Recurrent Selection on the Winters Sex-ratio Genes in Drosophila Simulans

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1534/genetics.109.109587</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:10403691">http://nrs.harvard.edu/urn-3:HUL.InstRepos:10403691</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a></td>
</tr>
</tbody>
</table>
Recurrent selection for the Winters sex-ratio genes in *Drosophila simulans*

Submitted to Genetics

Sarah B. Kingan*,†, Daniel Garrigan†, and Daniel L. Hartl*

* Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, 02138 USA

† Department of Biology, University of Rochester, Rochester, NY 14627

1. Present Address: Department of Biology, University of Rochester, Rochester, NY 14627
Corresponding Author:

Sarah B. Kingan

University of Rochester
Department of Biology
River Campus, Box 270211
Rochester, NY 14627-0211
Phone: (585) 275-4509
Fax: (585) 275-2070
e-mail: skingan@mail.rochester.edu
**ABSTRACT**

*Drosophila simulans* is unusual in having at least three distinct systems of X chromosome meiotic drive. These selfish genetic elements have biased transmission during meiosis, resulting in an excess of female progeny. Here, we characterize naturally occurring genetic variation at the Winters *sex-ratio* driver (*Distorter on the X* or *Dox*), its progenitor gene (*Mother of Dox* or *MDox*), and its suppressor gene (*Not Much Yang* or *Nmy*), which have been previously mapped and characterized. We survey three North American populations as well as 13 globally distributed strains, and present molecular polymorphism data at the three loci. We find that all three genes show signatures of selection in North America, judging from levels of polymorphism and skews in the site-frequency spectrum. This signature of selection likely results from the biased transmission of the driver and selection on the suppressor for the maintenance of equal sex ratios. The timing of selection is more recent than the age of the alleles, suggesting that the driver and suppressor are coevolving under an evolutionary “arms race”. None of the Winters *sex-ratio* genes are fixed in *D. simulans*, and at all loci we find ancestral alleles, which lack the gene insertions and exhibit high levels of nucleotide polymorphism compared to the derived alleles. In addition, we find several “null” alleles that have mutations on the derived *Dox* background, which result in loss of drive function. We discuss the possible causes of the maintenance of presence-absence polymorphism in the Winters *sex-ratio* genes.
Meiotic drive can leave signatures in the genome similar to positive natural selection without increasing the fitness of an organism (Lyttle 1993). Drive elements are preferentially transmitted during meiosis by disrupting the development or function of sperm carrying the homologous chromosome (Zimmering et al. 1970, meiotic drive sensu lato), or by true chromosome segregation defects during meiosis (Sandler and Novitski 1957, meiotic drive sensu stricto; Tao et al. 2007a); we use the term in the former sense throughout this article. While drive elements may arise on any chromosome, sex-chromosome meiotic drive is more easily detected due to the impact it has on the sex ratio, and a sex-linked driver is also more likely to invade a population than an autosomal driver (Hurst and Pomiankowski 1991). A chromosomal driver must maintain tight linkage with an insensitive target locus lest it drive against itself, a condition ensured by the lack of recombination between sex chromosomes (Charlesworth and Hartl 1978). Because of the impact drive elements have on sex ratios, sex-linked drivers are often referred to as “sex-ratio distorters” and the phenotype of skewed progeny sex ratios is termed “sex-ratio”. The mere transmission advantage of a driver, unless balanced by some detrimental fitness effect or masked by a suppressor, can cause it to sweep through a population in a manner similar to a positively selected mutation (Edwards 1961; Vaz and Carvalho 2004).

Obviously, a complete sweep of a sex-linked driver dooms a male-less (or female-less) population to extinction (Hamilton 1967), and natural selection strongly favors genetic factors that suppress drive and restore Mendelian segregation. Fisher (1930) presented a qualitative argument for the maintenance of an equal sex ratio, which predicts selection on any heritable variant that increases the production of the rarer sex. Fisher’s principle has been formalized mathematically and demonstrated empirically (e.g. Bodmer and Edwards 1960; Carvalho et al. 1998). Suppressors have been identified in a wide variety of meiotic drive systems and are predicted to be strongly favored by natural selection for the maintenance of equal sex ratios (reviewed by Jaenike 2001). Furthermore, the
evolution of enhancer genes on a suppressed driving chromosome may enable drivers to evade suppression, setting off another bout of selection for Fisherian sex ratios through suppression (Hartl 1975).

Meiotic drive is widespread, with systems identified in mammals, insects, and plants (Jaenike 2001). Drosophila is the most extensively studied insect taxon, and sex-chromosome meiotic drive systems have been identified in more than a dozen species (Jaenike 2001). The rapid evolution of suppressors by Fisherian selection results in a cryptic sex-ratio distorter, which may be identified only when the association between the driver and suppressor is lost, such as in hybrids between species or populations that do not share meiotic drive systems (Mercot et al. 1995). The coevolutionary arms race driving strong selection on drivers and suppressors likely contributes to Haldane’s rule (the preferential sterility of inviability of heterogametic hybrids) and is a leading explanation for the importance of X-linked loci in causing hybrid male sterility (Frank 1991; Hurst and Pomiankowski 1991; Presgraves 2008; Tao et al. 2007b). Two recently characterized hybrid male sterility factors, which are also sex-ratio distorters, are evidence of a possible link between meiotic drive and speciation (Orr and Irving 2005; Phadnis and Orr 2008; Tao et al. 2001).

Drosophila simulans is unusual in having at least three distinct X-linked drive systems, termed Paris, Durham, and Winters (Tao et al. 2007a). Here, we focus on the Winters sex-ratio (SR), whose driver and suppressor have been mapped to the gene level and whose molecular and cellular features have been elucidated (Tao et al. 2007a; Tao et al. 2007b). Two genes, Distorter on the X (Dox) and Mother of Dox (MDox) are required for sex-ratio distortion (Tao et al. 2007a, Y. Tao personal comm.). Dox is a duplicate copy of MDox, which is located 70 kilobases (kb) proximal from its progenitor locus on the X chromosome (Figure 1). A dominant suppressor, called Not Much Yang (Nmy) evolved on chromosome 3R as a retrotransposed copy of Dox (Tao et al. 2007b). Nmy likely suppresses Dox through an RNA interference mechanism by forming a double stranded RNA with homology to the distorter RNAs (Tao et al. 2007b). The genes of the Winters sex-ratio are not found in D.
Drosophila melanogaster, which diverged from D. simulans ~2.3 million years ago (L1 et al. 1999). Initial surveys of the genes in the simulans clade indicate that a functional Nmy gene is present in D. mauritiana (Tao et al. 2007b). Thus, the Winters genes are more than 250,000 years old, the speciation time of D. simulans, D. mauritiana, and D. sechellia (McDermott and Kliman 2008).

