Excess of Nonsynonymous Polymorphism at Acph-1 in Different Gene Arrangements of Drosophila subobscura

Àurea Navarro-Sabaté, Montserrat Aguadé, and Carmen Segarra

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

Nucleotide variation in the Acph-1 gene region was analyzed in a natural population of Drosophila subobscura from Bizerte (Tunisia). The lines studied differed in their gene arrangement for segment I of the O chromosome: 21 lines were O₃¹₄−₈, 21 were O₃¹₄−₂₃, and 3 were O₃⁵₄. According to chromosomal phylogenies, O₃¹₄ is a central arrangement from which O₃¹₄−₈ and O₃¹₄−₂₃ originated. Strong genetic differentiation at Acph-1 was detected among the different arrangements, which is reflected in strong linkage disequilibrium between the variants at informative polymorphic sites and the type of arrangement. Estimates of silent nucleotide diversity are slightly lower within O₃¹₄−₂₃ (NIₕ = 0.0166) than within O₃¹₄−₈ (NIₕ = 0.0228) or O₃⁵₄ (NIₕ = 0.0234). In contrast, nonsynonymous nucleotide diversity estimates (around 0.1%) are similar in the three arrangements. Most nonsynonymous rare variants are singletons, which results in highly significant Tajima’s neutrality tests within the young O₃¹₄−₈ and O₃¹₄−₂₃ arrangements. This test is not significant for nonsynonymous mutations within a large Spanish O₃⁵₄ sample. In addition, a significant and marginally significant excess of nonsynonymous polymorphism was detected by the McDonald and Kreitman test within O₃¹₄−₂₃ and O₃¹₄−₈, respectively. This excess results in a rather high neutrality index (NI = 5.25) when both arrangements are jointly analyzed, in contrast to its value within the old O₃¹₄ arrangement (NI = 1.74). The pattern of variation at Acph-1 within the young arrangements is unusual for nuclear genes and has the same characteristics previously detected in most genes of the mitochondrial genome. Assuming that most nonsynonymous mutations at Acph-1 are under weak negative selection, a smaller effective size of the young arrangements relative to O₃¹₄ might explain the observed results. The relatively low frequency of O₃¹₄−₈ and O₃¹₄−₂₃ in the distribution area of D. subobscura, the more recent origin of these arrangements relative to O₃⁵₄, and the suppression of recombination in heterokaryotypes might contribute to the relatively small effective size of the young arrangements. Therefore, present results indicate that the differences in effective size and recombination caused by chromosomal arrangements are modulating nonsynonymous variation at Acph-1.

Introduction

Deviations from the neutral model of molecular evolution (Kimura 1983) caused by a significant excess of nonsynonymous polymorphism have been extensively reported in the mitochondrial genome (mtDNA). Several organisms such as Drosophila, mice, marsh rats, and human present this excess of amino acid variation in mtDNA genes (Ballard and Kreitman 1994; Nachman, Boyer, and Aquadro 1994; Rand, Dorfman, and Kann 1994; Nachman et al. 1996; Rand and Kann 1996; Hasegawa, Cao, and Yang 1998; Kennedy and Nachman 1998; Wise, Sraml, and Eastal 1998). Different selective factors have been proposed to explain this excess of nonsynonymous polymorphism. First, balancing selection maintaining protein variation or, alternatively, positive selection fluctuating over time or space might contribute to an excess of amino acid variation (Gillespie 1991). Second, mild negative selection acting on replacement mutations can cause these mutations to be maintained in the population at low frequency, but rarely to become fixed (Ohba 1992). Third, a relaxation of selection acting in the past might maintain previously nontolerated nonsynonymous polymorphism (Takahata 1993).

In contrast to mtDNA genes, only a few nuclear Drosophila genes exhibit a significant excess of nonsynonymous polymorphism. In Drosophila melanogaster, this excess has been detected at Gld (Hamblin and Aquadro 1997), Pgi (McDonald and Kreitman unpublished, cited in Moriyama and Powell 1996), Pgm (Verrelli and Eanes 2000), and bcd (Baines et al. 2002). Indeed, an excess of nonsynonymous fixed differences is a much more common result in this species (Moriyama and Powell 1996; Fay, Wyckoff, and Wu 2002). The paucity of nuclear genes showing an excess of nonsynonymous polymorphism led Weinreich and Rand (2000) to suggest that the lack of recombination in the mtDNA and the distribution of the selective effects of newly arisen mutations might explain the commonly detected excess of nonsynonymous polymorphism in mitochondrial genes.

