Inferring the evolutionary history of *Drosophila americana* and *Drosophila novamexicana* using a multilocus approach and the influence of chromosomal rearrangements in single gene analyses

RAMIRO MORALES-HOJAS, CRISTINA P. VIEIRA and JORGE VIEIRA

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

Abstract

The evolutionary history of closely related organisms can prove sometimes difficult to infer. Hybridization and incomplete lineage sorting are the main concerns; however, genome rearrangements can also influence the outcome of analyses based on nuclear sequences. In the present study, DNA sequences from 12 nuclear genes, for which the approximate chromosomal locations are known, have been used to estimate the evolutionary history of two forms of *Drosophila americana* (*Drosophila americana americana* and *Drosophila americana texana*) and *Drosophila novamexicana* (*virilis* group of species). The phylogenetic analysis of the combined data set resulted in a phylogeny showing reciprocal monophyly for *D. novamexicana* and *D. americana*. Single gene analyses, however, resulted in incongruent phylogenies influenced by chromosomal rearrangements. Genetic differentiation estimates indicated a significant differentiation between the two species for all genes. Within *D. americana*, however, there is no evidence for differentiation between the chromosomal forms except at genes located near the X/4 fusion and Xc inversion breakpoint. Thus, the specific status of *D. americana* and *D. novamexicana* is confirmed, but there is no overall evidence for genetic differentiation between *D. a. americana* and *D. a. texana*, not supporting a subspecific status. Based on levels of allele and nucleotide diversity found in the strains used, it is proposed that *D. americana* has had a stable, large population during the recent past while *D. novamexicana* has speciated from a peripheral southwestern population having had an ancestral small effective population size. The influence of chromosomal rearrangements in single gene analyses is also examined.

Keywords: chromosomal rearrangements, *Drosophila americana*, *Drosophila novamexicana*, molecular phylogenetics

Received 23 December 2007; revision received 25 March 2008; accepted 7 April 2008

Introduction

Two of the main goals of systematic biology are describing biodiversity (delimiting species) and reconstructing the evolutionary history of organisms. However, these aims are not always simple to accomplish, especially in the case of closely related taxonomic groups (Wiens & Penkrot 2002). This is because events of hybridization with DNA introgression or incomplete lineage sorting can result in incongruent phylogenies obscuring the interpretation of species evolutionary histories (Buckley et al. 2006). Recombination is another feature exclusive of nuclear genomes that can lead to confounding effects when reconstructing phylogenies, although Posada & Crandall (2002) showed that when the sequences involved are closely related, the most common history is normally recovered. Also, many nuclear genomes show chromosomal rearrangements in the form of fusions or inversions that affect levels of gene flow through a reduction of recombination levels in heterozygous individuals (Noor et al. 2001; Rieseberg 2001; Navarro &
Barton 2003). This can result in population differentiation that could eventually lead to a speciation event (Rieseberg 2001; Navarro & Barton 2003). Thus, chromosomal rearrangements are expected to have an effect on the phylogenetic reconstruction of closely related species as well as within species, and contradictory results between single-locus analyses could be the consequence of the location of the genes nearby rearranged regions of the chromosomes.

Drosophila americana and Drosophila novamexicana are good models to test the possible effects of chromosomal rearrangements in phylogenetic reconstruction of closely related taxa. They are two closely related species that belong to the Drosophila virilis group of species (Throckmorton 1982). Both species are native to the USA, D. americana inhabiting the Central and Eastern regions from the South (Texas to the states around the Gulf of Mexico) to the North of the country (from Montana to Maine) (Patterson & Stone 1952), while D. novamexicana has a more restricted distribution in the Southwest (New Mexico, Arizona, Colorado and Eastern Utah) (Patterson & Stone 1949). The two species differ in their karyotype; D. americana has a fixed fusion of chromosomes 2 and 3 while D. novamexicana retained the ancestral karyotype where chromosomes 2 and 3 are unfused (Hughes 1939; Patterson 1941). There are also further karyotype differences between the two species in the frequency of some polymorphic chromosomal inversions. Also, within D. americana, two groups have been traditionally recognized according to the presence (Drosophila americana americana) or absence (Drosophila americana texana) of an X/4 chromosomal fusion. This X/4 fusion is likely maintained by selection, being the fusion frequent in the North of the distribution and rare in the South, and in heterozygous conditions over a wide geographical intermediate range (Vieira et al. 2001, 2003, 2006; McAllister 2002). Interestingly, there are populations of D. americana that have chromosomal rearrangements in common with D. novamexicana while other populations of D. americana show the alternative arrangement. For example, the inversion Xc is fixed in D. novamexicana and almost fixed in the Northern populations belonging to D. a. americana while it is almost absent in the Southern populations belonging to D. a. texana (Hsu 1952). Given that chromosomal rearrangements can reduce recombination levels to the point of promoting population differentiation and possibly speciation, it is reasonable to expect a priori that if a single gene that lies near the inversion Xc, for example, is used to reconstruct the evolutionary history of these taxa, D. novamexicana will cluster with D. a. americana and D. a. texana will show some degree of reciprocal monophyly with respect to D. a. americana. Earlier DNA studies have yielded inconsistent results on the relationships of these taxonomic units. And yet, the chromosomal location of the different DNA markers used in these studies has not been comprehensively taken into account. Using single mitochondrial DNA (mtDNA) sequences per taxon, Spicer & Bell (2002) results suggested that D. americana is a paraphyletic species with D. novamexicana being more closely related to D. a. americana than this to D. a. texana. Other studies using multiple sequences from nuclear genes, oskar and period, resulted in two divergent D. novamexicana monophyletic lineages within a paraphyletic D. americana clade (Hilton & Hey 1996, 1997). These two clades of D. novamexicana could represent either two differentiated populations (e.g. cryptic species) or be the result of balancing selection at the two loci studied (Hilton & Hey 1996, 1997). However, the presence of two divergent clades could also be the result of the position of the genes near chromosomal rearrangements that prevented normal levels of recombination in these genes between the two clades. Yet, there is no description of segregating inversions in D. novamexicana (Hsu 1952; Throckmorton 1982), thus, there is no reason to think that recombination in these genes has been reduced due to inversions. Recent phylogenetic analysis of two mitochondrial genes did not support the presence of two D. novamexicana lineages and a single monophyletic clade was placed within a paraphyletic D. americana. Furthermore, the two chromosomal forms of D. americana did not show any significant differentiation while D. novamexicana and D. americana showed significant differentiation (Caletka & McAllister 2004). Using microsatellites, Orsini et al. (2004) obtained a tree in which D. novamexicana was again monophyletic and more closely related to D. a. texana than the two forms of D. americana to each other. Furthermore, their analyses showed significant genetic differentiation between D. a. americana and D. a. texana at genome-wide level. This result is surprising given other earlier studies at population level that had not shown any significant genetic differentiation between D. a. americana and D. a. texana at several nuclear genes, with exception of those found at the base of the X chromosome (McAllister 2002; Vieira et al. 2003, 2006). A second recent analysis of microsatellites with a more comprehensive population survey showed no significant genetic differentiation between D. a. americana and D. a. texana (Schafer et al. 2006). These inconsistent results can be the consequence of chromosomal rearrangements.

