

# Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species

G C S Kuhn, C H Teo, T Schwarzacher and J S Heslop-Harrison

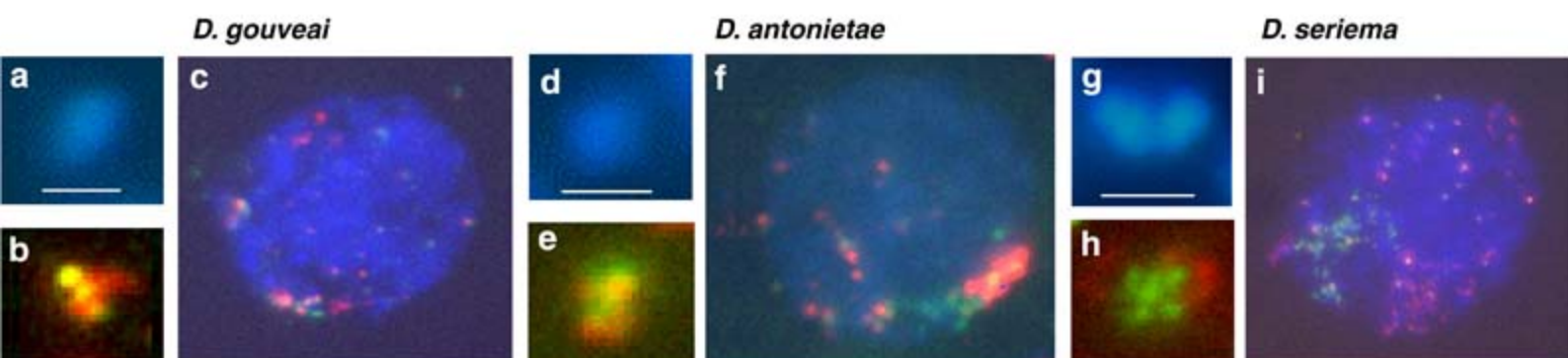


Figure 1. Double-target in situ hybridization with / (pBuM) (red) and DBC-150 (green) probes to the microchromosomes (left half of each image) or interphase nuclei (right half of each image) of *D. gouveai*, *D. antonietae* and *D. seriema*. Yellow signals show regions where the two satellite sequences are interspersed. In a, d and g, the microchromosomes are shown counterstained with DAPI (4',6'-diamidino-2-phenylindole, blue) and b, e and h show only the hybridization signals generated with the two satellite probes. Pictures of representative interphase nuclei (c, f and i) show overlays of the pBuM (red), DBC-150 (green) and DAPI DNA stain (blue). Scale bar: 1  $\mu$ m except 2  $\mu$ m in g. The color reproduction of this figure is available on the html full text version of the manuscript.

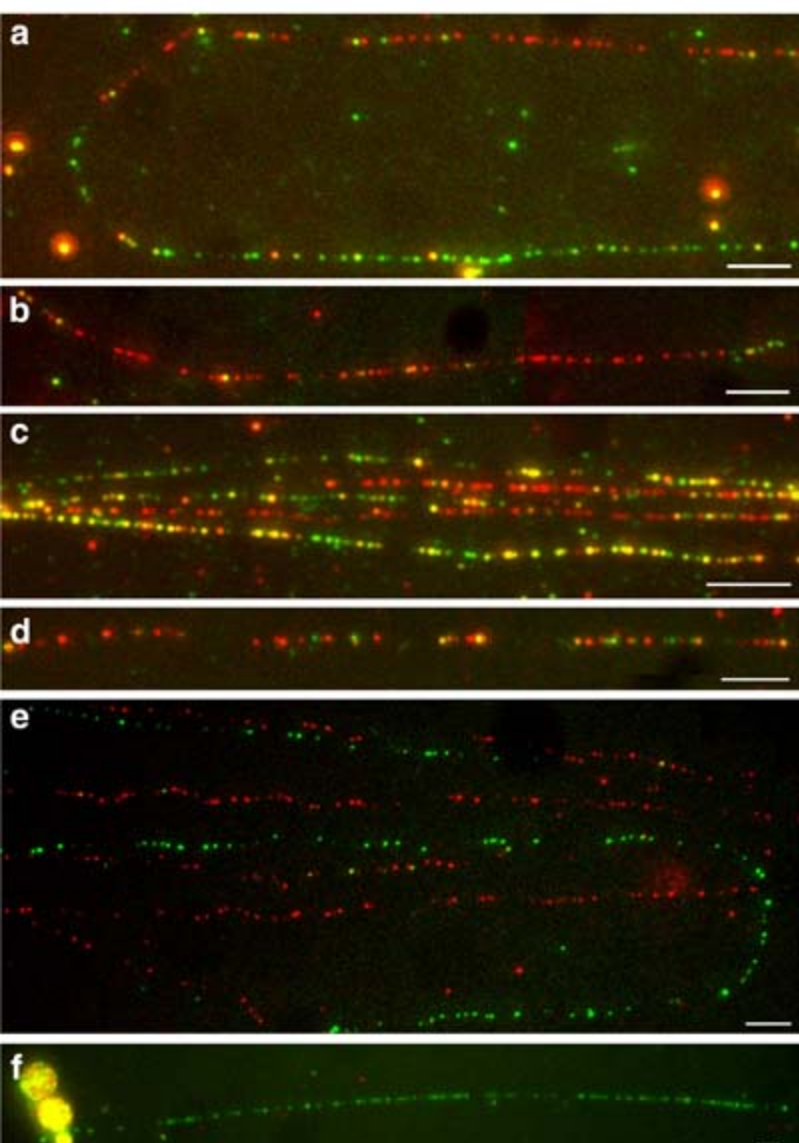


Figure 2. Extended DNA fibers hybridized with / (pBuM) (red) and DBC-150 (green) probes showing the organization of pBuM and DBC-150 satDNAs in representative fibers from *D. gouveai* (a–c), *D. antonietae* (d) and *D. seriema* (e and f). Interspersion and junctions are visualized by adjacent red and green hybridization sites, yellow where signals overlap. Scale bar represents 10 kb of stretched DNA. The color reproduction of this figure is available on the html full text version of the manuscript.

## ORIGINAL ARTICLE

Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* speciesGCS Kuhn<sup>1,2</sup>, CH Teo<sup>2</sup>, T Schwarzacher<sup>2</sup> and JS Heslop-Harrison<sup>2</sup><sup>1</sup>Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil and <sup>2</sup>Department of Biology, University of Leicester, Leicester, UK

Satellite DNA (satDNA) is a major component of genomes but relatively little is known about the fine-scale organization of unrelated satDNAs residing at the same chromosome location, and the sequence structure and dynamics of satDNA junctions. We studied the organization and sequence junctions of two nonhomologous satDNAs, pBuM and DBC-150, in three species from the neotropical *Drosophila buzzatii* cluster (*repleta* group). *In situ* hybridization to microchromosomes, interphase nuclei and extended DNA fibers showed frequent interspersions of the two satellites in *D. gouveai*, *D. antonietae* and, to a lesser extent, *D. seriema*. We isolated by PCR six pBuM × DBC-150 junctions: four are exclusive to *D. gouveai* and two are exclusive to *D. antonietae*. The six junction breakpoints occur at different positions within monomers, suggesting independent origin. Four junctions showed abrupt transitions between the

two satellites, whereas two junctions showed a distinct 10 bp tandem duplication before the junction. Unlike pBuM, DBC-150 junction repeats are more variable than randomly cloned monomers and showed diagnostic features in common to a 3-monomer higher-order repeat seen in the sister species *D. seriema*. The high levels of interspersions between pBuM and DBC-150 repeats suggest extensive rearrangements between the two satellites, maybe favored by specific features of the microchromosomes. Our interpretation is that the junctions evolved by multiples events of illegitimate recombination between nonhomologous satDNA repeats, with subsequent rounds of unequal crossing-over expanding the copy number of some of the junctions.

*Heredity* (2009) **102**, 453–464; doi:10.1038/hdy.2009.9; published online 4 March 2009

**Keywords:** repetitive DNA; higher-order repeats; unequal crossing-over; chromosomal evolution; microchromosomes; *buzzatii* cluster

## Introduction

Satellite DNAs (satDNAs) are non-protein-coding DNA elements consisting typically of a sequence motif (the monomer, between 100 and 500 bp long) that is highly repeated in arrays of 10<sup>4</sup>–10<sup>6</sup> tandemly organized, slightly varying, monomers. Arrays of a single family of satDNA may be located at one or more genomic sites, in the heterochromatic regions of chromosomes, near the centromere or in subtelomeric or intercalary regions; each species includes multiple, unrelated families of satDNA monomers (Ugarkovic and Plohl, 2002; Heslop-Harrison *et al.*, 2003). SatDNAs often represent more than 30% of total genomic DNA (Pons *et al.*, 2004). In *Drosophila*, changes in satDNA copy number during evolution positively correlate with genome sizes, which in turn might affect organismal traits of selective value,

such as developmental rate, body size and/or sperm size (Gregory and Johnston, 2008).

