An old bilbo-like non-LTR retroelement insertion provides insight into the relationship of species of the virilis group

Micael Reis, Cristina P. Vieira, Ramiro Morales-Hojas, Jorge Vieira

Instituto de Biologia Celular e Molecular (IBMC), University of Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal

ARTICLE INFO

Article history:
Received 10 April 2008
Received in revised form 18 July 2008
Accepted 5 August 2008
Available online 18 August 2008

Keywords:
Phylogeny
Xc inversion
Drosophila Americana
Deletion

1. Introduction

In Drosophila, most individual TE insertions usually segregate at less than 5% frequency (see references in Petrov et al., 2003; Biémont et al., 2003; Maside et al., 2005). This pattern is created by a balance between the spreading of TEs, and natural selection removing them (Charlesworth and Charlesworth, 1983; Kaplan and Brookfield, 1983; Langley et al., 1983). Nevertheless, in Drosophila melanogaster, individual non-LTR retroelements can segregate at high frequencies (Petrov et al., 2003), but they represent only about 18% of all retroelement copies (Bergman and Bensasson, 2007).

Few apparent fixations or near-fixations, at the population or species level, have been described in Drosophila, both in the recombining euchromatin (Maside et al., 2002; Petrov et al., 2003; Catania et al., 2004; Schlenke and Begun, 2004; Lipatov et al., 2005), non-recombining euchromatin (Bartolomé and Maside, 2004), and non-recombining heterochromatin (Maside et al., 2005; Junakovic et al., 2005).

In non-recombining regions, selection against the possible deleterious effects of TEs is less effective. Therefore, in non-recombining regions, TEs may reach fixation mostly due to drift (Bartolomé and Maside, 2004).

In recombining regions non-LTR retroelements may also reach fixation due to drift alone (Petrov et al., 2003). Non-LTR retroelements commonly generate 5′ truncated dead-on-arrival elements as a natural outcome of transposition (Luan et al., 1993). Since they do not encode functional proteins there is no selection against the deleterious effects of TE-generated proteins, as proposed by Nuzhdin (1999). Moreover, secondary deletions are common, and these may lower the strength of purifying selection acting on individual copies (Petrov et al., 2003). In agreement with this view, in D. melanogaster most apparent fixations found in euchromatic regions involve non-LTR retroelements (Petrov et al., 2003; Lipatov et al., 2005).

Occasionally, a TE insertion in recombining euchromatin may positively affect the fitness of its carriers. This is likely the case of the non-LTR retroelement Doc insertion that segregates at high frequency in D. melanogaster, which creates a truncated version of the putative choline transferase gene (Petrov et al., 2003; Aminetzach et al., 2005). Such insertion confers increased resistance to pesticide (Aminetzach et al., 2005). Furthermore, in D. melanogaster, insecticide resistance is correlated as well with the insertion of an Accord-like LTR retroelement into the 5′ region of the cytochrome P450 gene (Catania et al., 2004). In D. simulans, there is also an insertion of a Doc non-LTR retroelement into the 5′ region of the cytochrome P450 gene that confers resistance to pesticides (Schlenke and Begun 2004). In all cases, the reduction in variability
levels in the region encompassing the TE insertion provides evidence for a recent selective sweep (Catania et al., 2004; Schlenke and Begun, 2004; Aminetzach et al., 2005). In *D. melanogaster*, Maside et al. (2002) also reported the fixation of S-element insertions into the 5’ regions of three members of the *Hsp70* heat shock gene family. A significant reduction in variability levels indicative of a recent selective sweep was also found in this case (Maside et al., 2002). Nevertheless, it is unknown how the fitness of its carriers is positively affected.

The increase in frequency of the TE insertions described in *D. melanogaster* and *D. simulans* that correlate with insecticide resistance are recent. Indeed, those insertions are found near-fixation in non-African populations only (Catania et al., 2004; Schlenke and Begun, 2004; Aminetzach et al., 2005). There are no insertions of S-elements around *Hsp70* gene family members in *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. teissieri*, or *D. repleta* (Maside et al., 2002). Therefore this is also a recent event.

