Chromosoma (2009) 118:349–360
DOI 10.1007/s00412-008-0201-5

RESEARCH ARTICLE

Cloning and sequencing of the breakpoint regions of inversion *5g* fixed in *Drosophila buzzatii*

Olivia Prazeres da Costa · Josefa González · Alfredo Ruiz

Received: 7 November 2008 / Revised: 22 December 2008 / Accepted: 22 December 2008 / Published online: 7 February 2009

Abstract Chromosomal inversions are ubiquitous in *Drosophila* both as intraspecific polymorphisms and interspecific differences. Many gaps still remain in our understanding of the mechanisms that generate them. Previous work has shown that in *Drosophila buzzatii*, three polymorphic inversions were generated by ectopic recombination between copies of the transposon *Galileo*. In this study, we have characterized the breakpoint regions of inversion *5g*, fixed in *D. buzzatii* and absent in *Drosophila koepferae* and other closely related species. A novel approach comprising four experimental steps was used. First, *D. buzzatii* BAC clones encompassing the breakpoints were identified and their ends sequenced. Then, breakpoint regions were mapped at high resolution in the *Drosophila mojavensis* genome sequence. Finally, breakpoint regions were isolated by polymerase chain reaction in *D. buzzatii* and *D. koepferae* and sequenced. Our aim was to shed light on the mechanism that generated inversion *5g* and specifically to test for an implication of the transposon *Galileo*. No evidence implicates *Galileo* or other transposable elements in the origin of inversion *5g* that was generated most likely by two independent breaks and nonhomologous end-joining repair. Our results show that different inversion-generating mechanisms may coexist within the same lineage and suggest a hypothesis for the evolutionary time and mode of their operation.

Josefa González and Alfredo Ruiz contributed equally.

Communicated by B. Calvi

Electronic supplementary material The online version of this article (doi:10.1007/s00412-008-0201-5) contains supplementary material, which is available to authorized users.

O. Prazeres da Costa · J. González · A. Ruiz
Departament de Genètica i de Microbiologia,
Facultat de Biociències, Universitat Autònoma de Barcelona,
08193 Bellaterra, Barcelona, Spain

Present address:
J. González (✉)
Department of Biology, Stanford University,
Stanford, CA 94305, USA
e-mail: jgonzalp@stanford.edu

Introduction

Gross rearrangements are large-scale changes in chromosome structure that can be found as polymorphisms within species or as fixed differences between species. The occurrence of rearrangements in evolution has been known for a long time and, using cytological methods, the karyotypic evolution of many groups of plants and animals has been documented (Stebbins 1971; White 1973). The interest in chromosomal evolution has revived in recent years thanks to physical mapping and whole genome sequencing projects that allow us to compare the genomes of different species with an unprecedented resolution. The results of such comparative genomic analyses have shown an unexpected high rate of rearrangement fixation in many lineages and have demonstrated the remarkable flexibility of the eukaryotic genome (Ranz et al. 2001; Coghlan and Wolfe 2002; Eichler and Sankoff 2003; Coghlan et al. 2005; Nakatani et al. 2007; Bhutkar et al. 2008). How chromosomal rearrangements are generated and what are their functional consequences are long-standing but still controversial questions (Casals and Navarro 2007; Hurles et al. 2008).

Naturally occurring chromosomal inversions were first detected by Sturtevant (1917) as crossover suppressors in strains of *Drosophila melanogaster*. They were later found to be ubiquitous in *Drosophila* both as intraspecific poly-

morphisms (Sperlich and Pfriem 1986) as well as interspecific differences (Stone 1962; Powell 1997; Bhutkar et al. 2008). It is generally assumed that the origin and fixation of an inversion is a unique process and that two species bearing the same inversion share a common ancestor (Krimbas and Powell 1992; Wasserman 1992). Relying on this assumption, detailed inversion phylogenies have been elaborated in many species groups. The phylogeny produced by Wasserman (1992) for the *Drosophila repleta* species group comprises 70 species and includes nearly 300 inversions. Drosophila chromosomal elements often show contrasting patterns of chromosomal evolution (González et al. 2002). Most of the inversions (70.2%) in the repleta group are located on the dynamic chromosome 2 (that represents 23% of the euchromatin), whereas other chromosomes, such as chromosome 5 (containing 20.3% of the euchromatin), are remarkably conservative (6.4% of all inversions; Wasserman 1992).

The origin of *Drosophila* polymorphic inversions has been investigated in detail in a limited number of cases by isolating and sequencing the inversion breakpoint regions (see Ranz et al. 2007 for a review). Unequivocal evidence for the implication of transposable elements (TEs) has been found in three *Drosophila buzzatii* inversions: *2j* (Cáceres et al. 1999, 2001), $2q^7$ (Casals et al. 2003), and $2z^3$ (Delprat et al., in preparation). These three inversions were generated by non-allelic homologus recombination (or ectopic recombination) between copies of the transposon *Galileo* (Marzo et al. 2008) inserted in opposite orientation at two distant chromosomal sites. A polymorphic inversion of *Drosophila pseudoobscura*, *Arrowhead*, was also generated by ectopic recombination between 128 and 315-bp repeats, yet the nature of these repeats and their possible relation to an unidentified TE are obscure (Richards et al. 2005). TE copies have also been found at the breakpoints of two *Anopheles gambiae* inversions: *2Rd'* and *2La*. However, the implication of the TEs in the origin of these inversions is ambiguous (Mathiopoulos et al. 1998; Sharakhov et al. 2006). Another *A. gambiae* inversion, *2Rj*, seemingly arose by ectopic recombination between segmental duplications without the involvement of TEs (Coulibaly et al. 2007). Finally, no TEs or repeats of any kind are seemingly involved in the origin of three *D. melanogaster* inversions, *In(3L) Payne* (Wesley and Eanes 1994), *In(2L)t* (Andolfatto and Kreitman 2000), and *In(3R)Payne* (Matzkin et al. 2005). These inversions might have been generated by a mechanism of chromosomal breakage and repair by non-homologous end-joining (NHEJ; Ranz et al. 2007).

A number of inversions fixed between *Drosophila* species have also been investigated trying to elucidate how these rearrangements were generated (Cirera et al. 1995; Bergman et al. 2002; Richards et al. 2005; Ranz et al. 2007; Cirulli and Noor 2007; Runcie and Noor 2009; Bhutkar et al. 2008). Most of these studies did not detect any TEs at the inversion breakpoint regions, and only in a few cases were inverted repetitive sequences present at both co-occurrent breakpoint regions found (Richards et al. 2005; Ranz et al. 2007). In 18 out of 29 inversions fixed between *D. melanogaster* and *Drosophila yakuba*, breakpoint regions are associated with duplications of genes or other non-repetitive sequences, suggesting that these inversions arose by staggered breaks and NHEJ repair (Ranz et al. 2007). In the most comprehensive comparative analysis carried out so far using the 12 sequenced *Drosophila* genomes, Bhutkar et al. (2008) corroborated the high rate of inversion fixation in this genus but did not observe enrichment for repeat sequences in reused breakpoints. The absence of evidence for TE implication in the generation of fixed inversions contrasts with the results obtained analyzing polymorphic inversions. One explanation is that inversions generated by certain mechanisms are more likely to become fixed than those generated by other mechanisms. A second hypothesis is that breakpoint regions of fixed inversions, which are relatively old compared to polymorphic inversions, have been altered after the generation of the inversion so that the footprints of the generation mechanisms have been wiped out.