Apart from monomeric organization, satDNAs can show periodicity at the level of higher-order repeat (HOR) units, made up of multimers with a number of diverged monomers. A two-monomer (2-mer) HOR has been reported for scallop PmPst9 satDNA (Biscotti *et al.*, 2007), the plant rye has a highly abundant subtelomeric satellite sequence that is organized as a 3-mer HOR, where monomers are defined by *Hae*III restriction sites and 3-mer HOR by *Taq*I sites (Vershinin *et al.*, 1995), whereas centromeric repeats consisting of six monomers have been reported in rice (Lee *et al.*, 2006). In humans, the  $\alpha$ -satellite DNA comprises 171 bp monomers organized in a tandem, head-to-tail orientation, that can be arranged as several chromosome-specific HORs, such as the 6-mer HOR of chromosome 7 (Waye *et al.*, 1987). In insects, higher-order satDNA repeats have been reported in some beetle species (Pons *et al.*, 2002; Palomeque and Lorite, 2008) although there are no reports in *Drosophila*.

Satellite DNAs often exhibit a higher degree of sequence similarity among repeats from the same species compared with lower similarity between orthologous repeats, even in closely related species (Bachmann and

Correspondence: Dr GCS Kuhn, Departamento de Genética e Evolução, Universidade Federal de São Carlos, Via Washington Luís, Km 235, 13565-905, São Carlos, São Paulo, Brazil.

E-mail: gcskuhn@ufscar.br

Received 24 July 2008; revised 22 November 2008; accepted 23 December 2008; published online 4 March 2009

Sperlich, 1993; Mantovani *et al.*, 1997; Fernández *et al.*, 2001; Heslop-Harrison *et al.*, 2003; Rudd *et al.*, 2006). Models to explain this phenomenon, known as concerted evolution, assume that once mutations arise in individual repeats, repeated cycles of unequal crossing-over and gene conversion can duplicate and spread them simultaneously throughout a sequence family and a population of reproductive individuals (Dover, 1982; Kuhn *et al.*, 2008). Reproductive barriers would subsequently allow homogenization of different mutations in the genome of different populations and species. There is evidence that the homogenization process is more efficient within clusters of neighboring monomers than between more distal monomers located in the same chromosome or in nonhomologous chromosomes (Hall *et al.*, 2005; Rudd *et al.*, 2006).

There have been relatively few attempts to characterize the nature of the 'ends' of long tandem arrays of satDNAs. Whole-genome projects have many sharp junctions of satellites with generally uncharacterized genomic DNA, although accurate assembly of individual sequence reads into satDNA arrays may be extremely difficult (see Lee *et al.*, 2006 for care required for assembly of blocks of HORs). Nevertheless, in examples that have been verified, abrupt transitions involving satDNA  $\times$  satDNA or satDNA  $\times$  defined non-satDNA junctions have been reported in a number of organisms, such as humans (Gaff *et al.*, 1994), plants (Alkhimova *et al.*, 2004; Lee *et al.*, 2006) and insects (McAllister and Werren, 1999; Sun *et al.*, 2003; Krzywinski *et al.*, 2005).

We have been studying different aspects of the satDNA structure, organization and evolution in the *Drosophila buzzatii* cluster (*D. repleta* group). This is a monophyletic group comprising seven closely-related cactus-breeding species of South American origin: *D. buzzatii*, *D. koepferae*, *D. serido*, *D. antonietae*, *D. gouveai*, *D. seriema* and *D. borborema* (Manfrin and Sene, 2006). These species share at least five abundant unrelated satDNAs, three of them formally described (Kuhn and Sene, 2005; Kuhn *et al.*, 2007; Franco *et al.*, 2008). Two of these satDNA families with contrasting evolutionary and structural characteristics have been studied in more detail. The AT-rich pBuM family can be divided into two subfamilies: pBuM-1 consists of  $\alpha$  repeats of approximately 190 bp, whereas pBuM-2 consists of  $\alpha/\beta$  repeats of approximately 370 bp units, each consisting of an  $\alpha$  unit and an additional 180 bp unrelated  $\beta$  unit (Kuhn and Sene, 2005). A complex pattern of variation of pBuM repeats across species was detected concerning abundance, chromosomal distribution, long-range organization, homogenization rate and nature of homogenized repeats (Kuhn *et al.*, 2008). Intraspecific homogenization for a particular set of  $\alpha$  or  $\alpha/\beta$  variant occurred in the genome of five species of the cluster. The second family, DBC-150, is slightly GC rich (55.3% on average) and comprises repeat units of approximately 150 bp, showing no homology to pBuM repeats (Kuhn *et al.*, 2007). In contrast to the pBuM family, no intraspecific homogenization for particular DBC-150 variants was detected in any of the seven species of the *buzzatii* cluster.

The basic karyotype of the seven species of the *buzzatii* cluster comprises one pair of sex chromosomes, four pairs of telocentric autosomes of similar size and one pair of microchromosomes (also known as the sixth chromosome pair or 'dot' chromosomes). The amount of

heterochromatin present in the sex chromosomes and in the microchromosomes discriminates the karyotypes of most species of the cluster (reviewed by Kuhn *et al.*, 2007).

The DBC-150 family was found to be restricted to the microchromosomes, whereas the pBuM family had a broader chromosomal distribution (Kuhn *et al.*, 2007, 2008). Several *Drosophila* species present one pair of microchromosomes, which are typically heterochromatic, contain only a few genes and are stably transmitted to the progeny (Powell, 1997). In *D. melanogaster*, the analysis of thousands of normal flies conducted by Bridges (1935) revealed that the microchromosomes do not undergo crossing-over during meiosis under natural conditions (reviewed by Riddle and Elgin, 2006). Recent studies suggested that suppressed meiotic recombination is probably a characteristic of the microchromosomes found in the *buzzatii* cluster too (Casals *et al.*, 2005; Kuhn *et al.*, 2007).

We showed previously by independent single-color fluorescence *in situ* hybridization (FISH) experiments that the microchromosome pair of *D. gouveai*, *D. seriema* and *D. antonietae* is highly enriched with repeats from both pBuM and DBC-150 families (Kuhn *et al.*, 2007, 2008). In this paper, we investigated the organization of pBuM and DBC-150 repeats in the metaphase microchromosomes, interphase nucleus and extended DNA fibers by double-target FISH, and analyzed the structure and dynamics of sequence junctions between the two satellites, isolated by PCR. Altogether, the data showed chromosomal regions and DNA fibers where the two satellites are interspersed at the molecular level, and sequence analysis of six pBuM  $\times$  DBC-150 junctions isolated from *D. antonietae* and *D. gouveai* helped to identify possible factors associated with such organization.

## Materials and methods

### *Drosophila* strains

The following strains were analyzed: *D. antonietae* (J23A90M), *D. gouveai* (J79M4) and *D. seriema* (D73C3B). The geographical location of these strains in Brazil can be seen in Kuhn *et al.* (2008). The laboratory strains J79M4 and D73C3B are isofemale lines, derived from single females collected in the wild. The strain J23A90M was founded by several females.

### Isolation of satDNA sequence junctions

Genomic DNA was extracted using standard procedures from a sample of approximately 20 adult flies from each strain. Different combinations of pBuM and DBC-150 primers were used to amplify pBuM  $\times$  DBC-150 sequence junctions. The pBuM primers A2-F: 5'-CGGAGTATTTTCATTTCGAC-3' and A2-R: 5'-GGTATGCCATAAGAAGTCG-3' are specific for  $\alpha$  pBuM motifs present in both  $\alpha$  and  $\alpha/\beta$  pBuM repeats; BPS-F: 5'-GGCTCTAGCTAGGTTATCCT-3' and BPS-R: 5'-GAGGGTTTACGTACCTAAG-3' are specific for  $\beta$  motifs from  $\alpha/\beta$  repeats. The DBC-150 primers DBC-P1: 5'-TCCAACACCCGCTATAACG-3' and DBC-P2: 5'-CCAAAATTCTACGCCGAACG-3' are specific for DBC-150 repeats. The PCR was composed of a 25  $\mu$ l reaction mixture containing 30–50 ng of genomic DNA, 0.4  $\mu$ M of each primer, 1  $\times$  PCR buffer, 0.2 mM (dNTPs) and 1 U Taq DNA

polymerase. The PCR program consisted of a first denaturation step at 94 °C for 2 min followed by 30 cycles at 94 °C for 1 min, at 45 or 50 °C for 1 min and at 72 °C for 1 min, with a final extension step at 72 °C for 5 min. PCR products were cloned into the pGEM-T plasmid vector (Promega, Madison, WI, USA) before sequencing in both directions.

