The Relationship Between Allozyme and Chromosomal Polymorphism Inferred From Nucleotide Variation at the *Acph-1* Gene Region of *Drosophila subobscura*

Àurea Navarro-Sabaté, Montserrat Aguadé and Carmen Segarra

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, 08071 Barcelona, Spain

Manuscript received March 15, 1999
Accepted for publication July 1, 1999

ABSTRACT

The *Acph-1* gene region was sequenced in 51 lines of *Drosophila subobscura*. Lines differ in their chromosomal arrangement for segment I of the *O* chromosome (O_{st} and O_{3+4}) and in the *Acph-1* electrophoretic allele (*Acph-1*^{100}, *Acph-1*^{054}, and *Acph-1*$^{>100}$). The ACPH-1 protein exhibits much more variation than previously detected by electrophoresis. The amino acid replacements responsible for the *Acph-1*^{054} and *Acph-1*$^{>100}$ electrophoretic variants are different within O_{st} and within O_{3+4}, which invalidates all previous studies on linkage disequilibrium between chromosomal and allozyme polymorphisms at this locus. The *Acph-1*$^{>100}$ allele within O_{3+4} has a recent origin, while both *Acph-1*^{054} alleles are rather old. Levels of nucleotide variation are higher within the O_{3+4} than within the O_{st} arrangement except for nonsynonymous sites. The McDonald and Kreitman test shows a significant excess of nonsynonymous polymorphisms within O_{st} when *D. guanche* is used as the outgroup. According to the nearly neutral model of molecular evolution, this excess is consistent with a smaller effective size of O_{st} relative to O_{3+4} arrangements. A smaller population size, a lower recombination, and a more recent bottleneck might be contributing to the smaller effective size of O_{st}.

DROSOPHILA subobscura has a wide distribution over the Palearctic region and rather recently has colonized the American continent (Krimbas 1992). Chromosomal and allozymic polymorphisms have been extensively surveyed in natural populations of *D. subobscura*. The species harbors a very rich chromosomal polymorphism that affects all chromosomal elements except the dot-like element. Several of the described inversions and chromosomal arrangements present clear latitudinal clines not only in the Palearctic region (Krimbas 1992) but also in North and South America. The finding that these clines present the same direction in both hemispheres has been taken as evidence of their adaptive character (Prevosti *et al.* 1988).

The *O* chromosome, which corresponds to Muller's *D* element (Muller 1940), is by far the most polymorphic with 24 described inversions that form complex chromosomal arrangements with overlapping and nonoverlapping inversions. This chromosome, the longest of the complement with 25 sections (Kunze-Mühl and Müller 1958), has been subdivided into two segments: I and II. Segment I, which is the distal segment (sections 91–99), presents alternative gene arrangements, such as O_{st} and O_{3+4}. These arrangements exhibit clear latitudinal clines in Europe where O_{st} is the prevalent arrangement in northern populations where O_{3+4} is rare, while O_{3+4} is more frequent than O_{st} in southern populations.

Corresponding author: Carmen Segarra, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain. E-mail: carme@porthos.bio.ub.es

The advent of recombinant DNA technology has enabled the study of chromosomal polymorphism based on nucleotide variation at molecular markers closely linked to inversion breakpoints. Since the work by Aguadé (1988) in which restriction fragment length polymorphisms (RFLPs) at the *Adh* region were compared between *standard* and *In(2L)t* chromosomes of *D. melanogaster*, additional data in different chromosomal systems of the same and other Drosophila species have been reported (Aquadro *et al.* 1991; Bénassi *et al.* 1993; Popadić and Anderson 1994, 1995; Popadić *et al.* 1995, 1996; Wesley and Eanes 1994; Babcock and Anderson 1996; Hasson and Eanes 1996). In *D. subobscura*, the *rp49* (ribosomal protein 49) gene region, which maps at section 91C very close to one of the breakpoints of the O_3 inversion, has been used as a molecular marker of the O_{3+4} and O_{st} chromosomal arrangements. Both RFLP analysis (Rozas and Aguadé 1990; Rozas *et al.* 1995) and direct sequencing (Rozas and Aguadé 1993, 1994; Rozas *et al.* 1999) have shown that the O_{3+4} arrangement has a higher level of silent polymorphism at the *rp49* gene region than O_{st}. On the other hand, comparison of this gene region between *D. subobscura* and the closely related species *D. madeirensis* and *D. guanche* indicates that the O_3 inversion, which is present in these species but absent in current populations of *D. subobscura*, is the ancestral gene arrangement (Ramos-Onsins *et al.* 1998).

In *D. subobscura*, as in *D. melanogaster*, the acid phosphatase-1 gene (*Acph-1*) is tightly linked to *rp49*, and it is closer to the O_3 breakpoint than this latter gene

(Segarra *et al.* 1996). *Acph-1* is one of the several allozyme loci whose variation in natural populations of *D. subobscura* was studied by starch electrophoresis. This locus presents a common allele, *Acph-1*^{100}, with frequencies ranging between 85 and 90%, and a less common allele, named *Acph-1*^{054} by Loukas *et al.* (1979), with ~10% frequency. In addition, rarer alleles, mainly with a higher mobility than *Acph-1*^{100}, have also been detected. The location of *Acph-1* in segment I of the *O* chromosome motivated extensive studies on the possible association of the two more frequent *Acph-1* electrophoretic variants with the different chromosomal arrangements in that segment. In contrast to other loci also located in segment I, such as *Pept-1* (peptidase-1) and *Lap* (leucyl amino peptidase), no clear and consistent associations were detected for *Acph-1* (Loukas *et al.* 1979, 1980; Fontdevila *et al.* 1983; Larruga and Pinsker 1984).

