Changes in the Recombinational Environment Affect Divergence in the yellow Gene of Drosophila

Agustí Munté, Montserrat Aguadé, and Carmen Segarra

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

The complete coding region of the yellow (y) gene was sequenced in different Drosophila species. In the species of the melanogaster subgroup (D. melanogaster, D. simulans, D. mauritiana, D. yakuba, and D. erecta), this gene is located at the tip of the X chromosome in a region with a strong reduction in recombination rate. In contrast, in D. ananassae (included in the ananassae subgroup of the melanogaster group) and in the obscura group species (D. subobscura, D. madeirensis, D. guanche, and D. pseudoobscura), the y gene is located in regions with normal recombination rates. As predicted by the hitchhiking and background selection models, this change in the recombinational environment affected synonymous divergence in the y-gene-coding region. Estimates of the number of synonymous substitutions per site were much lower between the obscura group species and D. ananassae than between the species of the obscura group and the melanogaster subgroup. In fact, a highly significant increase in the rate of synonymous substitution was detected in all lineages leading to the species of the melanogaster subgroup relative to D. ananassae lineage. This increase can be explained by a higher fixation rate of mutations from preferred to unpreferred codons (slightly deleterious mutations). The lower codon bias detected in all species of the melanogaster subgroup relative to D. ananassae (or to the obscura group species) would be consistent with this proposal. Therefore, at least in Drosophila, changes in the recombination rate in different lineages might cause deviations of the molecular-clock hypothesis and contribute to the overdispersion of the rate of synonymous substitution. In contrast, the change in the recombinational environment of the y gene has no detectable effect on the rate of amino acid replacement in the Yellow protein.

Introduction

The recombination rate may differ considerably across the genome of a particular species. In Drosophila melanogaster, comparison of the cytological and the recombinational (or genetic) maps reveals important differences in the coefficient of exchange in different chromosomal regions (see Lindsley and Zimm 1992). An extremely low recombination rate has been detected in the centromeric regions, at least in some telomeric regions and in the dot-like element. In addition, interspecific differences in the recombination rates of particular chromosomal regions have also been detected between D. melanogaster and the closely related species D. simulans and D. mauritiana (True, Mercer, and Laurie 1996) and between D. melanogaster and D. yakuba (Takano-Shimizu 1999). However, a very strong suppression of recombination seems to be limited to regions around the telomeres and the centromeres, although this effect cannot be generalized to all chromosomal arms.

The rate of recombination in a certain genomic region has been shown to strongly affect the level of nucleotide polymorphism in that region (see Begun and Aquadro 1992; Aguadé and Langley 1994; Aquadro, Begun, and Kindahl 1994). In fact, a very low nucleotide diversity has been detected in regions with a strong reduction in the recombination rate, such as those located very near the telomere and the centromere of the X chromosome (Aguadé, Miyashita, and Langley [1989] and Langley et al. [1993], among others) or in the dot-like element (Berry, Ajioka, and Kreitman 1991) of D. melanogaster. This observation is consistent with the predictions of the hitchhiking (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989) and the background selection models (Charlesworth, Morgan, and Charlesworth 1993). Both models also predict an effect of the recombination rate on the fixation rate of mildly selected mutations (Birky and Walsh 1988; Charlesworth 1994). According to these models, an increase in the fixation rate of slightly deleterious mutations and a decrease in the fixation rate of slightly advantageous mutations is expected in regions with no recombination relative to regions with normal levels of recombination. These predictions are related to the Hill-Robertson effect (Hill and Robertson 1966), i.e., that natural selection acting on a locus affects the chance of fixation of alleles segregating at linked loci. This effect implies a relaxation of selection in regions with very low recombination, which can consequently be viewed as a reduction in the effective population size in regions with no recombination.

Kliman and Hey (1993) have shown that in D. melanogaster, genes located in regions with a drastic reduction in the recombination rate exhibit low codon bias. This observation is consistent with the Hill-Robertson effect, i.e., a reduced effect of selection in maintaining codon bias in regions of low recombination. The action of selection on codon usage was inferred by Akashi (1995), who reported a significantly higher ratio of polymorphism to divergence for preferred changes than for preferred changes in the D. simulans lineage. Akashi (1999) later confirmed the action of selection on synonymous sites by showing that preferred mutations segregated at higher frequencies and were more likely to become fixed than unpreferred mutations. If indeed weak selection discriminates between synonymous co-

Key words: yellow gene, melanogaster group of Drosophila, recombination rate, synonymous divergence, codon bias.

Address for correspondence and reprints: Carmen Segarra, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain. E-mail: carme@porthos.bio.ub.es.

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
ons, changes in effective population size in different lineages are expected to have an impact on the effectiveness of selection on synonymous sites in these lineages and thus on synonymous divergence. Akashi (1995) proposed that selection acting on codon bias was relaxed in the D. melanogaster lineage relative to the D. simulans lineage due to differences in the effective population sizes of both species. A similar approach was used by Takano (1998), who compared divergence between D. melanogaster and D. simulans at genes located in both species in regions with a strong reduction in recombination and at genes located in regions with normal levels of recombination.