### Sequence analysis

Multiple sequence alignments were performed using CLUSTALW (Thompson *et al.*, 1994) and improved by eye. The MEGA software version 4 (Tamura *et al.*, 2007) was used for the analysis of DNA primary structure, nucleotide variability, calculation of Kimura two-parameter (K-2P) distances (Kimura, 1980) and construction of neighbor-joining (NJ) dendrograms (Saitou and Nei, 1987). In addition to the NJ tree, a phylogenetical network was constructed using the Neighbor-Net method (Bryant and Moulton, 2004) combined with K-2P genetic distances. This method provides a more accurate representation of conflicting signals generated by recombination events, such as those involved in satDNA evolution.

### Molecular cytogenetics

The preparation of metaphase chromosomes and extended DNA fibers was carried out as described in Kuhn *et al.*, 2008. The interphase nuclei and metaphase chromosomes were obtained from neuroblasts of larvae at the third stage. DNA fibers were isolated from adult flies. Probes for the detection of pBuM and DBC-150 repeats were prepared by PCR labeling. The reactions consisted of a 25  $\mu$ l mixture containing 1–10 ng of plasmid DNA, 0.6  $\mu$ M of each primer, 1  $\times$  PCR buffer, 0.2 mM each dATP, TTP, dCTP and dGTP, 0.04 mM of digoxigenin-11-dUTP or biotin-16-dUTP (Roche, Indianapolis, IN, USA) and 1.5 U Taq DNA polymerase (Yorkshire Bioscience, York, UK). The M13 primers were used for labeling cloned  $\alpha/\beta$  or DBC-150 monomers. In addition, a  $\beta$ -probe lacking plasmid-flanking sequences was prepared with a PCR reaction containing an  $\alpha/\beta$  cloned monomer as template and the  $\beta$ -specific primers BPS-F and BPS-R.

Probes labeled with biotin and digoxigenin were detected with streptavidin-Alexa 594 and anti-digoxigenin-FITC, respectively. *In situ* hybridization experiments were performed as described previously (Kuhn *et al.*, 2008). After hybridization, the chromosomes (or DNA fibers) were counterstained with DAPI (4',6-diamidino-2-phenylindole) and analyzed with a Zeiss Axioplan epifluorescence microscope equipped with a Jenoptik ProgRes C12 digital image capture system. Images were processed in Adobe Photoshop CS2 using only cropping and adjustment functions that affected the whole image equally.

The length of single-DNA fibers was measured considering that 320 pixels (objective  $\times$  100) or 180 pixels (objective  $\times$  63) correspond to 10  $\mu$ m, which roughly corresponds to 29 kb of the extended DNA fiber.

## Results

### Organization of pBuM and DBC-150 repeats

Double-target FISH with DBC-150 and pBuM ( $\alpha/\beta$ ) probes revealed contrasting patterns of genomic organi-

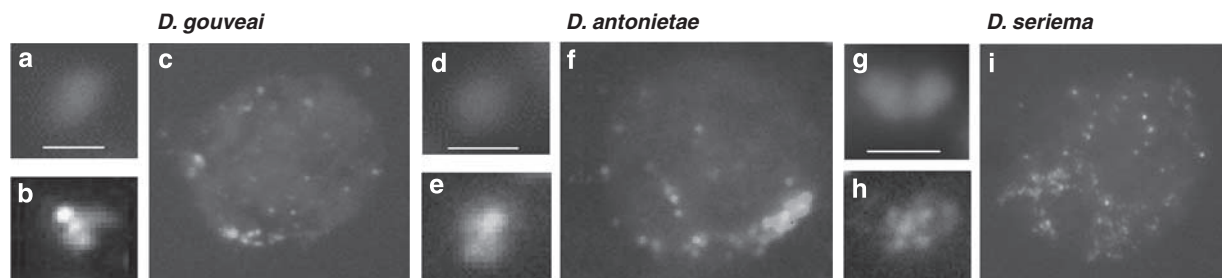
zation of the two satellite families (Figure 1). In *D. gouveai* and *D. antonietae*, the distribution of DBC-150 repeats overlaps with the distribution of pBuM repeats ( $\alpha$  or  $\alpha/\beta$ ) in some regions of the microchromosome (Figures 1a, b, d and e, yellow color showing overlapping hybridization). The submetacentric microchromosome of *D. seriema* showed pBuM and DBC-150 repeats occupying different chromosome domains (no yellow color in Figure 1h), with DBC-150 repeats clustering in the pericentromeric region and pBuM repeats clustering at the subtelomeric regions of both chromosome arms (Figures 1g and h). In *D. gouveai* and *D. antonietae*, zones of overlapping distribution between the two satDNAs were evident in the interphase nuclei (Figures 1c–f). In *D. seriema*, large arrays of pBuM and DBC-150 repeats are organized in distinct domains in the interphase nucleus, with overlapping distribution detected only in some punctuate genomic regions (Figure 1i).

High-resolution double-target FISH in extended DNA fibers revealed several examples where pBuM and DBC-150 repeats were found interspersed with each other within the same arrays (Figure 2). The same pattern of interspersions was observed with probes prepared with PCR using M13 or  $\beta$ -specific primers, or where labels were exchanged. The most common interspersions pattern observed in *D. gouveai*, *D. antonietae* and *D. seriema* consisted of long arrays (> 100 kb) comprising predominantly repeats of one family but with a few intruding repeats from the other (Figures 2a–e). In two cases, we detected adjacent arrays of pBuM and DBC-150 (Figures 2a and b). The highest level of interspersions of pBuM and DBC-150 repeats was found in *D. gouveai*, with repeats showing mutual arrangements in regions spanning more than 100 kb of the chromatin fiber (Figure 2c). In contrast, *D. seriema* showed continuous arrays of repeats from one satDNA with few (Figure 2e) or no intruding repeats from the other (Figure 2f).

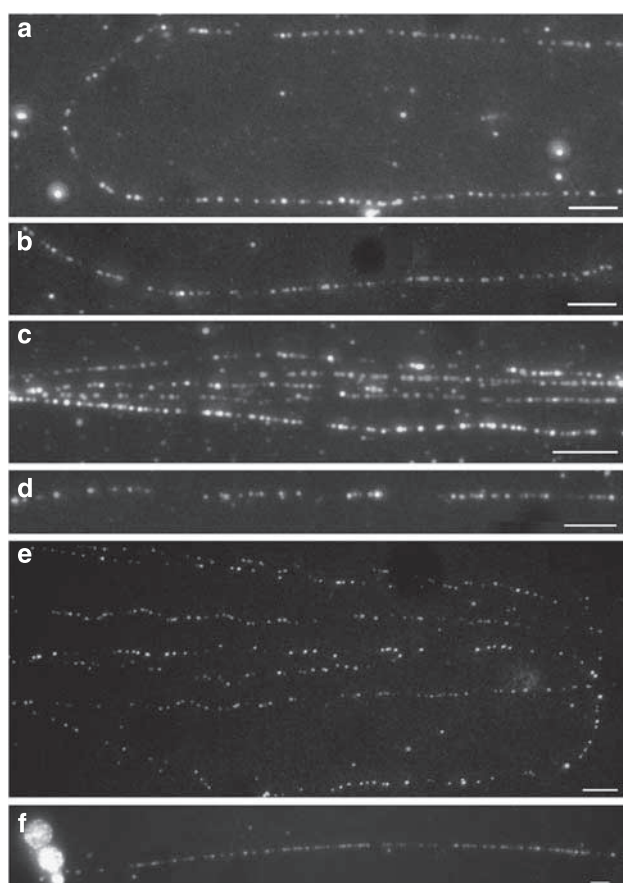
The pattern of interspersions of pBuM and DBC-150 in the same DNA fibers is consistent with the numerous gaps observed within pBuM arrays in our previous report (Kuhn *et al.*, 2008). Apparent gaps with no hybridization of either satDNA were also observed in some arrays. Some may be for technical reasons (for example, breakage of single fibers), but we also investigated possible interspersions with a major transposon family. The foldback element Galileo was found to accumulate in the heterochromatic regions of several chromosomes and was more pronounced in the microchromosomes of several species from the *buzzatii* cluster (Casals *et al.*, 2005). However, PCR experiments (data not shown) using single Galileo primers (G7 and E14 in Casals *et al.*, 2005) and a combination of Galileo and pBuM or DBC-150 primers showed little evidence of interspersions between Galileo and pBuM or DBC-150 arrays in genomic regions spanning 0.1–2 kb.

