Genet. Res., Camb. (2006), **87**, *pp.* 163–174. © 2006 Cambridge University Press
doi:10.1017/S0016672306008147 Printed in the United Kingdom

On the location of the gene(s) harbouring the advantageous variant that maintains the *X/4* fusion of *Drosophila americana*

CRISTINA P. VIEIRA, ANDRÉ ALMEIDA, JOÃO DANIEL DIAS AND JORGE VIEIRA*
Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

(*Received 9 June 2005 and in revised form 10 October 2005 and 9 March 2006*)

Summary

Weak selection is maintaining the *Drosophila americana X/4* fusion chromosomal frequency cline. The gene(s) harbouring the advantageous variant(s) that is responsible for the establishment and maintenance of this chromosomal frequency gradient must be located in a region of the *X* and/or *4th* chromosome that is genetically isolated between the *X/4* fusion and non-fusion forms. The limits of these regions must thus be determined before an attempt is made to identify these genes. For this purpose, the correspondence between the *D. virilis X* and *4th* chromosome genome scaffolds sequence and the *D. americana* gene order was established. Polymorphism levels and patterns at seven genes located at the base of the *D. americana X* chromosome, as well as three genes located at the base of the *4th* chromosome, were analysed. The data suggest that the *D. americana X/4* fusion is no more than 29 000 years old. At the base of the *X* chromosome, there is suppression of recombination within *X/4* fusion and non-fusion chromosomes, and little recombination between the two chromosomal forms. Apparent fixed silent and replacement differences are found in three of seven genes analysed located at the base of the *X* chromosome. There is no evidence for suppression of recombination between fusion and non-fusion chromosomes at the base of the *4th* chromosome. The advantageous variant responsible for the establishment in frequency and maintenance of the *X/4* fusion is thus inferred to be in the *D. americana X* centromere–inversion *Xc* basal breakpoint region.

1. Introduction

Understanding geographical patterns of adaptation and its molecular basis is one of the main aims of population genetics. *D. americana*, a species of the virilis group of species, that segregates for an *X/4* fusion (Muller's elements *A* and *B*, respectively; Muller, 1940) and several common chromosomal inversions (Hsu, 1952) is suitable for addressing this issue since some of the segregating chromosomal rearrangements may be maintained by local selection.

The *D. americana X/4* fusion is present as a shallow gradient, being very frequent in the north of the geographic distribution of the species and rare in the south (Vieira *et al.*, 2001; McAllister, 2002). Chromosomal inversion *Xc* is present in 94·6% of the *D. americana X/4* fusion chromosomes but in only about 7% of the non-fusion chromosomes (Hsu, 1952). Therefore, there is a gradient for the *Xc* inversion as well. Weak selection probably brought the *X/4* fusion to high frequency and is maintaining the *X/4* fusion *Xc* inversion chromosomal arrangement (Vieira *et al.*, 2001, 2003). The gene(s) harbouring the advantageous variant(s) responsible for the establishment of the polymorphic *X/4* chromosomal fusion must be in linkage disequilibrium with the *X/4* fusion chromosomal polymorphism and thus located in a region of the *X* and/or *4th* chromosome that is genetically isolated between the *X/4* fusion and non-fusion forms (Vieira *et al.*, 2001, 2003).

The observed strong association between inversion *Xc* and the *X/4* fusion suggests that the region where

* Corresponding author. Molecular Evolution Laboratory, Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823 Porto 4150-180, Portugal. Tel: +351 22 6074900. Fax: +351 22 6099157. e-mail: jbvieira@ibmc.up.pt

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

recombination is suppressed between fusion and non-fusion chromosomes may extend as far as the basal *Xc* inversion breakpoint. It should be noted that the regions around both *Xc* inversion breakpoints are probably regions of lowered recombination between *X/4* fusion *Xc* inverted chromosomes and non-fusion standard chromosomes. It is therefore also conceivable that the target of selection is located around the distal *Xc* inversion breakpoint (Andolfatto *et al*., 2001).

If the *D. americana X* chromosomal region between the centromere and the basal *Xc* inversion breakpoint is genetically isolated between the *X/4* fusion and non-fusion forms, then significant differentiation, including apparent fixed differences and few shared polymorphisms, should be found for genes located within this region, as for the *fused1* gene (Vieira *et al*., 2001). In this work we test this prediction by the analysis of seven gene regions located in this interval, in *X/4* fusion and non-fusion chromosomes. This is only feasible because it is possible to determine the *D. americana* gene order from the *D. virilis* gene order with a good degree of certainty, as done here.

From the DNA sequence data obtained here, it was also possible to obtain an estimate for the age of the *Xc* inversion and of the *X/4* fusion, as well of the relative role of these chromosomal rearrangements in the shaping of the recombination environment at the base of the *X* chromosome. Such knowledge is important for inferring the possible locations of the gene(s) harbouring the advantageous variant(s) that maintain the *X/4* fusion, and conceivably the *Xc* inversion of *D. americana*.

In sharp contrast to the base of the *X* chromosome, suppression of recombination at the base of the *4th* chromosome due to the presence of the *X/4* fusion alone may affect only a small region at the very base of the chromosome (McAllister & Charlesworth, 1999; McAllister, 2003). A gene located at the base of the *D. americana 4th* chromosome is analysed here, corroborating the previous observations.

2. Materials and methods

(i) *Random sequencing of* D. virilis *P1 clone inserts*

DNA extraction from large-insert *D. virilis* P1 clones was performed as described in Vieira *et al*. (1997a). The DNA was then partially restricted with *Sau*3AI and *Rsa*I, and an overhang A added to the DNA fragment ends. For this purpose, a reaction containing one unit of *Taq* DNA polymerase and the PCR buffer supplied by Fermentas (Fermentas, MD, USA), 3 mM Mg^{2+} and 10 mM dATP was added to the restricted DNA fragments, and the mixture incubated at 72 °C for 20 min. The samples were then separated in a 1·2% agarose gel and the fragments in the size range 1·5–3 kb excised from the gel. In order to purify these DNA fragments from the gel, the QIAEXII Gel Extraction Kit (QIAGEN, CA, USA) was used. The DNA fragments were then cloned using the TOPO-TA Cloning Kit for Sequencing from Invitrogen (Invitrogen, Spain). Positive colonies were grown in LB medium, and the DNA extracted using the QIAprep Spin Miniprep Kit (QIAGEN, CA, USA). The plasmid DNA of several colonies with different insert sizes or with different restriction patterns was sequenced using the universal primers M13F and M13R and the Applied Biosystems Model 310 DNA sequencing system with the ABI PRISM BigDye cycle-sequencing kit version 1.1 (Perkin Elmer, CA, USA). After visual inspection of the chromatograms, the DNA sequences were conceptually translated in all possible six open reading frames and compared with the *D. melanogaster* proteins using a BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/), in order to determine whether they contained fragments of *D. virilis* genes.