In this study, we isolated and sequenced the breakpoints of inversion *5g* fixed in *D. buzzatii*, a member of the repleta group (Wasserman 1992). This inversion is present in *D. buzzatii* but not in any of the closely related species, suggesting a relatively recent origin, approximately four million years ago (Fig. 1). In addition, it is the only inversion fixed in the conservative chromosome 5 of the *D. buzzatii* lineage (Wasserman 1992). We isolated and sequenced the *5g* inversion breakpoint regions in *D. buzzatii* and its closest relative *Drosophila koepferae* that possess the ancestral standard chromosome 5 arrangement (Fig. 1). This was accomplished with the aid of the *D. buzzatii* BAC library (González et al. 2005) and the whole genome sequence of *Drosophila mojavensis*, another

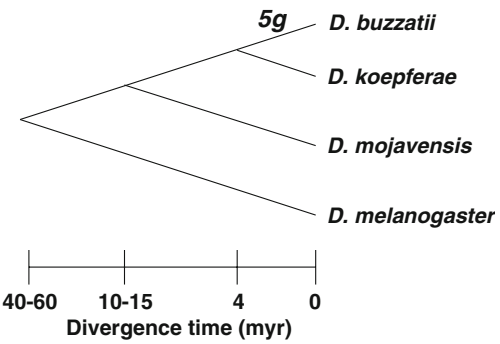

Fig. 1 Phylogenetic relationships among the four species used in this work. Divergence times are taken from Russo et al. (1995), Gomez and Hasson (2003), and Tamura et al. (2004)

member of the repleta group (Drosophila 12 Genomes Consortium 2007). By characterizing a comparatively young fixed inversion, we seek to determine the mechanism of its generation. Specifically, we wanted to test the hypothesis that the transposon *Galileo* generated this inversion very much like it generated the three *D. buzzatii* polymorphic inversions so far analyzed.

Materials and methods

Drosophila stocks

Stocks of three *Drosophila* species were used in this study: *D. mojavensis* (stock 15081-1352.22 from Catalina Island, California), *D. koepferae* (stock KO-2 from Sierra San Luis, Argentina), and *D. buzzatii* (stock st-1; González et al. 2005).

In situ hybridization

BAC clones from the *D. buzzatii* CHORI-225 library (González et al. 2005) and polymerase chain reaction (PCR) products generated in this work were used as probes for in situ hybridization. Polytene chromosome squashes, hybridization, and detection were carried out as previously described by Montgomery et al. (1987) and Ranz et al. (1997). All probes were labeled with biotin-16-dUTP (Roche) by random primed labeling, and detection was carried out with the ABC-Elite Vector Laboratories kit. Heterologous (interspecific) hybridizations were performed at 25°C, and homologous (intraspecific) hybridizations were completed at 37°C. Hybridization signals were localized on the polytene chromosomes using the cytological maps of *D. repleta* (Wharton 1942), *D. buzzatii* (Ruiz and Wasserman 1993; González et al. 2005), and *D. mojavensis* (Schaeffer et al. 2008). The chromosome maps of *D. buzzatii* are cut-and-paste reconstructions of the *D. repleta* maps according to the sequence of inversions proposed for their respective phylogenies (González et al. 2005).

PCR amplification

Primers (Table 1) were designed based on the genome sequence of *D. mojavensis* (Drosophila 12 Genomes Consortium 2007) or in sequencing products of *D. koepferae* or *D. buzzatii*. Genomic DNA or DNA from BAC clones was used as template. PCR amplification was carried out in a total volume of 50 μL, including 100–200 ng of DNA, 20 pM of each primer, 200 μM dNTPs, 1.5 mM $MgCl_2$, and 1 U Taq DNA polymerase. The following conditions were used: 30 cycles of 30 s at 94°C, 30 s at the annealing temperature 50–60°C, and 2–3 min at 72°C for extension. Annealing temperature and extension time depend on the primer pair and on the expected fragment length. Platinum Taq DNA Polymerase (Invitrogen) and Expand Long Template PCR System (Roche) were used to amplify long sequences.

DNA sequencing and sequence analysis

PCR products were cloned into the pGEM-T easy vector (Vector Systems I of Promega, Madison, WI, USA) and sequenced with T7 and SP6 primers. For subcloning, Bluescript II SK (Stratagene) was used as a vector and sequencing was performed using M13 universal forward and reverse primers. PCR products were gel-purified using QIAquick gel extraction kit (Qiagen) and directly sequenced with the same primers used for amplification.

Table 1 Primers used in this work for PCR amplification

Primer name	Sequence (5′-3′)	Primer name	Sequence (5′-3′)	Primer name	Sequence (5′-3′)
1F	ACCGTGTACTTGCAATCCAT	8F	CTGTTGTAGTATCAGCTGGT	A1	ACTATGACGACTACGAGGCG
1R	TGCCGGACTATACGATCTAG	8R	GACGACGATCTGGACAATGA	A2	ATACGCATTGGCATTCACAC
2F	CACGTCGCAGATTCTGTAAC	9F	ATCGCTTCTTCAAGGAGGTC	B1	ACAGAGATGAGCTCGAATCC
2R	CGGCACGGAGTTCATAATTA	9R	TTATACATCGCATGATCGAG	B2	GCATTGATTAGAGCATTGCC
3F	TGATTGGTGTCCGGATACTT	10F	GCCTCATACTGATCGGTGTT	B3	GTCTGCTGATACCTGCCATC
3R	TAATACAATAGCTGCACGCG	10R	CAGTTCATGAATGCTTCGAA	B4	GCCGCCATTGAAGAATATAC
4F	GCATTGATTAGAGCATTGCC	11F	ACATACGGATGCATAGTGAT	B5	CCGCACTATTTGATAGCCAC
4R	GCCGCCATTGAAGAATATAC	11R	CAACAGAGATGGACGACCAC	B6	GTGCATCATGGATGGCTT
5F	GGTGAACGTCCATCTCAAGT	12F	TTCTCCGTGACAATGCATAT	C1	CGAATCTACCTGTCTGCACA
5R	CTTCAAGTTGTCCTCTGGTC	12R	TATTGGCTGTCGGTCTGTCT	C2	ACATACGGATGCATAGTGAT
6F	CGGTGACATTCATCTTCTTG	13F	CAGCTTCAGCTATCCGATTG	C3	CCTAGTTTGCCTTATTCCGT
6R	ATAACTATTTACGCGCCCTA	13R	CAGCGTCTCCACCTACTTCT	C4	TGCTCTTCGGTGCTGTAG
7F	ACACGACGATTGGACAGTTC	14F	GAGGTGATCAATCCATTCTT	D1	ACATACATATGGTGCGTTGA
7R	CAGAGACCACCAGGTAAGCA	14R	CGCACGCACATAAATCTATA	D2	TATTGGCTGTCGGTCTGTCT

Sequencing results were assembled using the CAP3 Sequence Assembly program (Huang and Madan 1999). The inverted and standard arrangement breakpoint sequences were aligned using ClustalW (Chenna et al. 2003) and BLAST 2 sequences (Tatusova and Madden 1999). Annotation of tRNA genes in *D. mojavensis* was taken from the data produced by C. Bergman and D. Ardell using the combined evidence of two de novo tRNA gene prediction methods: tRNAscan-SE and Aragorn (http://www.bioinf.manchester.ac.uk/bergman/data/ncRNA/tRNA/). Sequence comparison to *D. mojavensis* genome was supported by DroSpeGe BLAST and GBrowse (http://rana.lbl.gov/drosophila/mojavensis.html). The breakpoint sequences were analyzed with REPuter (Kurtz et al. 2001), Einverted, Palindrome, Equicktandem (http://bioweb2.pasteur.fr/docs/EMBOSS/), and RepeatMasker (http://www.repeatmasker.org/) to identify repetitive sequences.

