DOI: 10.1534/genetics.107.073932

Positive Selection Near an Inversion Breakpoint on the Neo-*X* Chromosome of *Drosophila americana*

Amy L. Evans, Paulina A. Mena and Bryant F. McAllister[1]

Department of Biological Sciences and the Roy J. Carver Center for Comparative Genomics, University of Iowa, Iowa City, Iowa 52242

Manuscript received March 28, 2007
Accepted for publication July 25, 2007

ABSTRACT

Unique features of heteromorphic sex chromosomes are produced as a consequence of sex-linked transmission. Alternative models concerning the evolution of sex chromosomes can be classified in terms of genetic drift or positive selection being the primary mechanism of divergence between this chromosomal pair. This study examines early changes on a newly acquired chromosomal arm of the *X* in *Drosophila americana*, which was derived from a centromeric fusion between the ancestral *X* and previously autosomal chromosome *4* (element B). Breakpoints of a chromosomal inversion *In(4)a*, which is restricted to the neo-*X*, are identified and used to guide a sequence analysis along chromosome *4*. Loci flanking the distal breakpoint exhibit patterns of sequence diversity consistent with neutral evolution, yet loci near the proximal breakpoint reveal distinct imprints of positive selection within the neo-*X* chromosomal class containing *In(4)a*. Data from six separate positions examined throughout the proximal region reveal a pattern of recent turnover driven by two independent sweeps among chromosomes with the inverted gene arrangement. Selection-mediated establishment of an extended haplotype associated with recombination-suppressing inversions on the neo-*X* indicates a pattern of active coadaptation apparently initiated by *X*-linked transmission and potentially sustained by intralocus sexual conflict.

HETEROMORPHIC sex chromosomes exhibit a characteristic set of visible and functional differences relative to each other and to the autosomes. Unique features of the sex-chromosome pair arise following the acquisition of the primary gene, or genes, controlling sex determination by an autosomal pair (see Vallender and Lahn 2004; Bachtrog 2006; and references therein). Two distinct viewpoints predominate among hypotheses pertaining to the mechanisms that transform autosomes into heteromorphic sex chromosomes. One regards sex chromosomes as passively responding to deleterious mutations, with genetic drift being a sufficient mechanism for functional decay of the nonrecombining *Y* chromosome (Charlesworth and Charlesworth 2000; Gordo and Charlesworth 2001). A sensational prediction arising from the passive point of view is extinction as the eventual fate of every *Y* chromosome (Graves 2000) and even the males that carry them (Sykes 2004). An alternative viewpoint regards sex chromosomes as actively diverging from each other via sexually antagonistic selection (Rice 1996, 1998). Under antagonism, uniform selection pressure to optimize male fitness shapes the *Y* chromosome. While sexual antagonism generally is regarded as an important force leading to reductions in recombination between pairs of sex chromosomes (Charlesworth *et al.* 2005b), the relative impact of passive divergence between sex chromosomes driven by genetic drift *vs.* active divergence driven by sexual antagonism remains unclear.

Studies of recently derived sex chromosomes provide important data for evaluating models of sex chromosome evolution. The neo-*Y* chromosome of *Drosophila miranda* has been the focus of extensive analyses that reveal evidence of rapid degeneration of gene function (Steinemann *et al.* 1993; Bachtrog 2005; Bartolomé and Charlesworth 2006). However, due to the absence of recombination for this and other *Y*-linked regions, the primary force driving gene loss is unknown, because deleterious mutations could accumulate through genetic drift or through successive hitchhiking events with independent positively selected alleles. Although the neo-*Y* of *D. miranda* appears to have been affected by hitchhiking, the current pattern of sequence diversity (or lack thereof) only reveals evidence of the latest episode of positive selection (Bachtrog 2004). In theoretical models and empirical studies of divergence between evolving sex chromosomes, there has been less attention given to the role of the *X*. However, new insights gained from microarray studies suggest that the

Sequence data from this article have been deposited with the EMBL/NCBI Data Libraries under accession nos. EU069267–EU069360 and EU072745–EU072916.

[1]*Corresponding author:* Department of Biological Sciences, 143 Biology Bldg., University of Iowa, Iowa City, IA 52242-1324.
E-mail: bryant-mcallister@uiowa.edu

X may be more involved in its divergence from the *Y* than previously appreciated (see Vallender *et al.* 2005 for commentary).

Genomewide expression analyses indicate a difference in the gene content of the *X* chromosome relative to the autosomes (reviewed in Oliver and Parisi 2004). The heteromorphic *X* chromosome of *D. melanogaster* is deficient in the number of genes with male-biased expression, and it also shows a slight increase in the number of genes with female-biased expression (Parisi *et al.* 2003). Deficiency in genes expressed at a higher level in males than in females is regarded as evidence of demasculinization, which implies that the autosome from which the *X* arose contained a normal number of male-biased genes and, following the onset of *X*-linked transmission, gene expression was neutralized between sexes. Enrichment of female-biased genes on the heteromorphic *X* is regarded as evidence of feminization. These changes in gene expression may result from intralocus sexual conflict where an allele has antagonistic effects on female and male fitness (Rogers *et al.* 2003; Wu and Xu 2003; Oliver and Parisi 2004; Vallender *et al.* 2005; Vicoso and Charlesworth 2006). In contrast to autosomes, regions of the *X* chromosome that do not recombine with the *Y* are overrepresented twofold in females relative to males. This bias favors the accumulation of female-beneficial mutations on the *X* chromosome (Rice 1984); thus unique expression patterns for the *X* may represent a phenotype of optimized female fitness.

The newly acquired arm of the *X* chromosome in *D. americana* provides a system for examining immediate changes in response to selection pressures arising from *X* linkage. This species is segregating a centromeric fusion between the *X* and chromosome *4* (element B) and the frequency of this *X–4* centromeric fusion is positively correlated with latitude over a broad geographic region (McAllister 2002). Chromosome *4* experiences sex-linked transmission as a result of the centromeric fusion (McAllister and Charlesworth 1999). The ancestral unfused arrangement of chromosome *4* exists as a transient neo-*Y* chromosome; males transmit the chromosome as a neo-*Y* in the presence of the *X–4* fusion and as an autosome in the presence of an unfused *X*. Existence as an autosome provides selection for maintaining gene function on the unfused arrangement of chromosome *4*, and a study by Charlesworth *et al.* (1997) revealed no evidence for functional degeneration. At the base of the euchromatic region of this chromosome, a curious pattern of sequence diversity has been revealed for the *Adh* gene where neo-*X* alleles exhibit higher diversity than the neo-*Y* alleles (McAllister and Charlesworth 1999) and neo-*Y* alleles exhibit higher diversity than autosomal alleles (McAllister and Evans 2006). Furthermore, sequence diversity along unfused chromosome *4* is correlated with the level of *Y* linkage, indicating that polymorphism is inflated through some form of balancing selection in response to sex-linked transmission (McAllister and Evans 2006). Hitchhiking with sexually antagonistic alleles may account for the observed pattern of nucleotide diversity. Two plausible sources of antagonistic alleles exist in this system: masculinizing alleles on the neo-*Y* or feminizing (and demasculinizing) alleles on the neo-*X*.

Divergence associated with an inverted gene arrangement suggests that the neo-*X* is a target of novel selection pressures in *D. americana*. Nested paracentric inversions, *In(4)ab*, are associated with a unique neo-*X* haplotype within the *big brain* (*bib*) gene, which is located in the noninverted region between the centromere and the proximal inversion breakpoint (McAllister 2003). Measurements of recombination in laboratory crosses demonstrated that the inverted arrangement of the neo-*X* suppresses meiotic exchange with the transient neo-*Y* (McAllister 2003; McAllister and Evans 2006), consistent with models of sexual antagonism between proto-sex chromosomes that predict the emergence of recombination suppressors reducing exchange of sex-linked alleles (Rice 1987). Inversions may ultimately cause stratification in levels of sequence divergence between unique regions of *X* and *Y* chromosomes, which has been shown for the heteromorphic sex chromosomes in humans (Lahn and Page 1999; Ross *et al.* 2005). The nested *In(4)ab* complex defines a newly isolated *X*-specific region of the *D. americana* genome that is overrepresented in females, thus providing a substrate for detecting the influence of novel selection pressures.

Although inversion heterozygosity effectively eliminates the detection of meiotic recombinants in the laboratory, the influence of inversions on population genetic differentiation is generally less pronounced. An inverted region is prone to gene flux due to gene conversion and/or double crossovers between inverted and standard chromosomes (Navarro *et al.* 1997). Estimates of gene flux among paracentric inversions reveal a rate of nucleotide exchange about two orders of magnitude greater than the mutation rate (Betrán *et al.* 1997; Schaeffer and Anderson 2005). Such a high rate of gene flux causes substantial homogenization of neutral sequence diversity between polymorphic chromosomal arrangements and, ultimately, it will impede divergence through nucleotide substitution, because the probability of exchange of a new neutral mutation between classes exceeds the probability of its fixation within a class (Navarro *et al.* 2000). In contrast to the inverted region, the breakpoints of an inversion define two alleles at different loci maintained in complete linkage disequilibrium. While the breakpoints evolve independently within each chromosomal arrangement, owing to meiotic recombination within homozygotes, the breakpoints are protected from gene flux with other arrangements. Loci immediately flanking each of the

inversion breakpoints are also likely to be protected from gene flux (Navarro *et al.* 1997) and therefore provide the best historical record of the inversion. Previous analyses of sequences within the *bib* gene in *D. americana* revealed evidence of selection (McAllister 2003), and although the exact breakpoints of the *In(4)ab* complex were unknown, the closeness of *bib* to the most proximal breakpoint implicated the inversion, and/or its contents, as a target of selection.

Exact determination of the breakpoints of *In(4)ab*, coupled with population genetic analyses of flanking regions, would provide a more informed test of selection pressures within this newly *X*-linked genomic region. Comparative genomic analyses designed within the context of reference genome sequences of Drosophila and utilizing comparative genomic resources developed for *D. americana* represent an efficient method for isolating and characterizing the breakpoints of the inverted region of chromosome *4* and identifying flanking gene regions. A close relationship makes *D. americana* particularly amenable to genomic analyses informed by the genome sequence of *D. virilis* (Spicer and Bell 2002; Caletka and McAllister 2004). To provide resources for assembling the Drosophila genome sequences and to facilitate comparative analyses of close relatives (Markow *et al.* 2003), BAC libraries have been constructed for the sequenced species and for an additional set of closely related species (http://www.genome.gov/10001852). Close relatives of *D. virilis* for which BAC libraries are available include *D. americana* and two additional members (*D. novamexicana* and *D. littoralis*) of the virilis species group. These resources provide unique opportunities for utilizing the genome sequence of *D. virilis* to guide genome-level analyses in the other species, such as *D. americana*.

