Copyright © 2003 by the Genetics Society of America

Inferences on the Evolutionary History of the *Drosophila americana* Polymorphic *X/4* Fusion From Patterns of Polymorphism at the *X*-Linked *paralytic* and *elav* Genes

Cristina P. Vieira,[1] Paula A. Coelho and Jorge Vieira

Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto 4150-180, Portugal

Manuscript received November 6, 2002
Accepted for publication April 14, 2003

ABSTRACT

In Drosophila there is limited evidence on the nature of evolutionary forces affecting chromosomal arrangements other than inversions. The study of the *X/4* fusion polymorphism of *Drosophila americana* is thus of interest. Polymorphism patterns at the *paralytic* (*para*) gene, located at the base of the *X* chromosome, suggest that there is suppressed crossing over in this region between fusion and nonfusion chromosomes but not within fusion and nonfusion chromosomes. These data are thus compatible with previous claims that within fusion chromosomes the amino acid clines found at *fused1* (also located at the base of the *X* chromosome) are likely maintained by local selection. The *para* data set also suggests a young age of the *X/4* fusion. Polymorphism data on *para* and *elav* (located at the middle region of the *X* chromosome) suggest that there is no population structure other than that caused by the *X/4* fusion itself. These findings are therefore compatible with previous claims that selection maintains the strong association observed between the methionine/threonine variants at *fused1* and the status of the *X* chromosome as fused or unfused to the fourth chromosome.

POLYMORPHIC chromosomal arrangements other than inversions are rare in the genus Drosophila and are generally present only as fixed differences between species (Patterson and Stone 1952; Powell 1997). It is thus not surprising that, in Drosophila, while the maintenance of polymorphic chromosomal inversions has received much attention (Krimbas and Powell 1992; Powell 1997; Andolfatto *et al.* 2001), there is only limited evidence on the nature of the evolutionary forces affecting other types of chromosomal arrangements such as chromosomal fusions.

Drosophila americana is a species of the virilis group of Drosophila that possesses a derived *X/4* fusion chromosomal polymorphism (a fusion of Muller's elements A and B, respectively; Muller 1940; Throckmorton 1982). This fusion is distributed through a very wide cline along a latitudinal gradient being at a high frequency in the north of the United States and rare in the south of the United States (Vieira *et al.* 2001; McAllister 2002). For genes on chromosomes *X* [except for *fused1* (*fu1*)], *2*, *3*, and *4*, individuals with and without the *X/4* fusion are indistinguishable at the DNA level (Hilton and Hey 1996, 1997; McAllister and Charlesworth 1999; McAllister and McVean 2000; Vieira *et al.* 2001). These observations suggest that there is considerable gene flow between individuals with and without the *X/4* fusion. The original subspecies designation (*D. a. americana* and *D. a. texana*) based on the presence/absence of the *X/4* fusion is therefore unwarranted (McAllister 2002).

It is conceivable that heterozygosity for the *X/4* fusion may suppress crossing over between the centromere and the *X* chromosome proximal loci (Ashburner 1989, pp. 563–564) and this could explain the observed significant associations between the *X/4* fusion and single nucleotide polymorphisms at *fu1*, since this gene is located ~1–2 Mb away from the block of *X* chromosome centromeric heterochromatin (Vieira *et al.* 2001). Nevertheless, all fusion chromosomes carry a mutation of a methionine to a derived threonine at *fu1* (site 1633), whereas nonfusion chromosomes mostly carry the ancestral state. This amino acid replacement may be advantageous in the *X/4* fusion background or in the ecological conditions prevailing in more northerly areas. Selection maintaining this amino acid difference would reduce effective gene flow between arrangements and hence elevate divergence at linked silent sites between fusion and nonfusion chromosomes when compared with other genes in the *fu1*-centromere region (Charlesworth *et al.* 1997; Vieira *et al.* 2001). The analysis of additional genes in the *fu1*-centromere region, such as the *paralytic* (*para*) gene studied here, in principle, can help to elucidate these issues. Furthermore, the age of the *X/4* fusion chromosome has been tentatively esti-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AJ538200–AJ538219, AJ538221–AJ538252, AJ53254–AJ538256, and AJ538258–AJ538294.

[1] *Corresponding author:* Molecular Evolution Laboratory, Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, Porto 4150-180, Portugal. E-mail: cgvieira@ibmc.up.pt

mated as 0.61 MY (with a lower 95% limit of ∼0.27 MY; Vieira *et al.* 2001) under the assumption of no recombination between fusion and nonfusion chromosomes. Genes located closer to the centromere than *fu1* (such as *para*) can, in principle, also be used to get better estimates of the age of the *X/4* fusion.

Fusion chromosomes present 10 times less variability at *fu1* than nonfusion chromosomes do (Vieira *et al.* 2001). It is known that selection on linked sites can reduce levels of variability (reviewed in Charlesworth and Charlesworth 1998). It is thus of interest to gain insight into the level of crossing over between *fu1* and the centromere within *X/4* fusion chromosomes. This can be partially achieved by comparing *para* silent site variability levels in fusion and nonfusion chromosomes. Furthermore, since Vieira *et al.* (2001) hypothesized that, in the *fu1*-centromere region, crossing-over frequencies could be much lower within fusion than within nonfusion chromosomes due to a putative reduction in the amount of *X* chromosome pericentric heterochromatin in the former relative to the latter, in this work we have also experimentally determined their *X* chromosome heterochromatin levels.

The low variability levels found at *fu1* within fusion chromosomes, however, may not require an explanation in terms of reduced crossing-over frequencies at the base of the *X* chromosome within fusion chromosomes. Within *X/4* fusion chromosomes, there is clinal variation with respect to the three most common replacement polymorphisms but not silent site polymorphisms. In principle, these replacement polymorphisms could be maintained by a balance between gene flow and weak selection (Vieira *et al.* 2001). A recent increase in the frequency of these amino acid polymorphisms could thus be responsible, as well, for the low variability levels observed at *fu1* within fusion chromosomes. Population structure alone, nevertheless, could produce these patterns. There is no evidence for population structure in *D. americana*, but in most studies only a couple of populations have been studied (Hilton and Hey 1996, 1997; McAllister and Charlesworth 1999; McAllister and McVean 2000; Vieira *et al.* 2001). We have thus also analyzed the issue of population structure by studying the *elav* gene (located at about the middle of the *X* chromosome) in four *D. americana* populations including the two populations known to harbor the highest frequencies of these replacement polymorphisms as well as two populations in which their frequencies approach zero.