In *D. melanogaster*, about 90% of LTR retroelements were mobilized less than 92,600 years ago (Bergman and Bensasson, 2007). In contrast, for non-LTR retroelements, the average mobilization age is 782,400 years (Bergman and Bensasson, 2007). Only 12.7% of non-LTR copies are predicted to predate the *D. simulans*–*D. melanogaster* split estimated to have occurred 5.41 Mya (Bergman and Bensasson, 2007), and indeed most *D. melanogaster* TE insertions are not shared by closely related species (Caspi and Pachter, 2006). In *D. melanogaster*, only three TE fixations at the species level are estimated to be in between 3.5 to 10 million years old (Petrov et al., 2003).

In this study, we report the presence of a non-LTR retroelement insertion in the *CG1806*–*CG8002* intergenic region in five species of the virilis group, namely *D. littoralis*, *D. virilis*, *D. lummei*, *D. americana*, and *D. novamexicana*. In all species, except *D. americana*, the non-LTR retroelement insertion seems to be fixed. The comparison of species and individuals from the same species reveals that the non-LTR retroelement is mostly being slowly lost through the accumulation of small deletions, although there are instances where it was quickly lost through the accumulation of large deletions.

Non-LTR retroelements evolve primarily or even exclusively through vertical transmission (Malik et al., 1999). Therefore, it is unlikely that the insertion found in different species in the *CG1806*–*CG8002* intergenic region is the result of independent insertions. However, in order to validate this assumption, we compared the phylogenetic relationship obtained for the non-LTR retroelement insertion with that for the species where it is found, using previously published gene sequences (5 genes) as well as new sequences for two genes. It should be noted that the presence of a *bilbo*-like insertion in *D. littoralis* and its absence in *D. kanekoi* suggests that *D. littoralis* is more closely related to species of the *virilis* phylogroup than to species of the *montana* phylogroup. Nevertheless, traditionally, *D. littoralis* has been placed in the *montana* phylogroup, although the relationship of *D. littoralis* and species of the *montana* and *virilis* phylogroups is not yet well resolved (compare for instance Vieira (2002) and Spicer and Bell (2002)).

2. Materials and methods

2.1. DNA samples

To study the *bilbo* region the following strains were used: *D. littoralis* (BP41, Portugal), *D. kanekoi* 15010–1061.00 (Japan), *D. virilis* A12 (Japan), B15 (Japan), B42 (Mexico), W159 (Holland), 1432 (UK), 10518 (California, USA), 10519 (Japan), 10514 (laboratory strain), 10514 (Argentina), 10515 (Chile), and 10512 (Russia); *D. lummei* 200 (Russia); *D. novamexicana* 15010–1031.08 (New Mexico, USA) and 15010–1031.12 (Colorado, USA); *D. americana* G96.11 (Indiana, USA), NN97.4 (Nebraska, USA), ML97.4 (Louisiana, USA), and ML97.5 (Louisiana, USA). Furthermore, 95 *D. americana* flies were collected at the end of July and beginning of August 2004, from four collection sites (Lake Wappapello (N=35), Lake Ashbaugh (N=20), Lake Hurricane (N=9) and Howell Island (N=31)). Of these, 64 (68.4%) were males and 31 (31.6%) females. Most females (93.5%) were found to be inseminated. All wild-caught males as well as one son from each inseminated female were used for this study (93 individuals in total). Data was obtained for two genes (*CG7219* and *CG9631*) used in the phylogenetic reconstruction. The following strains were used: *D. montana* Mo1 (Finland), *D. kanekoi* 15010–1061.00 (Japan), *D. littoralis* BP41 (Portugal), *D. virilis* SBB (Japan), *D. lummei* 200 (Russia), *D. novamexicana* 15010–1031.00 (Colorado, USA), *D. americana* W14 (Missouri, USA) and *D. americana* texana ML97.5 (Louisiana, USA).

2.2. PCR amplification, cloning and sequencing of the *CG1806*–*CG8002* intergenic region in between primers *BilboF* and *BilboR*

The NN97.4 *D. americana* *CG1806*–*CG8002* intergenic region was obtained using primers cg1806F and cg8002R (supplementary Table 1). Standard amplification conditions were 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 48 °C for 45 s, and primer extension at 72 °C for 3 min. The 4000 bp amplification product was cloned using TOPO-TA Cloning Kit for Sequencing. Three colonies were sequenced to correct for possible nt mis-incorporations that may have occurred during the PCR reaction. Sequencing was performed by STABVIDA (Lisbon, Portugal), using ABI PRISM BigDye cycle-sequencing kit version 1.1 (Perkin Elmer, CA, USA) and specific primers (Supplementary Table 1) and the primers for the M13 forward and reverse priming sites of the pCR2.1 vector.