(ii) *Correspondence between the* D. virilis *polytene chromosome map and the* D. virilis *genome scaffolds*

The sequencing and assembly of the *D. virilis* genome was performed by Agencourt Bioscience Corporation, Beverly, MA, USA, using the Arachne assembler. For this work we have used the assembly made available by the Smith group (Agencourt Biosciences Corporation) on 12 July 2004. This work has been funded by the National Human Genome Research Institute (NHGRI), USA. The UCSC *D. virilis* Genome Browser has been developed by Angie Hinrichs, Brian Raney, Heather Trumbower, Robert Kuhn, Jennifer Jackson and Donna Karolchik and the initial Genome Browser Annotations made by the UCSC Genome Bioinformatics Group, University of California, Santa Cruz, CA, USA. The set of 27 948 scaffolds used in this work can be downloaded at http://rana.lbl.gov/drosophila/multipleflies.html. The *D. virilis* genome browser can also be found at the same web page. It should be noted that the annotation of the *D. virilis* genome is far from complete. For instance, the *Dpr8* gene here studied is not annotated but can be found by BLAST search on scaffold 5 (data not shown). This gene is probably the most basal gene of the *D. virilis X* chromosome. Furthermore, there are large gaps within scaffolds. For instance, the *Adh* gene previously located at cytological subdivision 49B should be present in scaffold 7, but BLAST searches (http://rana.lbl.gov/drosophila/multipleflies.html) performed using the *D. virilis Adh1 or Adh2* coding region (accession number U26846) produce hits (>99·5% identities over a region that is longer than 90 bp) only with four small scaffolds (463, 13 008, 15 971 and 26 327)

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

totalling 12 213 bp. There is, however, a 14 386 bp gap in scaffold 7 about 500 kb from *Lim3* towards the centromere. It is thus conceivable that the *Adh* region is located within this large gap.

Random sequences generated from previously mapped P1 clones (Vieira *et al*., 1997*a*), using the procedure described in Section 2(i), were used to perform a BLAST search against the *D. virilis* genome sequence, and thus make the correspondence between the *D. virilis* physical and polytene chromosome map (Gubenko & Evgen'ev, 1984). A literature search revealed further *D. virilis* genes and anonymous DNA fragments that had been mapped before, and for which a *D. virilis* or *D. melanogaster* sequence is available to perform the BLAST search against the *D. virilis* genome scaffolds.

(iii) *DNA samples, PCR amplification and DNA polymorphism analyses*

In order to test for significant differentiation between *X/4* fusion and non-fusion chromosomes at the base of the *D. americana X* chromosome, genomic DNA of single males from five *D. americana X/4* fusion strains (Gary, Indiana) and four non-fusion strains (Monroe, Louisiana, and Lone Star, Texas) were used to determine polymorphism levels at five genes located at the base of the *D. americana X* chromosome (*Dpr8*, *nej*, *CG18543*, *Cp36*, *Yp1*). The *runt* gene located elsewhere in the *X* chromosome has also been studied in single males from the above-mentioned strains, in order to compare patterns in different regions of this chromosome. In contrast to the *X*-linked genes, the *Lim3* gene located on the *4th* chromosome has been amplified from single females. The strains used in this work were originally established by Bryant McAllister (University of Iowa, Iowa, USA) and are described in detail in Vieira *et al*. (2001). For these strains, the status of the *X* chromosome as fused or unfused to the *4th* chromosome has been previously determined by cytological observation of mitotic chromosomes (Vieira *et al*., 2003). Chromosome squashes were prepared as described in Päällysaho (2001), stained in a 5% Giemsa (Merck Cat. No. 1.09204) solution for 5 min, and observed under the microscope, in order to determine whether the *X* chromosome carries the *Xc* inversion (data not shown). All *X/4* fusion chromosomes have the *Xc* inversion while all non-fusion chromosomes lack the *Xc* inversion.

The primers used and the corresponding annealing temperatures are presented in Supplementary Table 1. Primer sequences were designed, based on the obtained sequence of random fragments of the *D. virilis* P1 clones of interest or based on the whole *D. virilis* genome sequence (see Section 2(ii)). Standard amplification conditions were 35 cycles of denaturation at 94 °C for 30 s, primer annealing at the appropriate temperature for 45 s, and primer extension at 72 °C for 3 min. Direct sequencing of the PCR products was performed for every gene except *Lim3* after isolation of the amplification products from a 1·2% agarose gel using the QIAEXII Gel Extraction Kit (QIAGEN, CA, USA). The *Lim3* amplification products were cloned using the TOPO-TA Cloning Kit for Sequencing from Invitrogen (Invitrogen, Spain), since this gene is located on an autosome. At least three colonies were analysed for each individual in order to correct for possible PCR nucleotide misincorporations. Sequencing was performed with an Applied Biosytems Model 310 DNA sequencing system with the ABI PRISM BigDye cycle-sequencing kit version 1.1 (Perkin Elmer, CA, USA). The DNA sequences were deposited in GenBank (accession numbers for the *Dpr8* gene are DQ435550–DQ435558; for the *nej* gene are DQ435568–DQ435576; for the *CG18543* gene are DQ435532–DQ435540; for the *Cp36* gene are DQ435541–DQ435549; for the *Yp1* gene are DQ435586–DQ435594; for the *runt* gene are DQ435 577–DQ435585; for the *Lim3* gene are DQ435559–DQ435567). Analyses of DNA polymorphism were performed using the DnaSP (Rozas *et al*., 2003) and ProSeq version 2.43 (http://helios.bto.ed.ac.uk/evolgen/filatov/proseq.html) software.

(iv) In situ *hybridization*

The *in situ* hybridization technique was performed as described in Päällysaho (2001), using as a probe a purified 8·8 kb PCR amplification product that contains most of the *Yp1* gene and the upstream intergenic region. This region was amplified using primers 5′-GTGTGATGCTCCGTAAAG-3′ and 5′-CGCCC AATGTGATAACCA-3′ that were designed based on the *D. virilis* genome sequence (data not shown). Polytene chromosome squashes were prepared using strain G96.48 that carries both the *X/4* fusion and the *Xc* inversion (see above). We have located *Yp1* to region 15D/16A of the *D. novamexicana* polytene chromosome map of Vieira *et al*. (1997*b*). This map can be used since the *X* chromosome formula for both *D. novamexicana* and *D. americana* G96.48 strain is *Xabc* and the two species are closely related. It should be noted that our *in situ* localization differs from the one given by Wittkopp *et al*. (2003), based on a comparative approach.