MATERIALS AND METHODS

DNA samples, PCR amplification, and DNA polymorphism analyses: Genomic DNA of 32 single males from four populations collected in Puxico, Missouri (*PM99*), Lake Ashbaugh, Arkansas (*LA99*), Niobrara, Nebraska (*NN97*), and Gary, Indiana (*G96*), were used to determine the *elav* intron sequence. For 30 out of the 32 individuals analyzed the genomic sequences of *para* intron 3 and in the region between the alternatively spliced elements *c* and *d* were also determined. The *D. americana* populations studied here are described in detail in Vieira *et al.* (2001). For the *PM99* and *LA99* flies, the status of the *X* chromosome as fused or unfused to the fourth chromosome was previously determined (Vieira *et al.* 2001; McAllister 2002). For the *NN97* and the *G96* strains the status of the *X* chromosome was determined by cytological observation of mitotic chromosomes (see below).

The primers used are presented in Table 1 in the supplementary material (http://www.genetics.org/supplemental). Nested or seminested PCR was used for both genes. For the *elav* gene, the primers elF and el1177R were used in the first PCR amplification. The nested PCR was obtained with elAMF and el722R primers. To amplify *para* intron sequences between alternative spliced elements *c* and *d*, primers paraCF and paraCR were used for the first PCR. The sequenced region was obtained with paraC622F and paraCR primers. It should be noted that *D. virilis* intron 3 of *para* is the largest described that does not contain alternative exon sequences (Thackeray and Ganetzky 1995). To amplify this region the primers para34F and para34R were used for the first PCR. The sequenced region was obtained with para34F and para1388R primers. Standard amplification conditions were 35 cycles of denaturation at 94° for 30 sec, primer annealing at 50° for 45 sec, and primer extension at 72° for 3 min. Sequencing of the PCR products was performed with an Applied Biosytems (Foster City, CA) model 310 DNA sequencing system with the ABI PRISM BigDye cycle-sequencing kit (Perkin-Elmer). The DNA sequences were deposited in GenBank (accession numbers for the *elav* gene are AJ538263–AJ538294 and, for the *para* gene, AJ538200–AJ538219, AJ538221–AJ538252, AJ 53254–AJ538256, and AJ538258–AJ538262). Analyses of DNA polymorphism were performed using DnaSP (Rozas and Rozas 1999) and ProSeq version 2.43 (http://helios.bto.ed.ac.uk/evolgen/filatov/proseq.html) software.

Quantification of heterochromatin and *in situ* hybridizations: All fly stocks (http://www.uta.edu/biology/mcallister/bfmflies.html) were grown at 25° in standard media. For quantitative studies of arm length and sister chromatid separation, brains were treated with a hypotonic solution (0.5% sodium citrate) for 10 min before fixation with acetic acid (Carmena *et al.* 1993). Identification of the chromosomes and of heterochromatic regions was performed by 4′,6-diamidino-2-phenylindole staining, adapting the procedure described by Ashburner (1989, p. 10). Preparations were observed with a Zeiss Axioskop microscope, and images were acquired with a SPOT 2 camera (Diagnostic Instruments). Quantification analysis was performed using the University of Texas Health Science Center (San Antonio, TX) Image Tool version 3.00 for Windows (http://ddsdx.uthscsa.edu/dig/download.html).

In situ hybridizations with the polytene chromosomes of *D. americana* strain *NN97.4* were performed as described in Vieira *et al.* (1998) using biotinylated *para* and *elav* amplification products as probes.

RESULTS

Sequence data analyses: The status of the *X* chromosome as fused or unfused to chromosome *4* is known for all individuals analyzed (Vieira *et al.* 2001; see also materials and methods). The *NN97* and *G96* samples used in this work are made exclusively of individuals harboring the *X/4* fusion; the frequency of nonfusion chromosomes is <5% in these two populations (Vieira *et al.* 2001). About half of the individuals in our *PM99*

and *LA99* samples harbor the *X/4* fusion; the frequency of nonfusion chromosomes is ~45 and 51% for the *PM99* and *LA99* populations, respectively (Vieira *et al.* 2001; McAllister 2002).

We have localized the *elav* and *para* genes in *D. americana* to regions 11C and 19C of the *X* chromosome, respectively, using the *D. novamexicana* photographic polytene chromosome map of Vieira *et al.* (1997a) for reference, since these two taxa are homosequential for this region of the *X* chromosome. The *para* gene localization agrees well with that of Päällysaho (2001) who located this gene at the base of the *X* chromosome between *fu1* and the centromere.

We analyzed a region of ~740 bp of *elav* intron sequence in 32 males belonging to four *D. americana* populations (Figure 1). For the *para* gene we analyzed two regions totaling 1660 bp of *para* intron sequence (1200 bp of intron 3 sequence and 460 bp of intron sequence between alternatively spliced elements *c* and *d*; Figure 2) and 300 bp of *para* coding sequence (195 bp of exon 3 and 105 bp of alternatively spliced element *d* of *D. virilis*; Thackeray and Ganetzky 1995) in 30 males belonging to four *D. americana* populations (Figure 2). No synonymous and replacement variants are observed in the *para* coding region analyzed.

For *elav* and *para* genes, the permutation test of Hudson *et al.* (1992a) shows that there is no significant differentiation between populations when both *X/4* fusion and nonfusion chromosomes are considered separately (data not shown). For the population pairwise comparisons, the F_{ST} value (Hudson *et al.* 1992b) varies between 0.4–6.6% and 3–10.3% for *elav* and *para*, respectively. For *elav*, there is no significant differentiation between individuals with and without the fusion when their population origin is ignored (Table 1). It should be noted that the *elav* region analyzed is highly variable (Figure 1) and thus very informative. Furthermore, for *elav* there is also no significant differentiation between populations when both *X/4* fusion and nonfusion chromosomes are considered together (data not shown). For *para* there is, however, significant differentiation between individuals with and without the fusion when their population origin is ignored (Table 1). The permutation test of Hudson *et al.* (1992a) also shows significant differentiation between individuals with and without the *X/4* fusion at *fu1* and *Adh* (Table 1). In the latter, however, the F_{ST} value, the number of fixed differences, and the number of shared and exclusive polymorphisms suggest that there is very little differentiation, in agreement with the findings of McAllister and Charlesworth (1999). There is thus no evidence for any degree of population structure in *D. americana* other than that caused by the *X/4* fusion at the base of the *X* chromosome.

Table 2 shows the estimated level of silent site variability at the *X*-linked genes, *elav fused1* and *para*. For *elav*, the estimated level of variability is similar to that obtained for other genes (Hilton and Hey 1996, 1997; McAllister and Charlesworth 1999; Mcallister and McVean 2000; Vieira *et al.* 2001). There is little silent site variability in the *para* regions analyzed compared with the estimates obtained for *elav* (Table 2). Tests of neutrality (Table 3) show no significant deviations from neutrality in the *elav* data set and in both the fusion and nonfusion *para* data sets.