The present study intends to directly evaluate the effect of chromosomal rearrangements in the phylogenetic reconstruction of closely related taxa using two closely related species with well-characterized karyotypes. Multiple loci distributed throughout the entire genome are used and results are directly correlated to the known chromosomal location of the different genes. It is also intended to understand the evolutionary history of these taxa.
Materials and methods

Samples

The Drosophila americana isofemale lines used were established from individuals captured in different geographical populations (Niobrara, Nebraska, NN97; Platte River, Duncan, Nebraska, DN; Gary, Illinois, G96; Monroe, Louisiana, ML97; Lone Star, Texas, LP97; Puxico, Missouri, PM99; Lake Ashbaugh, Arkansas, LA99; Howell Island, Missouri, HI99; Floodgate Park, Arkansas, FP99). Given the positions of the genes in the Drosophila virilis genome, we can locate them in the D. americana chromosome map using the approach described in Vieira et al. (2006) and determine if they lie near a chromosomal rearrangement breakpoint or not (shown in Table 1). Genes on chromosomes 2 and 3 are either near the chromosomes 2/3 fusion, which is fixed in D. americana and absent in Drosophila novamexicana (Hsu 1952), or relatively far from inversion breakpoints to be affected by these. With respect to rearrangements in chromosome X, all samples used in the present study were males so as to avoid heterozygosity. The status of the X/4 fusion for the lines used had been previously established by cytological observation (Vieira et al. 2001, 2003). Thus, those samples from northern lines, G96 (G96.11, G96.21, G96.36, G96.47 and G96.48) and NN97 (NN97.2, NN97.4, NN97.8 and NN97.9), had the X/4 fusion and associated Xc inversion (being thus, Drosophila americana americana), while those samples used from the ML97 (ML97.3, ML97.4.2 and ML97.5) and LP97 (LP97.7) southern strains had unfused chromosomes 4 and X (being thus Drosophila americana texana). Regarding inversions in chromosome 4, samples used in the present study were tested for the presence or absence of inversion 4ab with the described marker BbrPI (McAllister 2003). All samples were homozygous for BbrPI- (standard chromosome 4) with the exception of NN97.2 that was BbrPI+/BbrPI- (4std/4ab) and C96.21 that also had the inversion. For these latter samples, sequences from genes on chromosome 4 were obtained from the 4ab chromosome. More details on the additional sequences used for the individual genes analyses can be found in Vieira et al. (2001, 2003) and McAllister (2002, 2003). Individuals belonging to D. novamexicana strains (1031.00, 1031.04, 1031.08 and 1031.12) were obtained from the Tucson stock centre. The chromosomal formula of D. novamexicana is Xabc 2abc 3a 4a and 5b, and no variation has been found (Hsu 1952). As outgroup, D. virilis strain 1051.87 was used.

DNA extraction, gene amplification and sequencing

DNA was extracted from individual flies using a standard phenol–chloroform method followed by an ethanol precipitation. Twelve nuclear genes were used in the present study, which were chosen because of their location relative to chromosomal rearrangements in D. americana (see Table 1). Each gene was amplified using specific primers designed using oligo version 1.4 (National Biosciences, Inc.) based on the D. virilis genome (sequenced recently by the Agencourt Bioscience Corporation and made available at the internet site http://genome.ucsc.edu/cgi-bin/hgGateway).

Table 1 Summary of the details of the gene sequences used in the present study. ‘Chromosome + rearrangement’ column shows the chromosome in which the gene is found and the rearrangements near it; N is the number of individuals analysed; H and π are the haplotype and nucleotide diversity, respectively.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Chromosome + rearrangement</th>
<th>Drosophila americana americana</th>
<th>Drosophila americana texana</th>
<th>Drosophila novamexicana</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>π N</td>
<td>H coding</td>
<td>Noncoding</td>
</tr>
<tr>
<td>CG3223</td>
<td>2; 2/3</td>
<td>9</td>
<td>0.80</td>
<td>0.0018</td>
</tr>
<tr>
<td>ebony</td>
<td>2; —</td>
<td>9</td>
<td>0.80</td>
<td>0.0092</td>
</tr>
<tr>
<td>CG13444</td>
<td>3; basal</td>
<td>9</td>
<td>0.83</td>
<td>0.0024 n.a.</td>
</tr>
<tr>
<td>transformer</td>
<td>3; —</td>
<td>26</td>
<td>1</td>
<td>0.0162</td>
</tr>
<tr>
<td>period</td>
<td>X; —</td>
<td>11</td>
<td>1</td>
<td>0.0061 n.a.</td>
</tr>
<tr>
<td>elav</td>
<td>X; —</td>
<td>20</td>
<td>0.99 n.a.</td>
<td>0.0223</td>
</tr>
<tr>
<td>yp1</td>
<td>X; Xc</td>
<td>11</td>
<td>1</td>
<td>0.0045</td>
</tr>
<tr>
<td>fused1</td>
<td>X; Xc; X/4</td>
<td>25</td>
<td>0.0029</td>
<td>0.0059</td>
</tr>
<tr>
<td>timeless</td>
<td>4; —</td>
<td>66</td>
<td>0.98</td>
<td>0.0099 n.a.</td>
</tr>
<tr>
<td>lim3</td>
<td>4; —</td>
<td>8</td>
<td>0.9</td>
<td>0.0293</td>
</tr>
<tr>
<td>bib</td>
<td>4; 4ab</td>
<td>67</td>
<td>0.93</td>
<td>0.0039</td>
</tr>
<tr>
<td>CG18397</td>
<td>4; 4ab</td>
<td>9</td>
<td>0.80</td>
<td>0.0037</td>
</tr>
<tr>
<td>Total</td>
<td>--</td>
<td>9</td>
<td>0.0052</td>
<td>0.0218</td>
</tr>
</tbody>
</table>
Details of gene amplifications are available upon request. Polymerase chain reaction (PCR) products were directly sequenced in the case of the X-linked genes following purification from the agarose gel using the QIAquick Gel Extraction Kit from QIAGEN. PCR products from those genes not in the X chromosome were extracted from the gel using the QIAEX II agarose gel extraction kit (QIAGEN) and then cloned with the TA cloning kit from Invitrogen. Positive colonies were picked randomly, grown in 5 mL of Luria-Bertani with ampicillin and plasmids were extracted using QIAprep Spin Miniprep Kit from QIAGEN. Three clones were sequenced for each sample to account for PCR misincorporations. Cycle sequencing was performed using ABI BigDye version 1.1 (Applied Biosystems) chemistry and reactions consisted of a first denaturation step at 96 °C for 2 min and 30 s followed by 25 cycles of 30 s at 96 °C, 15 s at 50 °C and 4 min at 60 °C. Primers used for sequencing were the specific primers used in each gene amplification. Sequencing products were run at StabVida Inc.