Signatures of positive selection have been previously detected at genomic regions linked to Drosophila sex-ratio distorters, but we present the first evidence of selection acting directly on a sex-ratio distorter gene and its suppressor gene. In Drosophila recens, driving X chromosomes show reduced nucleotide and haplotype variability relative to standard (non-driving) X chromosomes, and linkage disequilibrium extends over 130 cM of the driving chromosome (Dyer et al. 2007). The D. recens driver is located in a large chromosomal inversion and appears to be in the early stages of mutational degradation. In the Paris sex-ratio of D. simulans, Derome et al. (2004) found reduced haplotype diversity at the Nrg locus, which is closely linked to the Paris sex-ratio gene. In a later study, the group further localized the Paris driver to a pair of duplicated loci 150 kb apart, and demonstrated reduced haplotype diversity and linkage disequilibrium between variants associated with drive (Derome et al. 2008). In this study, we characterize patterns of genetic variation in natural populations of North American D. simulans and find signatures of strong positive selection at all three genes of the Winters sex-ratio.

MATERIALS AND METHODS

Population Samples: Samples from three North American populations of D. simulans were examined in this study (Supplementary Table 1). Two sets of isofemale lines were established from Massachusetts in September 2006: Tremont, collected in a backyard grape arbor on Tremont Street in Cambridge (n = 34), and Nicewicz, collected at the Nicewicz Family Farm in Bolton (n = 12), ~30 mi west of Cambridge. F1 males were frozen and used for DNA extraction. In addition, a set of isofemale lines collected in Winters, CA, in the summer of 1995 (Begun and Whitley 2000) was kindly donated
by Sergey Nuzhdin. We also obtained 13 lines of diverse geographic origins from the Tucson Species Stock Center: 5 African (Madagascar 14021.0251.196, 14021.0251.197, Kenya 14021.0251.199, Congo 14021.0251.184, and South Africa 14021.0251.169), 2 North American (California 14021.0251.194, unknown 14021.0251.195), 2 European (Scotland 14021.0251.216, Greece 14021.0251.181), and 4 Oceanian (New Guinea 14021.0251.009, New Zealand 14021.0251.007, Australia 14021.0251.176, New Caledonia 14021.0251.198). All strains were sampled randomly with respect to sex-ratio phenotype and genotype.

**Data Collection:** Genomic DNA was extracted from single males using a modified protocol of the Wizard Genomic DNA Purification Kit from Promega. From the Massachusetts populations F₁ males from wild-caught females were used, and so both autosomal alleles were included in our sample. All other stocks are inbred lab lines and the autosomal loci were found to be homozygous. Polymerase chain reaction was performed using Takara LA Taq polymerase according to manufacturer’s instructions. Previously published PCR primers for *Dox*, *MDox*, and *Nmy* were used that amplified complete genes as well as flanking sequence (Tao et al 2007a, Tao et al 2007b, Figure 1). Internal sequencing primers were used to obtain 2X coverage (forward and reverse reads) for PCR amplicons. Primers were designed using Primer3Plus (Untegasser et al. 2007) and Amplify v. 3.14 (Engels 2005). Sequencing was performed on an ABI3730 capillary sequencer according to manufacturer’s protocols. Sequences were edited using Sequencher v. 4.8 (Gene Codes Corp.) and aligned by eye with the aid of bl2seq program of the BLAST package (Tatusova and Madden 1999). Additional editing was performed using BioEdit (Hall, 2007). At the *Nmy* locus, singleton variants that were observed as heterozygous sites in chromatograms were confirmed with repeated PCR and sequencing. Two samples from the Tremont population, T44 and T62, are double heterozygotes at the *Nmy* locus; both heterozygous sites for these samples feature a singleton variant. Haplotype phase was resolved for these two samples by assuming that each singleton variant arose on the wildtype (e.g. most common) background, rather than the less likely case of both singletons having arisen on the same genetic
For each sample we collected 6.2 kb from Dox, 4.5 kb from MDox, and 7.5 kb from Nmy (Figure 1). A total of 1.6 Megabases (Mb) of resequence data were obtained.

**Data Analysis**: We calculated population genetic summary statistics using DnaSP (Rozas et al. 2003). The population mutation rate was estimated as the average pairwise diversity, $\pi$ (Tajima 1983) and Watterson’s estimator, $\theta_W$ (Watterson 1975) which is based on the number of segregating sites correcting for sample size. The site frequency spectrum was summarized by both Tajima’s $D$ (Tajima 1989) and Fay and Wu’s $H$ (Fay and Wu 2000). To summarize linkage disequilibrium (LD) across each gene, we estimated the statistic, $Z_{NS}$ (Kelly 1997), which is the average pairwise $R^2$ value among all variable sites (Hill and Robertson 1968) and $h$, the number of haplotypes. In order to calculate the age of the origin of the genes, we estimated divergence as the average number of nucleotide substitutions, $D_{XY}$ (Nei 1987 equation 10.20). The fit of various summary statistics to the standard neutral model was assessed through coalescent simulations using the observed number of segregating sites, the conservative assumption of no recombination, and 1000 simulations, as implemented in DnaSP. HKA tests were performed using the HKA software (Hey 2004; Hudson et al. 1987). Significance of HKA tests was determined from 10,000 coalescent simulations.

**Modeling selection**: A Bayesian approach was taken to estimate the time since selection on the Winters sex-ratio genes in each of the three North American populations, using coalescent simulations of neutral variants linked to a site under selection. The simulation has two phases (going forward in time), a complete selective sweep of a new beneficial variant followed by a neutral (recovery) phase. We used a modified version of a computer program by Przeworski (2003), which models the selected phase as the structured coalescent in which recombination between neutral variants and the site under selection is treated analogously to migration between demes (Kaplan et al. 1989). The neutral locus evolves according to the infinite sites model, with population mutation rate, $\theta = 4N\mu L$ (where $N$ is the effective population size, $\mu$ is the per site, per generation mutation rate, and $L$ is the length of the
sequence) and population recombination rate, $\rho = 4NrL$ (where $r$ is the per-site, per-generation recombination rate). Recombination between the neutral and selected sites occurs with rate $C = 4NrK$ (where $K$ is the distance between the closest neutral site and the selected site). The Bayesian method estimates the posterior probability distribution for the intensity ($4Ns$) and the time ($T$) since the completion of the selective sweep, using a summary likelihood method, in which the data are summarized by the number of segregating sites ($S$), number of haplotypes ($h$), and Tajima’s $D$.

