Heredity (2015) 114, 610–618

www.nature.com/hdy

ORIGINAL ARTICLE

A molecular perspective on a complex polymorphic inversion system with cytological evidence of multiply reused breakpoints

DJ Orengo[1], E Puerma[1], M Papaceit, C Segarra and M Aguadé

Genome sequence comparison across the Drosophila genus revealed that some fixed inversion breakpoints had been multiply reused at this long timescale. Cytological studies of Drosophila inversion polymorphism had previously shown that, also at this shorter timescale, some breakpoints had been multiply reused. The paucity of molecularly characterized polymorphic inversion breakpoints has so far precluded contrasting whether cytologically shared breakpoints of these relatively young inversions are actually reused at the molecular level. The E chromosome of *Drosophila subobscura* stands out because it presents several inversion complexes. This is the case of the $E_{1+2+9+3}$ arrangement that originated from the ancestral E_{st} arrangement through the sequential accumulation of four inversions (E_1, E_2, E_9 and E_3) sharing some breakpoints. We recently identified the breakpoints of inversions E_1 and E_2, which allowed establishing reuse at the molecular level of the cytologically shared breakpoint of these inversions. Here, we identified and sequenced the breakpoints of inversions E_9 and E_3, because they share breakpoints at sections 58D and 64C with those of inversions E_1 and E_2. This has allowed establishing that E_9 and E_3 originated through the staggered-break mechanism. Most importantly, sequence comparison has revealed the multiple reuse at the molecular level of the proximal breakpoint (section 58D), which would have been used at least by inversions E_2, E_9 and E_3. In contrast, the distal breakpoint (section 64C) might have been only reused once by inversions E_1 and E_2, because the distal E_3 breakpoint is displaced >70 kb from the other breakpoint limits.
Heredity (2015) **114,** 610–618; doi:10.1038/hdy.2015.4; published online 25 February 2015

INTRODUCTION

Cytological studies as well as comparative studies of the genome sequences from related species revealed that genome reorganization is widespread among taxa (for example, Krimbas and Powell, 1992; Newman *et al*., 2005; Clark *et al*., 2007; Ferguson-Smith and Trifonov, 2007; Schaeffer *et al*., 2008). Structural variation ranging from chromosome fusions to paracentric inversions underlies the detected reorganization. These studies also detected an uneven distribution of rearrangement breakpoints across the genome of both mammals and dipters, with some regions having suffered multiple disruptions (for example, Coluzzi *et al*., 2002; Pevzner and Tesler, 2003). This coincidental occurrence of a rearrangement breakpoint in the same region in two or more species is referred to as breakpoint reuse. In Drosophila, there is both cytological and molecular evidence that the breakpoints of some fixed inversions have been used more than once through evolutionary time (for example, Tonzetich *et al*., 1988; Ruiz and Wasserman, 1993; Ranz *et al*., 2007; Bhutkar *et al*., 2008; von Grotthuss *et al*., 2010). Even at the much shorter intraspecific timescale, a similar observation of some reused breakpoints has emerged from cytological studies of chromosomal inversion polymorphism in different Drosophila species (Kunze-Mühl and Müller, 1958; Aulard *et al*., 2002).

The identification and molecular characterization of inversion breakpoints constitute a starting point not only to address questions about the mechanisms for the origin of inversions but also to contrast whether breakpoints reused at the cytological level are or not reused at the molecular level. Polymorphic inversions offer an advantage over fixed inversions due to the shorter time elapsed since their origin. Indeed, extant sequences of polymorphic inversion breakpoints are expected to better reflect the sequences in the original inverted chromosome. Despite the extensive knowledge accumulated on inversion polymorphism in Drosophila, there are relatively few polymorphic inversion breakpoints molecularly characterized both in any particular species and across species (but see Wesley and Eanes, 1994; Andolfatto *et al*., 1999; Cáceres *et al*., 1999; Casals *et al*., 2003; Matzkin *et al*., 2005; Richards *et al*., 2005; Delprat *et al*., 2009; Corbett-Detig *et al*., 2012; Papaceit *et al*., 2013; Puerma *et al*., 2014). The number of molecularly characterized breakpoints drops even more when inversions with cytological evidence of reused breakpoints are considered (Corbett-Detig *et al*., 2012; Puerma *et al*., 2014). This paucity of molecularly characterized polymorphic inversion breakpoints has so far precluded to extensively contrast whether cytologically shared breakpoints of the relatively young polymorphic inversions are actually reused at the molecular level. It has also precluded ascertaining why disruptions occur repeatedly at multiply reused breakpoints.

Departament de Genètica, Facultat de Biologia, i Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
[1]These authors contributed equally to this work.
Correspondence: Professor M Aguadé, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 643, 08028 Barcelona, Spain.
E-mail: maguade@ub.edu
Received 9 September 2014; revised 12 December 2014; accepted 16 December 2014; published online 25 February 2015

Drosophila subobscura has a rich chromosomal inversion polymorphism in its five major chromosomes. Its E chromosome (Muller's C element) stands out because it presents several inversion complexes. One such case is the $E_{\underline{1+2+9+3}}$ arrangement that originated from the ancestral E_{st} arrangement through the sequential accumulation of four inversions sharing some of their breakpoints (E_1, E_2, E_9 and E_3; Figure 1). These characteristics render this chromosomal element particularly suitable not only to study the mechanisms underlying the origin of inversions through the molecular characterization of their breakpoints, but also to study breakpoint reuse, and more specifically to test whether breakpoints of polymorphic inversions are actually reused, or multiply reused, at the molecular level. Moreover, this characterization may provide new insights into the role played by asynapsis in inversion heterokaryotype breakpoints in the sequential accumulation of polymorphic inversions with shared breakpoints.

We recently characterized through chromosome walking the breakpoints of the two inversions leading from the ancestral arrangement of the E chromosome in the *D. subobscura* subgroup (E_{st}) to the $E_{\underline{1+2}}$ arrangement (Puerma *et al*., 2014). At the cytological level, inversions E_1 and E_2 share one breakpoint. Comparison of the three breakpoint regions in the two arrangements—AB, EF and GH in E_{st}, and AG, FB and EH in $E_{\underline{1+2}}$ (Figure 1)—pointed to each inversion having been generated by a different mechanism: E_1 through staggered double-strand breaks, and E_2 through ectopic recombination (Puerma *et al*., 2014). This study also revealed that the breakpoint reused at the cytological level had also been reused at the molecular level, even though we could not establish the order of occurrence of both inversions and, therefore, which breakpoint had been reused. The breakpoint reused would be the most proximal one corresponding to chromosomal section 58D if E_1 had occurred before E_2, whereas in the other case (E_2 before E_1) it would be the most distal one corresponding to chromosomal section 64C (Figure 1).