### Characterization of junctions between pBuM and DBC-150 satDNA families

PCR amplification of genomic DNA in *D. gouveai*, *D. antonietae* and *D. seriema* using combinations of DBC-150 and pBuM ( $\alpha$  and  $\alpha/\beta$ ) primers produced DNA fragments that varied in number and size according to the species and combination of primers tested. In total, we sequenced nine *D. antonietae* clones derived



**Figure 1** Double-target *in situ* hybridization with  $\alpha/\beta$  (pBuM) (red) and DBC-150 (green) probes to the microchromosomes (left half of each image) or interphase nuclei (right half of each image) of *D. gouveai*, *D. antonietae* and *D. seriema*. Yellow signals show regions where the two satellite sequences are interspersed. In **a**, **d** and **g**, the microchromosomes are shown counterstained with DAPI (4',6-diamidino-2-phenylindole, blue) and **b**, **e** and **h** show only the hybridization signals generated with the two satellite probes. Pictures of representative interphase nuclei (**c**, **f** and **i**) show overlays of the pBuM (red), DBC-150 (green) and DAPI DNA stain (blue). Scale bar: 1  $\mu\text{m}$  except 2  $\mu\text{m}$  in **g**. The color reproduction of this figure is available on the html full text version of the manuscript.



**Figure 2** Extended DNA fibers hybridized with  $\alpha/\beta$  (pBuM) (red) and DBC-150 (green) probes showing the organization of pBuM and DBC-150 satDNAs in representative fibers from *D. gouveai* (**a**–**c**), *D. antonietae* (**d**) and *D. seriema* (**e** and **f**). Interspersion and junctions are visualized by adjacent red and green hybridization sites, yellow where signals overlap. Scale bar represents 10 kb of stretched DNA. The color reproduction of this figure is available on the html full text version of the manuscript.

from six DNA bands and four primer combinations, ten *D. gouveai* clones derived from another six DNA bands and three primer combinations, and three *D. seriema* clones derived from two DNA bands and two primer combinations (Table 1). Sequence analysis of clones from the first two species showed that they belong to six

different satellite junctions, two of them exclusive to *D. antonietae* (junctions I and II), and four exclusive to *D. gouveai* (junctions III–VI) (Table 1). There was no direct relationship between the number of DNA bands, primer combination and the number of junctions. For example, sequences from junction II of *D. antonietae* were isolated from DNA bands of four different sizes generated with primers DBC-P1  $\times$  A2-R, sequences from junctions IV and V came from a same DNA band obtained with primers DBC-P1  $\times$  A2-R, and sequences from junction I, II and III came from distinct DNA bands generated with two different primer combinations (Table 1). These results also confirm that the satDNA junctions are not PCR artifacts.

Figure 3A shows a schematic representation of the nature and organization of pBuM and DBC-150 repeats involved in each satDNA junction type. DNA spacers were not detected between the two satellites. Instead, the junctions comprise only  $\alpha/\beta$  (pBuM family) and DBC-150 repeats, with a number of adjacent repeats of each satellite flanking the junction site. This situation thus allowed pBuM or DBC-150 primers to anneal at different 'junction repeats' (JRs), producing in some cases DNA products of different sizes derived from the same pBuM  $\times$  DBC-150 junction.

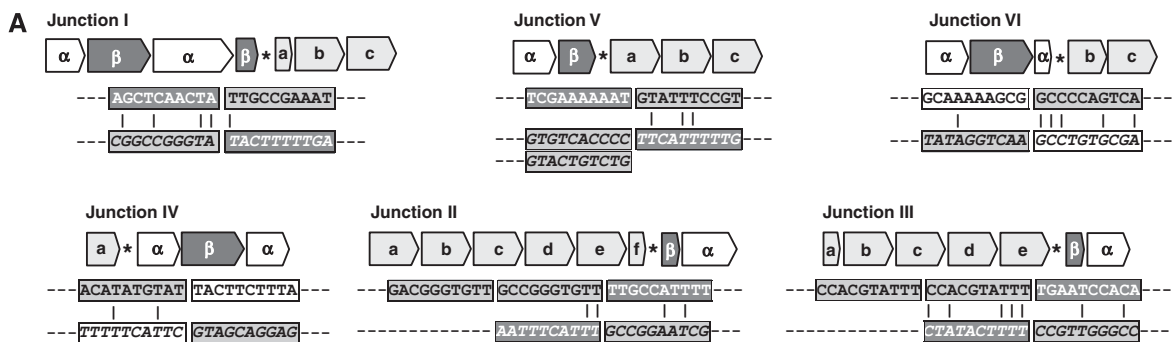
The sequences of junctions were aligned with the monomer sequences of DBC-150 and pBuM to characterize the molecular breakpoints (Figure 3A). Junctions I, IV, V and VI are abrupt transitions between pBuM and DBC-150 JRs. Junctions II and III are characterized each by a different 10 bp tandem duplication of a DBC-150 JR immediately bordering the junction. Each breakpoint occurred at a different position in both pBuM and DBC-150 monomers (Figures 3B and C), suggesting each was an independent recombination event. An alignment (Figure 3A) made with the 20 bp sequence of each junction with the predicted following nucleotide sequence of each repeat assuming no junction showed very low degree of sequence identity between the sequences involved in the junction (from 10% in junction IV to 35% in junction III). In four junctions (I, II, III and VI), 2–3 nucleotides immediately flanking the junction are ambiguous and could be classified as either belonging to a pBuM or a DBC-150 sequence.

None of the three clones from *D. seriema* obtained with primers DBC-P2  $\times$  BPS-F or DBC-P2  $\times$  A2 represents satDNA junctions. The clone Sma220/1 showed no

**Table 1** Satellite DNA sequences

Species	Primer combination	Annealing temp. (°C)	Size of DNA-sequenced products (bp)	Number of sequenced clones	Sequence annotation of clones	GenBank accession numbers	Structure <sup>a</sup>		Junction type
							DBC-150	pBuM	
<i>D. antonietae</i>	DBC-P1 × A2-R	54	868	1	Ant870-J2/1	EU543951	a, b, c, d, e, f	β, α	II
			604	3	Ant600-J2/1–3	EU543952– EU543954	c, d, e, f	β, α	II
			342	2	Ant340-J2/1,2	EU543956– EU543957	e, f	β, α	II
	DBC-P1 × BPS-R DBC-P2 × A2-F DBC-P2 × BPS-F	45	499	1	Ant500-J2/1	EU543955	c, d, e, f	β	II
			820	1	Ant820-J1/1	EU543949	a, b, c	α, β, α, β	I
			314	1	Ant310-J1/1	EU543950	a, b, c	β	I
<i>D. gouveai</i>	DBC-P1 × A2-R	54	704	1	Gou700-J3/1	EU543958	a, b, c, d, e	β, α	III
			440–444	4	Gou440-J3/1–4	EU543961– EU543964	d, e,	β, α	III
			446	1	Gou450-J4/1	EU543965	a	α, β, α	IV
	DBC-P2 × A2-F DBC-P1 × BPS-R	45	582	1	Gou580-J5/1	EU543966	a, b, c	α, β	V
			509	1	Gou510-J6/1	EU543967	b, c	α, β, α	VI
			336–345	2	Gou340-J3/1,2	EU543959– EU543960	d, e	β	III
<i>D. seriema</i>	DBC-P2 × BPS-F DBC-P2 × A2-F	45 54	217	1	Sma220/1	EU926749	—	—	—
			366–368	2	Sma370/1,2	EU926750– EU926751	a, b, c	—	—
<i>D. serido</i>	DBC-P1 × DBC -P2	60	372–374	12	Sdo370/1–12	EU543933– EU543944	a, b, c	—	—
			262	4	Sdo260/1–4	EU543945– EU543948	a/b, b/c	—	—

<sup>a</sup>See Figure 3.



