Reduced Sequence Variability on the Neo-*Y* Chromosome of
Drosophila americana americana

Bryant F. McAllister and Brian Charlesworth

Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637-1573 and Institute for Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

ABSTRACT

Sex chromosomes are generally morphologically and functionally distinct, but the evolutionary forces that cause this differentiation are poorly understood. *Drosophila americana americana* was used in this study to examine one aspect of sex chromosome evolution, the degeneration of nonrecombining *Y* chromosomes. The primary *X* chromosome of *D. a. americana* is fused with a chromosomal element that was ancestrally an autosome, causing this homologous chromosomal pair to segregate with the sex chromosomes. Sequence variation at the *Alcohol Dehydrogenase* (*Adh*) gene was used to determine the pattern of nucleotide variation on the neo-sex chromosomes in natural populations. Sequences of *Adh* were obtained for neo-*X* and neo-*Y* chromosomes of *D. a. americana*, and for *Adh* of *D. a. texana*, in which it is autosomal. No significant sequence differentiation is present between the neo-*X* and neo-*Y* chromosomes of *D. a. americana* or the autosomes of *D. a. texana*. There is a significantly lower level of sequence diversity on the neo-*Y* chromosome relative to the neo-*X* in *D. a. americana*. This reduction in variability on the neo-*Y* does not appear to have resulted from a selective sweep. Coalescent simulations of the evolutionary transition of an autosome into a *Y* chromosome indicate there may be a low level of recombination between the neo-*X* and neo-*Y* alleles of *Adh* and that the effective population size of this chromosome may have been reduced below the expected value of 25% of the autosomal effective size, possibly because of the effects of background selection or sexual selection.

S EX chromosomes generally exhibit a high degree of morphological and functional differentiation (Bull 1983). This heteromorphism is usually represented by a heterochromatic sex-limited chromosome (*Y* or *W*) that has few coding loci, but often contains an abundance of repetitive sequences. The chromosome (*X* or *Z*) that is shared between the two sexes contains a variety of functional genes and may exhibit some form of dosage compensation between the sexes. Sex chromosomes in a variety of taxa exhibit this pattern of heteromorphism, suggesting many independent origins of sex chromosomes from homologous chromosomal pairs (Bull 1983). Support for this hypothesis is provided by sequence comparisons that reveal independent autosome-to-sex chromosome transitions in mammals and birds (Graves 1995; Lahn and Page 1997; Fridolfsson *et al.* 1998). The absence of shared sequences between the *X* and *Y* chromosomes of Drosophila has, however, inspired a proposal of an alternative pathway for the evolution of its *Y* chromosome (Hackstein *et al.* 1996). A variety of models have been proposed to account for various aspects of the evolutionary transition of an identical chromosomal pair into

heteromorphic sex chromosomes (Muller 1914, 1918; Nei 1970; Charlesworth 1978, 1996; Rice 1987, 1996, 1998), although the mechanisms involved are poorly supported by empirical evidence.

Once a pair of homologues begins to segregate as proto-sex chromosomes, differentiation cannot occur until recombination in the heterogametic sex is restricted between the two sex chromosomes. Restricted recombination between the two sex chromosomes causes the genetic isolation of the sex-limited chromosome, and this isolation from recombination is an integral component of most models of *Y*-chromosome degeneration (Charlesworth 1996; Rice 1996). An empirical study of an artificial *Y* chromosome has confirmed the importance of this genetic isolation by demonstrating the accumulation of nonrecessive deleterious alleles in the absence of recombination (Rice 1994). Accumulation of deleterious alleles and subsequent loss in gene function on a nonrecombining chromosome must occur through an interaction between mutation, selection, and drift; alternative models of degeneration involve different relative contributions from these forces (Muller 1918; Nei 1970; Charlesworth 1978, 1996; Rice 1987, 1996, 1998; Charlesworth and Charlesworth 1997; Orr and Kim 1998). Two processes, background selection and selective sweeps, are likely candidates for the initial decay in gene function on a nonrecombining proto-*Y* chromosome when effective

Corresponding author: B. F. McAllister, Department of Biology, Box 19498, University of Texas, Arlington, TX 76019-0498.
E-mail: bryantm@exchange.uta.edu

population size (N_e) is very large, such as in many Drosophila species (Rice 1987, 1996; Charlesworth 1996). The stochastic accumulation of deleterious alleles caused by the successive loss of the class of chromosomes carrying the smallest number of deleterious mutations (Muller 1964), commonly known as Muller's ratchet, may also contribute to the degeneration of a nonrecombining proto-Y chromosome (Charlesworth 1978). The speed of this process is, however, highly dependent on N_e and characteristics of deleterious mutations (Charlesworth and Charlesworth 1997), so that it is not clear whether it would have a significant effect in groups like Drosophila, with their very large effective population sizes (Charlesworth 1996).

Under both background selection and Muller's ratchet, a nonrecombining Y chromosome has a low effective size, due to purifying selection continually removing chromosomes containing deleterious mutations from the population (Charlesworth *et al.* 1993; Rice 1998). When selection is sufficiently strong in relation to $1/N_e$, so that wild-type alleles always predominate at a given locus, mutation from wild-type to mutant allele is effectively irreversible. A sample of genes taken from a nonrecombining population ultimately traces its ancestry back to the relatively small genotypic class carrying the smallest number of such strongly selected deleterious mutations. This may facilitate the accumulation of mildly deleterious mutations (for which selection is weak in relation to $1/N_e$) on a proto-Y chromosome, because drift can cause such mutations to rise to high frequencies or even fixation (Charlesworth 1996) under conditions when they will be eliminated from a proto-X chromosome, because of the greater effectiveness of purifying selection in overcoming the effects of drift when N_e is large (Ohta 1992). Similarly, the fixation of advantageous mutations is inhibited on the proto-Y chromosome (Rice 1996; Orr and Kim 1998).

The selective sweep model proposes that selection acts in a positive way. An advantageous mutation occurring on a particular proto-Y chromosome may cause the fixation, or selective sweep, of the entire chromosome on which it occurs. Deleterious alleles at other loci on this chromosome would fix by hitchhiking with the beneficial allele (Rice 1987), assuming that the chromosome has a net fitness advantage. Fixation of deleterious alleles during a selective sweep may be facilitated by the masking effect of the X chromosome (Muller 1914, 1918; Fisher 1935; Nei 1970). The long-term consequence of these selective sweeps would be a locus-by-locus decay in genetic function on the proto-Y chromosome as successive alleles are fixed in the population.

Species with secondary sex chromosomes, formed by the fusion of an autosome with a primary sex chromosome, provide excellent models for studying sex chromosome evolution (Steinemann and Steinemann 1998). Several species of Drosophila have secondary sex
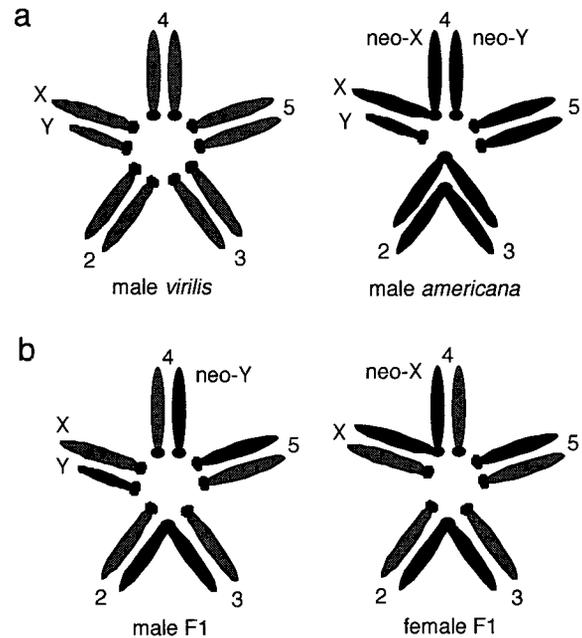

Figure 1.—Karotypes of *D. virilis, D. a. americana,* and F$_1$ hybrids. The dot chromosomes are excluded from the diagrams. (a) Chromosomal configurations in male *D. virilis* and male *D. a. americana.* All five chromosomal elements are present in *D. virilis.* The $X/4$ fusion, which created the neo-X and neo-Y, is represented in the karyotype of the male *D. a. americana* together with the fusion of chromosomes 2 and 3. (b) Chromosomal configurations of the male and female F$_1$ progeny derived from a cross between female *D. virilis* and male *D. a. americana.* Due to segregation in male *D. a. americana,* the neo-Y chromosome is inherited by male F$_1$ progeny and the neo-X chromosome is inherited by female F$_1$ progeny.

chromosomes that represent a gradation in sex chromosome differentiation, with older chromosomes exhibiting greater differentiation (Ashburner 1989; Charlesworth 1996; Powell 1997). The secondary sex chromosomes of *Drosophila americana americana* were formed by fusion of the primary X with chromosomal element 4 (Muller's element B), which was a freely segregating autosome in the ancestral state represented by *D. a. texana* and *D. virilis* (Hughes 1939; Stalker 1940; Throckmorton 1982). Following the original fusion event, chromosomes with the fused $X/4$ arrangement would have acquired sequence variability through recombination with unfused chromosomes in heterozygous females. As the $X/4$ fusion increased in frequency, the element 4 homologues began to segregate as sex chromosomes. In *D. a. americana* (see Figure 1a), neo-X chromosomes are represented by the element 4 arm of $X/4$ fusion chromosomes, and neo-Y chromosomes are represented by free fourth chromosomes, which are male-limited and segregate with the ancestral Y during male meiosis. Recombination has been restricted between the neo-X and neo-Y in *D. a. americana* since the fixation (or near fixation) of the $X/4$ fusion, due to the extremely low frequency of recombination in male

D. virilis (Kikkawa 1935), as in other species of Drosophila (Gethmann 1988). With this genetic isolation of the neo-*Y*, the process of degeneration was initiated.