Here, we have identified and characterized the breakpoints of inversions E_9 and E_3 that occurred sequentially and generated the $E_{\underline{1+2+9}}$ and $E_{\underline{1+2+9+3}}$ arrangements (Figure 1). As previously indicated, some breakpoints of inversions E_1, E_2, E_9 and E_3 are multiply shared at the cytological level (Figure 1). Indeed, if inversion E_1 had occurred first, the most proximal breakpoint at section 58D would have been shared by all four inversions, whereas the most distal breakpoint at section 64C would have been shared by inversions E_2 and E_3. Alternatively, if inversion E_2 had occurred first, the proximal breakpoint would have been shared by inversions E_2, E_9 and E_3, and the distal breakpoint by inversions E_1, E_2 and E_3. The molecular characterization of the breakpoints of inversions E_9 and E_3 in the corresponding ancestral and derived arrangements will not only allow establishing each

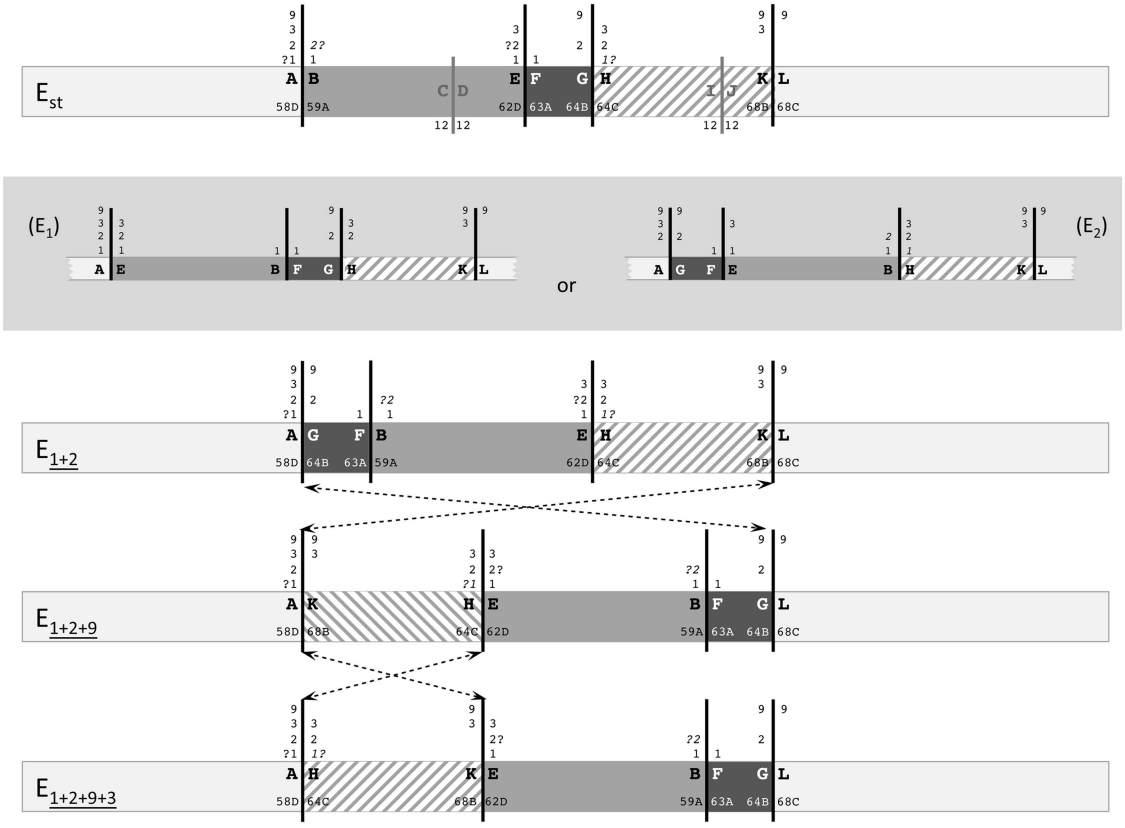

Figure 1 Schematic representation of E chromosomal arrangements of *D. subobscura*. The scheme includes four extant E chromosomal arrangements of *D. subobscura* as well as the two possible intermediate, and now extinct, arrangements connecting E_{st} and $E_{\underline{1+2}}$ (within a gray box). On the E_{st} chromosome scheme, black and gray vertical lines represent the cytological location of the breakpoints of five inversions, with the black lines referring to those of inversions included in the present study (E_1, E_2, E_9 and E_3). The fragments separating these inversion breakpoints are differentiated either by color (gray, dark gray) or texture (striped) to facilitate their identification in the different arrangements. Inversion breakpoints are labeled consecutively with pairs of capital letters (for example, AB, CD and EF) from the most proximal to the most distal breakpoint, as in Puerma *et al*. (2014), with numbers on both sides of each continuous vertical line referring to the inversions delimited by each breakpoint and map sections indicating their location on the Kunze-Mühl and Müller (1958) map. Discontinuous lines connecting two arrangements refer to the region inverted in each case.

inversion generating mechanism but also contrasting breakpoint reuse at the molecular level. Most importantly, it will provide novel molecular information concerning breakpoints that have been multiply reused at least at the cytological level.

MATERIALS AND METHODS

Drosophila strains

Four *D. subobscura* strains were used to identify the breakpoints of inversions E_9 and E_3 and to sequence the breakpoint regions: *ch cu*, OF21, OF82 and FO_12b, which are homokaryotypic for the E_{st}, $E_{\underline{1+2}}$, $E_{\underline{1+2+9}}$ and $E_{\underline{1+2+9+3}}$ chromosomal arrangements, respectively. OF strains were obtained through over 13 generations of sib mating from isofemale lines established upon collection in Observatori Fabra (Barcelona, Catalonia, Spain), as reported in Puerma *et al.* (2014). Strain FO_12b was made homokaryotypic for the $E_{\underline{1+2+9+3}}$ chromosomal arrangement upon 11 generations of sib mating from the segregating F_1 of a cross between a wild-caught male from Observatori Fabra and a strain homokaryotypic for the E_{st} arrangement (*ch cu*).

Identification of breakpoint regions: chromosome walk, *in situ* hybridization and sequencing

The starting point to identify breakpoint KL in E_{st} and $E_{\underline{1+2}}$ chromosomes (Figure 1) was a molecular marker previously located at section 68B of the E chromosome (Laayouni *et al.*, 2007). Our walk was based on both the *D. pseudoobscura* and *D. melanogaster* genome sequences, with oligonucleotides for PCR amplification designed using *D. subobscura* sequences whenever available (Barcelona Subobscura Initiative [BSI]). Labeled probes were *in situ* hybridized on polytene chromosomes of *D. subobscura* (see below), which allowed walking toward the breakpoint and to eventually cross it.

For breakpoint regions AK and GL in $E_{\underline{1+2+9}}$ chromosomes, and breakpoint regions AH and KE in $E_{\underline{1+2+9+3}}$ chromosomes (see Results section for the renaming of these breakpoints), the fragment spanning each breakpoint was PCR amplified using oligonucleotides anchored at its flanking regions (Supplementary Figure S1). Different Taq polymerases (GoTaq DNA polymerase from Promega Corporation, Fitchburg, WI, USA and TaKaRa DNA polymerase from Takara Bio Inc, Otsu, Japan) were used for PCR amplification according to the expected length of the fragment to be amplified.

All steps of the *in situ* hybridization procedure were performed as described in Montgomery *et al.* (1987). Hybridization signals were located on the cytological map of *D. subobscura* (Kunze-Mühl and Müller, 1958) with the standard arrangement for all chromosomes.

Only fragments spanning inversion breakpoints were sequenced upon their amplification by PCR (Supplementary Figure S1), using primer walking whenever necessary. MultiScreen PCR (Merck Millipore, Darmstadt, Germany) was used to purify amplicons before sequencing them with the ABI PRISM version 3.2 cycle sequencing kit (Thermo Fisher Scientific Inc, Waltham, MA, USA), with sequencing products separated on an ABI PRISM 3730 sequencer (Thermo Fisher Scientific Inc). All sequences were obtained on both strands and assembled using the DNASTAR package (Burland, 2000). When sequences could not be obtained directly from PCR products, we used the cloning and sequencing strategy described in Puerma *et al.* (2014).