This study leverages the physical anchoring of the *D. virilis* genome sequence on its polytene chromosomes to identify the breakpoints of the derived paracentric inversion, *In(4)a*, within the genome of *D. americana*. A strategy is described for isolating BAC clones containing the inversion breakpoints of *In(4)a*. Breakpoints at both ends of *In(4)a* are localized within small intergenic regions. Sequences of these intergenic regions on a chromosome with the *In(4)a* gene order reveal the presence of a shared dispersed repetitive element in which ectopic exchange apparently generated the inverted arrangement. Analysis of chromosomes from a natural population of *D. americana* demonstrates a complete association between *In(4)a* and the *X* chromosome mediated by the *X–4* centromeric fusion. Patterns of sequence diversity in gene regions near the inversion breakpoints reveal the influence of positive selection affecting the region surrounding the proximal breakpoint of the inverted arrangement. Moreover, the association of *In(4)a* with the neo-*X* chromosome has left an imprint of positive selection currently observed at the *bib* locus. Rapid change in the organization and nucleotide sequence of the neo-*X* through positive selection suggests a pattern of active coadaptation in response to incipient *X* linkage.

MATERIALS AND METHODS

Localization of inversion breakpoints: Based on the resolution of cytology, the standard gene order of chromosome *4* is homosequential in *D. americana* and *D. virilis* (Hsu 1952; B. F. McAllister and P. A. Mena, personal observation). Furthermore, the order is the same for 11 genetic markers within a 105-cM linkage map of the standard arrangement *D. americana* spanning ~75% of the polytene map from the centromere to *timeless* at 42E (McAllister and Evans 2006). By anchoring the assembled genome sequence of *D. virilis* using reference positions previously mapped along chromosome *4* (supplemental Table S1 at http://www.genetics.org/supplemental/ lists relevant reference points), a strategy was developed using targeted sequence analysis to localize the breakpoints of *In(4)a* in *D. americana*. Primer pairs were developed in the suspected vicinity of the *In(4)a* breakpoints using the corresponding genome sequence of *D. virilis*. For the proximal breakpoint region, putative orthologs of annotated genes in *D. melanogaster* were identified at ~200-kb intervals using the *bib* gene as a starting position and continuing to the distal end of scaffold 12723 (NCBI accession CH940654.1). At the distal-breakpoint region, a similar strategy was initiated in scaffold 12963 (NCBI accession CH940649.1) starting near an anchor point at cytological subdivision 43E and continuing in the proximal direction. For each pair of PCR primers, the *D. virilis* sequence was aligned with genomic and cDNA sequences of *D. melanogaster* to identify putative exon–intron boundaries and primers were designed ~1 kb apart in conserved regions of adjacent exons spanning a small intron. Supplemental Table S2 (http://www.genetics.org/supplemental/) lists the gene regions used in this analysis, primer sequences, and positions within the genome sequence of *D. virilis*.

On the basis of a coarse analysis of sequence differentiation between a small reference sample of inverted and standard chromosomes, six of the regions were selected as probes for screening a BAC library of *D. americana*. Printed filters of library Da_ABa were obtained from the Arizona Genomics Institute (Tucson, AZ). This library was constructed from a highly inbred strain (15010-0951.15, Tucson Drosophila Stock Center) of *D. americana* containing the *X–4* fusion arrangement and *In(4)ab*, which was verified by cytological analyses of polytene chromosomes. An unfused chromosome *4* with the standard gene order is present as a neo-*Y* chromosome in this line. PCR products from three regions near the proximal breakpoint (putative orthologs of Dmel\CG15828, Dmel\CG9171, and Dmel*DLP*) and three regions near the distal breakpoint (putative orthologs of Dmel*Trn-SR*, Dmel*dp*, and Dmel*Drp1*) were amplified from a standard laboratory line (NN97.4-red) containing *In(4)ab*. A Southern hybridization protocol from LI-COR Biosciences (Lincoln, NE) was followed for screening the BAC library. Images of hybridized probes were acquired by scanning with a LI-COR Odyssey infrared imager. A second library screen was performed with a single probe (putative ortholog of Dmel*raw*) located near the proximal breakpoint.

Positive clones were grown and DNA isolated following standard procedures for alkaline lysis. End sequences of clone inserts were obtained using standard T7 and BES_13R (a derivative of M13R; D. Kudrna, personal communication) primers. Edited and trimmed sequences were queried against the genome sequence of *D. virilis* in *blastn* searches using the

TABLE 1

Summary of regions analyzed on chromosome *4*

Region	Dmel ortholog	Alignment (bp)	Silent sites	*In(4)a* position	Distance to breakpoint
tim	*timeless*	516	117.1	Outside distal	3.44 Mb
15435	CG15435	706	252.6	Outside distal	433 bp
raw	*raw*	553	132.6	Inside distal	93.9 kb
8665	CG8665	779	291.6	Inside distal	913 kb
dp	*dumpy*	855	191.6	Inside proximal	91.8 kb
9171	CG9171	695	219.2	Outside proximal	242 kb
bib[a]	*big brain*	947	364.5	Outside proximal	643 kb
Adh[a]	*Alcohol dehydrogenase*	846	309.6	Outside proximal	2.09 Mb
18095	CG18095	669	195.1	Outside proximal	3.50 Mb
nmd	*no mitochondrial derivative*	653	148.7	Outside proximal	4.55 Mb

Regions contained within *In(4)a* are underlined.
[a] Sequence data only from McAllister (2003).

BLAST server available through FlyBase (Grumbling *et al.* 2006). Two clones, each containing a breakpoint region, were isolated. Representations of the corresponding genomic regions of *D. virilis*, which were initially oriented on the basis of end sequences, were examined using GBrowse available through FlyBase (Grumbling *et al.* 2006). Primer pairs were designed from putative coding regions in the genome sequence of *D. virilis* and used to screen the clone DNA as sequence tagged sites (STSs). Supplemental Table S3 (http://www.genetics.org/supplemental/) describes each primer pair. Presence and absence of STSs in the clones localized each breakpoint to an interval between adjacent genes in the genome sequence of *D. virilis*. Amplification across the breakpoints was achieved using the Expand Long Template PCR system (Roche, Indianapolis) and template DNA from the clones and from two inbred lines: NN97.4-red, containing *In(4)ab* and ML97.5-pur, containing the standard gene order. Products were cloned into pGEM T-Easy Vector (Invitrogen, Carlsbad, CA) and purified clone DNA was sequenced (ABI 3730 with BigDye chemistry) using primer walking to extend through each insert. The distal breakpoint in the strain used for construction of the BAC library contains an ~5.5-kb-derived insertion relative to strain NN97.4, so the complete sequence of the repeats in the insert of the BAC clone was not determined. Analyses of the sequences were performed with Sequencher (GeneCodes, Ann Arbor, MI) for contig assembly and ORF detection, MultiPIPMaker (Schwartz *et al.* 2000) for sequence alignments, and *blastn* for similarity searches on the genome sequence of *D. virilis* (using the FlyBase server, Grumbling *et al.* 2006) and other sequence data (using the NCBI server).

A set of PCR primers was developed for detecting the presence of the inverted or standard chromosomal arrangements on the basis of DNA sequence. One primer is located in the sequence outside of the proximal breakpoint and the other two are located inside and near each end of the inverted region. The three primer sequences are: 10728amF2, GAT ATG TTA CCG AGC TCC TT; 17840-bpF, CGA ACA ACT TAC CGA TCG TG; and bwa-bpR, CGC AGA ACA AAC ACG TCT G. With all primers combined in a single PCR using 60° annealing temperature and a 2.5-minute extension time, inversion status is evident by amplification of an ~1660-bp product from the inverted arrangement and/or an ~620-bp product from the standard arrangement.

Inversion polymorphism: Both standard and inverted arrangements of chromosome *4* are present in populations of *D. americana*. Frequencies of the two arrangements and their associations with the *X* chromosome were measured in a sample (IR) of wild-caught flies originally described by McAllister and Evans (2006). Male flies collected in 2004 from the IR locality (GPS: 41.779° N, 91.715° W) near Iowa City, Iowa, were mated individually in the lab with an inbred line (ML97.5-pur) of *D. americana* that has the standard gene order of chromosome *4*. Polytene chromosomal preparations from F_1 larvae, with gender identified from the size of the gonads, were obtained following standard methods (Kennison 2000). Presence of *In(4)ab* on the paternal chromosome was evident by the formation of overlapping inversion loops. Transmission of chromosome *4* in the wild-caught males was determined from genotypes at seven microsatellite loci (Gpdh, V68-86.1, V93-93, V68-62, V68-4, V68-74, and V71-6; Schlötterer 2000; McAllister and Evans 2006) in pooled DNA samples of six adult F_1 males and six adult F_1 females. Sex-limited alleles at each heterozygous microsatellite locus are indicative of an *X–4* fusion in the wild-caught male, whereas shared alleles between males and females are indicative of autosomal transmission of chromosome *4*. Each wild-caught male was also crossed with the V46 line of *D. virilis* to obtain F_1 progeny for sequence analysis.

The procedures described above determine the linear arrangement of a single chromosome *4* and its linkage relationship with the sex chromosomes, and furthermore, each chromosome is transmitted to a hybrid genetic background where species-specific primers can be used to directly obtain sequences along the entire chromosome as a single haplotype. Insufficient numbers of hybrid progeny with *D. virilis* were obtained from the wild-caught males, so these same procedures were applied to single males from different isofemale lines derived from individual females collected at the same locality to increase the number of characterized chromosomes for sequence analysis.

DNA sequence variation: Assays were developed for obtaining PCR product from the allele of *D. americana* in F_1 hybrid flies using species-specific primers that discriminate against the allele of *D. virilis*. Sequences from a reference set of inbred lines were aligned with the sequence of *D. virilis* and primer pairs were designed on the basis of conservation among sequences of *D. americana* in regions containing at least one discriminatory 3′ nucleotide relative to *D. virilis*. Table 1 lists the gene regions examined, lengths of the sequenced regions, and distances to the closest inversion breakpoint estimated from the assembled genome sequence of *D. virilis* (primer

sequences included in supplemental Table S2 at http://www.genetics.org/supplemental/). Standard methods were followed for PCR amplification, column purification, and direct sequencing of PCR product on an ABI 3730. Sequences were obtained for eight gene regions from a set of 35 chromosomes independently derived from the IR population.