According to these observations, changes in the recombinational environments of particular genes in different species are expected to affect their rates of nucleotide substitution, specially if in one species genes are located in a region with no recombination and in the other they are in a region with normal levels of recombination. This is the case for the y and sc genes that in D. melanogaster are located at the tip of the X chromosome in a region with a strong reduction in recombination, while in D. subobscura their location is separated from the centromere by more than one euchromatic region. Divergence at y and sc between D. melanogaster and D. subobscura (Munte´, Aguadé, and Segarra 1997) is characterized by a high number of synonymous substitutions per site and very different codon biases in the two species (much lower in D. melanogaster than in D. subobscura). As all the melanogaster subgroup species are homossexual for the tip of the X chromosome (Ashburner 1989a), the location of y and sc at the telomere can be traced back to at least 6–15 MYA, which is the proposed divergence time of these species (Lachaise et al. 1988). The observed pattern of divergence between D. melanogaster and D. subobscura could therefore be explained by an increase in the fixation rate of slightly deleterious mutations in the D. melanogaster lineage, at least during this period (Munte´, Aguadé, and Segarra 1997). However, the lack of data from an outgroup species prevented us from performing a relative-rate test of this hypothesis and also prevented us from analyzing the putative effect of the change in the recombinational environment on nonsynonymous mutations in Drosophila.

In order to perform relative-rate tests, information from a species of the melanogaster group (not included in the melanogaster subgroup) with a nontelomeric position of the y or the sc gene would be required. Inspection of the genetic map of D. ananassae (Moriwaki and Tobari 1975) showed that in this species y is not located in the telomeric region, which motivated sequencing of the y gene in this species. In fact, the pattern of divergence in y between D. ananassae and D. subobscura is expected to be different from the pattern previously detected between D. melanogaster and D. subobscura if indeed recombination affects divergence. In addition, the study was extended (Munte´ 1999) to D. simulans and D. mauritiana, species included in the melanogaster complex, and to one species of each of the other species clusters (D. yakuba and D. erecta) described in the melanogaster subgroup. All of these species that have a telomeric location of the y gene and are included in the melanogaster subgroup are expected to show a pattern of divergence similar to that detected between D. melanogaster and D. subobscura. Finally, the y gene was also sequenced in D. pseudoobscura to determine whether the pattern of divergence previously detected between D. subobscura and D. melanogaster was also shared by another species of the obscura group (not included in the subobscura cluster) in which the y gene is not located in the telomeric region (Segarra et al. 1995) and no strong reduction in the recombination rate of that region is thus expected.

The results reported here extend those recently published by Takano-Shimizu (1999) that include the partial sequencing of the y gene in different species of the melanogaster subgroup. In that study, a difference in patterns of synonymous substitution between the D. melanogaster and D. yakuba lineages was detected. This difference could not be explained by mutational bias and indicated a reduction in the efficacy of selection in the D. melanogaster lineage relative to the D. yakuba lineage. This reduction was consistent with the difference in the recombination rate detected at the tip of the X chromosome in these species, which is about one order of magnitude larger in D. yakuba than in D. melanogaster (Takano-Shimizu 1999).

Present results revealed an effect of the recombination rate on synonymous divergence in Drosophila. A significant increase in the fixation rate of synonymous mutations was detected in the lineages leading to the studied species of the melanogaster subgroup relative to the D. ananassae lineage. Therefore, drastic interspecific changes in the recombinational environments of particular genes might cause strong deviations in the constancy of the synonymous substitution rate and thus on the molecular-clock hypothesis, at least in Drosophila. In contrast, the change in the recombinational environment of the y gene has no detectable effect on the fixation rate of nonsynonymous mutations in the different lineages.

Materials and Methods
DNA Sequencing

The D. pseudoobscura line used in the present study was kindly provided by R. C. Lewontin, and those of D. ananassae, D. erecta, and D. yakuba were provided by F. Lemeunier. Genomic DNA from single males of each species was purified according to Ashburner (1989b). The complete coding region of the y gene was PCR-amplified in two fragments, one including the first exon and the other including the second exon. Primers used for this purpose in D. pseudoobscura were 5′-AACACCACGAACCACTCAAG-3′ and 5′-GCTACCACGAGTTTGTTTG-3′ for the first exon and 5′-ATTGGACGGAATTCACTGAC-3′ and 5′-GTTGTTGCTTTAAGAATTC-3′ for the second exon. The primers used to PCR-amplify the y gene region in the melanogaster group species were 5′-CGCCACCGGTCCACAGAG-3′ and 5′-GCTA-
CCGCGGATTTTGTITGG-3’ for the first exon and 5’-ACATCggCAATTTGACGCAGCAAC-3’ and 5’-ATGCTGATGATGCCAACANCC-3’ for the second exon. However, in some of the melanogaster group species these primers gave more than one amplification product despite different attempts to optimize the PCR conditions. In these cases, the nested PCR procedure (Roux 1996) was followed. After the first amplification with multiple products, an additional round of amplification was performed using internal primers that were specific for the ends of the desired fragment. This reamplification procedure produced in all cases a single PCR product suitable for sequencing. PCR products were purified with the QIAquick PCR purification kit (Qiagen) and used as templates for sequencing. Primers designed approximately every 350 nt were used to sequence both strands of the complete y-gene-coding region. For D. ananassae only, the putative last three codons of the second exon were not sequenced, as a primer covering these sites had to be used in the reamplification by the nested PCR procedure. Sequencing reactions were performed with the ABI PRISM Dye Terminator Cycle Sequencing kit (Perkin Elmer) following the manufacturer’s instructions. Sequencing products were separated on an ABI PRISM 377 (Perkin Elmer) automated sequencer.