Results

The aim of the study was to identify, clone, and sequence the breakpoint sequences of *D. buzzatii* inversion *5g*. For consistency to previous studies (Cáceres et al. 1999; Casals et al. 2003), the breakpoint regions in the standard (non-inverted) arrangement are denoted as AB (distal breakpoint) and CD (proximal breakpoint). In the inverted chromosome, the breakpoint regions become AC (distal breakpoint) and BD (proximal breapoint) because segment BC is inverted (see Fig. 2). To clone and sequence the breakpoint regions, a novel approach comprising four experimental steps was applied: (1) identification of BAC clones encompassing the breakpoints; (2) BAC end sequencing; (3) high-resolution mapping of the breakpoint regions in *D. mojavensis*; and (4) isolation of the breakpoint regions in *D. buzzatii* and in *D. koepferae*. This methodology is potentially applicable to a considerable number of species because the genomes of 12 *Drosophila* species have been sequenced (Drosophila 12 Genomes Consortium 2007) and BAC libraries have been produced for many other Drosophila species (Arizona Genomics Institute: http://www.genome.arizona.edu/; CHORI: http://bacpac.chori.org/).

Identification of BAC clones encompassing the breakpoints

Wasserman (1962) located the breakpoints of inversion *5g* to sections D2/3 and F2 of the polytene map of *D. buzzatii* chromosome 5. We used the BAC-based physical map of the *D. buzzatii* genome (González et al. 2005) to identify BAC clones encompassing the breakpoints. Several BAC clones were selected from contigs 978 and 1052 mapping near chromosomal sections D2/3 and F2, respectively. These BAC clones were hybridized to the polytene chromosomes of *D. buzzatii* and *D. koepferae*. Hybridization of BAC clones 16F24 and 22C13 showed one signal in band D2d on chromosome 5 in *D. buzzatii* and two signals of similar intensity on bands D2d and F3c of chromosome 5 in *D. koepferae* (orthologous to 90C1 and 97C1 in the new *D. mojavensis* map; Fig. 3) Thus, they must contain the distal breakpoint region (AC). Likewise, BAC clones 20J13 and 20J14 hybridized to band F3c of chromosome 5 in *D. buzzatii* and to bands D2d and F3c of chromosome 5 in *D. koepferae* (Fig. 3). Therefore, they must contain the proximal breakpoint region (BD).

BAC end sequencing

We sequenced the ends of the four BAC clones of *D. buzzatii* bearing the inversion breakpoints and mapped the BAC end sequences to the *D. mojavensis* genome sequence. This species is the closest relative of *D. buzzatii* (Fig. 1) among the 12 species whose genome has been fully sequenced (Drosophila 12 Genomes Consortium 2007), and we used it for a high-resolution mapping of the *5g* inversion

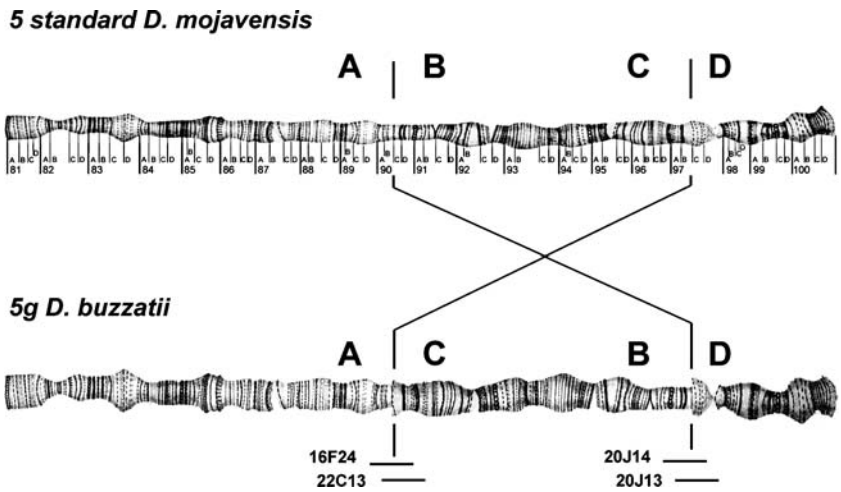

Fig. 2 Polytene chromosome maps of the *5 standard* arrangement of *D. mojavensis* (Schaeffer et al. 2008) (*top*) and the *5g* arrangement of *D. buzzatii* (*bottom*). The distal and proximal breakpoint regions of inversion *5g* are indicated in the standard chromosome (AB and CD, respectively) and in the inverted chromosome (AC and BD, respectively). The *D. buzzatii* BAC clones used for mapping and isolating the breakpoints are shown under the *5g* chromosome

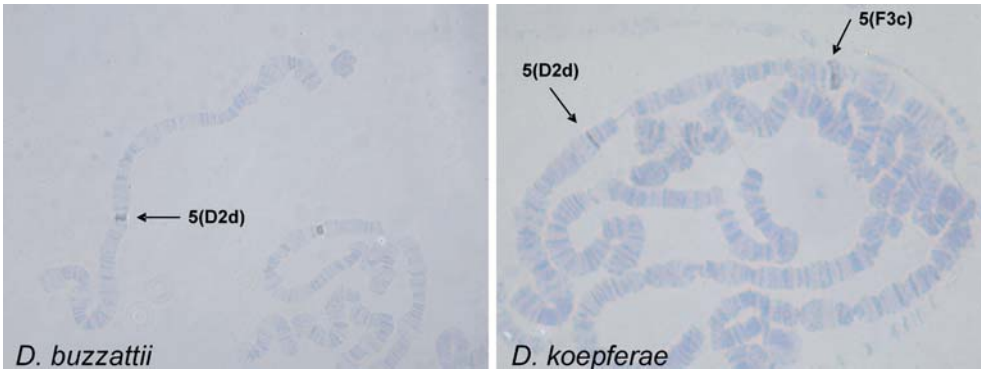

Fig. 3 In situ hybridization results produced with *D. buzzatii* BAC clone 22C13 in *D. buzzatii* (*left*) and *D. koepferae* (*right*)

breakpoints. Only four reliable BAC end sequences were obtained (Table 2). These sequences were mapped to *D. mojavensis* scaffold 6496, corresponding to chromosome 5 (Schaeffer et al. 2008). Using the highest scoring matches, we identified the distal and proximal breakpoint regions as those downstream of BAC end 16F24-T7 and upstream of BAC end 20J14-SP6, respectively (Table 2).

