidence that the protein-coding genes have evolved by recurrent selective sweeps, although with the low levels of recombination across the dot chromosome, it is possible that a single sweep in a protein-coding gene could produce the pattern we observe. However, a potentially more likely explanation is that positive selection on noncoding
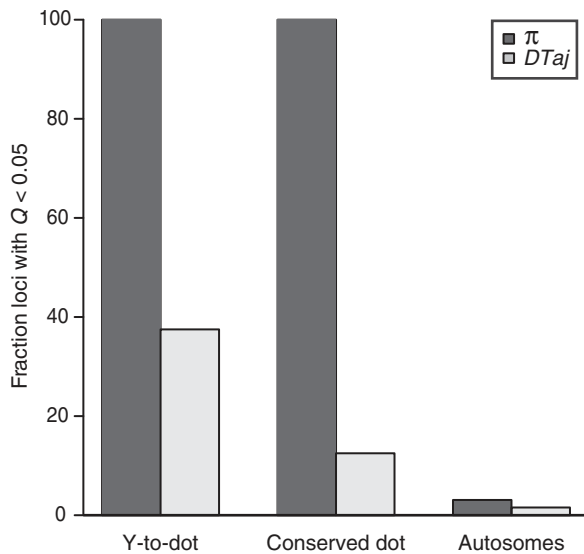
FIG. 3. Fit of the exponential growth model to Y-to-dot, dot (non-Y), and autosomal data. Barplot showing the fraction of loci (with $S > 0$) that reject the growth model (at FDR < 5%, or $Q < 0.05$) for π and D_{Taj}.

FIG. 4. The marginal posterior distribution for the time since the last selective sweep (t_{sweep}) in $4N_e$ generations. We infer the time to the most recent selective sweep in an expanding population (see Materials and Methods) for the Y-to-dot region as 0.0578 (95% CI t_{sweep} 0.0118–1.881) $4N_e$ generations ago.

regions of the translocation may instead drive the departure from neutrality. The Y-to-dot translocated region shrank drastically since moving to the dot chromosome. The reduction in size of the region may have been accompanied by recurrent positive selection favoring shorter introns and a loss of repetitive DNA in intergenic regions. To test this hypothesis, we simulated selective sweeps in a population that is expanding exponentially, according to the demographic model that is consistent with the autosomal data (described above). These simulations assume that there is no recombination at the sampled locus. Based on the results of our recombination analysis (discussed earlier), we focus on the concatenated Y-to-dot translocation data set. Using the Approximate Bayesian Computation (ABC) method and rejection sampling conditional on summaries of the Y-to-dot frequency spectrum (see Materials and Methods), we estimate the time since the last selective sweep to be 0.0578 (95% CI t_{sweep} 0.0118–1.881) $\times 4N_e$ generations (fig. 4). Assuming an N_e of 1.8 million (Haddrill et al. 2010) and five generations per year, the last sweep occurred approximately 83,200 (95% CI 16,920–2,708,640) years ago. We do not find evidence for a recent selective sweep using the concatenated data from the conserved region of the dot or the entire dot data set (supplementary fig. S3, Supplementary Material online).