The secondary sex chromosomes in *D. a. americana* apparently represent a very early stage of differentiation; therefore, they should provide a good model for examining the mechanisms that lead to the initial loss in gene function on *Y* chromosomes. There is no cytological or genetic evidence for degeneration of the neo-*Y* or for dosage compensation on the neo-*X* in *D. a. americana* (Bone and Kuroda 1996; Marin *et al.* 1996; Charlesworth *et al.* 1997a). One explanation for the lack of differentiation between the neo-sex chromosomes of *D. a. americana* is that the recent fixation of the *X/4* fusion has provided little opportunity for differentiation to occur. The argument is supported by the absence of the *X/4* fusion in *D. a. texana* (Throckmorton 1982), coupled with DNA sequence data indicating that *D. a. americana* and *D. a. texana* are very closely related (Tominaga and Narise 1995; Hilton and Hey 1996, 1997; Nurminsky *et al.* 1996). The presence of the *X/4* fusion chromosome in *D. a. americana* is currently the only diagnostic feature of the two subspecies. Previous studies indicated that the *X/4* fusion chromosome is fixed throughout the range of *D. a. americana* in the north central to northeastern United States, whereas this arrangement is not present in *D. a. texana* in the south central to southeastern United States (Hsu 1952; Patterson and Stone 1952; Throckmorton 1982). A hybrid zone involving the *X/4* fusion is present in the parapatric region where the ranges of the two subspecies overlap (Throckmorton 1982). Because of the presence of the hybrid zone, an alternative explanation for the lack of degeneration of the neo-*Y* in *D. a. americana* is that gene flow from *D. a. texana* provides a "fresh" input of unfused autosomes into the pool of neo-*Y* chromosomes (Charlesworth *et al.* 1997a).

In this study, a population analysis of DNA sequences is presented for the *Alcohol Dehydrogenase* (*Adh*) gene of *D. a. americana* and *D. a. texana*. The *Adh* gene is located at cytological band 49B, which is near the centromere of chromosomal element 4 in *D. virilis*, *D. a. americana*, and *D. a. texana* (Gubenko and Evgen'ev 1984; Nurminsky *et al.* 1996; Charlesworth *et al.* 1997a). Sequence data from *Adh* provide a direct measure of differentiation between neo-*X* and neo-*Y* chromosomes of *D. a. americana* and the autosomes of *D. a. texana*. Furthermore, the pattern of sequence variation on the neo-*Y* chromosomes of *D. a. americana* can be examined to determine if hitchhiking due to background selection, Muller's ratchet, and/or selective sweeps has influenced nucleotide variability on this chromosome. These processes are expected to reduce nucleotide variability on this chromosome; background selection is, however, unlikely to cause a detectable departure from neutrality in the pattern of nucleotide variation at segregating sites, whereas selective sweeps can have a strong effect

(Braverman *et al.* 1995; Charlesworth *et al.* 1995; Simonsen *et al.* 1995). The fact that the reductions in N_e in both Muller's ratchet and background selection result from a similar process suggests that they will have similar effects on patterns of variability, although this remains to be investigated formally.

MATERIALS AND METHODS

Strains: Laboratory strains of *D. virilis* and *D. a. americana* were obtained from the National Drosophila Species Resource Center (Charlesworth *et al.* 1997a). In addition to these laboratory lines, which represent a subspecies-wide sample and have been maintained in culture for many years, samples of *D. a. americana* and *D. a. texana* were collected from natural populations for this study. A sample of *D. a. americana* was collected during October 1996 from bait cups containing fermented banana hung among a dense stand of sandbar willows (*Salix exigua*) in a marsh located ~10 km west of Gary, Indiana. A sample of *D. a. texana* was similarly collected during April 1997 among large black willows (*S. nigra*) growing in a cove of a lake about 5 km west of Lone Star, Texas. Each line was established from either a single wild-caught female or a wild-caught pair. The Gary sample is referred to as G96.# (with # being the line identification), whereas the Lone Star sample is referred to as LP97.#.

Karyotypic analyses were performed on the newly collected lines to verify the presence or absence of the *X/4* fusion chromosome. The laboratory lines were also karyotyped recently (Charlesworth *et al.* 1997a). Larvae were cultured at low density on banana medium maintained at 18°. Ganglia from third instar larvae were dissected in Ringer's solution, placed in a 0.07 m sodium citrate solution for 15 min, fixed in a 1:1 solution of 100% ethanol:45% acetic acid, and stained in 2% orcein. Several cells from at least two larvae were scored for each line.

DNA extraction and sequencing: Hybrids between *D. a. americana* and *D. virilis* are easy to obtain, so that DNA was extracted from hybrid offspring and used to obtain templates for sequencing. This procedure allowed for the use of species-specific primers so that single alleles of *D. a. americana* could be sequenced and assigned to neo-*X* or neo-*Y* chromosomes. Males from the *D. a. americana* lines were mated to females from a strain (V.46) of *D. virilis* carrying multiple visible mutations (see Charlesworth *et al.* 1997a). Segregation in male *D. a. americana* carrying the *X/4* fusion results in the F$_1$ female offspring carrying the neo-*X* chromosome, whereas F$_1$ males carry the neo-*Y* chromosome (Figure 1b). F$_1$ hybrids were also used to obtain a sequencing template from *D. a. texana* by crossing either a male or female to the V.46 strain of *D. virilis*. Single F$_1$ flies were used for DNA extraction in all cases. Each fly was homogenized in 50 μl of a buffer containing 0.15 m NaCl, 0.1 m EDTA (pH 8.0), 0.5% SDS, and 150 μg/ml protease K. The solution was incubated at 55° for 1 hr and extracted once with phenol/chloroform (1:1) and once with chloroform alone. The DNA was ethanol precipitated and resuspended in TE.

Nucleotide sequences of the *Adh* locus from species in the virilis group were obtained from GenBank (Nurminsky *et al.* 1996). To obtain sequences from single alleles, differences between the *Adh* sequences of *D. virilis* and *D. a. americana*/ *D. a. texana* were identified. Primers were designed in these regions, allowing for the use of PCR in the specific amplification of *D. a. americana*/ *D. a. texana* alleles in the F$_1$ hybrids. Two overlapping regions of the *Adh* gene were amplified (Figure 2): the 5′ end by Adh5f, 5′ CCG ACT AGA AAG CAT CAC, and
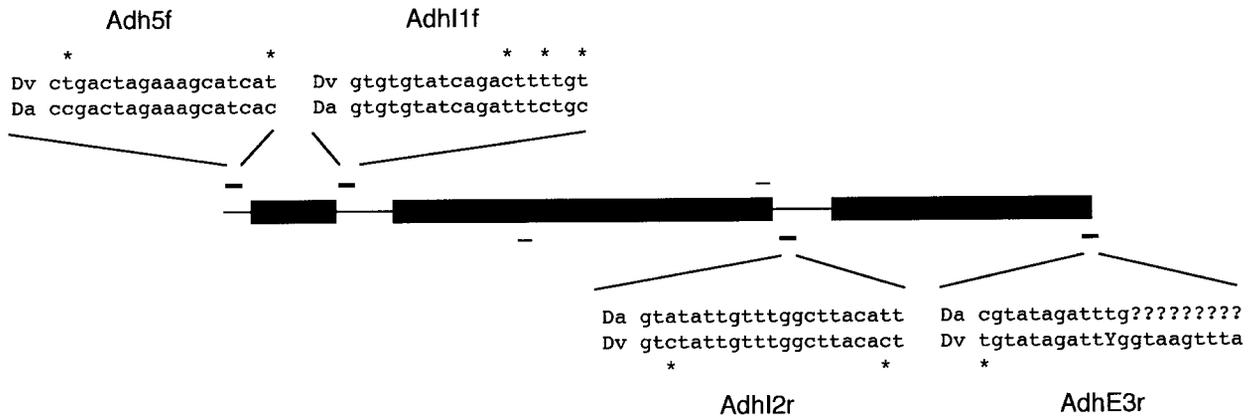
Figure 2.—Structure and analysis of the *Adh* gene. Sequences were obtained for an 884-bp region of the gene, from 7 bases upstream of the initiation codon to 7 bases prior to the stop codon. The three exons are indicated by black boxes and the noncoding regions are represented by the line. Positions of the two primer pairs, Adh5f/AdhI2r and AdhI1f/AdhE3r, used for amplification of the region from genomic DNA, are indicated. Sequences of *D. virilis* (Dv) and *D. a. americana/texana* (Da) are indicated for the primer regions, where asterisks denote the sites of nucleotide differences between the primer sequences and the *D. virilis* sequence. Positions of two internal sequencing primers are also indicated by light lines.