3. Results

(i) *The* D. virilis *and the* D. americana X *and* 4th *chromosome*

Although a set of 27 948 annotated *D. virilis* scaffolds is available (see Section 2), the origin and orientation of each scaffold is unknown. Nevertheless, even a

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

relatively small number of genes or anonymous DNA sequences previously mapped by *in situ* hybridization to the *D. virilis* polytene *X* and *4th* chromosomes according to the map of Gubenko & Evgen'ev (1984) can be used to determine the location and orientation of the largest (and most informative) scaffolds. A literature search was thus performed to identify mapped *D. virilis* genes or anonymous DNA sequences (mostly derived from the *D. virilis* P1 clones previously mapped by Lozovskaya *et al.* (1993) and Vieira *et al.* (1997*a*)), for which a *D. virilis* or the orthologous *D. melanogaster* sequence is available to perform a BLAST search against the available *D. virilis* genome scaffolds. In this work, further *D. virilis* DNA sequences were also generated from previously mapped P1 clones (1–33; 12–59; 14–31; 14–77; 20–36; 20–66; 20–73; 71–76; it should be noted that only the P1 clones 1–33 and 14–77 are located at the base of the *D. virilis* and *D. americana* *X* chromosome). For scaffolds that are linked to the chromosomal map by more than one data point it is also possible to determine the orientation of the scaffolds (see for instance scaffold 5 on the *X* chromosome; Fig. 1). In some cases, although a given scaffold is supported by a single data point, its orientation can also be inferred because of spatial constraints imposed by the neighbouring scaffolds (see for instance scaffold 31 and the neighbouring scaffolds 67 and 37 in Fig. 1). No genes have been mapped to *D. virilis* regions 3A–3C and 4B–4C of the *X* chromosome or to the telomeric region of the *4th* chromosome but it is conceivable that these regions are contained within large scaffolds that could only fit into these chromosomal regions. Therefore, the chromosomal origin of the 39 scaffolds (numbered from 0 to 38) larger than 1 Mb was inferred by performing BLAST searches against the *D. melanogaster* genome using as a query a block of sequence (larger than 10 kb) of the scaffold of interest (data not shown). This methodology works because of the general rule (there are only a few exceptions) of chromosomal arm gene content conservation within *Drosophila* (reviewed in Hartl & Lozovskaya, 1995). Scaffold 36 can in principle fit into either the 3A–3C or the 4B–4C region of the *D. virilis* *X* chromosome while scaffold 23 must come from within the region 40A–40F (*D. virilis* chromosome *4*).

The common ancestor of *D. virilis* and *D. americana* lived about 5 Mya (Nurminsky *et al.*, 1996; Spicer & Bell, 2002; Caletka & McAllister, 2004). It is thus not surprising that most primers designed based on the *D. virilis* coding genome sequence work well on *D. americana*. It is also not surprising that the *X* and *4th* *D. virilis* gene order can be converted into the *D. americana* gene order by assuming at most three and two inversions respectively (Fig. 2; Hsu, 1952). The inversion breakpoints shown in Fig. 2 are tentative due to difficulties in accurately relating Hsu's *D. virilis* drawn map to the polytene chromosome map of Gubenko & Evgen'ev (1984). Nevertheless, it is expected that the breakpoints are at most one cytological subdivision removed. In the case of the basal *Xc* inversion breakpoint, the resolution can be obtained at the molecular level since it is shown here that the gene *Yp1* is located at the base of the *X* chromosome in *D. americana*, and, at the chromosomal scale, near *fused1*. According to Päällysaho (2001), at the chromosomal level the *cac* gene (also known as *Dcma1a* or *nbA*) is located near the *fused* gene in *D. virilis* but far away from it in *D. americana*. Therefore the basal *Xc* inversion breakpoint must be located in the region *cac*-*Yp1* (see Fig. 2). In *D. virilis*, *Yp1* has been located to chromosomal subdivision 15A (Wittkopp *et al.*, 2003). This fits with the previously estimated interval 14A–15C for one of the *Xc* inversion breakpoints (Vieira *et al.*, 1997*b*). Using the *D. virilis* Genome Browser at http://rana.lbl.gov/drosophila/multipleflies.html and the coordinates scaffold_10 : 1 450 000–1 600 000, it can be seen that the region between *cac* and *Yp1* is 86 kb long. Therefore *Yp1* must be less than 86 kb away from the basal *Xc* inversion breakpoint.

For the base of the *X* and *4th* chromosome, the genes here analysed are between genes that have now been mapped by *in situ* hybridization in *D. virilis* and *D. americana* (Fig. 2). Therefore the order of the genes studied is known in both *D. virilis* and *D. americana*. It should be noted that the results obtained by *in situ* hybridization are not always reliable. For instance, the published cytological locations for *D. virilis* *Cp36* and *run* genes (Konsolaki *et al.*, 1990; Vieira & Charlesworth, 1999) are incompatible with the *D. virilis* DNA genome sequence (data not shown). This may be due to the poor definition of some chromosomal regions, or the presence of repetitive DNA in the probes used for *in situ* hybridizations (Vieira *et al.*, 1997*a*).

(ii) *Variability levels*

Amino acid polymorphism was detected at three of the seven new data sets produced (*Dpr8*, *nej*, *CG18543*, *Cp36*, *Yp1*, *runt* and *Lim3*). Only two of the 10 amino acid variants detected were not singletons. At *Yp1* one variant was detected twice (an alanine by a serine detected in both the *X*/*4* fusion and non-fusion background), while at *CG18543* one fixed amino acid difference was found between the *D. americana* fusion and non-fusion forms (an asparagine and a serine in the *X*/*4* fusion and non-fusion form, respectively).

For each of the seven *D. americana* genes located at the base of the *X* chromosome (*Dpr8*, *nej*, *CG18543*, *Cp36* and *Yp1* plus the previously published data sets for *fused1* and *para*), the average number of pairwise