High-resolution mapping of the breakpoint regions in *D. mojavensis*

Once the *5g* breakpoint regions were identified in scaffold 6496 of the *D. mojavensis* genome, which represents the standard (non-inverted arrangement), we walked along this scaffold to locate the breakpoints precisely within these regions. PCR probes were designed step by step further in the direction of the breakpoint, and these probes were hybridized to the chromosomes of *D. buzzatii* (Table 3). Hybridization to the inverted chromosome 5 of *D. buzzatii* allowed us to assign each gene to region A, B, C, or D. The distal breakpoint (AB) was localized ~78 kb downstream of BAC end 16F24-T7, between genes *Sox15* (*CG8404*) (A) and *CG8394* (B). The proximal breakpoint (CD) was localized ~40 kb upstream of BAC end 20J14-SP6 between genes *CG30081* (C) and *CG15121* (D).

The intergenic region between *Sox15* and *CG8394* is short enough for a PCR-based isolation of the breakpoints (see below). However, the intergenic region between *CG30081* and *CG15121* is relatively long (~7.5 kb). In order to pinpoint the CD breakpoint within this region, a methodology based on PCR with primers anchored in conserved sequences was applied. PCR probe 11 (*CG30081*) could be amplified using DNA from BAC clones 16F24 and 22C13 (AC region in *D. buzzatii*) as template, but not with that of BAC clones 20J13 and 20J14 (BD region in *D. buzzatii*). Thus, this PCR probe was assigned to region C. The opposite amplification pattern was observed for PCR probe 12: Amplification was successful with clones 20J14 and 20J13 (BD), but not with clones 22C13 and 16F24 (AC). This probe was therefore assigned to region D. Hence, the location of the proximal breakpoint was narrowed down to a 3.5-kb region (Table 3).

Isolation of the breakpoint regions in *D. buzzatii* and in *D. koepferae*

The breakpoint regions in the inverted *5g* chromosome of *D. buzzatii* (AC and BD) were isolated by PCR. Primers (Table 1) were designed using the breakpoint sequences of *D. mojavensis* scaffold 6496 and combined in the appropriate orientation. Because nucleotide substitutions might be present between *D. mojavensis* and *D. buzzatii* sequences, we tried several primer sequences and combinations to successfully isolate the *D. buzzatii* breakpoints (see Fig. 4).

For PCR amplification of the distal breakpoint in *D. buzzatii* (AC), primers A2 and C2, designed based on the *D. mojavensis* genome sequence, were combined and DNA of BAC clone 16F24 encompassing this breakpoint was used as template. The PCR product was cloned and sequenced (3,662-bp; GenBank accession number FJ534379).

The first attempt to amplify the proximal breakpoint in *D. buzzatii* (BD) by PCR was unsuccessful. Primers could not be anchored in conserved gene regions because the

Table 2 Results of BLAST searches carried out with *D. buzzatii* BAC end sequences against the *D. mojavensis* genome

BAC end	Best hit coordinates	Identity (bp)	Strand	Score	*E* value
20J14-T7	11405327–11405179	146/149	+/–	272	2e–71
20J14-SP6	20736095–20735861	229/235	+/–	418	e–116
16F24-T7	11172225–11172591	338/367	+/+	498	e–139
22C13-T7	11196915–11197033	106/119	+/+	127	5e–28

All coordinates correspond to scaffold_6496 (chromosome 5; Schaeffer et al. 2008)

Table 3 Mapping of PCR probes to pinpoint the *5g* inversion breakpoints on the *D. mojavensis* genome

Probe	*D. mojavensis* Scaffold_6496 coordinates	Size (bp)	Gene	DNA template	Hybridization to *D. buzzatii* chromosome 5
1F-1R	11203487–11202606	881	*CG8485*	BAC 22C13	D2d
2F-2R	11236627–11239372	2,745	*Mtor* (*CG8274*)	BAC 22C13	D2d
3F-3R	11253151–11249832	3,319	*Sox15* (*CG8404*)	*D. mojavensis*	D2d
4F-4R	11257111–11255484	1,627	*CG8394*	*D. mojavensis*	F3c
5F-5R	11263639–11261171	2,468	*CG30069*	*D. mojavensis*	F3c
6F-6R	20634256–20631400	2,856	*CG9313*	*D. mojavensis*	D2d
7F-7R	20643385–20642095	1,290	*CG3216*	*D. mojavensis*	D2d
8F-8R	20664492–20662325	2,167	*Pros29* (*CG9327*)	*D. mojavensis*	D2d
9F-9R	20677506–20674591	2,915	*CG9346-CG31232*	*D. mojavensis*	D2d
10F-10R	20686635–20683504	3,131	*CG30296*	*D. mojavensis*	D2d
11F-11R	20693138–20693824	686	*CG30081*	BAC 22C13	ND
12F-12R	20696142–20696603	461	*CG30081-CG15121*	BAC 20J14	ND
13F-13R	20707363–20704519	2,844	*CG15122*	*D. mojavensis*	F3c
14F-14R	20735970–20733302	2,668	*Smooth* (*CG9218*)	*D. mojavensis*	F3c

ND not determined

breakpoint falls in a relatively large intergenic region which is likely to be only partially conserved. Therefore, we amplified region B in *D. koepferae* with primers B1 and B2 designed in gene *CG8394* of *D. mojavensis* and the 1,354-bp product was sequenced. Then, a new primer (B3) was designed in region B of *D. koepferae*. We expected this primer to work in *D. buzzatii* because *D. buzzatii* and *D. koepferae* are close relatives (see Fig. 1) and the probability of nucleotide changes in the primer sequence is lower. We combined this primer with a primer of region D of *D. mojavensis* (D2) to amplify the breakpoint BD in *D. buzzatii*. BAC clone 20J14 of *D. buzzatii* was used as template DNA for PCR amplification. The resulting PCR product was cloned and sequenced (4,831-bp; GenBank accession number FJ534380).

We also successfully sequenced the breakpoint regions in *D. koepferae*. Breakpoint region AB was amplified with primer A1, designed in the *D. mojavensis* genome sequence, and primer B4 designed in region B of *D. buzzatii*. Genomic DNA of *D. koepferae* was used as template and the 2,267-bp product was sequenced. In order to assemble the AB region with the B region previously sequenced in this species, we designed one primer in the AB region, primer B5, and another one in the B region,