The previously published sequences of D. melanogaster (Geyer, Spana, and Corces 1986; EMBL accession number X04427), D. subobscura (Munte, Aguadé, and Segarra 1997; EMBL accession number Y13909), and D. madeirensis and D. guanche (Munte, Aguadé, and Segarra 2000; EMBL accession numbers AJ289812 and AJ289813, respectively) were also used in this study. The newly reported sequences are deposited in the EMBL data library under accession numbers AJ300666–AJ300671.

**In Situ Hybridization**

PCR-amplified fragments including the second exon of the y-gene-coding region of each species of the melanogaster group were used as probes for in situ hybridization on the corresponding polytene chromosomes. The λDsubRA2.1 recombinant phage of D. subobscura was used as a probe for in situ hybridization on D. madeirensis and D. guanche polytene chromosomes. Probes were labeled by nick translation using 16-bio-dUTP (Roche) as the labeled nucleotide. Prehybridization and hybridization conditions were as referenced in Segarra et al. (1995).

**Sequence Analysis**

Sequences were multiply aligned with the CLUSTAL W program (Thompson, Higgins, and Gibson 1994), and subsequently the alignment obtained was optimized manually. Sequences were edited with the MacClade program (Maddison and Maddison 1992). Divergence at the y-gene-coding region was estimated by the number of synonymous differences per synonymous site (p_s) and the number of nonsynonymous differences per nonsynonymous site (p_a) as proposed by Nei and Gojobori (1986). These estimates were then corrected for multiple hits at the same site (K_s and K_a estimates, respectively) according to different models of nucleotide substitution. As divergence at nonsynonymous sites was low, the p_a estimates were directly corrected by the one-parameter model developed by Jukes and Cantor (1969). In contrast, the K_s estimates were obtained by the Li, Wu, and Luo (1985) method with the modification proposed by Li (1993). This method is specific for coding regions and assumes the two-parameter model of nucleotide substitutions (Kimura 1980) to correct for multiple hits.

Three different methods were used to estimate the extent of codon bias present at the y gene in the different species: the scaled χ^2 (Shields et al. 1988), the codon bias index (CBI) (Morton 1993), and the effective number of codons (ENC) (Wright 1990) methods. Both a higher scaled χ^2 value and a higher CBI value indicate a stronger deviation from random use of synonymous codons. In contrast, the highest ENC value (61) indicates that all synonymous codons are used equally, and the lowest value (20) indicates that only one codon is used in each synonymous class.

The relative-rate test (Sarich and Wilson 1973) was used to contrast putative deviations of the molecular clock hypothesis. This test compares the rates of substitution in two lineages and requires information for a third species that is used as the outgroup. Three different methods were applied to the y gene data set to perform the relative-rate test. Wu and Li’s (1985) method is based on the difference in the corrected estimates of the number of substitutions per site between two lineages (K_{13} - K_{23}) and its standard deviation (SD). The test statistic z ([K_{13} - K_{23}] /SD) is normally distributed when the total number of differences detected in the two lineages compared is larger than 20. We also used the method of Muse and Weir (1992) based on likelihood estimates. The evolutionary model proposed by Muse and Gaut (1994) for coding regions was used as the probabilistic model for the process of nucleotide substitution. All parameters were estimated independently for each branch. The estimate of the likelihood ratio test statistic (λ) distributed as a χ^2 with 1 df was used to assess the constancy of the substitution rate. The K2WULI and HYPHY programs were used to perform the Wu and Li (1985) and the Muse and Weir (1992) tests, respectively. The Tajima (1993) test is based on the number of observed substitutions in each of two lineages. A χ^2 test with 1 df was used to check whether these numbers differed significantly. This test was applied independently for synonymous substitutions and for amino acid replacements. When two substitutions were detected in the same codon, all possible pathways were considered in order to compute the number of synonymous substitutions. Only those cases that allowed an unambiguous classification of the substitutions were taken into account. Codons differing by three substitutions were not considered in the analysis.