Based on this sequence primers *BilboF* and *BilboR* (supplementary Table 1) were designed. When these primers are used an amplification fragment is obtained from all species of the *virilis* phylogroup (*D. virilis*, *D. lummei*, *D. americana*) except from *D. novamexicana*. In the latter species, the region where primer *Bilbo* is located is missing. Therefore, for *D. novamexicana*, primer *BilboNov* was used. The conditions for amplification, cloning and sequencing are those described above.

For *D. littoralis* and *D. kanekoi*, primers *virmojCG1806F* and *virmojCG8002R* (supplementary Table 1) were designed based on the *virilis* and *D. majavensis* genomes (http://flybase.bio.indiana.edu/). PCR amplifications were performed using system 3 protocol of the Roche Expand long template PCR system (Roche, Germany). The amplification products were cloned using Topo XL PCR cloning Kit (Invitrogen, Spain). DNA sequencing was performed as described above. All DNA sequences except that for *D. kanekoi* were deposited in GenBank (accession numbers are FJ014927–FJ014941; the *D. kanekoi* sequence could not be deposited in GenBank since the studied region is smaller than 50 bp; see Results section).

2.3. PCR amplification, cloning and sequencing of the genes *CG7219* and *CG9631*

In order to obtain a species phylogeny, nt sequences for genes *CG7219* and *CG9631* were obtained for *D. montana*, *D. kanekoi*, etc.
D. littoralis, D. virilis, D. lummei, D. novamexicana, D. americana americana and D. americana texana. Primer sets 7219F-R and 9631F-R were designed based on the D. virilis and D. mojavensis genomes (Supplementary Table 1). Amplification conditions are those described above, with annealing temperature of 56 °C and 50°C for genes CG7219 and CG9631, respectively. Cloning and sequencing was performed as described above. We used the universal primers for the priming sites present in the vector arms and primers indicated in Supplementary Table 1. The DNA sequences were deposited in GenBank (accession numbers for CG7219 are FJ006536–FJ006543 and for CG9631 are FJ006544–FJ006551).

2.4. Phylogenetic analyses

Given the short length of the bilbo-like copy found in the genome of D. littoralis, the evolutionary relationships between the copies from the different species were estimated using a distance method (Minimum Evolution, ME) and Maximum Likelihood (ML) including and excluding the sequence from D. littoralis, respectively. Before running the phylogenetic analyses, levels of saturation were checked with DAMBE (Xia and Xie, 2001) by plotting the number of observed transitions and transversions against the corrected genetic distance (TN93). Some saturation was observed in the transitions when the bilbo-like sequence from D. littoralis was included (not shown), thus, the ME analysis was run excluding transitions. Heuristic searches were run using ME and ML as optimality criteria in PAUP* v4.0b10 (Swoford, 2002). The starting trees were obtained via stepwise addition, with random addition of sequences and 100 replicates. Tree-bisection–reconnection was used as the branch-swapping algorithm.

The models of evolution implemented in ME and ML analyses were obtained with the Akaike Information Criterion (AIC) applied in ModelTest 3.7 (Posada and Crandall, 1998). To test the robustness of the trees 1000 bootstrap replicates were run.

The species phylogeny was estimated using MrBays 3.1.2 (Ronquist and Huelsenbeck, 2003). Only those species for which bilbo-like copies had been sequenced were included in the analysis, with D. mojavensis as outgroup. The phylogeny was estimated using sequences from seven genes (16S-12S mtDNA, Adh, fu, Gpdh and Nona from the GenBank; CG9631 and CG7219 sequenced by our lab) in a total dataset of 8649 bp. The Bayesian analysis was run with the combined dataset partitioned according to genes. The evolutionary models for each gene partition were found using the Akaike Information Criterion (AIC) as implemented in Modelltest 3.7. Two simultaneous runs of 1×10^6 generations with four chains each (1 cold and 3 heated chains) were set up. Tree-bisection–reconnection was used as the branch-swapping algorithm. The consensus tree was obtained with the Akaike Information Criterion (AIC) as implemented in Modeltest 3.7. Two simultaneous runs of 1×10^6 generations with four chains each (1 cold and 3 heated chains) were set up. The Bayesian input was that obtained with the Bayesian analysis.