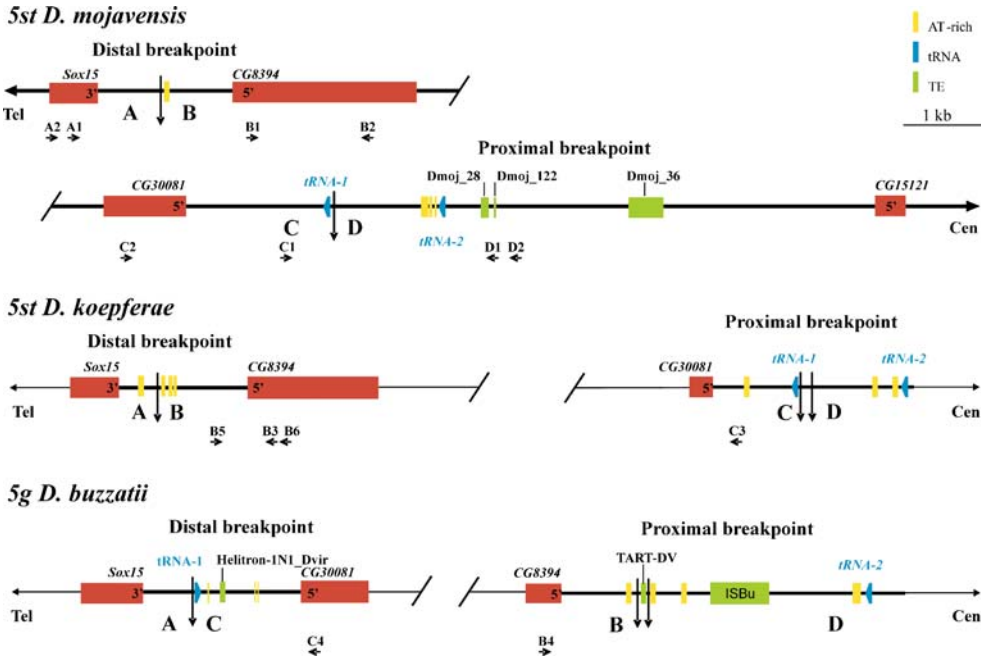

Fig. 4 Sequence organization of the *5g* inversion breakpoint regions in *D. mojavensis* (*top*), *D. koepferae* (*middle*), and *D. buzzatii* (*bottom*). Sequenced intergenic regions are represented by *thick horizontal lines* and sequenced portions of ORFs by *red rectangles* with 5′ or 3′ ends indicating the direction of transcription; tRNA genes are shown in *blue* with an *arrowhead* indicating their orientation; TE insertions are depicted in *green*; AT-rich blocks are indicated by *yellow boxes*. The localization of PCR primers used to amplify the breakpoint regions in *D. buzzatii* and *D. koepferae* is shown by *horizontal arrows* under the sequence used to design them

primer B6. The 465-bp amplification product was sequenced and the three fragments were assembled to produce a 3,948-bp sequence (GenBank accession number FJ534377). Finally, we amplified the proximal breakpoint region in *D. koepferae* (CD). We first amplified a 2,386-bp segment using primers C1 and D1, both of them designed in the *D. mojavensis* genome sequence. We then used this sequence to design another primer, primer C3. Primer C3 was combined with primer C4 designed in region C of *D. buzzatii* to amplify a fragment of 499 bp. Both fragments were assembled to produce a 2,865-bp sequence (GenBank accession number FJ534378).

Breakpoints sequences: annotation and analysis

We identified and isolated the two *5g* breakpoint sequences in three species: *D. mojavensis*, *D. koepferae* (both representing the standard, non-inverted, arrangement), and *D. buzzatii* (bearing the inverted chromosome). These sequences were annotated with the aid of the DroSpeGe Browser for *D. mojavensis* annotation and also by similarity searches using BLAST against the *D. mojavensis* and *D. melanogaster* genomes. Other bioinformatic tools to uncover repeats and TEs were also used (see "Materials and methods"). Figure 4 depicts the molecular organization of the *5g* breakpoint regions in the standard and inverted arrangements.

The distal breakpoint in the non-inverted arrangement (AB) falls in the intergenic region between genes *Sox15* and *CG8394*. The size of this intergenic region (from the *Sox15* STOP codon to the initial ATG codon of *CG8394*) is 1,633 bp in *D. mojavensis* and 1,652 bp in *D. koepferae*. In the latter species, this intergenic region contains four small blocks of AT-rich sequence (190 AT nucleotides out of 199 in total) and a $(CCA)_{11}$ imperfect microsatellite. A small block of AT-rich sequence is also found in the homologous region in *D. mojavensis*.

The proximal breakpoint in the non-inverted arrangement (CD) was located in the intergenic region between genes *CG30081* and *CG15121* which is 7,362 bp long in *D. mojavensis*. Two tRNA genes have been annotated in *D. mojavensis* within this region on the minus (−) strand, D. moj_His_GTG_14000067 and D.moj_His_GTG_14000068 (http://www.bioinf.manchester.ac.uk/bergman/data/ncRNA/tRNA/). Henceforth, we will refer to them as tRNA-1 and tRNA-2, respectively (Fig. 4). In addition, three TE fragments have been annotated using ReAS: Dmoj_28 (105 bp), Dmoj_122 (41 bp), and Dmoj_36 (389 bp). Using probe 12, the breakpoint was further mapped to the region between *CG30081* and tRNA-2. This 2,171-bp region contains immediately downstream of the tRNA-2 gene three small blocks of AT-rich sequence (totaling 156 bp with 146 AT nucleotides).

In *D. koepferae*, we sequenced 2,865 bp from the CD breakpoint. This sequence includes the beginning of *CG30081* (306 bp) and two tRNA genes putatively orthologous to those present in the homologous region of *D. mojavensis* and located in the same orientation (− strand). The intergenic region between *CG30081* and tRNA-2 is 2,477 bp long and contains three AT-rich blocks of sequence, one of them 452 bp downstream of tRNA-1 and the other two immediately downstream of tRNA-2 (Fig. 4).

We sequenced 3,662 bp from the AC breakpoint of *D. buzzatii*. This sequence contains the end of *Sox15* coding region (positions 1–789) and the beginning of *CG30081* (position 2,800–3,662). A his-tRNA, presumably orthologous to tRNA-1 in the *D. koepferae* CD sequence, is also present in the + strand (position 1,372–1,487). Alignment of *D. buzzatii* AC and *D. koepferae* AB sequences showed significant similarity reaching position 1,349. This observation locates the breakpoint in a 22-bp segment between positions 1,350 and 1,371 because the similarity with the sequence surrounding the tRNA that presumably belongs to C starts at site 1,372. A 53-bp segment with similarity to Helitron-1N1_Dvir is found 250 bp downstream of the breakpoint (Fig. 4). Two microsatellites, $(CATA)_6$ and $(TATG)_5$, and three small blocks of AT-rich sequence are also found in the *Sox15-CG30081* intergenic region (Fig. 4).

The BD sequence in *D. buzzatii* (4,662-bp) contains the beginning of *CG8394* (positions 1–353) and a his-tRNA gene in the − strand (position 4,238–4,309), presumably orthologous to tRNA-2 in the *D. koepferae* CD sequence. In addition, two TE-related sequences were annotated: a 77-bp fragment of a LINE-like element (TART-DV) and an ISBu2-like element of *D. buzzatii* (positions 2,241–2,990). When the *D. buzzatii* BD sequence was aligned to the *D. koepferae* AB sequence, the similarity extended well beyond the *CG8394* coding sequence until position 1,351. Likewise, alignment with *D. koepferae* CD showed a small block of similarity (positions 1,451–1,607) and a larger one at the end of the BD sequence (positions 3,797–4,313) that includes the his-tRNA. These observations place the breakpoint in a 99-bp segment (positions 1,352–1,450) that includes the LINE-like fragment. The BD breakpoint region also contains two microsatellites, $(TGG)_{16}$ and $(CAA)_{15}$, and four small blocks of AT-rich sequence (Fig. 4).