Sequence analysis

All breakpoint regions were annotated with genes by comparison with the *D. pseudoobscura* genome of FlyBase (http://flybase.org/) using BLAST tools and analyzed to detect repeated motifs using RepeatMasker. The newly sequenced breakpoint regions as well as those of the extended $E_{\underline{1+2}}$ breakpoint regions were compared using the Align Sequences Nucleotide BLAST utility at NCBI webpage to accurately map each breakpoint and to determine putative duplications resulting from the inversion process.

RESULTS

Identification and characterization of inversion E_9 breakpoints

According to available cytological information (Kunze-Mühl and Müller, 1958), inversions E_2 and E_9 (and possibly E_1; see Introduction and Figure 1) share their proximal breakpoint at section 58D. The previously sequenced ~7.1-kb fragment that spans this breakpoint in $E_{\underline{1+2}}$ (AG in Figure 2 in Puerma *et al.*, 2014) was PCR amplified using DNA from strain OF21 ($E_{\underline{1+2}}$) and used as a probe for *in situ* hybridization on $E_{\underline{1+2+9}}$

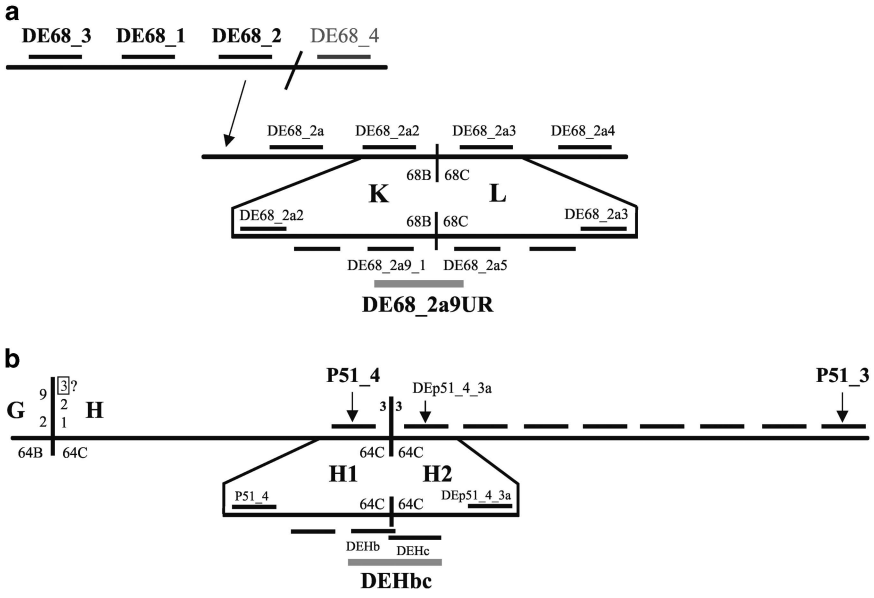

Figure 2 Chromosome walks. Schematic representation of chromosome walks performed on E_{st} (*ch cu*) chromosomes to identify the KL (**a**) and H1H2 (**b**) breakpoint regions (not at scale). Horizontal lines above the long horizontal lines that represent *D. subobscura* chromosomes indicate probes used in the first steps, whereas horizontal lines below the corresponding chromosome indicate those used in the final steps. Only the names of the most informative probes are shown. Probes spanning the breakpoints are highlighted in gray. In each breakpoint region, a thick vertical line represents the breakpoint itself. In the case of the KL breakpoint region (**a**), the inclined line at the upper scheme separates two noncontiguous regions in *D. subobscura*, whereas the arrow indicates the location of the corresponding probe in the lower scheme. In the case of the H1H2 breakpoint region (**b**), the vertical line on the left side of the *D. subobscura* chromosome corresponds to the GH breakpoint that according to cytological information would not only be shared by inversions E_1, E_2 and E_9 but also by inversion E_3.

chromosomes (strain OF82). This probe gave two strong signals at sections 58D next to 68B (breakpoint AK) and 64B next to 68C (breakpoint GL) (Figure 1 and Supplementary Figure S2), confirming that it included the proximal breakpoint of inversion E_9.

To identify the distal breakpoint of inversion E_9 (KL in E_{st} and $E_{\underline{1+2}}$; Figure 1), a chromosome walk was started from a molecular marker previously located at section 68B of the E chromosome (Laayouni *et al.*, 2007). Four probes were designed, two at each side of the marker, on the *D. pseudoobscura* genome sequence: probes DE68_1, DE68_2, DE68_3 and DE68_4 (Figure 2 and Supplementary Figure S3). Probes were amplified using DNA from the *ch cu* (E_{st}) strain and subsequently hybridized on $E_{\underline{1+2+9}}$ chromosomes. Only the first three probes hybridized close to the inversion breakpoint, probes DE68_3 and DE68_1 at section 68A/B and probe DE68_2 at section 68B (Supplementary Table S2), indicating the direction in which to proceed with our walk. Moreover, a break of collinearity between the *D. subobscura* and *D. pseudoobscura* genomes was detected between fragments DE68_2 and DE68_4 (Supplementary Figure S3 and Supplementary Table S1). We designed four new probes based on the *D. melanogaster* genome sequence: DE68_2a, DE68_2a2, DE68_2a3 and DE68_2a4 (Figure 2 and Supplementary Figure S3). When *in situ* hybridized on $E_{\underline{1+2+9}}$ chromosomes, the first two probes gave a single strong signal at section 68B next to 58D, whereas the latter two probes did at section 68C next to 64B (that is, at a certain distance), indicating that probes DE68_2a2 and DE68_2a3 flanked the KL breakpoint (Figures 1 and 2; Supplementary Figures S3 and Supplementary Table S1). Four additional probes were designed in this ~30-kb-long interval and subsequently hybridized on $E_{\underline{1+2+9}}$ chromosomes, which allowed narrowing down the KL breakpoint region to an ~8-kb-long fragment flanked by probes DE68_2a9_1 and DE68_2a5 (Figure 2 and Supplementary Figure S3). A new probe (DE68_2a9UR) was designed anchored at genes *Sox14* and *Phm* (Figure 2 and Supplementary Figure S3) that gave a single signal when *in situ* hybridized on $E_{\underline{1+2}}$ chromosomes, and two distant signals (at sections 68B and 68C) on $E_{\underline{1+2+9}}$ chromosomes both when amplified using DNA from strains *ch cu* (E_{st}) and OF21 ($E_{\underline{1+2}}$) (Supplementary Figure S2).

Upon identification of inversion breakpoint regions on $E_{\underline{1+2}}$ chromosomes (AG and KL in Figure 1) and sequencing of the KL fragment in both E_{st} (*ch cu*) and $E_{\underline{1+2}}$ (OF21) chromosomes, fragments spanning the E_9 breakpoint regions in the $E_{\underline{1+2+9}}$ arrangement (AK and GL in Figure 1) were amplified with the corresponding combination of oligonucleotides (Supplementary Figure S1) and subsequently *in situ* hybridized on both $E_{\underline{1+2}}$ and $E_{\underline{1+2+9}}$ chromosomes. In both cases, the amplified fragments gave one strong signal on $E_{\underline{1+2+9}}$ chromosomes and two strong distant signals on $E_{\underline{1+2}}$ chromosomes at the expected locations, confirming that the breakpoints were included in the amplified fragments (Figure 1 and Supplementary Table S1). Each of these probes gave an additional weak signal at the alternative breakpoint on $E_{\underline{1+2+9}}$ chromosomes (see next paragraph and Supplementary Figure S2). The fragments spanning the breakpoints were completely sequenced in the OF82 strain ($E_{\underline{1+2+9}}$).