As expected considering its location, there is clear evidence for recombination at the *elav* gene (Table 4). The *para* gene is located closer to the *X* centromeric block of heterochromatin than *elav* is, but at this gene there is evidence for recombination within fusion chromosomes. There is, however, no evidence for recombination within nonfusion chromosomes, but the reduced number of informative segregating sites precludes a firm conclusion (Table 4).

Heterochromatin content at the base of the *X* chromosome: Vieira *et al.* (2001) hypothesized that heterochromatin levels could be reduced at the base of the *X* chromosome of *X/4* fusion chromosomes compared with those of nonfusion chromosomes as a result of the *X* chromosome *4* translocation. The greater proximity of the centromere could cause suppression of recombination in the proximal euchromatin of fusion chromosomes. To test this hypothesis, we compared heterochromatin levels at the base of the *X* chromosome of fusion and nonfusion chromosomes (Table 5). There are no estimates of the DNA content of individual *D. americana* chromosome arms. If we assume, however, that the genome size of all species of the virilis phylad is similar, a 1% difference in the amount of *X* heterochromatin corresponds to ~210 kb (Vieira *et al.* 1997b). There are no significant differences in *X* heterochromatin content within fusion and nonfusion chromosomes (ANOVA *F*-test; not shown) or between fusion and nonfusion chromosomes (ANOVA *F*-test; Table 5).

DISCUSSION

Gene flow between fusion and nonfusion chromosomes at the base of the *X* chromosome: Patterns of *X* chromosome heterochromatin are different between fusion and nonfusion chromosomes (Figure 3). This is not surprising despite the young age of the *X/4* fusion since heterochromatin patterns evolve rapidly (Spradling 1994). Furthermore, it is possible that the presence of the fusion itself changes heterochromatin patterns on fusion chromosomes.

Significant differentiation has been found between fusion and nonfusion chromosomes at both *fu1* (Vieira *et al.* 2001) and *para* (Table 1). It should be noted that there is no evidence for population structure in *D. americana.* The significant differentiation observed at *fu1* and *para* is thus likely due to low levels of recombination between fusion and nonfusion chromosomes, in a way similar to that of the inverted and standard chro-

		11	11	1122222222222333333333333333333333334444	4444444444444455555555	55555556666666666677
		122444445568933	66	8814557788890001122333555566777788990001	2234444448888890122234	55667780114567777805
		45905645671806202	03	4761161201363896809125013725168915670233	3881237802478338237775	07681242010762589989
NN97.2	F	AAA1GGTGGAGTCA1A(AT)$_4$G-		C11AA-CATGGGGG1TGGAAATCAATAT-CTGG-AGTTC A$_4$	T 1AGGAG-GGAAT1ATGCGA-	TTTTGAAAGCGGTACTC1A
NN97.4	F	.............T-..	..	.2.G.........2........CA.............	.2G..G....TG.-.C.T..AT	....A..G..........2.
NN97.8	F	.T.2A...A.....2..	..	T2.......A..TT-.....GG..C....T.........	..G.........G...C.T...	..C.................
NN97.9	F	.T..A........T-.(AT)$_5$..		.3.........TTT-.....G..GC...1.ATT1..A...	..G..G.....G.-.C.T...	.............A......T
G96.11	F	.T..........T-..	..	.3.G........T-GA...G...C...............A$_5$	.G....................	....
G96.21	F	.T.-.........T-..	..	.2..G.......TT-.....G...C.........T..T A$_2$CA	.G..G.....G.-.C.....	.......G..........-.
G96.36	F	.T..A.......T-..	..	.2.G..............G...C..........A...	A .G...........C.T..AT	.............A.......
G96.47	F	GT..A............	..	.2................CGG..C..............	.2G..G....TG.-.C.T...	.......G............
G96.48	F	.T..A........T-..	..	.............T-.....G...C..........A...	..G..G.....G.-.......	.......G............
LA99.10.11	F	.T..A....CTCTT-..	..	.2.....T..........CGG..C..............	..G..G.....G.-.C.T...	.A.C...G............
LA99.30.3	F	.T..A..C......2..	..	.2.........T-.....G...C...............	..G..G.....G...C.T...	.....T......C-......
LA99.34.6	F	.T..A.........2G.	..	.3................G...C...1.ATT1.A....	..G.TG.1T..G...C.T...	..................G3.
LA99.38.11	F	.TC2AA...CTCT....	..	.3....A...........G...C...............	..G........G...C.T...	..C........A........
LA99.42.11	F	.T..A........T-..	..	...G.........2.....G...C.C.........C..	..G.TG.1TT.G........	...........C......4.
PM99.2.25	F	.T..A........T-..	..	.3...1..C..TTT-.....G...C.......A.-.....	..G..G.....G.-.C.TA..	....
PM99.16.21	F	.T..A......CTT-..	A(TG)$_2$	.4...1......A.2.....G.T.C...............A$_5$	..G..G.....G.-.C.....	....
PM99.12.12	F	.T..A............	..	.3........ATTT-....CGG..C..............	..G..G.....G.-.C.T...	.A........C.........
PM99.30.21	F	.T..A........T-..	..	.3...1..C..TTT-.....G...C.......A.-.....	..G..G.....G.-.C.TA..	....
PM99.32.21	F	.T..A....CTCT....	..	.3...........A-.....G...CA.............	.2G..G....TG.-.C.T..AT	....A...AG..........
PM99.4.12	F	.T..A........T-..	..	.2.G.........T-.....G...C..G.......T...A$_2$CA	..G..GT....G..GC.TA..	C...........C-......
LA99.8.2	nF	.T..A....CTCT....	..	.2...........T-.....G...C...............A$_3$	..G..GT....G...C.T...	.............A.......
LA99.28.11	nF	.T..A......CTT-.(AT)$_3$..		.3.G.......TTT-.....G..GC..............	..G..G.....G.-.C.....	.A.........A........
LA99.30.2	nF	.T..A............	..	.--.........A.TT-.....GG..C..C..........	..G.........C..C.....	....
LA99.40.21	nF	.T..A............	..	.2.........TT-.....GG..C..............	.-G..GT....G...C.T...	.A............G....
LA99.50.11	nF	.T..A.........2..	..	.2...........T-.....G...C.......A......A$_5$	..G............CTT...	.............A.......
LA99.54.11	nF	.T..A....CTCT.-..	..	.32G....C...TT-.....G...C..........A...	..G............C.TA..	.A..................
PM99.2.21	nF	.T..A....CTCTT-..	A.	.3..............A..G...C...............	.2G..G.....G.-.C.T...	.........AG.........
PM99.4.14	nF	.T..A.C......T-..	..	.2.............2.....G...C.C............A$_3$	..G..G.....G.-.C.....	.A....C....A........
PM99.8.11	nF	.T..A........T-..	..	T2.........TT-.....G...C..G.............A$_3$	..G..G.....G.-.C.....	.................G...
PM99.16.22	nF	.T..A.........2..	..	.2.G.......T.T-...G.G...C..............	A .GA.......G...C.TACAT	.............A....G..
PM99.42.21	nF	.T..A........T-..	..	T2.........TTT-.....GG..C..............	..G............C.T..AT	.......G...A......4.
PM99.44.23	nF	.T..A....CTCT....	..	.2.............2.....G...C..G...........A$_3$	..G..G.....G.-.C.....	.A.....G...A........
D virilis		----A......--.3.-	A.	.-3.G2........----....A..C.TA2ACCA2.C...1	..C.TG.2AC..A2.C...-AT	.......-.TA.A----.5.