Phylogenetic analysis was performed with the neighbor-joining method (Saitou and Nei 1987) using as genetic distance either the observed number of synon-
Table 1
Synonymous Divergence at the y Gene

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sub</td>
<td>0.0125</td>
<td>0.0673</td>
<td>0.3092</td>
<td>0.4928</td>
<td>0.7966</td>
<td>0.7859</td>
<td>0.8325</td>
<td>0.8332</td>
<td>0.8218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mad</td>
<td>0.0108</td>
<td>0.0748</td>
<td>0.3189</td>
<td>0.4974</td>
<td>0.7959</td>
<td>0.7865</td>
<td>0.8344</td>
<td>0.8351</td>
<td>0.8224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gua</td>
<td>0.0704</td>
<td>0.0781</td>
<td>0.3130</td>
<td>0.4841</td>
<td>0.7879</td>
<td>0.7861</td>
<td>0.8302</td>
<td>0.8309</td>
<td>0.8144</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pse</td>
<td>0.4084</td>
<td>0.4219</td>
<td>0.4101</td>
<td>0.4956</td>
<td>0.7806</td>
<td>0.7631</td>
<td>0.8152</td>
<td>0.8133</td>
<td>0.8172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ana</td>
<td>0.8458</td>
<td>0.8588</td>
<td>0.8075</td>
<td>0.8450</td>
<td>0.7461</td>
<td>0.6928</td>
<td>0.7765</td>
<td>0.7670</td>
<td>0.7586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ere</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.6973</td>
<td>0.1831</td>
<td>0.1837</td>
<td>0.1683</td>
<td>0.1784</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yak</td>
<td>2.0861</td>
<td>2.1190</td>
<td>2.1441</td>
<td>1.9619</td>
<td>1.4226</td>
<td>0.1782</td>
<td>0.2109</td>
<td>0.2033</td>
<td>0.2108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sim</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.9169</td>
<td>0.1911</td>
<td>0.2069</td>
<td>0.0286</td>
<td>0.0675</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mau</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.8382</td>
<td>0.1703</td>
<td>0.1928</td>
<td>0.0265</td>
<td>0.0649</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mel</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.7565</td>
<td>0.1864</td>
<td>0.2068</td>
<td>0.0661</td>
<td>0.0573</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—Numbers of synonymous differences per site \( (p_s) \) are shown above the diagonal, and the corrected estimates \( (K_s) \) according to Li (1993) are shown below the diagonal. NA = not applicable. Estimates in bold correspond to synonymous divergence between species of the melanogaster and obscura groups. Drosophila species names are as follows: sub, D. subobscura; mad, D. madeirensis; gua, D. guanche; pse, D. pseudoobscura; ana, D. ananassae; ere, D. erecta; yak, D. yakuba; sim, D. simulans; mau, D. mauritiana; mel, D. melanogaster.

y noname differences per site \( (p_s) \) or the corrected number of nonsynonymous substitutions per site \( (K_s) \). Sites with alignment gaps in any sequence were not considered when genetic distances were computed (complete deletion option). Neighbor-joining trees were reconstructed with the MEGA program (Kumar, Tamura, and Nei 1994). Parsimony analysis was applied either for all substitutions in the y-gene-coding region or for amino acid replacements in the Yellow protein. In both cases, sites with indels were also excluded from the analysis. The PAUP (Swofford 1998) and MacClade (Maddison and Maddison 1992) computer programs were used for parsimony analysis. Maximum-likelihood trees were also reconstructed from information at the y-gene-coding region with the quartet puzzling approach as implemented in the PUZZLE program (Strimmer and von Haeseler 1996; Strimmer, Goldman, and von Haeseler 1997) and using the Hasegawa, Kishino, and Yano (1985) model of nucleotide substitutions. This approach was also used to estimate the branch lengths of the inferred phylogeny and to test the molecular-clock hypothesis for whole trees.

Results

In Situ Hybridization

In situ hybridization analysis (not shown) confirmed the telomeric position of the y gene in all of the studied species of the melanogaster subgroup. In contrast, in D. ananassae, included in the ananassae subgroup, the y gene exhibited a nontelomeric position on the XL chromosomal arm, as expected according to the genetic map available for this species (Moriwaki and Tobari 1975). These results indicate that the recombinational environment of the y gene differs between D. ananassae and the species of the melanogaster subgroup. In fact, it has been reported that the y and scute (sc) genes are separated by 0.2 map units in D. ananassae (C. W. Hinton, cited in Moriwaki and Tobari 1975). If the organization of the y-achaeate-sc complex was the same in D. ananassae as in D. melanogaster, these estimates would support the hypothesis that the recombination rate in the y gene region was higher in the former species. In the four studied species of the obscura group, like in D. ananassae, the y gene has a nontelomeric position (Segarra et al. 1995).

Sequence Divergence

The y-gene-coding region had the same length in all the studied species of the melanogaster subgroup (1,623 sites when the stop codon was not considered). Drosophila ananassae differed from these species by three indels that were 6, 12, and 3 nt long. In this species, the y-gene-coding region was 1,638 nt long assuming that the stop codon position was conserved in all melanogaster group species. The y-coding region was 1,704 nt long in the species of the subobscura cluster (D. subobscura, D. madeirensis, and D. guanche). The D. pseudoobscura coding region differed from the other species of the obscura group by four indels (with lengths of 6, 12, 6, and 36 nt) and was 1,680 nt long.

The multiple alignment of the 10 sequences included a total of 1,749 sites. This number dropped to 1,599 when all sites with alignment gaps were excluded from the analysis. A total of 530 sites presented a different variant in at least one species, and 446 of these sites were parsimony-informative.