The *Acph-1* gene was first cloned and sequenced in *D. melanogaster* (Chung *et al.* 1996) and subsequently in the three species of the subobscura cluster (*D. subobscura*, *D. madeirensis*, and *D. guanche*; Navarro-Sabaté *et al.* 1999). The gene is organized into five exons interrupted by four rather short introns. The encoded protein (447 amino acids long in the obscura group species) has a signal peptide in the N-terminal end and a transmembrane domain in the carboxyl end, both of which are cleaved off to give the mature peptide. Both immunological studies (MacIntyre *et al.* 1978) and analysis of divergence at nonsynonymous sites (Navarro-Sabaté *et al.* 1999) indicate that ACPH-1 is a rapidly evolving enzyme.

Here, we report the study of nucleotide variation at the *Acph-1* gene region in a natural population of *D. subobscura* from El Pedroso, which is well characterized at the chromosomal and allozymic levels (Fontdevila *et al.* 1983; Rodríguez-Trelles *et al.* 1996). Lines from this natural population differing in their chromosomal arrangement for segment I of the *O* chromosome (O_{st} and O_{3+4}) and also in the *Acph-1* electrophoretic allele (*Acph-1*^{100}, *Acph-1*^{054}, and *Acph-1*$^{>100}$) have been studied in order to relate nucleotide polymorphism with chromosomal and allozyme polymorphism.

This study aims therefore at inferring the gene history of the *Acph-1* alleles studied, in particular in those resulting in the same and different electrophoretic variants in the two chromosomal arrangements. Only a sequencing study can reveal all nucleotide and amino acid variation and will thus contribute to understanding the presence of shared electrophoretic alleles between arrangements. We have found that for ACPH-1, electrophoresis detects only a minor fraction of the ACPH-1 protein haplotypes. The data have also revealed that the *Acph-1*^{054} and *Acph-1*$^{>100}$ alleles have a distinct origin within each arrangement. In addition, chromosomal polymorphism may affect the level and pattern of nucleotide variation at linked loci. The footprint left by the origin and posterior expansion of the O_{st} and O_{3+4} gene arrangements was already detected at the *rp49* gene region (Rozas and Aguadé 1993, 1994; Rozas *et al.* 1999). Present data at the *Acph-1* locus confirm the monophyletic character of O_{st} and O_{3+4}, and nucleotide diversity at the *Acph-1* noncoding regions further supports the proposed estimates of the age of these arrangements. Finally, in contrast to *rp49* that is monomorphic at the protein level, the rather high level of variation at the ACPH-1 protein has enabled us to detect an excess of nonsynonymous polymorphism in the O_{st} gene arrangement relative to O_{3+4}.

MATERIALS AND METHODS

Fly samples: The natural population of *D. subobscura* analyzed was sampled in El Pedroso. The isochromosomal lines studied were a subset of those described in Rozas *et al.* (1995), which had been established using the *O* chromosome *Va*/*Ba* balancer stock (Sperlich *et al.* 1977). Both the *O* chromosome arrangement and the *Acph-1* electrophoretic variant were determined for each line. Lines were grouped in two chromosomal classes (O_{st} and O_{3+4}) according to their arrangement for segment I of the *O* chromosome. In fact, all O_{3+4} lines were O_{3+4+7} (segment I_{3+4} + segment II_7) and all O_{st} lines were O_{st} (segment I_{st} + segment II_{st}) except line J34 that was O_7 (segment I_{st} + segment II_7).

Twenty-one O_{st} and 20 O_{3+4} lines were studied; within each gene arrangement lines were randomly sampled. All lines in these samples had the *Acph-1*^{100} electrophoretic variant except lines J34ST and J61ST within O_{st} and lines J70 and J72 within O_{3+4}, which had the *Acph-1*^{054} variant. Most lines in these samples were collected in autumn 1989. However, the lack of seasonal differentiation detected at the *rp49* gene region (Rozas *et al.* 1995) enabled us to include some lines sampled in other seasons of the same year.

These random samples were enlarged with six *Acph-1*^{054} lines: four O_{st} (J4ST, J8ST, A1ST, and A4ST) and two O_{3+4} (J77 and A10). Moreover, one O_{st} (J56ST) and three O_{3+4} (J21, A11, and A12) lines that presented a rare electromorph with a higher mobility than *Acph-1*^{100} were also included in the study. Finally, to corroborate the amino acid replacement responsible for the different electrophoretic variants, one *Acph-1*^{054}/O_{3+4} line from Barcelona, Spain and one *Acph-1*$^{>100}$/O_{3+4} line from Central Europe were also analyzed.