Analysis of the breakpoint regions in the 12 *Drosophila* species sequenced

We analyzed whether the genes flanking the *5g* inversion breakpoints were syntenic in the 12 *Drosophila* species sequenced. The distal breakpoint is located in the intergenic region between genes *Sox15* and *CG8394* (Fig. 4). These two genes are closely linked in 11 of the 12 *Drosophila* species sequenced according to Flybase (www.flybase.org).

The only apparent exception is *Drosophila simulans* where these two genes have not been annotated. However, a careful inspection of the region suggests that they are indeed present, although a 3-kb-long assembly gap in this genomic region of *D. simulans* obscures their detection. We conclude that the region between *Sox15* and *CG8394* has been conserved across the evolution of the 12 *Drosophila* species sequenced.

The proximal breakpoint of the *5g* inversion is located between two *tRNA* genes that are flanked by genes *CG30081* and *CG15121* (Fig. 4). We have determined the gene order of this region in the 12 *Drosophila* genomes (Electronic supplementary material Fig. S1). In the nine species of the *Sophophora* subgenus, the organization is *CG11007-tRNA-CG15121-CG15122* with the *tRNA* missing in *Drosophila ananassae*. The gene order in *D. mojavensis*, *CG30296-CG30081-tRNA-tRNA-CG15121-CG15122* (Fig. 4; Electronic supplementary material Fig. S1), reveals three alterations in comparison to the organization found in the species of the *Sophophora* subgenus. First, the gene *CG30081* seems to have transposed into the region (in *D. melanogaster CG30081* is nested within *CG8092* in a distant region of chromosome 2R). This transposition is shared by the three species in the *Drosophila* subgenus and thus must be old. Second, in *D. mojavensis*, there are two tRNA genes instead of only one as in all other *Drosophila* genomes (except *D. ananassae*). The new tRNA may have arisen by a relatively recent duplication or transposition event as it is exclusive of *D. mojavensis*. Finally, *D. mojavensis CG30081* is not flanked by *CG11007* but by *CG30296*, indicating the presence of a chromosomal rearrangement breakpoint (Electronic supplementary material Fig. S1).

Discussion

No evidence for the implication of the transposon *Galileo* in the generation of the *5g* inversion

Several TE families have been shown to induce chromosomal rearrangements in laboratory populations of *Drosophila* (Lim and Simmons 1994). Among these families, the *P* element stands out as one of those especially prone to induce rearrangements (Engels and Preston 1984). However, the evidence for an implication of TEs in the origin of natural *Drosophila* inversions is minimal and appears to be restricted to some of the polymorphic inversions still segregating in natural populations (see "Introduction"). So far, no positive evidence for generation by TEs has been obtained for any fixed inversion. Three polymorphic inversions of *D. buzzatii*, *2j*, $2q^7$, and $2z^3$, have been generated by the transposon *Galileo* (Cáceres et al. 1999, 2001; Casals et al. 2003; Delprat et al., in preparation), a relative of the *D. melanogaster P* and *1360* elements recently classified within the *P* superfamily (Marzo et al. 2008). In each case, *Galileo* copies were found at both inversion breakpoint junctions in all chromosomes with the inverted arrangement, and the pattern of target site duplications flanking the inversion indicated that it was generated by an ectopic recombination event. In all three cases, other TE copies were also found inserted in the breakpoint regions within or near the *Galileo* copies. These TEs are secondary colonizers of the breakpoints that are inserted after the generation of the inversion and accumulate in these regions due to the reduction of recombination in the heterokaryotypes. This consistent pattern provides a benchmark for testing the implication of TEs in the origin of other inversions.

Here, we have isolated and sequenced the breakpoints of the fixed paracentric inversion *5g* distinguishing *D. buzzatii* from its close relative *D. koepferae* (Fig. 1). Our aim was to shed light on the mechanism that generated this inversion and specifically to test for an implication of *Galileo*. The results clearly show that this is not the case. No *Galileo* copies or even fragments with similarity to *Galileo* were observed at any of the two breakpoints of inversion *5g*. The possibility that *Galileo* did in fact generate inversion *5g* but the responsible *Galileo* copies were deleted from the inverted chromosomes afterwards seems very unlikely. There is a high intrinsic rate of nonfunctional DNA loss in *Drosophila* compared to mammals (Petrov et al. 1996; Petrov and Hartl 1998; Singh and Petrov 2004), but the half-life of such DNA (i.e., the expected time until 50% of the sequence has been eliminated by deletion) is still 12–14 myr (Petrov et al. 1996, 2000; Petrov 2002). Because the *5g* inversion must be relatively young (<4 myr, Fig. 1), we expected to find at least partial or defective *Galileo* copies in the breakpoints of this inversion if this element was responsible for its generation. Thus, we must reject a role for *Galileo* in the generation of inversion *5g*.

5g inversion was most likely generated by staggered breaks and NHEJ

We then looked for the presence of other repetitive sequences that could have acted as substrates for ectopic recombination. We did find a few TE copies besides *Galileo* inserted in the inversion breakpoints of the *D. buzzatii 5g* chromosome. A small segment (53 bp) with similarity to Helitron-1N1 Dvir (Kapitonov and Jurka 2007a) was found ~250 bp downstream of the distal breakpoint, and a 750-bp segment with similarity to ISBu2 (Cáceres et al. 2001) was observed ~800 bp downstream of the proximal breakpoint. These two elements are Helitrons (Kapitonov and Jurka 2007b; Yang and Barbash 2008), a subclass of DNA transposons that replicate using a rolling-

circle mechanism (Wicker et al. 2007) that are extremely abundant in *Drosophila* species (up to 6000 copies per haploid genome; Yang and Barbash 2008). The two sequences are 86% identical over the aligned region (53 bp) and are inserted in the same orientation. Neither their localization (Fig. 4) nor their orientation (see below) supports a role for these TEs in the generation of inversion *5g*. ISBu copies have been often found inserted in the breakpoint regions of *D. buzzatii* polymorphic inversions as a result of secondary colonization of the breakpoints (Cáceres et al. 2001; Casals et al. 2003; Delprat et al., in preparation). Thus, this seems the most plausible interpretation for the presence of these TEs in the *5g* breakpoint regions. A similar explanation may apply to the small segment (77 bp) found in the proximal breakpoint with similarity to the non-LTR retrotransposon TART-DV (Casacuberta and Pardue 2003). TART along with other non-LTR retrotransposons is a normal constituent of *Drosophila* telomeres (Pardue et al. 2005; Villasante et al. 2007) and so far has never been implicated in the origin of chromosomal rearrangements. Although this insertion seems to be right in the proximal breakpoint junction, no traces of a similar copy were found at the distal breakpoint region, and thus, there is no evidence to implicate this element in the generation of inversion *5g*.

Two highly similar (94% identical) tRNA copies were found in opposite orientation in the *D. buzzatii 5g* breakpoints, one at each breakpoint. At first glance, this observation might suggest that the *5g* inversion was generated by ectopic recombination between these two tRNA copies. tRNA genes have been previously implicated in the origin of chromosomal rearrangements by ectopic recombination in yeast (Szankasi et al. 1986; Kellis et al. 2003). Ectopic recombination requires the presence of homologous sequences in opposite orientation at two sites in the parental chromosome (Petes and Hill 1988). The arrangement of the two tRNA copies in the parental non-inverted chromosome is inconsistent with this hypothesis. In chromosome 5 of both *D. koepferae* and *D. mojavensis*, two tRNA genes are found in the proximal breakpoint region (in the minus strand), but none is observed at the distal breakpoint region (Fig. 4). The presence of one of these tRNA genes in the distal breakpoint of *D. buzzatii* (in the plus strand) indicates that the proximal breakpoint falls right between the two tRNA genes (Fig. 4) and that these genes are not responsible for the generation of the *5g* inversion.