Table 1 shows the distance matrix with the estimates of the number of synonymous differences per synonymous site \( (p_s) \) above the diagonal and the corrected estimates \( (K_s) \) according to Li (1993) below the diagonal after excluding all sites with gaps from the analysis. \( p_s \) estimates between the obscura group species and D. ananassae ranged between 0.4841 and 0.4956. In contrast, estimates between the obscura group species and those included in the melanogaster subgroup were much higher and ranged between 0.7631 and 0.8351. Among the lowest estimates were those found in comparisons of the obscura group species and D. yakuba (from 0.7631 to 0.7865). Consequently, no corrected estimates \( (K_s) \) could be obtained for most comparisons between species of the obscura group and the melanogaster subgroup, as the method of correction for multiple hits involves logarithmic functions that are inapplicable when the argument is zero or negative.

As divergence at synonymous sites is negatively correlated with codon bias (Sharp and Li 1989), the de-
Divergence in the y Gene in Drosophila

Relative-Rate Tests

The average synonymous divergence at y between D. ananassae and the obscura group species (K_s = 0.84) was similar to the average divergence reported for 24 genes sequenced in D. melanogaster and either D. pseudobscura or D. subobscura (K_s = 0.81; Zeng et al. 1998). On the other hand, our results indicated that divergence at synonymous sites between the obscura group species and D. ananassae was lower than that between species of the obscura group and those of the melanogaster subgroup despite the divergence time being the same in both comparisons. The contrast between both observations could be explained by an acceleration of the synonymous substitution rate at y in the lineages of the melanogaster subgroup species. To further investigate this result, relative-rate tests were performed to contrast the constancy of the substitution rate in the different lineages. Table 4 summarizes the results obtained by the relative-rate tests proposed by Wu and Li (1985) and by Muse and Weir (1992) when all substitutions in the y-gene-coding region were considered. All analyses were performed in sets of three species, of which one was the outgroup. When the rates of substitution in two lineages were compared, their closest relative among the species studied was chosen as the outgroup. However, in comparisons between the D. yakuba or the D. erecta lineages and lineages of the other species of the melanogaster subgroup, D. ananassae was used as the outgroup, as the gene tree inferred from divergence in the y gene region (see below) differs from the phylogenetic relationships generally accepted for the species of the melanogaster subgroup (Lachaise et al. 1988). On the other hand, any of the obscura group species could be used as the outgroup in comparisons between D. ananassae and the melanogaster subgroup species. Although table 4 shows only the results when D. subobscura was the outgroup, similar results were obtained with any other obscura group species. A highly significant difference in the rate of substitution (P < 0.001) was detected in the lineage of D. ananassae relative to any of the lineages of the melanogaster subgroup species by either the genetic distance method (Wu and Li 1985) or the maximum-likelihood approach (Muse and Weir...
The negative sign of the \( K_{13} - K_{23} \) difference indicated a higher substitution rate in the lineages of the melanogaster subgroup species. In addition, significant deviations were detected (0.05 > \( P > 0.01 \)) in the \( D. yakuba \) lineage relative to those of the other species of the melanogaster subgroup except in one of the comparisons performed with the likelihood approach. Even in this case, the probability value (\( P = 0.0525 \)) was close to the critical value. The observed deviation could be explained by a lower substitution rate in the \( D. yakuba \) lineage. Finally, none of the other tests performed were significant. In summary, a highly significant increase in the rate of substitution was detected in the \( y \)-gene-coding region in the melanogaster subgroup species relative to \( D. ananassae \), and within this subgroup a significant increase was also detected in the lineages of the melanogaster complex species and \( D. erecta \) relative to \( D. yakuba \).

In order to identify which kind of substitution was responsible for the detected deviations of the constancy in the nucleotide substitution rate, the Tajima (1993) test was applied independently for synonymous substitutions and amino acid replacements (table 4). In all cases, the observed deviations could be attributed to differences in the fixation rate of synonymous mutations, as none of the tests performed for replacement substitutions detected any deviation from a constant rate of substitution.

Therefore, the performed relative-rate tests indicated a significant increase in the fixation rate of synonymous mutations in all lineages of the melanogaster subgroup species relative to \( D. ananassae \) and, to a lesser extent, of the \( D. melanogaster, D. simulans, D. mauritiana \), and \( D. erecta \) lineages relative to \( D. yakuba \).

**Phylogenetic Analysis**

Figure 1a shows the parsimony tree obtained when all substitutions in the \( y \)-gene-coding region were considered after excluding from the analysis those sites with indels. The tree was reconstructed using the obscura group species as outgroups. All nodes were supported by bootstrap percentages higher than 95%. The maximum-likelihood and neighbor-joining trees based on nucleotide substitutions were similar to the parsimony tree (results not shown). Branches leading from the root to any of the melanogaster subgroup species were unusually long. In fact, when the root of the tree was positioned at the midpoint of the longest branch, \( D. ananassae \) clustered with the obscura group species instead of with the other species of the melanogaster group. The neighbor-joining tree based only on synonymous differences gave a similar result, which indicates that differences in branch length are mainly due to synonymous substitutions. This conclusion was further supported by the comparison of the trees reconstructed from nucleotide substitutions in the \( y \)-gene-coding region (fig. 1a) and from amino acid replacements at the Yellow protein (fig. 1b). In this tree, the numbers of substitutions that accumulated from the ancestor of the melanogaster group to any of the species in this group including \( D. ananassae \) were much more alike, as expected from the equivalent divergence time. A similar result was obtained with the neighbor-joining method using \( K_a \) as the genetic distance (result not shown).