Sequences obtained in the present study have been deposited in GenBank and have accession nos EU545829–EU545978. Despite numerous attempts, the lim3 sequence for strain NN97.9 and the sequence of elav from strain ML97.5 were not possible to obtain due to unknown reasons. Thus, the data corresponding to these genes from these two strains were treated as missing in the total evidence analyses. For the genes transformer (tra), period, elav, yp1, fused1, timeless, lim3 and bib, additional sequences from other studies and available in the GenBank were included in the analyses (McAllister & McVean 2000; Vieira et al. 2001, 2003, 2006; McAllister 2003; Maside et al. 2004). Sequences were checked for reading errors with bioedit version 5.0.9 (Hall 1999) before being aligned by eye with PROSEQ version 2.9 (Filatov 2002).

**Phylogenetic and network analyses**

Maximum-likelihood (ML) (Felsenstein 1981) and maximum-parsimony (MP) (Farris 1970) phylogenetic analyses were run in **PAUP** version 4.0b10 (Swofford 2002) for the combined data set and for each gene separately. Bayesian analyses were also run in **MrBayes** 3.1.2 (Ronquist & Huelsenbeck 2003) with the combined data set partitioned according to genes. Concatenated sequences were obtained from the same strain, and mostly from the same individual. Before the total evidence analyses were run, a homogeneity partition test, the incongruence length difference test (ILD) (Mickevich & Farris 1981; Farris et al. 1994, 1995), was performed as implemented in **PAUP** version 4.0b10 to check for possible incongruent phylogenetic signals among the 12 gene partitions. One thousand replicates were performed. Recombination within species will dissociate loci histories within species rendering the ILD uninformative. However, given the a priori taxonomic status of the lines, it could be the case that recombination would not obscure totally the intertaxa phylogenetic signal and that all genes would still recover two clades corresponding to *D. novamexicana* and *D. americana*. Furthermore, if the subspecific status reflects some degree of differentiation (despite previous studies that show no differentiation at several nuclear loci, see Introduction), it would also be reasonable to expect that with enough sequence data, the estimated phylogeny will recover the subspecies as monophyletic clades. Thus, by running the ILD test, it was intended to examine if there was congruent phylogenetic signal in each gene to recover the monophyly of the taxonomic units.

For the ML and Bayesian analyses, the evolutionary models that best fitted our data were found using the Akaike information criterion (AIC) as implemented in **MODELLTEST** 3.7 (Posada & Crandall 1998). In the ML analysis, heuristic searches were run with the starting tree obtained via stepwise addition, with random addition of sequences and 10, 50 or 100 replicates depending on the computational weight. For the total evidence analysis, 50 replicates of the random addition of sequences were run. Tree-bisection–reconnection was used as the branch-swapping algorithm. Maximum-parsimony heuristic tree searches were run with tree-bisection–reconnection as the branch-swapping algorithm. The starting tree was obtained via stepwise addition with random addition of sequences and 10, 50 or 100 replicates were used depending on the computational burden. For the total evidence analysis, 100 replicates of the random addition of sequences were run. All characters were treated as unordered and given equal weights, and gaps were treated as missing data. In the Bayesian analysis, 10 × 10^6 generations with four chains each (one cold and three heated chains) were set up. Trees were sampled every 100th generation and the first 25 000 trees were discarded as ‘burn-in’. The remaining trees were used to compute the Bayesian posterior probabilities of each clade of the consensus tree.

To test the robustness of the trees, 1000 bootstrap (Felsenstein 1985) replicates were run and zero-branch-length tests were performed in **PAUP** version 4.0b10. The Bremer support or decay index (DI) (Bremer 1994) and the partition Bremer support (PBS) (Baker & DeSalle 1997; Baker et al. 1998) were also estimated for the total evidence consensus MP tree to measure levels of support of each gene partition for the different nodes using **TREEROT** version 2.0 (Sorenson 1999). In the PBS analysis, a heuristic search under parsimony criterion was run, with the starting tree obtained via stepwise addition, random sequence addition and 100 replicates.

The Shimodaira–Hasegawa test, implemented in **PAUP** version 4.0b10, and the approximately unbiased test (Shimodaira 2002), performed with **conse**l (Shimodaira &
Hasegawa (2000), were used to evaluate two alternative tree topologies, the ML phylogeny obtained with the combined data set and a phylogeny in which *D. americana* and *D. novamexicana* were forced to show reciprocal monophyly.

Statistical parsimony networks were estimated using tcs version 1.21 (Clement et al. 2000) for the mitochondrial data of Caletka & McAllister (2004) and each nuclear gene used in this study. Networks were estimated with the 95\% parsimony connection limit and treating indels as missing data with the exception of *ebony* for which the network was also estimated using a connection limit set to 90\% in order to obtain a fully connected network.

Genetic differentiation, genetic diversity and estimation of time of divergence

The degree of taxa differentiation was evaluated for each gene by calculating the $F_{ST}$, $K_{ST}$ (Hudson et al. 1992) and $S_{nn}$ (Hudson 2000) parameters with proseq version 2.9 (Filatov 2002) and dnaplus version 4.10.8 (Rozas et al. 2003). The significance of the subdivision parameters was evaluated with a permutation test (1000 replicates) as performed in these programs. These different genetic differentiation indexes were used to ensure that differentiation was detected independently of the assumptions of each test and the characteristics of our sequence data. Thus, $F_{ST}$ is most appropriate when diversity is low (Nagylaki 1998), $K_{ST}$ was introduced as a more powerful test for sequence data and low sample sizes, while $S_{nn}$ uses a nearest-neighbour statistic approach and has been described to perform well under a wide range of sample sizes and levels of variation (Hudson 2000). The nucleotide ($\pi$) and haplotype diversity ($H$) were also estimated for intron and exons separately in each gene using dnaplus version 4.10.8.

The age of divergence between *D. americana* and *D. novamexicana* was estimated for each gene independently. Since the sample size of *D. novamexicana* was small, it is more probable to overestimate the differentiation between the two species if the derived (using *D. virilis* sequence as outgroup) mutations in *D. novamexicana* are included in the calculations. This is because some of the apparent fixed derived nucleotides in *D. novamexicana* could be polymorphic. This sample size bias can be corrected by using only the synonymous derived mutations present in the *D. americana* branch. The synonymous divergence ($K_s$) between *D. americana* and *D. virilis* multiplied by the number of synonymous sites analysed and divided by two gave the number of synonymous mutations in the *D. americana* lineage since the separation from *D. virilis* about 5 million years ago (Caletka & McAllister 2004). Then it was calculated the expected time of divergence between *D. americana* and *D. novamexicana* using the number of fixed derived nucleotides in *D. americana* with respect to *D. novamexicana*.