One intron within the Y-to-dot translocation (*ORY* intron 3) was polymorphic for a 90-bp deletion, a 69-bp deletion, a 1-bp deletion, and an 8-bp insertion. Of the 56 lines without missing data at this locus, 21 lines have the 90-bp deletion and 2 additional lines have the 69-bp deletion, which is nested within the boundaries of the former deletion (supplementary fig. S4, Supplementary Material online). Moreover, *D. pseudoobscura* is fixed for at least six indels in this region compared with *D. miranda*. The polymorphic deletions appear in all of our sampled populations—that span much of the range of *D. pseudoobscura*—ruling out population structure as a
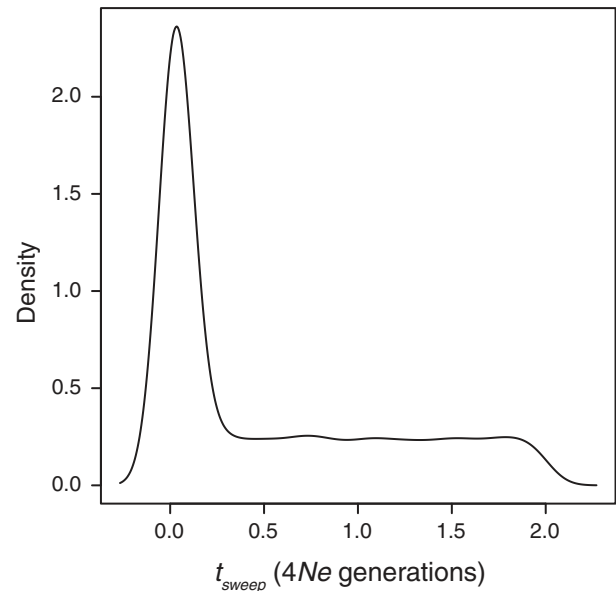
cause for the haplotype configuration. We performed tests of neutrality based on haplotype frequencies in this intron assuming an demographic model of population expansion consistent with *D. pseudoobscura* autosomal data (see Materials and Methods; Innan et al. 2005). The haplotype configuration in this intron is unlikely ($P = 0.0001$) in an expanding population given $n = 56$, $S = 5$ (assuming an average autosomal $\theta = 0.015$; we obtained the same results for $\theta = 0.002$). If positive selection is acting on this locus, the large 90-bp deletion—which is the most frequent haplotype—would be at a higher frequency than expected under neutrality. This appears to be the case: the frequency of the most frequent haplotype ($M = 21/56$) is unexpectedly high ($P = 0.018$) under neutrality, in an expanding population. The haplotype diversity for the locus is unexpectedly low ($H = 0.7272$), but the number of haplotypes is not unexpected ($K = 6$; $P = 0.804$). Nucleotide diversity across the Y-to-dot translocation in lines containing the large deletion haplotype is lower than in lines without the deletion ($\pi_{\text{deletion}} = 0.0002$; $\pi_{\text{no deletion}} = 0.0006$). These results are consistent with positive selection acting on the deletion polymorphism in this intron. The acceptance rate for this combination of parameters under the growth model was an order of magnitude higher than simulations with the same parameters but under a model of constant population size (growth = 0.0671; constant 0.0035).

Purifying Selection on the Dot

Background selection and models of local adaptation are expected to increase F_{ST} where there are low levels of variation in a structured population (Charlesworth et al. 1997; Stephan et al. 1998). All loci surveyed here have low F_{ST} values between

populations (supplementary table S4, Supplementary Material online), indicating that populations are relatively homogeneous. Population homogeneity could be a signature of positive selection sweeping selected variants across subpopulations; however, there is evidence for considerable gene flow in *D. pseudoobscura* (Riley et al. 1989; Schaeffer and Miller 1992a; Kovacevic and Schaeffer 2000; Schaeffer et al. 2003), which also leads to population homogeneity. Therefore, based on population structure, we cannot exclude the possibility that background selection contributes to the reduction in variation in the Y-to-dot region. However, because there are relatively few genes in the translocated region, there is a relatively small target of purifying selection. Moreover, there is a severe skew in the frequency spectrum of the translocated region—a pattern unexpected under background selection (Charlesworth et al. 1995) where there is a low gene density. The conserved region of the dot chromosome, however, has ~90 genes, many with housekeeping functions. While there was a significant reduction in levels of neutral variability under both the standard neutral and exponential growth models for this part of the data set, the overall degree of skew in the frequency spectra did not reject either of these models ($D_{\text{Taj}} = -1.589$; $P_{\text{growth}} = 0.2392$; $P_{\text{constant}} = 0.0790$)—a pattern consistent with the effects of background selection. Because of the low level of recombination across the dot chromosome, the effects of background selection on the conserved region of the dot could impact the Y-to-dot translocation.

Discussion

The dot chromosome of *D. pseudoobscura* is unique in that it has acquired the ancestral Y chromosome (Larracuente et al. 2010). Although the dot chromosome is nonrecombining in *D. melanogaster* (Ashburner 1989), more recent empirical studies have indicated that recombination in the form of gene conversion, and possibly a low level of crossing over, has affected the dot chromosomes of *D. melanogaster* and *D. simulans* (Jensen et al. 2002; Wang et al. 2002; Wang et al. 2004). We find evidence for a low level of recombination mostly as gene conversion events on the dot chromosome of *D. pseudoobscura* but with some evidence for crossing over—with as many as 170 conversion events for each crossover. The low level of recombination on the dot is expected to affect the ability of natural selection to work efficiently on this chromosome. Indeed, the efficacy of selection appears reduced on the dot chromosome as seen in its higher nonsynonymous divergence compared with the other chromosomes and lower levels of codon bias (Haddrill et al. 2007; Singh et al. 2008; Betancourt et al. 2009; Arguello et al. 2010). Interestingly, the few recombination events that we did detect are restricted to the part of the dot chromosome that does not include the ancestral Y chromosome.