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

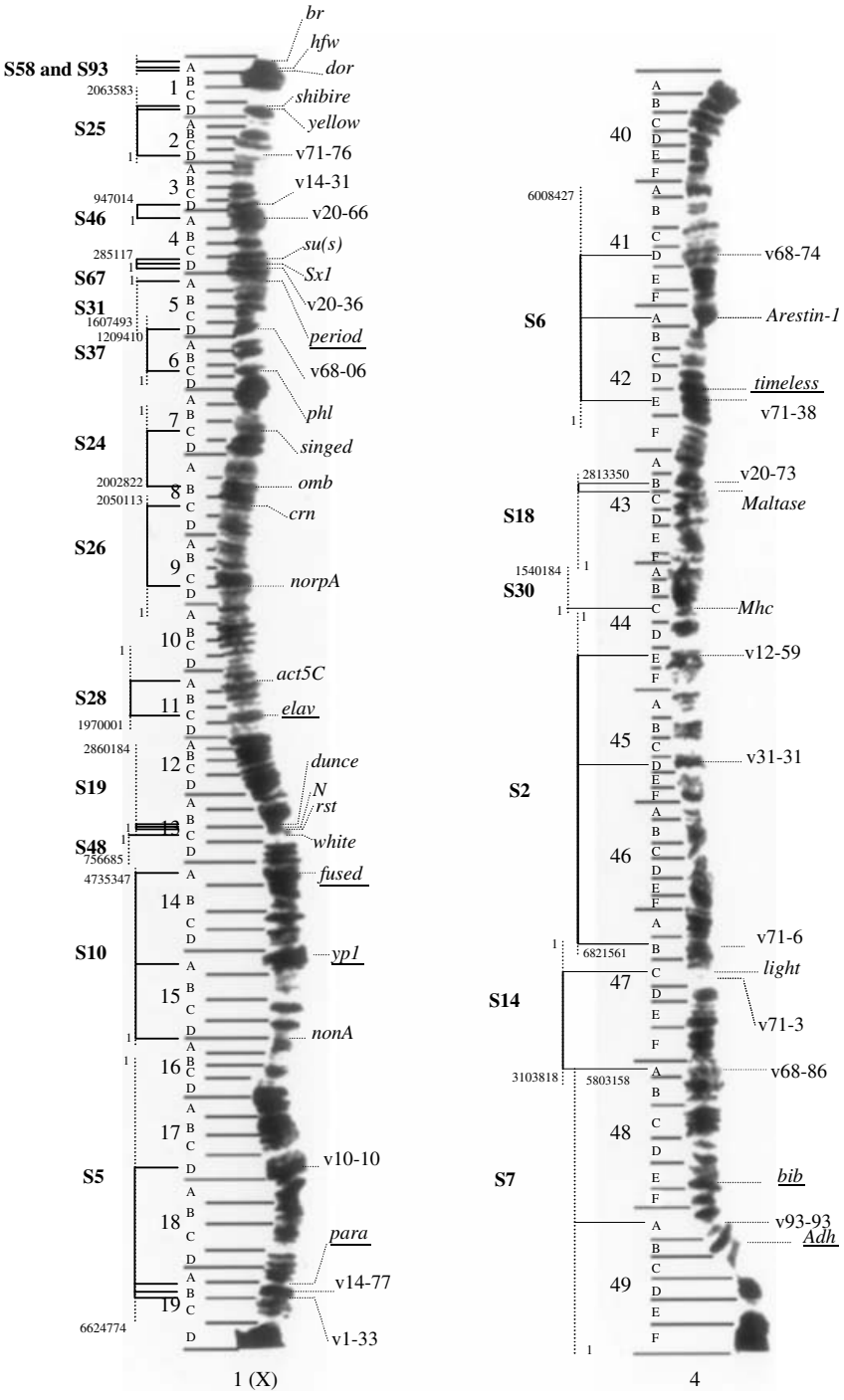

Fig. 1. Relationship between *D. virilis* chromosome *X* and *4* sequence scaffolds (http://rana.lbl.gov/drosophila/multipleflies.html#Dvir) and the polytene chromosome map of Gubenko & Evgen'ev (1984). Scaffold names are indicated by the prefix S. Numbers associated with each scaffold give information on the size and orientation of the scaffold (in bp). The unbroken line represents the area of the polytene chromosome that is known to be contained in the scaffold. The dotted line represents the area of the polytene chromosome that is estimated to be contained in a given scaffold assuming that for a given chromosome every subdivision has approximately the same DNA content. At the right side of each chromosome the names of the genes that have been used to anchor the scaffolds are shown. Genes for which there is an available *D. americana* population data set are shown underlined. *D. virilis* P1 clones are indicated by the prefix v.

silent (synonymous plus intron sites) differences per silent site (Nei, 1987), as well as Watterson's estimator of this quantity (θ; Watterson, 1975) were computed (Table 1). There is evidence for differentiation between the *X*/*4* fusion and non-fusion chromosomal forms of *D. americana* at five of the seven genes analysed (see below), thus the two chromosomal forms were analysed separately. On average the *X*/*4*

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

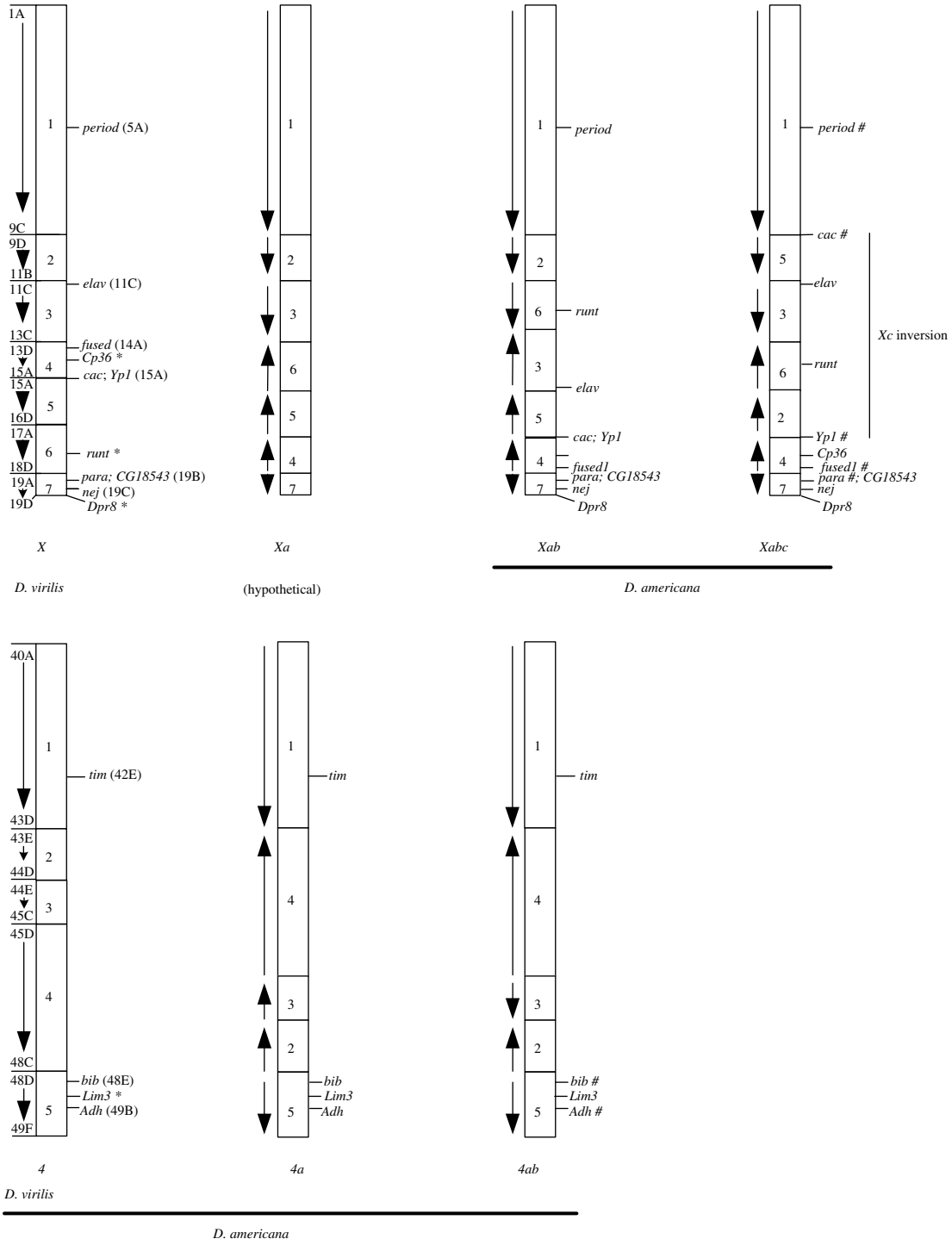

Fig. 2. Common *D. americana* chromosome *X* and *4* types. *D. virilis* cytological positions according to the polytene chromosome map of Gubenko & Evgen'ev (1984) are indicated in order to allow the correspondence between the *D. virilis* genome sequence information shown in Fig. 1 and the regions of the common chromosomal *D. americana* types. Inversion breakpoints are tentative due to difficulties in accurately relating Hsu's (1952) drawn map of *D. virilis* to the polytene chromosome map of Gubenko & Evgen'ev (1984). Conserved blocks between different *D. americana* chromosomal types are indicated by numbers. Arrowheads show the relative orientation of the conserved blocks. #Known (determined by *in situ* hybridization to polytene chromosomes; McAllister, 2003; Päällysaho, 2001; Vieira *et al*., 2001, 2003; this work) gene positions. All other gene locations are inferred from the cytological work of Hsu (1952) and the comparative map of *D. virilis*/*D. novamexicana* (Vieira *et al*., 1997*b*). *Inferred from the *D. virilis* genome sequence.