Pairwise comparison of the AG, AK and GL breakpoint regions allowed detecting the presence of the A part not only in the AG and AK regions but also in the GL region (Figure 3). Actually, the A

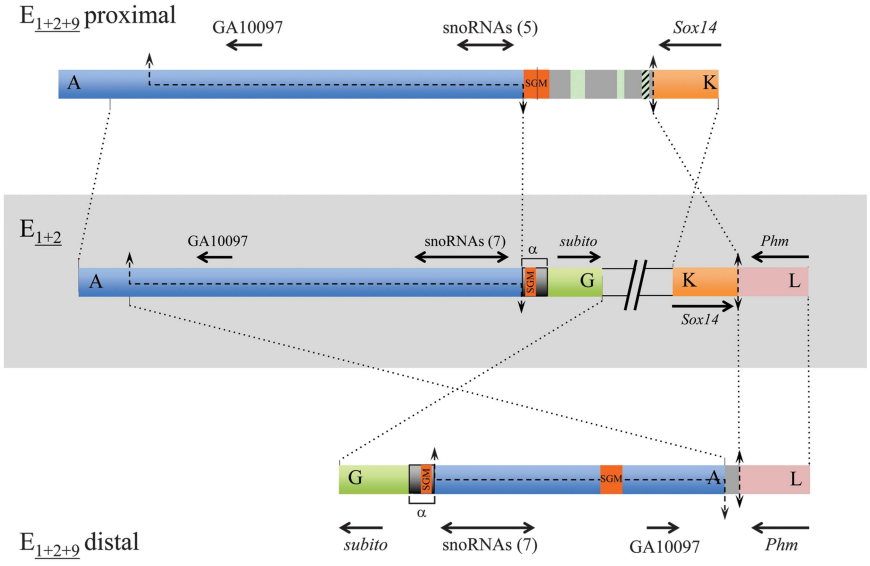

Figure 3 E_9 inversion breakpoints. Schematic representation of the E_9 inversion breakpoint regions on $E_{\underline{1+2}}$ and $E_{\underline{1+2+9}}$ arrangements with breakpoints labeled as in Figure 1. Thick colored bars represent the different breakpoint regions. The central part shows, within a light gray rectangle, the scheme of both breakpoints of the E_9 inversion in gene arrangement $E_{\underline{1+2}}$ (AG and KL), separated by two inclined lines. Schemes of the proximal and distal breakpoints in gene arrangement $E_{\underline{1+2+9}}$ (AK and GL) are represented above and below, respectively, of the $E_{\underline{1+2}}$ scheme. The lengths of the four sequenced breakpoint regions are AG, ~14.4-kb; KL, ~3.8-kb; AK, ~17.0-kb; and GL, ~13.0-kb. Black discontinuous lines along a chromosomal region represent staggered breaks with their limits indicated by arrows. Vertical double-headed arrows indicate cut-and-paste breakpoints. Dotted lines between arrangements indicate the limits and orientation of homologous regions. On each flanking region, the names of the orthologous coding regions in either *D. melanogaster* or *D. pseudoobscura* are given, with black arrows indicating their sense and approximate size. Double-headed horizontal arrows refer to the multiple snoRNAs generated from the corresponding *Uhg* gene introns, with their number given in parenthesis. Orange boxes labeled SGM refer to either near canonical copies of the SGM transposable element, or remnants thereof within a larger repeat sequence named α-motif in Puerma *et al.* (2014) and marked with an α. Gray regions present in the E_9 breakpoints (AK and GL) in the $E_{\underline{1+2+9}}$ arrangement are intervening regions not present in any of the $E_{\underline{1+2}}$ breakpoint regions that in the proximal breakpoint present different-sized fragments of the *gypsy* (green) and *Pao* (striped green) transposable elements.

fragment present in the GL region extended further upstream than that previously sequenced in the E_{st} and $E_{\underline{1+2}}$ arrangements (Puerma *et al.*, 2014), which prompted us to also sequence the extended stretch in the E_{st}, $E_{\underline{1+2}}$ and $E_{\underline{1+2+9}}$ arrangements. In all three cases (Figure 3), the duplicated A fragment (~7.8-kb long) does not only contain the orthologs of the snoRNA genes GA29824 to GA29818 that are encoded in introns of the *Uhg5* gene but also that of the GA10097 gene. The presence of this duplicated region in the AK and GL breakpoints accounts for the weak additional signal detected when probes containing these breakpoints are hybridized on $E_{\underline{1+2+9}}$ chromosomes (Supplementary Figure S2).

Pairwise comparison of the KL (~3.8-kb long), AK and GL breakpoint regions revealed the precise nucleotide at which the KL region had been broken in the inversion process (Figure 3). The K (~1.9-kb long) and L (~2.0-kb long) sequenced fragments contain the *Sox14* gene (partially) and the *Phm* gene (partially), respectively.

In the AK (~18.2-kb long) breakpoint region, there is an intervening stretch (3.6-kb long) between the flanking parts present in the non-inverted breakpoints—AG (~14.4-kb long) and KL; Figure 3—This stretch includes a rather conserved SubobscuraGuancheMadeirensis (SGM) transposable element (Miller *et al.*, 2000) and fragments of other transposable elements (Figure 3). It should be noted that the ~700 bp α-motif present in the G part of the AG breakpoint region—and also in that of the GL breakpoint region—constitutes the E_2 proximal breakpoint itself (Puerma *et al.*, 2014). The E_2 breakpoint motif (Figure 2 in Puerma *et al.*, 2014) exhibits only two fragments with some similarity to the SGM element, unlike the nearly canonical SGM element detected both at the AK breakpoint region and within the A part of the GL breakpoint region (Figure 3).

Identification and characterization of inversion E_3 breakpoints

According to available cytological information (Kunze-Mühl and Müller, 1958), inversion E_3 shares its proximal breakpoint at section 58D with those of inversions E_2 (or E_1 and E_2; see Introduction and Figure 1) and E_9, whereas it shares its distal breakpoint at section 64C with those of inversions E_2 (or E_1 and E_2; see Introduction and Figure 1). Since inversion E_3 occurred on an $E_{\underline{1+2+9}}$ chromosome, the fragment spanning the proximal breakpoint in this arrangement (AK in Figure 1) was PCR amplified using DNA from strain OF82 ($E_{\underline{1+2+9}}$) and subsequently *in situ* hybridized on $E_{\underline{1+2+9+3}}$ chromosomes (strain FO_12b), where it gave two strong signals at section 58D next to 64C (breakpoint AH) and 68B next to 62D (breakpoint KE) on $E_{\underline{1+2+9+3}}$ chromosomes (Figure 1 and Supplementary Figure S4), confirming that it included the proximal breakpoint of inversion E_3. It should be noted that in both gene arrangements probe AK gave a fainter signal at section 64B next to 68C due to the presence of part A in the GL breakpoint as a result of its duplication during the E_9 inversion process (Figure 3 and Supplementary Figure S4).