At site 10: 1 – CAAAGTTTGA; 2- CAAAGTTTAA; site 92: 1 – GGCAGCGAGCTGCCATTAAAAGC; 2 – GGCATCGAGCTGCCATTAAAAGC; 3 – GGCCTTGAGTTGCCATTAGAAGC; site 187: 1 – GCTTATAGAGAAGGGTCTTGAGTATATTG; 2 – GCCTATAGAGAAGGGTCTTGAGTATATTG; 3 – GCCTATAGAAAAGGGTCTTGAGTATATTG;4 -TCCTATAGAGAAGGGTCTTGAGTATATTG; site 216: 1 – CTTCGGTGCTTACAGTCCGA; 2 – CTTCTGTGCTTACAGTCCGA; 3 – GGTCGGTGCTTACAGTCCGA; site 256: 1 – CCTAAATATGATC; 2 – CTTAAATCTGACC; site 309: 1 – GTA; 2 – CTA; site 371: 1 – ACC; 2 – CTG; site 385: 1 – AACCTGAGAAA; 2 – AGGCACATCTA; site 413: 1 – CAAGCTAAT; site 428: 1 – CGATATGTT; 2 – CGATATGTA; site 448: 1 – CAAATATCGACAAGAAATTATAAAACCTGGG; 2 – CAAATATCGACAAGAAATGATAAATTATTAA; site 493: 1 – AGAAG; 2 – GGTGT; site 708: 1 – TTGTAACTAA; 2 – TTGTAAATAA; 3 – TTCTAACTAA; 4 – TCGTAACTAA; 5 – TTGAAACTAA.

FIGURE 1.—*D. americana X/4* fusion (F) and *X* nonfusion (nF) *elav* intron haplotypes. Dots represent the same nucleotide as in the first sequence, and a dash represents an indel. See MATERIALS AND METHODS for population codes.

		Intron 3																																Intron between element *c* and *d*											
		262	268	305	405	426	429	444	447	638	683	725	733	764	785	863	868	910	918	964	975	978	1119	1135	1149	1155	1195	1196	1217	1248	1250	1259	1299	1310	1362	128	238	269	288	307	323	351	405	429	450	481	485
NN97.2	F	C	C	G	$(CT)_6$	T	C	T	T	T	A	$(TA)_2$	T	G	G	A	T	A_4	$(TA)_9$	T	G	A	A_{12}	A	A	C	C	C	G	A	A	-	T	C	C	A	A	T	G	T	G	A	C	A	-	C	G
NN97.4	F	.	.	.	$(CT)_5$	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	G	A_{11}	.	T	T	.	.	A	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	A
NN97.8	F	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	A
NN97.9	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G96.11	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	A_{10}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
G96.21	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G96.36	F	.	.	.	.	.	T	.	.	.	C	.	.	A	.	T	.	.	.	.	.	G	A_{11}	.	T	T	.	.	A	.	.	.	C	.	.	C	.	.	A	.	.	.	.	.	.	.	A
G96.47	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	A_2	$(TA)_{10}$	.	.	G	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
G96.48	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	G	.	G	.	.	T	T	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LA99.10.11	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LA99.30.3	F	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	.	A
LA99.34.6	F	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	A	.	A_{11}	.	T	T	.	.	A	.	.	.	.	G	.	.	.	.	A	.	.	.	.	.	.	.	A
LA99.38.11	F	.	T	.	$(CT)_7$	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
LA99.42.11	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
PM99.2.25	F	.	T	.	CTCC$(CT)_5$	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.16.21	F	.	.	.	.	.	.	A	.	.	C	.	.	A	.	T	.	.	$(TA)_{10}$	.	.	G	A_{11}	T	T	T	.	.	A	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.12.12	F	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	A_2	$(TA)_{10}$	.	.	G	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.30.21	F	.	.	.	.	.	.	.	.	A	C	.	.	A	.	.	.	.	.	.	.	.	A_{10}	.	T	T	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.32.21	F	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	.	A
PM99.4.12	F	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	.	.	.	.	.	.	A_{11}	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
LA99.8.2	nF	T	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	$(TA)_8$	.	.	.	A_9	.	T	T	.	.	A	.	.	.	.	.	A	.	T	.	A	.	.	.	A	.	.	.	A
LA99.28.11	nF	.	.	.	$(CT)_5$	.	.	.	.	.	C	.	.	.	.	.	.	.	$(TA)_{10}$	.	.	G	A_{13}	.	T	T	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	C	1	T	A
LA99.40.21	nF	.	.	A	.	.	.	.	.	.	C	.	.	.	.	.	.	.	$(TA)_8$	.	.	G	.	.	T	T	.	.	A	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	A
LA99.50.11	nF	.	.	.	.	C	.	.	.	.	C	.	.	.	A	.	.	.	.	.	.	.	.	.	T	T	.	.	A	G	C	1	C	.	.	.	.	.	A	.	.	.	.	.	.	.	A
LA99.54.11	nF	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A_{11}	.	T	T	A	T	A	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	A
PM99.4.14	nF	.	.	.	$(CT)_5$	.	.	.	.	.	C	TA	-	.	.	.	.	.	.	.	.	.	A_{11}	.	T	T	.	.	A	.	.	.	.	.	.	.	.	A	A	.	.	.	.	C	.	.	A
PM99.8.11	nF	.	.	.	.	.	.	.	.	.	C	$(TA)_4$	.	.	.	.	A	A_8	$(TA)_7$	.	.	.	A_{13}	.	T	T	.	.	A	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.16.22	nF	.	.	.	$(CT)_4$	.	.	.	.	.	C	.	.	.	.	.	.	.	$(TA)_{11}$	.	.	.	.	.	T	T	.	.	A	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	A
PM99.42.21	nF	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	$(TA)_8$	.	.	.	A_9	.	T	T	.	.	A	.	.	.	C	.	.	.	.	A	A	.	.	.	.	C	.	.	A
PM99.44.23	nF	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	$(TA)_{10}$	.	.	.	A_{10}	.	T	T	.	.	A	.	.	.	.	.	A	.	.	.	A	.	.	.	.	C	1	T	A
D. virilis	nF	.	.	.	-	.	.	.	.	.	C	$(TA)_3$	.	.	.	.	.	A_2	.	.	.	.	A_{14}	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	2	.	A

intron 3 at site 1259: 1 - TGTCTATGTATA; sequence between element *c* and *d* at site 450: 1- AATCCGT; 450 - 2 - (AATCCGT)$_2$AATTCGT

FIGURE 2.—*D. americana X/4* fusion (F) and *X* nonfusion (nF) *para* haplotypes. Dots represent the same nucleotide as in the first sequence, and a dash represents an indel. See MATERIALS AND METHODS for population codes.