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

Table 1. *Summary of* D. americana *silent site sequence variation*

Gene	Chromosome		Non-fusion form	*X/4* fusion form
Dpr8	*X*	S	1	0
		π	0·0144	0
		θ	0·0157	0
		L	34·83	34·70
nej	*X*	*S*	5	2
		π	0·0044	0·0017
		θ	0·0048	0·0017
		L	574·33	574·33
CG18543	*X*	*S*	18	9
		π	0·0140	0·0052
		θ	0·0150	0·0062
		L	653·00	696·07
para[a]	*X*	*S*	20	24
		π	0·0031	0·0033
		θ	0·0042	0·0040
		L	1700·83	1707·83
fused1[a]	*X*	*S*	37	4
		π	0·0152	0·0021
		θ	0·0178	0·0018
		L	782·11	789·37
Cp36	*X*	*S*	2	1
		π	0·0058	0·0020
		θ	0·0054	0·0024
		L	201·33	201·33
Yp1	*X*	*S*	13	10
		π	0·0219	0·0166
		θ	0·0228	0·0154
		L	311·75	312·53
runt	*X*	*S*	2	5
		π	0·0041	0·0091
		θ	0·0038	0·0084
		L	284·17	285·17
Lim3	*4*	*S*	48	51
		π	0·0333	0·0343
		θ	0·0345	0·0363
		L	760·00	767·00

[a] Vieira *et al*. (2001, 2003). *X*-linked genes are presented starting with the most centromeric and finishing with the most telomeric. *S* is the number of segregating sites; π (Nei, 1987) is the average number of pairwise nucleotide differences per base pair; θ is Watterson's estimator based on the number of segregating sites (Watterson, 1975). *L* is the number of silent sites analysed. Sample sizes for the new data sets but for *Lim3* are five and four for *X/4* fusion and non-fusion individuals, respectively. Strain G96.47 was not included in the *Lim3 X/4* fusion data set because at the *bib* gene it shows a nucleotide variant that has been found to be associated with compound inversion *4ab* (see Section 3(iii)). Thus, sample sizes for *Lim3* are four for both *X/4* fusion and non-fusion individuals.

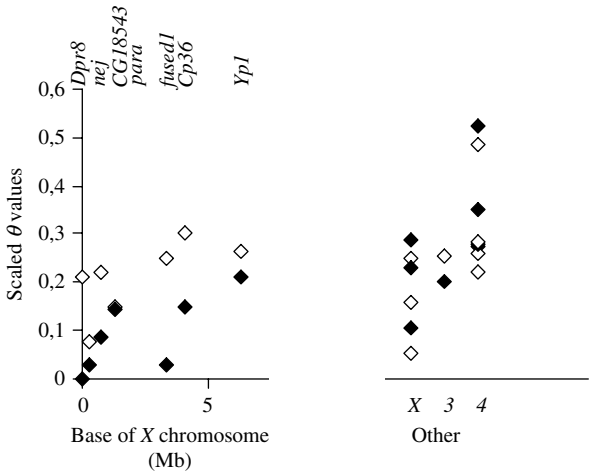

Fig. 3. Scaled *D. americana* silent site variability values (silent site θ values divided by silent site divergence between *D. americana* and *D. virilis*) for genes at the base of the *X* chromosome and elsewhere in the genome. Distance is from the *Dpr8* gene that is closest to the *X* chromosome centromere. The *Yp1* gene is close to the basal *Xc* inversion breakpoint (see Section 3(i) for details). Data source is shown in Table 2. θ values for *X*-linked genes have been multiplied by 4/3 in order to account for the different population sizes of *X*-linked and autosomal genes. We are thus assuming the same effective size for females and males, which requires in particular the same variance in offspring number (Charlesworth, 2001; Laporte & Charlesworth, 2002). Black diamonds, *D. americana X/4* fusion form; white diamonds, *D. americana* non-fusion form.

fusion form shows about half (49%) of the variability of the non-fusion form. The θ estimate was also divided by the estimated silent site divergence between *D. americana* and *D. virilis*, in order to account for different levels of purifying selection acting at silent sites (Fig. 3). The results show that, at the base of the *X* chromosome of both *D. americana* chromosomal forms, there is an increase in the scaled value of θ from the *X* centromere inward (Spearman's non-parametric coefficient of correlation is $r=0{\cdot}847$ ($P<0{\cdot}05$) and $r=0{\cdot}886$ ($P<0{\cdot}05$) for the *D. americana X/4* fusion and non-fusion form respectively). It should be noted that there is no significant correlation between gene location and silent site divergence between *D. virilis* and *D. americana* (data not shown). For comparison, the scaled value of θ is also shown for genes located elsewhere in the genome (Fig. 3).

The region of the *Lim3* gene here analysed (first intron) is highly variable and thus highly informative. The θ value is close to that observed for other genes not located at the base of the *X* chromosome such as *elav*, *bib*, *Adh* or *transformer* genes (Table 1).

Several tests of neutrality (Tajima's *D* (Tajima, 1989), Fu and Li's *D* and *F* tests using a *D. virilis* sequence as an outgroup (Fu & Li, 1993), Kelly's *Zns* (Kelly, 1997) and Wall's *B* and *Q* tests (Wall, 1999)) were performed for the new data sets (the *X/4* fusion and the non-fusion chromosomal data sets were treated separately). None of the statistical tests performed showed significant deviations from neutrality under the conservative assumption of no recombination (data not shown). The power of the tests may,

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

Table 2. *Genetic differentiation and divergence between* D. americana *individuals with* (*from the northeastern population*) *or without* (*from the south*) *the X/4 chromosomal fusion*