Natural Selection on the Y-to-dot Translocation

We find significantly reduced nucleotide variability across the dot chromosome of *D. pseudoobscura* (table 1). Though the demographic history of *D. pseudoobscura* likely has involved population expansion, we show that this history alone cannot explain the reduction in variation (using a model inferred from empirical data by Haddrill et al. 2010; fig. 3). However, the conserved part of the dot chromosome and the Y-to-dot translocation appear to have different evolutionary histories: notably, the frequency spectrum at the Y-to-dot translocated region indicates that positive selection may contribute to the departure from neutrality. Our data are consistent with a history of selective sweeps in the Y-to-dot translocated region. We estimate the most recent sweep to be approximately 83 kya (95% credible interval of 17–2,700 kya; fig. 4). We do not find evidence for recent selective sweeps outside of the Y-to-dot translocated region. This, coupled with the evidence for recombination in the part of the dot outside the translocation, indicates that the two parts of this chromosome are distinct and do not always evolve as a single linkage group. We do not know where on the dot chromosome the formerly Y-linked genes reside. It may be that they map to the small, heterochromatic arm of the chromosome, whereas the gene-dense region is on the long arm. Rare recombination events may increase the independence of these two parts of the chromosome.

Recurrent selective sweeps may not offer the only explanation for patterns of diversity we see across the Y-to-dot translocation of *D. pseudoobscura*. Background selection, or purifying selection against strongly deleterious mutations, can reduce levels of neutral variability (Charlesworth et al. 1995) and skew the frequency spectrum toward rare alleles in regions of low recombination (Kaiser and Charlesworth 2009). While patterns of low F_{ST} on the dot chromosome offer greater support for a model of selective sweeps than background selection, F_{ST} is only expected to be elevated under background selection if the population is structured (Charlesworth et al. 1997; Stephan et al. 1998). Extensive gene flow among *D. pseudoobscura* populations has been described at other autosomal loci and so this assumption may be violated. Pervasive selection against weakly deleterious mutations (e.g., transposable elements [TEs] where the product of the selection and dominance coefficients $\approx 2 \times 10^{-4}$) can also reduce levels of neutral variability and skew the frequency spectrum toward rare alleles (Charlesworth et al. 1995). Upon moving to the dot chromosome, the Y chromosome brings a high density of TEs and other repetitive elements. Many of these elements on the Y chromosome, however, are truncated and degenerated forms of relic TEs (Kurek et al. 2000). While the low gene density in the Y-to-dot translocation offers few targets, the high gene density of the conserved region of the dot chromosome offers many targets for purifying selection. Background selection is expected to have a larger impact in regions of low recombination where there are many targets for purifying selection (Kaiser and Charlesworth 2009). The reduced variation in the conserved region of the dot chromosome is consistent with the effects of background selection and this may impact the Y-to-dot translocation.

What might drive selective sweeps of the Y-to-dot translocation? Although the translocation event itself may have caused a selective sweep, it occurred sufficiently long ago that

it would not have left a signature in our data. Instead, the signature of sweeps that we see in the Y-to-dot region must be a result of subsequent events. We do not find evidence that the loci surveyed in this study experience positive selection at the protein level. Instead, selective sweeps may be driven by positive selection on the noncoding regions. The Y-to-dot translocation was followed by a ~10-fold reduction in intron size and reduction in intergenic space in the formerly Y-linked region (Carvalho and Clark 2005). Because of its male–male transmission and lack of recombination, the Y chromosome has a reduced efficacy of selection (reviewed in Charlesworth and Charlesworth 2000), allowing the expansion of introns to the megabase size range on the Y chromosomes of *Drosophila* species (Kurek et al. 2000). Transcribing megabase pair-long introns may come at a cost to the cell (Carvalho and Clark 1999), especially in an active, gene-dense part of the genome (Prachumwat et al. 2004) like the dot chromosome. While the size of the introns on the Y chromosome may have escaped selection, the quadrupled N_e and potential for infrequent recombination on the dot chromosome may have allowed for the shrinking of introns. Positive selection favoring deletions in introns may have triggered recurrent selective sweeps that shortened the Y-to-dot region over time, thus reducing levels of neutral variability. This hypothesis is supported by our observation of a large deletion polymorphism segregating at higher-than-expected frequencies in an intron of *ORY*—a gene in the Y-to-dot translocation. This region is also fixed for several indels between *D. pseudoobscura* and *D. miranda*.