3. Results and discussion

3.1. A bilbo-like non-LTR retroelement insertion is found in five species of the virilis group in the intergenic region CG1806–CG8002

While studying the CG1806–CG8002 X chromosomal region in D. americana, using BLASTX and the NCBI non-redundant protein database, a defective non-LTR retroelement showing similarity with bilbo (from D. subobscura; accession number AAB92389), and trim (from D. miranda; accession number CAA41923), was identified (Fig. 1). The D. americana copy is likely defective, since frame shifts are observed relative to the reported open reading frame for the two other elements (Fig. 1). This is not surprising since non-LTR retroelements commonly generate 5’ truncated dead-on-arrival elements as a natural outcome of transposition (Luan et al., 1993). Furthermore, secondary deletions are common in this kind of TEs (Petrov et al., 2003). Since the homology is highest with the D. subobscura non-LTR bilbo retroelement, we named the insertion found in D. americana bilbo-like.

The observed amino acid similarity (about 50% conserved amino acids and conservative changes between bilbo, trim and the bilbo-like element here described) means that in the D. americana strain NN97.4 this non-LTR retroelement is, at least, 375 bp long. A large insertion is present in the D. americana copy relative to the reported D. subobscura bilbo and D. miranda trim amino acid sequences that may be also defective (Fig. 1). This insertion may be part of the open reading frame of the D. americana copy, although it does not show similarity to other reported TE amino acid sequences.

We also looked for the presence of the bilbo-like insertion in other species of the virilis group. Despite several attempts, the intergenic region of interest could only be amplified in D. littoralis, D. kaneko, D. virilis, D. lummei, and D. novamexicana. In all species, except D. kaneko a bilbo-like insertion could be found in the CG1806–CG8002 intergenic region (Fig. 2). The above mentioned insertion of unknown origin found in D. americana strain NN97.4 is also observed (with varying sizes) in D. virilis, D. lummei, and D. novamexicana (Fig. 2).

3.2. bilbo-like non-LTR retroelements are rare in D. virilis

Using TBLASTX and the 12 available Drosophila genome sequences (http://flybase.bio.indiana.edu/), we estimated the number of TEs showing similarity to the bilbo-like retroelement in species from the
melanogaster, obscura, willistoni (subgenus Sophophora), repleta, virilis and Hawaiian (subgenus Drosophila) groups (Table 1). In D. virilis, only one such insertion is found, while none is found in D. mojavensis, a species that has been estimated to be diverging from D. virilis for 26 million years (My) (time estimated for the virilis-repleta radiation; Powell, 1997). In contrast, in Hawaiian Drosophila many copies showing similarity with the bilbo-like insertion are found. Given that this group has been diverging from the virilis-repleta group for over 30 My (Russo et al., 1995; Powell, 1997), it is tempting to conclude that bilbo-like elements have remained at very low frequencies, for the last 26 My, in the Drosophila subgenus. Nevertheless, large differences are found for the estimated number of sequences showing similarity with the bilbo-like insertion in the closely related species from the melanogaster sub-group (Table 1). Such a conclusion cannot be thus, taken at present.

3.3. Within species variability

For D. virilis, the data presented in Fig. 2 suggests little variation regarding the size of the bilbo-like element. Indeed, the four D. virilis strains analysed differ by few and small (<100 bp) deletions. When using 13 D. virilis strains from very diverse geographic origin (Europe, Asia, North and South America; see Materials and methods section for details) and primers BilboF and BilboR, a PCR amplification product of about 800 bp is always obtained. The predicted size, based on the D. virilis genome sequence is 778 bp. For D. americana, however, there is pronounced variation regarding the size of the bilbo-like insertion (Fig. 2). In this species, there is an X/4 fusion Xc inversion north–south shallow cline (Vieira et al., 2006). These chromosomal arrangements are frequent in the north of the geographic distribution and are almost absent in the south of the distribution (Vieira et al. 2001; McAllister 2002; Vieira et al. 2006). Weak selection is likely responsible for the maintenance of the cline (Vieira et al. 2001, 2003, 2006).