The $K_s$ and the number of synonymous sites were determined using dnaplus version 4.10.8.

A second approach to estimate the age of divergence between *D. americana* and *D. novamexicana* was also used based on Stephens & Nei (1985). The number of nucleotides per site for each species and the net nucleotide difference per site between both species were estimated using formulae (1) and (3) and using the concatenated genes. This method corrects for intraspecific polymorphism before speciation time, and thus for shared ancestral polymorphisms.

Results

Sequence diversity

Gene sequence details and genome localizations are summarized in Table 1 and the polymorphic sites of each gene are available in Tables S1–S12, Supplementary material. Nucleotide and haplotype diversity ($\pi$ and $H$, respectively) of the taxa under study at each gene are also shown in Table 1. Overall levels of genetic diversity are much lower in *Drosophila novamexicana* (0.0018 and 0.0024 in coding and noncoding regions, respectively) than in either of the chromosomal forms of *Drosophila americana* (*Drosophila americana* americana 0.0052 and 0.0218 and *Drosophila americana* texana 0.0067 and 0.0318 in coding and noncoding regions, respectively). This could be argued to be a bias due to the smaller sample size; however, *D. a. texana* with similar number of individuals sequenced has $\pi$ values considerably higher. There are two exceptions: CG13444 shows similar values of $\pi$ between *D. a. americana* and *D. novamexicana*, while *D. a. texana* shows higher level of diversity; period shows similar values of $\pi$ in both, *D. americana* and *D. novamexicana*. In the case of CG13444, given that there is no differentiation between both forms of *D. americana*, the reduction of diversity in *D. a. americana* is most likely a random effect of the sample. In the case of period, it is the nonsynonymous diversity of *D. novamexicana* (0.0047) that shows an increased level with respect to that in *D. americana* (0.0020 and 0.0017 in *D. a. americana* and *D. a. texana*, respectively).

Phylogenetic relationships of the taxa

The total length of the combined data set alignment was of 8126 nucleotides. The possible conflict between the phylogenetic signal of the different nuclear genes was tested with the ILD. Result of the 1000 partition-homogeneity replicates was significant ($P = 0.001$), thus indicating a conflict between the signals of the 12 genes. Since one of the purposes of the analysis was to evaluate the monophyly of the three taxa (*D. novamexicana, D. a. americana* and *D. a. texana*) and the conflict between gene signals could
come from within species/subspecies recombination (see below), we proceeded with the total evidence approach, although individual gene phylogenies and networks were also evaluated. Nevertheless, this significant result suggested that interspecific phylogenetic signal was obscured by a stronger intraspecific recombination signal.

Phylogenetic analysis using MP of the combined data set resulted in two most parsimonious trees of a length of 1234 (the strict consensus is shown in Fig. 1a). The number of variable characters was 817 of which 265 were parsimony informative. The trees had a consistency index (CI) of 0.6994 (CI excluding uninformative characters 0.4379, rescaled CI 0.3882), and a retention index (RI) of 0.5552. In the consensus tree, the two species, *D. americana* and *D. novamexicana*, form two exclusive, highly supported monophyletic sister clades supporting their separation.
as different species. However, there is no evidence in the tree for the existence of two differentiated clades within *D. americana* corresponding to the chromosomal forms *D. a. americana* and *D. a. texana*. Furthermore, the relationships of the strains within *D. americana* show low bootstrap support values suggesting little within species resolution. This lack of resolution is most likely the result of recombination between genomes.

The total evidence ML analysis was run with settings corresponding to the GTR + I + G substitution model, which best fitted the data. The parameters of the model were set to nucleotide frequencies of A = 0.2778, C = 0.2391, G = 0.2320 and T = 0.2511, six nucleotide substitution types with rate parameters of R(A-C) = 0.6315, R(A-G) = 1.9424, R(A-T) = 1.3467, R(C-G) = 0.7786, R(C-T) = 2.7637 and R(G-T) = 1, an assumed proportion of invariable sites of R(G-T) = 1, an assumed proportion of invariable sites of 0.7762, and a gamma distribution of variable sites with rate parameters of R(A-C) = 0.6315, R(A-G) = 1.9424, R(A-T) = 1.3467, R(C-G) = 0.7786, R(C-T) = 2.7637 and R(G-T) = 1, an assumed proportion of invariable sites of 0.7762, and a gamma distribution of variable sites with shape parameter \( \alpha \) of 0.6703. The score of the best tree found was \(-\ln \text{Likelihood} = 18624.67859\) (Fig. 1b). The obtained ML tree differs from the MP tree in the monophyly of *D. americana* with respect to *D. novamexicana*. Two basal clades are present in the ML tree, one that includes some individuals of *D. a. americana* and *D. a. texana*, and a second clade containing *D. novamexicana* and the remaining *D. americana*. Within this latter clade, *D. novamexicana* appears monophyletic and with high zero-branch-length test support. However, this basal split of *D. americana* into two clades is not supported by zero-branch-length tests. In addition, this ML tree is not significantly different from a tree constrained to have *D. americana* as monophyletic sister clade to *D. novamexicana* (\(-\ln \text{Likelihood} = 18624.68981\), Shimodaira–Hasegawa test \( P = 0.484\), approximately unbiased gave a \( P = 0.517\) for the constrained tree vs. a \( P = 0.483\) for the ML tree). Relationships between *D. americana* strains are better supported by the zero-branch-length test than by bootstrap in the MP analysis (Fig. 1b). Nevertheless, three clades within *D. americana* (NN97.2, G96.47) (G96.11, LP97.7) and (ML97.3, ML97.4.2) were highly supported in both analyses. It is significant that a strain from *D. americana* (G96.11) is closest to a *D. a. texana* (LP97.7).

The partitioned Bayesian analysis was run with independent model settings for each gene (data not shown). The consensus of the 75001 trees showed a topology very similar to that of the MP analysis (Fig. 1c). Therefore, *D. novamexicana* and *D. americana* formed two reciprocal monophyletic clades (with 100% and 71% posterior probabilities for the *D. novamexicana* and *D. americana* clades, respectively). The *D. americana* clade is further subdivided into two groups but not corresponding to chromosomal forms. Again, G96.11 is placed together with LP97.7 with high posterior probability (PP 100%), and G96.48 is placed with high PP as basal to the clade containing other X/4 unfused lines.