Conclusions

It is intriguing that after millions of years with exclusive male–male transmission on a Y chromosome, testis-specific genes—some of which are essential for male fertility—could move to an autosome. The heterochromatic nature of the dot and Y chromosomes may have contributed to the success of the translocation: the gene regulatory environment may be sufficiently similar in heterochromatic regions of the genome. Also intriguing is the great reduction in size of the region since moving to the dot chromosome. Our data are consistent with the idea that this transformation was accompanied by positive selection, possibly favoring introns of shorter size. This exceptional case of karyotype evolution offers clues about how selection can rapidly remodel heterochromatic regions of the genome.

Materials and Methods

Fly Strains

We surveyed dot chromosome variation in 64 lines of *D. pseudoobscura* from nine different populations, spanning the geographic range of the species (supplementary table S2, Supplementary Material online). All samples were from isofemale lines that were inbred for 10 or 11 generations by Richard P. Meisel (except for the Flagstaff line). Genomic DNA was isolated from multiple male flies using a phenol chloroform DNA extraction.

Sequencing

Primers were designed using the CAF1 *D. pseudoobscura* assembly (supplementary table S3, Supplementary Material online; Drosophila 12 Genomes Consortium 2007). For each line, PCR resequencing was performed on 20 PCR products spanning the dot chromosome: 11 from the Y-to-dot translocated region and 9 from the rest of the dot. The PCR conditions for each reaction were at least 40 cycles of 95 °C for 30 s, 55 °C for 45 s, and 72 °C for 1 min. Unincorporated nucleotides were removed from PCR reactions using Exonuclease I/Shrimp alkaline phosphatase clean up prior to the sequencing reaction. PCR resequencing was done using the ABI Prism Big Dye cycle sequencing kit according to manufacturer's protocol and sequencing reactions were purified using a Sephadex column. Both the forward and reverse strands of each PCR product were sequenced using an ABI 3730 automated sequencer. A total of 10,593 bp were sequenced in each of 64 lines of *D. pseudoobscura*: 5,898 bp were from the Y-to-dot translocated region and 4,695 were from the rest of the dot chromosome. Traces were edited and aligned using Sequencher version 4.7 (Gene Codes, Ann Arbor, MI). Sequences were exported, concatenated (for each analysis), and formatted using custom PERL scripts.

Polymorphism Analysis

For each locus sampled and the concatenated dataset, θ_w (population mutation rate per silent nucleotide site), π (measure of nucleotide diversity per silent site), Tajima's D, haplotype diversity, and—for the concatenated data set—θ_h (measure of diversity per site that depends on high frequency variants) were estimated using the program "compute" in the analysis version 0.8.0 package associated with the libsequence version 1.7.0 library (Thornton 2003). For samples with segregating sites, F_{ST} was calculated using DNAsp version 4.10.3 (Rozas et al. 2003). Indels were eliminated from all analyses except the haplotype analysis described later. To estimate recombination rate parameters and to analyze the amount of LD across the sampled regions, the individual loci were concatenated. Because not every line has complete sequence for each locus, we imputed the missing data using fastPHASE (Scheet and Stephens 2006). The concatenated data set included 19 loci (*kl-3_2* was excluded), totaling 10,007 bp, and 40 SNPs; sites with multiple segregating mutations and gaps were eliminated. The relative location of each fragment was determined using the CAF1 Whole Genome Shotgun assembly. For the Y-to-dot translocated region, gaps with estimated length were factored into the distance; however, gaps of unknown length were not (e.g., gaps between scaffolds). Therefore, the distances in the Y-to-dot translocated region and thus the total length of the dot chromosome are approximations and should be considered to be minimum distances.