DNA sequencing: Genomic DNA of the different isochromosomal lines was purified according to Kreitman and Aguadé (1986). The complete *Acph-1* gene region (~2.2 kb) was amplified by polymerase chain reaction (PCR). Amplification primers located in the 5′ and 3′ flanking regions were: 5′-TCCTATGGTCAACGCCTATCG-3′ and 5′-GTTTTTCATTACCAAATGCAC-3′, respectively. The best PCR conditions were as follows: 28–34 cycles of 94° for 45 sec, 54° for 45 sec, and 72° for 2.5 min. After purification with Qiaquick columns, PCR products were used as templates for directly sequencing both strands of the amplified region with primers designed approximately every 300 nucleotides. Cycle sequencing reactions were performed using the Perkin Elmer (Norwalk, CT) cycle sequencing kit following the manufacturer's instructions. The reaction products were analyzed on a Perkin Elmer ABI PRISM 377 automated DNA sequencer.

Partial sequences were assembled using Staden's (1982) programs. Sequences were multiply aligned manually and the multiple alignment was edited with the MacClade program (Maddison and Maddison 1992). The sequences of *D. ma-*

deirensis and *D. guanche* (EMBL accession nos. Y18840 and Y18841, respectively) were used in the interspecific comparisons. The sequences reported here will appear in the EMBL database library under accession nos. AJ389424 to AJ389476.

Data analysis: Nucleotide polymorphism was estimated as the number of segregating sites (S), the average number of pairwise nucleotide differences (k), the nucleotide diversity (π), or average number of nucleotide differences per site (Nei 1987), and the heterozygosity per site (θ) expected under the neutral model at mutation-drift equilibrium given the observed S value (Watterson 1975). The pairwise nucleotide difference distribution or mismatch distribution was also analyzed. According to the neutral model with no recombination, this distribution is Poisson-like in expanding or growing populations, while it fits to a geometric distribution in constant size populations (Slatkin and Hudson 1991; Rogers 1995). The shape of the distribution was characterized by the raggedness, r, statistic (Harpending *et al.* 1993; Harpending 1994), which measures the smoothness of the distribution that is smaller in expanding than in constant size populations.

The level of genetic differentiation was estimated as the average number of nucleotide substitutions per site (d_{xy}) between arrangements (Nei 1987). Genetic differentiation between arrangements was contrasted by the permutation test proposed by Hudson *et al.* (1992). The hypergeometric distribution was used to test whether the number of detected shared silent polymorphisms (sites segregating for the same two nucleotides in both arrangements) could be explained by parallel mutations (Rozas and Aguadé 1994; Rozas *et al.* 1999).

Linkage disequilibrium was analyzed between pairs of informative sites (sites where the less frequent variant is present at least twice in the sample). The χ^2 test was used to detect significant linkage disequilibrium and the Bonferroni procedure to correct for multiple tests was applied (Weir 1996). In addition, the sign test on D (Lewontin 1995) was applied to search for overall evidence of linkage disequilibrium in the complete data set. According to this test, linkage disequilibrium between independent pairs of polymorphic sites are analyzed; adjacent sites were chosen for simplicity. Polymorphic sites with single variants are also included in the analysis. The likelihood ratio statistic, G, is used to determine whether the observed number of positive and negative D values differs from that expected under the null hypothesis of site independence. Recombination events between polymorphic sites were identified by the four-gamete test proposed by Hudson and Kaplan (1985).