Although most studies of nucleotide variation in Drosophila nuclear genes correspond to species of the melanogaster group, some surveys have been carried out in species of the obscura group. Only the analysis of nucleotide variation in the Acph-1 gene of D. subobscura has revealed an excess of nonsynonymous polymorphism (Navarro-Sabaté, Aguadé, and Segarra 1999a). In this case, however, the region under study maps at section 91C, very close to one of the breakpoints of the O3 inversion. As a consequence, variation in the Acph-1 gene was analyzed separately for lines differing by their chromosomal arrangement for segment I (sections 91–99) of the O chromosome: O₄ and O₅₄. The excess of nonsynonymous polymorphism was significant in the O₄ arrangement, but not in O₅₄, which suggested an effect of chromosomal inversion polymorphism on amino acid variation, at least at Acph-1. In addition, assuming that most nonsynonymous mutations at Acph-1 are under weak negative selection, the contrasting results between O₅₄ and O₄ are consistent with a much smaller effective size of O₄ relative to O₅₄ (Navarro-Sabaté, Aguadé, and Segarra 1999a).
Chromosomal phylogeny in *D. subobscura* (Krimbas 1992) indicates that O3+4 is a central arrangement from which several arrangements, such as O3+4+8 and O3+4+23, originated, giving rise to the O3+4+4 inversion complex. Therefore, it can be inferred unambiguously that these two arrangements are younger than O3+4. The presence of such a complex system of gene arrangements in segment I of the O chromosome of *D. subobscura* offers the opportunity to further analyze the putative effect of chromosomal inversions on nonsynonymous variation. Indeed, differences in the effective size of these arrangements are expected as a result of differences in their frequency, and even in their age. The effective size of young arrangements might be strongly affected by the extreme bottleneck produced in their origin, which should cause a reduction in the long-term effective size of the arrangement.

Here, we have studied nucleotide polymorphism at *Acph-1* in the young O3+4+8 and O3+4+23 arrangements. The quite low average frequency of these two arrangements in the distribution area of *D. subobscura* and their derived character suggest that their effective sizes should be lower than the effective size of the old O3+4 arrangement. These differences in the effective size and the well-established role of inversions as suppressors of recombination might affect nonsynonymous variation if nonsynonymous mutations are under weak selection, as shown in the case of O3+4. The results obtained support this prediction. A significant and marginally significant excess of nonsynonymous polymorphism is detected by the McDonald and Kreitman test (McDonald and Kreitman 1991) in the young arrangements O3+4+23 and O3+4+8, respectively. In addition, highly significant Tajima’s D statistics (Tajima 1989) for nonsynonymous mutations were detected within both gene arrangements. Therefore, variation at *Acph-1* within the young arrangements O3+4+23 and O3+4+8 presents the same pattern previously detected for most mitochondrial genes of *D. melanogaster*. According to these results, it can be inferred that differences in effective size and recombination caused by chromosomal arrangements are modulating nonsynonymous variation in a nuclear gene.

**Materials and Methods**

**Fly Samples**

Forty-five isochromosomal lines of *D. subobscura* were established from flies collected at Bizerte (Tunisia) in 1996. Most of these lines had been previously analyzed for variation at the *rp49* gene region (Rozas et al. 1999). Lines differed in their arrangement for segment I of the O chromosome: 21 lines were O3+4+8, 21 were O3+4+23, and 3 were O3+4. All O3+4+23 lines presented an additional inversion (inversion 6) in segment II; therefore, their complete gene arrangement for the O chromosome was O3+4+23+6.

**In Situ Hybridization**

Polytene chromosomes slides were obtained after dissection of the salivary glands from third-instar larvae heterozygous for different arrangements of segment I: O3+4/O3+4, O3+4+8/O3+4+23, O3+4+4/O3+4+8, and O3+4+23/O3+4+8. The DNA of the recombinant phage λ*Ds*ub*Acph*-1.1 (Navarro-Sabaté, Aguadé, and Segarra 1999b) was labeled with biotin-16-dUTP and used as probe. Prehybridization, hybridization, and detection conditions were as described in Segarra and Aguadé (1992).

**DNA Sequencing**

Genomic DNA of the isochromosomal lines was purified according to Ashburner (1989). The complete *Acph-1* gene region, approximately 2.2 kb long, was amplified by polymerase chain reaction (PCR). Amplification primers and PCR conditions were as described in Navarro-Sabaté, Aguadé, and Segarra (1999a). The PCR products were purified with QIAquick columns (Qiagen) and used as templates for sequencing with primers designed at approximately every 300 nucleotides. Both DNA strands were completely sequenced. The cycle sequencing kit from PerkinElmer was used to perform the sequencing reactions according to the manufacturer’s instructions. The reaction products were run on an Abi Prism 377 automated DNA sequencer.

Complete sequences of the 45 Bizerte lines were multiply aligned manually with the SeqApp (version 1.9) program (Gilbert 1993), and the sequence of line B2ST (Navarro-Sabaté, Aguadé, and Segarra 1999b) was used as the reference sequence. A multiple alignment including the 53 Spanish lines previously analyzed for the *Acph-1* gene region (Navarro-Sabaté, Aguadé, and Segarra 1999a; European Molecular Biology Laboratory (EMBL) accession numbers AJ389424 to AJ389476) was also obtained. Alignments were edited with the MacClade program (Maddison and Maddison 1992) prior to analysis.