Polymorphism on the autosomes (Hamblin and Aquadro 1999; Schaeffer et al. 2001; Machado et al. 2002; Schaeffer et al. 2003; Haddrill et al. 2010) and X chromosome (Kovacevic and Schaeffer 2000; Machado et al. 2002; Haddrill et al. 2010) was summarized from the literature. For the third chromosome data borrowed from Schaeffer et al. (2003), we averaged

within-inversion summary statistics across five inversion types (AR, PP, ST, CH, and TL) rather than using estimates of diversity across inversion types because of population structure among third chromosome inversion types.

Divergence

The orthologous sequences for all 20 dot-linked loci in *D. miranda* (which shares the Y-to-dot translocation) were obtained by blasting a *D. miranda* Illumina short-read sequence assembly (Zhou and Bachtrog 2012). Divergence between *D. pseudoobscura* and *D. miranda* was calculated as the average number of nucleotide substitutions per site using DNAsp version 4.10.3 (Rozas et al. 2003). Divergence between *D. pseudoobscura* and *D. miranda* for the X chromosome and autosomes was summarized from the literature (Hamblin and Aquadro 1999; Kovacevic and Schaeffer 2000; Machado et al. 2002; Schaeffer et al. 2003; Haddrill et al. 2010).

Recombination

All analyses of recombination and LD were performed on the concatenated data set both with missing and imputed data, and no large difference between the results were found; only the results with the imputed data set are reported. The minimum number of recombination events were estimated using the RecMin software (http://www.stats.ox.ac.uk/~myers/RecMin/, last accessed January 14, 2014) as R_m (Hudson and Kaplan 1985) and R_h (Myers and Griffiths 2003). Pairwise estimates of r^2 and D' were obtained using the genetics package version 1.3.2 in R, and heat maps were drawn using the LDheatmap package in R (Shin et al. 2006). Estimates from the program "rsq" in the analysis version 0.8.0 package associated with the libsequence version 1.7.0 library (Thornton 2003) were used for the plots of r^2 as a function of the distance between SNPs. MaxHap software (Hudson 2001; http://home.uchicago.edu/~rhudson1/source/maxhap.html, last accessed January 14, 2014) was used to estimate ρ (4*Nr*; *N* is effective population size and *r* is the per generation per base rate of crossing over) and the rate of gene conversion to crossovers, f (g/ρ; where g is the probability per generation of a gene conversion at a particular site). A grid of 500 ρ values ranging from $\rho = 0.000001$ to $\rho = 0.1$ and 800 values of f from 0 to 1,000 was searched, with points equally spaced on a log scale. Mean conversion tract lengths ranging from 50 to 600 bp were considered, incrementing by 50 bp.

Neutral Coalescent Simulations

Neutral coalescent simulations were performed using a custom C++ script that uses the libsequence version 1.7.0 library (Thornton 2003). We simulated 10,000 genealogies using θ_w drawn from a random uniform distribution corresponding to the range of the observed autosomal θ_w at silent sites ($\theta_w \sim U(0.0037, 0.0359)$) summarized from the literature (Hamblin and Aquadro 1999; Schaeffer et al. 2001; Machado et al. 2002; Schaeffer et al. 2003).

Demographic History

We simulated data under a population expansion model consistent with the demographic history in Haddrill et al. (2010) using ms. To assess the fit of the expansion model to autosomal data from all studies, these simulations were performed 10,000 times and an empirical cumulative probability function (ecdf) on the distribution of π and Tajima's D was used to calculate two-sided P values in R. The FDR was calculated using the p.adjust function in R (Benjamini and Hochberg 1995). We also compared the variance in summary statistics between the observed and simulated data.