**B**  
**Alpha/beta (pBuM)**  
TACTGTCGCAAAAAGCG\***J6**\*GCCTGTGCGAAAAAATTGACTATTTTCGAAGTTTAAACAAGCTATAACCGGAGTATTTTCATTCC\***J4**\*GACTTCTTTATGGCATACCATTTTAGAAGCGTCTTTATCGGAGAACTCAGATATATCTAAGGATCTGGCATGGTCTAAGAACTTTCCGAAATATTCACATAAATCCCA**TAAAAATATGACCAATTGAAC**TTTAAATGGCTCTAGATAGGTTATCCTTTGTCCGATTTTAAAGCTCAACTA\***J1**\*TACTTT\***J3**\*TGAATCGACATAGACAGTTCTTTTGATATAAACACAAAAAATTTTCTCGAAAAAAT\***J5**\*TTCATTT\***J2**\*TTGCCATTTTAAACCATCTTAGGTAACGTAACCCCTGGT

**C**  
**DBC-150**  
TTCGGCGCAGAAAATTGGGTGTCACCCC\***J5**\*GTATTT\***J3**\*CCGTTGGCCGGGTTATAGGTCAA\***J6**\*GCCCCAGTCACATATGTAT\***J4**\*GTAGCAGGAGCGCCGGGTG\***J1**\*TT\***J2**\*GCCGAAATCGCGGTCCGGATATATACCGGTACTGTCCAGATCCAACCCCGGTATAACG

**Figure 3** (A) Schematic representation of the α/β (pBuM) and DBC-150 junction repeat units (JR) found in the six different junctions (I–VI). Neighbor DBC-150 JRs are distinguished by letters (a–f), which are not meant to be equivalent between different junctions. A 20 bp sequence spanning the junction is shown in detail below the representation of each junction. The upper strand represents the actual sequence junction, and the lower strand represents the predicted nucleotide sequence continuation of each repeat reconstructed based on DBC-150 and pBuM consensus sequences. In Junction V, two possibilities for the predicted nucleotide sequences of the DBC-150 JR are given (see also Figure 4). The sequences were shaded with different colors corresponding to each family. (B) Location of the six junctions in an α/β consensus sequence derived from α/β junctions. The β sequence is in bold. (C) Distribution of the six junctions in a DBC-150 junction consensus sequence.

significant sequence similarity to pBuM, DBC-150 or any other sequence available in the GenBank (July 2008). The two clones Sma370/1-2 contain only DBC-150 neighbor repeats.

#### Nature and organization of DBC-150 and pBuM junction repeats

The six DBC-150 JRs from *D. gouveai* and *D. antonietae* share some nucleotide features (including nucleotide substitutions and indels) that could not be found or are found at low frequencies in a collection of 35 randomly isolated DBC-150 monomers from all seven species of the *buzzatii* cluster (see nucleotide alignment in Figure 4), indicating that the same pool of DBC-150 variants was involved in the formation of all six satDNA junctions. Among these features is a long, almost perfect, palindrome sequence 5'-CCGGA(TA)<sub>n</sub>CCGG-3' (positions 116–130 in Figure 4) present in at least 14 of the 20 isolated DBC-150 JRs (its presence or absence could not be determined in some JRs directly adjacent to the break-points, such as those in junctions II and IV). Another hallmark of DBC-150 JRs is a deletion of an approximately 50 bp segment spanning from the end of a repeat unit (positions 141–162) to the beginning of the following repeat unit (positions 1–27) (Figure 4). The average nucleotide variability between neighbor DBC-150 JRs in *D. gouveai* and *D. antonietae* is also higher (~13 and ~12%) compared to that observed between randomly isolated monomers from each species (~7 and ~10%).

Nucleotide comparisons and NJ trees showed that  $\alpha/\beta$  (pBuM) JRs in *D. gouveai* and *D. antonietae* belong to the same  $\alpha/\beta$  variant type characteristic of each species ( $\alpha/\beta$  I and III in Kuhn *et al.*, 2008; data not shown). Therefore, the two junctions of *D. antonietae* and the four junctions of *D. gouveai* probably originated independently after the split of these two species from their common ancestor. Within each species,  $\alpha/\beta$  JRs are indistinguishable from a collection of 7 or 36 randomly isolated monomers from *D. antonietae* and *D. gouveai*, respectively. Therefore, in contrast to DBC-150 JRs, no atypical nucleotide features can be related to the  $\alpha/\beta$  JRs.

The DBC-150 junction repeats are genetically related to a DBC-HOR present in *D. serido*

PCR with DBC-150 primers (DBC-P1/P2) produced DNA fragments corresponding to the sizes of monomers and multimers of DBC-150 repeats in all species of the *buzzatii* cluster (Kuhn *et al.*, 2007). However, in addition to these DNA fragments, *D. serido* showed a rather prominent DNA fragment at a size of approximately 400 bp, which roughly corresponds to the size of DBC-150 trimers. We sequenced 12 clones from these DNA bands, derived from two independent PCR experiments. Sequence analysis revealed that they comprise three adjacent DBC-150 repeats (a, b and c), which are arranged in the form of a 3-mer HOR (Supplementary material S2). The average nucleotide variability between the a, b, c subrepeats is  $9.4 \pm 1.8\%$  and between the 3-mer HORs, the average nucleotide variability is  $8 \pm 0.9\%$ . The average nucleotide variability of the 3-mer HOR is slightly lower than that observed among random DBC-150 monomers ( $9.2 \pm 1.5\%$ ). The sequence analysis from four additional clones of *D. serido* obtained from an approximately 300 bp PCR DNA band with same

primers revealed dimers originated by recombination between a, b, c subrepeats (see below).

Consensus sequences derived from each a, b, c subrepeats from the 12 *D. serido* 3-mer HOR are shown in the nucleotide alignment of Figure 4c. The 3-mer HOR of *D. serido* shares several nucleotide features in common to DBC-150 JRs, including the 5'-CCGGA(TA)<sub>n</sub>CCGG-3' palindromic sequence, the 50 bp deletion and several other nucleotide substitutions that are frequent among DBC-150 JRs of *D. gouveai* and *D. antonietae* (Figure 4b). Therefore, the origin of these structural changes predates the cladogenesis of the three species.

#### Phylogenetic relationships of DBC-150 sequences

The NJ tree in Figure 5a shows that DBC-150 JRs from *D. gouveai* and *D. antonietae* are grouped into two main branches, one composed of repeats with the palindromic sequence and another without it. The latter group is clustered in a separate branch within the main branch leading to 35 DBC-150 monomers from all seven species of the *buzzatii* cluster. Repeats from different junctions and species were clustered together, confirming previous observations of a general lack of species specificity of repeats belonging to the DBC-150 satDNA family (Kuhn *et al.*, 2007).

The a and b subrepeats from the 3-mer HOR of *D. serido* were clustered in the same branch containing DBC-150 JRs presenting the palindrome, whereas the c subrepeat was clustered with the DBC-150 monomers (Figure 5a). This result raises the possibility that the satDNA junctions involved DBC-150 repeats with a certain level of higher-order organization.

Five (out of six) repeats from junction II were clustered in two specific branches. One branch comprises repeats a, c and e, whereas the other comprises repeats b and d (Figure 5a). Such alternating pattern of interrepeat variation suggests that the DBC-150 JRs from junction II (panels a–f in Figure 3A) are members of a 2-mer HOR. In fact, the average nucleotide variability among all DBC-150 JRs from junction II is 12.2%, but falls to the half (6.0%) when the repeat size includes the 2-mer HOR (ab, cd, ef). Evidence of HORs was not detected in the remaining satDNA junctions, although we cannot exclude out the possibility that the number of adjacent DBC-150 JRs was too small for their detection.

The same DBC-150 sequences were used to construct a phylogenetic network based on the Neighbor-Net method. The resulting network (Figure 5b) displayed splits representing the same main groups of repeats illustrated in the NJ tree. The presence of several boxes throughout the network illustrates conflicting splits more likely generated by recombination events between repeats.

#### Recombination between satDNA repeats

Recombination events between neighbor DBC-150 satDNA repeats were detected seven times in our data. All the recombination events resulted in total reciprocal exchanges. In the satDNA junctions, the events involved repeats e/c (Rec1) and e/f (Rec2) from junction II (*D. antonietae*), and b/d (Rec3) from junction III (*D. gouveai*) (Supplementary material S1). The average pair-wise variability values between repeats involved in Rec1, Rec2 and Rec3 are 5.2, 13.7 and 6.1%, respectively. Three clones representing 3-mer HORs of *D. serido* showed





recombination between b/c (Rec4-6), and one clone showed recombination between a/b repeats (Rec7) (Supplementary material S2). The average nucleotide variability between repeats involved in Rec4-6 is 10.9% and in Rec7 is 14.4%.