The molecular-clock hypothesis for the whole tree inferred from nucleotide substitutions was tested with the maximum-likelihood approach. Constancy of substitution rates could be rejected from the value of the
Absolute Substitution Rates

The melanogaster subgroup species after the split of D. melanogaster were considered (λ = 36.27; P = 0.0000). In contrast, the molecular-clock hypothesis could not be rejected either for the obscura group species (λ = 1.12, P = 0.5702) or for the melanogaster subgroup species (λ = 2.02, P = 0.5684).

Absolute Substitution Rates

The phylogenetic analysis (fig. 1a) indicated that most substitutions accumulated in the branch leading to the melanogaster subgroup species after the split of D. ananassae. This result suggested an episodic acceleration of the substitution rate in the stem lineage of the melanogaster subgroup species. Following the approach of Friedrich and Tautz (1997), the absolute substitution rate in this lineage was estimated and compared with the absolute rates in other lineages. The maximum-likelihood branch length of the stem lineage was 0.1722 (equivalent to 223 substitutions in fig. 1a). The proposed divergence times for the melanogaster group and the melanogaster subgroup species are 17–20 Myr and 6–15 Myr, respectively (Lachaise et al. 1988). According to the upper limits of these estimates, the absolute substitution rate in the stem lineage turned out to be 34.4 × 10⁻⁹ substitutions per site per year (15.7 × 10⁻⁹ substitutions per site per year when the lowest estimates of the divergence time were considered). In contrast, the absolute rate for the species of the melanogaster cluster (D. melanogaster, D. simulans, and D. mauritiana) that diverged 2.5 MYA (Lachaise et al. 1988) was 3.7 × 10⁻⁹ substitutions per site per year. This estimate was quite similar to that obtained for the D. ananassae lineage assuming a divergence time of 20 Myr (4.0 × 10⁻⁹ substitutions per site per year). Therefore, the absolute substitution rate in the stem lineage of the melanogaster subgroup species increased about 4–10 times relative to the other lineages.

Discussion

The obscura group species and D. ananassae share a nontelomeric position of the y gene region, in contrast to the melanogaster subgroup species, for which the gene maps at the tip of the X chromosome. According to the most parsimonious criterion, the y gene region would have been positioned very near the telomere before the split of the melanogaster subgroup species but after the split of D. ananassae. This change in the chromosomal location of the y gene (likely due to a paracentric inversion) caused a drastic change in its recombinational environment, since, in the melanogaster subgroup species, it moved to a region with a putative strong reduction in the recombination rate. According to the differences in recombination rate across the X chromosome of D. melanogaster (Kliman and Hey 1993), the rate at the y gene region may differ by even more than two orders of magnitude between the melanogaster subgroup species and both D. ananassae and the obscura group species. Although spontaneous recombination has been detected in D. ananassae males (Moriwaki and Tobari 1975), this characteristic is not expected to affect the recombination rate for X-linked loci.

As predicted by the hitchhiking and background selection models, this change in the recombinational environment may affect the fixation rate of mildly selected mutations in this gene region (Birky and Walsh 1988; Charlesworth 1994). Specifically, if synonymous and/or nonsynonymous mutations were weakly selected, nucleotide divergence at genes located in regions with a strong reduction in the recombination rate would be affected. As it is generally accepted that in the genus Drosophila, weak selection is acting on synonymous mutations (Kliman and Hey 1993; Akashi 1995; Powell and Moriyama 1997; Comeron, Kreitman, and Aguadé 1999), such an effect can be expected at least on synonymous divergence.
Recombination and Synonymous Substitutions

Results from the present study strongly support the hypothesis that the recombination rate affects nucleotide divergence. First, estimates of synonymous differences per site between the obscura group species and D. ananassae were much lower than those between the obscura group and the melanogaster subgroup species. Second, the relative-rate test showed an increase in the rate of fixation of synonymous mutations in all lineages leading to the melanogaster subgroup species relative to that of D. ananassae. Third, codon bias of the \( y \) gene was much higher in the obscura group species and D. ananassae than in the melanogaster subgroup species. Fourth, the gene tree showed that the branches leading from the root to any of the melanogaster subgroup species were unusually long when the root of the tree was positioned between the obscura and the melanogaster species groups.

All these results are concordant with a relaxation of selection acting on synonymous mutations in the lineage leading to the melanogaster subgroup species. This reduction in the effectiveness of selection has contributed to an increase in the fixation rate of mutations from preferred to unpreferred codons (slightly deleterious mutations) in this lineage and thus to a decrease in codon bias in the species of the melanogaster subgroup. The telomeric location of the \( y \) gene in these species in a region with a strong reduction in the recombination rate can explain the detected increase in the fixation rate of synonymous mutations.