Gene	Chromosome	Nucleotide differences			F_{st}[a]	K_{st}[b]	S_{nn}[c]	*Ks*	Standardized *Ks*[d]
		Shared	Exclusive	Fixed					
Dpr8	*X* (*C-Xc*)	0	2	0	0	n.s.	n.s.	0·0072	0·072
nej	*X* (*C-Xc*)	0	7	0	0·146	n.s.	n.s.	0·0036	0·045
CG18543	*X* (*C-Xc*)	4	21	3	0·415	$P<0·01$	$P<0·0001$	0·0150	0·166
para[e]	*X* (*C-Xc*)	4	34	0	0·252	$P<0·0001$	$P<0·0001$	0·0044	0·119
fused1[f]	*X* (*C-Xc*)	1	51	7	0·573	$P<0·0001$	$P<0·0001$	0·0214	0·229
Cp36	*X* (*C-Xc*)	0	3	0	0·175	n.s.	$P<0·01$	0·0047	0·196
Yp1	*X* (*C-Xc*)	2	23	4	0·464	$P<0·01$	$P<0·0001$	0·0400	0·351
elav[e]	*X*	29	50	0	0·012	n.s.	n.s.	0·0216	0·118
runt	*X*	2	3	0	0·124	n.s.	n.s.	0·0076	0·075
period[g]	*X*	40	57	0	0·012	n.s.	n.s.	0·0297	0·196
Lim3	*4*	20	58	0	n.a.	n.s.	n.s.	0·0331	0·257
bib[h,i]	*4*	23	37	0	0·053	n.s.	n.s.	0·0336	0·270
Timeless[i]	*4*	7	24	0	0·038	n.s.	n.s.	0·0467	0·435
Adh[j]	*4*	17	22	0	0·013	n.s.	n.s.	0·0225	0·252
transformer[k]	*3*	31	30	0	0·024	n.s.	n.s.	0·0203	0·202

C-Xc, genes located between the *X* chromosome centromere and the basal breakpoint of inversion *Xc*.
n.s., not significant ($P>0·05$).
n.a., not available, i.e. the estimated F_{st} value is negative.
[a] Hudson *et al*. (1992*a*).
[b] Hudson *et al*. (1992*b*).
[c] Hudson (2000).
[d] Silent site divergence per site between *D. americana* with and without the *X/4* fusion divided by the silent site divergence per site between *D. americana* and *D. virilis*.
[e] Vieira *et al*. (2003).
[f] Vieira *et al*. (2001).
[g] Hilton & Hey (1996).
[h] Four Gary sequences from chromosomes that likely have compound inversion *4ab* were excluded (see text for details).
[i] McAllister (2003).
[j] McAllister & Charlesworth (1999).
[k] McAllister & McVean (2000).

however, be small with sample sizes of five and four *X/4* fusion and non-fusion chromosomes respectively.

(iii) *Differentiation between* D. americana X/4 *fusion and non-fusion chromosomal forms*

Significant differentiation has been observed only between *D. americana X/4* fusion (only samples from the northeastern branch of the distribution were used) and non-fusion chromosomal forms for genes at the base of the *X* chromosome (five of the seven gene regions analysed show evidence for significant differentiation; Table 2) and for the gene *bib* on the *4th* chromosome (McAllister, 2003). There is no evidence for differentiation at *bib*, however, when the four sequences (22% of the sequences) which show a variant that has been found to be associated with compound inversion *4ab* (McAllister, 2003) are removed (inversion *4ab* is known to be present at 20% frequency in the northeastern branch; Hsu, 1952). The differentiation detected at *bib* is thus probably the result of the presence of the *4ab* inversion in McAllister's *X/4* fusion data set. A third gene at the base of *4th* chromosome (the *Lim3* gene that is located in between *bib* and *Adh*; Fig. 2) has been studied here. No evidence for differentiation has been found (Table 2), confirming that there is little or no suppression of recombination between fusion and non-fusion chromosomes at the base of the *4th* chromosome.

A total of 14 apparent fixed differences (including two amino acid differences, one at *fused1* (Vieira *et al*. 2001) and another at *CG18543*) have been found in the seven sampled gene regions at the base of the *X* chromosome. These findings are compatible with reduced recombination levels between *D. americana X/4* fusion and non-fusion chromosomal forms. It should be noted, however, that 11 polymorphisms are shared between the two chromosomal forms. Within chromosomal forms, the frequency of sites that have been hit by mutation more than once, i.e. the frequency of sites with three or more variants, is only 0·6% of the polymorphic sites. Therefore it seems unlikely that these 11 shared polymorphisms are homoplasies. The data thus imply a low level of

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

Table 3. *Age estimates for the X/4 fusion and Xc inversion. Lower and upper 95% bounds from the Poisson distribution are also shown*

Chromosomal region	No. of apparent fixed differences		Estimated age (thousand years)
	Observed	Expected	
(*X/4* fusion, *Dpr8*, *nej*, *CG18543*, *para*)	2	3·26 (1–6)	0 (0–29)
(*fused1*, *Cp36*, *Yp1, Xc* inversion)	10	3·00 (0–6)	308 (176–441)

recombination (either crossing over or gene conversion) between the two chromosomal forms at the base of the *X* chromosome. Using the approach of Betran *et al*. (1997), a 23 bp putative gene conversion tract is detected for one *X/4* fusion chromosome at gene *CG18543*. At this gene, three of the four polymorphic sites shared between the two chromosomal forms are located in this 23 bp region. For the other genes located at the base of the *X* chromosomes, no putative gene conversion tracts are identified with this approach.

(iv) *Age estimates for the* X/4 *fusion and* Xc *inversion*

In *D. americana*, the chromosomal region between the *X* chromosome heterochromatin block and the basal *Xc* inversion breakpoint is probably larger than 6 Mb. This estimate is based on the observation that in *D. virilis* the region between *Dpr8* and *para* is 1·4 Mb long and that the region between *fused* and *Yp1* is 3 Mb long (see Section 2(ii)). These chromosome blocks are conserved in *D. americana*, so they are expected to have about the same size in the two species (see Fig. 2). In *D. americana*, however, the distance between *para* and *fused1* is difficult to estimate. Nevertheless, considering the size of the other blocks and the *in situ* localizations of the *para* and *fused1* genes, it is probably about 2 Mb.

Age estimates for the *Xc* inversion and *X/4* fusion can in principle be obtained using the observed number of silent site apparent fixed differences between individuals with the *Xc* inversion *X/4* fusion and individuals without the *Xc* inversion and fusion, under the following evolutionary model. We assume that the *Xc* inversion is older than the *X/4* fusion, since this chromosomal rearrangement is present in both *D. americana* and *D. novamexicana*, while the *X/4* fusion is found only in the former species (Hsu, 1952). It is assumed that the *Xc* inversion suppresses recombination between *Xc* inverted and non-inverted chromosomes at least in the *fused1*-*Xc* basal inversion breakpoint region (about 3 Mb in size) but not as far as the *para* region (about 4·6 Mb away from the basal *Xc* inversion breakpoint). No significant differentiation was found between fusion and non-fusion fourth chromosomes at the *Adh* locus (McAllister & Charlesworth, 1999), which is about 1 Mb from the centromeric heterochromatin block on chromosome four (Vieira *et al*., 2001). It is thus assumed that the *X/4* fusion suppresses recombination between *Xc* inverted *X/4* fusion and *Xc* inverted non-fusion chromosomes only in the *Dpr8–para* region (about 1·4 Mb long).