AdhI2r, 5′ TTA CAT TCG GTT TGT TAT ATG, and the 3′ end by AdhI1f, 5′ GTG TGT ATC AGA TTT CTG C, and AdhE3r, 5′ ATT TGA ATG GTT TAG ATA TGC. To obtain a template for sequencing, DNA from the F_1 flies was used in 25-μl reactions containing 1/10 reaction buffer (Perkin Elmer, Norwalk, CT), 2.5 mm $MgCl_2$, 0.1 μm each dNTP, 0.2 μM each primer, and 2 units *Taq* polymerase. After an initial 95° denaturation for 2 min, the reactions were cycled 30 times using the following parameters: 95°, 0.5 min; 54°, 1 min; and 72°, 1 min. Upon obtaining reaction products, 20 μl of each reaction was column purified (QIAGEN, Valencia, CA). The four primers used to amplify the products and two additional internal primers were used in 10-μl cycle sequencing reactions. Approximately 30 ng of each purified PCR product was mixed with 4 μl of ABI PRISM dye terminator reaction mix (Perkin Elmer) and 3 pg of the appropriate oligonucleotide. Reactions were cycled 25 times with the following parameters: 95°, 30 sec; 54°, 30 sec; and 60°, 3 min. Products were ethanol precipitated and run on either an ABI 373 or ABI 377 automated DNA sequencer.

Statistical analyses of sequence data: The DNA sequences were visually aligned using SeqPup (D. Gilbert, Indiana University) and compared using SITES (Hey and Wakely 1997). The standard estimates of nucleotide diversity, π and θ, and their expected errors, were calculated according to Tajima (1993). These are based respectively on the average pairwise differences present in the samples and the number of segregating sites in the samples. To determine the level of sequence divergence that was present among samples, the net number of pairwise differences (*d*) was estimated (Nei 1987, p. 276). The *D* statistics of Tajima (1989) and Fu and Li (1993) were used to determine if the distribution of variation within samples was consistent with a neutral model. Both statistics were based on all the observed nucleotide variants. The *D. virilis* sequence was used as the outgroup to determine the ancestral state for the Fu and Li test.

It was also necessary to determine the statistical significance of an observed difference in variability between the neo-*X* and neo-*Y* chromosomes. An appropriate standard statistical test has not been devised to compare levels of nucleotide diversity between samples, so a resampling method was used. A pooled set of neo-*X* and neo-*Y* chromosomes corresponding to the sequences in the actual samples was randomly sampled with replacement into two groups of the same sizes as the original samples. For each iteration of this resampling, the mean number of pairwise differences between sequences was calculated for the two randomized samples, along with the difference in their means. Repetition of the resampling process 1000 times provided an estimate of the probability of obtaining neo-*X* and neo-*Y* samples from this population with a difference in mean pairwise difference greater than or equal to the observed value. Because the neo-*X* and neo-*Y* sequences were obtained from a single population of flies, it is appropriate to apply this test, which only considers the sampling error associated with the estimated diversities for these two samples and disregards the evolutionary stochastic error. The resampling procedure is simply a test of the null hypothesis that the neo-*X* and neo-*Y* alleles are sampled from one homogenous population at the locality where the flies were collected.

Coalescent simulations: In addition, coalescent simulations were used to examine the time of origin of the neo-sex chromosomes and the extent of any reduction in effective population size for the neo-*Y* chromosome. We assume that the centric fusion generating the neo-*X* chromosome reached fixation T time units ago (Figure 3), where time is measured in units of the mean time to coalescence of a strictly neutral autosomal locus. One coalescent time unit is thus equal to $2N_e$ generations, where N_e is the effective population size with respect to autosomal genes, assumed to be constant throughout (Hudson 1990). The time taken for fixation of the fusion is assumed to be negligible compared with coalescent time, and is treated as zero in what follows. Prior to the origin of the neo-*Y* chromosome, the locus was inherited as an autosomal gene and hence had a coalescent time of unity.

After the fixation of the fusion, the neo-*X* and neo-*Y* alleles are assumed to have coalescent times that are fractions t_x and t_y of the autosomal rate (Figure 3). In the absence of sexual selection, which would cause the variance in reproductive success of males to be greater than that of females, t_x is 0.75, but it may approach or even exceed 1 if there is intense sexual selection (Charlesworth 1994; Caballero 1995; Nagylaki 1995). We similarly expect t_y to be equal to 0.25 in the absence of sexual selection, but to be reduced below this by sexual selection. Hitchhiking due to selective sweeps or background selection acting on the neo-*Y* would reduce t_y even more. In the simulations, we assumed that $t_x = 0.75$ or 1.0 and consid-
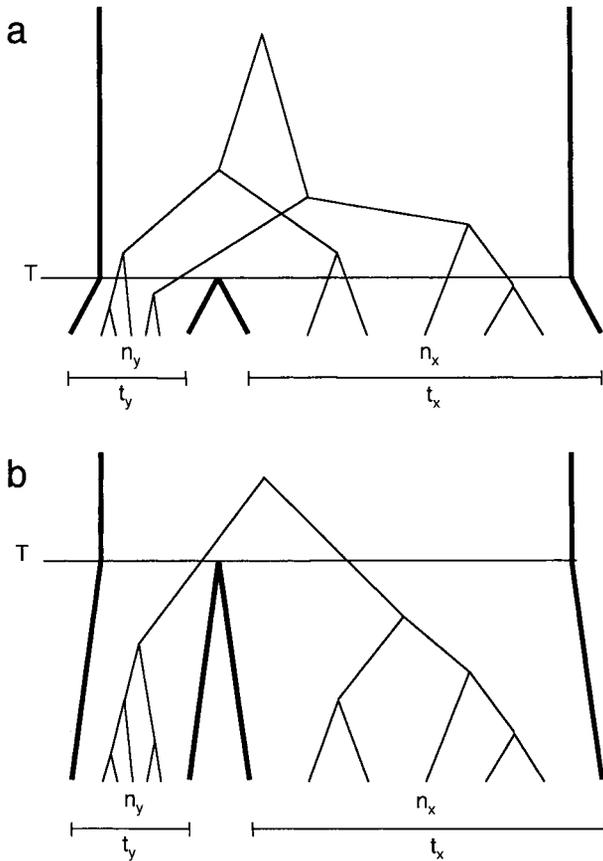
Figure 3.—Representation of the genealogical process without recombination as modeled in the coalescent simulations. Fixation of the *X/4* fusion and isolation between the neo-*Y* and neo-*X* chromosomes occurs at time *T* before the samples of n_x and n_y neo-*X* and neo-*Y* alleles were collected. Prior to this the alleles were present in one population, but afterward the two populations were separate, each with their respective coalescent times, t_x and t_y. (a) Idealized genealogy of alleles with coalescent events occurring both after and prior to the fusion at time *T*. (b) Idealized example of "premature coalescence," where all the neo-*X* and neo-*Y* alleles descend from a single common ancestor following the fixation of the *X/4* fusion at *T*.

ered the effect of varying t_y between 0.25 and 0.01. The two values for t_x span the range between no sexual selection and intense sexual selection, and the values for t_y span the range from no sexual selection or hitchhiking to intense effects of such factors. The neutral mutation rate scaled in units of coalescent time is denoted by $M = 2 N_e u$, where *u* is the neutral mutation rate per generation, summed over all relevant sites.