**Data Analysis**

Data analysis was performed with the DnaSP, version 3.4, program (Rozas and Rozas 1999). Nucleotide polymorphism was estimated as: *S* (the number of segregating sites), *k* (the average number of pairwise nucleotide differences), *π* (nucleotide diversity; Nei 1987) and *θ* (the heterozygosity per site expected in a population at mutation-drift equilibrium given the observed *S* value; Watterson 1975). Genetic differentiation between arrangements or between populations was estimated by the average number of nucleotide differences per site (*dxy*; Nei 1987). The *K*′ statistic as proposed by Hudson, Boos, and Kaplan (1992) was used to determine whether arrangements or populations were significantly differentiated. The number of shared polymorphisms between arrangements (sites with the same nucleotide variants segregating in two different arrangements) expected by recurrent mutation was estimated assuming a hypergeometric distribution of mutations. Gene conversion tracts between arrangements were identified according to Betrán et al. (1997). Linkage disequilibrium between pairs of parsimonious informative sites (sites where the less frequent variant is present at least twice in the sample) was analyzed by the *χ*² test. The percentage of significant
comparisons and the average $r^2$ (Hill and Robertson 1968) between pairs of informative sites were used as measures of linkage disequilibrium. The recombination parameter $C = 2N_e \theta$ in Drosophila, where $N_e$ is the effective size and $\theta$ is the recombination rate between adjacent sites) was inferred according to Hudson (1987), and the minimum number of recombination events was inferred according to Hudson and Kaplan (1985).

Different neutrality tests (Tajima 1989; McDonald and Kreitman 1991) were performed to determine whether the level and pattern of nucleotide variation detected within gene arrangement conformed to neutral model expectations. Statistical significance of the Tajima’s $D$ observed values was obtained by coalescent simulation with no recombination (Hudson 1990) after 10,000 replicates. The probability for each test statistic was obtained as the frequency of replicates with a value lower than the observed statistic (one-tailed test).

The genealogy of the Acph-1 sequences was reconstructed by the Neighbor-Joining method (Saitou and Nei 1987) implemented in the Mega program, version 2.1 (Kumar et al. 2001). Genetic distances between sequences were estimated according to the Kimura two-parameter model (Kimura 1980) to correct for multiple hits. Sites with polymorphic indels were excluded from the analysis (complete deletion option). Bootstrap percentages for the different nodes were obtained after 500 replicates.

The sequences reported here have been deposited in the EMBL database library under accession numbers AJ563964 to AJ564008.

Results

In Situ Hybridization

The location of the Acph-1 gene in the inversion loop of the heterokaryotypes for the arrangements studied is shown in figure 1. In most nuclei of O3,4/O4 heterokaryotypes, the hybridization signal appears in the synaptic region of one of the inversion loops; this is in agreement with the location of Acph-1 inside and very close to one of the breakpoints of inversion 3. In O3,4/O3,4+23 and O3,4+23/O3,4+8 heterokaryotypes, the hybridization signal is detected inside the inversion loop and close to one of its ends, but in most nuclei the signal appears in the synaptic region of the loop. Finally, Acph-1 maps in a rather central position of the inversion loop present in O3,4/O3,4+8 heterokaryotypes.

Nucleotide Variation

The multiple alignment of the 45 sequences from the Bizerte sample had 2,169 sites. Nine polymorphic indels with one, two, or seven nucleotides were detected in noncoding regions. These polymorphic indels presented two or three variants segregating in the data set. All polymorphisms with three variants affected the length of runs of a single nucleotide. In addition, line TB21 presented in the third exon of the coding region a 16-nt long deletion that causes a frameshift mutation. After excluding all sites with alignment gaps, a total of 184 nucleotide polymorphisms were detected among the 2,135 sites of the data set (fig. 2). Three different nucleotides were segregating in six sites; thus the minimum number of mutations was 190.

Estimates of nucleotide polymorphism within each chromosomal arrangement are presented in table 1. These estimates were obtained independently for noncoding sites (flanking regions plus introns), silent sites (noncoding sites plus synonymous sites in the coding region), and synonymous and nonsynonymous sites of the coding region. Line TB21 with a frameshift mutation in the coding region was excluded from this analysis, because, for this line, the classification of sites in these different categories is ambiguous. For the O3,4+23 lines from Bizerte, estimated nucleotide diversity was similar (both for all sites and for silent sites) to that present within the O3,4+4 arrangement either from Bizerte or from Spain ($\pi = 0.0123$ and $\pi_{silent} = 0.0219$; Navarro-Sabaté, Aguadé, and Segarra 1999a). In contrast, variation within O3,4+23 is slightly lower when all sites, silent sites, or only synonymous sites in the coding region are considered. In contrast, nucleotide diversity at nonsynonymous sites is similar in the three gene arrangements.