The selection model has the following parameters: $N$, effective population size; $s$, the selection coefficient; $\mu$, mutation rate; $r$, recombination rate; and $T$, time since fixation of the beneficial variant. The posterior probability distributions for model parameters was generated using a rejection algorithm (Tavare et al. 1997). Briefly, parameter values are sampled from a prior distribution, a genealogy is simulated with the sampled parameters, and $S$ segregating sites are placed randomly onto the genealogy. The data summaries described above are calculated from the simulated genealogy and compared to the summaries from the observed data. Parameter values that generate the observed number of haplotypes and a Tajima’s $D$ values within some user-specified interval ($\epsilon$) are accepted and output to the posterior. To capture uncertainty in model parameters, the prior distribution of $\mu$, $r$, and $N$ are gamma-distributed whereas $s$ is sampled from a uniform prior.

**Choice of prior distributions of parameters.** In an effort to insure that the prior distribution of model parameters accurately reflect neutral variation in North American populations of *D. simulans*, we calculated the mean $\theta_W$ and $\rho$ (Hudson 1987) for 29 loci on the X and chromosome 3R sequenced in the same Winters, CA population (Begun and Whitely 2000, see Supplementary Table 2). We used gamma-distributed priors for $N$, $r$, and $\mu$ that yielded priors of the model parameters, $\theta$ and $\rho$ with these empirically observed means. We estimated $\theta_W$ and $\rho$ separately for loci on the X versus 3R and included all variable sites. The empirically estimated mean per site $\theta_W$ and $\rho$ for the X loci are 0.00488 and 0.01947 and for the 3R loci are 0.01029 and 0.08431 (Supplementary Table 2). The inheritance
scalar for the effective size of the X chromosome to that of the autosomes is accounted for in the joint prior probability distribution of \( \theta \) and \( \rho \). The analysis outputs time scaled in units of 4N generations and that scaling can be considered arbitrary. To avoid confusion, we have reported scaled times in units of N generations.

Mutation rate was calculated from whole-genome divergence between \( D. \) simulans and \( D. \) melanogaster. Begun et al. (2007) estimated lineage-specific divergence for \( D. \) simulans in 10 kb windows across the entire genome. We calculated \( \mu \) for each window as the lineage-specific divergence divided by 2.3MY, the divergence time for \( D. \) simulans and \( D. \) melanogaster (Li et al. 1999). Assuming 10 generations per year, this calculation gives a median per-site, per-generation mutation rate for chromosomes X and 3R of 1.2 x 10^{-9} and 1.0 x 10^{-9}, respectively. These estimates are within the range of other estimated mutation rates for Drosophila (Andolfatto and Przeworski 2000), but slightly lower than a commonly used mutation rate based on synonymous sites only (Sharp and Li 1989). If we assume there is a single effective population size for a population, the per-site, per-generation \( r \) can be calculated as \( (\rho * \mu) / \theta \). For the Winters, CA population data (Begun and Whitley 2000), we calculated \( r = 4.8 \times 10^{-9} \) for the X and \( r = 8.2 \times 10^{-9} \) for 3R. The prior distributions of \( \mu \) for the X and 3R were gamma with shape parameter 12 and 10, respectively, and scale parameter 10^{-10}; thus, the means of these distributions are 1.2 x 10^{-9} and 1.0 x 10^{-9}, respectively (Table 2). The prior distributions of \( r \) for the X and 3R were gamma distributed with shape parameter 48 and 82, respectively, and scale parameter 10^{-10}; thus, the means of these distributions are 4.8 x 10^{-9} and 8.2 x 10^{-9}, respectively. The prior for the selection coefficient, \( s \), was uniform between 5 x 10^{-4} and 0.5.

We chose to estimate \( r \) using a population genetic estimate (\( \rho \)) rather than genetic map data for several reasons. First, recombination rates estimated from genetic maps are systematically higher than those estimates from population genetic parameters (Andolfatto and Przeworski 2000; O'Reilly et al. 2008). While this pattern may be shaped by selection, demographic factors such as population
bottlenecks or population expansions may also increase levels of LD in natural populations (Stumpf and McVean 2003). Secondly, recombination rate in Drosophila is sensitive to maternal age, temperature, and genetic background and recombination estimates in laboratory stocks do not take into account these biological factors (Ashburner et al. 2005). Third, our use of the lower, population-based estimates of recombination is conservative with regards to the estimated strength of selection and timing of selection (i.e. time since selection may be over estimated and strength of selection may be under estimated).

RESULTS

Ancestral alleles observed at all loci: For each of the three sequenced loci we observe multiple chromosomes that lack the gene insertion, which represent the ancestral state of each locus (Supplementary Table 1). For convenience we refer to the presence of the gene insertion as the “derived” allele. At the Dox locus, four strains (two from Madagascar, one from New Caledonia, and one from New Zealand) lack the 3833 bp Dox gene insertion; at MDox, four samples lack the 3549 bp gene insertion (two from Madagascar, one from Congo, and one from New Zealand); and at Nmy, two North American samples lack the 2041 bp gene insertion (one individual each from Winters, CA, and the Tremont population from Massachusetts).

Null mutations at Dox: Three different alleles at Dox were observed that have the derived gene insertion but have lost their ability to drive (see Figure 2). The wild-type allele is the functional distorter Dox and is present in 75% of the sampled lines (n = 53). A previously characterized null allele dox[del105] is present in 3 copies (4%) (Tao et al. 2007a). This allele has a 105 bp deletion overlapping intron 2 and exon 3, which removes a region that is critical for distortion. Ten samples (14%) have the dox[del150] null allele, which has a total of 150 bp deleted in exon 4, including one large 135 bp deletion and two smaller deletions of 12 bp and 3 bp. We found a single copy of
*dox[del585]*, which shares the exon 4 deletions with *dox[del150]* but has an additional 435 bp deletion spanning exon 1 and intron 1. We tested the ability of *dox[del150]* and *dox[del105]* to distort sex ratios in a non-suppressing *nmy* background, where *nmy* is a loss-of-function mutant of the *Nmy* gene (Tao et al. 2007b). These crosses yielded progeny with equal sex ratios (see Supplementary Table 3 and Supplementary Figure 1). We assume that *dox[del585]* is a loss-of-function mutant because it shares the *dox[del150]* deletions in addition to the large deletion in exon 1.