Modeling Selective Sweeps

We estimated the time since the most recent selective sweep using an ABC method with rejection sampling (Pritchard et al. 1999; Przeworski 2003; Larracuente and Clark 2013). We simulated selective sweeps in an expanding population: a population was expanding and at some time, t_{sweep} $4N_e$ generations in the past, all remaining lineages coalesced. For the population expansion phase of the simulations, we used the model described in the previous section, which we show accounts for levels of autosomal variability. Selective sweep simulations were performed using a custom C++ script that uses the libsequence version 1.7.0 library (Thornton 2003). For these simulations, we assumed that there is no recombination between any loci—we focused on contrasting the Y-to-dot translocated region with the conserved region of the dot chromosome but also analyzed the entire concatenated dot chromosome data set. A rejection sampling technique was used to obtain m samples from the joint posterior distribution of parameters in the expansion with sweep model. Because selective sweeps are expected to both reduce levels of neutral variability and create a skew toward rare variants in the frequency spectrum, we chose to summarize the data by the pairwise nucleotide diversity at silent sites (π) and a summary of the frequency spectrum, Tajima's D (Tajima 1989). We also used two complementary and informative summaries of the frequency spectrum—θ_h (Fay and Wu 2000) and the number of singletons (S_1)—because they are sensitive to intermediate and rare variants, respectively.

The rejection sampling algorithm is as follows:

1) Draw θ_w and t_{sweep} from prior distributions.
2) Simulate genealogies using the coalescent under a population expansion with selective sweep model based on empirical sample size.
3) Accept θ_w with the probability $u = Po(S, 4\theta\tau)/Po(S, S)$; otherwise return to step 1.
4) Place mutations on the tree according to the infinite sites model and calculate summary statistics for simulated genealogies.
5) Accept or reject chosen parameter values conditional on $|\pi_{Obs} - \pi_{Sim}| \leq \varepsilon$, $|D_{Obs} - D_{Sim}| \leq \varepsilon$, $|\theta_{h\ Obs} - \theta_{h\ Sim}| \leq \varepsilon$, and $|S_{1\ Obs} - S_{1\ Sim}| \leq \varepsilon$.
6) Return to step 1 and continue simulations until m desired samples from the joint posterior probability distribution are collected.

The prior distributions used were $\theta \sim \gamma(1, 0.1)$ and $t_{\text{sweep}} \sim U(1 \times 10^{-7}, 2)$. For simulations reported here, ε was set to 10% of the observed values of the summary statistics (in step 4) and m was set to 10,000. Our empirical data were silent sites from the concatenated data set of Y-to-dot chromosome loci sampled in this article. These simulations were also performed using silent sites from the whole dot chromosome concatenated data set and the concatenated data set for just the part of the dot chromosome not involved in the translocation.

Haplotype Tests

We performed coalescent-based tests of neutrality conditioning on haplotype configuration using haploconfig (http://www.stanford.edu/group/rosenberglab/haploconfig.html, last accessed January 14, 2014; Innan et al. 2005). Simulated haplotypes were used to calculate the probability that the observed haplotype configuration is unlikely given the sample size (n), the number of segregating sites (S), and the given mutation rate parameter (θ) in an expanding population (at time $t N$ generations ago, the population size was $N^{-\lambda t}$ where λ is the growth rate parameter, assuming no recombination). Two-sided P values were calculated based on 10,000 accepted simulated genealogies (where S has been accepted based on the given θ). Two-sided P values were also calculated for the number of haplotypes (K), the frequency of the most frequent haplotype (M), and the haplotype diversity (H) given S, θ, and n. The simulations were performed for the *ORY* intron where we find deletions segregating at high frequencies assuming $\lambda = 20$, $S = 5$, $n = 56$ (eight samples were dropped due to missing data within the locus) and for two values of θ: $\theta = 0.015$ (average autosomal θ) and $\theta = 0.002$ (estimated θ for both the dot chromosome overall and the *ORY* locus). We also performed simulations under a model of constant population size.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Arguello JR, 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.

Ashburner M. 1989. *Drosophila*: a laboratory handbook. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 57:289–300.

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

Bridges CB. 1935. The mutants and linkage data of chromosome four of *Drosophila melanogaster*. *Biol Zh.* 4:401–420.

Carvalho AB, Clark AG. 1999. Intron size and natural selection. *Nature* 401:344.

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

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

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res.* 70:155–174.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141:1619–1632.

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

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78:737–756.

Gao JJ, Watabe HA, Aotsuka T, Pang JF, Zhang YP. 2007. Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the Old World species. *BMC Evol Biol.* 7:87.

Gatti M, Pimpinelli S. 1983. Cytological and genetic analysis of the Y chromosome of *Drosophila melanogaster*. I. Organization of the fertility factors. *Chromosoma* 88:349–373.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. *Genome Biol.* 8:R18.