Linkage disequilibrium analysis between pairs of informative sites within each arrangement was also performed. In O3,4+4+8, 162 of 1,431 comparisons (11%) were significant ($P < 0.05$) by the $\chi^2$ test, but only 2% of the comparisons were significant when the Bonferroni procedure (Weir 1996) to correct for multiple tests was applied. In O3,4+4+23 the extent of linkage disequilibrium was similar: the percentages of significant comparisons were 12% (99 of 820 comparisons), and only 1% with the Bonferroni correction. The lack of strong linkage disequilibrium is consistent with the rather high recombination detected within gene arrangements. Estimates of the recombination parameter (Hudson 1987) were $C = 0.092$ within O3,4+4+8 and $C = 0.029$ within O3,4+4+23. In contrast, the minimum number of recombination events (Hudson

FIG. 1.—Location of the Acph-1 gene revealed by in situ hybridization in the different heterokaryotypes for the arrangements studied of D. subobscura: A, O3,4/O4; B, O3,4/O3,4+23; C, O3,4+23/O3,4+4+8; D, O3,4/O3,4+4+8. Hybridization signals are indicated with arrowheads.
and Kaplan 1985) detected in the data set was 19 within O$_3^+$4 and 14 within O$_3^+$4+$^+$23.

Genetic Differentiation Between Arrangements

The $d_{xy}$ and $K^*_s$ estimates in the three pairwise comparisons between the O$_3^+$4, O$_3^+$4+$^+$8, and O$_3^+$4+$^+$23 arrangements are shown in table 2. The three arrangements are significantly differentiated, although no fixed differences between them were detected. Otherwise, the O$_3^+$4 lines from Bizerte are not genetically differentiated from the Spanish O$_3^+$4 lines ($P = 0.434$). The number of silent polymorphisms shared between arrangements in Bizerte is much higher than that expected by recurrent mutation (table 2).

Genetic exchange between arrangements can explain...
the relatively high number of shared polymorphisms. Despite the lack of fixed differences and the presence of shared polymorphisms, genetic differentiation between arrangements prevails because of the relatively high number of almost fixed differences between them. For instance, in 17 of the 49 polymorphisms shared between O3\textsuperscript{+4+8} and O3\textsuperscript{+4+23}, the most frequent variant differs between the two arrangements. Genetic differentiation between arrangements may also be detected by the analysis of linkage disequilibrium between nucleotide polymorphic sites and the type of chromosomal arrangement. Indeed, the presence of fixed or almost fixed variants between arrangements at a particular site should cause a strong association between the variants at this site and the type of chromosomal arrangement. This analysis was performed for all informative polymorphic sites and the chromosomal arrangement, in a sample including the O3\textsuperscript{+4+23} and O3\textsuperscript{+4+8} lines studied. The extent of linkage disequilibrium was estimated by the percentage of comparisons that were significant after applying the \( \chi^2 \) test and by the average \( r^2 \) value. Despite the lack of fixed differences and the presence of shared polymorphisms between O3\textsuperscript{+4+23} and O3\textsuperscript{+4+8} (likely due to gene conversion), 28 of 69 comparisons (about 41%) were significant when all 21 lines of each arrangement O3\textsuperscript{+4+23} and O3\textsuperscript{+4+8} were considered (table 3). This rather high percentage contrasts with that around 10% detected within arrangement, which reflects that recombination is restricted in heterokaryotypes. Therefore, the maintenance of both chromosomal arrangements in the population contributes to an overall reduction of recombination in the Acph-1 gene region, as shown by the high percentage of significant associations between the alternative variants of the informative sites and the chromosomal arrangement.

The analysis of linkage disequilibrium between nucleotide polymorphic sites and gene arrangements was extended to the Ost and O3\textsuperscript{+4} arrangements from the Spanish population (Navarro-Sabaté, Aguadé, and Segarra 1999a). As stated above, the rationale for this analysis is that high levels of linkage disequilibrium in joint samples of two arrangements (relative to linkage disequilibrium within each arrangement) indicate a strong genetic differentiation between them and thus a strong reduction of recombination in heterokaryotypes. For the results to be comparable, the analysis was conducted with samples where the two arrangements considered were equifrequent. The highest genetic differentiation was found in the samples including the O3\textsuperscript{+4} and the lowest differentiation in the sample with the O3\textsuperscript{+4} and O3\textsuperscript{+4+8} lines (table 3).