For the distal breakpoint, the fragment spanning it in $E_{\underline{1+2}}$ and $E_{\underline{1+2+9}}$ chromosomes (EH and HE, respectively, in Figure 1) was PCR amplified using DNA from strain OF21 ($E_{\underline{1+2}}$) and subsequently *in situ* hybridized on $E_{\underline{1+2+9+3}}$ chromosomes (strain FO_12b), where it gave only one signal at section 64C close to section 62D (Figure 1 and Supplementary Table S1), indicating that it did not span the distal breakpoint of inversion E_3. To identify this breakpoint region, we took advantage of the previous chromosome walk that we had performed to identify the distal breakpoint (GH) of inversions E_1 and E_2 (Supplementary Figure S7 in Puerma *et al.*, 2014), a walk that extended over 700 kb outside the inversion (that is, on its H part that is the one that at the cytological level is reused by inversion E_3; Figure 1). To identify the new breakpoint region that was renamed H1H2, two probes from our previous walk that are at an appreciable distance from the GH breakpoint—probes P51_4 and P51_3 approximately 74-kb and 168-kb apart of that breakpoint (Supplementary Figure S7 in Puerma *et al.*, 2014)—were *in situ* hybridized on $E_{\underline{1+2+9+3}}$ chromosomes. Probe P51_4 mapped at section 64C in the proximity of section 62D (that is, in the non-inverted H1 region; Supplementary Figure S1), whereas probe P51_3 did at section 64C in the proximity of section 58D (that is, in the inverted H2 region) (Supplementary Figure S1 and Table S1). The seven new probes designed in that ~94-kb-long interval mapped, like probe P51_3, at section 64C in the proximity of section 58D, like probe P51_3 (Figure 2, and Supplementary Figure S5 and Supplementary Table S1), indicating that the breakpoint was located between probes P51_4 and DEp51_4_3a. Upon narrowing down the breakpoint region with three new probes, a final probe (DEHbc) was designed to span the breakpoint (Figure 2 and Supplementary Figure S5). When this probe was *in situ* hybridized on $E_{\underline{1+2+9+3}}$ and $E_{\underline{1+2+9}}$ chromosomes, it gave two signals on the former chromosomes—at section 64C next to 62D (KH1) and at section 64C next to 58D (AH2), respectively; Supplementary Figures S1 and S4—while it gave a single signal at section 64C close to 62D (H2H1) on $E_{\underline{1+2+9}}$ chromosomes (Supplementary Figures S1 and S4), confirming that it contained the distal breakpoint of the E_3 inversion.

Upon identification of inversion E_3 breakpoint regions on $E_{\underline{1+2+9}}$ chromosomes—AK and H2H1 (~6.3-kb long) in Figure 4—and sequencing of the H2H1 fragment in $E_{\underline{1+2+9}}$ (OF82) chromosomes, fragments spanning the $E_{\underline{1+2+9+3}}$ breakpoint regions (AH2 and KH1) were amplified with the corresponding combination of oligonucleotides (Supplementary Figure S1) and subsequently *in situ* hybridized on both $E_{\underline{1+2+9}}$ and $E_{\underline{1+2+9+3}}$ chromosomes. In both cases, the amplified fragments gave two signals on $E_{\underline{1+2+9}}$ chromosomes, confirming that the breakpoints were included in the amplified fragments (Supplementary Figure S4). Moreover, these fragments (AH2 and KH1) gave also two signals on $E_{\underline{1+2+9+3}}$ chromosomes, suggesting the presence of a duplicated region in the E_3 breakpoints (see below), and, similarly to fragment AK (see above), fragment AH2 gave an additional weak signal at section 64B next to 68C of the $E_{\underline{1+2+9+3}}$ arrangement (Supplementary Figure S4). Fragments AH2 (~8.3-kb long) and KH1 (~14.3-kb long) were completely sequenced in the FO_12b ($E_{\underline{1+2+9+3}}$) strain.

The H2H1 breakpoint region contains the orthologs of genes GA19540, GA15025, GA24519 and *OstDelta* (Figure 4). Pairwise comparison of the H2H1, AH2 and KH1 breakpoint regions allowed detecting the presence of an ~3.6-kb fragment in the central part of the H2H1 region in both breakpoints of the $E_{\underline{1+2+9+3}}$ arrangement (Figure 4). Pairwise comparison of the AK, AH2 and KH1 regions revealed the precise nucleotide at which the AK region had been broken in the inversion process (Figure 4). Moreover, a LINE element was detected between the K and H1 parts of this breakpoint, with a canonical SGM element inserted within the LINE element (Figure 4).

DISCUSSION

Origin of inversions in *D. subobscura*

An important question concerning the origin of chromosomal inversions, and more specifically of those that increase in frequency and either become polymorphic or attain fixation, refers to the relative importance of ectopic recombination between transposable elements (or other repetitive sequences) and of staggered double-strand breaks (Ranz *et al.*, 2007) in their generation. A large number of breakpoints need to be molecularly characterized to reliably quantify the contribution of the different mechanisms generating inversions in any particular species or species group. In Drosophila, the currently largest

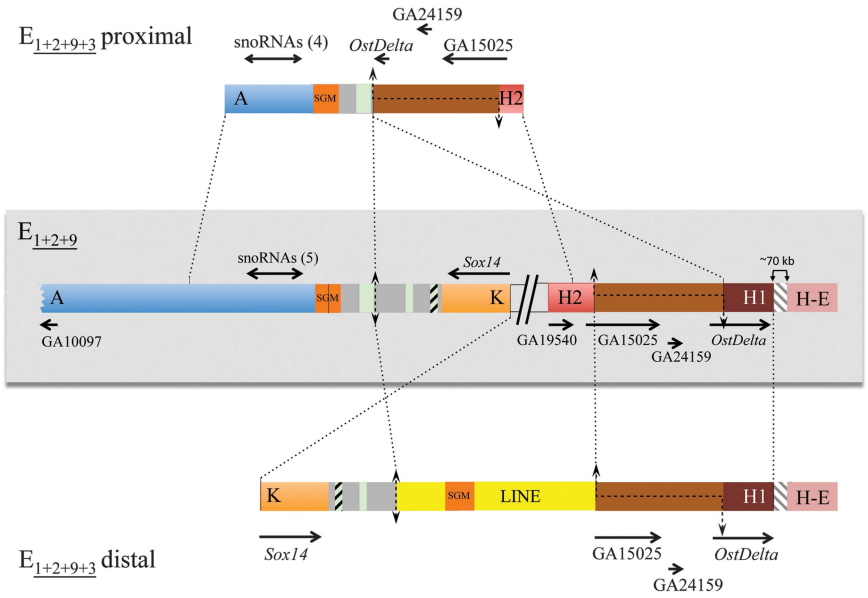

Figure 4 E_3 inversion breakpoints. Schematic representation of the E_3 inversion breakpoint regions on E_{1+2+9} and $E_{1+2+9+3}$ arrangements with the proximal breakpoint labeled as in Figure 1 (AK) and the distal breakpoint relocated to over 70 kb from the HE breakpoint and, therefore, renamed H2H1 (see text and Figure 2). Thick colored bars represent the different breakpoint regions. The central part shows, within a light gray rectangle, the scheme of both breakpoints of the E_3 inversion in gene arrangement E_{1+2+9} (AK and H2H1), separated by two inclined lines. Schemes of the proximal and distal breakpoints in gene arrangement $E_{1+2+9+3}$ (AH2 and KH1) are represented above and below, respectively, of the E_{1+2+9} scheme. The gray striped region present in the distal breakpoint (in both E_{1+2+9} and $E_{1+2+9+3}$ arrangements) indicates the ~70-kb-long fragment separating the H part of the HE region from the H1 part of the H2H1 region. Black discontinuous lines along a chromosomal region represent staggered breaks with their limits indicated by arrows. Vertical double-headed arrows indicate cut-and-paste breakpoints. Dotted lines between arrangements indicate the limits and orientation of homologous regions. On each flanking region, the names of the orthologous coding regions in either *D. melanogaster* or *D. pseudoobscura* are given, with black arrows indicating their sense and approximate size. Double-headed horizontal arrows refer to the multiple snoRNAs generated from the corresponding *Uhg* gene introns, with their number given in parenthesis. Orange boxes labeled SGM refer to near canonical and canonical copies of the SGM transposable element, whereas green and striped green boxes refer to different-sized fragments of the *gypsy* and *Pao* transposable elements, respectively, and the two yellow boxes correspond to the two parts of a canonical LINE element. See Figure 3 legend for the gray boxes content.