#### At least one of the junctions is repeated

The seven clones spanning the junction II of *D. antonietae* (Table 1) represent at least six different sequences showing variability values ranging from 0.0 to 3.2% (average  $0.9 \pm 0.2\%$ ). As the clones were derived from a strain founded by several females (collection date 22 March 1998), they may represent either different alleles or different paralogous sequences.

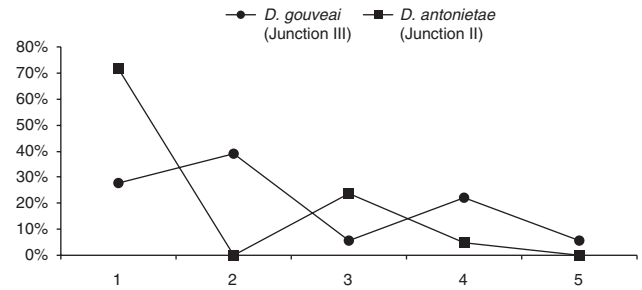
The seven clones spanning the junction III of *D. gouveai* were isolated from a strain founded by a single female collected recently from nature (collection date 27 February 2002). Either one (homozygous) or two (heterozygous) different types of sequences would be expected if they represent different alleles. Notably, these seven clones represent seven different sequences showing pairwise variability values ranging from 0.3 to 4.4% (on average  $1.9 \pm 0.4\%$ ), so there are at least five copies of junction III in the genome of *D. gouveai*. The higher nucleotide variability of junction III compared to junction

II clones (Figure 6) probably reflects a more ancient origin of junction III.

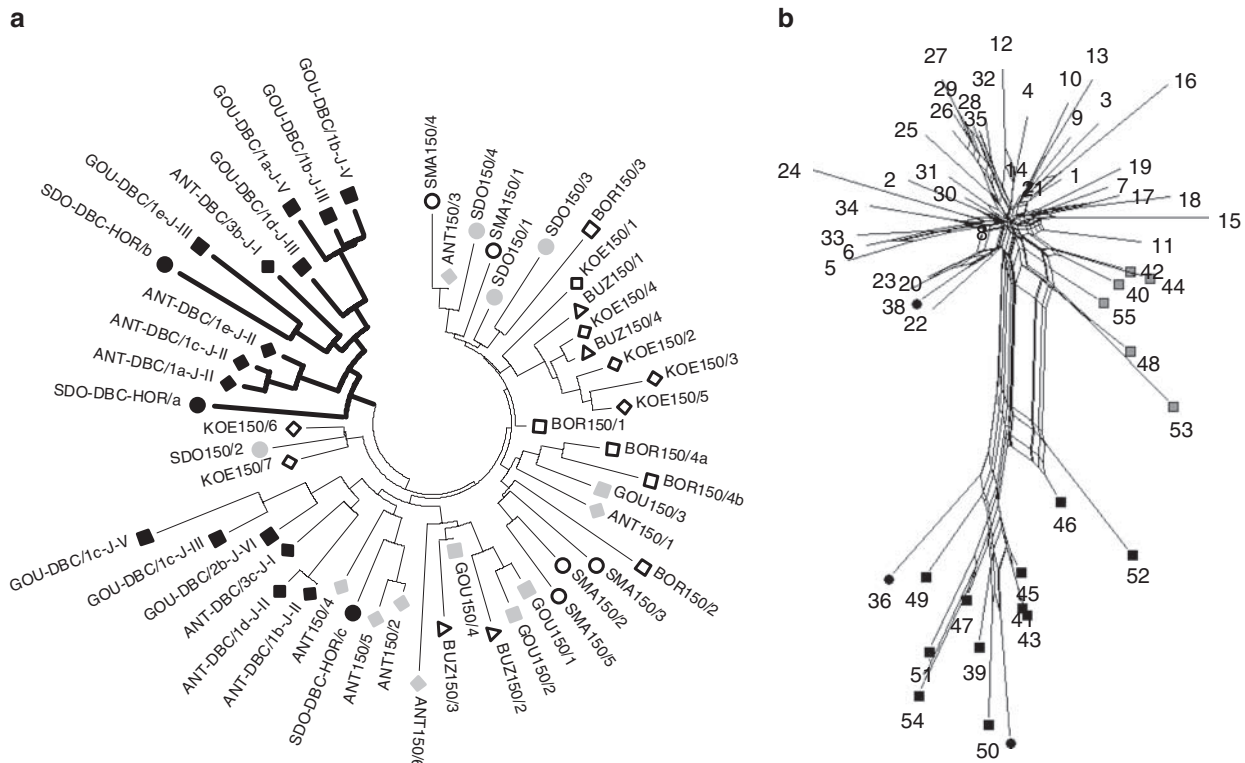
## Discussion

### Interspersion of satellite DNA families

Using FISH on metaphase and interphase chromosomes, we showed that the localization of DBC-150 and pBuM satDNA families overlaps in some regions of the



**Figure 6** Graph showing the frequency distributions (y axis) of pairwise nucleotide variability frequencies (x axis) between full DBC-150 × pBuM junction sequences, calculated individually for *D. gouveai* (junction III) and *D. antonietae* (junction II).



**Figure 5** (a) Neighbor-joining tree containing (1) consensus sequences derived from each DBC-150 junction repeat (JR) present in junctions I, II, III, V and VI isolated from *D. antonietae* (ANT-DBC) and *D. gouveai* (GOU-DBC); (2) consensus sequences derived from each a, b and c subrepeats from *D. serido* 3-mer higher-order repeats (HORs) (SDO-DBC-HOR/a–c) and (3) 35 randomly isolated DBC-150 monomers from *D. antonietae* (ANT150), *D. gouveai* (GOU150), *D. serido* (SDO150), *D. seriema* (SMA150), *D. koepferae* (KOE150), *D. borborema* (BOR150) and *D. buzzatii* (BUZ150). Partial DBC-150 JR sequences with less than 80 bp were not included in generation of the tree (for example, GOU-DBC/1a-J-IV). Different symbols in the beginning of each sequence were added to facilitate discrimination of repeats from each species and junction. Black symbols represent DBC-150 JRs and *D. serido* 3-mer HORs; gray and white symbols represent DBC-150 monomers. The branch leading to DBC-150 JRs containing the 11 bp palindromic sequence is shown in bold. (b) Neighbor-Net network with the same 55 sequences shown in (a). Squares represent junction sequences (sequences featuring palindromes are black-shaded) and circles represent a, b and c subrepeats from *D. serido* 3-mer HORs. In (a) and (b) genetic distances were calculated according to Kimura two-parameter (K-2P) model.

microchromosomes of *D. gouveai* and *D. antonietae* and high-resolution FISH revealed extended DNA fibers containing interspersed repeats from both the pBuM and DBC-150 families (Figures 1 and 2). Notably in *D. gouveai*, mutual arrangements of DBC-150 and pBuM repeats were found in regions spanning more than 100 kb of the chromatin fiber. The species *D. seriema* showed the lowest level of interspersion of both satellites (Figures 1 and 2).

Nonhomologous satDNA families present in the same species and chromosomal region(s) are typically organized in separate arrays. For example, in *D. melanogaster*, a different set of satDNA families are often located near one another on the same chromosome (Lohe *et al.*, 1993) and sequence analysis of a 31 kb DNA centromeric region from the X-derived *Dp1187* minichromosome revealed that the highly repeated satellites AATAT and TTCTC are organized in uniform arrays directly juxtaposed, with neither interspersion nor intervening DNA (Sun *et al.*, 2003). In humans, the centromeric region of each chromosome contains at least one  $\alpha$ -satellite DNA together with a unique combination of other satDNAs. These satDNAs are organized in discrete and uniform arrays that can be directly adjacent to each other, such as satellites 1, 3 and  $\beta$  in chromosome 22 (Shiels *et al.*, 1997).

Reports of interspersion of satDNAs usually involve repeats with marked homologies, such as satellites I and II in the beetle *Tribolium madens* (Zinic *et al.*, 2000); pBuM-1 and pBuM-2 satellite subfamilies in the *D. buzzatii* cluster (Kuhn *et al.*, 2008) or *psr2* and *psr18* in the wasp *Nasonia vitripennis* (Reed *et al.*, 1994). Interspersion of nonhomologous satellites is rarely reported, although was observed between the pSc200 and pSc250 satellites in subteleric regions of rye chromosomes (Alkhimova *et al.*, 2004).