TABLE 1

Differentiation between chromosome types for the *X*-linked *elav*, *fu1*, and *para* genes and the fourth chromosome *Adh* gene

Gene	F_{ST} values[c]	No. of polymorphisms		No. of fixed differences	N_{LD}	Permutation test[d]
		Shared	Exclusive			
elav	0.019	31	56	0	0	$P > 0.05$
fu1[a]	0.570	1	51	7	9	$P < 0.001$
para	0.263	4	36	0	3	$P < 0.001$
Adh[b]	0.013	17	22	0	0	$P < 0.05$

N_{LD} is the number of sites showing strong association with the *X/4* fusion.

[a] Based on the G96 *X/4* fusion and the FP99 *X* free sequence data of Vieira *et al.* (2001).

[b] Based on the G96 *X/4* fusion and the LP97 nonfusion fourth chromosome sequence data of McAllister and Charlesworth (1999).

[c] Hudson *et al.* (1992b).

[d] Hudson *et al.* (1992a).

mosomal arrangements around inversion breakpoints (Andolfatto *et al.* 2001), as well as the result of the original selective sweep that brought the *X/4* fusion to a high frequency. There are, nevertheless, shared polymorphisms between fusion and nonfusion chromosomes at both *fu1* (Vieira *et al.* 2001) and *para* (Table 1). Since the *X/4* fusion is a unique event, this observation indicates that gene flow between the two chromosome types cannot be completely suppressed in the *fu1*-centromere euchromatic region. The shared polymorphisms may be the result of gene conversion between *X/4* fusion and nonfusion chromosomes, as in the regions around the inversion breakpoints where gene conversion is more important than crossing over (Navarro *et al.* 1997). Although in the *para* data sets the observed associations between variants in *X/4* fusion chromosomes (between sites 1149 and 1299) are suggestive of gene conversion between nonfusion and *X/4* fusion chromosomes, no gene conversion tracts are detected using the Betrán *et al.* (1997) test. Since statistical tests for departures from neutrality based on levels of association between sites (Kelly 1997; Wall 1999; Table 3) are not significant, it seems that variability at *para X/4* fusion chromosomes is not greatly increased due to gene conversion between the two chromosome types.

It is conceivable that there are chromosome pairing difficulties at the base of the *X* chromosome in heterozygotes for the *X/4* fusion and nonfusion *X* chromosomes that would result in little gene flow between the two types of *X* chromosomes (Vieira *et al.* 2001). For the *fu1*-centromere region the number of migrants per generation between the population of fusion and nonfusion chromosomes can be estimated from F_{ST} values (Hudson *et al.* 1992b). This value is also an estimate of the number of recombinant individuals per generation in the *fu1*-centromere region. For both the *para* and *fu1* loci the number of recombinant individuals in this region is <1 (0.94 and 0.88, respectively). In general, it has been stated that one migrant per generation is enough to prevent the effects of genetic drift among populations (Hedrick 2000, p. 289). Therefore, although a particular theoretical model at equilibrium is being used to get these estimates and they should thus be cautiously interpreted, it is conceivable that genetic drift could account for the observed significant differentiation between fusion and nonfusion chromosomes in the *fu1*-centromere region, together with the effect of the original selective sweep that brought the *X/4* fusion to a high frequency. A direct estimate of crossing over between fusion and nonfusion chromosomes, however, is still required to validate the estimated levels of gene flow. Under the assumption of less than one migrant per generation between fusion and nonfusion chromosomes in the *fu1*-centromere region, significant associations between single nucleotide polymorphisms in this region and the status of the *X* chromosome as fused or unfused to the fourth chromosome are expected. So far, significant associations have been found at *fu1* and *para* for nine (Vieira *et al.* 2001) and three single nucleotide polymorphisms, respectively (Table 1).

Of all the *fu1* and *para* single nucleotide polymorphisms surveyed so far, the methionine/threonine replacement variants at *fu1* site 1633 show the strongest association with the status of the *X* chromosome as fused or unfused to the fourth chromosome. In a sample of 48 fusion and 47 nonfusion chromosomes only 6.3% of the *X/4* fusion chromosomes are associated with the *fu1* 1633 variant that is present at 100% frequency within nonfusion chromosomes (Vieira *et al.* 2001). Selection could thus also play a role in the maintenance of this association. Since there is no suppression of crossing over at the base of the *X* chromosome within both fusion and nonfusion chromosomes (see below), if this hypothesis were true, in principle, divergence between fusion and nonfusion chromosomes at *fu1*-linked silent sites should be higher than that observed for other genes in the *fu1*-centromere region. The comparison of the *fu1* and *para* loci could thus be informative. The average

TABLE 2

Summary of *D. americana* silent site sequence variation at three *X*-linked genes

		F	nF	Both
elav (11C)	S	76	42	82
	π	0.0224 ± 0.0104	0.0196 ± 0.0090	0.0219 ± 0.0092
	θ	0.0313 ± 0.0110	0.0215 ± 0.0082	0.0326 ± 0.0095
	θ_L	0.0160	0.0094	0.0161
	θ_U	0.0670	0.0520	0.0576
	L	685	646	624
	N	20	12	32
	K	0.1807	0.1832	0.1864
fu1 (18C)[a]	S	4	37	
	π	0.0021 ± 0.0013	0.0152 ± 0.0076	
	θ	0.0018 ± 0.0011	0.0178 ± 0.0076	
	θ_L	0.0004	0.0075	
	θ_U	0.0074	0.0500	
	L	789.37	782.11	
	N	10	9	
	K	0.0908	0.0961	
para (19C)	S	24	20	
	π	0.0033 ± 0.0017	0.0031 ± 0.0016	
	θ	0.0040 ± 0.0015	0.0042 ± 0.0019	
	θ_L	0.0019	0.0018	
	θ_U	0.0089	0.0116	
	L	1707.83	1700.83	
	N	20	10	
	K	0.0356	0.0369	

F, *X/4* fusion chromosomes; nF, nonfusion chromosomes. S is the number of segregating sites; π (Nei 1987) is the average number of pairwise nucleotide differences per base pair, and θ is Watterson's estimator of $3N_e\mu$ (where N_e is the effective population size and μ the neutral mutation rate) based on the number of segregating sites (Watterson 1975). For θ_L and θ_U, the 95% confidence intervals of θ were calculated according to Kreitman and Hudson (1991). L is the number of silent sites analyzed. N is the sample size. K is silent site divergence between *D. americana* and *D. virilis* after Jukes-Cantor correction (Jukes and Cantor 1969). For *fu1* and *para*, no estimate is given for fusion and nonfusion chromosomes together since there is significant differentiation between the two chromosome types at these loci. The standard deviations of π and θ due to stochastic factors, including sampling variance, were calculated according to Nei (1987, pp. 254–258) and Tajima (1993, pp. 37–59) under the conservative assumption of no recombination. Gene locations are shown in parentheses.