**Insertion-deletion polymorphism:** Insertion-deletion (indel) polymorphisms at the *Dox* locus were already discussed in the context of loss-of-function mutations. At *MDox*, we observe one copy of *MDox[del105]* which has a 105 bp deletion that spans exons 2 and 3, one copy of *MDox[ins135]*, which has a total of 135 bp inserted into exon 3, and one copy of *MDox[ins32]* which has 32 bp inserted in exon 1. The functional implications of these mutations are not known. In some cases, the same indel polymorphisms were observed at *Dox* and *MDox*, and evidently derive from gene conversion between the two paralogs (see next section). In addition to indel polymorphism in the *MDox* gene sequence, we observe variable numbers and lengths of the 360 bp repetitive elements that flank the *MDox* gene (Tao et al. 2007a). (Copies of this element also flank the *Dox* gene and may facilitate gene conversion between the paralogs). The New Zealand and Kenyan samples have an additional full-length repeat element 5' of the *MDox* gene, and one of the Madagascar samples (14021.0251.197) is missing the two 3’ repeat elements. At *Nmy*, three samples (two from Madagascar and one from Congo) have a 6 bp insertion in one of the inverted repeats necessary for suppression by *Nmy*; we refer to this allele as *Nmy[ins6]*. Two of these three samples (the Congolese sample and one Madagascar sample, 14021.0251.196) also have a 201 bp insertion adjacent to a deletion of 77 bp between the inverted repeats, which is in the putative loop region of the RNA secondary structure (Tao et al. 2007a). The functional implications of these mutations at *Nmy* are not known.

**Nucleotide polymorphism and divergence:** Estimates of nucleotide polymorphism for the full dataset at all three genes are relatively low, but not unusually so compared to other datasets for *D*. 

12
simulans (ANDOLFATTO 2001; BEGUN and WHITLEY 2000). Importantly, the derived alleles have significantly reduced levels of nucleotide polymorphism compared to ancestral alleles (Table 1, Figure 3). Derived alleles have 2.22% of the ancestral allele diversity at Dox when measured as $\pi$ (4.38% when measured as $\theta_W$), and the corresponding parameters estimated for $\text{MDox}$ are 0.99% (4.62%) and for $\text{Nmy}$ 2.29% (14.65%). To test the statistical significance of the reduction, we performed pairwise HKA tests (HUDSON et al. 1987) for each locus, in which levels of polymorphism and divergence are compared for derived and ancestral alleles (Figure 3). Divergence was measured from $D.\ melanogaster$ in the region flanking the genes. Deviation from neutral expectations is significant for all three loci ($\text{Dox}: X^2 = 57.84, P < 0.0001; \text{MDox}: X^2 = 35.05, P < 0.0001; \text{Nmy}: X^2 = 13.716, P < 0.0036$).

To determine whether the Winters $\text{SR}$ genes show signatures of positive selection, we conducted multilocus HKA tests in which we compared polymorphism and divergence at each of the three Winters $\text{SR}$ genes in the three North American populations to that of 13 unrelated loci sampled in the same Winters, CA population (Table 3). For our “neutral” set of loci, we chose a subset of the 29 loci sampled by Begun and Whitley (2000) that had the largest number of sampled chromosomes ($n = 8$). The Winters $\text{SR}$ genes are predicted to be non-protein coding RNA genes (TAO et al. 2007a; TAO et al. 2007b) so we included all variable sites in our analysis because we cannot restrict our analysis to synonymous sites only, whose evolution is least likely to be influence by non-neutral processes (ANDOLFATTO 2005; HALLIGAN and KEIGHTLEY 2006). The original Begun and Whitely (2000) study analyzed only synonymous sites, so we reanalyzed all sites in their data in order to directly compare the datasets. A multi-locus HKA test on these 13 loci does not show any departure from neutral expectations ($X^2 = 17.99, P < 0.0764$, Table 3). However, when we include the Winters $\text{SR}$ genes we observe significant deviation from the neutral expectations in all but one test (Table 3). We first conducted nine tests where we added data for a single Winters $\text{SR}$ locus from a North American population to the 13 Begun and Whitely (2000) loci. All nine tests are significant except when we
added *Nmy* from the Winters population ($X^2 = 20.28, P = 0.0903$). For the *Nmy* data, we conducted two additional tests for the Tremont and Winters populations where we excluded the single ancestral allele present in each population. Both of these tests are significant (Winters: $X^2 = 59.92, P < 0.0001$; Tremont: $X^2 = 94.52, P < 0.0001$). (Here we report the uncorrected $P$-value but all tests remain significant at $P < 0.0011$ after a Bonferonni correction for multiple tests.) If positive selection has acted on the Winters *SR* genes, we expect to see deviation in the test in the direction of elevated divergence and reduced polymorphism at the Winters *SR* genes. In five of the 11 tests conducted, the Winters *SR* gene showed the largest deviation from neutral expectations in both polymorphism and divergence. In the remaining five significant tests, the Winters *SR* gene had the largest deviation from neutral expectations for divergence but not polymorphism. Moreover, these deviations were in the direction of reduced polymorphism and elevated divergence.

**Site-frequency spectrum:** Non-neutral processes such as natural selection or non-equilibrium demography shape the site-frequency spectrum, which is commonly summarized by the statistics Tajima’s $D$ (TAJIMA 1989) and Fay and Wu’s $H$ (FAY and WU 2000). Tajima’s $D$ ($TD$) is a summary of the folded frequency spectrum and compares two estimates of nucleotide polymorphism, $\pi$ and $\theta_w$, yielding a negative value if a locus has an excess of low frequency variants and a positive value if a locus has an excess of intermediate frequency variants. Fay and Wu’s $H$ ($FWH$) is a summary of the unfolded frequency spectrum and is sensitive to the frequency of derived mutations such that it is negative when there is an excess of high frequency derived variants. Both of these statistics are commonly used as tests of selection where a negative value for each is compatible with a locus having experienced a selective sweep.