The two strains (NN97.4 and G96.11) known to carry the X/4 fusion and the Xc inversion (Vieira et al., 2001) show a longer insertion than the other individuals analysed. The Xc inversion but not the X/4 fusion is found in D. novamexicana (Hsu, 1952). In this species the size of the bilbo-like insertion is almost identical to that found in the D. americana individuals harbouring the Xc inversion (Fig. 2). Thus, little DNA loss occurred at this region since D. novamexicana speciation. Based on DNA sequence data, age estimates in between

<table>
<thead>
<tr>
<th>Species</th>
<th>Total number of hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>&gt;5</td>
</tr>
<tr>
<td>D. simulans</td>
<td>&gt;25</td>
</tr>
<tr>
<td>D. sechellia</td>
<td>0</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>&gt;60</td>
</tr>
<tr>
<td>D. erecta</td>
<td>0</td>
</tr>
<tr>
<td>D. ananassae</td>
<td>&gt;100</td>
</tr>
<tr>
<td>D. pseudoobscura</td>
<td>&gt;30</td>
</tr>
<tr>
<td>D. persimilis</td>
<td>&gt;65</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>0</td>
</tr>
<tr>
<td>D. mojavensis</td>
<td>0</td>
</tr>
<tr>
<td>D. virilis</td>
<td>1</td>
</tr>
<tr>
<td>D. grimshawi</td>
<td>&gt;55</td>
</tr>
</tbody>
</table>

Table 1

Estimated number of bilbo-like copies in the genome of 12 Drosophila species

Fig. 2. Schematic representation of the region in between primers BilboF and BilboR in species of the virilis group. The bilbo insertion is highlighted in gray. The squared pattern indicates the region of the putative bilbo-like element that shows no homology to other TE amino acid sequences. In black is shown the CG1806 (located at the left)-CG8002 (located at the right) intergenic region. White spaces indicate deletions. The arrow shows the transcription orientation of the bilbo-related open reading frame.
300 and 400 thousand years have been proposed for the last common ancestor of *D. americana* and *D. novamexicana* (Caleta and McAllister, 2004; Vieira et al., 2006). It should be noted that the *bilbo*-like insertion is located at most 80 kb away from the basal Xc inversion breakpoint (Vieira et al., 2006).

The above observations suggest that in *D. americana*, when using the primers BilboF/BilboR, the presence of a PCR amplification product of about 1400 bp may be a molecular marker for the Xc inversion. In order to test this hypothesis, we first typed 93 wild-caught individuals from different latitudes for the presence of the Xc/fusion using previously described markers for this chromosomal rearrangement (Table 2; Vieira et al., 2001, 2006; Morales-Hojas et al., 2008). As expected, the above mentioned Xc fusion north–south gradient is observed. From north to south, the estimated frequency of the Xc/fusion based on the CG18543 BstBI and fused1 Clal markers, is respectively: Howell Island (100%) and 96.6%, Lake Wappapelo (82.9%) and 74.3%, Lake Ashbaugh (55%) and 55%, and Lake Hurricane (44.4% and 33.3%).

According to the cytological observations of Hsu (1952), based on a large number of individuals, 94.6% of the *D. americana* Xc/fusion chromosomes present the Xc inversion. In contrast, only about 7% of non-fusion chromosomes present the Xc inversion. When 65 wild-caught *D. americana* individuals that can be unambiguously typed as having the Xc/fusion are used, a PCR amplification product of about 1400 bp can be obtained from all individuals except two (H17 and W36:2; Table 2) that show a band of about 1200 bp (Table 2). Among 3.51 individuals were expected not to have the Xc inversion. When 18 individuals that can be unambiguously typed as not having the Xc/fusion are used, only one individual (H22.10; Table 2) shows an amplification product of about 1400 bp (Table 2). 126 chromosomes were expected to have the Xc inversion.

67% of the non-fusion chromosomes showed an amplification fragment of about 450 bp that, is compatible with no *bilbo*-like insertion. These individuals are presumed to be similar to the wild-caught individuals L475 and W48 as well as to the individuals from strains ML97.4 and ML97.5, shown in Fig. 2. 28% of the *D. americana* non-fusion chromosomes showed an amplification product larger than 450 bp and smaller than 1400 bp.