### Partition Bremer support analysis and gene genealogies

In order to evaluate the support of each individual gene to the monophyly of *D. americana* and *D. novamexicana*, a partition Bremer support analysis was run. The supports of each gene partition to the different nodes of the strict consensus MP phylogeny are shown in Table 2. All genes support the exclusive monophyly of *D. novamexicana* (node B) with the exception of *fused1* and *transformer* that showed PBS values of 0 and –0.5, respectively. The phylogeny and network of *fused1* (19 parsimony-informative sites), when a larger data set was used, showed
that three lines of *D. novamexicana* (nova00, nova04 and nova08) were identical and monophyletic with a sample of *D. a. americana* (G96.11); the other *D. novamexicana* sample (nova12) was identical to a *D. a. texana* (LA99.54.11) and one nucleotide different from the other *D. novamexicana* lines (Fig. 2). This sharing of alleles explains the lack of support for an exclusive monophyletic clade of *D. novamexicana*. Network analysis puts the *D. novamexicana* alleles in an ancestral position (nova12 and LA99.54.11 had the highest outgroup probability), which is supported by the presence of ancestral characters at the variable positions with respect to *D. americana* (data not shown).

Twenty-one of the 25 lines identified as having the Xc/4 fusion (*D. a. americana*, marked in grey in Fig. 2) showed two alleles differing by two nucleotides (Fig. 2). The other four lines of the X/4 fusion form showed two other alleles (one allele represented twice in the sample) exclusive of this chromosomal form, and a shared allele with *D. novamexicana*.

The MP analysis of *transformer* (22 parsimony-informative sites) resulted in a gene tree with no resolution (tree not shown). The ML phylogeny, however, showed the *D. novamexicana* sequences as basal and paraphyletic with significant zero-branch-length test although no bootstrap support (Fig. 3). The basal position of *D. novamexicana* in this gene is corroborated by the fact that the nucleotides that differentiate this species from *D. americana* are all ancestral in *D. novamexicana* (Table 4). Furthermore, network analysis also resulted in the positioning of *D. novamexicana* alleles as ancestral with the highest outgroup probability (results not shown). Network analysis also showed a great level of homoplasy within *D. americana* most likely the result of normal levels of recombination. Since this gene is located towards the distal end of chromosome 3, and no chromosomal rearrangement is known to be near it, it is expected that recombination occurs at normal levels. The negative PBS value of *transformer* with respect to the monophyly of *D. novamexicana* is probably the result of this lack of phylogenetic resolution (in the case of MP) or the paraphyly shown by the ML analysis.

The exclusive monophyly of *D. americana* with respect to *D. novamexicana* (node N in Fig. 1a) was not supported by
five genes, which gave negative PBS values. Most of them showed low negative values (from –0.2 to –1), however, *ebony* gave a strong negative value (–9). When *ebony* was used alone in phylogenetic analyses, both methods of reconstruction (ML and MP) resulted in a basal polytomy with three major clades (Fig. 4a). One of them is highly supported by bootstrap and zero-branch-length tests and corresponds to *D. novamexicana*. The other two clades comprise the *D. americana* lines. Thus, there is no support from *ebony* for an exclusive monophyletic clade of *D. americana*. This gene is highly variable (Table 1; 46 parsimony-informative sites) and the statistical parsimony analysis resulted in a series of unconnected subnetworks at the 95% connection limit. In order to have a connected network, the connection limit had to be set to 90%, but the number of ambiguities increased greatly (Fig. 4b). This gene has been involved in the differentiation in colouration between *D. novamexicana* and *D. americana* (Wittkopp et al. 2003). However, there are also differences in colouration within *D. americana* so that *D. a. texana* flies are darker, and the western branch *americana* populations are still lighter than the eastern populations (Throckmorton 1982). Nevertheless, there is no evidence for *D. americana* chromosomal forms or populations clustering in the tree nor in the network.

It is interesting to note that apart from *ebony*, the other genes giving negative PBS values to the monophyly of *D. americana* are located on chromosomes X and 4, whose fusion and associated Xc inversion differentiate the northern and southern populations. Of these, *yp1* phylogenetic analyses (42 parsimony-informative sites) resulted in a clear division between the two subspecies (Fig. 5). Furthermore, the *D. novamexicana* clade was embedded within the *D. a. americana* lines and its monophyly was not highly supported by bootstrap. *Timeless* also shows negative PBS values for almost all nodes, being the only exception the node corresponding to the common ancestry of *D. novamexicana*. There was little resolution within the *D. americana* clade, with almost no bootstrap support for any clustering. The most likely explanation for this result is that levels of recombination within this gene are expected to be normal given its chromosomal location, which together with its higher level of polymorphism (28 parsimony-informative sites) makes the phylogenetic signal less clear in *D. americana*.

**Genetic differentiation**

To evaluate whether the taxa showed genetic differentiation, the *F*$_{ST}$, *K*$_{ST}^{*}$, and *S*$_{mn}$ values for each pair of taxa were estimated for every gene (Table 3). Results of the three tests demonstrate that *D. novamexicana* and *D. americana* are differentiated taxa. All comparisons of *D. novamexicana* with each of the chromosomal forms of *D. americana* were significant, although those with *D. a. texana* showed lower levels of significance (*P* < 0.05), but this is likely the result of the few samples being compared. Comparisons of the two chromosomal forms of *D. americana* resulted in nonsignificant values for most of the genes indicating a lack of differentiation. Only *yp1* and *fused1* showed highly significant differentiation in all three estimates between *D. a. americana* and *D. a. texana*. Furthermore, the only three
fixed differences over all genes between the two forms are found in yp1. These genes are located in the chromosomal region between the basal breakpoint of inversion Xc and the fusion point of chromosomes X and 4, two linked chromosomal arrangements that are almost fixed in D. a. americana and entirely (X/4 fusion) or almost (inversion Xc) absent in D. a. texana. Thus, reduced levels of recombination between the fused, inverted (D. a. americana) and nonfused, noninverted (D. a. texana) chromosomes have allowed the differentiation between these forms. Nevertheless, fused1 and yp1 also show a significant differentiation between D. novamexicana and D. a. americana even though they share the inversion Xc. This further reinforces the specific status of D. novamexicana. In this respect, it is also

Fig. 4 Strict consensus of five MP trees (a) and statistical parsimony network (b) obtained with ebony sequences. On branches of (a) are shown the bootstrap values for the different nodes. Sample names and colour-coding follow those of Fig. 1.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd
noteworthy that the number of fixed differences in \( \text{yp1} \) between \( D. \text{novamexicana} \) and \( D. \text{a. americana} \) was lower than that found with respect to \( D. \text{a. texana} \) or between the two chromosomal forms of \( D. \text{americana} \) (Table 4). This could be the result of the lower sample size of \( D. \text{a. texana} \) than of \( D. \text{a. americana} \), but other genes with identical sample sizes do not show such differences. Significant genetic differentiation between the two chromosomal forms of \( D. \text{americana} \) at either one or two of the three estimates were obtained in \( \text{CG13444}, \text{transformer} \) and \( \text{CG18397} \) genes, however, the significance was marginal around the 5% limit. Moreover, there were no fixed differences between the two forms in these genes.