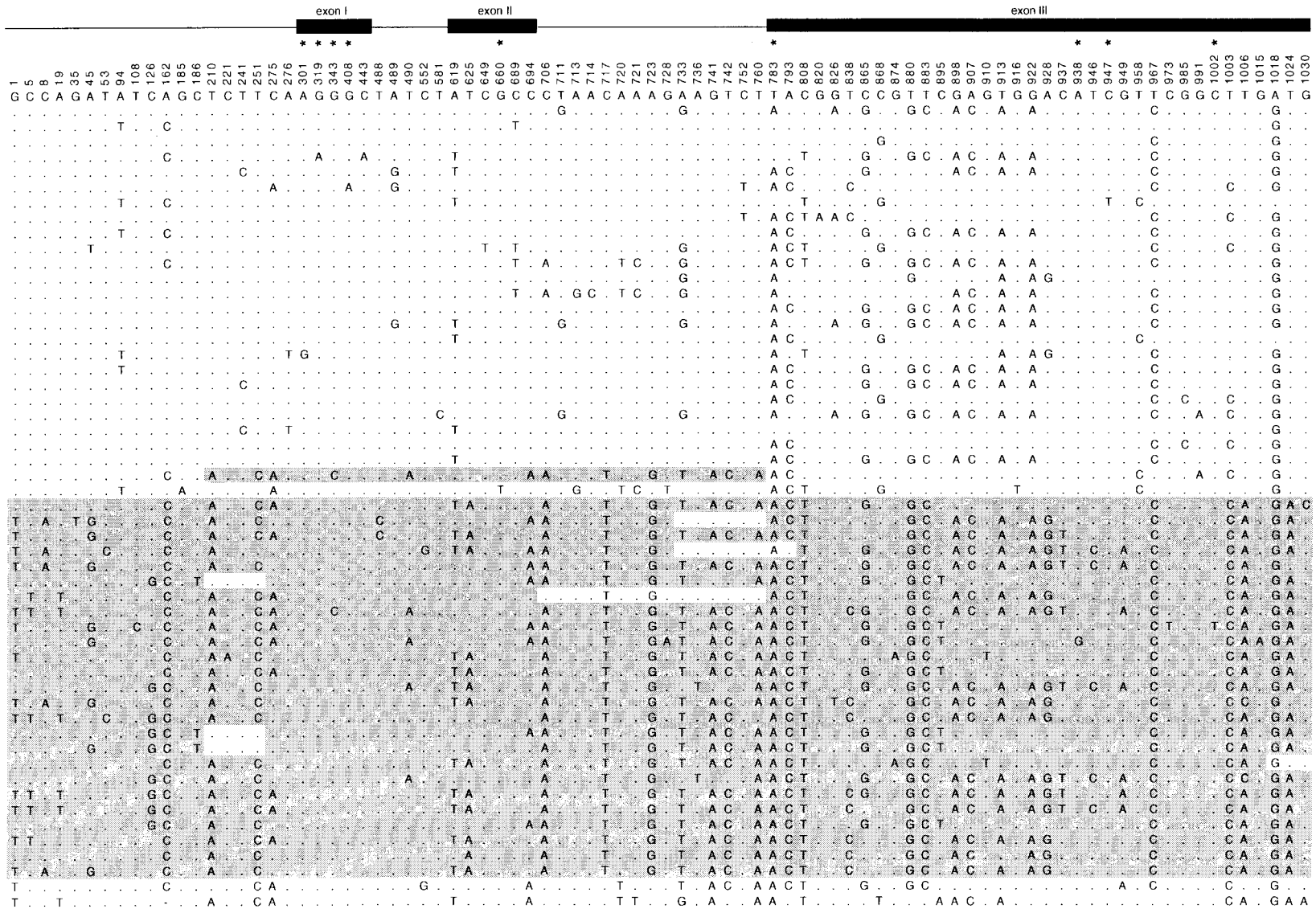
Different neutrality tests were performed to determine whether the observed data conformed to the predictions of the neutral model of molecular evolution. Tajima's (1989) test, based on intraspecific data, compares the observed polymorphism frequency spectrum with that expected under neutrality. Negative values of Tajima's D statistic indicate an excess of polymorphisms segregating at low frequency in the data set. The Fu and Li (1993) tests compare independent estimates of θ assuming neutrality. Their D statistic is based on the number of mutations in the internal and external branches of the gene genealogy, while the F statistic compares the average number of pairwise differences (k) and the number of mutations in external branches of the genealogy. An outgroup species is needed to estimate the number of mutations in external branches, and both the sequences of *D. madeirensis* and *D. guanche* were used for this purpose. In the Fu and Li test without outgroup (D^* and F^* statistics) the number of mutations in the external branches is inferred from the number of singletons or polymorphic variants that are present only once in the sample. Negative values of the different Fu and Li's statistics indicate an excess of unique polymorphisms in the data set. Tajima's and Fu and Li tests assume no recombination between sites and they are conservative for genomic regions with recombination.

The test proposed by Hudson *et al.* (1987), or the Hudson-Kreitman-Aguadé test, requires data from two genomic regions on the intraspecific variation in at least one species and on the interspecific divergence. This test determines whether the level of polymorphism and the level of divergence are proportional in both regions as expected from neutral predictions. The test assumes free recombination between both genomic regions and no recombination between sites of the same region; however, the test is conservative when these assumptions do not hold. The McDonald (1996) statistical test contrasts the heterogeneity in the distribution of polymorphism and divergence across a DNA region. Putative heterogeneity is analyzed by the number of runs detected in the sample, where a run is defined as a set of one or more polymorphic (or fixed) sites preceded and followed by at least one fixed (or polymorphic) site. Directional and balancing selection cause the number of runs detected in the sample to be smaller than that expected under neutrality, which is tested by Monte Carlo simulations. The McDonald and Kreitman (1991) test determines whether the ratio of nonsynonymous to synonymous polymorphisms within species is the same as the ratio of nonsynonymous to synonymous substitutions between species, as expected from neutral predictions. The F_s test statistic (Fu 1997) is based on the probability of having no fewer haplotypes or alleles than those observed in the data set. An excess of rare alleles relative to the number expected under neutral predictions is reflected by a large negative value of the F_s statistics.

The DnaSP program (Rozas and Rozas 1997) was used to estimate nucleotide polymorphism, linkage disequilibrium, recombination, the raggedness statistic, and genetic differentiation between arrangements and to detect gene conversion tracts. The permutation test for genetic differentiation (Hudson *et al.* 1992) was performed with the Permtest program provided by the authors. The DnaSP program was also used to carry out all neutrality tests, except the Monte Carlo simulations of the McDonald's test that were performed with the DNA Runs program (McDonald 1996). In addition, computer simulations based on the coalescent algorithm for no recombination described by Hudson (1990) and implemented in the DnaSP program were used to estimate (after 10,000 replicates) the probability of the observed Tajima's D and raggedness statistics. The computer program SignTestLD (S. Ramos-Onsins, personal communication) was used to perform the sign test on D (Lewontin 1995). Critical values of the observed F_s statistic were obtained by Monte Carlo simulations (after 5000 replicates) according to the computer program of Fu (1997). Phylogenetic analysis to reconstruct the gene genealogy of the studied lines was performed with the MEGA program (Kumar *et al.* 1994).