From north to south, the estimated frequency of the Xc inversion based on the *bilbo*-related 1400 bp amplification product is: Howell Island (96.6%), Lake Wappapelo (74.3%), Lake Ashbaugh (55%), and Lake Hurricane (33.3%) (Table 2).

### Table 2
Survey of 93 wild-caught individuals from different populations for the presence of the Xc/fusion, using previously described markers (the CG18543 BstBI and fused1 Clal markers), and for the size of the PCR fragment amplified using primers BilboF and BilboR

<table>
<thead>
<tr>
<th>Population</th>
<th>Marker</th>
<th>Bilbo size N</th>
<th>Ind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howell Island</td>
<td>F F</td>
<td>1400 27</td>
<td>1.10, 2.3, 4, 15, 6, 7, 9, 10, 11, 12, 13, 10, 14, 15, 10, 18, 20, 22, 23, 25, 24, 25, 26, 28, 29, 29, 30, 32, 33</td>
</tr>
<tr>
<td>Lake Wappapelo</td>
<td>F F</td>
<td>1400 25</td>
<td>1, 2, 4, 6, 11, 14, 15, 17, 18, 24, 26, 28, 29, 29, 30, 32, 33</td>
</tr>
<tr>
<td>Lake Ashbaugh</td>
<td>F F</td>
<td>1400 9</td>
<td>4, 6, 9, 15, 20, 22, 31, 36, 37</td>
</tr>
<tr>
<td>Lake Hurricane</td>
<td>F F</td>
<td>1400 3</td>
<td>15, 16, 17</td>
</tr>
</tbody>
</table>

Marker 1—CG18543 BstBI marker; Marker 2—fused1 Clal marker; biollo size—Approximate size in bp using primers BilboF and BilboR; N—Number of individuals; Ind—Individuals typed; F—Xc/fusion; NF—non-fusion.

3.4. Estimating the age of the *bilbo*-like insertion

Non-LTR retroelements evolve primarily or even exclusively through vertical transmission (Malik et al., 1999). The *bilbo*-like element here described is a non-LTR retroelement, and thus, is unlikely that the insertion found in the *CG1806–CG8002* intergenic region in five species of the *virilis* group is due to multiple independent insertions. Furthermore, *bilbo*-like elements are rare in the *D. virilis* genome (see above) and thus, it seems unlikely that they would insert in the same genome region by chance. The similar size and deletion pattern found in *D. americana*, *D. novamexicana*, and *D. lummei* is another argument favouring the single insertion hypothesis (Fig 2). In *D. littoralis* and those species, the neighbouring 5′ region shown in Fig. 2 can be unambiguously aligned (data not shown). The ML and ME phylogenies based on the *bilbo*-like insertion nt sequence data fits well the inferred species relationships (see below), although we could not resolve the placement of *D. lummei* (Fig. 3). *D. americana* individuals presenting the Xc inversion and *D. novamexicana* cluster together with high bootstrap support, as expected since they share the Xc inversion and this insertion is at most 80 kb away from the basal Xc inversion breakpoint (Vieira et al., 2006). It should be noted that when reconstructing the *bilbo*-like insertion phylogeny we are assuming that the entire region evolves at the same rate. Although, this assumption seems reasonable, given that we are analyzing a defective TE which is likely evolving according to neutrality, the Likelihood Ratio Test rejects the molecular clock hypothesis (*P*=0.0374). Overall, it seems very likely that the presence of the *bilbo*-like element in the five *Drosophila* species is not due to multiple independent insertions. Nevertheless, although unlikely, it is also possible that the integrase encoded by this *bilbo*-like element is sequence specific and always targets the same integration site.