[a] From Vieira *et al.* (2001).

number of silent site differences between fusion and nonfusion chromosomes at *fu1* and *para* is, however, proportional to the average number of silent site differences between *D. americana* and *D. virilis* at these loci (16.46/7.45 and 68.85/59.63, respectively; $P > 0.05$; contingency table *G*-test). Nevertheless, when we compare the degree of association [using either D' or R^2 values (Lewontin 1988)] of *fu1* and *para* variants with the status of the *X* chromosome as fused or unfused to the fourth chromosome, significant stronger associations are found for *fu1* than for *para* variants ($P < 0.05$ in both cases; Mann-Whitney *U*-test). Since *para* is located between *fu1* and the centromere, in the absence of selection-distorting patterns of variability at *fu1*, the opposite pattern would be expected. Thus some evidence suggests that the methionine/threonine variants at *fu1* site 1633 and the status of the *X* chromosome as fused or unfused to the fourth chromosome is maintained by selection. This conclusion, nevertheless, should be interpreted with caution since the assumption of independence of the Mann-Whitney *U*-test may be violated if common *fu1*-derived variants have a correlated genetic history as the result of the original selective sweep that brought the *X/4* fusion to a high frequency. It should be noted, however, that on average only two of the six *fu1* most common derived variants are expected to have been associated with this event (Vieira *et al.* 2001).

Crossing-over levels within fusion and nonfusion chromosomes: Fusion chromosomes present 10 times less variability at *fu1* than nonfusion chromosomes do and this difference has been shown to be significant using a coalescent approach (Vieira *et al.* 2001). Furthermore, when the Hudson-Kreitman-Aguadé (HKA) test (Hudson *et al.* 1987) is performed on *para-fu1* and *elav-fu1* sequences from *X/4* fusion chromosomes, using

TABLE 3

Summary of four neutrality tests: no significant deviations from neutrality detected

	elav:	*para*			
		Intron 3		Intron between elements *c* and *d*	
Statistical test	All chromosomes	F	nF	F	nF
Tajima's D[a]	−1.38	−0.58	−1.34	−0.53	−0.74
Kelly's Z_{ns}[b]	0.036	0.099	0.152	0.083	0.086
Wall's B[c]	0.070	0.118	0.250	NA	NA
Wall's Q[c]	0.126	0.167	0.385	NA	NA

F, *X/4* fusion chromosomes; nF, nonfusion chromosomes; NA, not applicable.
[a] Tajima (1989).
[b] Kelly (1997).
[c] Wall (1999).

D. virilis sequences and silent sites only, the results are significant (Table 6). No other HKA tests using *para*, *elav*, and *fu1* sequences are significant. These results thus strongly suggest that there is a polymorphism deficit at *fu1 X/4* fusion chromosomes. Adaptive or purifying selection on linked sites can reduce levels of variability (reviewed in Charlesworth and Charlesworth 1998), and in Drosophila reductions in the amount of pericentric heterochromatin are known to cause suppression of recombination in proximal euchromatin (Yamamoto and Miklos 1978). The amount of *X* chromosome centromeric heterochromatin, however, is similar for fusion and nonfusion chromosomes (Table 5), suggesting that crossing-over levels in the *fu1*-centromere region are similar in the two types of chromosomes. Whether there is significant linkage disequilibrium between *fu1* and *para* variants within fusion chromosomes could not be tested, however, because of the small number of segregating sites found at *fu1* within fusion chromosomes (Table 2). Levels of silent site variability at *para* (located between *fu1* and the centromere) are, however, similar in fusion and nonfusion chromosomes (Table 2). It is thus highly unlikely that crossing-over frequencies are different at the base of the *X* chromosome in fusion and nonfusion chromosomes. Since the HKA tests using *para* and *elav* sequences or *para* and *fu1* nonfusion chromosomes are nonsignificant (Table 6), the low levels of silent site variability at *para* can be attributed to high degree of constraint on synonymous and intron sites. This result was unexpected, since the level of conservation in the intron regions between *D. virilis* and *D. melanogaster* is so low that, except for the conserved 5′ and 3′

TABLE 4

Summary of recombination statistics

	elav:	*para*[a]	
	Both	F	nF
Si	43	8 + 3	3 + 3
4GT	300 (903)	12 (31)	0 (6)
R_m	14	2	0
LD_{FB}	7 (903)	5 (31)	0 (6)
LD_{dist}	2, 4, 7, 9, 11, 13, 15	6, 62, 68, 107, 176	—

F, *X/4* fusion chromosomes; nF, nonfusion chromosomes. Si is the number of informative segregating sites; 4GT is the number of pairwise comparisons presenting the four gametic types; R_m is the minimum number of recombination events (Hudson and Kaplan 1985); LD_{FB} is the number of sites showing significant linkage disequilibrium using Fisher's exact test after Bonferroni correction for multiple comparisons; LD_{dist} is the distance in base pairs between sites showing significant linkage disequilibrium. The total number of pairwise comparisons is shown in parentheses.

[a] The two *para* regions (intron 3 and the intron region between elements *c* and *d*) were analyzed separately and the results combined.

TABLE 5

Amount of *X* centromeric heterochromatin expressed as prometaphase *X* chromosome length of heterochromatin over the total length of the *X* chromosome arm of *D. americana*

Chromosomes	Strain	N	*X* chromosome heterochromatin content (%)
Fusion	*NN97.2*	22	49.28 ± 4.75
	NN97.4	19	52.77 ± 5.27
	NN97.8	35	50.00 ± 5.08
	NN97.9	21	49.93 ± 4.28
	G96.11	31	49.94 ± 3.46
	G96.21	37	51.13 ± 5.00
	G96.36	30	49.05 ± 4.93
	G96.46	25	50.85 ± 5.72
	G96.48	43	50.74 ± 4.57
Average			50.41
Nonfusion	*ML97.3*	39	50.59 ± 3.57
	ML97.5	19	52.77 ± 5.27
	LP97.7	20	51.13 ± 4.66
	CD97.5	18	52.94 ± 3.66
Average			51.86

N, the number of mitoses analyzed.