Since each chromosomal rearrangement is probably a unique event, and recombination is assumed to be regionally suppressed between standard arrangement chromosomes and rearranged chromosomes, a fraction of the observed apparent fixed differences between *Xc*-inverted *X/4* fusion and non-inverted non-fusion chromosomes could be explained by the capture of low-frequency variants by the rearranged chromosome when it first arose. Therefore, the expected number of variants that are captured by the *Xc* inversion or *X/4* fusion and that are expected to be absent in a sample of gametes without the chromosomal rearrangement being considered was calculated according to the formula given in Vieira *et al*. (2001), assuming that the ancestral population had variability levels close to those observed in the non-inverted non-*X/4* fusion chromosomes. It should be noted that there are large variances associated with the estimated θ values used in these calculations (Tajima, 1993), especially for those genes located in regions with low levels of recombination as seems to be the case in *D. americana* for the *Dpr8–para* chromosomal region (see Section 4). Therefore it is desirable to obtain an estimate based on several gene regions rather than just one. For a given chromosomal region, the expected number of apparent fixed differences for each gene is then added. The value, together with the lower and upper 95% bounds from a Poisson distribution, is shown in Table 3.

For each chromosomal region the difference between the observed and expected number of apparent fixed differences is then calculated and divided by the total number of silent sites considered. For each gene the rate of silent site divergence (with Jukes–Cantor correction) between *D. americana* non-fusion non-inverted chromosomes and *D. virilis* is then calculated. Since the different genes seem to have different rates of evolution (data not shown), for each region an average of the different rates weighted by the number of silent sites analysed for each gene was calculated. The estimated rate of divergence per silent

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

site between *Xc* inverted *X/4* fusion chromosomes and non-inverted non-fusion chromosomes is then divided by this number. The age estimates shown in Table 3 are obtained assuming that *D. americana* and *D. virilis* shared the last common ancestor about 5 Mya (Nurminsky *et al.*, 1996; Spicer & Bell, 2002; Caletka & McAllister, 2004). These are rough estimates of the true age of the chromosomal rearrangements. Rare recombination between chromosomes with different rearrangements could also have an impact on the age estimates. Nevertheless, both chromosomal rearrangements seem to be young, and the *Xc* inversion seems to be older than the *X/4* fusion, thus validating one of the original assumptions of the model.

4. Discussion

The annotated *D. virilis* euchromatic genome sequence is now available. The *D. virilis* scaffolds located here on the polytene chromosome map of this species contain most polytene chromosome subdivisions of the *X* and *4th* chromosomes. *D. virilis* and *D. americana* are closely related species (the common ancestor of the two species lived about 5 Mya; Nurminsky *et al.*, 1996; Spicer & Bell, 2002; Caletka & McAllister, 2004). The *D. virilis X* and *4th* chromosome gene order can be easily converted into the *D. americana* gene order by assuming at most (since some inversions are polymorphic) three and two chromosomal inversions, respectively (Fig. 2; Hsu, 1952). The exact location of the chromosomal breakpoints requires, nevertheless, additional molecular work. The close proximity of the two species also means that most primers designed based on the *D. virilis* coding regions work well in *D. americana*. For the purpose of performing studies on particular regions of the *X* and *4th* chromosomes in species of the americana complex, the *D. virilis* genome sequence together with the correspondence presented here is thus very useful.

The analysis of seven genes at the base of the *D. americana X* chromosome revealed that for both *X/4* fusion and non-fusion chromosomes there is a significant positive correlation between gene location (starting at the centromere) and levels of silent site polymorphism (Fig. 3). Since this pattern is observed for both chromosomal forms, it is not a consequence of the recent history of the *X/4* fusion. It should also be noted that there is no significant correlation between gene location and silent site divergence between *D. virilis* and *D. americana* (data not shown). The observed pattern is compatible with background selection and hitchhiking in regions experiencing low recombination levels (reviewed in Charlesworth & Charlesworth, 1998). It is thus possible that the base of the *D. americana X* chromosome is a region with suppressed levels of recombination, as is the case for *D. melanogaster* and many other species (Charlesworth & Charlesworth, 1998). It is unclear how far the region with suppressed recombination levels extends. Nevertheless, given the observed scaled levels of variability (Fig. 3), the *fused1–Yp1* region may already experience recombination levels close to those experienced by genes elsewhere in the genome. In order to confirm these inferences recombination rates should ideally be directly estimated for this region. There are, however, very few visible markers available for *D. americana*.

It seems reasonable to assume that the *Xc* inversion is older than the *X/4* fusion since the former is present in both *D. americana* and *D. novamexicana* while the *X/4* fusion is found only in the former species (Hsu, 1952). The estimated relative ages for these chromosomal arrangements are in agreement with this assumption (Table 3). The *D. americana* and *D. novamexicana* species pair seem to have been diverging, however, for about 380 000 years (Caletka & McAllister, 2004), although a large variance is certainly associated with this estimate. Nevertheless, Spicer & Bell (2002) argue that the divergence time between this species pair could be as much as 1 million years. It should be noted that there is no convincing evidence in the published literature for naturally occurring hybridization between *D. americana* and *D. novamexicana* which might explain the apparent discrepancy.

It is thus proposed that about 308 000 years ago the *Xc* inversion arose on a standard *D. americana X* chromosome (an *Xab* chromosome; see Fig. 2). This event would have suppressed recombination in a large region around the basal *Xc* inversion breakpoint between *Xc*-bearing chromosomes and standard chromosomes. Inversions alone are known to suppress recombination far away from inversion breakpoints in females heterozygous for inversions, as happens with the *D. americana* compound inversion *4ab* (McAllister, 2003). It must be assumed, however, that recombination is not suppressed at the very base of the *X* chromosome in order to explain why different age estimates are obtained when two different regions at the base of the *X* chromosome are considered.

The *Xc* inversion could have increased in frequency due to genetic drift, selection or both. It is easier to explain the occurrence of the *X/4* fusion on an *Xc* inversion background if it assumed that the *Xc* inversion at that time was already common in some part of the *D. americana* geographic range. The *X/4* fusion is proposed to have occurred very recently, no more than 29 000 years ago. This estimate is lower than previously reported (Vieira *et al.*, 2001, 2003). The previous estimates are, however, based on single gene fragments that are several megabases away from the

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

X chromosome centromeric heterochromatin block. Furthermore, the possible influence of the *Xc* inversion was not accounted for in the previous analyses.

It is argued that the *X/4* fusion would suppress recombination only at the very base of the *X* chromosome between *Xc* inverted chromosomes and *Xc* inverted *X/4* fusion chromosomes, since no differentiation has been detected at the base of the *4th* chromosome between the two chromosomal forms (McAllister & Charlesworth, 1999; McAllister, 2003; see also Table 2). It should be noted that the genes analysed at the base of the *4th* chromosome are, however, further away from the centromere than the genes analysed at the base of the *X* chromosome. Although the *X/4* fusion is assumed to be a unique event, the *Xc* inverted *X/4* fusion chromosomes would thus recover variability by recombining with *Xc* inverted non-fusion chromosomes but at the very base of the *X* chromosome. In the latter region, however, silent site variability levels are already low due to background selection and hitchhiking in this region of the chromosome inferred to have low recombination levels (see above).