The main reason for re-evaluating the phylogeny of the *virilis* group using previously published and new data was getting a good estimate for the speciation ages, and thus for the age of the *bilbo*-like insertion, under the hypothesis of a single ancestral insertion. This is an important aspect of this study as very few very old fixations have been so far reported in *Drosophila* (Petrov et al., 2003). An over-estimation of the *virilis* species ages could result in an over-estimation of the *bilbo*-like insertion age. Furthermore, the phylogenic relationship of *D. littoralis* (showing evidence for a *bilbo*-like insertion) and *D. kanekoi* (not showing any evidence for a *bilbo*-like insertion) is not well resolved (compare for instance Vieira, 2002, Spicer and Bell, 2002). The estimated Bayesian phylogeny is shown in Fig. 4. All nodes show high Bayesian Posterior Probabilities, indicating a strong support for the topology obtained. In the phylogeny, *D. kanekoi* is basal to *D. littoralis*, which has a most recent common ancestor with the *virilis* subphylog. This result contrasts with the taxonomic classification of *D. kanekoi* and *D. littoralis* within the montana phyad, which should have been reflected in the phylogeny as a monophyletic clade including
D. montana, D. kanekoi and D. littoralis and showing reciprocal monophyly with the virilis subphyllad. Being D. kanekoi older than the D. littoralis/virilis subphyllad lineage suggests that, the bilbo-like retroelement was inserted in the genome of the D. kanekoi lineage. However, it is not possible to reject an earlier insertion of the bilbo-like element in the genome of the D. kanekoi/D. littoralis ancestor with a later loss of the region in D. kanekoi after it diverged from the other lineages.

Results of the molecular dating are shown in the form of a chronogram in Fig. 4. Our age estimates correlate well with geological events that could have been responsible for putative allopatric speciation events. The origin of the virilis group has been placed in East Asia (Throckmorton, 1982; Vieira and Charlesworth, 1999; Wang et al., 2006). According to our age estimates, the ancestral species of D. montana phylad (represented by D. montana in the present analysis), formed by Nearctic species with the exception of D. montana that has a Nearctic and Palearctic distribution, started to diverge from the other species around 10 Mya (6.2–15.9 Mya 95% confidence interval; Fig. 4). There is evidence suggesting that the Bering Strait possibly started to open up for the first time since the middle Cretaceous period (105 Mya) around 4.8 to 7.4 Mya (Marincovich and Gladenkov, 1999), an event that could have disrupted a putative ancestral population of the virilis group distributed throughout both continents. D. kanekoi speciated around 8.4 Mya (5.3–12.5 Mya 95% confidence interval) after the ancestral species of the montana subphyllad diverged. This is a species found so far in Japan, and although it is also likely to be found in adjacent mainland Asia (Throckmorton, 1982), there are still no records from there. Given this known distribution, it is possible that the speciation of D. kanekoi is the result of the final opening of the Japan Sea that occurred at the end of the mid Miocene (Jolivet et al., 1994), isolating its ancestor from the mainland populations.

D. littoralis is found widespread in Europe (Throckmorton, 1982). The ancestry of this species diverged from the virilis lineage (D. virilis has been suggested to have originated in South East Asia, Throckmorton, 1982) around 7.7 Mya (5–11.2 Mya 95% confidence interval) when there was an increase in aridity in the Asian interior (9–8 Mya) after a significant increase in the Tibetan Plateau occurred 10–8 Mya (Guo et al., 2002; Zhisheng et al., 2001). This event could have separated in two an ancestral population expanding from East Asia to Europe. D. virilis is proposed to have originated in South East China where wild populations are found (and from where it has expanded in recent times to other parts of the world as a ‘domestic’ species) (Vieira and Charlesworth, 1999; Mirol et al., 2007) and D. lummei is found from Fennoscandia to Russia and Georgia, and Japan assuming thus that it is present also in central Asia (Throckmorton, 1982). From the late-Miocene to mid-Pliocene there is a period of climate instability in East Asia with alternate cool-warm periods until the definitive onset of the Northern Hemisphere Glaciation around 3.4–3.2 Mya (Guo et al., 2002; Jiang et al., 2005; Vandenberghe et al., 2004; Zhisheng et al., 2001). The speciation of D. virilis occurred within this period of climatic instability around 3.8 Mya (3.05274–4.79544 Mya 95% confidence interval) and could be the result of expansion–contraction events of their ecological niches. D. lummei and the American species of the virilis subphyllad share a most recent common ancestor, which could have been distributed throughout both continents until the setting of the Northern Hemisphere Glaciation around 3.4–3.2 Mya that disjoined both populations. In concordance with this event, D. lummei is estimated to have speciated around 3 Mya (2.5–3.4 Mya 95% confidence interval). The estimated time of divergence of D. novamexicana from D. americana is around 1.5 Mya (1.1–2.1 Mya 95% confidence interval). The discrepancy between this estimate and that
of Caletka and McAllister (2004), Vieira et al. (2006), and Morales-Hojas et al. (2008), on the order of 300–400 thousand years ago, may be due to the use of two individuals for D. americana (representing the standard chromosome arrangement and the X/4 inversion Xc inverted arrangement). Under the phylogenetic model used, all within species variability post-dates speciation, and this may be an unrealistic assumption for closely related species such as D. americana and D. novamexicana. The mid- to late-Pleistocene is a period characterised by pluvial–interpluvial cycles in Southwest North America that has very likely influenced the evolution and diversification of these species as it did for other North American species (Ayoub and Riechert, 2004; Castoe et al., 2007).