Most studies on satDNAs involve the characterization of repeat units, their chromosomal location and their evolutionary dynamics across different species. In contrast, there are only few data sets about the structure and organization of satDNA arrays at the molecular level (Shiels *et al.*, 1997; Zinic *et al.*, 2000; Alkhimova *et al.*, 2004; Lee *et al.*, 2006). It remains to be investigated whether the interspersion noted here is a regular occurrence or associated to some peculiar features of the *Drosophila* microchromosomes, such as highly heterochromatic nature, suppressed meiotic recombination and presence of few genes (reviewed by Riddle and Elgin, 2006), allowing a more complex interplay between satDNAs without deleterious effects.

#### Illegitimate recombination and the origin of satDNA $\times$ satDNA junctions

In mammals and plants, *in vivo* and *in vitro* experiments showed that efficient recombination between sequences requires a minimum of approximately 200 bp of uninterrupted homology (Rubnitz and Subramani, 1984; Opperman *et al.*, 2004). Below this threshold, sequence divergence has an inhibitory effect on homologous recombination and can decrease threefold with a divergence level as low as 0.16%. In our data, all recombination products between DBC-150 direct repeats involved sequences with more than 5% sequence divergence (reaching 14% in Rec7). In another study,

Okumura *et al.* (1987) detected recombination between human  $\alpha$ -satellite DNA repeats ( $\sim$ 171 bp long) showing 20–30% sequence divergence. Therefore, recombination among neighbor satDNA repeats seems to require less homology than in other systems.

Very little sequence identity (<35%) exists over 20 bp regions spanning the  $\alpha/\beta \times$  DBC-150 junctions, and the  $\alpha/\beta$  and DBC-150 JRs do not share any segment of 20 bp with more than 70% sequence identity, ruling out the involvement of homologous recombination in generation of the satDNA interspersion. pBuM and DBC-150 repeats do not have structural, sequential or terminal features of transposons, so we do not consider transposase-mediated transposition as a mechanism for amplification of repeats or generation of the interspersion patterns. Therefore, our results indicate that alternative illegitimate recombination mechanisms created the satDNA junctions.

Illegitimate recombination can occur by different mechanisms (reviewed by Müller *et al.*, 1999). In the 'breakage-and-rejoining' model, free DNA ends can be joined either by direct ligation of blunt ends or by pairing of single-stranded cohesive ends mediated by short stretches (2–6 bp) of homology (microhomologies). In the 'copy-choice' model, illegitimate recombination can occur during DNA replication, when the polymerization of a new DNA strand is interrupted at certain sites and continues at a different template position. Such slippage events can be facilitated by the presence of palindrome sequences forming secondary structures, such as that found in one of the junctions between satellites pSc200 and pSc250 in rye (Alkhimova *et al.*, 2004).

Sequence analysis of six different DBC-150  $\times$  pBuM satDNA junctions revealed that they consist of a number of  $\alpha/\beta$  and DBC-150 JRs flanking abrupt transitions where there are no intervening bases between the two satellite sequences. The absence of intervening bases in our junctions favors the 'breakage-and-rejoining' model as the main mechanism creating at least the six isolated satellite junctions. This is further supported by the fact that microhomologies between DNA ends were found in four junctions (I, II, III and VI). Illegitimate recombination is likely to occur during DNA transcription or replication, when DNA topoisomerases introduce nicks on one or both DNA strands to change the superhelical state of DNA (Wang, 1985), leaving temporarily free 3' DNA ends. In our data, such satDNA exchanges might have been facilitated by the fact that some arrays are very close to each other (Figures 2a and b).

Two of the junctions (II and III) can also be distinguished by the presence of a tandem duplication of different DBC-150 JR's 10 bp segment, immediately adjacent to the junction. Interestingly, a 9 bp tandem duplication of the human  $\alpha$ -satellite monomer was also found in abrupt junctions between this satellite and satellite III (Gaff *et al.*, 1994), pointing to a similar recombination pathway.

All the DBC-150 JRs feature in at least one of the bordering repeats an almost perfect palindrome sequence 5'-CCGGA(TA)<sub>n</sub>CCGG-3', which was not found in a collection of 35 DBC-150 monomers (Figures 4 and 5). The involvement of palindromes in recombination events between repeats has been suggested by other studies (Stary and Sarasin, 1992; Reed *et al.*, 1994). The 11 bp gap observed in the region of the palindrome

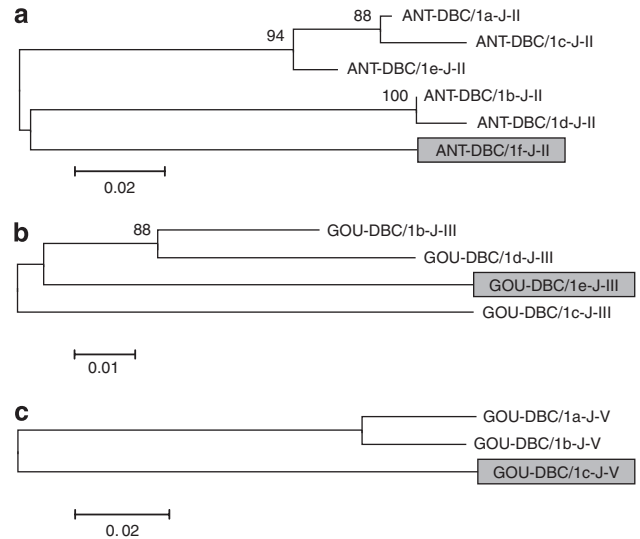
can be explained by the fact that palindromes form hairpin structures that can be deleted during DNA replication, following repair mechanisms (Trinh and Sinden, 1993).

We isolated two different pBuM  $\times$  DBC-150 satDNA junctions of *D. antonietae* and another four of *D. gouveai* through PCR. The junctions were found in pBuM repeats, which are characteristic of each species, suggesting that they probably originated after the split of the two species from their common ancestor. Each junction originated by independent recombination events. Taking into account the interspersion level observed between pBuM and DBC-150 repeats in some arrays (Figure 2), it is realistic to expect that additional junctions are present in the genome of these two species. First, they might contain divergent pBuM and DBC-150 sequences that could have missed amplification due to our PCR settings (primer composition and annealing temperatures). Second, different junctions might be present in the same DNA band, as we showed twice in our data (Table 1). Finally, we also found evidence that at least one junction (junction III of *D. gouveai*) is repeated. Unequal crossing-over between homologous sequences flanking the junction is the most likely mechanism responsible for the propagation and consequently increase of the interspersion organization of satDNA junctions. Examples of reiteration of specific satDNA  $\times$  satDNA junctions are scarce in the literature. In humans, a junction between satellite III and  $\alpha$ -satellite DNAs is present in at least three chromosomes (Gaff et al., 1994), and in the beetle *T. madens*, there are multiple copies of a junction between satellites I and II (Mravinac and Plohl, 2007).

#### Evolutionary dynamics of satDNA $\times$ satDNA junctions

Unequal crossing-over between sister chromatid repeats has been proposed to be a mechanism that promotes both homogenization and amplification/deletion of satDNA repeats during evolution (Smith, 1976). The model also predicts that repeats immediately adjacent to satDNA  $\times$  satDNA or satDNA  $\times$  non-satDNA junctions should present higher levels of sequence divergence compared to repeats from more central regions of the array, where unequal crossing-over is more likely to happen. There are very limited data sets on the literature concerning the analysis of satDNA junctions under this context, but they generally support the expectations of the model (McAllister and Werren, 1999; Schueler et al., 2001; Mravinac and Plohl, 2007). The homogeneity of a junction between the AATAT and TTCTC satellites in the *D. melanogaster* Dp1187 minichromosome is an exception (Sun et al., 2003). Probably other molecular mechanisms, such as replication slippage, might be important in the homogenization and amplification of simple satDNA repeats.