In addition, within the melanogaster subgroup species, an increase in the rate of synonymous substitution was detected in the lineages leading to D. melanogaster, D. simulans, D. mauritiana, and D. erecta relative to D. yakuba. This result is consistent with that of Takano-Shimizu (1999), who reported that the recombination rate at the tip of the \( X \) chromosome in D. yakuba was about one order of magnitude higher than that in D. melanogaster. However, the estimated recombination rate in the D. yakuba lineage was still much lower than the recombination rates that D. melanogaster exhibited in the central portion of the \( X \) chromosome. This difference in the recombination rate at the tip of the \( X \) chromosome causes a difference in the pattern of synonymous substitutions in the \( y \) gene region between the D. melanogaster and the D. yakuba lineages that cannot be explained by mutational bias (Takano-Shimizu 1999). In fact, the pattern of synonymous substitutions indicates a reduction in the efficacy of selection in the D. melanogaster lineage relative to the D. yakuba lineage consistent with the higher recombination rate in the latter lineage.

Other alternative explanations could also account for the observed results. According to the nearly neutral mutation hypothesis (Ohta 1992), a higher rate of fixation of slightly deleterious mutations is expected in lineages with lower effective population sizes. Consistent with this hypothesis, a fast rate of molecular evolution has been detected in the Hawaiian Drosophila lineages that apparently have experienced severe bottlenecks during their evolutionary histories (DeSalle and Templeton 1988; Ohta 1993). Therefore, lower effective population sizes of the melanogaster subgroup species (not so important for D. yakuba) relative to D. ananassae and the obscura group species would explain the observed results. Alternatively, a shift in the mutation pressure in a particular lineage may also contribute to an acceleration of molecular evolution in that lineage. For instance, the episodic acceleration of the rDNA substitution rate detected in the stem lineage of the Diptera was explained by a change in the directional mutation pressure (Friedrich and Tautz 1997).

Changes in the effective population size or in the mutational bias in particular lineages are expected to affect all genes of the genome, and not only those that have changed their recombinational environment. Unfortunately, few nuclear genes have been analyzed in the same set of species reported here, since although the melanogaster subgroup species have been extensively studied, most of these studies have not included D. ananassae. Only the studies of the D1 and D2 domains of the 28S rDNA gene (Pélandakis, Higgins, and Solignac 1991) and of the Amy gene family (Inomata, Tachida, and Yamazaki 1997) have included this species. In the former study, the gene tree did not reflect any acceleration in the fixation rate of mutations in the lineages of the melanogaster subgroup species, although D. ananassae clustered with the obscura group species but with very low bootstrap support. In the Amy study, a completely different result was obtained, since the phylogram based on synonymous substitutions showed a long branch leading to the ananassae subgroup species relative to the branches leading to the species of the melanogaster subgroup or to the obscura group species. In addition, codon bias is much lower in D. ananassae (G+C content at third codon positions is 73.4\%) than in the species of the melanogaster subgroup and the obscura group (mean value = 91.3\%). In fact, this is the expected result for a gene located in a region of low recombination in D. ananassae but not in the other species of the melanogaster group and those of the obscura group. The organization of the Amy gene family in D. ananassae is complex, as it includes two main clusters with different gene copies (Da Lage et al. 1992; Da Lage, Maczkowiak, and Cariou 2000). Interestingly, one of these clusters maps very close to the base of the 3L chromosomal arm in a region that very likely exhibits a strong reduction in recombination (centromeric effect). Different arguments discussed by Da Lage, Maczkowiak, and Cariou (2000) suggest the ancestral character of the centromeric cluster. A cluster with two Amy genes (Amy3 and Amy4) with a centromeric location is also present in D. kikkawai (Inomata and Yamazaki 2000), a species of the montium subgroup included in the melanogaster group. A low codon bias and an increase in the fixation rate of synonymous mutations in the lineage leading to these genes was also reported in the latter study. These observations would again be consistent with the effect of the different recombinational environments on gene divergence.
Consequently, the pattern of divergence in the y gene seems to be specific of this locus and is consistent with an effect of the recombination rate on synonymous divergence, at least in the genus Drosophila. This effect was clearly detected in genes with strong differences in recombination rate in different lineages. It is expected to decrease with divergence and to be weaker the lower the difference in recombination rate. In addition, interspecific differences in the recombinational environment might cause deviations of the molecular-clock hypothesis and contribute to the overdispersion of the fixation rate of synonymous mutations.

The proposed effect of recombination on synonymous divergence in Drosophila seems to contradict previous results where no correlation was found between divergence and recombination rate either for genes with different recombination rates across the genome of D. melanogaster and D. simulans (Begun and Aquadro 1992) or for genes that had changed their recombinational environments in these species, like the anon1E9 locus (Schmidt et al. 1999). It is likely that the divergence time between D. melanogaster and D. simulans is not long enough for the effect of recombination on synonymous divergence to be detectable.

Recombination and Amino Acid Replacements

In contrast to synonymous substitutions, the change in the recombination rate of the y gene region does not seem to affect divergence of the Yellow protein in the different lineages. This is a rather conserved protein, and the number of amino acid replacements detected in pairwise comparisons was generally low. That might have affected the power of Tajima’s (1993) test. However, even in those cases where the number of substitutions was high, and therefore the power of the test was also high, Tajima’s test did not detect any departure from constancy of amino acid replacement rates (table 4).