4. Conclusion

Overall, our results suggest that the reason for the sharing of a bilbo-like insertion in the CG1806-CG8002 intergenic region by five species of the virilis group is vertical transmission, and not multiple independent insertions. Therefore, the presence of a bilbo-like insertion in this region in D. littoralis and species of the virilis phylad, in contrast with its absence in D. kaneko, suggests that D. littoralis is more closely related to species of the virilis phylad than to D. montana. It should be noted that traditionally, D. littoralis has been placed in the montana phylad (Spicer and Bell, 2002). Nevertheless, the relationship of D. littoralis and species of the montana and virilis phylads are far from being resolved (compare for instance Vieira, 2002, Spicer and Bell, 2002). Here we show, based on the analysis of a 7 gene dataset, that, as suggested by the presence of the bilbo-related insertion, D. littoralis is more closely related to species of the virilis phylad than to species of the montana phylad.

Our re-evaluation of the virilis group of Drosophila, as well as geological data, supports an estimated age for the fixation of the non-LTR bilbo-like retroelement insertion in the CG1806-CG8002 intergenic region of approximately 7.5 million years. Therefore the bilbo-like insertion here described is one of the oldest ones so far reported in Drosophila (see Petrov et al., 2003). Given its age, single occurrence in D. virilis, and degeneration, it is not possible to determine whether this insertion went to fixation due to positive selection or genetic drift alone. Both explanations seem plausible since, in Drosophila, non-LTR retroelement fixations due to both drift (Petrov et al., 2003) and positive selection (Petrov et al., 2003; Schlenke and Begun, 2004; Aminetzach et al., 2005; Lipatov et al., 2005) have been described. As shown here, there is at least one very old fixation of a TE in the virilis group of species. It is important to determine how many such TE fixations are present in species of the virilis group in order to establish the generality of the observations made so far in species of the melanogaster group.

In the virilis group of species, the bilbo-like insertion is mostly being slowly lost through accumulation of small deletions. This conclusion is supported by several lines of evidence: a) only three smaller than 100 bp alignment gaps are needed to align the bilbo-like sequences from D. lummei, and D. novamexicana. The estimated divergence for this species pair is around 3 million years; b) the bilbo-like insertion shows a similar size in all D. americana individuals expected to have the Xc inversion; c) the same is true for D. virilis where all individuals have a bilbo-like copy of similar size.

Extensive DNA loss is, nevertheless, also observed in D. littoralis where large deletions seem to have deleted most of the bilbo-like insertion. 7.5 million years were thus enough to delete most of the element in one lineage, but not in the other. It is also remarkable that 67% of the D. americana individuals without the Xc inversion no longer show any evidence for a bilbo-like insertion in the CG1806-CG8002 intergenic region. In this species and D. novamexicana the presence of a PCR amplification product of about 1400 bp (when using the primers BilboF/BilboR) is a molecular marker for the Xc inversion. For how long a given TE insertion remains at a given location depends largely on the frequency of large deletions.

Acknowledgments

This work was funded by the project POCI/BIA-BCM/59399/2004, funded by POCI 2010, co-funded by FEDER funds. R. M-H. is funded by the FCT fellowship SFRH/BPD/24329/2005. The fly collection was funded by APBRF. We thank Raquel Molarinho for technical assistance.

Appendix A. Supplementary data


References