Three chromosomal classes are present in the IR population: *X–In(4)a* is an inverted chromosome *4* fused with the *X* (we are not completely certain of the presence of *In(4)b* in all of these cases), *X–4std* is a standard chromosome *4* fused with *X*, and *Unf 4std* is a standard chromosome *4* independent of the *X*. Because all chromosomes used as templates for sequencing were obtained from the F_1 progeny of individual males, and each male has an unfused arrangement of chromosome *4*, whereas this arrangement is rare among females in this population, the number of sequences obtained from each chromosomal class is not representative of their frequencies in the population. Population samples representative of the chromosomal arrangement frequencies were constructed by random selection of chromosomes from each class (Hudson *et al.* 1994). A combined sample representing the neo-*X* contains an equal number ($n = 9$) of inverted and standard chromosomes, which corresponds with their nearly equal frequencies estimated from wild-caught males. A representative sample of the entire population was constructed on the basis of the estimated frequency of 97% fused *X* chromosomes and assuming a 1:1 sex ratio (McAllister and Evans 2006). This reconstructed random sample contains 25 chromosomes: 36% ($n = 9$) *X–In(4)a*, 36% ($n = 9$) *X–4std*, and 28% ($n = 7$) *Unf 4std*.

Determination of haplotype status at the *bib* locus for each IR chromosome was assessed by digestion with *Bbr*PI following previously described methods (McAllister 2003). Available sequences of the *Adh*, *bib*, and *tim* gene regions (McAllister 2003) were also included in the analyses. Arrangement of the chromosomes from which these sequences were obtained was determined by PCR analysis of the *In(4)a* proximal breakpoint. Since these sequences represent a combination of samples where the frequencies of the different chromosomal arrangements are not well characterized, only analyses based on chromosomal classification were performed.

Analyses of sequence alignments, with the othologous region from the genome sequence of *D. virilis* included as an outgroup, were performed using *DnaSP* ver. 4.10.4 and by independent analyses of its output (Rozas *et al.* 2003). Numbers of segregating mutations, the number that resulted in amino acid replacements, pairwise diversity (π) and heterozygosity (θ) at silent sites (Watterson 1975; Tajima 1983, respectively), haplotype diversity (Nei and Tajima 1981), and average linkage disequilibrium (Kelly 1997) were obtained for each sample. Statistics describing the distribution of variation were also obtained, including *D* (Tajima 1989) and *H* (Fay and Wu 2000). Statistical significance of observed values, testing fit with the neutral model, was obtained by coalescent simulation on the basis of the number of segregating sites and assuming no recombination. Polymorphism within each of the *X–4* fusion classes and divergence relative to *D. virilis* for each gene region was tested for homogeneity within the HKA framework (Hudson *et al.* 1987). In each case, sequences from the *tim* gene region from the same sample of chromosomes as the test locus were used to standardize polymorphism and divergence. Multiplicity issues affect the interpretation of individual neutrality tests reported here as type I error rates for single comparisons (Benjamini and Hochberg 1995); however, while statistical power for rejecting neutrality with error rates corrected for the overall experiment (*e.g.*, Bonferroni correction) was sacrificed by examining many loci and sample configurations with multiple measures, composite results guided inferences of selection.

Measures of divergence among each chromosomal class and for each chromosomal class of *D. americana* compared to *D. virilis* were obtained from the net number of nucleotide substitutions (Nei 1987). A weighted average of net substitutions per site over all sequenced gene regions was used to construct a distance matrix including each chromosomal class and *D. virilis*, and a neighbor-joining tree with estimated branch lengths was constructed in PAUP* version 4.0b10 (Swofford 2002).

No evidence of heterogeneity was detected between standard chromosomes that are either fused with the *X* or not; therefore, a set of analyses was performed contrasting inverted and standard chromosomes. For each gene region, sequence diversity was standardized by dividing total pairwise diversity within each class by the average number of nucleotide differences between *D. americana* compared to *D. virilis*. A standardized measure of substitution rate within the inverted and standard chromosomal classes was obtained by comparison with *D. virilis* using the following: $(f_i - f_{i+k})/f_{i+k}$, where f_i is the number of fixed differences between a sample (i) and the outgroup and f_{i+k} is the number of fixed differences between a combination of samples i and k compared to the outgroup. This estimates the proportion of unique fixed differences within a subsample (i) of sequences relative to the number of fixed differences observed for a larger sample ($i + k$), which increases due to substitution (including specific loss of ancestral variants) within the subsample. Net divergence and differentiation between the inverted and standard arrangement were measured using *DnaSP*, and a permutation test of homogeneity between classes measured by K_{st} was performed with 10,000 replicates (Hudson *et al.* 1992). The average value of absolute D' (Lewontin 1964) for nucleotide variants within each region relative to the inverted and standard arrangements was measured to reveal overall associations with the chromosomal forms. Segregating variants within each chromosomal class and differences between classes were identified as ancestral or derived using parsimony criteria (Charlesworth *et al.* 2005a).

Coalescent simulations modeling recovery from a population bottleneck after reducing to a single chromosome were used to estimate ages of monophyletic haplotypes. Simulations of haplotype origin were performed with *ms* (Hudson 2002) by reducing the population to the reciprocal of the effective size estimated by the scaled mutation parameter ($N\mu$) and assuming a mutation rate of 5.8×10^{-9} per silent site (Haag-Liautard *et al.* 2007). Assessment of fit between silent pairwise diversity in the observed data *vs.* 10,000 simulated samples was obtained by varying the time since the bottleneck (T) and estimating the interval of T where 95% of simulated datasets contained the observed value of π.

RESULTS

Characterization of the inversion breakpoints: Inspection of inversion loops formed in polytene chromosomes of inversion heterozygotes indicated the breakpoints of the large 4a inversion were near subdivisions 48C/D (proximal breakpoint) and near subdivision 44A (distal breakpoint) on the chromosomal map of *D. virilis* (Figure 1A). Breakpoints for the smaller nested inversion, *In(4)b*, are located near 45C and 44E. The *bib* gene was previously localized by *in situ* hybridization to 48E (McAllister 2002) and its position along with other physically mapped markers near each of the inferred locations of the inversion breakpoints provided

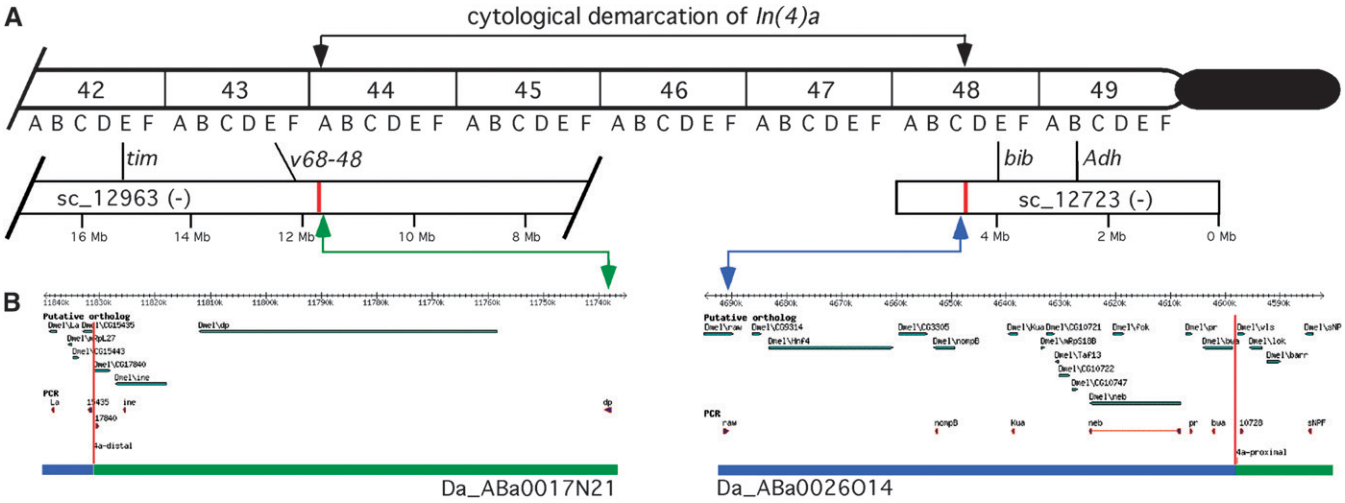

FIGURE 1.—Overview of *In(4)a* breakpoints isolated in BAC clones of *D. americana*. (A) Identified cytological positions of breakpoints anchored to scaffolds 12963 and 12723 of the assembled genome sequence of *D. virilis*. Additional mapped positions used in orienting the scaffolds are listed in supplemental Table S1 (http://www.genetics.org/supplemental/). (B) GBrowse image of annotated genome sequence of *D. virilis* in regions corresponding to the contents of BAC clones Da_ABa0017N21 and Da_ABa0026O14. Positions of probes *dp* and *raw*, which were used to isolate the BAC clones, are indicated by arrowheads connecting with genomic scaffolds. STS markers are annotated in the GBrowse image and delineated by the contents of the two BAC clones. Red lines in each panel identify the positions of the isolated breakpoints.

multiple points for associating polytene chromosomal position with the genome sequence of *D. virilis* (Figure 1A, supplemental Table S1 at http://www.genetics.org/supplemental/). This framework was used to develop PCR primers for obtaining rough estimates of sequence variability among a small reference set of inverted chromosomes and measuring sequence differentiation between the inverted and standard arrangements of *D. americana* (data not shown). The positions of the *In(4)a* breakpoints were approximately localized from this preliminary analysis of patterns of polymorphism and divergence.

The sequence analysis guided probe selection for screening a BAC library of *D. americana* constructed from a strain with the *In(4)ab* arrangement by the Arizona Genomics Institute. Three probes considered as being near each inversion breakpoint were combined and used to screen the BAC library. Fifty-five end sequences were obtained from 30 clones isolated from the library, and on the basis of the localization of these sequences within the genome sequence of *D. virilis*, one clone (Da_ABa0017N21) contained a single breakpoint from the inverted arrangement. This clone contained the region corresponding to the probe *dp*, which was used for screening the library, and one end sequence aligned in scaffold 12963 of *D. virilis* near the inferred position of the distal breakpoint, and the other end sequence aligned in scaffold 12723 near the inferred position of the proximal breakpoint. The other breakpoint was not evident among the clones isolated from the library in the initial screen, so a second screen was performed using a probe developed from the gene *raw* located in the region of the proximal breakpoint. A clone (Da_ABa0026O14) hybridizing with *raw* and containing the proximal breakpoint was identified. Using the organization of putative orthologs in the genome of *D. virilis* as sites for the development of STSs, the proximal breakpoint of *In(4)a* was located between *bwa* and *vls* in the upstream region of both genes, and the distal breakpoint was located between putative orthologs of Dmel\CG17840 and Dmel\CG15435, also in their upstream regions (Figure 1B). Primers anchored within the coding sequence of these genes were used to amplify the intervening sequence from clone DNA and genomic DNA containing inverted and standard gene arrangements.