**Populational history and times of divergence**

As indicated by Avise (2000), a combination of the \( H \) and \( \pi \) can be used to infer populational history. As shown in Table 1, \( D. \text{novamexicana} \) shows relatively high \( H \) and low \( \pi \) values indicating that the population is in the process of growing from an ancestral reduced effective population size. This could be the case if \( D. \text{novamexicana} \) speciated...
recently from a peripheral, isolated population of *D. americana*. The case of *D. americana* is the opposite. Both chromosomal forms show, in most of the genes, high $H$ and $\pi$ values, suggesting a stable and large population effective size for a considerable period of time. Nevertheless, the pattern is different in *D. a. americana* for some genes, such as those two that showed differentiation between the chromosomal forms. In the case of *fused1* and *yp1*, $\pi$ was reduced compared to *D. a. texana* while having relatively high $H$ values. This is again an indication of a population growing from an ancestral small size (Avise 2000). Given that values for the other genes suggest a large, stable population, results from these two genes are most likely showing the recovery of nucleotide diversity after its loss. As discussed above, these genes are located in the region between the fusion point of X chromosome and the basal breakpoint of inversion Xc, and were likely affected by both rearrangements and/or the effects of a recent selective sweep that would have eliminated their nucleotide diversity (Vieira et al. 2006; Maside & Charlesworth 2007).

In order to estimate the age of divergence between the two species, it has to be assumed a molecular clock. Nevertheless, the molecular-clock hypothesis was rejected by the likelihood-ratio test [$-\ln L(\text{clock}) = 18643.57196$; $-\ln L(\text{no clock}) = 18624.67859$, $\chi^2 = 3391.40934$, d.f. = 16, $P = 0.0002$], indicating that not all sequences are evolving at a homogeneous rate along all branches of the phylogeny. This was also corroborated by the fact that the majority of the fixed differences between *D. novamexicana* and *D. americana* were derived in the former species (Table 4). Assuming an equal rate of occurrence of new mutations along the branches leading to each species, and being 24 the number of parsimony-informative mutations that could be classified as ancestral or derived (using *D. virilis* as outgroup), it would be expected that approximately 12 of these derived mutations should be found in each of the species. However, the observed numbers of derived mutations in *D. novamexicana* and *D. americana* were 21 and 3, respectively, which is significantly different from the neutral expectation ($q^2 = 13.5$, d.f. = 1, $P = 0.0002$). This could be an indication of an increase of the evolutionary rate in *D. novamexicana*. Nevertheless, it cannot be ruled out the possibility of the result being biased due to the smaller sample size of *D. novamexicana*.

Given these known limitations, we still proceeded to estimate the ages of divergence based on synonymous rates of evolution. Based on the $K_s$ between *D. americana* and *D. virilis* and the number of synonymous, derived mutations of *D. americana* with respect to *D. novamexicana* (in the case of those genes that showed significant differentiation between the chromosomal forms, the age was estimated for each form separately), the age of the cladogenesis event was estimated to have occurred between 833 000 and 266 000 years ago (Table 4). It was only possible to estimate the time of divergence in the case of *yp1* and *lim3* since in the other 10 genes there were no fixed synonymous, derived mutations in *D. americana* with respect to *D. novamexicana*.

Table 4  Estimated times of divergence based on the fixed number of derived mutations in *Drosophila americana* with respect to *Drosophila novamexicana*. $N_f$ and $N_d$ indicate the number of fixed differences (all and silent, respectively) between *D. americana* and *D. novamexicana* (in brackets the derived mutations in *D. novamexicana*). $L_s$ is the number of silent sites in each gene (including *Drosophila virilis*, *D. americana* and *D. novamexicana*). $K_s$ is the silent site divergence estimates for *D. americana* and *D. virilis* (am-vir) and *D. americana* and *D. novamexicana* (am-nov). In the Time column are the estimates of the age of divergence between *D. americana* and *D. novamexicana*; n.a. indicates that it was not possible to estimate the age of divergence given the lack of derived mutations in the *D. americana* lineage.

<table>
<thead>
<tr>
<th>Genes</th>
<th>$N_f$</th>
<th>$N_d$</th>
<th>$L_s$</th>
<th>Time (My)</th>
<th>am-vir</th>
<th>am-nov</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG3223</td>
<td>0</td>
<td>0</td>
<td>288.35</td>
<td>0.05191</td>
<td>0.0603</td>
<td>n.a.</td>
</tr>
<tr>
<td>ebony</td>
<td>4 (4)</td>
<td>4 (4)</td>
<td>242.04</td>
<td>0.09673</td>
<td>0.05154</td>
<td>n.a.</td>
</tr>
<tr>
<td>CG13444</td>
<td>4 (3)</td>
<td>2 (2)</td>
<td>139.23</td>
<td>0.04549</td>
<td>0.01748</td>
<td>n.a.</td>
</tr>
<tr>
<td>transformer</td>
<td>1</td>
<td>0</td>
<td>306.19</td>
<td>0.08625</td>
<td>0.01148</td>
<td>n.a.</td>
</tr>
<tr>
<td>period</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>170.88</td>
<td>0.18387</td>
<td>0.03199</td>
<td>n.a.</td>
</tr>
<tr>
<td>elav</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>520.00</td>
<td>0.18571</td>
<td>0.02750</td>
<td>n.a.</td>
</tr>
<tr>
<td>yp1†</td>
<td>1 (1)†</td>
<td>7 (4)†</td>
<td>310.75</td>
<td>0.09761†</td>
<td>0.11467†</td>
<td>0.04095†</td>
</tr>
<tr>
<td>fused1†</td>
<td>0†</td>
<td>0†</td>
<td>194.00</td>
<td>0.11036†</td>
<td>0.11619†</td>
<td>0.01621†</td>
</tr>
<tr>
<td>timeless</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>103.41</td>
<td>0.10573</td>
<td>0.03801</td>
<td>n.a.</td>
</tr>
<tr>
<td>lim3</td>
<td>9 (7)</td>
<td>9 (7)</td>
<td>583.00</td>
<td>0.12922</td>
<td>0.04487</td>
<td>0.266</td>
</tr>
<tr>
<td>bib</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>232.13</td>
<td>0.11889</td>
<td>0.03152</td>
<td>n.a.</td>
</tr>
<tr>
<td>CG18397</td>
<td>0</td>
<td>0</td>
<td>411.17</td>
<td>0.08867</td>
<td>0.02235</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

†Calculated only taking into account the X/4-fused form *Drosophila americana americana*; ‡calculated using only the nonfused form *Drosophila americana texana*. 