RESULTS

Nucleotide polymorphism: The multiple alignment of the *Acph-1* gene region analyzed in the 51 lines from the El Pedroso population included 2170 sites: 829 are located in noncoding regions (flanking regions and introns) and 1341 correspond to the coding region. A total of 171 nucleotide polymorphic sites were detected among the 2134 sites compared after excluding all sites with alignment gaps (Figure 1). Insertion / deletion polymorphisms were also detected but only in noncoding regions. The longest length polymorphism, located in


exon I
exon II
exon III
B2ST
J2ST
J7ST
J9ST
J10ST
J11ST
J12ST
J13ST
J50ST
J51ST
J52ST
J53ST
J54ST
J55ST
J57ST
J58ST
J59ST
J60ST
A2ST
A3ST
J34ST/054
J61ST/054
J4ST/054
J8ST/054
A1ST/054
A4ST/054
J56ST/>100
J16(3+4)
J18(3+4)
J20(3+4)
J24(3+4)
J25(3+4)
J30(3+4)
J71(3+4)
J73(3+4)
J74(3+4)
J75(3+4)
J76(3+4)
J78(3+4)
J79(3+4)
J80(3+4)
J81(3+4)
J82(3+4)
J83(3+4)
A13(3+4)
J70(3+4)/054
J72(3+4)/054
J77(3+4)/054
A10(3+4)/054
J21(3+4)/>100
A11(3+4)/>100
A12(3+4)/>100
D. madeirensis
D. guanche