Overall, we did not find evidence for inverted repetitive sequences in the breakpoint regions, suggesting that a mechanism other than ectopic recombination may be responsible for the generation of this inversion. Using genomic sequences, Ranz et al. (2007) analyzed the breakpoint regions of 29 inversions fixed between *D. melanogaster* and *D. yakuba*. They found that 18 of them (~62%) were associated with duplications of genes or intergenic regions at both co-occurrent breakpoint regions. Sequences from both breakpoints were duplicated in six of the inversions, whereas in the remaining 12 inversions, only sequences from one of the two breakpoints were duplicated. They proposed a model of staggered breaks (either isochromatid or chromatid) and repair by NHEJ as the most likely mechanism for inversion generation. The variation in the size of the duplications would be explained by the variable distance between the staggered breaks. Those cases in which sequences from only one of the two breakpoints were duplicated could be caused by staggered breaks in only one of the breakpoints and a single break in the other. This model of breakage (either staggered or not) and NHEJ is, at this point, the most likely hypothesis to explain how inversion *5g* was generated. The absence of duplications of gene or intergenic sequences suggests that either a single break occurred at each breakpoint or the short distance between staggered breaks coupled with subsequent nucleotide evolution made the small duplications undetectable. The susceptibility of DNA to breakage is known to depend on its base composition. AT-rich sequences show an increased probability of breaks, in particular when they are palindromic and thus capable of forming hairpin or cruciform secondary structures (Schwartz et al. 2006; Zhang and Freudenreich 2007; Durkin and Glover 2007; Lukusa and Fryns 2008). We found several AT-rich small blocks of sequence in the breakpoints of the *5g* inversion (Fig. 4), and it is possible that these AT-rich blocks could have enhanced the susceptibility of these breakpoint regions to breakage. In addition, it must be recalled that the *5g* proximal breakpoint seems to coincide with a particularly dynamic region in the *D. mojavensis* genome (Electronic supplementary material Fig. S1). The region contains a translocated gene (*CG30081*), a recently duplicated or translocated tRNA gene and a rearrangement breakpoint besides the *5g* breakpoint. Thus, this chromosomal region may be considered as "fragile."

Coexistence of different inversion-generating mechanisms within the same lineage and its implications

The variety of molecular mechanisms for the generation of *Drosophila* inversions in nature is striking and raises questions about their relative contribution and the evolutionary time and mode of their operation. The apparent differences in the responsible mechanism between polymorphic and fixed inversions (see "Introduction") are intriguing, although it is probably too soon to draw any firm conclusion. The results presented here on the fixed *5g* inversion most likely produced by breakage and NHEJ repair and previous results on the three polymorphic *D.*

buzzatii inversions generated by the transposon *Galileo* show that different mechanisms can operate within a single lineage. Why such a contrast between fixed and polymorphic inversions? One hypothesis is that inversions generated by breakage and NHEJ repair have a higher probability of fixation than those generated by ectopic recombination. We consider this explanation unlikely, although more information is needed to reject it. The fact that the three polymorphic inversions occurred on the dynamic chromosome 2 whereas the inversion analyzed here occurred in the more conservative chromosome 5 suggests another more likely hypothesis. We can explain these observations by assuming that the breakage and NHEJ repair mechanism generates inversions with a basal or background rate in all lineages and at all times. Double-strand breaks are produced in several ways in all cells, and the machinery necessary to deal with these lesions is conserved from yeasts to vertebrates (Pastink et al. 2001; Sonoda et al. 2006). This mechanism would explain most of the inversions present in the repleta group in conservative chromosomes, e.g., inversion *5g* analyzed here or inversion *Xe* fixed in *D. mojavensis* (Cirulli and Noor 2007; Runcie and Noor 2009). On the other hand, TE activity would explain, for instance by means of the ectopic recombination mechanism, a local or temporary increase in the rate of inversion occurrence. This mechanism would be responsible for the generation of most of the chromosome 2 inversions in the repleta group, including the three polymorphic inversions of *D. buzzatii*. TE activity is likely to vary considerably between lineages and between times and it may also vary between chromosomes due to the accumulation of TE copies in the inverted segments of polymorphic chromosomes as has been observed for *Galileo* (Casals et al. 2005) and another six transposon families (Casals et al. 2006) in *D. buzzatii*. This hypothesis may be tested by characterizing the breakpoints of fixed inversions in different chromosomes in *D. buzzatii* and other species.

Functional consequences of the *5g* inversion

Inversions are considered to play a role in the adaptation of species to their environments and in reproductive isolation between species (Noor et al. 2001; Coghlan et al. 2005; Hoffmann and Rieseberg 2008). However, the molecular mechanisms by which inversions could affect fitness are still unclear. One possibility is that the localization of the inversion breakpoints near or inside genes could affect their function or expression profile. The analysis of the breakpoint regions of inversion *5g* showed that both breakpoints are located in intergenic regions and therefore do not disrupt the coding region of any of the flanking genes. The same scenario was found when the breakpoint regions of the other three inversions sequenced in *D. buzzatii* were analyzed (Cáceres et al. 1999; 2001; Casals et al. 2003; Delprat et al., in preparation). For the proximal breakpoint of one of them, inversion *2j*, it was shown that the expression level of the gene located immediately outside the inversion was reduced in strains carrying the inversion (Puig et al. 1994). This silencing effect was not caused by the inversion itself, but by one of the TEs inserted at the breakpoint junctions. This particular kind of position effect is not likely to be acting in the case of *5g* inversion, since no TEs were found close to the inversion breakpoints. However, the *5g* inversion may still be affecting the expression of the neighboring genes, for example by disrupting or changing the location of *cis*-regulatory elements. The availability of the sequence of inversion breakpoint regions, as described in this paper, will allow the study of the position effects of natural inversions which was previously hindered by the lack of molecular studies.

Acknowledgments We thank Oriol Calvete, Alejandra Delprat, Barbara Negre and Marta Puig for technical support and comments on a previous version of the manuscript. Dmitri Petrov lent us generously his lab to carry out the final part of this project. Work supported by a PIF fellowship from the UAB awarded to O. P. da Costa and grant BFU2005-02237 from the Dirección General de Investigación (Ministerio de Educación y Ciencia, Spain) awarded to A. Ruiz.