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd
The method of Stephens & Nei (1985) would also correct for the smaller sample size of *D. novamexicana* since it corrects for ancestral polymorphisms present before speciation. Using this method, the intraspecific number of nucleotide substitution per site (d) would be 0.0142 and 0.0020 for *D. americana* and *D. novamexicana*, respectively. Given this intraspecific variability, the net number of nucleotide differences between both species (δam-nov) was estimated to be 0.0097. We used the average d between *D. americana* and *D. virilis* to estimate the rate of nucleotide substitution per site per year using the 5-million-year divergence time between both species. Given that these species have been evolving for a considerable period of time, it is expected that the intraspecific polymorphism of each will not bias the mutation rate (λ, in Stephens & Nei 1985) calculation, estimated to be of 1.25 × 10^{-8} per year. Given δam-nov and λ, it was estimated that *D. americana* and *D. novamexicana* diverged 0.388 million years ago. This date is similar to that estimated above using lim3.

**Discussion**

*The taxonomic status of Drosophila americana and Drosophila novamexicana*

Extensive debate regarding the taxonomic status of *Drosophila americana* and *Drosophila texana* and their evolutionary relationship with *Drosophila novamexicana*, has been going on in recent times (Tominaga & Narize 1995; Hilton & Hey 1996, 1997; Schlötterer 2000; McAllister 2002; Spicer & Bell 2002; Vieira et al. 2003, 2006; Caletka & McAllister 2004; Orsini et al. 2004; Schäfer et al. 2006). Despite these studies obtained some incongruent results, it is now generally accepted that the two subspecies of *Drosophila americana* are in fact chromosomal forms and not differentiated taxa. However, all studies resulted in the apparent paraphyly of *D. americana* with respect to *D. novamexicana*. This lack of reciprocal monophyly was suggested to be the result of a large effective population size and high sequence diversity in *D. americana* which obscured the independence of the *D. novamexicana* lineage (Caletka & McAllister 2004). When using multiple loci, gene histories are expected to be concordant between species and discordant within species (with respect to species exclusivity, following de Queiroz 1998). However, all species have most likely gone through a period of nonexclusive gene genealogies at some point in their history due to incomplete lineage sorting, even if they showed diagnostic morphological characters (Neigel & Avise 1986). This scenario could be common in those cases where a new species arises as a distinct small range population (i.e., peripheral population) from a species with a larger geographical range and population size (Wiens & Penkrot 2002). In such cases, the new species will become exclusive relatively fast while the ancestral, larger species will retain a paraphyletic status (e.g., Funk et al. 1995; Talbot & Shields 1996; Hedin 1997; Paetkau 1999). Nevertheless, in the case of *D. novamexicana* and *D. americana*, we argue that even though there is evidence for incomplete lineage sorting (*D. novamexicana* shares alleles with *D. americana* in some genes which are not likely the result of recent introgression), the reciprocal monophyly of these species is recovered with enough DNA data. Thus, the requirement of reciprocal monophyly for the recognition of these two taxa as genealogical species (Shaw 1998, 2001) is met, contrary to what had been previously suggested (Caletka & McAllister 2004).

Having established in this study the monophyly of *D. novamexicana* using nuclear genes, it is noteworthy the contrasting results obtained in other studies using the genes *period* and *oskar* (Hilton & Hey 1996, 1997). These suggested the existence of two divergent, polyphyletic clades of *D. novamexicana*, which were explained by either the existence of two cryptic species or the existence of balancing selection maintaining two divergent clades (Hilton & Hey 1996). New sequence data obtained for *period* in the present investigation, using four of the strains that fell in different clades, does not support the division of *D. novamexicana* into two clades. There is no evidence for genetic differentiation within *D. novamexicana* either. Nevertheless, it is noteworthy that the nonsynonymous diversity in *D. novamexicana* is higher than in *D. americana*, which could support the balancing selection hypothesis of Hilton & Hey (1996).

Traditionally, *D. americana* is being divided into two subspecies according to the presence of a fusion of chromosomes X and 4 in northern populations (*D. a. americana*) and its absence in the southern ones (*D. a. texana*) (Throckmorton 1982). However, recent analyses have established that the fusion is associated to a latitudinal cline, existing flies heterozygous for the X/4 fusion (Vieira et al. 2001; McAllister 2002), and there is no evidence for a genetic differentiation between the forms (Caletka & McAllister 2004; Schäfer et al. 2006), except at genes located at the base of the X chromosome (Vieira et al. 2006; Maside & Charlesworth 2007). The present study further supports these findings. Phylogenetic and network analyses give no evidence for the existence of two monophyletic, independent clades corresponding to the chromosomal forms. Also, population genetic tests result in a lack of significant genetic differentiation, with the exception of two genes, *yp1* and *fused1*. Nonetheless, these genes reflect the different evolutionary history of a chromosomal rearrangement (the inversion *Xc* and maybe the X/4 Fusion) rather than a real subspecific status. This is in accordance to what other investigations have argued (Vieira et al. 2001, 2003, 2006; McAllister 2002). Thus, it is concluded that the *D. a. americana* and the *D. a. texana* should be treated as suggested...
before as chromosomal forms of the same species rather than subspecies, a taxonomic status that gives the idea of a genetic differentiation.

**Evolutionary history of D. americana subphylad**

The evolutionary histories of *D. novamexicana* and *D. americana* appear to be different. Results show that *D. americana* has had a large, stable population and there is no evidence that indicates any recent significant reduction in population size. Large genetic diversity combined with also large haplotype diversity is an indication of such a populational history (Avise 2000). Nevertheless, two genes (*fused1* and *yp1*) show a relatively reduced genetic diversity combined with a high haplotype diversity in *D. a. americana*. This is a signature that could be associated with a recent bottleneck (Avise 2000) and can be related to the occurrence of the Xc inversion and/or X/4 fusion in the northern populations of *D. americana* (Vieira et al. 2006). Chromosomal rearrangements are unique events that occur in single chromosomes reducing to zero the genetic and haplotype diversities of the rearranged chromosome segments affected. As individuals with the rearranged chromosomes increase in numbers in the population, so do the haplotype and genetic diversities, although the recovery of the former occurs faster. These two genes are located in the chromosome X segment that lies between the two rearrangements, thus making it difficult to dissociate the effects of each event on the genes.

Conversely, there is evidence in the results for a reduced ancestral population size in *D. novamexicana*. Nevertheless, and given that haplotype diversity is not completely recovered for half of the genes, it is possible to suggest that the actual effective population size of this species remains moderate.