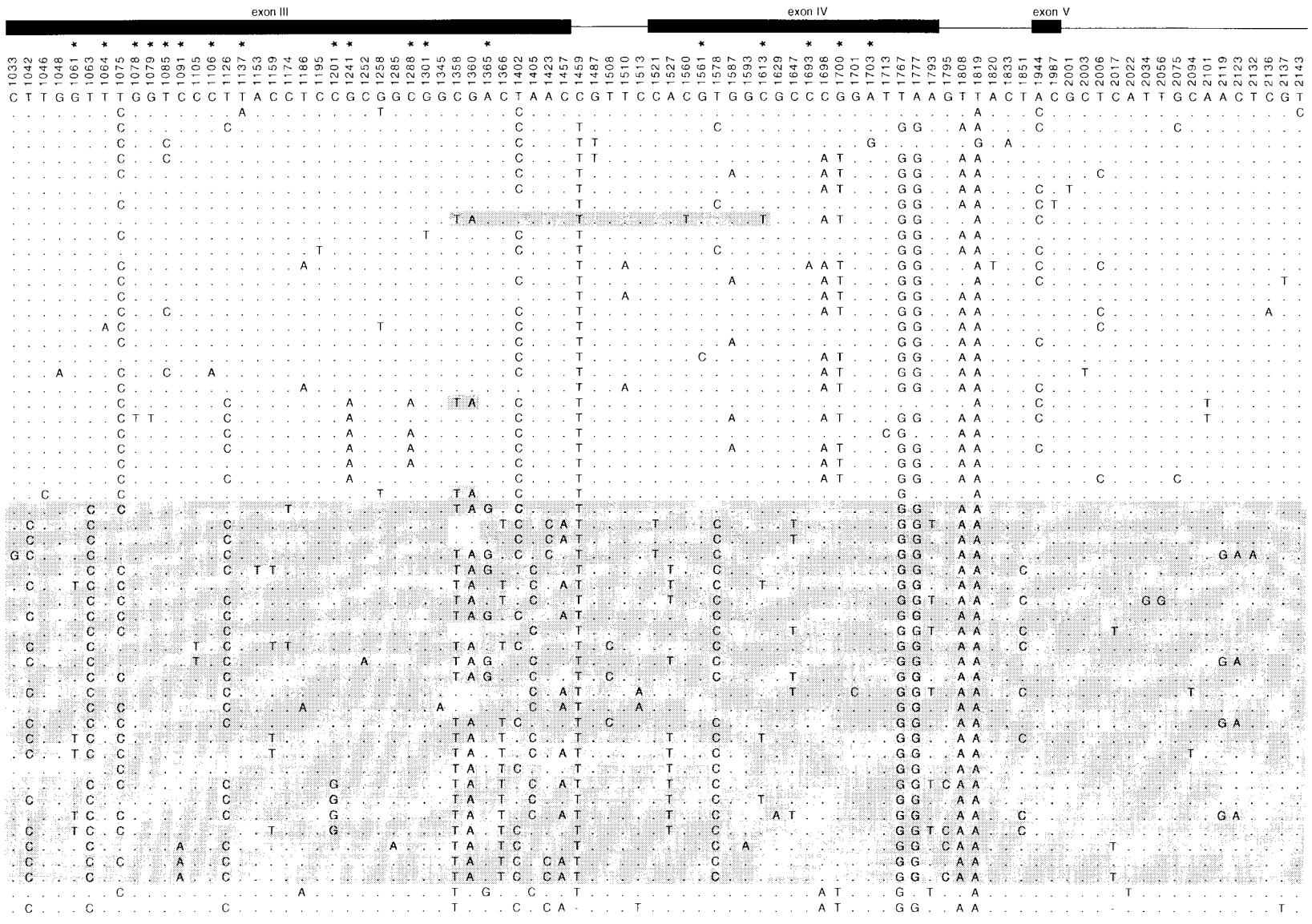


Figure 1.—Nucleotide polymorphic sites in the 51 lines from El Pedroso (Galicia, Spain). Numbering of sites is given using as reference the previously published sequence of *D. subobscura* (B2ST, EMBL accession no. Y18839). Site 1 in the present alignment corresponds to site 565 in the reference sequence. Nucleotides identical to the reference sequence are indicated by a dot. Black boxes in the upper part of the figure indicate the coding exons; a thin line shows flanking regions and introns. Replacement polymorphic sites in the coding region are indicated by an asterisk. Gene conversion tracts transferred from O_{3+4} to O_{st} lines are shown by shaded boxes, and those transferred from O_{st} to O_{3+4} lines are shown by white boxes. The last two rows of the alignment include the information for these polymorphic sites in the outgroup species *D. madeirensis* (EMBL accession no. Y18840) and *D. guanche* (EMBL accession no. Y18841). Identification numbers of each line are those used by Rozas and Aguadé (1994). ST and (3+4) after the identification number indicate gene arrangements O_{st} and O_{3+4}, respectively. The identification number of lines *Acph-1*^{054} and *Acph-1*$^{>100}$ includes after a slash (/) its electrophoretic mobility (054 and >100, respectively). The remaining lines are *Acph-1*^{100}. The first 21 O_{st} sequences and the first 20 O_{3+4} sequences with their identification numbers in bold are those included in the random samples.

the 3′ flanking region, includes the motif AATCGTGTT that is repeated once, twice, or three times in the different lines.

Population parameters were estimated only for the random samples of each chromosomal arrangement (Table 1). Among the 2145 sites scored in the random samples, 92 and 91 segregating sites (S) were detected within O_{st} and within O_{3+4} arrangements, respectively. All segregating sites presented only two variants within each arrangement. In contrast, there were four sites with three variants when all lines in the random samples were considered. On the other hand, the number of singletons varied considerably between O_{st} and O_{3+4} (40 and 27, respectively). The other estimates of polymorphism for all sites were lower within O_{st} than within O_{3+4}. Although estimates of nucleotide diversity (π) and of θ were similar within O_{3+4}, θ was generally larger than π within O_{st}. The higher number of singletons within O_{st} accounts for this difference since singletons have a larger effect increasing θ than increasing nucleotide diversity (as π considers not only the number of polymorphisms but also their frequency). A higher level of nucleotide polymorphism within O_{3+4} was also detected when only noncoding regions, silent sites (noncoding plus synonymous sites), or synonymous sites in the coding region were considered. In contrast, estimates of nonsynonymous polymorphism were higher within O_{st} than within O_{3+4}.

The distribution of nucleotide diversity (π) across the region studied was analyzed by the sliding window approach and is represented in Figure 2 for each arrangement. Both arrangements present a peak in the level of polymorphism around position 900. This window (100 sites) with the largest π value includes 10 segregating sites within O_{st} and 13 within O_{3+4}, all of them synonymous. The most striking difference between both graphs is the presence in O_{3+4} of a second peak around site 1370. The window encompassing this second peak in the O_{3+4} sample includes 8 segregating sites, 1 of them nonsynonymous.

The average number of nucleotide differences between sequences differing in chromosomal arrangement was 40.974, which gives an estimate of the average number of nucleotide substitutions per site between arrangements, d_{xy} (Nei 1987), equal to 0.019. In fact, there were 3 fixed differences between arrangements, while the number of sites that were monomorphic in O_{3+4} but polymorphic in O_{st} was 73, and the number of sites that were monomorphic in O_{st} but polymorphic in the alternative arrangement was 72. Significant differentiation between arrangements was detected by the permutation test proposed by Hudson *et al.* (1992) using K_s* as the test statistic: $P = 0.000$ after 1000 replicates. The two arrangements were considered separately in all posterior analyses due to their significant genetic differentiation.

Despite differentiation between arrangements, there were 19 (17 silent) shared polymorphisms, *i.e.*, sites segregating for the same two nucleotides in both arrangements. The hypergeometric distribution was applied to test whether the observed number of shared polymorphisms could be explained by parallel mutations that had arisen independently in both arrangements. According to the number of silent sites (1135) and the number of polymorphic silent sites in each arrangement (73 in O_{st} and 84 in O_{3+4}), the expected number of silent shared polymorphisms for $P > 0.05$ is only ≤ 8. Therefore, the high number of observed shared polymorphisms has to be explained by genetic exchange, most likely by gene conversion, between both arrangements.

The algorithm proposed by Betrán *et al.* (1997) was used to detect gene conversion tracts between arrangements. This analysis was performed using all the lines, as some lines not included in the random samples showed evidence of gene conversion by visual inspection. There were 75 informative sites (Betrán *et al.* 1997) in the complete data set. The probability of a site being informative of a conversion event or ψ (psi) was 0.01075. A total of 16 gene conversion tracts (4 within O_{st} and 12 within O_{3+4}) were identified (Figure 1). The number of gene conversion events, however, might be lower since different lines presented the same tracts.