Replacements Polymorphism

The ACPH-1 preprotein has 447 residues in all lines except TB21. The frameshift mutation present in this line results in a truncated preprotein with 130 residues; thus this line can be considered a null allele for Acph-1. When this line is aligned in frame with the other lines, a single amino acid replacement (site 1537 in fig. 1) is detected, and no additional changes that disrupt the coding region (as nonsense mutations) are present. After excluding this null allele from the analysis, a total of 22 polymorphic residues were detected in the Bizerte sample: 2 within O3\textsuperscript{+4}, 11 within O3\textsuperscript{+4+23}, and 10 within O3\textsuperscript{+4+8} (fig. 4). These amino acid polymorphisms result in three different protein haplotypes within O3\textsuperscript{+4}, 10 within O3\textsuperscript{+4+23}, and 10 within O3\textsuperscript{+4+8}. The G/V (glycine/valine) polymorphism at residue 387 is shared by the O3\textsuperscript{+4+8} and O3\textsuperscript{+4+23} arrangements. However, the presence of a V in TB167 (O3\textsuperscript{+4+23}) can be explained by a gene conversion tract detected in this line. In contrast, three different variants were detected
Table 2
Genetic Differentiation Between Gene Arrangements

<table>
<thead>
<tr>
<th></th>
<th>O3+4 and O3+8</th>
<th>O3+4/O3+23</th>
<th>O3+4/O3+23</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{xy}$</td>
<td>0.0144</td>
<td>0.0150</td>
<td>0.0136</td>
</tr>
<tr>
<td>$K_r$</td>
<td>3.254</td>
<td>3.012</td>
<td>3.117</td>
</tr>
<tr>
<td>$P(K_r &gt; K_r^0)$ (observed)</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Fixed differences</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of silent shared polymorphisms</td>
<td>28</td>
<td>27</td>
<td>49</td>
</tr>
<tr>
<td>$P$ value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Expected number =</td>
<td>3.576</td>
<td>3.751</td>
<td>9.565</td>
</tr>
</tbody>
</table>

$^a$ P value based on the hypergeometric distribution.

$^b$ Expected number of silent shared polymorphism based on the hypergeometric distribution.

The McDonald and Kreitman (1991) test of neutrality (MK test) was also performed within each gene arrangement using the sequence of D. guanche (Navarro-Sabaté, Aguadé, and Segarra 1999b) in the interspecific comparison (table 5). This test is based on the comparison of the ratio of nonsynonymous to synonymous polymorphisms ($P_{nonsyn/Fsyn}$) and the ratio of nonsynonymous to synonymous fixed differences between species ($P_{nonsyn/Fsyn}$). Both ratios are expected to be equal under strict neutrality and putative deviations from this expectation can be determined by a G-test of independence. The test was significant ($P = 0.0374$) within O3+4+23 and marginally significant ($P = 0.0647$) within O3+4+8. In fact, the $P_{nonsyn/Fsyn}$ ratio within gene arrangements (0.28 and 0.24 within O3+4+8 and O3+4+23, respectively) was much higher than the $P_{nonsyn/Fsyn}$ ratio (0.06 in both comparisons). This deviation is also reflected in the neutrality index (NI; Rand and Kann 1996) estimated within O3+4+8 and O3+4+23 (table 5). This index is expected to be 1 under strict neutrality, and values higher than 1 indicate that the $P_{nonsyn/Fsyn}$ ratio is higher than the $P_{nonsyn/Fsyn}$ ratio.

The MK tests indicated a stronger deviation from predictions of the neutral model. The test was performed independently for silent and nonsynonymous sites (table 4). The low number of O3+4+4 lines from Bizerte prevented performance of this analysis within this chromosomal arrangement in this population. Thus, neutrality tests for this gene arrangement were performed with the Spanish random sample (Navarro-Sabaté, Aguadé, and Segarra 1999a). Even under the conservative assumption of no recombination, the test for nonsynonymous mutations is highly significant within the young arrangements: O3+4+23 and O3+4+8. In contrast, the test is not significant for nonsynonymous mutations within O3+4+23. None of the tests performed in the different arrangements was significant for silent mutations. The negative value of the significant Tajima’s $D$ statistic indicated a significant excess of low frequency (mostly singletons) nonsynonymous mutations in the data set. Although Tajima’s $D$ was also negative for silent sites within O3+4+23 and O3+4+8, the relative excess of low frequency silent mutations was not significant in these arrangements. In addition, the number of singleton and non-singleton polymorphisms differs significantly for nonsynonymous and synonymous mutations both within O3+4+23 ($\chi^2 = 11.27$, 1 df, $P = 0.0008$) and O3+4+8 ($\chi^2 = 5.37$, 1 df, $P = 0.0205$). The same test is not significant within O3+4 ($\chi^2 = 0.13$, 1 df, $P = 0.7156$). These results clearly indicate a different pattern of variation for silent and nonsynonymous mutations in the young arrangements (O3+4+23 and O3+4+8).