The presence of both the *Xc* inversion and *X/4* fusion would strongly suppress recombination between *Xc* inverted *X/4* fusion chromosomes and non-inverted non-fusion chromosomes in the region in between the *X* chromosome block of heterochromatin and the basal *Xc* inversion breakpoint. There is, nevertheless, evidence for 11 shared polymorphisms in this chromosomal region, which suggests some recombination between the two chromosomal forms. In the central part of the geographic range females heterozygous for the *Xc* inversion are probably common. Thus it is probable that some recombination occurs between the two chromosomal forms even if recombination is highly suppressed. Since there are high levels of gene flow between populations (Vieira *et al*., 2003 and references therein), evidence for recombination is to be found throughout the geographic range.

That at the base of the *X* chromosome, *X/4* fusion chromosomes show significantly less silent site variability than non-fusion chromosomes ($P < 0{\cdot}05$; non-parametric sign test) can be explained by the young age of the *Xc* inversion and *X/4* fusion and a putative historically low effective population size (N_e) for *Xc* inverted chromosomes and for *Xc* inverted *X/4* fusion chromosomes.

The *Xc* inverted *X/4* fusion chromosomes would have increased in frequency due to selection since the observed *D. americana X/4* fusion *Xc* inversion north–south frequency cline seems to be maintained by selection (Vieira *et al*., 2001, 2003). No similar molecular clines have been detected for any of the genes not located at the base of the *X* chromosome studied so far (Vieira *et al*., 2001, 2003). This chromosomal rearrangement is a young evolutionary event and it is possible that some molecular signature of the selective sweep that brought the *X/4* fusion to a high frequency is still detectable.

Since there is no evidence, so far, for suppression of recombination between fusion and non-fusion chromosomes at the base of the *4th* chromosome, the target of selection is very likely to be located in the *D. americana X* centromere–inversion *Xc* basal breakpoint region. Nevertheless, if, as argued, the *X/4* fusion suppresses recombination between *Xc* inverted *X/4* fusion and *Xc* inverted non-fusion chromosomes only at the very base of the *X* chromosome, then it is most likely that the target of selection is located at the very base of the *X* chromosome. Otherwise the advantageous mutation could have escaped to an *Xc* inverted non-fusion chromosome background by recombination during the selective sweep phase.

The *D. americana X* centromere–inversion *Xc* basal breakpoint region comprises about 450 genes. Further studies are needed to determine which of these 450 genes are responsible for the establishment and maintenance of the *X/4* chromosomal fusion. We may then start understanding what phenotypic characteristic natural selection is acting upon.

Supplementary Table 1. *Primer list and corresponding annealing temperatures*

Gene	Primer set	Annealing temperature (°C)
Dpr8	GCTCTTTTACGATACCTTC/GTCATTATTGTCTTGTTTT	51
nej	CGTTTTGCTGCCATTTCA/ACCATCCACTGTTTGAGG	48
	TGGGACAATAGGTATGCC/CCCTACCCCCCACCAAAT	52
	CGAACAGTGGATGGTCAGA/GCGACAGGTTGCTTCAGTG	49
CG18543	CGTTGGCGTGCGACATTC/CGAGTTGTTTTGTTTGAC	48
	CGGCAGATTTCCAGTAGG/TGTAAGGGCGTCCAGTAT	48
Cp36	CACATCACGCATCCACACA/GGGGCGACACCACATTGAT	52
Yp1	GCCCCACCACCCGTATGC/TCCAGACCAGTGACACGACG	57
	ACTTTGGACAGGATGAGG/TTGCTGCTGCTGTTGCTGC	52
runt	TCACCATTGCCACCTATCC/CACCCAGACCCACAGACGA	48
Lim3	TTGAATCTGTTTTGTCGTC/CTTCCGTTTTCTCATCTGC	47

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press

Brian Charlesworth and two anonymous reviewers made useful suggestions that improved this manuscript. C.P.V. benefited from the Post-Doctoral grant SFRH/BPD/5592/2001 from Fundação para a Ciência e Tecnologia during part of this work. This work has been partially funded by Fundação para a Ciência e Tecnologia (research project POCTI/BSE/37421/2001 and POCI/BIA-BCM/59399/2004), FEDER, The University of Porto and the Ilídio Pinho foundation. Sequence data from this article have been deposited with the GenBank data library under accession numbers DQ435532–DQ435594.

References

Andolfatto, P., Depaulis, F. & Navarro, A. (2001). Inversion polymorphisms and nucleotide variability in Drosophila. *Genetical Research* **77**, 1–8.

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

Caletka, B. C. & McAllister, B. F. (2004). A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. *Molecular Phylogenetics and Evolution* **33**, 664–670.

Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genetical Research* **77**, 153–166.

Charlesworth, D. & Charlesworth, B. (1998). Sequence variation: looking for effects of genetic linkage. *Current Biology* **8**, R658–R661.

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

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

Hartl, D. L. & Lozovskaya, E. R. (1995). *The Drosophila Genome Map: A Practical Guide*. Austin, Texas: R. G. Landes.

Hilton, H. & Hey, J. (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.

Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014.

Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992*a*). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**, 138–151.

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

Hsu, T. C. (1952). Chromosomal variation and evolution in the virilis group of *Drosophila*. *University of Texas Publications* **5204**, 35–72.

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

Konsolaki, M., Komitopoulou, K., Tolias, P. P., King, D. L., Swimmer, C. & Kafatos, F. C. (1990). The chorion genes of the medfly, *Ceratitis capitata*. I. Structural and regulatory conservation of the *s36* gene relative to two *Drosophila* species. *Nucleic Acids Research* **18**, 1731–1737.

Laporte, V. & Charlesworth, B. (2002). Effective population size and population subdivision in demographically structured populations. *Genetics* **162**, 501–519.

Lozovskaya, E. R., Petrov, D. A. & Hartl, D. L. (1993). A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning. *Chromosoma* **102**, 253–266.

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. & Charlesworth, B. (1999). Reduced sequence variability on the neo-*Y* chromosome of *Drosophila americana americana*. *Genetics* **153**, 221–233.

McAllister, B. F. & McVean, G. A. (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. In *New Systematics* (ed. J. Huxley), pp. 185–268. Oxford: Clarendon Press.

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

Nurminsky, D. I., Moriyama, E. N., Lozovskaya, E. R. & Hartl, D. L. (1996). Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *alcohol dehydrogenase* gene. *Molecular Biology and Evolution* **13**, 132–149.

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.

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

Spicer, G. S. & Bell, C. D. (2002). Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial *12s* and *16s ribosomal RNA* genes. *Annals of the Entomological Society of America* **95**, 156–161.

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

Vieira, J. & Charlesworth, B. (1999). *X* chromosome DNA variation in *Drosophila virilis*. *Proceeding of the Royal Society of London, Series B* **266**, 1905–1912.

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

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

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

Vieira, C. P., Coelho, P. A. & Vieira, J. (2003). 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. *Genetics* **164**, 1459–1469.

Wall, J. D. (1999). Recombination and the power of statistical tests of neutrality. *Genetical Research* **74**, 65–79.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–275.

Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. (2003). *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences of the USA* **100**, 1808–1813.

https://doi.org/10.1017/S0016672306008147 Published online by Cambridge University Press