The coalescent model is similar to that previously employed for two subpopulations connected by migration (Hudson 1990; Wakeley 1996; Nordborg 1997), except that we allow for a finite time since the split between the two subpopulations. At the start of a replicate run, a sample of n_x and n_y genes is assumed to be drawn from neo-*X* and neo-*Y* chromosomes derived from a natural population. At an arbitrary time *t* prior to sampling ($t \leq T$), let there be n_{xt} and n_{yt} genes from the sample present in the neo-*X* and neo-*Y* subpopulations, respectively. The rates per unit coalescent time of coalescent events within the neo-*X* and neo-*Y* subpopulations are $k_{xt} = n_{xt} (n_{xt} - 1)/2t_x$ and $k_{yt} = n_{yt} (n_{yt} - 1)/2t_y$, respectively.

Let R_x be the rate (in units of coalescent time) at which

alleles that are currently in the neo-*X* subpopulation are derived from the neo-*Y* population, and R_y be the corresponding rate at which alleles in the neo-*Y* population are derived from the neo-*X* subpopulation. If unfused fourth chromosomes are rare, there are approximately three times as many neo-*X* chromosomes as neo-*Y* chromosomes. It is thus reasonable to assume that $R_x = 0.333 \ R_y$. At time *t*, the net rates at which recombination events occur are thus $n_{xt}R_x$ and $n_{yt}R_y$ for the neo-*X* and neo-*Y*, respectively. If these rates are of order 1, the corresponding distributions of the waiting times to recombination events are exponential with expectations of $1/(n_{xt}R_x)$ and $1/(n_{yt}R_y)$. The probability that a given event occurs in a given subpopulation and is a coalescent or a recombination event is the ratio of the rate of the event to the sum of the rates for all four types of event (Hudson 1990). The latter is given by the sum of the rates of the events in question. A uniform random number is thus compared with the probabilities of each class of event to determine what happens at the time point following *t*. An exponentially distributed random number is chosen, from a distribution of mean one, and divided by the sum of the current rate parameters to determine the time from *t* to the next coalescent or recombination event.

The genes within a given subpopulation are classified as having been present in that population in the initial sample or as having entered the subpopulation from the other one by recombination. Records are kept of the numbers of genes in each of these classes and of the numbers of genes within each of them that have previously experienced a coalescent event with a given class of gene. Mutations are laid down on internal and external branches of the gene tree connecting the set of sampled alleles, according to Poisson distributions with means 0.5 *M* times the lengths of the appropriate branches. Singleton variants correspond to mutations on external branches; fixed mutations correspond to mutations that occur on branches that are not shared by genes sampled from different subpopulations. This procedure is repeated until time *T* is reached or until coalescence of all the neo-*X* genes or neo-*Y* genes with other genes present in the same initial subpopulation ("premature coalescence"). Premature coalescence corresponds to a situation in which one or more of the following possibilities are realized: (i) all variants on the neo-*X* or all neo-*Y* are found only within their own subpopulation; (ii) there are fixed differences between the neo-*X* and neo-*Y* chromosomes; or (iii) there are no segregating sites within either or both of the neo-*X* and neo-*Y* samples (Figure 3b). Because any one of these events is incompatible with the observed data, runs of this type were not used for calculating the statistics of interest.

After time *T* is reached in the absence of premature coalescence, coalescent events proceed at a rate determined by the autosomal rate, and no further recombination events are allowed, because the genes now all form part of the same population. As before, however, records are still kept of the numbers of genes that derived from the original set of neo-*X* or neo-*Y* genes and of the numbers that have experienced a coalescent event with a gene in a given category.

RESULTS

Patterns of DNA sequence variability: Sequences of an 884-bp region of the *Adh* gene were obtained for 6 neo-*X* and 5 neo-*Y* chromosomes from the laboratory strains of *D. a. americana*, 19 neo-*X* and neo-*Y* chromosomes from the recently collected *D. a. americana* G96 sample, and 10 autosomes from the recently collected *D. a. texana* LP97 sample. A total of 41 variable nucleo-

tide sites were identified upon comparison of the 884-bp sequences from these 59 chromosomes, and these are presented in Figure 4. Of the 41 variable sites, 8 occur in the 130 bp of noncoding sequence, 32 are present at silent sites of 252 codons, and 1 is a replacement substitution. The one replacement substitution is a threonine-to-isoleucine replacement on the neo-*Y* chromosome of G96.40. The variable sites are distributed very evenly among the chromosomes (Figure 4), indicating very little haplotype structure in the samples and suggesting a historically high level of recombination in this region. There are 50 different haplotypes represented by the 59 sequences, and six instances where identical sequences were obtained from two different chromosomes: D.am.6X/G96.12X, D.am.0Y/LP97.02, G96.14Y/G96.30Y, D.am.0X/G96.06Y, G96.21Y/L97.08, and G96.03Y/G96.48Y. In another case, four different chromosomes, D.am.3X, D.am.7X, G96.13X, and G96.45X, share the same sequence. A total of 21 haplotypes are present on 25 neo-*X* chromosomes, 22 haplotypes are present on 24 neo-*Y* chromosomes, and 10 haplotypes on 10 autosomes of *D. a. texana*. Haplotype structure in the samples of neo-*X* and neo-*Y* chromosomes of *D. a. americana* and autosomes of *D. a. texana* is, therefore, very similar.

Another case where identical sequences were obtained involves the neo-*X* and neo-*Y* chromosomes of G96.31. This is the only line (out of 18) from the G96 sample where the F$_1$ female and male share the same sequence. Under the experimental methods that were followed to obtain the sequences, identical sequences can be obtained from F$_1$ males and females if the *X/4* fusion is not present in the *D. a. americana* male used in the original cross with *D. virilis*. Given the low haplotype structure observed in the samples, it is improbable that a neo-*X* and neo-*Y* from the same strain would share the same sequence, so it seems likely that the *X/4* fusion was not present in the parental male of G96.31 used for this cross. Cytological examination of the G96.31 line confirmed that the *X/4* fusion is polymorphic in this line; therefore, these sequences were excluded from the analyses. The line was established from a wild-caught pair, and apparently one of the three *X* chromosomes present in the two individuals used to establish the line was not fused to the fourth chromosome. Our sequence and cytological analyses indicate that the other 22 lines of *D. a. americana* are fixed for the *X/4* fusion chromosome. If the unfused *X* chromosome in G96.31 represents a low-level polymorphism, the frequency of the unfused *X* chromosomes in the population at Gary, Indiana is apparently <1.5%. Cytological analyses confirmed the absence of the *X/4* fusion chromosome in the LP97 sample of *D. a. texana*.

On the basis of the DNA sequences that were obtained, the different samples are very similar with respect to the variable nucleotide sites present in each. No fixed differences were found between any of the

```
position        1111112222333333344444445555556666777777
position        2581556893991224468004579044690229044788
position        4252701946228826694503812324753216263551 7
Noncode         I===III======================IIII======
Synonyms        =SSS===SSSSSS=SSSSSSSSSSSSSSSSS====SSSSSSS
Replacmnt       ===========R============================
Ancestral       GCTTTTAGCCGGGCGCCTGAGGCCCTCTCGCACTCCCAGCG

D.am.0X         ACTTTTAGCAGGGCGCCTGGGGCCCTCCCGCACTCCCAGCG
D.am.1X         G--A----------------------T---------T-T-
D.am.6X         G--A----------------------T-----T------
D.am.9X         GTC-------------------------------------
D.am.3X         ----------A------------TT-C-------C-T---T-
D.am.7X         ----------A------------TT-C-------C-T---T-

G96.03X         ---A--------A-T-A-----------------------
G96.36X         ---A------CA----------------------------
G96.40X         ---A------T-----------------T-----T------
G96.14X         --CA----T-------------------T-------T--
G96.12X         G--A----------------------T-----T------
G96.48X         G--A-GT----------A-A--------------------
G96.46X         G-CAC----------------------------------T-
G96.30X         G-C---------A-T-------T---------A------TA
G96.23X         -----------AA-----A------C-----G---------
G96.06X         -----------AA-----A------C----------T----
G96.24X         -----------A-TT------T-------------------
G96.33X         ------------------------C-T-------------
G96.21X         ------------------------C---------T----
G96.27X         --------------------CT--------G-------T-
G96.38X         --------------------CT-------T------
G96.11X         ----------A-----------TT-C-T------------
G96.13X         ----------A------------TT-C-------C-T---T-
G96.45X         ----------A------------TT-C-------C-T---T-
G96.10X         ---------T-----------------T-------------

D.am.9Y         G---------------------------------------
D.am.1Y         G-CA----T-------------T-----------------
D.am.6Y         G--A-----CA---------------C-------------
D.am.3Y         ---A----------------------T---------------T-
D.am.0Y         ---A-GT-----------------C-------------T-

G96.21Y         --C------C----------------C-------------
G96.03Y         --C-------------------------------------
G96.47Y         --C-----------------T-------------------
G96.41Y         --C-------------------C-T---------------
G96.38Y         --C---------A---------T-----------------
G96.36Y         -----------A-T---------------------T----
G96.27Y         ----------------T----------T-------------
G96.06Y         ----------------------------------------
G96.33Y         ------------------------C---------------
G96.13Y         ---------------------T--------A------T-
G96.48Y         --C-------------------------------------
G96.10Y         G-CA------------------C-T---------T-T-
G96.40Y         G--A---------T-------------TA---------T-
G96.45Y         G--A--------------A----TC-------------T-
G96.11Y         ---A------------------C-------C-T---T-
G96.14Y         ---A-----------------------------------T-
G96.30Y         ---A-----------------------------------T-
G96.12Y         ---A------------T-----------------T----T-
G96.46Y         ---A--------A-T-------T---------------T-T-

LP97.01         -----------AA-----A--------------------
LP97.03         -----------A-----------------T--------T-T-
LP97.05         -------C--------T-----------T----------T-
LP97.08         --C------C---------------C---------------
LP97.12         ---A------------------T----C-----------T-T-
LP97.13         ---A----C-----T----------C-T------------
LP97.02         ---A-GT------------------C-------------T-
LP97.09         ---A-----C----------------------T----
LP97.07         G--A----T-----------------C---------T----
LP97.04         ---A----T-------T--------C-----G-------T-
```