This result suggests that rates of amino acid replacement are insensitive to differences in the effective population size of the y gene region due to changes in the recombination rate. Zeng et al. (1998) also detected that in Drosophila, unlike in mammals (Gillespie 1989; Ohta 1995), rates of amino acid replacement were not overdispersed. They concluded that effective population sizes in Drosophila were large enough for most nonsynonymous mutations to be effectively deleterious and therefore they do not become fixed.

There are, however, some reported examples in Drosophila where weak selection seems to be acting on nonsynonymous mutations. In this sense, an excess of nonsynonymous polymorphisms relative to nonsynonymous fixed differences has been detected in different mitochondrial genes of Drosophila where there is no recombination (Kaneko et al. 1993; Ballard and Kreitman 1994; Rand, Dorfman, and Kann 1994; Rand and Kann 1996) and also in a few nuclear genes located in regions with rather low recombination rates (Simmons et al. 1994; Hamblin and Aquadro 1997). There might be no contradiction between both sets of results, as selection decreases the probability of fixation of weak deleterious mutations more effectively than it decreases the probability that these mutations segregate within populations (Kimura 1983). In fact, a reduction in the recombination rate (and consequently in the effective population size) of a particular gene region might cause a longer persistence in the populations of slightly deleterious mutations segregating at low frequency and therefore having a low probability of becoming fixed. Weak selection acting on deleterious mutations would therefore more likely be detected in regions of low recombination (and effective population size) by the polymorphism to divergence ratio than by changes in the fixation rate.

On the other hand, the fractions of slightly deleterious nonsynonymous mutations may differ substantially among different genes depending on the functional constraints of the encoded protein. As the Yellow protein is a rather conserved protein, purifying selection against nonsynonymous mutations in the y gene might be strong. Information from genes with higher nonsynonymous substitution rates than y would help to ascertain whether in those cases the change in the effective population size caused by a change in the recombinational environment may also affect the rate of fixation of amino acid replacements.

Recombination and Phylogenetic Reconstruction

The phylogenetic relationships of most of the studied species have been extensively analyzed. After Lachaise et al. (1988), the generally accepted phylogeny points to a closer relationship between D. yakuba and the species of the melanogaster complex, with D. erecta being a more distant species. Molecular phylogenies such as those based on divergence in Adh (Jeffs, Holmes, and Ashburner 1994; Russo, Takezaki, and Nei 1995) or Amy (Shibata and Yamazaki 1995) also support these relationships. However, in some other reported phylogenies, D. yakuba and D. erecta cluster together as sister taxa in the gene tree, e.g., in those based on DNA-DNA hybridization (Caccione, Amato, and Powell 1988), on nucleotide divergence in mitochondrial genes (Nigro, Solignac, and Sharp 1991), on nucleotide divergence in the 28S rDNA (Pêlandakis, Higgins, and Solignac 1991), and on synonymous divergence in the Amy genes (Inomata, Tachida, and Yamazaki 1997). The tree based on divergence at the Yellow protein also supports a closer relationship between D. yakuba and D. erecta.

On the other hand, the phylogeny based on total divergence (or only on synonymous divergence) of the y gene region is not consistent with the generally accepted relationships of the species studied. Thus, when the tree is rooted at the midpoint of the longest branch, D. ananassae clusters with the obscura group species instead of clustering with the other members of the melanogaster group. In addition, the position of D. yakuba in the gene tree as the most divergent taxa within the melanogaster subgroup was also unexpected. These distortions in the inferred phylogeny are a consequence of the inequality of the substitution rates (in fact, of the synonymous substitution rates) in the different lineages due to the change in the recombinational environment.
of the y gene region. Thus, the episodic acceleration of the substitution rate in the stem lineage of the melanogaster subgroup species would have occurred after the y gene changed its location to a region with a strong reduction in recombination.

The effect on synonymous divergence and on phylogenetic reconstruction detected in the y gene may not be restricted to this gene. In fact, the number of genes with strong interspecific differences in the recombinational environment might not be negligible. The extensive reshuffling detected within the basic chromosomal elements (Muller 1940) in studies of chromosomal evolution in Drosophila (Segarra et al. 1995; Segarra, Ribó, and Aguadé 1996) supports this idea. Thus, the particular sets of genes located in regions with very reduced recombination and a strong telomeric or centromeric effect may be different in different Drosophila species. Molecular phylogenies based on genes that have suffered such a change in the recombinational environment have to be interpreted with caution, at least in Drosophila. In addition, the effect of recombination on synonymous divergence indicates that putative changes in the recombination rate have to be considered when trees based on different genes are not concordant.

Acknowledgments

We thank J. Rozas for critical comments on the manuscript. We also thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. This work was supported by a predoctoral fellowship from Ministerio de Educación y Ciencia, Spain, to A.M. and by grants PB94-923 and PB97-918 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 1997SGR-59 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain, to M.A.

LITERATURE CITED


D IETHARD TAUTZ, reviewing editor

Accepted February 5, 2001