We calculated $TD$ and $FWH$ at the Winters *SR* genes for a sample of all chromosomes, for each of the three North American populations and five African samples, and for the derived and ancestral alleles (Table 1). In the full dataset for all loci, we observe significantly negative Tajima’s $D$ values at each gene ($Dox$: $-2.19, P < 0.00001$; $MDox$: $-2.58, P < 0.00001$; $Nmy$: $-2.76, P < 0.00001$). For the
North American populations, all but two population samples for which we could conduct tests have significantly negative Tajima’s $D$ values ($Dox$: Nicewicz, -2.17, $P = 0.003$; Tremont, -0.29, n.s.; Winters, -0.34, n.s.; $MDox$: Nicewicz, -2.09, $P < 0.00001$; Tremont, -1.98, $P = 0.008$; Winters, -1.14, $P < 0.05$; $Nmy$: Nicewicz, n.a.; Tremont, -2.88, $P < 0.00001$; Winters, -2.32, $P = 0.008$). The African sample has a significantly positive $TD$ at $Dox$ (1.82, $P = 0.001$) and $TD$ values close to zero for the other loci ($MDox$: -0.36, n.s.; $Nmy$: 0.32, n.s.). At all loci, the derived alleles have significantly negative $TD$’s ($Dox$: -1.48, $P = 0.041$; $MDox$: -2.40, $P < 0.00001$; $Nmy$: -2.69, $P < 0.00001$) whereas the ancestral alleles have $TD$’s close to zero ($Dox$: 0.35, n.s.; $MDox$: -0.27, n.s.; $Nmy$: n.a.). Samples for which $TD$ values could not be calculated due to lack of segregating sites or too few samples are indicated with “n.a.” In summary, $TD$ estimates are compatible with positive selection acting at all three Winters $SR$ genes. At each gene, samples including all chromosomes as well as only the derived alleles show significantly negative $TD$ values. Because the site frequency spectrum is sensitive to population pooling (Hammer et al. 2003; Ptak and Przeworski 2002), the estimates for individual North American populations minimizes this problem (but may not eliminate it as the geographic scale of population structure in North American $D. simulans$ is not well understood). For the individual populations, we observe significantly negative $TD$ values for all tests except for Tremont and Winters at $Dox$. This pattern is not likely to result from demographic forces such as population growth because significantly negative $TD$ values are not observed at any of the reanalyzed Begun and Whitley (2000) loci (Supplementary Table 2), which were sampled in the same Winters, CA population.

For the complete dataset, we observe significant $FWH$ at $Nmy$, and marginal significance at the driver loci ($Dox$: -25.95, $P = 0.067$; $MDox$: -24.57, $P = 0.052$; $Nmy$: -91.63, $P < 0.00001$). Similarly, North American populations and samples of derived alleles also have significant $FWH$ at $Nmy$ only ($Dox$: Nicewicz, n.a.; Tremont, 0.06, n.s.; Winters, n.a.; derived, 0.03, n.s.; $MDox$: Nicewicz, 0.15, n.s.; Tremont, 0.18, n.s.; Winters, 0.15, n.s.; derived, 0.15, n.s.; $Nmy$: Nicewicz, n.a.; Tremont, -36.19, $P =
0.005; Winters, -72.88, $P < 0.00001$, derived, -29.72 $P = 0.003$). None of the tests are significant for the African sample or the ancestral alleles.

**Gene conversion between Dox and MDox:** Alignment of the paralogous region of the Dox and MDox loci reveal three gene-conversion tracts. The dox[del150] allele has a sequence motif of three deletions and a cluster of 5 single nucleotide polymorphisms (SNPs) that is shared with the wild-type MDox haplotype. In addition, we find one MDox haplotype that resembles the wild-type Dox haplotype in that it lacks these same three deletions and the SNP motif. Finally, the 105 bp deletion that characterizes the dox[del105] allele is also found in one MDox haplotype. These gene-conversion tracts were identified by eye and confirmed with the method of Betran et al. (1997) using the DnaSP software.

**Linkage disequilibrium:** Positive selection on a beneficial mutation can cause linked neutral variants to increase in frequency along with the selected site, which results in elevated levels of linkage disequilibrium across the genomic region. To test for elevated levels of LD at the Winters SR genes, we summarized LD as the average pairwise $R^2$ value across each gene, $Z_N S$ (Kelly 1997). We also tested for a reduction in the number of haplotypes ($h$), which results from hitchhiking by positive selection (Nielsen 2005). The results of this test are largely parallel with the estimates of $Z_N S$ (Table 1). In the complete dataset, we observe significantly elevated LD at all three loci (Dox: 0.52 $P = 0.003$; MDox: 0.32, $P = 0.046$; Nmy: 0.40, $P = 0.01$). Six of the ten populations for which we could calculate $Z_N S$ show elevated LD (Dox: Nicewicz, 0.80, $P = 0.007$; Tremont, 0.35, n.s.; Winters, 0.76, $P = 0.038$; Africa, 0.97, $P = 0.003$; MDox: Nicewicz, 0.83, $P = 0.015$; Tremont, 0.25, n.s.; Winters, n.a.; Africa, 0.37, n.s.; Nmy: Nicewicz, n.a.; Tremont, 0.84, $P < 0.00001$; Winters, 1.00, $P < 0.00001$, Africa, 0.47, n.s.).

Several factors besides selection may increase levels of LD in a sample. These include pooling derived and ancestral alleles (particular when alleles differ by large genomic insertions that may inhibit recombination), paralogous gene conversion, and pooling samples from different biological
populations. To explore these affects, we first calculated $Z_{NS}$ separately for derived and ancestral alleles. The sample including all derived alleles at $Nmy$ showed elevated LD ($n = 113, Z_{NS} = 0.41, P = 0.005$) but we see no significant $Z_{NS}$ values at other loci (Table 1). When we exclude the ancestral alleles in the Tremont and Winters populations at $Nmy$ (no ancestral alleles were observed at $Dox$ or $MDox$ in North America), the signature of LD is no longer evident (Tremont, $Z_{NS} = 0.0002$, n.s.; Winters $Z_{NS} = n.a.,$ no segregating sites), meaning that the elevated LD was caused by the presence of the single divergent ancestral allele. Next, gene conversion between $Dox$ and $MDox$ may have introduced several non-independent mutations, which will initially be in linkage disequilibrium with each other until the association is eroded by recombination or mutation. We performed a second analysis of LD after encoding all mutations within gene conversion tracts as a single mutation. This reanalysis only differed from our initial analysis in the LD estimates at $Dox$ and $MDox$, and resulted in no significant LD in the North American populations or the total sample of derived alleles (data not shown). Finally, pooling among subpopulation can result in spuriously high levels of LD (HARLTL and CLARK 2007). The African samples includes several lines from populations which are genetically differentiated from each other (BAUDRY et al. 2006), which may be the cause of the elevated LD in the complete dataset at each locus as well as the African sample at $Dox$. In summary, we do not observe elevated LD in samples of derived alleles in our North American populations after correcting for gene conversion or excluding ancestral gene copies.