The McDonald and Kreitman (1991) test of neutrality (MK test) was also performed within each gene arrangement using the sequence of D. guanche (Navarro-Sabaté, Aguadé, and Segarra 1999b) in the interspecific comparison (table 5). This test is based on the comparison of the ratio of nonsynonymous to synonymous polymorphisms ($P_{nonsyn/Fsyn}$) and the ratio of nonsynonymous to synonymous fixed differences between species ($P_{nonsyn/Fsyn}$). Both ratios are expected to be equal under strict neutrality and putative deviations from this expectation can be determined by a G-test of independence. The test was significant ($P = 0.0374$) within O3+4+23 and marginally significant ($P = 0.0647$) within O3+4+8. In fact, the $P_{nonsyn/Fsyn}$ ratio within gene arrangements (0.28 and 0.24 within O3+4+8 and O3+4+23, respectively) was much higher than the $P_{nonsyn/Fsyn}$ ratio (0.06 in both comparisons). This deviation is also reflected in the neutrality index (NI; Rand and Kann 1996) estimated within O3+4+8 and O3+4+23 (table 5). This index is expected to be 1 under strict neutrality, and values higher than 1 indicate that the $P_{nonsyn/Fsyn}$ ratio is higher than the $P_{nonsyn/Fsyn}$ ratio.
Nucleotide Variation Within and Between Arrangements

Nucleotide variation at Acph-1 has been analyzed in four chromosomal arrangements segregating in natural populations of D. subobscura: O₃+₄, O₄+₂₃, O₃+₄+₈, and O₃+₄+₂₃. Because of the mutational process and the suppression of recombination in inversion heterokaryotypes, some genetic differentiation between arrangements is expected. However, differentiation may be counterbalanced by genetic exchange between arrangements. It has been proposed that the level of genetic exchange should be higher for markers located in central positions of the inversion loop than for markers located near the breakpoints (Rozas and Aguadé 1994; Navarro et al. 1997; Navarro, Barbadilla, and Ruiz 2000; Andolfatto, Depaulis, ...
and Navarro 2001). Indeed, near breakpoints, genetic exchange should be possible only by gene conversion, whereas in the central region of the inversion loop not only gene conversion but also double crossover can contribute to the genetic exchange. The genetic exchange is considered to be eventually null at the inversion breakpoints themselves where chromosomes cannot establish synopsis properly (Wesley and Eanes 1994; Andolfatto, Wall, and Kreitman 1999; Cáceres et al. 1999; Cáceres, Puig, and Ruiz 2001).

Empirical data obtained for markers located at different positions relative to the breakpoints of, for instance, the In(3L)Payne inversion of D. melanogaster support the proposed differential genetic exchange along the inversion (Hasson and Eanes 1996). Acph-1 is located inside and rather close to one of the ends of the inversion loop formed in O_{3+4+23}/O_{3+4} and O_{3+4+23}/O_{3+4+8} heterokaryotypes. This is not the case for the O_{3+2}/O_{3+4+8} heterokaryotypes, in which Acph-1 maps in a central position of the inversion loop (fig. 1). Consequently, genetic differentiation at Acph-1 is expected to be higher between O_{3+4+23} relative to both O_{3+4} and O_{3+4+8} than between the two latter arrangements. Although the \( d_{xy} \) estimates are similar in the three comparisons between these arrangements, the data on linkage disequilibrium between nucleotide polymorphic sites and the type of arrangement (table 3) are consistent with this expectation. Linkage disequilibrium in the O_{3+4} and O_{3+4+8} joint sample (and thus genetic differentiation between O_{3+4} and O_{3+4+8}) is lower than that detected in the O_{3+4} and O_{3+4+23} or in the O_{3+4+23} and O_{3+4+8} joint samples. In addition, the highest level of linkage disequilibrium was found in the samples including the O_{2a} lines. This result was expected because Acph-1 maps very close to one of the breakpoints of inversion 3 and would thus be asynaptic in heterokaryotypes between O_{2a} and any of the O_{3+4} arrangements (fig. 1 and results not shown). Therefore, the level of genetic exchange between O_{2a} and the other chromosomal arrangements is expected to be considerably reduced at Acph-1.

Prevalence of genetic differentiation despite the presence of genetic exchange between arrangements is also reflected in the genealogy inferred from nucleotide variation at Acph-1. The clustering of lines into two main groups (O_{a} and O_{3+4}) is consistent with restricted genetic exchange between these chromosomal classes. In contrast, O_{3+4} lines do not completely cluster according to their gene arrangement, which is consistent with a higher level of genetic exchange within O_{3+4} than between O_{a} and O_{3+4}. However, genetic exchange within O_{3+4} has not been high enough to erase completely the original chromosomal relationships among lines, as reflected in the partial clustering of lines by gene arrangement. A similar result was reported earlier, according to variation at the rp49 gene region (Rozas et al. 1999) that is tightly linked to Acph-1.