Figure 4.—Nucleotide polymorphisms in the *Adh* gene region of *D. a. americana* and *D. a. texana*. Positions of the variable nucleotide sites are presented relative to the sequenced region. Changes are indicated as noncoding (I), synonymous (S), or replacement (R). The ancestral state at each variable nucleotide position is based on sequences from *D. virilis* and *D. lummei* (Nurminsky *et al.* 1996). Complete sequences have been deposited in GenBank under accession nos. AF136650–AF136708.

TABLE 1

Average and net number of pairwise sequence differences at *Adh*

	Lab-X	G96-X	G96-Y	Lab-Y	LP97
Lab-X	6.93	7.02	5.91	5.87	7.03
G96-X	−0.24	7.59	6.24	6.39	7.08
G96-Y	−0.03	−0.03	4.95	5.13	5.72
Lab-Y	−0.20	−0.01	0.05	5.20	5.84
LP97	0.37	0.09	0.04	0.04	6.40

Average pairwise differences within samples are along the diagonal, average pairwise differences between samples are above the diagonal, and net number of nucleotide differences between samples are below the diagonal.

samples, and only two nucleotide site variants were identified that were present in more than one sequence but limited to a single sample (positions 196 and 451 in Figure 4). The divergence between samples was quantified using *d*, a measure of the mean number of net nucleotide differences between a pair of populations (Nei 1987). As can be seen in Table 1, comparisons between samples of *D. a. americana* generally yield negative values for *d*, which indicates there is a greater average number of differences within samples than between samples. The comparisons between the samples of *D. a. americana* and *D. a. texana* have positive values for *d*, but these are very small. Given that the variance of *d* is large when its value is small (Takahata and Nei 1985), inferences regarding the relationships among these samples are unreliable. What is clear from these comparisons, however, is that there is no quantitative divergence between the neo-*X* and neo-*Y* chromosomes of *D. a. americana.* Furthermore, there is no divergence between either of these chromosomes and the homologous locus in *D. a. texana.*

Sequence diversity per nucleotide site within each sample was estimated by the standard measures θ, based

on the number of segregating sites, and π, based on average pairwise differences among sequences (Tajima 1993). The laboratory strains represent individual isolates from throughout the range of *D. a. americana* and provide a standard for estimating the total variability within the subspecies, whereas the G96 sample was obtained from multiple flies collected at a single locality. As is shown in Table 2, measures of variability for both the neo-*X* and neo-*Y* are relatively consistent between the established laboratory strains of *D. a. americana* and the recently collected G96 sample. The estimates of π, however, are in closer agreement than those for θ. For example, π is 0.0056 and 0.0059 for the G96 and Lab samples of the neo-*Y*, whereas θ is 0.0071 and 0.0060, respectively (Table 2). The comparison between the Lab and G96 samples reveals that nucleotide diversity in the population at Gary, Indiana is similar to that present in the subspecies as a whole, indicating there have been no recent bottlenecks that have influenced nucleotide diversity in the population at Gary.

There is reason to suspect that sequence variability at *Adh* on the neo-*X* chromosome of *D. a. americana* may have been recently affected by selection. The *X/4* fusion was presumably derived through a single mutational event. If some form of directional selection was responsible for the increase in frequency of this chromosome, sequence variation at *Adh* may have been affected. Because this locus is located relatively close to the centromere on the *X/4* fusion chromosome, current variation at this locus would reflect the frequency of recombination between this locus and the centromere, and the time required for fixation of the *X/4* fusion (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992). Both the Tajima (1989) and Fu and Li (1993) statistical tests were used to examine the pattern of sequence variation in the G96-X sample for significant departure from neutrality. Although θ was greater than π for the G96-X sample, yielding a Tajima's *D* of −0.88 and a Fu and Li's *D* of −0.59, both test statistics

TABLE 2

Nucleotide diversity at the *Adh* locus in *D. a. americana* and *D. a. texana*

		Segregating sites		Total		Silent			
Sample	N	Silent	Replacement	π/site ± SE	θ/site ± SE	π/site ± SE	θ/site ± SE	Tajima's *D*	Fu and Li's *D*
Lab-X	6	15	0	0.0078 ±0.0043	0.0074 ±0.0039	0.0219 ±0.0120	0.0207 ±0.0108	0.338	0.167
G96-X	19	34	0	0.0086 ±0.0042	0.0110 ±0.0041	0.0239 ±0.0117	0.0307 ±0.0115	−0.878	−0.586
G96-Y	19	22	1	0.0056 ±0.0029	0.0071 ±0.0028	0.0153 ±0.0078	0.0190 ±0.0075	−0.828	−0.300
Lab-Y	5	11	0	0.0059 ±0.0034	0.0060 ±0.0034	0.0164 ±0.0096	0.0167 ±0.0094	−0.109	0.105
LP97	10	21	0	0.0072 ±0.0037	0.0084 ±0.0038	0.0202 ±0.0105	0.0234 ±0.0105	−0.650	−0.643

are well within the limits of the neutral model (the α = 0.05 value for Tajima's D is -1.80 and for Fu and Li's D is -1.89). Also, selection on the $X/4$ fusion would be expected to cause a reduction in sequence variation at *Adh*, but the observed variation was very high in the G96-X and Lab-X samples, $\sim$14% greater than the observed diversity at *Adh* on the autosomes of *D. a. texana* (Table 2). Nucleotide diversity at *Adh* on the neo-*X* chromosome is also relatively consistent with the levels of silent nucleotide variability observed at other loci in the genome of *D. a. americana* and *D. a. texana* (Hilton and Hey 1996, 1997; B. F. McAllister, unpublished data).

Nucleotide variability on the neo-*Y* of *D. a. americana* would not have been influenced directly by the fixation of the $X/4$ fusion chromosome. The population of neo-*Y* chromosomes currently present in *D. a. americana* represents a subpopulation of the freely segregating autosomes that was present before the fusion of this element with the *X* chromosome, but these chromosomes have now been "captured" by the $X/4$ fusion. However, the selective sweep model of *Y*-chromosome degeneration (Rice 1987) predicts that the pattern of sequence diversity at *Adh* on the neo-*Y* chromosome would be distorted by the bottleneck in population size caused by a selective sweep (Braverman *et al.* 1995; Simonsen *et al.* 1995). Negative values for the *D* statistics were generally observed for the two samples of neo-*Y* chromosomes (Table 2). For the G96-Y sample, Tajima's D was -0.83 and Fu and Li's D was -0.30; both are consistent with neutrality. On the basis of these statistical tests, there is no evidence for any selective sweeps having occurred on the neo-*Y* chromosome of *D. a. americana.*

Although the G96-X and G96-Y sequences were obtained from the same sample and primarily from the same strains of flies, there is a striking difference in the level of nucleotide variability present in these two chromosomal populations. Nucleotide diversity on the neo-*Y* is $\sim$65% (with a 95% CI of $\sim\pm$27%) of that observed on the neo-*X*, regardless of whether the comparison is based on π or θ (Table 2). The average pairwise difference among the neo-*X* sequences from the G96 sample is 7.58 bp for the entire 884-bp region, whereas the neo-*Y* sequences from the G96 sample have an average pairwise difference of 4.95 bp. The sequences from these two G96 samples were used in a resampling scheme to determine if the observed 2.64-bp difference in nucleotide diversity between the neo-*Y* and neo-*X* chromosomes is statistically significant (see materials and methods). We investigated the probability of observing a greater difference in the amount of nucleotide diversity than is observed for the G96-X and G96-Y sequences by repeatedly reconstructing and comparing two random samples of these sequences. Two out of 1000 replicates were obtained with a difference between the random samples that is greater than or equal to the observed 2.64-bp difference in mean pairwise diversity between the G96-X and G96-Y samples. This result pro-

vides strong statistical support ($\alpha \leq 0.002$) for rejecting the null hypothesis that these two samples of alleles were obtained from a single homogenous population, thus indicating that average nucleotide diversity on the neo-*Y* is significantly lower than that on the neo-*X.*