data set corresponds to the 29 fixed inversions that differentiate *D. melanogaster* and its close relatives *D. simulans* and *D. yakuba* (Ranz *et al.*, 2007). The presence of duplicated inverted non-repetitive fragments in the breakpoints of 17 inverted chromosomes led the authors to propose the staggered-break mechanism as the prevalent mechanism originating inversions in the melanogaster species group. The recent detection, also through genome sequence comparison, of inverted duplications at the breakpoints of 5 out of the 8 studied inversions segregating in natural populations of *D. melanogaster* (Corbett-Detig *et al.*, 2012) would support the Ranz *et al.* (2007) proposal for the melanogaster species group. This proportion does not seem, however, to hold for the repleta group of species, neither for fixed inversions—with inverted duplications in 3 out of 10 fixed inversions (Calvete *et al.*, 2012; Guillén and Ruiz, 2012)—nor for polymorphic inversions in *D. buzzatii*, where the generating mechanism of the three characterized inversions was ectopic recombination between transposable elements (Cáceres *et al.*, 1999; Casals *et al.*, 2003; Delprat *et al.*, 2009).

In the species here studied, *D. subobscura*, the number of polymorphic inversions with breakpoints molecularly characterized in both inverted and non-inverted chromosomes is still relatively low for any formal quantitation of the generating mechanisms: one inversion—O_3—in Muller's E element (Papaceit *et al.*, 2013) and four inversions—E_1, E_2, E_3 and E_9—in Muller's C element (Puerma *et al.*, 2014; present work). This number allows, however, a qualitative evaluation of their relative importance in this species. In the three previously characterized inversions—O_3, E_1 and E_2—, only in the E_2 case was ectopic recombination between inverted repeated elements—α-motifs—considered the generating mechanism. In contrast, inverted duplicates of rather small fragments were detected in O_3 and E_1, suggesting the staggered-break mechanism of origin (Papaceit *et al.*, 2013; Puerma *et al.*, 2014). In the case of inversions E_9 and E_3 here studied, we have detected inverted duplicates in the corresponding derived chromosomal arrangement in both cases, that is, in E_{1+2+9} and $E_{1+2+9+3}$ chromosomal arrangements, respectively (Figures 3 and 4). Our results from five inversions are similar to those obtained from eight polymorphic inversions in *D. melanogaster* (Corbett-Detig *et al.*, 2012), suggesting that also in *D. subobscura* staggered double-strand breaks and subsequent repair might be the prevalent mechanism generating inversions. Although the number of polymorphic inversions with breakpoints molecularly characterized is rather similar in both species, it should be noted that in contrast to the 29 fixed inversions with breakpoints molecularly characterized since the *D. melanogaster* – *D. yakuba* split (Ranz *et al.*, 2007), no fixed inversion breakpoint has been yet characterized in the *D. subobscura* lineage. Indeed, even if we previously characterized the breakpoint regions of an inversion fixed since the split of the *D. melanogaster* and *D. subobscura* lineages (Cirera *et al.*, 1995), the phylogenetic comparative analysis of these regions has allowed us to infer that it did occur in the melanogaster group and more specifically in the ancestor of the melanogaster subgroup (results not shown).

The extent and content of the fragments duplicated in the inversion process differs between the two previously characterized *D. subobscura* inversions—O_3 and E_1—and those here characterized—E_9 and E_3. In both O_3 and E_1, a rather small fragment was duplicated: an ~300-bp-long intergenic fragment in the O_3 case (Papaceit *et al.*, 2013),

and an ~400-bp-long repeat named β-motif in the E_1 case (Puerma *et al*., 2014). During the E_9 inversion process, an ~8-kb-long fragment that included one protein-coding gene (GA10097/CG10131) and one snoRNAs generating gene (*Uhg5*) became duplicated and is now present in the two breakpoints of the derived $E_{\underline{1+2+9}}$ arrangement (Figure 3). As the duplication includes the complete GA10097/CG10131 coding region and its 5′ regulatory region, the presence of two copies in the derived arrangement might have some phenotypic effect in case that the encoded protein (likely a 3-hydroxyacyl-CoA dehydrogenase involved in fatty acid metabolism) had a dose-dependent effect. Although the *Uhg5* gene is also duplicated in the derived $E_{\underline{1+2+9}}$ arrangement, the number of snoRNAs encoded in each copy (five and seven in the AK and GL breakpoint regions, respectively) is either smaller or equal than in the $E_{\underline{1+2}}$ arrangement (seven in the AG breakpoint region; Figure 3). Thus, only five of the seven *Uhg5* encoded snoRNAs are present twice in the inverted arrangement (results not shown). During the E_3 inversion process, an ~3.5-kb-long fragment became duplicated and is now present in the two breakpoints of the derived $E_{\underline{1+2+9+3}}$ arrangement (Figure 4). In this case, the duplicated fragment includes three coding regions, with the *OstDelta* and the GA15025 genes being truncated in the proximal AH2 breakpoint and the distal KH1 breakpoint (Figure 4), respectively, which would result in a single complete and functional copy of these genes in the $E_{\underline{1+2+9+3}}$ arrangements and two complete and functional copies of the GA24519/CG14488 gene.

In all four *D. subobscura* inversions generated by staggered double-strand breaks, the fragment duplicated in the inverted arrangement corresponds to only one of the breakpoints in the non-inverted arrangement. The other breakpoint is clearly delimited between two adjacent nucleotides in some cases—for example, in the distal KL breakpoint of inversion E_9 (Figure 3) and in the proximal AK breakpoint of inversion E_3 (Figure 4)—, whereas in other cases a small fragment of the single-strand template might have been lost during the repair stage (for example, in the distal breakpoint of inversion O_3; Papaceit *et al*., 2013). This observation of an asymmetric stagger in the two breakpoints of an inversion is not unique to *D. subobscura* inversions but extensive to most molecularly characterized inversions. Among the detected exceptions in inverted chromosomes with inverted duplicates corresponding to both breakpoints in non-inverted chromosomes are the five *D. melanogaster* polymorphic inversions—In(1)A, In(1)Be, In(2R)NS, In(3R)K and In(3R)Payne (Matzkin *et al*., 2005; Corbett-Detig *et al*., 2012)—, six of the 18 inversions with inverted duplicates fixed between *D. melanogaster* and *D. yakuba* (Ranz *et al*., 2007), and the 2q inversion fixed in *D. mojavensis* (Guillén and Ruiz, 2012).

Transposable elements and inversions

Although not directly involved in the origin of the here studied inversions, transposable elements—mostly defective—have been detected in the breakpoint regions of inversions E_9 and E_3. Although transposable elements can transpose to any region of the genome, their rate of excision varies with recombination rate (Charlesworth and Langley, 1989). In inversion heterokaryotypes, recombination in regions affected by chromosomal inversions is reduced, with the highest reduction near the breakpoints (Navarro *et al*., 1997). Transposable elements detected at breakpoint regions will not only include those copies present in the unique chromosome involved in the inversion process but also younger copies accumulated due to the highly suppressed recombination in inversion heterokaryotypes (Charlesworth *et al*., 1997). As both point and length mutations will lead to transposon degeneration through time, sequence comparison of transposable elements, or remnants thereof, detected at or near inversion breakpoints would thus allow detecting the periods of active transposition of particular TEs, and it might also be indicative of the relative age of inversions.