Complete sequences of both breakpoint regions from the inverted and the standard gene arrangements were obtained and compared to reveal the structure of the sequence at the breakpoints (Figure 2; annotated sequences provided as supplemental material at http://www.genetics.org/supplemental/). A repetitive sequence is shared between both breakpoint regions of the inverted arrangement; however, this sequence is absent in the corresponding regions of the chromosome containing the standard arrangement and also of the genome sequence of *D. virilis*. The repeat sequence at the proximal breakpoint shows features generally associated with transposable elements. An 869-bp internal region is flanked by 240-bp terminal inverted repeats (Figure 2B). There is no evidence of an open reading frame within the sequence and no similar sequences were detected in searches against GenBank except in the genome of *D. virilis*, where many copies of the sequence are dispersed throughout the genome. In comparisons with other copies of the repeat, the

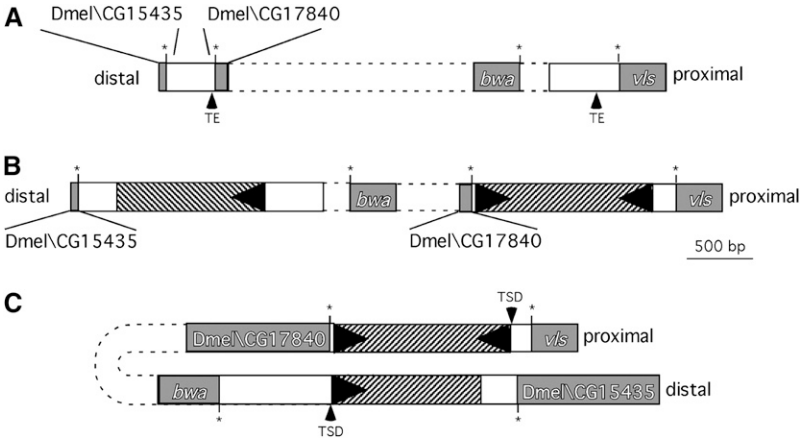

FIGURE 2.—Structure of the sequences at the breakpoints of *In(4)a*. For each putative ortholog, the gene identification for *D. melanogaster* is indicated in addition to an asterisk identifying the position of the putative start codon within the sequence of *D. americana*. Dashed horizontal lines are not drawn to indicated scale. (A) Structure of the sequence in standard arrangement of *D. americana*, which also corresponds to the genome sequence of *D. virilis*. Arrowheads indicate the positions of the TE insertion sites. (B) Structure of the sequence in a representative chromosome containing the *In(4)a* gene arrangement. Crosshatching indicates the orientation of inserted sequences with terminal repeat structures marked with triangles. (C) Folding of the chromosome illustrating sequence identity at the breakpoint regions. Positions of the putative target site duplication (TSD) are consistent with the disruption of a previously intact element. Annotated sequence alignments are available as supplemental material at http://www.genetics.org/supplemental/.

sequence at the proximal breakpoint appears to be a canonical element, whereas the repeat sequence at the distal breakpoint is a rearranged variant (Figure 2B). Features of this repetitive element indicate similarity to miniature inverted repeat transposable elements (MITEs) identified in a variety of organisms, including Drosophila (YANG *et al.* 2006).

Orientation of the sequence of the inverted arrangement is consistent with the independent insertion of two copies of the MITE followed by intrachromosomal exchange within the repeat as a cause of the chromosomal rearrangement. Close proximity to the putative start codon of several genes is a remarkable feature of these insertions and the subsequent inversion (Figure 2). In the most extreme case, the MITE inserted 20-bp upstream of the putative start codon of the apparent ortholog of Dmel\CG17840. The other insertion occurred 168-bp upstream of the *vls* gene. Orientation of the genes flanking this repeat sequence is consistent with simple folding of the chromosome coupled with intrachromosomal recombination within the MITE sequence to generate *In(4)a* (Figure 2C). An imperfect target site duplication (TSD), CACMTTTT, which would have formed upon the insertion of the full-length proximal element and is currently identifiable at the proximal end of the proximal insertion and at the proximal end of the distal insertion, provides direct evidence of the rearrangement. The exchange point responsible for the inversion falls somewhere within the MITE sequences, but without comparable ancestral sequences the exact position cannot be localized. On the basis of the position of the breakpoints in the genome sequence of *D. virilis*, this inversion reoriented the contents of an ~13-Mb region of chromosome *4*.

Polymorphism for *In(4)a*: The *In(4)a* arrangement segregates in natural populations of *D. americana* and previous studies revealed an association between this inversion and the *X–4* centromeric fusion (BLIGHT 1952; MCALLISTER 2003). A sample of chromosomes in the F_1 progeny of wild-caught males provided material to further examine this association. Sex-linked transmission of microsatellite markers on chromosome *4* was detected in the progeny of 48 wild-caught males, thus the *X–4* fusion was present in all males collected at the IR locality.

Linear arrangement of the chromosomes was determined by the presence/absence of inversion loops in the F_1 larvae of wild-caught males crossed to a standard laboratory stock. Due to the presence of the *X–4* centromeric fusion and the absence of crossing over in these males, the entire *X–4* arrangement transmits to daughters, and unfused chromosome *4* (and the *Y*) transmits to sons. Of the female larvae resulting from 41 different wild-caught males, the *In(4)ab* arrangement was observed for 20 families and the standard chromosome *4* arrangement for 21 families. This indicates the presence of the inverted and standard arrangements at approximately equal frequencies (95% C.I.: 35–65%) on the neo-*X* chromosome. On the other hand, the inverted arrangement of chromosome *4* was not observed for any of the F_1 male larvae (37 families), thus indicating a statistically significant (Fisher's exact test; $P < 1 \times 10^{-6}$) absence of the inverted arrangement on chromosomes not fused with the *X*. This analysis of flies collected from the IR locality provides further demonstration that *In(4)ab* is completely associated with the *X* chromosome and subject to *X*-linked transmission.

Amplification using primers anchored within the sequence rearranged by *In(4)a* was used to directly assess the presence of the inverted or standard arrangement. Combined analyses of microsatellite loci to determine sex linkage, polytene chromosome squashes to determine gene order, and PCR analyses to determine the inverted or standard arrangement of DNA sequence at the proximal breakpoint revealed complete agreement in the recognition of *In(4)a* from polytene chromosomes and PCR amplification (Table 2). The PCR assay invariably identified presence/absence of *In(4)a* in this sample of chromosomes, thus demonstrating equivalence (Fisher's exact test, $P < 1 \times 10^{-6}$) between

TABLE 2

Sex-linked transmission and gene order of chromosomes in males from IR population

Sex linkage	Polytene	PCR of proximal breakpoint	
		In(4)a	*4std*
X linked	*In(4)ab*	7	0
X linked	*St(4)*	0	9
Y linked	*St(4)*	0	15

cytological identification of the inversion and orientation of the underlying DNA sequence at the proximal breakpoint.

Sequence variation flanking the breakpoints: Nucleotide variation assayed from population samples provides a substrate for detecting the influence of natural selection within or near sequenced regions. Eight regions along chromosome *4*, including positions flanking both inversion breakpoints, were sequenced for the chromosomes in the IR sample. The arrangement of each of these chromosomes was determined from cytological and/or PCR analyses, and sequences from each chromosomal class were analyzed as individual groups. The neo-*X* population and the entire population were analyzed as reconstructed random samples on the basis of the observed frequencies of 50% standard and 50% *In(4)a* gene arrangements associated with the *X–4* fusion, which represents 97% of *X* chromosomes. Table 3 reports measures of nucleotide variability at these eight loci distributed along chromosome *4*, including *tim*, which is located 3.44 Mb toward the telomere from the distal inversion breakpoint. Direct ascertainment of meiotic recombination within inversion heterozygotes identified 6.9% ($n = 172$) recombinants at the *tim* locus relative to markers that are completely linked with the inversion (McAllister 2003; data not shown), therefore, it is expected to be minimally affected by the inversion. Correspondingly, sequence diversity at *tim* exhibits no evidence of departure from neutral expectations (Table 3) and sequence differentiation among classes does not differ significantly from zero (Table 4), so sequences of this gene region were used in HKA tests of polymorphism and divergence at other loci.

A distinct contrast exists in patterns and overall level of nucleotide variation on standard and inverted chromosomes. Variable nucleotide sites were observed in each gene region on standard chromosomes with either a fused or unfused centromere; both classes exhibit high haplotype diversity and low linkage disequilibrium, and statistical analyses fail to reject the standard neutral model in all tests (Table 3). On the other hand, gene regions flanking the proximal breakpoint on the inverted arrangement exhibit patterns inconsistent with neutrality. Moreover, all loci inside *In(4)a* or within 3 Mb of its breakpoints exhibit significant differentiation measured by K_{st} between inverted and standard chromosomes (Table 4).

Nucleotide variation is absent in the sequenced region of *dp* among the nine *In(4)a* chromosomes in the IR sample (Table 3). The HKA test ($\chi^2 = 8.2$, $P < 0.01$) indicates a deficit of nucleotide variability at *dp* among inverted chromosomes using *tim* as the control (Table 4 reports standardized polymorphism). While the single haplotype observed for the *dp* gene region is extraordinary for the highly variable genome of *D. americana*, low haplotype diversity and high linkage disequilibrium (within the gene region and with the inverted and standard chromosomal arrangements) typifies gene regions throughout (*raw* and 8665) and immediately flanking (15435 and 9171) *In(4)a* on chromosomes containing this inversion (Tables 3 and 4).

Nucleotide diversity at the 9171 gene region is not reduced significantly among *In(4)a* chromosomes in the IR sample, although of the regions assayed it is closest to *dp* on the inverted arrangement, estimated from the genome sequence of *D. virilis* to be 242 kb from the breakpoint and 334 kb from *dp*. Distribution of ancestral and derived variants among *X–In(4)a* sequences indicates a pattern that is consistent with this gene being near a causative locus for a recent selective sweep occurring within the inverted chromosomal class. Haplotype structure at 9171 is dominated by two closely related haplotypes present among seven inverted chromosomes in the sample, whereas two distinct haplotypes are present on the other two inverted chromosomes. A significant excess of high-frequency-derived nucleotide variants is revealed by Fay and Wu's *H* statistic (Table 3; $H = -10.3$, $P < 0.001$). The single haplotype at *dp* combined with the high-frequency-derived haplogroup at 9171 is indicative of a beneficial variant near *dp* associated with the inverted arrangement having been recently swept through this chromosomal class.