Results of coalescent simulations: As described in materials and methods, coalescent simulations were performed to investigate the following questions: (1) the value of the time T since the origin of the neo-*X* and neo-*Y* chromosomes and (2) the effective population size of the neo-*Y*, which is inversely proportional to t_y. There are several features of the data that need to be reproduced with reasonable probability by values of these parameters: (i) the relatively low proportion of neo-*Y* chromosome variants that are unique to this subpopulation; (ii) the relatively high proportion of unique neo-*X* chromosome variants; (iii) the high proportion of singletons among the unique neo-*Y* chromosome variants; (iv) the low diversity on the neo-*Y* chromosome compared with the neo-*X*; and (v) the lack of fixed differences between the neo-*Y* and neo-*X* chromosomes.

To ask these questions, the simulation program determined the following statistics, conditioning an assumed values of the scaled mutation rates and recombination rates described in materials and methods: (i) the proportion (P_{uy}) of runs in which the proportion of neo-*Y* chromosome variants unique to the neo-*Y* chromosome was equal to or less than the proportion observed; (ii) the proportion (P_{ux}) of runs in which the proportion of neo-*X* chromosome variants unique to the neo-*X* chromosome was equal to or greater than the proportion observed; (iii) the proportion (P_s) of runs in which the proportion of singleton variants among variants unique to the neo-*Y* chromosome was equal to or greater than that observed; (iv) the proportion (P_y) of runs in which the fraction of all segregating sites represented by neo-*Y* chromosomal variants was equal to or less than that observed; and (v) the proportion (P_f) of runs in which the number of fixed differences between the neo-*X* and neo-*Y* was equal to or less than the number observed. In addition, P_p, the proportion of runs in which "premature coalescence" occurred (see materials and methods and Figure 3b), was recorded. Values of P_f and P_p are not shown in the tables, because P_f is always close to one and P_p is always zero or small, unless recombination is zero and T is large.

As can be seen from Table 3, P_{uy} and P_s tend to decline with increasing T when R_y is zero, and P_y declines when t_y is larger. Recombination is expected to reduce P_{ux} and P_y and to increase P_s, as was observed in the simulation results. An overall assessment of the extent to which a given parameter set is compatible with the observed variants segregating at silent sites in the samples of neo-*Y* and neo-*X* alleles can be obtained by calculating the proportion (P_c) of runs in which the criteria used

TABLE 3

Results of coalescent simulations

T	t_y	P_{uy}	P_{ux}	P_s	P_y	P_c
$M = 3.5,\ R_y = 0$						
0.10	0.25	0.094	0.736	0.142	0.478	0.021
	0.10	0.207	0.907	0.192	0.854	0.100
	0.01	0.477	0.997*	0.403	1.000*	0.372
0.20	0.25	0.028	0.903	0.071	0.620	0.006
	0.10	0.090	0.985	0.114	0.929	0.047
	0.01	0.322	1.000*	0.261	0.995*	0.250
0.30	0.25	0.010	0.963*	0.057	0.657	0.004
	0.10	0.053	0.995*	0.076	0.938	0.025
	0.01	0.215	0.999*	0.162	0.981*	0.150
$R_y = 10$						
0.10	0.25	0.222	0.340	0.370	0.209	0.023
	0.10	0.431	0.522	0.372	0.513	0.089
0.20	0.25	0.209	0.336	0.352	0.208	0.019
	0.10	0.433	0.546	0.384	0.530	0.117
0.50	0.25	0.219	0.329	0.384	0.204	0.026
	0.10	0.411	0.563	0.404	0.545	0.119
1.0	0.25	0.194	0.336	0.402	0.217	0.018
	0.10	0.409	0.554	0.401	0.531	0.113
5.0	0.25	0.197	0.363	0.364	0.250	0.031
	0.10	0.369	0.617	0.416	0.578	0.120
$R_y = 100$						
0.1	0.25	0.303	0.095	0.631	0.058	0.008
	0.10	0.469	0.110	0.544	0.117	0.022
5.0	0.25	0.297	0.147	0.632	0.090	0.013
	0.10	0.307	0.189	0.614	0.152	0.022
$M = 5,\ R_y = 10$						
0.1	0.25	0.186	0.299	0.304	0.179	0.012
	0.10	0.414	0.5291	0.328	0.513	0.075
5.0	0.25	0.189	0.367	0.314	0.226	0.018
	0.10	0.220	0.376	0.365	0.260	0.091
$R_y = 100$						
0.1	0.25	0.291	0.070	0.616	0.041	0.007
	0.10	0.461	0.093	0.518	0.090	0.011
5.0	0.25	0.268	0.137	0.632	0.097	0.014
	0.10	0.368	0.182	0.572	0.131	0.018

T is the time since the origin of the fusion (in units of the autosomal coalescent time); R_y is the rate at which a neo-*Y* allele moves to the neo-*X* chromosome, measured on the time-scale of the autosomal coalescent process ($2N_e$ generations); t_y is the coalescent time for genes on the neo-*Y* relative to the autosomal coalescent time; P_{uy} is the fraction of runs that yielded a proportion of neo-*Y* variants unique to the neo-*Y* as small as or smaller than observed; P_{ux} is the fraction of runs that yielded a fraction of neo-*X* variants unique to the neo-*X* as large as or larger than observed; P_s is the fraction of runs in which there were a proportion of singleton variants among the unique neo-*Y* variants as large as or larger than observed; P_y is the fraction in which the ratio of the number of segregating sites on the neo-*Y* to the total for both chromosomes was as small or smaller than observed; P_c is the fraction of runs in which all of the criteria for P_{uy}, P_{ux}, P_s, and P_y were met. Note that P_{uy}, P_{ux}, P_s, and P_y were calculated only from runs where there was no premature coalescence (see materials and methods). * The probability of the observed value of the sample statistic is <0.05.

to establish the values of P_{uy}, P_{ux}, P_s, and P_y are all met, and when premature coalescence did not occur. If $P_c \leq 0.05$, we may consider the parameter set in question to be incompatible with the data at the 5% probability level. While this criterion is *ad hoc*, it seems to capture most features of the data that are relevant to the parameters of interest and is computationally straightforward. We chose to use the number of segregating sites as the basis for our test statistics, rather than pairwise difference measures, because these have superior statistical properties in the context of subdivided populations (Wakeley 1997).

The top section of Table 3 shows the results for the case when no recombination between the neo-*X* and neo-*Y* is allowed, corresponding to a state of complete fixation of the *X*/4 fusion and no introgression from *D. a. texana*. Most of the runs were done with an *M* value of 3.5; this corresponds to a θ value of 7 for the *Adh* silent sites, which is in the middle range of the values estimated for the sum over all silent sites for the samples of *americana* neo-*X*, neo-*Y*, and *texana* alleles. Sample sizes of 25 and 24 were used for the neo-*X* and neo-*Y* alleles, respectively, corresponding to the pooled data for the *americana* sequences. A coalescent time of 0.75 for the *X* chromosome relative to the autosomes was used, corresponding to the case of no sexual selection (Charlesworth 1994; Caballero 1995; Nagylaki 1995). One thousand runs were performed for each parameter set. The simulation results were compared with the observed number of three segregating sites unique to the neo-*Y*, two of which were singletons, and a value of 0.409 for the proportion of all variants that are found on the neo-*Y*.

With no recombination, it is not hard to account for the proportion of variants that are unique to the neo-*Y* chromosome sample, except when *T* is fairly large (≥ 0.2), unless t_y is very small. The frequency of neo-*Y* singletons is also sensitive both to *T* and the effective population size of the neo-*Y* chromosome subpopulation; the region of parameter space for which $T \geq 0.30$ and $t_y \geq 0.10$ is ruled out by this statistic (P_s) alone. The proportion of neo-*X* variants that are unique to the neo-*X* becomes significantly too large for large *T* or small t_y. The observed value of the fraction of all variants that are found on the neo-*Y* chromosome is compatible with a wide range of values of *T* and t_y, although P_y becomes >0.95 for small t_y and large *T*, which implies that the observed value is significantly too large for these parameters. The observed number of fixed differences is also compatible with the data for the range of parameters shown, although P_f increases as *T* decreases. Overall, only $T \leq 0.10$ and $t_y = 0.1$ are compatible with the data. This remains true even if widely different *M* values are used (data not shown). One possibility is that strong sexual selection may cause the coalescent time for *X*-linked genes relative to autosomal genes to be much higher than the value of 0.75 assumed in Table 3. This

was tested by runs in which t_x was set to 1, but this made only a minor difference to the results.