Linkage disequilibrium for the random samples of each chromosomal arrangement was analyzed for all pairs of informative polymorphic sites. In the O_{st} random sample, 146 out of 1326 (11%) comparisons showed a significant association by the χ^2 test ($P < 0.05$). The number of significant comparisons dropped to 15 (1%) after applying the Bonferroni procedure to correct for multiple tests (Weir 1996). In the O_{3+4} random sample, 231 out of 2016 comparisons (11%) were significant by the χ^2 test, but only 20 (1%) remained significant by using the Bonferroni correction. However, these percentages are not very informative since, as pointed out by Lewontin (1995), some tests of association cannot give a significant result even with the more extreme disequilibrium. Therefore, linkage disequilibrium was also analyzed with the sign test on D (Lewontin 1995). Within O_{st}, 91 independent pair comparisons were performed. The observed number of positive D's was 31 and the expected number was 22.61 ($G = 3.86$; $0.025 < P < 0.05$). Thus a significant excess of coupling linkage in this gene arrangement was detected. A similar result was found within the O_{3+4} arrangement ($G = 4.22$; $0.025 < P < 0.05$).

In addition, some clustering of linkage disequilibria was detected both within O_{st} and within O_{3+4} arrangements at the beginning of exon 3 (Figure 3). In O_{st}, all comparisons between synonymous sites 865, 880, 883, 898, 907, 913, and 922 were significant ($0.001 < P < 0.01$) by the χ^2 test, although only the association between sites 898-907 and 913-922 was significant with the Bonferroni correction. The information in these sites

TABLE 1

Estimates of polymorphism within gene arrangement

Arrangement		Total (2145 sites)	Silent (1135 sites)	Noncoding (804 sites)	Coding region: Synonymous (331 sites)	Coding region: Nonsynonymous (1010 sites)
O_{st} (n = 21)	S	92	73	34	39	19
	Singletons	40	25	16	9	15
	k	20.838	18.070	6.781	11.2858	2.7672
	π	0.0097	0.0159	0.0084	0.0341	0.0027
	θ	0.0119	0.0179	0.0117	0.0327	0.0052
O_{3+4} (n = 20)	S	91	84	37	47	7
	Singletons	27	24	12	12	3
	k	26.332	24.8580	9.679	15.1829	1.4644
	π	0.0123	0.0219	0.0120	0.0459	0.0014
	θ	0.0120	0.0209	0.0130	0.0400	0.0019

forms two haplotypes (CTTGATG and GGCACAA) that segregate at intermediate frequencies within O_{st}. In this arrangement, disequilibrium between these sites is not complete and, in fact, recombination events have been detected between sites 865-880 and 880-898 when applying the four-gamete test (Hudson and Kaplan 1985). In O_{3+4}, 10 of the 20 significant disequilibria with the Bonferroni correction affect synonymous sites 898, 907, 913, 922, and 928 (Figure 3). At these sites two haplotypes (GATGA and ACAAG) in complete disequilibrium and segregating at intermediate frequencies were detected within the O_{3+4} random sample.

Replacement polymorphism: Twenty-seven out of 447 amino acid residues were polymorphic among the 51 lines (Figure 4). All sites segregated for only two variants. Although 2 residues (18 and 358) were polymorphic for the same variants in both arrangements, the presence of an alanine at site 18 (line A4ST) and of a valine at site 358 (line J50ST) can be explained by the gene conversion tracts detected in these lines. The first three replacement polymorphisms (4, 10, and 18) affect the signal peptide of the preprotein that is not included in the mature protein. All other amino acid polymorphisms should affect the mature protein since the transmembrane domain is encoded by exon 5 (Chung *et al.*

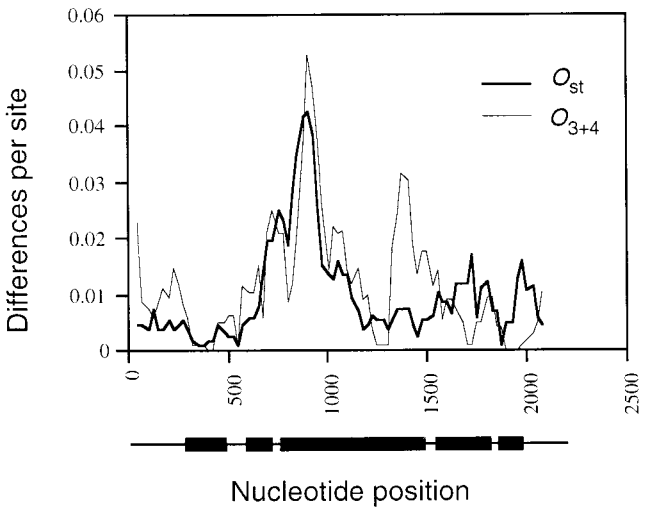


Figure 2.—Sliding window analysis of the distribution of nucleotide diversity (π) within the random samples of the O_{st} and O_{3+4} arrangements. Windows include 100 sites with successive displacements of 25 sites. The *x*-axis values indicate nucleotide position across the *Acph-1* gene region and those in the *y*-axis indicate nucleotide diversity. Solid boxes in the lower part of the figure indicate the coding exons; a thin line shows flanking regions and introns.

	865	880	883	898	907	913	922	928	frequency
ST	C	T	T	G	A	T	G	A	9
ST	G	G	C	A	C	A	A	.	9
ST	G	.	.	A	C	A	A	.	1
ST	.	G	.	.	.	A	A	G	1
ST	.	.	.	A	C	A	A	.	1
ST	.	.	.	.	.	A	A	G	1
(3+4)	G	G	C	.	.	.	.	.	7
(3+4)	.	G	C	A	C	A	A	G	5
(3+4)	G	G	C	A	C	A	A	G	6
(3+4)	.	G	C	.	.	.	.	.	2
D. madeirensis	G	G	C	.	.	.	.	.	
D. guanche	.	.	.	A	C	A	.	.	