Reconstructed samples representing estimated frequencies of the different chromosomal arrangements in the population provide little indication of the putative sweep affecting the inverted chromosomes (Table 3). A significantly positive Tajima's *D* statistic ($D = 2.59$, $P < 0.01$) is obtained for the *dp* gene region upon analysis of neo-*X* chromosomes as a group, owing to half the sample being the single haplotype associated with *In(4)a* containing 10 fixed differences relative to haplotypes present among the standard arrangement. However, the neutral model is not rejected for the sample representing the entire population where only 36% of chromosomes contain the *In(4)a* gene arrangement. Fay and Wu's *H* statistic appears to have the greatest power to detect a sweep prior to fixation of a beneficial mutation (Zheng *et al.* 2006), which should be the case when considering this putative sweep in the context of the entire population, yet gene region 9171 within Fus–*In(4)a* remains the only gene region and

TABLE 3

Measures of sequence diversity and patterns of sequence variability in samples and reconstructed populations

Region	Sample	n	η	Syn	Rep	π	θ	H_{div}	Z_{nS}	D_T	H_{FW}
tim	*X–In(4)a*	9	16	16	0	0.042	0.050	1.00	0.138	−0.78	3.36
	X–4std	9	17	16	1	0.045	0.050	1.00	0.105	−0.63	1.89
	Fus pop[a]	18	22	21	1	0.043	0.052	0.99	0.068	−0.78	1.86
	Unf 4std	16	24	23	1	0.040	0.059	0.99	0.075	−1.33	1.93
	Population[a]	25	24	23	1	0.041	0.052	0.99	0.053	−0.82	2.00
15435	*X–In(4)a*	9	7	5	1	0.009	0.009	0.86	0.392	0.35	1.81
	X–4std	10	17	4	1	0.016	0.022	0.98	0.128	−1.33	−0.80
	Fus pop[a]	18	27	10	2	0.023	0.029	0.96	0.120	−0.88	1.78
	Unf 4std	16	13	8	1	0.014	0.014	0.97	0.117	−0.36	−1.17
	Population[a]	25	31	15	2	0.022	0.030	0.97	0.080	−1.05	0.01
raw	*X–In(4)a*	9	7	5	2	0.013	0.014	0.83	0.193	−0.61	0.94
	X–4std	10	20	16	4	0.036	0.043	1.00	0.133	−0.58	3.29
	Fus pop[a]	18	25	20	5	0.033	0.044	0.96	0.107	−0.97	2.80
	Unf 4std	16	19	13	6	0.027	0.030	0.99	0.060	−0.62	0.30
	Population[a]	25	28	21	7	0.033	0.042	0.98	0.062	−0.86	2.83
8665	*X–In(4)a*	9	17	6	1	0.017	0.018	0.86	0.465	−0.33	−3.42
	X–4std	10	12	10	0	0.012	0.014	0.93	0.163	−0.49	−0.18
	Fus pop[a]	18	27	16	1	0.030	0.026	0.94	0.204	0.50	2.41
	Unf 4std	15	17	13	0	0.016	0.017	0.95	0.158	−0.23	−1.67
	Population[a]	25	34	22	1	0.030	0.030	0.96	0.129	−0.11	2.33
dp	*X–In(4)a*	9	0	0	0	0.000	0.000	0.00	NA	NA	NA
	X–4std	10	4	3	1	0.005	0.006	0.71	0.239	−0.76	−1.51
	Fus pop[a]	18	22	12	1	0.033	0.018	0.89	0.785	2.59*	0.34
	Unf 4std	16	9	9	3	0.009	0.009	0.83	0.157	−0.59	1.75
	Population[a]	25	17	16	1	0.032	0.022	0.81	0.411	1.39	1.87
9171	*X–In(4)a*	9	15	7	0	0.020	0.025	0.69	0.514	−0.94	−10.3**
	X–4std	10	19	9	1	0.031	0.029	1.00	0.157	0.38	3.64
	Fus pop[a]	18	31	15	1	0.045	0.040	0.93	0.170	0.50	2.88
	Unf 4std	16	27	14	2	0.028	0.034	1.00	0.076	−0.89	3.88
	Population[a]	25	36	17	1	0.044	0.042	0.96	0.124	0.10	5.07
18095	*X–In(4)a*	9	19	13	3	0.034	0.030	0.94	0.274	0.40	2.61
	X–4std	9	30	16	11	0.032	0.036	1.00	0.154	−0.95	4.44
	Fus pop[a]	18	34	17	13	0.033	0.036	0.97	0.118	−0.71	4.08
	Unf 4std	16	32	19	8	0.033	0.037	1.00	0.130	−0.82	2.43
	Population[a]	25	38	17	16	0.031	0.030	0.99	0.081	−0.87	3.43
nmd	*X–In(4)a*	9	6	6	0	0.016	0.015	0.81	0.625	0.47	1.39
	X–4std	10	6	6	0	0.017	0.014	0.82	0.426	0.72	−1.24
	Fus pop[a]	18	8	8	0	0.017	0.016	0.82	0.275	0.27	−0.63
	Unf 4std	16	9	9	0	0.017	0.019	0.91	0.256	−0.37	−0.15
	Population[a]	25	11	11	0	0.017	0.020	0.84	0.241	−0.44	−0.35

The following measures are reported for each sample and reconstructed population: n, number of chromosomes analyzed; η, total number of mutations; Syn, number of synonymous mutations; Rep, number of nonsynonymous mutations; π, pairwise diversity per silent site (synonymous and noncoding); θ, heterozygosity per silent site; H_{div}, haplotype diversity; Z_{nS}, average linkage disequilibrium; D_T, Tajima's D; H_{FW}, Fay and Wu's *H*. Regions contained within *In(4)a* are underlined. Statistical significance: $*P < 0.01$; $**P < 0.001$.

[a] Reconstructed random samples: Fus pop = 9× *X–In(4)a*, 9× *X–4std*; Population = 9× *X–In(4)a*, 9× *X–4std*, 7× *Unf 4std*.

sample for which the neutral model is rejected by the *H* statistic (Table 3).

The influence of *In(4)ab* on sequence variation in *D. americana* was originally detected by haplotype structure and population differentiation at the *bib* gene region (McAllister 2003), which is located in the interval between the proximal breakpoint and the centromere and farther from the inversion than 9171. The breakpoint PCR assay was used to examine the DNA samples from which sequences were previously obtained (McAllister and Charlesworth 1999; McAllister 2003) and to determine the gene order for these chromosomes. *In(4)a* is present on 83% ($n = 12$) of fused fourth chromosomes obtained from two localities in

TABLE 4

Contrasts of sequence measures for inverted and standard chromosomal arrangements

	Standard polymorphism		Standard substitution		d_a:	K_{st}:	Average $\|D'\|$:
Region	*In(4)a*	*St(4)*	*In(4)a*	*St(4)*	*In(4)a*:*St(4)*	*In(4)a*:*St(4)*	*In(4)a*:*St(4)*
tim[a]	0.40	0.45	0.14	0.00	0.000	0.000	0.480
tim	0.38	0.39	0.11	0.00	0.000	−0.006	0.405
15435	0.12	0.17	0.24	0.00	0.006	0.332**	1.000
raw	0.15	0.35	0.18	0.00	0.004	0.222**	0.944
8665	0.16	0.25	0.00	0.09	0.011	0.419**	0.957
dp	0.00*	0.11	0.57	0.14	0.012	0.727**	1.000
9171	0.16	0.24	0.09	0.00	0.012	0.346**	0.919
bib[a]	0.04*	0.25	0.23	0.00	0.007	0.143**	1.000
Adh[a]	0.23	0.22	0.06	0.00	0.001	0.041**	0.698
18095	0.33	0.35	0.00	0.00	0.000	−0.002	0.421
nmd	0.16	0.16	0.08	0.00	0.000	0.015	0.580

Regions contained within *In(4)a* are underlined. Statistical significance in HKA and differentiation tests: *$P < 0.01$; **$P < 0.001$.

[a] Sequence data from McAllister (2003).

Nebraska and it is present on 13% (n =15) of fused fourth chromosomes from the eastern G96 population (Gary, IN). These frequencies, in combination with the estimate of 50% *In(4)a* for the Iowa population described here, are indicative of the inversion being present at the highest frequency in western populations and the lowest frequency in eastern populations, which is also suggested by a previous survey of inversion polymorphism (Hsu 1952).

Reanalyses of sequence data from a combined sample representing both western and eastern populations are included as supplemental Table S4 (http://www.genetics.org/supplemental/). The one notable result is the significant reduction in nucleotide diversity among inverted chromosomes for the *bib* gene region, which is revealed by the HKA test ($\chi^2 = 8.0$, $P < 0.01$) using *tim* sequences from the same chromosomes as a control (Table 4). Although locus-specific reductions in nucleotide diversity are indicated only for the *dp* and *bib* gene regions, chromosomes containing the *In(4)a* arrangement exhibit a lower overall level of sequence diversity than standard chromosomes (Table 4; Wilcoxon signed-rank test; $W = 42$, $P < 0.05$). This reduction in variability may result from the recent origin of the inversion, from subsequent sweeps within the inverted arrangement, or from a combination of these and other effects.

Long-range haplotype structure and recombination: Upon formation of *In(4)a*, this rearrangement would have been associated with a single haplotype of chromosome *4*. PCR amplification of the identified rearrangement at the proximal breakpoint provides direct evidence that all inverted chromosomes originate from a single haplotype. Reestablishment of nucleotide variation on chromosomes containing *In(4)a* could have occurred through recombination; or in regions of completely restricted recombination between rearrangements, new mutations have been the only source of variability among inverted chromosomes. Therefore, the pathway through which nucleotide variants have been acquired by inverted chromosomes can be inferred from haplotype structure relative to chromosomal arrangement.

The homogenizing effect of recombination is evident at most loci flanking the inversion, and to a limited extent, even within the inverted region. Of all segregating sites at *tim*, about half are shared between chromosomes with the inverted and standard arrangements (Table 5). Only about a third of segregating sites are shared between these arrangements at *Adh*, 18095, and *nmd*, which are located a similar distance from an inversion breakpoint, but on the opposite end toward the centromeric region of chromosome *4*. Combined inhibition of crossing over by the inversion and by the centromere potentially reduces the overall rate of recombination between different arrangements at proximal loci. Notably, no recombinants were obtained in this region in a previous experiment using females heterozygous for *In(4)ab* and the standard arrangement and respectively fused and not fused with the *X* (McAllister 2003); however, the rate of exchange is unknown for inversion heterozygotes when centromeric arrangement is the same [*i.e.*, *X*–*In(4)ab*/*X*–4std]. Although lower proportions of shared variability indicate lower rates of exchange for loci in the proximal region compared to *tim*, the realized level of exchange at the two most proximal loci (18095 and *nmd*) appears sufficient to completely homogenize sequence diversity among the different arrangements as revealed by K_{st} measures that do not differ significantly from zero (Table 4).