The results here presented are consistent with a recent cladogenesis event of *D. novamexicana* and *D. americana*. The age of divergence between these two species was estimated to be 0.866 million years and 0.266 million years for *fused1* and *lim3*, respectively, with one method and 0.388 with Stephens & Nei (1985) method. Although the rejection of a molecular clock in the present analysis makes these estimates questionable, they are nonetheless consistent with other ages of divergence estimated formerly (0.380 million years, Caletka & McAllister 2004). Furthermore, other lines of evidence support the recent age of these species such as the estimated age for the inversion Xc (0.308 million years, Vieira et al. 2006). This inversion is fixed in *D. novamexicana* and present in 95% of the X/4 chromosomes in the northern populations of *D. americana* (X/4 fusion form), although almost absent (7%) in the southern *D. americana* (non-X/4 fusion form) (Hsu 1952). The most parsimonious scenario for the presence of this rearrangement in these two taxa would be to assume that it arose previous to the speciation event. In addition, the reduced number of fixed differences between *D. americana* and *D. novamexicana*, and the presence of common alleles in some genes, is also consistent with the recent separation of the two taxa.

The restricted geographical location of *D. novamexicana* to the western slopes of the Rocky Mountains suggests that *D. novamexicana* evolved from a marginal, peripheral population of the *D. americana*/*D. novamexicana* ancestral species. The mid- to late-Pleistocene is a period characterized by pluvial–interpluvial cycles in Southwest North America that have influenced the evolution and diversification of many North American species (e.g. Ayoub & Riechert 2004; Castoe et al. 2006). Apart from the divergence of *D. novamexicana* and *D. americana*, there is no evidence for geographical structuring within these species which is consistent with other studies in *D. americana* (Tominaga & Narize 1995; Hilton & Hey 1996, 1997; McAllister 2002; Vieira et al. 2003, 2006; Caletka & McAllister 2004; Schäfer et al. 2006). Thus, the effect of the pluvial–interpluvial period of the Pleistocene appears to be restricted to the most Southwestern population of the ancestral species. It is also evident from the results that *D. novamexicana* evolved in complete isolation from *D. americana* since there is no conclusive evidence for introgression of *D. americana* alleles. Overall, three alleles were shared by both species, one in CG18397 and two in *fused1*. The two common alleles shared in *fused1* show a basal position (one being the most ancestral allele) in the network and phylogeny that could suggest that they are the result of incomplete lineage sorting rather than of hybridization with introgression. The shared allele in CG18397 does not show an ancestral position in the network, although it is the most basal allele within *D. novamexicana*, the other two alleles differing by one and two mutations. There is no evidence for natural hybrids between both species, and there is no evidence either for mitochondrial introgression in those lines used in the present study and previous ones (Spicer & Bell 2002). These would suggest that the presence of common alleles is due to an incomplete lineage sorting. Furthermore, it has been pointed out that the continental divide has been an effective barrier to gene flow during the Pleistocene in a number of organisms (Morafka et al. 1992). Nevertheless, it cannot be ruled out the possibility of rare hybridization events occurring or having occurred. Further data would be required in order to discriminate between the two possibilities.

**Effects of chromosomal rearrangements in phylogenetic reconstruction**

Chromosomal rearrangements are very common in the genomes of the genus *Drosophila*, in which more than half...
of its species show them and their frequency in natural populations is high (Powell 1997). They are thought to promote reproductive isolation within and between species by reducing recombination along large sections of the genome (e.g. Dobzhansky 1949; Navarro & Barton 2003). Using D. americana and D. novamexicana as model species, we have evaluated the effects that known chromosomal rearrangements can have on the evolutionary history reconstruction of closely related species. Recombination has the effect of maintaining the cohesion of the species. Thus, Paterson (1985) defined the species as a ‘field for recombination’. As expected, when recombination is suppressed, alleles will differ between the chromosomal forms (inverted and standard) and the reconstruction of the evolutionary history of the taxa involved will reflect this in the phylogeny. This is observed in some of the genes used in the present study. But even more, if a new species arises from a population with a given chromosomal inversion, when the speciation event is recent as in the case of D. novamexicana and D. americana, the new species will cluster together with the alleles from the chromosome with the rearrangement, and the sister species will appear as paraphyletic. The phylogeny of fused1, for example, shows a clustering (and reduction of diversity) of the majority of the alleles from X/4 fused chromosomes, which could be explained by different events. Fused1 is located between the basal breakpoint of inversion Xc and the base of the fused X chromosome. These two rearrangements (Xc inversion and X/4 fusion) occurred independently, although are tightly linked, and it has been argued that the Xc inversion arose before the fusion (Vieira et al. 2006). Since rearrangements appear on single chromosomes, the nucleotide variability is reduced to zero in the rearranged portion of the chromosome and to some extent around the breakpoints. As the rearrangements start to increase in number in the population, variability levels start to recover. However, it is not clear that these alone are affecting fused1 variability levels since it has been argued that a more recent independent selective sweep could also have affected variability levels of fused1 (Vieira et al. 2001, 2006). The phylogenetic analyses of yp1 also resulted in a clear division between the two chromosomal forms D. a. americana and D. a. texana, with the D. novamexicana clade clustering with the D. a. americana lines. This gene is found near the breakpoint of inversion Xc (Vieira et al. 2006), which is fixed in D. novamexicana and present in 95% of D. a. americana and almost absent in D. a. texana (Hsu 1952). Therefore, the basal division of the phylogeny is reflecting presence and absence of the inversion rather than the presence or absence of the X/4 fusion. If yp1 was affected by the presence or absence of the X/4 fusion, D. novamexicana should appear together with D. a. texana lines since they share the lack of fused chromosomes. Thus, previous phylogenetic and phylogeographical analyses giving controversial results can be better understood with chromosomal data of the taxa under study.

Acknowledgements

This work was funded by the project POCI/BIA-BCM/99399/2004, funded by POCI 2010, cofunded by FEDER funds. R. M-H. is funded by a FCT fellowship number SFRH/BPD/24329/2005. We thank Raquel Molarinho for technical assistance.

References

Bremer K (1994) Branch support and tree stability. Cladistics, 10, 295–304.
Funk DJ, Futyma DJ, Ortí G, Meyer A (1995) A history of host associations and evolutionary diversification for Ophiura...
Supplementary material

The following supplementary material is available for this article:

Table S1 Polymorphic sites in CG3223. Dots represent same nucleotide as in Drosophila virilis and hyphens represent indels.

Table S2 Polymorphic sites in ebony.

Table S3 Polymorphic sites in CG13444.

Table S4 Polymorphic sites in transformer.

Table S5 Polymorphic sites in period.

Table S6 Polymorphic sites in elav.

Table S7 Polymorphic sites in yp1.

Table S8 Polymorphic sites in fused1.

Table S9 Polymorphic sites in timeless.

Table S10 Polymorphic sites in lim3.

Table S11 Polymorphic sites in bib.

Table S12 Polymorphic sites in CG18397.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03796.x

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.