**Age of derived alleles:** To estimate the age of the genes, the nucleotide divergence between the flanking sequence in the ancestral and derived alleles was calculated at each locus. From the sequence divergence, the age can be estimated as $t = d/(2\mu g)$ (where $d$ is the divergence per site, $\mu$ is the per-site per-generation mutation rate, and $g$ is generation time in years). We used the mutation rates calculated above for the modeling of selection. The per site divergence between the ancestral and derived alleles for $Dox$, $MDox$, and $Nmy$ are 0.0467, 0.0198, and 0.0165, yielding age estimates of 1.96 MY, 832,000 years and 817,000 years, respectively. Based on this result, the $Dox$ gene appears to
be much older than the other genes. It is possible that the duplication and transposition event that created the Dox gene may also be associated with extensive sequence changes, particularly in the repetitive sequences that flank the gene. A more accurate method of dating the Dox gene insertion is to determine the divergence between Dox and MDox at the gene insertion sequence, which is 0.0206, giving an age of 864,000 years, an estimate that is closer to the estimated ages of the other two genes. At MDox and Dox, we observed no shared polymorphisms and 22 and 77 fixed differences, respectively, between the ancestral and derived alleles. At Nmy, there are 12 shared polymorphism and 45 fixed differences—these shared polymorphisms result from a recombination event in the middle of the sequenced region such that sample T37a has the ancestral haplotype at the Nmy gene and a derived haplotype in the region distal to the gene.

**Timing of selection:** To estimate the time since selection on the three genes of the Winters sex-ratio, we implement a model of a selective sweep followed by a neutral (recovery) phase in each of the three North American populations (Przeworski 2003). We assume the selective sweep was complete and therefore restrict our analysis to the derived alleles at each gene, which leads us to exclude one ancestral Nmy allele from each of the Tremont and Winters population. We were unable to perform the analysis for the Nicewicz population at the Nmy locus, because only one segregating site is present and Tajima’s $D$ could not be calculated. By assuming fixation, we may be upwardly biasing our estimates of the time since selection at Nmy (in North America, Dox and MDox are fixed in our sample so this is less likely to be a problem at these loci). If ancestral alleles are segregating in the population, recombination between derived and ancestral alleles may introduce mutations onto the derived background, which would make derived alleles seem more diverse, and it would appear that selection occurred longer ago than it actually did. In view of the results actually obtained, therefore, excluding the ancestral Nmy sequences is conservative. In addition, the presence of gene conversion between Dox and MDox results in conservative estimates of time since selection. Gene conversion increases the
number of segregating sites by introducing multiple non-independent mutations, thus increasing the length of the recovery phase after selection is complete.

We generate 1000 sets of model parameters that are compatible with our data summaries at each locus. For five of the datasets, we accepted simulated Tajima’s $D$ values within $\varepsilon = 0.1$ of the observed data. However, three datasets ($D_{ox}$-Tremont, $D_{ox}$-Winters, and $N_{my}$-Winters) exhibited low acceptance rates, which led us to increase $\varepsilon$ to 0.5. The fit of the selection model to the data summaries can be evaluated based on the shape of the posterior distribution for $T$, the time since the sweep in coalescent time units of $N$ generations (see Supplementary Figure 2). If the posterior is flat, it suggests that selection is either too old to be detected (i.e. more than $4N$ generation ago), or else did not occur (PRZEWORSKI 2003). Based on an effective population size on the order of $1 \times 10^6$ years and 10 generations per year, we should be able to detect selection that occurred up to 4 million generations, or 400,000 years ago. All eight datasets are compatible with the model of selection (Supplementary Figure 2). The median time since selection for $D_{ox}$ and $M_{Dox}$ ranges from $0.0304 \times N$ generations to $0.0348 \times N$ generations (Table 4). At $N_{my}$, selection is more recent, with a median time of $0.0068 \times N$ generations for the Tremont population and $0.0164 \times N$ generations for the Winters population. The time since selection in years can be calculated as $t = TNg$ where $T$ is the time in coalescent time units, $N$ is the effective population size, and $g$ is the generation time in years, in this case 0.1, or 10 generations per year. At $D_{ox}$ and $M_{Dox}$, selection occurred around 3,000 years ago, with median times ranging from 2,800 years for the Tremont population at $M_{Dox}$ to 3,500 years ago for the Nicewicz population at $D_{ox}$ (Table 4). Selection in the Tremont population at $N_{my}$ is most recent (median time = 1,600 years), while in the Winters population at $N_{my}$ the median time since selection is 3,800 years. Importantly, the 95% credible interval for all eight datasets excludes the origin of the genes more than 250,000 years ago (TAO et al. 2007a). Selection most likely occurred less than 14,000 years ago, well after the genes of the Winters SR had evolved in the ancestor of the $D. simulans$ clade.
DISCUSSION

In this study, we characterize patterns of genetic variation in North American populations of *D. simulans* at the genes of the Winters sex-ratio, one of three X-linked meiotic drive systems in this species (*Tao et al. 2007a*). We find that the presence of all genes—the distorser locus, *Dox* its progenitor gene, *MDox*, and the suppressor, *Nmy*—are polymorphic in this species. The frequencies of the ancestral form of the driver loci (the allele which lacks the gene insertion) are highest in African and Oceanean samples, while ancestral *Nmy* is rare in the North American samples and absent in samples from other geographic localities. We also find evidence of gene conversion between *Dox* and *MDox*, the paralogous gene pair responsible for sex-ratio distortion in this system. Finally, we find several loss-of-function mutations on the derived *Dox* background, consistent with virtually complete suppression of the *sex-ratio* system in North American populations.