Excess of Nonsynonymous Polymorphism

Under predictions of the neutral model, the ratio between the number of polymorphic variants within species and the number of fixed differences between species is expected to be similar for synonymous and nonsynonymous mutations. This prediction results in a neutrality index equal to 1. Neutrality indexes considerably higher than 1 have been detected in O_{3+4+23} and O_{3+4+8}. Indeed, this index is 5.25 when both young arrangements are jointly analyzed. This deviation from strict neutrality is statistically significant, as revealed by the MK test (table 5). Also, a significant MK test associated with a neutrality index higher than 1 (NI = 9.01) was previously detected in the O_{a} arrangement (Navarro-Sabaté, Aguadé, and Segarra 1999a). In contrast, the MK test is not significant for the old O_{3+4} arrangement. Neutrality indexes higher than 1 have been interpreted mainly as resulting from an excess of nonsynonymous polymorphism. Alternatively, they could be due to an excess of fixed synonymous differences, but this latter explanation should hold for all chromosomal arrangements and thus should not cause differences between arrangements. The contrasting results obtained by the MK test in O_{3+4} versus O_{3+4+23}, O_{3+4+8}, and O_{a} indicate that different

### Table 4

<table>
<thead>
<tr>
<th>Tajima’s Test</th>
<th>Nonsynonymous sites</th>
<th>Synonymous sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrangement</td>
<td>( d_{xy} )</td>
<td>( d_{xy} )</td>
</tr>
<tr>
<td>O_{3+4}</td>
<td>-0.6625</td>
<td>0.2652</td>
</tr>
<tr>
<td>O_{3+4+23}</td>
<td>-1.4037</td>
<td>0.0580</td>
</tr>
</tbody>
</table>

* \( d_{xy} \) values under no recombination (one-tailed test).
* \( d_{xy} \) data from Navarro-Sabaté, Aguadé, and Segarra (1999a).

### Table 5

McDonald and Kreitman’s Test

<table>
<thead>
<tr>
<th>Fixed Polymorphic</th>
<th>Synonymous</th>
<th>Nonsynonymous</th>
</tr>
</thead>
<tbody>
<tr>
<td>O_{3+4} (old arrangement)</td>
<td>35</td>
<td>47</td>
</tr>
<tr>
<td>O_{3+4+23} (young arrangement)</td>
<td>35</td>
<td>46</td>
</tr>
<tr>
<td>O_{3+4+8} (young arrangement)</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

* \( G \) and \( P \) values from the MK test (two-tailed).
* \( NI \) data from Navarro-Sabaté, Aguadé, and Segarra (1999a).
constraints act on nonsynonymous mutations in the two sets of arrangements. The different constraints might be a priori attributable to differences in the strength of positive selection favoring or maintaining amino acid variation in the studied arrangements. In fact, an excess of replacement polymorphism is expected from positive selection fluctuating over time or space (Gillespie 1991). The excess of nonsynonymous polymorphism detected at Pgm in D. melanogaster (Verrelli and Eanes 2000) has been attributed to this kind of selection. Also, positive selection was proposed to explain the different ratio of replacement to silent polymorphism detected at fused-1 in standard X chromosomes and in chromosomes carrying a centric fusion between the X chromosome and the fourth chromosome of D. americana (Vieira, McAllister, and Charlesworth 2001). It does not seem likely, however, that positive selection can explain the results observed at Acph-1. All amino acid polymorphisms within O₃+₄+₂₃ and all but one within O₃+₄+⁴ are singletons, which results in highly significant Tajima’s D statistics for nonsynonymous variation (table 4). In addition, a significant negative Tajima’s D was detected within O₈ (Navarro-Sabaté, Agudé, and Segarra 1999a and present data). If positive selection were maintaining nonsynonymous variation at Acph-1, then a higher proportion of nonsynonymous polymorphisms segregating at intermediate frequencies would be expected.

The excess of nonsynonymous polymorphism detected at Acph-1 can be more likely explained according to the nearly neutral model of molecular evolution (Ôhta 1992), and assuming that most replacement mutations at Acph-1 are slightly deleterious. Apart from the variants responsible for the polymorphism at residues 205 and 241 of the ACPH-1 protein, all singleton nonsynonymous variants detected within O₃+₄+₈ and O₃+₄+₂₃ are derived and absent from the other arrangements. Therefore, it can be inferred that these variants appeared by mutation and were kept at low frequency within each arrangement by weak negative selection; i.e. nonsynonymous mutations are slightly deleterious. The highly significant negative Tajima’s D statistic values for nonsynonymous mutations also support this argument. Tajima’s D statistic is, however, highly sensitive to demographic factors such as population expansions and population subdivision. The presence of the different chromosomal arrangements clearly causes genetic substructuring at Acph-1, which in turn may affect Tajima’s D statistic. Nevertheless, demographic factors are expected to affect different regions of the genome or different kinds of sites to a similar extent. The different pattern of variation detected for nonsynonymous and silent mutations is not consistent with demographic explanations.