Given the evidence that some X chromosomes in *D. a. americana* are unfused, it is important to investigate the consequences of recombinational exchange between the centromere and *Adh.* Such recombination can occur in females that lack the fusion, which may result from incomplete fixation of the fusion in *D. a. americana* or introgression from *D. a. texana.* This would result in a transfer of alleles between the neo-X (fused fourth chromosomes) and neo-Y (unfused fourth chromosomes) subpopulations, at a rate determined jointly by the frequency of crossing over between the centromere and *Adh* in females and the frequency of unfused X chromosomes.

The results of simulations that include recombination are shown in the lower part of Table 3. The overall effect of recombination is to increase the frequency of runs that meet the test criteria, especially for large *T.* If an *M* value of 3.5 is assumed, the only cases in which the probability of meeting the criteria for acceptance are when $t_y < 0.25$. Values of *M* that are much smaller than 3.5 deviate significantly from the observed number of segregating sites in the neo-X chromosome sample (data not shown). Similarly, an *M* value of 5 gives a worse fit than $M = 3.5$. Somewhat counterintuitively, the best fit is obtained when $R_y = 10$ for a wide range of *T* values. Substantially higher R_y values lead to a lack of overall agreement with the data. No very precise conclusion concerning the likely value of *T* can thus be drawn from these results. A very high recombination rate ($\geqslant 100$), coupled with a large T (>1) is, however, inconsistent with the observed fraction of all segregating sites that are on the neo-Y chromosome, given the significantly lower diversity for the neo-Y chromosome than for the neo-X (see above). Overall, the results suggest the occurrence of a low frequency of exchange of *Adh* alleles between the neo-X and neo-Y chromosomes, with a significant reduction in the effective population size of the neo-Y chromosome to well below 25% of the autosomal value.

DISCUSSION

Sequence data from the *Adh* gene of *D. a. americana* reveal two important features of the neo-sex chromosomes: effectively no sequence divergence is present between the neo-X and neo-Y chromosomes, and sequence diversity is significantly lower on the neo-Y chromosome relative to the neo-$X.$ These findings have implications for the progression of sex chromosome differentiation as represented by the secondary sex chromosomes of *D. a. americana.* The neo-sex chromosomes of *D. a. americana* are currently in an incipient stage of differentiation, given the observed lack of divergence between the neo-X and neo-$Y.$ Substantial reduction in nucleotide diversity on the neo-Y relative to the

neo-X suggests that N_e for the neo-Y chromosome has been reduced well below the expected 25% of the autosomal value. We consider both of these points in turn.

Lack of divergence between neo-X and neo-Y alleles: This is the first study to provide direct evidence that the neo-sex chromosomes of *D. a. americana* share a high degree of sequence similarity. Because *Adh* is located on both the neo-X and neo-Y chromosomes, divergence at this locus provides a direct measure of the degree of separation between these two chromosomes, but even a low frequency of recombination between the neo-X and neo-Y chromosomes would retard this sequence divergence (Hudson 1990; Wakeley 1996; Nordborg 1997). The simulation results (Table 3) are consistent with the presence of rare recombination events: R_y values of ~ 10 provided a somewhat better fit to the observed sequence data than zero recombination, although the latter cannot be ruled out if the fusion is of recent origin ($T \leq 0.1$) and t_y is ≤ 0.1. However, a high rate of exchange ($R_y \gg 10$) seems to be ruled out.

The value of R_y is of the order of the product of the effective population size and the rate of exchange between the centromere and the neo-Y copy of *Adh.* Because silent site diversities at *Adh* and other loci (Hilton and Hey 1996, 1997; B. F. McAllister, unpublished data) suggest an autosomal effective population size of $>1,000,000$ in *D. a. americana*, the frequency of exchange in this chromosomal interval must be of the order of 1 in 100,000 or less. This is an average recombination rate for all potential sources of exchange between fourth chromosomes that are fused to the X and unfused fourth chromosomes. The simulation results therefore suggest the possibility of a cumulative effect of rare recombination events between neo-X and neo-Y chromosomes.

There are several possible sources of recombination events. Measured rates of recombination in male *D. virilis* are very low (Kikkawa 1935), but rare recombination events in males may be a source of exchange between neo-X and neo-Y chromosomes if the fusion has gone to complete fixation. The observed polymorphism of the $X/4$ fusion provides another opportunity for rare recombination events. Previous studies of *D. a. americana* indicate that the $X/4$ fusion chromosome is fixed throughout its range, with the exception of the relatively narrow hybrid zone (Hsu 1952; Throckmorton 1982). As already noted, a single line out of 22 collected from the population at Gary, Indiana is polymorphic for the $X/4$ fusion, suggesting that $<1.5\%$ of the X chromosomes in this population are not fused with chromosome *4.* Such a low frequency polymorphism of the $X/4$ fusion provides an opportunity for neo-Y chromosomes to undergo infrequent recombination when present in females heterozygous for the fusion. Because *Adh* is close to the centromere at cytological band 49B (Gubenko and Evgen'ev 1984; Nurminsky *et al.* 1996; Charlesworth *et al.* 1997a), and recombination is sup-

pressed near translocation breakpoints (Ashburner 1989), it may experience only a low frequency of exchange in such females, accounting for the low value of R_y that we have estimated. Loci that are located more distally on chromosome *4* should exhibit greater effects of recombination, because there would be a greater opportunity for crossing over when an unfused chromosome *4* was present in a female where exchange could occur. Sequence data from two additional regions of chromosome *4* are consistent with this prediction (B. F. McAllister, unpublished data).

While increasing the rate of recombination causes an increase in similarity for a given *T*, even with an extended amount of time, it does not do so in such a way that all criteria can be satisfied for an arbitrarily large *T*. This is understandable from the fact that recombination between the two chromosome classes behaves like a conservative migration process connecting two partially isolated subpopulations; use of equations (A1) of Nordborg (1997) shows that the difference in the equilibrium level of nucleotide diversities between two subpopulations connected by conservative migration is small in comparison with the divergence between the subpopulations, even if they differ substantially in size. Inclusion of recurrent gene flow from *texana* into *americana*, another possible source of unfused fourth chromosomes, would behave in a similar way to recombination. It is therefore essentially impossible to produce both a lack of divergence between the two chromosome classes and a reduction in diversity within the neo-*Y* chromosomes, with very large values of both *T* and recombination rates.

The only known difference between *D. a. americana* and *D. a. texana* is the *X*/*4* chromosomal fusion, suggesting a relatively recent origin of this chromosomal arrangement, especially as it is also absent from close relatives such as *D. novamexicana* and *D. virilis.* The maintenance of a steep cline between the fused arrangement of *D. a. americana* and the unfused arrangement of *D. a. texana* (Patterson and Stone 1952; Throckmorton 1982) indicates the operation of selection across the hybrid zone against immigrant karyotypes. The probable close linkage of the *Adh* gene to the centromere of chromosome *4* suggests that sequence divergence at *Adh* could have proceeded between the two subspecies even in the face of gene flow, due to the elimination of immigrant genotypes, if sufficient time had elapsed since the origin of the fusion (Barton 1986; Barton and Bengtsson 1986; Charlesworth *et al.* 1997b). The lack of such divergence suggests that *T* may in fact be of the order of the coalescence time or less. Absence of sequence divergence between *D. a. americana* and *D. a. texana* has also been observed for the *period* gene (Hilton and Hey 1996) on the distal end of the *X* chromosome, the *oskar* gene on chromosome *2* (Hilton and Hey 1997), and two separate genes, *seven in absentia* and *transformer*, on the distal end of chromosome *3* (B. F. McAllister, unpublished data). This consistency in the lack of sequence differentiation documented for *Adh* and the other loci suggests that the recent separation between the two taxa has had a greater impact than current gene flow on the distribution of nucleotide site variants between *D. a. americana* and *D. a. texana.*

Reduction of diversity on the neo-*Y* chromosome: The finding that sequence diversity is reduced 34% on the neo-*Y* chromosome relative to the neo-*X* is very interesting in light of the evidence for possible ongoing recombination between the neo-*Y* and neo-*X* chromosomes. Established sex chromosomes should have a *Y* chromosome with $\sim$33% of the sequence diversity present on the *X*, unless strong selection greatly increases the variance in male reproductive success relative to that in females (Charlesworth 1994; Caballero 1995; Nagylaki 1995). The coalescent simulations suggest that the best fit to the data is provided by a model in which N_e for the neo-*Y* is reduced to 10% or less of the autosomal value. The facts that the neo-*Y* chromosomes have few unique variants and that there is no evidence for a drastic loss of variability or distortion of the allele frequency spectrum, militate against any selective sweeps having occurred in the neighborhood of the *Adh* locus as postulated by Rice (1987) in his model for the degeneration of *Y* chromosomes. The reduction in N_e for the neo-*Y* chromosome must therefore have been caused by either strong sexual selection, background selection (Charlesworth *et al.* 1993), or temporally fluctuating selection (Gillespie 1994, 1997), because these would not necessarily produce such effects.