All three genes of the Winters *sex-ratio* show signatures consistent with recent positive selection. In this context, we use the term “selection” to also include the transmission-ratio advantage of the meiotic drive locus. The evidence for selection is two-fold. First, nucleotide variability on the derived allele background is greatly reduced compared to the ancestral allele background (Table 1, Figure 3). Second, all genes show skews in the site-frequency spectrum with an excess of low-frequency variants observed in all genes, and an excess of high-frequency derived variants observed at *Nmy*. These site-frequency skews are reflected in significant negative Tajima’s *D* and Fay and Wu’s *H* statistics (Table 1). Both of these patterns are consistent with a hitchhiking model where a new mutation has rapidly increased in frequency in a population due to natural selection or biased transmission during meiosis. In addition, we find our data to be compatible with a coalescent model of a recent selective sweep at all loci that occurred well *after* the origins of the genes (Table 4 and Figure 4). In fact, the 95% credible interval for the time since selection at all loci is more recent than the split between *D. simulans* and *D. mauritiana*, ~250,000 years ago (*McDermott and Kliman 2008*). This
result is consistent with theoretical prediction that meiotic drive systems experience repeated bouts of drive and suppression, and thus multiple rounds of selection (FRANK 1991).

Selection on the Winters sex-ratio is older than on the Paris sex-ratio, the other system in D. simulans that has been extensively studied. Derome et al. (2008) estimated that selection acted on the Paris driver only 88 years ago, based on an analysis of linkage disequilibrium across a region linked to the driver. Our results indicated selection acted less than 15,000 years ago, with an average age across loci of 3,000 years. Consistent with this estimate, we do not observe elevated linkage disequilibrium in derived gene copies at any of the Winters sex-ratio genes, after correcting for gene conversion between Dox and MDox. Significant linkage disequilibrium would be a signature of very recent selection. This signal is absent; whereas the signal of reduced polymorphism and skewed site frequencies are evident. At the time that selection was most likely acting on the genes of the Winters sex-ratio, the geographic range of D. simulans was restricted to Africa, the Indian ocean islands, and Eurasia (Lachaise et al. 1988). North America was likely settled ~500 years ago during the European colonization of the New World, facilitated by commensalism with humans (Lachaise and Silvain 2004). Interestingly, the most recent round of selection on the Winters SR occurred around the time of the expansion into Eurasia, 6,500-5,000 years ago (LACHAISE and SILVAIN 2004). Female biased populations have higher growth rates than populations with even sex ratios (HAMILTON 1967), suggesting that the unleashing of the Winters driver and the resulting excess of females may have facilitated the colonization of new habitats. However, due to the large credible intervals of the estimated time since selection, we cannot exclude the possibility that selection occurred when the species range was restricted to Africa.

Could other evolutionary forces besides selection have caused these departures from neutral patterns? Demographic forces such as population-size changes or population subdivision can have profound effects on genetic variation. However, these factors shape variation across all loci whereas selection targets particular genes or functional regions. Patterns of variation at Dox, MDox, and Nmy are unusual when compared to other loci sampled in North American populations (BEGUN and
In all three populations, each gene has either reduced polymorphism or elevated divergence, or both, as evidenced by significant multi-locus HKA tests (Table 3). Population growth can result in skews in the site-frequency spectrum similar to what we observed (i.e., an excess of rare variants and negative Tajima’s D). However, previous work indicates that populations of *D. simulans* were been subject to a population bottleneck during their colonization of the New World (Wall *et al.* 2002). Recent population bottlenecks are expected to result in an excess of intermediate frequency variants (Wakeley 2009), whereas we observe a dearth in our data. Indeed, the Tajima’s D estimates for the Begun and Whitley data (2000) are slightly positive, consistent with a population bottleneck (Supplementary Table 2). Combined with our detailed knowledge of the function of these genes (Tao *et al.* 2007a; Tao *et al.* 2007b), we are confident that the observed departures from neutral equilibrium expectations at the genes of the Winters sex-ratio are due to selection.

If all three genes show signatures of positive selection, why are they not fixed in the species? Even under a simple model of selective neutrality and drift, neutral mutations are not expected to persist beyond 4N generations, or roughly 400,000 years in the case of *D. simulans* if we assume 10 generations per year and an effective population size on the order of one million (Hartl and Clark 2007). Four copies of the ancestral distorter alleles were found in African and Oceanean populations and two ancestral suppressors were found in North America. Polymorphism at the suppressor can be explained from a simple model of selection to maintain Fisherian sex ratios. Assume, after Fisher (1930), that the total reproductive value of males and females is equal, \[ \sum_{i=1}^{m} W_i = \sum_{j=1}^{f} W_j \] where \( W_i \) is the fitness of the \( i \)th male, \( W_j \) is the fitness of the \( j \)th female, and there are \( m \) males and \( f \) females in the population. If we apportion fitness evenly among individuals of each sex, the fitness of each male is then simply equal to the sex ratio, \( W_i = f/m \). In a female-biased population, members of the “rarer sex” (males) have higher fitness. Under a model where a sex-ratio distorter invades a population and fixes due to its transmission advantage, selection on a new suppressor is frequency dependent. At low
frequency, a population is female-biased and selection for the maintenance of equal sex ratios is strong; but at high frequency, most copies of the distorter are masked, the population sex ratio is close to 50/50, and selection is much weaker. This result explains why selection for Fisherian sex ratios may be inefficient at removing the last few copies of a non-suppressor allele, even though under a deterministic model, the suppressor will eventually fix (Vaz and Carvalho 2004). In addition, selection is expected to be even less efficient at purging null suppressors if the functional suppressor is dominant, as vanishingly few individuals will express sex-ratio. This verbal model makes many simplifying assumptions such as panmixia, infinite population size, no pleiotropic fitness effects of drivers or distorters, and dominant suppression, but it could nevertheless explain the presence of ancestral Nmy alleles in North American populations that are fixed for the derived allele at both Dox and MDox.