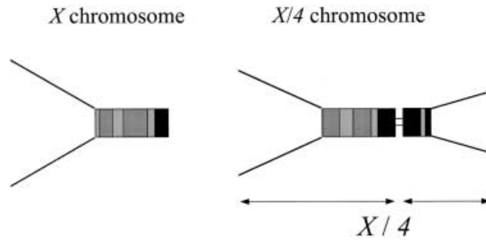

Figure 3.—Schematic representation of the centromeric and pericentromeric heterochromatin patterns for nonfused *X* and fused *X/4* chromosomes. The black, dark-gray, and light-gray boxes correspond to different staining intensities. Black boxes represent the brightest bands and light-gray boxes the least-stained ones.

splice sites, the alignments are ambiguous (Thackeray and Ganetzky 1995). Levels of polymorphism found at *fu1* in nonfusion chromosomes are similar to those reported for genes located elsewhere in the genome (Hilton and Hey 1996, 1997; McAllister and Charlesworth 1999; McAllister and McVean 2000; Vieira *et al.* 2001; McAllister 2002; see also results). In Drosophila and many other genera, variability levels and crossing-over levels are correlated (reviewed in Charlesworth and Charlesworth 1998). Therefore, the above-mentioned observations suggest that there is no suppression of crossing over at the base of the *X* chromosome within both fusion and nonfusion chromosomes. In *D. virilis*, a species closely related to *D. americana*, there is also little or no suppression of recombination at the base of the *X* chromosome (Vieira and Charlesworth 1999). Moreover, the *D. americana Adh* locus located on the fourth chromosome at ~1 Mb away from centromeric heterochromatin also shows variability levels compatible with no suppression of recombination (McAllister and Charlesworth 1999). The low levels of silent site variability observed at *fu1* in fusion chromosomes relative to nonfusion chromosomes is thus due to factors other than low levels of recombination at the base of the *X* chromosome.

Low levels of crossing over between fusion and nonfusion chromosomes (see above) and high levels of crossing over within fusion and nonfusion chromosomes imply that any selective sweep in the *fu1*-centromere region should affect only one type of *X* chromosome (either fusion or nonfusion chromosomes) and that levels of variability should be affected only in the vicinity of the selection target.

Frequency clines within *X/4* fusion chromosomes: Vieira *et al.* (2001) previously noted that among chromosomes with the *X/4* fusion, there are significant correlations between latitude and longitude and the frequency of the three most common amino acid polymorphisms (at positions 442, 1609, and 2157) at the *fu1* gene. All three replacement variants are derived and are likely younger than the *X/4* fusion since they are common only in fusion chromosomes. In contrast, there is no evidence for clinal patterns for silent variants within the fusion chromosomes or for nonfusion chromosomes. This evidence suggests that these clines are the result of differential selection pressures in different parts of the species range, although the role of population structure could not be completely ruled out at the time. The analysis of the highly polymorphic, and thus highly informative, *elav* gene shows that there is no significant population structure in *D. americana*. No population structure has been detected when either the *para* gene of *X/4* fusion chromosomes or the nonfusion chromosomes are analyzed. It should be noted that the *elav* and *para* data sets include individuals from the *NN97* and *G96* populations in which the three replacement variants are most common and from the *LA99* and *PM99* populations in which these replacement variants are present at very low frequency (Vieira *et al.* 2001). If the amino-acid gradients were due to population structure, we should thus have detected it. Therefore, differential selection pressures in different parts of the species range of *D. americana* seem to maintain the frequency gradients for the three most common replacement polymorphisms within fusion chromosomes. Furthermore, the level of variability at *para* within fusion and nonfusion chromosomes is similar, indicating a comparable effec-

TABLE 6

***HKA* tests using the *X*-linked *elav*, *fu1*, and *para* genes**

	F					nF				
	S	*N*	*L*	$K_{a\text{-}v}$	HKA	*S*	*N*	*L*	$K_{a\text{-}v}$	HKA
elav-fu1	41[a]–4	32–10	624–789.37	88.63–67.45	$P < 0.05$	41[a]–37	32–9	624–782.11	88.63–70.33	$P > 0.05$
para-fu1	24–4	20–10	1707.83–789.37	59.24–67.45	$P < 0.05$	20–37	10–9	1700.83–782.11	60.41–70.33	$P > 0.05$
para-elav	24–41[a]	20–32	1707.83–624	59.24–88.63	$P > 0.05$	20–41[a]	10–32	1700.83–624	60.41–88.63	$P > 0.05$

F, *X/4* fusion chromosomes; nF, nonfusion chromosomes; *S*, number of silent segregating sites; *N*, *D. americana* sample size; *L*, number of silent sites analyzed; $K_{a\text{-}v}$, average number of silent site differences between *D. americana* and *D. virilis*.

[a] Since for *elav* there is no evidence for genetic differentiation between *X/4* fusion and nonfusion chromosomes (in contrast with *fu1* and *para* genes), and since the levels of silent site variability at *para* are similar for *X/4* fusion and nonfusion chromosomes, the effective population size for *elav* is inferred to be double that for *fu1* and *para*. The number of *elav* segregating sites used for the *HKA* tests ($S = 41$) is thus half of those shown in Table 2 for a sample of 32 sequences and 624 silent sites analyzed.

tive population size for the two types of *X* chromosomes. Since the *para* gene is located closer to the *X* chromosome centromere than the *fu1* gene is, and the observed low level of nucleotide variation at *fu1* within *X/4* fusion chromosomes is incompatible with a selective sweep occurring 0.27 MYA (the lower 95% limit for the age of the *X/4* fusion; Vieira *et al.* 2001), it is very likely that the low variability observed at *fu1* for fusion chromosomes is due to the recent spread of gametes with the three most common replacement polymorphisms. As predicted by this hypothesis, the southernmost sample of fusion chromosomes (where these replacement variants are absent) is the most variable at *fu1* gene silent sites (Vieira *et al.* 2001).

The age of the *X/4* fusion: On the basis of the apparent synonymous site substitution frequency between fusion and nonfusion chromosomes at *fu1*, the age of the *X/4* fusion has been previously estimated as 0.61 MY (Vieira *et al.* 2001). At the *para* gene, no apparent silent site fixed differences have been found between fusion and nonfusion chromosomes out of 17,083 silent sites analyzed (Table 1). Assuming a Poisson distribution, an expected maximum of 2.99 apparent silent site fixed differences is compatible with the observed value of no apparent silent site fixed differences. Using the same approach as in Vieira *et al.* (2001), 0.57 apparent silent site fixed differences are expected to have occurred due to the putative selective sweep that brought the *X/4* fusion to high frequency in the northerly areas of the *D. americana* distribution. A maximum of 2.42 apparent silent site fixed differences is thus expected to have occurred in the neutral period that followed this putative selective sweep. To estimate the age of the *X/4* fusion from the *para* data set we use a substitution rate (3×10^{-3}/site/MY) that is 3.3 times smaller than that used for *fu1* since the level of *para* silent site divergence is 3.3 times less than that of *fu1* synonymous site divergence. The *para* data set thus suggests that the *X/4* fusion is younger than 0.47 MY. This value is compatible with that estimated from the *fu1* data set (0.61 MY with a lower 95% limit of ~0.27 MY; Vieira *et al.* 2001).