Two variable loci within the inversion, *raw* and 8665, exhibit evidence of exchange between inverted and standard arrangements. A single shared variant is pre-

TABLE 5

Shared, unique, and fixed variants in comparisons of inverted and standard chromosomal arrangements

		Unique ancestral variants		Unique derived variants		Derived fixed differences	
Gene	Shared	*X–In(4)a*	*St(4)*	*X–In(4)a*	*St(4)*	*X–In(4)a*	*St(4)*
tim[a]	15	0	1	1	19	0	0
tim	15	0	1	1	13	0	0
15435	0	0	2	7	21	3	0
raw	1	0	2	5	24	0	0
8665	2	0	0	7	16	0	1
dp	0	0	0	0	9	8	2
9171	5	0	1	10	26	1	0
bib[a]	0	0	4	4	52	1	0
Adh[a]	13	0	1	1	25	0	0
18095	17	0	0	2	26	0	0
nmd	3	0	1	3	6	0	0
Total	57	0	12	40	218	13	3

Regions contained within *In(4)a* are underlined.
[a] Sequence data from McAllister (2003).

sent within the sequences of *raw* and two shared variants are present within the sequences of 8665 (Table 5). Shared variants within 8665 occur at two nucleotide sites separated by a single invariant site and are detected as a gene conversion tract by the method of Betrán *et al.* (1997). Presence of little shared variation between the arrangements indicates that *In(4)a* is a weakly permeable barrier to exchange inside the inversion within ~1 Mb of the breakpoints, which has resulted in some disruption of linkage disequilibria with the inverted and standard arrangements (Table 4). Divergence between the chromosomal arrangements, however, suggests evolutionary independence (Table 4).

The sequenced region 15435 is only 400 bp outside the distal inversion breakpoint and of the regions assayed it is closest to a breakpoint, so it should provide the best record of the evolutionary history of the inversion. A common haplogroup for 15435 exists among the inverted chromosomes, which is evidenced by the absence of shared variants and the presence of three fixed differences relative to the standard arrangement (Table 5). All variable sites at 15435 are, therefore, completely associated (average $|D'| = 1.0$) with either the inverted or standard chromosomes (Table 4). As expected, proximity to the inversion breakpoint apparently averts exchange between inverted and standard arrangements, thus preserving historical associations with the inversion and protecting newly derived mutations from being distributed between the arrangements.

Fixed nucleotide differences between the arrangements potentially originated prior to the inversion and were present on the original inverted haplotype, but were subsequently lost or at least not sampled from the standard arrangement. These ancestral polymorphisms would upwardly bias estimates of age using fixed differences (Charlesworth *et al.* 2005a). Uniquely segregating variants, on the other hand, are derived through new mutations arising on the inverted arrangement. The observed level of uniquely derived silent diversity ($\pi = 0.009$) at 15435 associated with *In(4)a* is consistent with an expected reduction in heterozygosity by 3/4 for an *X*-linked locus (which is the case for the inverted arrangement) when calibrated from the observed diversity among standard chromosomes ($\pi = 0.015$), suggesting persistence of the inverted arrangement for a sufficient amount of time to effectively achieve mutation/drift equilibrium. Coalescent simulations indicate $1.0N_e$ generations as the most likely interval for acquiring the observed pairwise diversity following a complete bottleneck defining the origin of *In(4)a*. Furthermore, <5% of simulated samples contain $\pi \geq 0.009$ following $0.24N_e$ generations after a bottleneck, which provides an estimate of the minimum generations required to accumulate the observed pairwise diversity among *In(4)a* chromosomes.

The *dp* region is contained within the inversion and all segregating variants are either fixed between the arrangements or segregating only among standard chromosomes (Table 5). Absence of derived nucleotide variants among the inverted chromosomes indicates the most likely age for the *dp* haplotype is zero generations ago, which is inconsistent with the age of *In(4)a* predicted from region 15435. However, a maximum plausible age of the *dp* haplotype is not obtained from coalescent simulations, because zero polymorphism is not unexpected under mutation/drift equilibrium given the low heterozygosity predicted from the observed silent diversity ($\pi = 0.009$) at this locus among standard chromosomes. Using the method of Fu (1996) and calibrating θ ($3N\mu = 1.296$) from the *dp* region of standard chromosomes, a value of $1.12N_e$ generations is obtained as the maximum estimate of the time since the most recent common ancestor of the *dp* haplotype on inverted chromosomes with 95% confidence. Therefore,

credible intervals of the ages of a common haplotype at these two regions overlap, possibly indicating the origin of *In(4)a*, but the calculation assumes neutrality and disregards evidence of a deficit of polymorphism at *dp*, suggesting a shortened genealogy due to a recent sweep. Standard chromosomes are also assumed to represent an accurate estimate of expected heterozygosity and the HKA test yields a marginally nonsignificant result, so the interval estimated for *dp* is extremely wide and likely an overestimate of the plausible age of this haplotype.

Fixed differences also provide little inference on the age of *In(4)a* or of the common haplotype at *dp*. Eight derived variants are fixed on the inverted arrangement (Table 5). The mutations responsible for these fixed differences may have occurred prior to the formation of the inversion. Alternatively, these may be mutations that were derived on the inverted arrangement or were acquired through recombination and subsequently fixed during a secondary sweep of the *dp* haplotype. The inability to determine when these mutations occurred would bias any inference concerning the age of the inversion using numbers of fixed differences. A method for inferring the proportion of fixed differences arising from ancestral mutations was recently proposed by Charlesworth *et al.* (2005a); however, these data violate both the assumption of neutrality and of equal effective population sizes and are further complicated by the possibility of exchange between the arrangements.

A telling contrast exists between 9171 and *bib*, which are both located outside the proximal breakpoint. A reduction in disequilibrium with the arrangements due to five shared variants at 9171 is consistent with recombination in this gene region (Tables 4 and 5). Although *bib* is located farther from the breakpoint (and the *dp* haplotype), the *bib* gene region does not contain any shared variants between the inverted and standard arrangements (Table 5). Therefore, *dp* and *bib* both exhibit locus-specific reductions in nucleotide diversity on chromosomes with *In(4)a* and nucleotide variants in both are completely associated with either the inverted or standard arrangement (Table 4) and therefore with each other. A derived nucleotide variant in the *bib* gene defines an acquired restriction site for *Bbr*PI previously associated with *In(4)ab* in assays of laboratory lines of *D. americana* (McAllister 2003). This association also holds for chromosomes sampled widely from natural populations. Compilation of haplotype structure including the centromeric arrangement, the *Bbr*PI restriction site in *bib*, and the gene arrangement of chromosome *4* reveals a complete three-way association (Table 6; $G^2 = 128.9$, $P < 0.001$). Observed silent diversity ($\pi = 0.003$) in the *bib* gene region for chromosomes containing *In(4)a* is consistent with coalescence $0.07N_e$ (C.I.: 0.03–0.17) generations ago calibrating heterozygosity on the observed silent diversity ($\pi = 0.033$) among standard chromosomes and predicting a recovery size reduced by 3/4 due to *X* linkage. Therefore, the likely age of the association of a single *bib* haplotype, including the *Bbr*PI restriction site with *In(4)a* is intermediate between the older coalescent event defining the origin of the *In(4)a* haplotype and the contemporary sweep eliminating variation at *dp* near the proximal breakpoint.

TABLE 6

Presence of *bib* haplotype on different chromosomal arrangements in samples of *D. americana*

Sample	*bib* haplotype	Chromosomal arrangement		
		X–In(4)a	*X–4std*	*Unf 4std*
IR, Iowa	*Bbr*PI+	9	0	0
	*Bbr*PI−	0	10	17
G96, Indiana	*Bbr*PI+	2	0	0
	*Bbr*PI−	0	13	15
NN, Nebraska	*Bbr*PI+	6	0	0
	*Bbr*PI−	0	1	4
DN, Nebraska	*Bbr*PI+	4	0	0
	*Bbr*PI−	0	1	5
Total	*Bbr*PI+	21	0	0
	*Bbr*PI−	0	25	41

Sequence divergence of *In(4)a*: A striking feature of the nucleotide differences between inverted and standard chromosomes is the lineage-specific accumulation of derived variants associated with *In(4)a* when variable sites are polarized with the sequence of *D. virilis* (Table 5). A neighbor-joining tree constructed from pairwise measures of net substitutions and presented as Figure 3 also illustrates the accelerated divergence of the inverted arrangement. No divergence is observed between fourth chromosomes either fused or not fused with the *X* but having the standard gene order. From the common node of *D. americana* alleles, the Fus–*In(4)a* arrangement contains a branch fourfold greater in length than these standard arrangements, indicating an elevated rate of substitution within the inverted chromosomal class.

A standardized measure of the substitution rate for inverted and standard chromosomes was developed to examine the pattern among loci and to directly compare arrangements. This measure of standardized substitution is based on the number of unique fixed variants within a class relative to the number of shared fixed differences compared to an outgroup. Results are presented for the comparison of the inverted and standard arrangements of chromosome *4* (Table 4). A significant excess of substitutions is observed for the inverted arrangement for these 10 loci distributed along chromosome *4* (Wilcoxon signed-rank test; $W = 38$, $P < 0.05$). Compared to *D. virilis*, more fixed differences are present on the inverted than the standard arrangement at each region except 8665 within the inversion and 18095 near the centromere. Accumulated substitutions

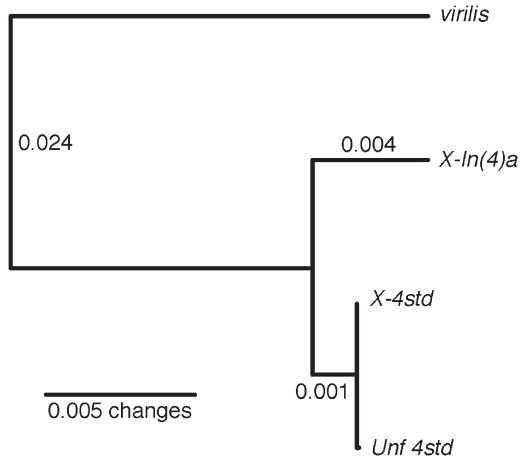

Figure 3.—Neighbor-joining tree constructed from pairwise estimates of net divergence and rooted with *D. virilis*. Branch lengths represent net rate of substitution per site estimated from an overall alignment of 6972 nucleotides.

on the inverted arrangement are elevated the most at *dp*, *bib*, and at the two sequenced regions flanking the distal breakpoint. Each of these loci represents a position identified as having undergone a bottleneck through a single haplotype, therefore causing fixation of derived variants associated with the surviving haplotype. Many of these variants clearly existed prior to events isolating inverted and standard chromosomes, because overall about half of these sites are still segregating the ancestral and derived variants among standard chromosomes (contrast of unique ancestral variants with derived fixed differences in Table 5).