This study of *D. a. americana* provides an initial assessment of the early processes involved in *Y*-chromosome degeneration. The possibility of a low frequency of recombination between the neo-*Y* and neo-*X* chromosomes is consistent with previous analyses indicating that the neo-*Y* chromosome is still capable of functioning in the homozygous condition (Charlesworth *et al.* 1997a) and that the neo-*X* chromosome does not show signs of dosage compensation (Bone and Kuroda 1996; Marin *et al.* 1996). The loss of sequence diversity on the neo-*Y* is currently the only apparent consequence of partial genetic isolation between these chromosomes, indicating a strong reduction in effective population size of the neo-*Y* chromosomes as postulated in standard models of *Y*-chromosome degeneration (Charlesworth 1996; Rice 1996). Additional sequence data are needed to confirm these conclusions.

232 B. F. McAllister and B. Charlesworth

LITERATURE CITED

Ashburner, M., 1989   *Drosophila, A Laboratory Handbook.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Barton, N. H., 1986   The effects of linkage and density-dependent regulation on gene flow. Heredity **57:** 415–426.

Barton, N. H., and B. O. Bengtsson, 1986   The barrier to genetic exchange between hybridising populations. Heredity **57:** 357–376.

Bone, J. R., and M. I. Kuroda, 1996   Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila. Genetics **144:** 705–713.

Braverman, J. M., R. H. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995   The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

Bull, J. J., 1983   *Evolution of Sex Determining Mechanisms.* Benjamin/Cummings, Menlo Park, CA.

Caballero, A., 1995   On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics **139:** 1007–1011.

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

Charlesworth, B., 1994   The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

Charlesworth, B., 1996   The evolution of chromosomal sex determination and dosage compensation. Curr. Biol. **6:** 149–162.

Charlesworth, B., and D. Charlesworth, 1997   Rapid fixation of deleterious alleles can be caused by Muller's ratchet. Genet. Res. **70:** 63–73.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993   The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Charlesworth, B., D. Charlesworth, J. Hnilicka, A. Yu and D. S. Guttman, 1997a   Lack of degeneration of loci on the neo-Y chromosome of *Drosophila americana americana.* Genetics **145:** 989–1002.

Charlesworth, B., M. Nordborg and D. Charlesworth, 1997b   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., B. Charlesworth and M. T. Morgan, 1995   The pattern of molecular variation under the background selection model. Genetics **141:** 1619–1632.

Fisher, R. A., 1935   The sheltering of lethals. Am. Nat. **69:** 446–455.

Fridolfsson, A. K., H. Cheng, N. G. Copeland, N. A. Jenkins, H. C. Liu *et al.*, 1998   Evolution of the avian sex chromosomes from an ancestral pair of autosomes. Proc. Natl. Acad. Sci. USA **95:** 8147–8152.

Fu, Y.-X., and W.-H. Li, 1993   Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gethmann, R. C., 1988   Crossing over in males of higher Diptera (Brachycera). J. Hered. **79:** 344–350.

Gillespie, J. H., 1994   Alternatives to the neutral theory, pp. 1–17 in *Non-Neutral Evolution*, edited by B. Golding. Chapman and Hall, New York.

Gillespie, J. H., 1997   Junk ain't what junk does: neutral alleles in a selected context. Gene **205:** 291–299.

Graves, J. A. M., 1995   The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. BioEssays **17:** 311–321.

Gubenko, I. S., and M. B. Evgen'ev, 1984   Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica **65:** 127–139.

Hackstein, J. H., R. Hochstenbach, E. Hauschteck-Jungen and L. W. Beukeboom, 1996   Is the Y chromosome of Drosophila an evolved supernumerary chromosome? Bioessays **18:** 317–323.

Hey, J., and J. Wakeley, 1997   A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hilton, H., and J. Hey, 1996   DNA sequence variation at the *Period* locus reveals the history of species and speciation events in the *Drosophila virilis* group. Genetics **144:** 1015–1025.

Hilton, H., and J. Hey, 1997   A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. Genet. Res. **70:** 185–194.

Hsu, T. C., 1952   Chromosomal variation and evolution in the *virilis* group of *Drosophila.* Univ. Tex. Publ. **5204:** 35–72.

Hudson, R. R., 1990   Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, Vol. 7, edited by J. Antonovics and D. Futuyma. Oxford University Press, Oxford.

Hughes, R. D., 1939   An analysis of the chromosomes of the two sub-species *Drosophila virilis virilis* and *Drosophila virilis americana.* Genetics **24:** 811–834.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989   The "hitch-hiking effect" revisited. Genetics **123:** 887–899.

Kikkawa, H., 1935   Crossing-over in the male of *Drosophila virilis.* Cytologia **6:** 190–194.

Lahn, B. T., and D. C. Page, 1997   Functional coherence of the human Y chromosome. Science **278:** 675–680.

Marin, I., A. Franke, G. J. Bashaw and B. S. Baker, 1996   The dosage compensation system of Drosophila is co-opted by newly evolved X chromosomes. Nature **383:** 160–163.

Maynard Smith, J., and J. Haigh, 1974   The hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–35.

Muller, H. J., 1914   A gene for the fourth chromosome of *Drosophila.* J. Exp. Zool. **17:** 325–336.

Muller, H. J., 1918   Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. Genetics **3:** 422–499.

Muller, H. J., 1964   The relation of recombination to mutational advance. Mutat. Res. **1:** 2–9.

Nagylaki, T., 1995   The inbreeding effective population number in dioecious populations. Genetics **139:** 473–485.

Nei, M., 1970   Accumulation of nonfunctional genes on sheltered chromosomes. Am. Nat. **104:** 311–322.

Nei, M., 1987   *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nordborg, M., 1997   Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Nurminsky, D. I., E. N. Moriyama, E. R. Lozovskaya and D. L. Hartl, 1996   Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *Alcohol Dehydrogenase* gene. Mol. Biol. Evol. **13:** 132–149.

Ohta, T., 1992   The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **13:** 263–286.

Orr, H. A., and Y. Kim, 1998   An adaptive hypothesis for the evolution of the Y chromosome. Genetics **150:** 1693–1698.

Patterson, J. T., and W. S. Stone, 1952   *Evolution in the Genus Drosophila.* MacMillan, New York.

Powell, J. R., 1997   *Progress and Prospects in Evolutionary Biology, The Drosophila Model.* Oxford University Press, Oxford.

Rice, W. R., 1987   Genetic hitch-hiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics **116:** 161–167.

Rice, W. R., 1994   Degeneration of a nonrecombining chromosome. Science **263:** 230–232.

Rice, W. R., 1996   Evolution of the Y sex chromosome in animals. Bioscience **46:** 331–343.

Rice, W. R., 1998   Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome. Proc. Natl. Acad. Sci. USA **95:** 6217–6221.

Simonsen, K. L., G. A. Churchill and C. F. Aquadro, 1995   Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

Stalker, H. D., 1940   Chromosome homologies in two sub-species of *Drosophila virilis.* Genetics **26:** 575–578.

Steinemann, M., and S. Steinemann, 1998   Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. Genetica **102/103:** 409–420.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992   The effect of strongly selected substitutions on neutral polymorphism—analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

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

Tajima, F., 1993   Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.

Takahata, N., and M. Nei, 1985   Gene genealogy and variance of interpopulational nucleotide differences. Genetics **110:** 325–344.

Throckmorton, L. H., 1982   The *virilis* species group, pp. 227–296

in *The Genetics and Biology of Drosophila*, Vol. 3b, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, New York.

Tominaga, H., and S. Narise, 1995   Sequence evolution of the *Gpdh* gene in the *Drosophila virilis* species group. Genetics **96:** 293–302.

Wakeley, J., 1996   The variance of pairwise differences in two populations with migration. Theor. Popul. Biol. **49:** 39–57.

Wakeley, J., 1997   Segregating sites in Wright's island model. Theor. Popul. Biol. **53:** 166–174.

Communicating editor: W. F. Eanes