In our data,  $\alpha/\beta$  repeats bordering the pBuM  $\times$  DBC-150 junctions from *D. antonietae* and *D. gouveai* cannot be distinguished from a collection of random isolated  $\alpha/\beta$  monomers from each species, indicating similar rates of mutation and homogenization in JRs and non-JRs. In contrast, neighbor DBC-150 JRs showed on average higher interrepeat variability compared to random isolated DBC-150 monomers from each species. Although we lack evidence of functionality for pBuM



**Figure 7** Neighbor-joining tree containing DBC-150 junction repeats (JR) a–f from junction II (a), b–e from junction III (b) and a–c from junction V (c). The last repeat from each junction is gray-shaded. Bootstrap values  $>70\%$  indicated at branches; scale bar represents genetic distances calculated according to Kimura two-parameter (K-2P) model.

or DBC-150 sequences, it is unlikely that selection has played a major role in the conservation of  $\alpha/\beta$  JRs present in all six independent junctions. A possible explanation is that some satDNA junctions we sequenced came from small ‘islands’ of DBC-150 repeats located within long pBuM arrays. FISH on DNA fibers supports this pattern of organization. In such a situation, the homogenization rate of  $\alpha/\beta$  JRs is not expected to differ from the overall rate of homogenization in the array, because they are separated by only a small cluster of DBC-150 repeats. In contrast, unequal exchanges and consequently homogenization between adjacent DBC-150 JRs are expected to occur at a lower rate, with mutations tending to accumulate at a higher rate (Dover, 1982). We detected three recombination events between DBC-150 JRs (Supplementary material S1) and one of them just involved the last two repeats before the junction (e/f in junction II). However, such recombination events could have happened before the DBC-150 transposition event and the formation of junction II. In three junctions where the number of neighbor JRs enabled interrepeat comparisons, we verified that the last DBC-150 JR before the junction is also the most divergent repeat in two cases (Figure 7). Therefore, the dynamics of DBC-150 JRs is generally in agreement with the unequal crossing-over model, but the dynamics of  $\alpha/\beta$  JRs is clearly not.

#### Acknowledgements

We thank Fabio M Sene and Maura H Manfrin (University of São Paulo, Brazil) for their contribution to this work. G Kuhn was supported at the University of Leicester by a postdoctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Proc. 1158/06-6.

## References

- Alkhimova OG, Mazurok NA, Potapova TA, Zakian SM, Heslop-Harrison JS, Vershinin AV (2004). Diverse patterns of the tandem repeat organization in rye chromosomes. *Chromosoma* **113**: 42–52.
- Bachmann L, Sperlich D (1993). Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis* and *D. obscura*. *Mol Biol Evol* **10**: 647–659.
- Biscotti A, Olmo E, Canapa A, Barucca M, Teo CH, Schwarzacher T et al. (2007). Repetitive DNA, molecular cytogenetics and genome organization in the King Scallop (*Pecten maximus*). *Gene* **406**: 91–98.
- Bridges CB (1935). The mutants and linkage data of chromosome four of *Drosophila melanogaster*. *Biol Zh (Moscow)* **4**: 401–420.
- Bryant D, Moulton V (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255–265.
- Casals F, Cáceres M, Manfrin MH, González J, Ruiz A (2005). Molecular characterization and chromosomal distribution of Galileo, Kepler and Newton, three foldback transposable elements of the *Drosophila buzzatii* species complex. *Genetics* **169**: 2047–2059.
- Dover GA (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111–117.
- Fernández R, Barragán MJL, Bullejos M, Marchal JA, Martínez S, de la Guardia RD et al. (2001). Molecular and cytogenetic characterization of highly repeated DNA sequences in the vole *Microtus cabreræ*. *Heredity* **87**: 637–646.
- Franco FF, Sene FM, Manfrin MH (2008). Molecular characterization of SSS139, a new satellite DNA family in sibling species of the *Drosophila buzzatii* cluster. *Genet Mol Biol* **31**: 155–159.
- Gaff C, du Sart D, Kalitsis P, Iannello R, Nagy A, Choo KHA (1994). A novel nuclear protein binds centromeric alpha satellite DNA. *Hum Mol Genet* **3**: 711–716.
- Gregory TR, Johnston JS (2008). Genome size diversity in the family Drosophilidae. *Heredity* **101**: 228–238.
- Hall SE, Luo S, Hall AE, Preuss D (2005). Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. *Genetics* **170**: 1913–1927.
- Heslop-Harrison JS, Brandes A, Schwarzacher T (2003). Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species. *Chromosome Res* **11**: 241–253.
- Kimura M (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111–120.
- Krzywinski J, Sangaré D, Besansky NJ (2005). Satellite DNA from the Y chromosome of the malaria vector *Anopheles gambiae*. *Genetics* **169**: 185–196.
- Kuhn GCS, Franco FF, Manfrin MH, Moreira-Filho O, Sene FM (2007). Low rates of homogenization of the DBC-150 satellite DNA family restricted to a single pair of microchromosomes in species from the *Drosophila buzzatii* cluster. *Chromosome Res* **15**: 457–469.
- Kuhn GCS, Sene FM (2005). Evolutionary turnover of two pBuM satellite DNA subfamilies in the *Drosophila buzzatii* cluster (*repleta* group): from *alpha* to *alpha/beta* arrays. *Gene* **349**: 77–85.
- Kuhn GCS, Sene FM, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS (2008). Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res* **16**: 307–324.
- Lee HR, Neumann P, Macas J, Jiang J (2006). Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Mol Biol Evol* **23**: 2505–2520.
- Lohe AR, Hilliker AJ, Roberts PA (1993). Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* **134**: 1149–1174.
- Manfrin MH, Sene FM (2006). Cactophilic *Drosophila* in South America: a model for evolutionary studies. *Genetica* **126**: 1–19.
- Mantovani B, Tinti F, Bachmann L, Scali V (1997). The Bag320 satellite DNA family in *Bacillus* stick insects (Phasmatodea): different rates of molecular evolution of highly repetitive DNA in bisexual and parthenogenetic taxa. *Mol Biol Evol* **14**: 1197–1205.
- McAllister BF, Werren JH (1999). Evolution of tandemly repeated sequences: what happens at the end of an array? *J Mol Evol* **48**: 469–481.
- Mravincac B, Plohl M (2007). Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. *Gene* **394**: 45–52.
- Müller AE, Kamisugi Y, Grüneberg R, Niedenhof I, Hörold RJ, Meyer P (1999). Palindromic sequences and A+T-rich DNA elements promote illegitimate recombination in *Nicotiana tabacum*. *J Mol Biol* **291**: 29–46.
- Okumura K, Kiyama R, Oishi M (1987). Sequence analyses of extrachromosomal Sau3A and related family DNA: analysis of recombination in the excision event. *Nucleic Acids Res* **15**: 7477–7489.
- Opperman R, Emmanuel E, Levy AA (2004). The effect of sequence divergence on recombination between direct repeats in *Arabidopsis*. *Genetics* **168**: 2207–2215.
- Palomeque T, Lorite P (2008). Satellite DNA in insects: a review. *Heredity* **100**: 564–573.
- Pons J, Bruvo B, Petitpierre E, Plohl M, Ugarkovic D, Juan C (2004). Complex satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): structural features and evolution. *Heredity* **92**: 418–427.
- Pons J, Juan C, Petitpierre E (2002). Higher-order organization and compartmentalization of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. *Chromosome Res* **10**: 597–606.
- Powell JR (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press: NY.
- Reed KM, Beukeboom LW, Eickbush DG, Werren JH (1994). Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. *J Mol Evol* **38**: 352–362.
- Riddle NC, Elgin SCR (2006). The dot chromosome of *Drosophila*: insights into chromatin states and their change over evolutionary time. *Chromosome Res* **14**: 405–416.
- Rubnitz J, Subramani S (1984). The minimum amount of homology required for homologous recombination in mammalian cells. *Mol Cell Biol* **4**: 2253–2258.
- Rudd MK, Wray GA, Willard HF (2006). The evolutionary dynamics of alpha-satellite. *Genome Res* **16**: 88–96.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Schueler MG, Higgins AW, Rudd KM, Gustashaw K, Willard HF (2001). Genomic and genetic definition of a functional human centromere. *Science* **294**: 109–115.
- Shiels C, Coutelle C, Huxley C (1997). Contiguous arrays of satellites 1, 3, and b form a 1.5-Mb domain on chromosome 22p. *Genomics* **44**: 35–44.
- Smith GP (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528–535.
- Szary A, Sarasin A (1992). Molecular analysis of DNA junctions produced by illegitimate recombination in human cells. *Nucleic Acids Res* **20**: 4269–4274.
- Sun X, Le HD, Wahlstrom JM, Karpen GH (2003). Sequence analysis of a functional *Drosophila* centromere. *Genome Res* **13**: 182–194.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTALW: improving the sensitivity of progressive sequence alignment

- through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Trinh TQ, Sinden RR (1993). The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* **134**: 409–422.
- Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles—causes and effects. *EMBO J* **21**: 5955–5959.
- Vershinin A, Schwarzacher T, Heslop-Harrison JS (1995). The large scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. *Plant Cell* **7**: 1823–1833.
- Wang JC (1985)). DNA topoisomerases. *Annu Rev Biochem* **54**: 665–697.
- Waye JS, England SB, Willard HF (1987). Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome. *Mol Cell Biol* **7**: 349–356.
- Zinic SD, Ugarkovic D, Cornudella L, Plohl M (2000). A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. *Chromosome Res* **8**: 201–212.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)