DISCUSSION

This study demonstrates the utility of the 12 sequenced Drosophila genomes for launching investigations within other closely related species. Using the genome sequence of *D. virilis* as a guide, and leveraging against cytological reference points, clones containing the inversion breakpoints of *In(4)a* were efficiently isolated from a BAC library of *D. americana*. Furthermore, the annotated genome sequence of *D. virilis* served as a reference for localizing the breakpoints to relatively small intergenic intervals. DNA sequences of these intervals clearly reveal insertions of two MITE elements and an intrachromosomal exchange within this dispersed repeat as the mechanism generating this inversion. Exchange within a dispersed repeat is also the mechanism responsible for the origin of three different inversions in *D. buzzatii* (Cáceres *et al.* 1999; Casals *et al.* 2003) and the Arrowhead inversion in *D. pseudoobscura* (Richards *et al.* 2005). In contrast, the two common chromosome *3* inversions in *D. melanogaster* both appear to have originated by breakage and repair in regions without shared repeat sequences (Wesley and Eanes 1994; Matzkin *et al.* 2005). Through either mechanism, all of these inversions share a common feature of originating through a single intrachromosomal recombination at nonhomologous sites. Our finding adds to the growing list of inversions having breakpoints clearly localized within repeated sequences, thus directly supporting the hypothesis that repetitive sequences mediate change in genome organization (Finnegan 1989; Montgomery *et al.* 1991).

While the mutational mechanisms that generate inversions are being clarified through molecular analysis, the role of selection in the persistence of these rearrangements remains unclear (Andolfatto *et al.* 2001). Identification of the breakpoints responsible for *In(4)a* enabled detailed analyses of sequence variation and haplotype structure to reveal imprints of selection along the chromosome. In separate samples of *D. americana*, the inverted arrangement is only observed in combination with the *X–4* centromeric fusion; therefore, the breakpoints of the inversion and allelic variants associated with the breakpoints are completely linked with the *X* chromosome. While this observation is consistent with the inverted arrangement arising on the *X–4* arrangement, cytologically defined *In(4)a* and its corresponding rearrangement of the sequence at the proximal breakpoint is also present on autosomal chromosome *4* in *D. novamexicana*, an allopatric sister species of *D. americana* (Hsu 1952; A. L. Evans, P. A. Mena and B. F. McAllister, unpublished data). Comparative evidence indicating *In(4)a* was present in the common ancestor of *D. novamexicana* and *D. americana* corroborates the sequence data from the three loci (15435, raw, and 8665) examined around the distal breakpoint, suggesting a relatively old inversion. Unfortunately, any evidence that positive selection initially favored the inverted arrangement following its origin has been lost.

While the history of selection on *In(4)a* itself has been obscured by time, multiple imprints of selection clustered around its proximal breakpoint suggest a pattern of coadaptation within the class of chromosomes it defines. Models of coadaptation developed over the past 50 years demonstrate the selective benefit of an inversion as a recombination suppressor associated with coordinately selected alleles (Kimura 1956; Kirkpatrick and Barton 2006). Isolated "islands" of strong linkage disequilibrium provide evidence for coadaptation (Schaeffer *et al.* 2003), and the complete association among the *X* chromosome, the *Bbr*PI^{+} haplotype at *bib* and *In(4)a* (including all nucleotide variants within the sequenced regions of *dp* and 15435) is an example of such interspersed disequilibrium in *D. americana*. Apparent establishment of these allelic associations through sequential events of positive selection further implies coadaptation. The following findings within the proximal region of chromosomes containing *In(4)a* are consistent with a coadaptation model including the rearrangement history, the targets of selection, and the patterns of recombination suppression within this newly *X*-linked

genomic region: (i) deficiency of nucleotide polymorphism within the *dp* gene region on the *In(4)a* arrangement and complete association of its nucleotide variants with the rearrangements of chromosome *4*, (ii) excess of high-frequency-derived nucleotide variants within the 9171 region on the *In(4)a* arrangement and incomplete association of its nucleotide variants due to recombination relative to the rearrangements of chromosome *4*, (iii) deficiency of nucleotide polymorphism within the *bib* gene region on the *In(4)a* arrangement and complete association of its nucleotide variants with the rearrangements of chromosome *4*, and (iv) lack of associations between nucleotide polymorphisms at proximal loci and the rearrangements of chromosome *4*.

A model of historical events affecting *In(4)a* is presented in Figure 4 beginning with the coexistence of autosomes containing this inversion with chromosomes being either fused or not fused to the *X* having the standard gene order. The diagram indicates the predicted influence of chromosomal rearrangements on patterns of nucleotide variation through their effects as barriers to exchange when each derived rearrangement replaces and recombines with existing arrangements. First, the association of *In(4)a* with the neo-*X* is postulated through a unique recombination event in the interval between *bib* and the proximal inversion breakpoint (Figure 4A). This would have generated a single haplotype of the *X–4* fusion chromosome containing *In(4)a*, thus initiating the *X*-linked transmission that currently persists. This derived arrangement acquired existing nucleotide variation in proximal and distal regions through recombination with the two ancestral arrangements from which it arose; however, the original association of the haplotype in the *bib* gene region has been retained on this derived arrangement (Figure 4B).

Several factors are potentially responsible for the retention of the haplotype association at the *bib* locus: recombination suppression due to *In(4)a* and the centromere, coexistence of the alternative arrangements containing *In(4)a* for a brief interval, or selection favoring particular allelic associations within this derived arrangement. These associations extend into the *X* chromosome (Vieira *et al.* 2001, 2006), including an inversion, *In(X)c*, on this chromosomal arm of the *X–4* fusion (Blight 1955). Although the presence of selected alleles in or near the *bib* gene is uncertain, because a combination of the other two factors is sufficient to account for associations in this region, the apparent deficiency of nucleotide diversity in the *bib* region implies that positive selection favored coupling *In(4)a* with the *X* chromosome. Furthermore, the fact that *bib* experiences >3% recombination relative to the alternative centromeric arrangements in females homozygous for the standard gene arrangement (McAllister and Evans 2006) evokes doubt over the complete neutrality of this region, because linkage disequilibrium would decay at a substantial rate (~0.97 per generation) given a similarly high rate of exchange among inverted chromosomes.

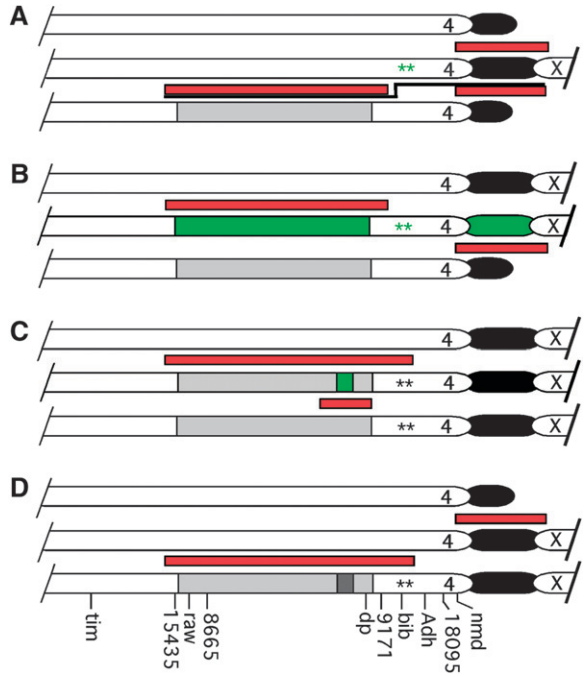

Figure 4.—Evolutionary model of historical rearrangements, patterns of recombination, and turnover within chromosome *4* of *D. americana*. Red blocks indicate recombination barriers between arrangements. Green features are putative targets of positive selection and the asterisks indicate a candidate region for feminization. See text for details. (A) Origin of the *Fus-In(4)a* arrangement through unique recombination event. (B) Recovery of variability on *Fus-In(4)a* through exchange with parental chromosomes. (C) Origin of *In(4)b* with hitchhiking. (D) Arrangements of chromosome *4* commonly observed in contemporary populations.

The model also illustrates the effects of the small *In(4)b* rearrangement arising within *In(4)a* (Figure 4C), although the breakpoints of this inversion have yet to be identified. This smaller inversion is found nearly in complete association with *In(4)a* in *D. americana* and it has not been identified in *D. novamexicana* (Hsu 1952; P. A. Mena and B. F. McAllister, unpublished data). Origin of *In(4)b* nested within *In(4)a* and selection-mediated replacement of the existing inverted chromosomes would have generated a large hitchhiking event within this chromosomal class due to suppression of recombination by the small inversion, thus accounting for the monomorphism observed for the *dp* gene region. Preliminary analyses of nucleotide variability indicate that the monomorphism observed within the *dp* gene extends over a 2.5-Mb region inside the proximal breakpoint of *In(4)a*, thus encompassing the chromosomal region containing *In(4)b* (A. L. Evans and B. F. McAllister, unpublished data). The inference that *In(4)b* caused a recent strong sweep near the proximal breakpoint and eliminated variation at the *dp* gene

region also explains the elevated frequency of derived nucleotides at 9171, because the distance between this region and *In(4)b* appears sufficiently large that recombination with the ancestral arrangement containing only *In(4)a* is plausible.

This study reveals a pattern of coadaptation characterized by selection-mediated turnover within the inverted class and apparently initiated with the coupling of *In(4)a* and the *X* chromosome. New haplotypes have arisen at different positions along the inverted chromosome and replaced ancestral haplotypes. The overall effects of these advances are evident in the accelerated rate of substitution observed for the inverted arrangement of the neo-*X*. Differential success of inverted *vs.* standard arrangements of the neo-*X* chromosome remain unclear, although the apparent west–east cline and comparison of contemporary and historical collections of *D. americana* indicates a possible increase in frequency of the inverted arrangement over the past 50 years (McAllister 2003).