The inferred evolutionary history of the polymorphic *X/4* fusion of *D. americana*: The *X/4* fusion of *D. americana* is a relatively young event and is likely not older than 0.5 MY. This event did not lead to a significant loss of *X* chromosome heterochromatin or to a change in recombination levels within *X/4* fusion chromosomes at the base of the *X* chromosome. There is, however, evidence for reduced gene flow in a 2-Mb region at the base of the *X* chromosome between fusion and nonfusion chromosomes, likely due to regional pairing difficulties between the two chromosomal types. In contrast, a region of chromosome *4* that is ~1 Mb away from the block of heterochromatin shows only very weak signs for suppression of recombination between fusion and nonfusion chromosomes. If the *X/4* fusion itself is advantageous in the northerly geographic areas of the *D. americana* distribution, then the frequency of this chromosomal arrangement may have rapidly increased in frequency soon after this chromosomal arrangement took place. Alternatively, a neutral *X/4* fusion may have persisted for some time at a low frequency in a restricted geographical distribution. Subsequently, a mutation that is advantageous only in this genetic background or in the northerly geographic areas of the *D. americana* distribution took place within *X/4* fusion chromosomes, bringing this chromosomal rearrangement to a high frequency in these localities. The inferences made here on the level of crossing over between fusion and nonfusion chromosomes in the *fu1*-centromere region suggests that any loci in this region (including *fu1*) could have influenced the increase in frequency of the *X/4* fusion. The high levels of crossing over within fusion chromosomes imply that any selective sweep in this region should affect levels of variability only in the vicinity of the selection target. Selection may maintain the very strong association between the *fu1* methionine/threonine variants at site 1633 and the status of the *X* chromosome as fused or unfused to the fourth chromosome. Thus, *fu1* may be one of the genes responsible for the maintenance of the *X/4* fusion gradient. As suggested before, the cline for the *X/4* fusion is thus very likely maintained by a balance between gene flow and weak selection on the karyotypes themselves or on associated genes (Barton and Gale 1993; Vieira *et al.* 2001). The significant correlations between latitude and longitude and the frequency of the three most common amino-acid polymorphisms (at positions 442, 1609, and 2157) at *fu1 X/4* fusion chromosomes is likely due to differential selection pressures in different parts of the species range and happened later in the *X/4* fusion history.

We thank B. Charlesworth and B. McAllister for comments on an early version of this manuscript. We also thank I. Gordo for helping us calculate the confidence intervals for Watterson's estimate of θ. C. P. Vieira is supported by the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BPD/5592/2001). P. A. Coelho is supported by FCT (SFRH/BPD/5647/2001). This work has been partially funded by FCT (research project 37421/BSE/2001).

LITERATURE CITED

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

Ashburner, M. (Editor), 1989 Staining mitotic chromosomes with Hoechst 33258, p. 10 in *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Barton, N. H., and K. S. Gale, 1993 Genetic analysis of hybrid zones, pp. 13–45 in *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison. Oxford University Press, Oxford.

Betran, 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.

Carmena, M., J. P. Abad, A. Villasante and C. Gonzalez, 1993 The *Drosophila melanogaster dodecasatellite* sequence is closely linked to the centromere and can form connections between sister chromatids during mitosis. J. Cell Sci. **105:** 41–50.

Charlesworth, D., and B. Charlesworth, 1998 Sequence variation: looking for effects of genetic linkage. Curr. Biol. **8:** R658–R661.

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

Hedrick, P. W., 2000 *Genetics of Populations*. Jones & Bartlett, Boston.

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

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

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

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, 1992a A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992b Estimation of levels of gene flow from population data. Genetics **132:** 583–589.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, p. 21 in *Mammalian Protein Metabolism*, edited by H. N. Munro. Academic Press, New York.

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

Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh*-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

Krimbas, C. B., and J. R. Powell, 1992 *Drosophila Inversion Polymorphism*. CRC Press, Boca Raton, FL.

Lewontin, R. C., 1988 On measures of gametic disequilibrium. Genetics **120:** 849–852.

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

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 G. A. McVean, 2000 Neutral evolution of the sex-determining gene *transformer* in Drosophila. Genetics **154:** 1711–1720.

Muller, H. J., 1940 Bearings of the Drosophila work on systematics, pp. 185–268 in *New Systematics*, edited by J. Huxley. Clarendon Press, Oxford.

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

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

Päällysaho, S., 2001 In situ hybridisation analysis of the X-linked genes in the species of the virilis group of Drosophila. Genetica **1772:** 1–7.

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

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

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Spradling, A. C., 1994 Transposable elements and the evolution of heterochromatin. Soc. Gen. Physiol. Ser. **49:** 69–83.

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

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

Thackeray, J. R., and B. Ganetzky, 1995 Conserved alternative splicing patterns and splicing signals in the Drosophila sodium channel gene *para*. Genetics **141:** 203–214.

Throckmorton, L. H., 1982 The virilis species group, pp. 227–296 in *The Genetics and Biology of Drosophila*, Vol. 3b, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, New York.

Vieira, J., and B. Charlesworth, 1999 X chromosome DNA variation in *Drosophila virilis*. Proc. R. Soc. Lond. Ser. B **266:** 1905–1912.

Vieira, J., C. P. Vieira, D. L. Hartl and E. R. Lozovskaya, 1997a Discordant rates of chromosome evolution in the *Drosophila virilis* species group. Genetics **147:** 223–230.

Vieira, J., C. P. Vieira, D. L. Hartl and E. R. Lozovskaya, 1997b A framework physical map of *Drosophila virilis* based on P1 clones: applications in genome evolution. Chromosoma **106:** 99–107.

Vieira, J., C. P. Vieira, D. L. Hartl and E. R. Lozovskaya, 1998 Factors contributing to the hybrid dysgenesis syndrome in *Drosophila virilis*. Genet. Res. **71:** 109–117.

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

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–275.

Yamamoto, M., and G. L. Miklos, 1978 Genetic studies on heterochromatin in *Drosophila melanogaster* and their implications for the functions of satellite DNA. Chromosoma **66:** 71–98.

Communicating editor: M. Aguadé