In *D. subobscura*, the characterization of the E_2 breakpoints revealed that the breakpoints were within a repeated motif (α-motif) that exhibited two regions with low similarity to the SGM (Miller *et al*., 2000) family of transposable elements, indicating that they contained remnants of old copies of this element. Although the α-motif was present at each proximal and distal breakpoints of the E_2 inversion (Figure 2 in Puerma *et al*., 2014), it was only present at the proximal breakpoint in the $E_{\underline{1+2}}$ arrangement (Figure 3). Moreover, it was also present in the GL, but not at the AK breakpoint of inversion E_9, which allowed us to place this motif outside of the fragment duplicated during the inversion process (Figure 3). Surprisingly, at the same location of the AK breakpoint region, we detected a nearly complete canonical copy of the SGM element, which would be the result of a novel transposition to this location that might have been fostered by the staggered-break mechanism (Onozawa *et al*., 2014). We also detected two even more conserved copies of the SGM element, one in the distal (GL) breakpoint of inversion E_9—within the A part duplicated fragment—and a second one in the distal (KH1) breakpoint of inversion E_3—within a canonical and complete copy of the LINE transposable element. Although there is no evidence from our data for SGM transposition in the period immediately preceding the E_2 inversion, and soon thereafter, this family seems to have been active at least in the period between the occurrence of inversion E_9 and some time after that of the E_3 inversion. We can also infer from our data that LINE elements were active most possibly after the occurrence of inversion E_3.

Two bouts of SGM activity are now documented in the subobscura subgroup lineage (Miller *et al*., 2000; present work). The first one would have occurred before the split of *D. guanche* from the *D. subobscura-D. madeirensis* lineage as previously inferred from its detection in the P-neogene cluster present in all three species (Miller *et al*., 2000), inference that is supported by the presence of old degenerated copies at both breakpoints of the E_2 inversion (Figure 5; Puerma *et al*., 2014). The presence of near canonical and canonical copies of the SGM element at the breakpoints of both the E_9 and E_3 inversions, points to a second more recent bout of SGM (as well as of LINE) activity possibly after the *D. subobscura-D. madeirensis* split.

Contrasting breakpoint reuse at the cytological and molecular levels

In Drosophila, paracentric inversions constitute the main source of gene reorganization within each chromosomal element. Classical cytological studies of chromosomal polymorphism (that is, of inversions segregating in natural populations) revealed that some breakpoints are shared by two or more inversions (for example, Kunze-Mühl and Müller, 1958; Aulard *et al*., 2002), which could also be ascertained from comparisons of species of the same group or subgroup. Comparative genomic analysis across the complete, or partial, Drosophila phylogeny also revealed that some genes had flanked two or more chromosomal breaks (Bhutkar *et al*., 2008; von Grotthuss *et al*., 2010). To assess reuse at a fine molecular level (that is, more precisely than in either of the above-mentioned types of studies), detailed sequence comparison of the breakpoint regions in non-inverted and inverted chromosomes is needed, and preferably of polymorphic inversions because, as evolutionary change accumulates with time, breakpoint sequences better reflect the inversion process the shorter the time elapsed since their origin.

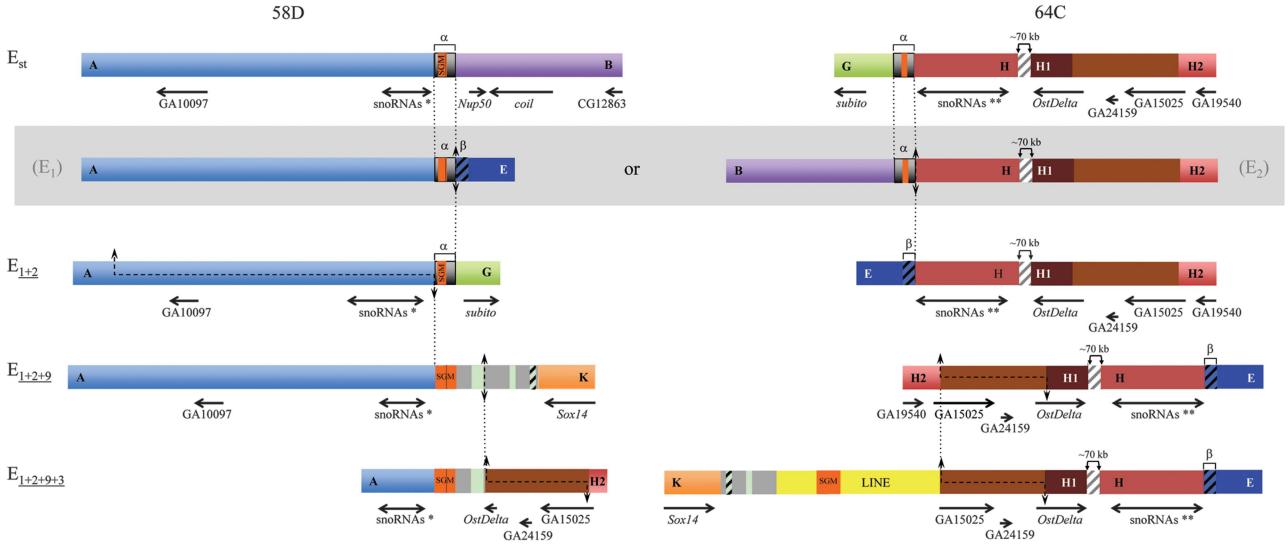

Figure 5 Comparison of two reused breakpoints. The regions spanning the two breakpoints (at sections 58D and 64C, respectively) that have been multiply reused, at least at the cytological level (see Figure 1), in extant chromosomal arrangements E_{st}, E_{1+2}, E_{1+2+9} and $E_{1+2+9+3}$ as well as in the two possible intermediate, and now extinct, arrangements connecting E_{st} and E_{1+2} (within a gray box). Given that inversions occurred sequentially, vertical discontinuous lines connect a particular breakpoint in the corresponding non-inverted and inverted chromosomes. The blue striped fragments refer to a repeat motif, named β-motif in Puerma *et al.* (2014) and, therefore, marked with a β. The asterisk (*) and double asterisks (**) are meant to differentiate the snoRNAs generated from the *Uhg5* and *Uhg1* gene introns, respectively. The sequenced AB fragment is ~18.3-kb long. See legends of Figures 3 and 4 for all other notations.

Our previous characterization of the breakpoints of inversions E_1 and E_2 in *D. subobscura* was one of the first studies that molecularly characterized a polymorphic inversion breakpoint cytologically shared by two inversions (Corbett-Detig and Hartl, 2012; Puerma *et al.*, 2014). Ours was a difficult endeavor because both inversions occurred sequentially and only the E_{st} and E_{1+2} arrangements segregate in extant populations. Although we could not establish which breakpoint had been reused—that is, whether it was the proximal AB or the distal GH breakpoint (Figure 1)—, we could, however, infer that the E_2 break occurred somewhere within an ~700-bp-long repeat motif—α-motif; Figure 2 in Puerma *et al.* (2014)—and the E_1 break most possibly at this motif distal limit. These results indicate that the shared breakpoint (be it the proximal or the distal breakpoint) was reused not only at the cytological level but also at the molecular level.