Understanding the presence of null alleles of Dox and MDox is more complex. Under simple, single-population models of sex-chromosome drive, polymorphism between driving (SR) and standard (ST) X chromosomes can result from three conditions (Vaz and Carvalho 2004). First, the transmission advantage of an SR chromosome may be balanced by deleterious effects of either the driving locus itself or linked variants. Experimental work in a variety of Drosophila species indicates that when mated multiply, SR males suffer reduced fertility as well as reduced sperm competitive ability; these are examples of pleiotropic effects of the drive locus due to reduced sperm production (Jaenike 2001). Linked deleterious mutations may affect either male or female fitness and are common when driver elements occur in chromosomal inversions. In D. recens, females homozygous for SR chromosomes have reduced fertility, presumably due to a mutation at an unrelated locus trapped in the large inversion which contains the drive locus (Dyer et al. 2007). The last two conditions for SR/ST polymorphism require the evolution of suppressors by selection for Fisherian sex ratios or genomic conflict, which mask the expression of drive. If suppression is complete (i.e., suppressors are fixed) the
meiotic drive system is essentially “dead” and both loci evolve neutrally. If the suppression is partial (i.e., suppressors are polymorphic) polymorphism in the driver may be maintained.

For the Winters sex-ratio, we may argue against an offsetting deleterious effect based on several lines of evidence. First, the distorter is not located within a chromosomal inversion and is unlikely to be associated with deleterious variants. Second, theoretical work indicates that SR chromosomes balanced by deleterious effects cannot reach a frequency high enough to skew sex ratios and induce selection for suppressors (Vaz and Carvalho 2004). So the mere presence of \( N_{my} \) indicates the \( D_{ox}/M_{Dox} \) is not maintained as a balanced polymorphism. However, rejection of this hypothesis requires careful measurement of the fitness of each genotype. Interestingly, experiments suggest that the fertility of males expressing drive may be reduced relative to that of males with suppressed drivers (TAO et al. 2007b). Although rates of female remating in \( D. \ simulate \) is low (MARKOW 1996), in a female biased population, sperm limitation may be an issue for males. A difficulty in testing this hypothesis stems from the fact that small fitness effects may have important consequence in natural populations yet be undetectable in the laboratory.

The partial suppression hypothesis seems unlikely because, although \( N_{my} \) is not fixed, the frequency of males homozygous for non-suppressing \( N_{my} \) is very low. Based on the observed allele frequencies in our sample, non-suppressing males are expected to occur at 0.6% in the Winters population and at 0.02% in the Tremont population. Thus, the “neutral” explanation seems most likely as it is supported by the presence of loss-of-function mutations on the derived \( D_{ox} \) background and the near complete suppression of driving chromosomes based on observed allele frequencies in our sample.

Our inability to distinguish among these three hypotheses for the polymorphism in the Winters driver is complicated by the fact that \( D. \ simulate \) violates many assumptions of the simple population-genetic models implicit in the discussion above. The species exhibits high levels of population structure, particularly in Africa (Hamblin and Veuille 1999), and it is possible that the ancestral alleles
were sampled in populations that do not exchange migrants with populations that currently harbor the Winters sex-ratio genes. More extensive population sampling of the Madagascar, Congolese, New Caledonia, and New Zealand populations may shed light on this possibility. The possibility of competitive exclusion of the Winters driver by the Paris driver also exists. Notably, the frequency of the Paris driver is highest in central Africa and the Indian Ocean islands (Jutier et al. 2004), where, based on our coarse global sampling, ancestral copies of the Winters driver are found. Consistent with the competitive exclusion hypothesis, the intensity of drive is higher in the Paris system than in the Winters system, ~96% versus ~81% (Montchamp-Moreau et al. 2006; Tao et al. 2007b). Neither driver appears to be a balanced polymorphism that would limit the spread of the drivers through the population, so differential intensity of drive would in large part determine the frequency of the drivers in the population (Thomson and Feldman 1975). Testing the competitive exclusion hypothesis will require more extensive population sampling of the Winters driver, particularly in Africa and the Indian Ocean islands, as well as competition experiments between the two drivers in population cages in the laboratory.

Our analysis indicates that selection is much more recent than the actual origin of the Winters sex-ratio genes about 850,000 years ago. The date is based on sequence analysis and is consistent with the species distribution of the genes. All are absent in D. melanogaster but preliminary data indicates that the genes are present in D. mauritiana (Tao et al. 2007b, Kingan, unpublished data). Moreover, the D. sechellia Y chromosome is sensitive to drive by Dox (Tao et al. 2007a). An old origin but recent selection is suggestive of a genetic “arms-race” model for the evolution of drivers and suppressors, whereby multiple rounds of suppression and distortion occur due to ongoing genetic conflict between the loci (Frank 1991). In fact, the structure of the driving locus for Winters supports this “arms-race” model. Dox may have evolved as an enhancer or modifier of an original distorter, most likely MDox, which had been suppressed by an unknown locus (or an earlier form of Nmy). The most recent suppressor, Nmy, may then have evolved to suppress the new, compound distorter. This model is
testable by substituting chromosomes with a derived \textit{MDox} and ancestral \textit{Dox} into a variety of autosomal backgrounds. If drive is observed for some genotypes, it would confirm that \textit{MDox} was once able to drive alone. In addition, if there is polymorphism in the drive phenotype, one may be able to map the original suppressor of \textit{MDox}.

The Winters sex-\textit{ratio} is not the only trans-specific meiotic drive system: in mice, stalk-eyed flies, and Drosophila, shared drive systems are found in multiple closely related species (Jaenike 2001). The genomic conflict that results from a single meiotic drive system can have profound effects on patterns of genomic diversity in multiple species over a period of millions of years. On the molecular level, these patterns are indistinguishable from those caused by adaptation based on novel variation. It is only with a detailed understanding of the functional importance of genomic regions that one can attribute genomic signatures of selection to processes that increase the fitness of individual organisms.
ACKNOWLEDGMENTS

We thank Yun Tao for generously sharing research materials, fly stocks, and unpublished results as well as his expertise and insight into this system; also Luciana Araripe, Horacio Montenegro, Kalsang Namgyal, Erik Dopman, and Nguyen Nguyen for technical assistance, and Noemi Velazquez for administrative assistance. The Nicewicz family farm kindly gave us access to their farm for fly collections. We are grateful to John Wakeley and Molly Przewoski for help with the coalescent modeling. Daven Presgraves and Yun Tao provided thoughtful comments, which greatly improved the manuscript. This work was supported by NIH grant GM065169 to D.L.H and an NSF Graduate Research Fellowship to S.B.K.
REFERENCES


PTAK, S. E., and M. PRZEWORSKI, 2002 Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet 18: 559-563.