Haddrill PR, Loewe L, Charlesworth B. 2010. Estimating the parameters of selection on nonsynonymous mutations in *Drosophila pseudoobscura* and *D. miranda*. *Genetics* 185:1381–1396.

Hamblin MT, Aquadro CF. 1999. DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics* 153:859–869.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res.* 8:269–294.

Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics* 159:1805–1817.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.

Innan H, Zhang K, Marjoram P, Tavare S, Rosenberg NA. 2005. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics* 169:1763–1777.

Jensen MA, Charlesworth B, Kreitman M. 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* 160:493–507.

Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* 25:9–12.

Kovacevic M, Schaeffer SW. 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* 156:155–172.

Kurek R, Reugels AM, Lammermann U, Bunemann H. 2000. Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of *Drosophila* sp. *Genetica* 109:113–123.

Larracuente AM, Clark AG. 2013. Surprising differences in the variability of Y chromosomes in African and cosmopolitan populations of *Drosophila melanogaster*. *Genetics* 193:201–214.

Larracuente AM, Noor MA, Clark AG. 2010. Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. *Mol Biol Evol.* 27:1612–1620.

Larsson J, Chen JD, Rasheva V, Rasmuson-Lestander A, Pirrotta V. 2001. *Painting of fourth*, a chromosome-specific protein in *Drosophila*. *Proc Natl Acad Sci U S A.* 98:6273–6278.

Larsson J, Meller VH. 2006. Dosage compensation, the origin and the afterlife of sex chromosomes. *Chromosome Res.* 14:417–431.

Machado CA, Hey J. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc Biol Sci.* 270:1193–1202.

Machado CA, Kliman RM, Markert JA, Hey J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol.* 19:472–488.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.

Myers SR, Griffiths RC. 2003. Bounds on the minimum number of recombination events in a sample history. *Genetics* 163:375–394.

Ortiz-Barrientos D, Chang AS, Noor MA. 2006. A recombinational portrait of the *Drosophila pseudoobscura* genome. *Genet Res.* 87: 23–31.

Prachumwat A, DeVincentis L, Palopoli MF. 2004. Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. *Genetics* 166:1585–1590.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol.* 16:1791–1798.

Przeworski M. 2003. Estimating the time since the fixation of a beneficial allele. *Genetics* 164:1667–1676.

Riddle NC, Elgin SC. 2006. The dot chromosome of *Drosophila*: insights into chromatin states and their change over evolutionary time. *Chromosome Res.* 14:405–416.

Riley MA, Hallas ME, Lewontin RC. 1989. Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. *Genetics* 123:359–369.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

Schaeffer SW. 2002. Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genet Res.* 80:163–175.

Schaeffer SW, Aquadro CF, Anderson WW. 1987. Restriction-map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Mol Biol Evol.* 4:254–265.

Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, Kim K, Nelson JG, Anderson WW. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci U S A.* 100:8319–8324.

Schaeffer SW, Miller EL. 1992a. Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics* 132:471–480.

Schaeffer SW, Miller EL. 1992b. Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* 132:163–178.

Schaeffer SW, Walthour CS, Toleno DM, Olek AT, Miller EL. 2001. Protein variation in Adh and Adh-related in *Drosophila pseudoobscura*. Linkage disequilibrium between single nucleotide polymorphisms and protein alleles. *Genetics* 159:673–687.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 78:629–644.

Shin J, Blay S, McNeney B, Graham J. 2006. LDheatmap: an R function for graphic display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Soft.* 16:Code Snippet 3.

Singh ND, Larracuente AM, Clark AG. 2008. Contrasting the efficacy of selection on the X and autosomes in Drosophila. *Mol Biol Evol.* 25: 454–467.

Stephan W, Xing L, Kirby DA, Braverman JM. 1998. A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc Natl Acad Sci U S A.* 95:5649–5654.

Sun FL, Cuaycong MH, Craig CA, Wallrath LL, Locke J, Elgin SC. 2000. The fourth chromosome of *Drosophila melanogaster*: interspersed euchromatic and heterochromatic domains. *Proc Natl Acad Sci U S A.* 97:5340–5345.

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

Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.

Vicoso B, Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* 499:332–335.

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 JJ, Long M. 2004. Nucleotide variation and recombination along the fourth chromosome in *Drosophila simulans*. *Genetics* 166:1783–1794.

Zhou Q, Bachtrog D. 2012. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337:341–345.