References

Andolfatto P, Kreitman M (2000) Molecular variation at the In(2L)t proximal breakpoint site in natural populations of *Drosophila melanogaster* and *D. simulans*. Genetics 154:1681–1691

Bergman CM, Pfeiffer BD, Rincon-Limas DE et al (2002) Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome. Genome Biol 3:Research0086–6

Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF, Gelbart WM (2008) Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. Genetics 179:1657–1680

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

Cáceres M, Puig M, Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. Genome Res 11:1353–1364

Casacuberta E, Pardue ML (2003) Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the virilis group. Proc Natl Acad Sci U S A 100:3363–3368

Casals F, Navarro A (2007) Chromosomal evolution: inversions: the chicken or the egg? Heredity 99:479–480

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

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

Casals F, González J, Ruiz A (2006) Abundance and chromosomal distribution of six *Drosophila buzzatii* transposons: BuT1, BuT2, BuT3, BuT4, BuT5, and BuT6. Chromosoma 115:403–412

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 31:3497–3500

Cirera S, Martin-Campos JM, Segarra C, Aguade M (1995) Molecular characterization of the breakpoints of an inversion fixed between *Drosophila melanogaster* and *Drosophila subobscura*. Genetics 139:321–326

Cirulli ET, Noor MA (2007) Localization and characterization of X chromosome inversion breakpoints separating *Drosophila mojavensis* and *Drosophila arizonae*. J Heredity 98:111–114

Coghlan A, Wolfe KH (2002) Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. Genome Res 16:857–867

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet 21:673–682

Coulibaly MB, Lobo NF, Fitzpatrick MC et al (2007) Segmental duplication implicated in the genesis of inversion 2Rj of *Anopheles gambiae*. PLoS ONE 2:e849

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203–218

Durkin SG, Glover TW (2007) Chromosome fragile sites. Annu Rev Genet 41:169–192

Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosome evolution. Science 301:793–797

Engels WR, Preston CR (1984) Formation of chromosome rearrangements by *P* factors in *Drosophila*. Genetics 107:657–678

Gómez GA, Hasson E (2003) Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii*. Mol Biol Evol 20:410–423

González J, Ranz JM, Ruiz A (2002) Chromosomal elements evolve at different rates in the *Drosophila* genome. Genetics 161:1137–1154

González J, Nefedov M, Bosdet I et al (2005) A BAC-based physical map of the *Drosophila buzzatii* genome. Genome Res 15:885–892

Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst 39:21–42

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868–877

Hurles ME, Dermitzakis ET, Tyler-Smith C (2008) The functional impact of structural variation in humans. Trends Genet 24:238–245

Kapitonov VV, Jurka J (2007a) Helitrons in fruit flies. Repbase Reports 7:129

Kapitonov VV, Jurka J (2007b) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23:521–529

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254

Krimbas CB, Powell JR (1992) Drosophila inversion polymorphism. CRC, Boca Raton

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642

Lim JK, Simmons MJ (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. Bioessays 16:269–275

Lukusa T, Fryns JP (2008) Human chromosome fragility. Biochim Biophys Acta 1779:3–16

Marzo M, Puig M, Ruiz A (2008) The foldback-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the Drosophila genus. Proc Natl Acad Sci U S A 105:2957–2962

Mathiopoulos KD, della Torre A, Predazzi V, Petrarca V, Coluzzi M (1998) Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction. Proc Natl Acad Sci U S A 95:12444–12449

Matzkin LM, Merritt TJ, Zhu CT, Eanes WF (2005) The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion In(3R)Payne in *Drosophila melanogaster*. Genetics 170:1143–1152

Montgomery E, Charlesworth B, Langley CH (1987) A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet Res 49:31–41

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17:1254–1265

Noor MA, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci U S A 98:12084–12088

Pardue ML, Rashkova S, Casacuberta E, DeBaryshe PG, George JA, Traverse KL (2005) Two retrotransposons maintain telomeres in *Drosophila*. Chromosome Res 13:443–453

Pastink A, Eeken JC, Lohman PH (2001) Genomic integrity and the repair of double-strand DNA breaks. Mutat Res 480–481:37–50

Petes TD, Hill CW (1988) Recombination between repeated genes in microorganisms. Annu Rev Genet 22:147–168

Petrov DA (2002) DNA loss and evolution of genome size in *Drosophila*. Genetica 115:81–91

Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol 15:293–302

Petrov DA, Lozovskaya ER, Hartl DL (1996) High intrinsic rate of DNA loss in *Drosophila*. Nature 384:346–349

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. Science 287:1060–1062

Powell JR (1997) Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press, Oxford

Puig M, Caceres M, Ruiz A (1994) Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. Proc Natl Acad Sci U S A 101 (24):9013–9018

Ranz JM, Segarra C, Ruiz A (1997) Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* species group. Genetics 145:281–295

Ranz JM, Casals F, Ruiz A (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. Genome Res 11:230–239

Ranz JM, Maurin D, Chan YS et al (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol 5:e152

Richards S, Liu Y, Bettencourt BR et al (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. Genome Res 15:1–18

Ruiz A, Wasserman M (1993) Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. Heredity 70:582–596

Runcie DE, Noor MA (2009) Sequence signatures of a recent chromosomal rearrangement in *Drosophila mojavensis*. Genetica. Jul 26 (in press)

Russo CA, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of *Drosophilid* species. Mol Biol Evol 12:391–404

Schaeffer SW, Bhutkar A, McAllister BF et al (2008) Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics 179:1601–1655

Schwartz M, Zlotorynski E, Kerem B (2006) The molecular basis of common and rare fragile sites. Cancer Lett 232:13–26

Sharakhov IV, White BJ, Sharakhova MV et al (2006) Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. Proc Natl Acad Sci U S A 103:6258–6262

Singh ND, Petrov DA (2004) Rapid sequence turnover at an intergenic locus in *Drosophila*. Mol Biol Evol 21:670–680

Sonoda E, Hochegger H, Saberi A, Taniguchi Y, Takeda S (2006) Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. DNA Repair 5:1021–1029

Sperlich D, Pfriem P (1986) Chromosomal polymorphism in natural and experimental populations. In: Ashburner M, Carson HL, Thompson JN Jr. (eds) The genetics and biology of *Drosophila*. 3rd rd edn. Academic, NY, pp 257–309

Stebbins GL (1971) Chromosomal evolution in higher plants. Arnold, London

Stone WA (1962) The dominance of natural selection and the reality of superspecies (species groups) in the evolution of *Drosophila*. Univ Texas Publ 6205:507–537

Sturtevant AH (1917) Genetic factors affecting the strength of linkage in *Drosophila*. Proc Natl Acad Sci U S A 3:555–558

Szankasi P, Gysler C, Zehntner U, Leupold U, Kohli J, Munz P (1986) Mitotic recombination between dispersed but related tRNA genes of *Schizosaccharomyces pombe* generates a reciprocal translocation. Mol Gen Genet 202:394–402

Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol Biol Evol 21:36–44

Tatusova TA, Madden TL (1999) BLAST 2 sequences—a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE, de Pablos B (2007) *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. Genome Res 17:1909–1918

Wasserman M (1962) Cytological studies of the repleta group of the genus *Drosophila*: V. The mulleri subgroup. Univ Tex Publ 6205:85–117

Wasserman M (1992) Cytological evolution of the *Drosophila repleta* species group. A: *Drosophila* inversion polymorphism (edited by Krimbas CB and Powell JR). CRC, Boca Raton, FL, pp 455–452

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

Wharton LT (1942) Analysis of the repleta group of *Drosophila*. Univ Texas Pub 4228:23–59

White MJD (1973) Animal cytology and evolution. Cambridge University Press, Cambridge

Wicker T, Sabot F, Hua-Van A (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Yang HP, Barbash DA (2008) Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. Genome Biol 9:R39

Zhang H, Freudenreich CH (2007) An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. Mol Cell 27:367–379