In the present study, we contrast at the molecular level the multiple cytological and sequential reuse of the breakpoints at sections 58D and 64C by inversions E_1, E_2, E_9 and E_3 (Figures 1 and 5). The breakpoint shared by inversions E_1 and E_2 (be it the proximal or the distal breakpoint) was previously considered an example of strict reuse (even if not at the strictest 1-bp level) as opposed to broad sense reuse adopted in interspecific genome comparisons (breakpoint flanked by the same gene; von Grotthuss *et al.*, 2010). Indeed, breakpoints of inversions E_1 and E_2 were narrowed down to repeats α and β that ranged in size between 400 and 700 bp (Figures 1 and 5; Puerma *et al.*, 2014). In the case of the proximal breakpoint at section 58D, the distal limit of the region duplicated in the E_9 inversion process is actually the proximal limit of the α-motif present in the E_{1+2} arrangement (Figures 3 and 5), whereas the proximal E_3 breakpoint is displaced ~1.5 kb from that limit (Figures 4 and 5). In the case of the distal breakpoint at section 64C, both limits of the E_3 breakpoint are displaced >70 kb from the other breakpoint limits (Figures 4 and 5).

It should be noted that even if the four inversions considered segregate in natural populations, they are not extremely young (Puerma *et al.*, 2014) raising the possibility of length mutations (insertions and deletions of different size and kind) having blurred the actual limits of breakpoints (see above for transposable elements and other intervening sequences; Figures 3–5). If possible indel events occurring during or after the inversion process are considered, our data would support the strict reuse of the proximal breakpoint at section 58D by inversions either E_1, E_2, E_9 and E_3, or E_2, E_9 and E_3 (Figure 5). In the case of the distal breakpoint at section 64C, our data would support its strict reuse by inversions E_1 and E_2—but not by inversion E_3—if E_2 had occurred before E_1, whereas in the alternative case (E_1 before E_2) our data would similarly not support the cytological observation that indicated its reuse by inversions E_2 and E_3.

Our inference of strict multiple reuse of the proximal breakpoint at section 58D is supported at the broad sense by the presence of gene *Uhg5* flanking this breakpoint (Figure 5). Similarly, our inference of strict reuse of the distal breakpoint at section 64C only by inversions E_1 and E_2, if E_2 had preceded E_1, is supported at the broad sense by gene *Uhg1* flanking the breakpoint of only these inversions. *Uhg* genes would seem prone to breakage maybe as a result of the secondary structures that might be adopted by their snoRNAs encoding introns. It is also important to consider that the inversions generating the different arrangements of this complex system occurred sequentially. The presence in inversion heterokaryotypes of asynapsed regions near breakpoints might add some tension to those regions and render them more breakage prone than completely synapsed regions. The added fragility associated with *Uhg* genes and with asynapsed regions might not, however, fully explain the strict reuse, detected one or multiple times, of the breakpoints here considered.

In summary, we have characterized the breakpoints of four inversions involved in a complex inversion system and shown that these inversions originated more frequently by the staggered-break mechanism than by repeat-mediated ectopic recombination. Moreover, we have shown that the transposable elements that accumulate at the breakpoint regions as a result of reduced recombination in heterokaryotypes can be indicative of the periods of their active

transposition. Finally, we have shown that one of the multiply shared breakpoints at the cytological level—at section 58D—has also been multiply reused at the molecular level, whereas the other one—at section 64C—might have been reused only once at this level.

DATA ARCHIVING

Sequences newly obtained have been deposited in the EMBL/GenBank Data Libraries under accession numbers LM999978 to LM999984, and updated accession numbers LK022764.2 and LK022779.2.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank David Salguero for his excellent technical assistance, and Servei de Genòmica, Serveis Cientifico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. We also thank reviewers for their critical comments. This paper was prepared with full knowledge and support of the Barcelona Subobscura Initiative (BSI). This work was supported by grants BFU2012-35168 from Ministerio de Economía y Competitividad, Spain, and 2009SGR-1287 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Generalitat de Catalunya, Spain to MA.

Andolfatto P, Wall JD, Kreitman M (1999). Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics* **153**: 1297–1311.

Aulard S, David JR, Lemeunier F (2002). Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genet Res* **79**: 49–63.

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

Burland TG (2000). DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol* **132**: 71–91.

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.

Calvete O, González J, Betrán E, Ruiz A (2012). Segmental duplication, microinversion and gene loss associated with a complex inversion breakpoint region in *Drosophila*. *Mol Biol Evol* **29**: 1875–1889.

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.

Charlesworth B, Langley CH (1989). The population genetics of Drosophila transposable elements. *Annu. Rev Genet* **23**: 251–287.

Charlesworth B, Langley CH, Sniegowski PD (1997). Transposable element distributions in Drosophila. *Genetics* **147**: 1993–1995.

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

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA *et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203–218.

Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415–1418.

Corbett-Detig RB, Cardeno C, Langley CH (2012). Sequence-based detection and breakpoint assembly of polymorphic inversions. *Genetics* **192**: 131–137.

Corbett-Detig RB, Hartl DL (2012). Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genet* **8**: e1003056.

Delprat A, Negre B, Puig M, Ruiz A (2009). The transposon *Galileo* generates natural chromosomal inversions in Drosophila by ectopic recombination. *PLoS One* **4**: e7883.

Ferguson-Smith MA, Trifonov V (2007). Mammalian karyotype evolution. *Nat Rev Gen* **8**: 950–962.

Guillén Y, Ruiz A (2012). Gene alterations at Drosophila inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics*. **13**: 53.

Krimbas CB, Powell J (1992). *Drosophila inversion polymorphism*. CRC: Boca Raton, FL, USA.

Kunze-Mühl E, Müller E (1958). Weitere Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. *Chromosoma* **9**: 559–570.

Laayouni H, García-Franco F, Chávez-Sandoval BE, Trotta V, Beltran S, Corominas M *et al.* (2007). Thermal evolution of gene expression profiles in *Drosophila subobscura*. *BMC Evol Biol* **7**: 42.

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.

Miller WJ, Nagel A, Bachmann J, Bachmann L (2000). Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. *Mol Biol Evol* **17**: 1597–1609.

Montgomery EB, 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 *D. melanogaster*. *Genet Res* **49**: 31–41.

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

Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, McGrath SD *et al.* (2005). A genome-wide survey of structural variation between human and chimpanzee. *Genome Res* **15**: 1344–1356.

Onozawa M, Zhang Z, Kim YJ, Goldberg L, Varga T, Bergsagel PL *et al.* (2014). Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc Natl Acad Sci USA* **111**: 7729–7734.

Papaceit M, Segarra C, Aguadé M (2013). Structure and population genetics of the breakpoints of a polymorphic inversion in *Drosophila subobscura*. *Evolution* **67**: 66–79.

Pevzner P, Tesler G (2003). Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc Natl Acad Sci USA* **100**: 7672–7677.

Puerma E, Orengo DJ, Salguero D, Papaceit M, Segarra C, Aguadé M *et al.* (2014). Characterization of the breakpoints of a polymorphic inversion complex detects strict and broad breakpoint reuse at the molecular level. *Mol Biol Evol* **31**: 2331–2341.

Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J *et al.* (2007). Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol* **5**: 1366–1381.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R *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.

Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM *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.

Tonzetich J, Lyttle TW, Carson HL (1988). Induced and natural break sites in the chromosomes of Hawaiian Drosophila. *Proc Natl Acad Sci USA* **85**: 1717–1721.

von Grotthuss M, Ashburner M, Ranz JM (2010). Fragile regions and not functional constraints predominate in shaping gene organization in the genus Drosophila. *Genome Res* **20**: 1084–1096.

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

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