Selection mediated buildup of this complex chromosomal arrangement involving the centromeric fusion with the *X* and two nested inversions is effectively the same process thought to underlie the formation of gene complexes causing segregation distortion (Thomson and Feldman 1974; Dyer *et al.* 2007). We have not detected any gross distortion of sex ratio in *D. americana*, and the inverted form is the majority arrangement of the neo-*X* in northwestern populations, so the coadaptation observed within this chromosomal arrangement does not appear to result from a sex ratio distorting mechanism such as commonly found in species of Drosophila (Jaenike 1996). In the case of chromosome *4* in *D. americana*, as with other sex chromosomes, sex-linked transmission may promote coadaptation through intralocus sexual conflict (Bull 1983; Rice 1987). Establishment of the inversion complex may be favored because it reduces the influx of masculinized alleles accumulating among *Y*-linked unfused fourth chromosomes. Due to the transient nature of *Y* linkage in *D. americana*, conditions favoring accumulation of male-benefit alleles are strongest near the centromere of the unfused fourth chromosome (McAllister and Evans 2006). However, absence of sequence differentiation at the two most proximal loci regardless of centromeric arrangement indicates a sufficient level of exchange among all three arrangements to homogenize sequence variation in this region (Figure 4D). Because of the evidence of flux in the region with the greatest potential for masculinization, the establishment of this *X*-linked inversion complex as protection against masculinized alleles appears unlikely. A similar argument also applies to protection against the influx of passively accumulating deleterious variation on unfused chromosome *4* (Orr and Kim 1998).

Feminizing selection is a possible cause of the apparent coadaptation of the neo-*X*. Relative to autosomal chromosome *4*, the *X–4* arrangement is present twice as often in females where it is cotransmitted with maternal factors. This bias should advance sexually antagonistic alleles that increase female fitness at the expense of male fitness (Rice 1984; Gibson *et al.* 2002). Although *X*-linked feminized alleles may arise on chromosome *4*, widespread polymorphism for the *X–4* centromeric fusion creates ample opportunities for meiotic exchange with the transient neo-*Y* (McAllister 2002, 2003; McAllister and Evans 2006). Suppression of recombination by *In(4)ab* protects any feminized content from being lost from the neo-*X*. Feminization of a locus near the *bib* gene on the *X–4* arrangement would have favored the initial association with *In(4)a* and would preserve the associations identified by this analysis (Figure 4). Further analyses of the region of chromosome *4* proximal to *In(4)a*, including the *bib* gene, are needed to identify direct targets of sex-specific selection pressures responsible for the divergence of this neo-*X* chromosome.

Comments from members of the McAllister lab, an anonymous reviewer, and D. Begun improved the presentation of these results. The National Human Genome Research Institute funded production of the publicly available genome sequence and the BAC library used in the study. This article is based upon work supported by the National Science Foundation under grant no. DEB-0420399 and the Roy J. Carver Charitable Trust under grant no. 05-2045.

LITERATURE CITED

Andolfatto, P., F. Depaulis and A. Navarro, 2001 Inversion polymorphisms and nucleotide variability in Drosophila. Genet. Res. **77:** 1–8.

Bachtrog, D., 2004 Evidence that positive selection drives *Y*-chromosome degeneration in *Drosophila miranda*. Nat. Genet. **36:** 518–522.

Bachtrog, D., 2005 Sex chromosome evolution: molecular aspects of *Y*-chromosome degeneration in *Drosophila*. Genome Res. **15:** 1393–1401.

Bachtrog, D., 2006 A dynamic view of sex chromosome evolution. Curr. Opin. Genet. Dev. **16:** 578–585.

Bartolomé, C., and B. Charlesworth, 2006 Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. Genetics **174:** 2033–2044.

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

Betrán, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146:** 89–99.

Blight, W. C., 1955 A cytological study of linear populations of *Drosophila americana* near St. Louis, Missouri. Ph.D. Dissertation, Washington University, St. Louis.

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

Caletka, B. C., and B. F. McAllister, 2004 A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. Mol. Phylogenet. Evol. **33:** 664–670.

Casals, F., M. Cáceres and A. Ruiz, 2003 The Foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of Drosophila buzzatii. Mol. Biol. Evol. **20:** 674–685.

Cáceres, M., J. M. Ranz, A. Barbadilla, M. Long and A. Ruiz, 1999 Generation of a widespread Drosophila inversion by transposable element. Science **285:** 415–418.

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

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

Charlesworth, B., C. Bartolomé and V. Noël, 2005a The detection of shared and ancestral polymorphisms. Genet. Res. **86:** 149–157.

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

Dyer, K., B. Charlesworth and J. Jaenike, 2007 Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc. Natl. Acad. Sci. USA **104:** 1587–1592.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Finnegan, D. J., 1989 Eukaryotic transposable elements and genome evolution. Trends Genet. **5:** 103–107.

Fu, Y.-X., 1996 Estimating the age of the common ancestor of a DNA sample using the number of segregating sites. Genetics **144:** 829–838.

Gibson, J. R., A. K. Chippindale and W. R. Rice, 2002 The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc. R. Soc. Lond. Ser B **269:** 499–505.

Gordo, I., and B. Charlesworth, 2001 The speed of Muller's ratchet with background selection, and the degeneration of *Y* chromosomes. Genet. Res. **78:** 149–161.

Graves, J. A. M., 2000 Human *Y* chromosome, sex determination, and spermatogenesis—a feminist view. Biol. Reprod. **63:** 667–676.

Grumbling, G., V. Strelets, and The FlyBase Consortium, 2006 FlyBase: anatomical data, images and queries. Nucleic Acids Res. **34:** D484–D488.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature **445:** 82–85.

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

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the Superoxide Dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Jaenike, J., 1996 Sex-ratio meiotic drive in the *Drosophila quinaria* group. Am. Nat. **148:** 237–254.

Kimura, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10:** 278–287.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation, and speciation. Genetics **173:** 419–434.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.

Kennison, J. A., 2000 Preparation and analysis of polytene chromosomes, pp. 111–117 in *Drosophila Protocols*, edited by W. Sullivan, M. Ashburner and R. S. Hawley. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human *X* chromosome. Science **286:** 964–967.

Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

Markow, T. A., B. F. McAllister, and T. C. Kaufman, 2003 A white paper requesting BAC library construction: Drosophila as a model for comparative genomics. http://www.genome.gov/10001852.

Matzkin, L., T. J. S. Merritt, C.-T. Zhu and W. F. Eanes, 2005 The structure and population genetics of the breakpoints associated with the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster*. Genetics **170:** 1143–1152.

McAllister, B. F., 2002 Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline. Genome **45:** 13–21.

McAllister, B. F., 2003 Sequence differentiation associated with an inversion on the neo-*X* chromosome of *Drosophila americana*. Genetics **165:** 1317–1328.

McAllister, B. F., and B. Charlesworth, 1999 Reduced sequence variability on the neo-*Y* chromosome of *Drosophila americana americana*. Genetics **153:** 221–233.

McAllister, B. F., and A. L. Evans, 2006 Increased nucleotide diversity with transient *Y* linkage in *Drosophila americana*. PLoS ONE **1:** e112.

Montgomery, E. A., S.-M. Huang, C. H. Langley and B. H. Judd, 1991 Chromosome rearrangements by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. Genetics **129:** 1085–1098.

Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics **146:** 695–709.

Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal region in Drosophila. Genetics **155:** 685–698.

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

Nei, M., and F. Tajima, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

Oliver, B., and M. Parisi, 2004 Battle of the Xs. BioEssays **26:** 543–548.

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

Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley *et al.*, 2003 Paucity of genes on the Drosophila *X* chromosome showing male-biased expression. Science **299:** 697–700.

Rice, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution **38:** 735–742.

Rice, W. R., 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution **41:** 911–914.

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.

Richards, S., Y. Liu, B. R. Bettencourt, P. Hradecky, S. Letovsky *et al.*, 2005 Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. Genome Res. **15:** 1–18.

Rogers, D. W., M. Carr and A. Pomiankowski, 2003 Male genes: X-pelled or X-cluded? BioEssays **25:** 739–741.

Ross, M. T., D. V. Gratham, A. J. Coffey, S. Scherer, K. McLay *et al.*, 2005 The DNA sequence of the human *X* chromosome. Nature **434:** 325–337.

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

Schaeffer, S. W., M. P. Goetting-Minesky, M. Kovacevic, J. R. Peoples, J. L. Graybill *et al.*, 2003 Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. Proc. Natl. Acad. Sci. USA **100:** 8319–8324.

Schaeffer, S. W., and W. W. Anderson, 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics **171:** 1729–1739.

Schlötterer, C., 2000 Microsatellite analysis indicates genetic differentiation of the neo-sex chromosomes in *Drosophila americana americana*. Heredity **85:** 610–616.

Schwartz, S., Z. Zhang, K. A. Frazer, A. Smit, C. Riemer *et al.*, 2000 PipMaker A Web Server for Aligning Two Genomic DNA Sequences. Genome Res. **10:** 577–586.

Spicer, G. S., and C. D. Bell, 2002 Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12s and 16s ribosomal DNA. Ann. Entomol. Soc. Am. **95:** 156–161.

Steinemann, M., S. Steinemann and F. Lottspeich, 1993 How *Y* chromosomes become genetically inert. Proc. Natl. Acad. Sci. USA **90:** 5737–5741.

Swofford, D. L., 2002 *Phylogenetic Analysis using Parsimony* (*and Other Methods*). Sinauer Associates, Sunderland, MA.

Sykes, B., 2004 *Adam's Curse: The Science that Reveals Our Genetic Destiny*. W. W. Norton, New York.

Tajima, F., 1983 Evolutionary relationships of DNA sequences in finite populations. Genetics **105:** 437–460.

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

Thomson, G. J., and M. W. Feldman, 1974 Population genetics of modifiers of meiotic drive. II. Linkage modification in the segregation distortion system. Theor. Popul. Biol. **5:** 155–162.

Vallender, E. J., and B. T. Lahn, 2004 How mammalian sex chromosomes acquired their peculiar gene content. BioEssays **26:** 159–169.

Vallender, E. J., N. M. Pearson and B. T. Lahn, 2005 The *X* chromosome: not just her brother's keeper. Nat. Genet. **37:** 343–345.

Vicoso, B., and B. Charlesworth, 2006 Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. **7:** 645–653.

Vieira, C., A. Almeida, J. D. Dias and J. Vieira, 2006 On the location of the gene(s) harboring the advantageous variant that maintains the *X/4* fusion of *Drosophila americana*. Genet. Res. **87:** 163–174.

Vieira, J., B. F. McAllister and B. Charlesworth, 2001 Evidence for selection at the *fused1* locus of *Drosophila americana*. Genetics **158:** 279–290.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wesley, C. S., and W. F. Eanes, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91:** 3132–3136.

Wu, C.-I., and E. Y. Xu, 2003 Sexual antagonism and X inactivation—the SAXI hypothesis. Trends Genet. **19:** 243–247.

Yang, H-P, T-L Hung, T-L You and T-H Yang, 2006 Genomewide comparative analysis of the highly abundant transposable element *DINE-1* suggests a recent transpositional burst in *Drosophila yakuba*. Genetics **173:** 189–196.

Zheng, K., Y.-X. Fu, S. Shi and C.-I. Wu, 2006 Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics **174:** 1431–1439.

Communicating editor: D. Begun