The pattern of variation in the young O₃+₄+₂₃ and O₃+₄+⁴ arrangements, and even in O₈, has the same signature as previously reported for most genes of the mitochondrial genome (mtDNA) in D. melanogaster and other organisms. This signature has been mainly explained according to the nearly neutral model and assuming that nonsynonymous mutations are weakly selected. Although theoretical models based on weak selection have recently been developed (McVean and Charlesworth 2000; Comeron and Kreitman 2002), the study of the joint effects of selection, drift, and recombination on weakly selected mutations requires further attention. However, as discussed by Weinreich and Rand (2000), two characteristics of the mitochondrial genome that may contribute to the detected pattern of variation are the lack of recombination and its lower effective size relative to the nuclear genome. Therefore, it is tempting to argue that these characteristics may also contribute to the pattern of variation detected at Acph-1 in the young O₃+₄+₂₃ and O₃+₄+⁴ arrangements, and even in O₈. The lack of recombination in mtDNA might be related to the reduction in recombination at Acph-1 in heterokaryotypes. In addition, a reduction in recombination can be viewed as a reduction in effective size as a consequence of the Hill-Robertson effect (Hill and Robertson 1966). In contrast, the effective size of these arrangements might be low given their frequency and their young character. In any case, a small effective size results in a longer persistence of slightly deleterious mutations in the population (i.e., negative selection against slightly deleterious mutations is relaxed).

The frequencies of O₃+₄+₈ and O₃+₄+₂₃ in Bizerte are 0.58 and 0.31, respectively (Rozas et al. 1999). These frequencies are in the upper limit of the frequencies of these arrangements in the Palearctic region. O₃+₄+₂₃ has a very localized distribution and it is present in appreciable frequencies only around Tunisia and in England. O₃+₄+₈ has a wider distribution, but its frequency rarely is larger than 0.1 except in the north of Africa, where it reaches the highest frequency in Tunisia. Therefore, it can be inferred that the average frequencies of these arrangements in the whole distribution area of D. subobscura are lower than the frequencies in Bizerte, with the lowest average frequency for O₃+₄+₂₃. The frequency of O₃+₄ is 0.79 in the Spanish sample, but only 0.11 in Bizerte (Rozas et al. 1999). The lack of genetic differentiation between both populations indicates that migration is high enough to homogenize the gene content within arrangement (see also Rozas et al. 1995). O₃+₄+₂₃ is an arrangement that reaches its highest frequencies (up to 0.9) in Iran and the Canary Islands (Krimbas 1992). Only in Tunisia and the north of Europe is its frequency lower than 0.1. In northwest Africa and the Mediterranean, frequencies from 0.2 to 0.7 are common. Therefore, the average frequency of O₃+₄+₂₃ in the distribution area of D. subobscura is expected to be slightly but not much lower than its frequency in Spain. The different average frequencies of these arrangements in the distribution area of D. subobscura allows us to predict that the effective size of O₃+₄+₂₃ is considerably higher than effective sizes of O₃+₄+₈ and O₃+₄+₂₃. In addition, these are the youngest arrangements, which may be why their long-term effective size was smaller than that of O₃+₄+₂₃. It has to be also considered that recombination at Acph-1 is only likely in homokaryotypes, which are expected to be rare for low-frequency arrangements. In this sense, the three-fold difference in the recombination parameter between O₃+₄+₂₃ and O₃+₄+₂₃ (see Results) would be consistent with the expected frequencies of the O₃+₄+₂₃ homokaryotypes in Bizerte (0.34 and 0.1, respectively).
All arrangements except O3þ4 exhibit an excess of nonsynonymous polymorphism at Acph-1. This excess should be also expected in other genes associated with the different arrangements. Previous results at rp49 do not support this expectation, because in this gene no nonsynonymous polymorphism was detected in any of the studied gene arrangements (Rozas et al. 1999). However, it has to be considered that the distribution of the selective coefficients acting against nonsynonymous mutations may differ among genes depending on the selective coefficients acting against nonsynonymous polymorphism. Constraints on the ACPH-1 protein would likely be relatively low, as indicated by the rather high nonsynonymous divergence at Acph-1 (Navarro-Sabaté, Aguadé, and Segarra 1999b) and the detection of one line (TB21) with a null allele. The low functional constraints on the ACPH-1 protein could have greatly contributed to the detected results. The effect of chromosomal polymorphism on nonsynonymous variation would thus vary with the level of constraints of the encoded protein, and such an effect might be negligible in genes subject to strong purifying selection, like rp49.

Acknowledgments

We thank J. Rozas for critical comments on a previous version of the manuscript, as well as D. M. Rand and the anonymous reviewers for their constructive comments. We also thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. This work was supported by a predoctoral fellowship from Comission Interdepartamental de Recerca i Innovació Tecnológica, Catalonia, Spain to À.N.-S. and by grants PB94-923 and PB97-918 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 1995SGR-577, 1997SGR-59, and 1999SGR-25 from Comission Interdepartamental de Recerca i Innovació Tecnológica, Catalonia, Spain, to M.A. Cáceres, M., M. Puig, and A. Ruiz. 2001. Molecular characterization of two natural hotspots in the Drosophila buzzatii genome induced by transposon insertions. Genome Res. 11:1353–1364.


Literature Cited


David Rand, Associate Editor
Accepted June 10, 2003