Figure 3.—Nucleotide polymorphic sites at the beginning of exon 3 that show clustering of linkage disequilibria in O_{st} and/or O_{3+4} random samples. The sites affected within each arrangement have been shaded differentially. Numbers in the last column indicate the absolute frequency of each haplotype in the random samples. The corresponding haplotypes for *D. madeirensis* and *D. guanche* are also shown.

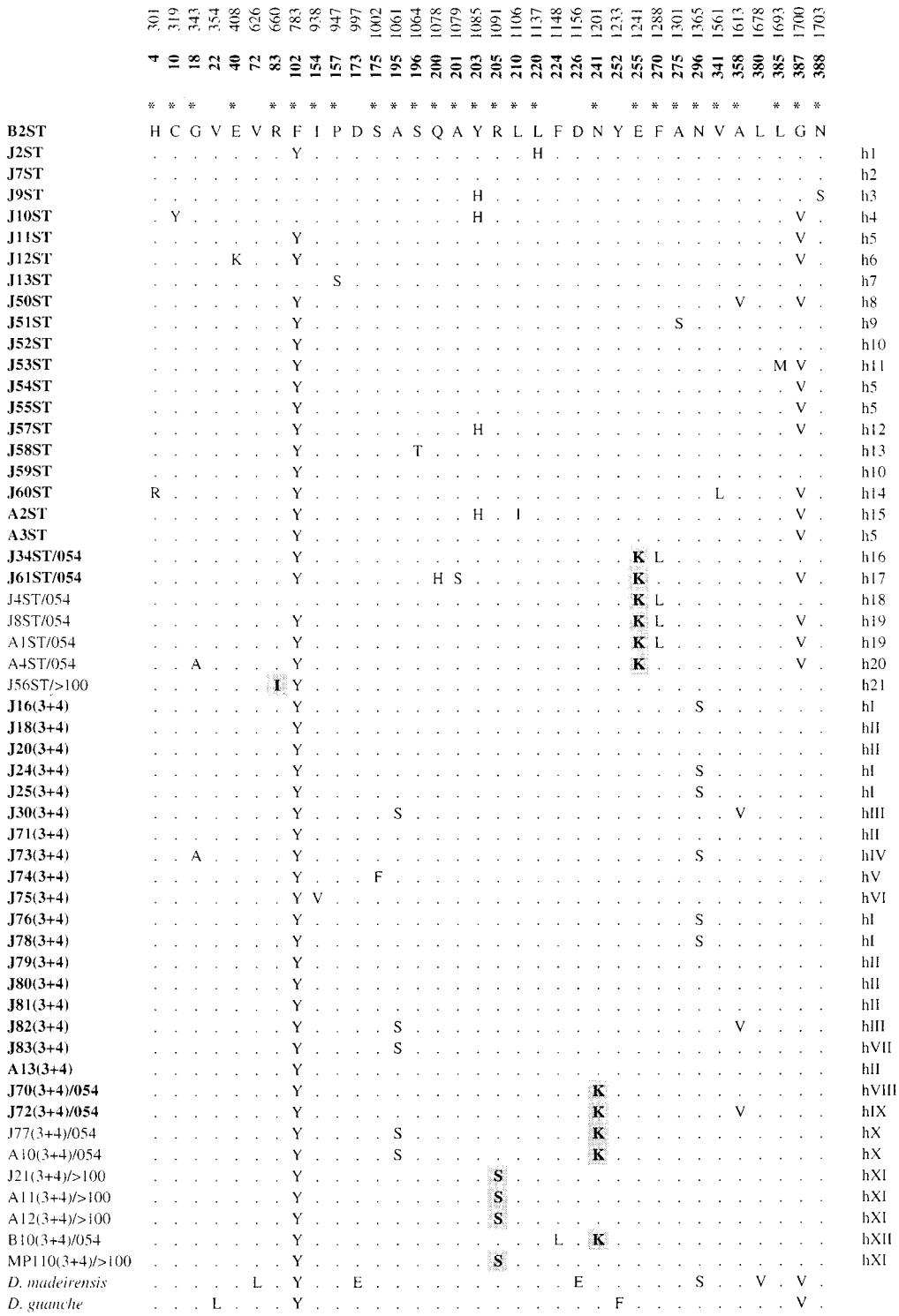


Figure 4.—Replacement polymorphisms at the ACPH-1 protein in the 51 lines studied from the El Pedroso population using as reference the previously published sequence of *D. subobscura* (B2ST, EMBL accession no. Y18839). Lines are named as in Figure 1. Two additional lines [B10(3 + 4)/054 from Barcelona and MP110(3 + 4)/>100 from Central Europe] sequenced to corroborate the amino acid replacement responsible for the corresponding electrophoretic alleles are also included. Fixed differences relative to *D. madeirensis* and *D. guanche* are shown in the last two rows of the alignment. The numbering at the top corresponds to nucleotide sites as in Figure 1. Numbers in boldface (below those of nucleotide sites) correspond to amino acid residues of the ACPH-1 preprotein. Asterisks indicate replacement polymorphic sites in the El Pedroso sample as depicted in Figure 1. The last column of the figure shows an identification number for each protein haplotype within O_{st} and within O_{3+4}. Haplotype 10 (h10) in O_{st} is identical to haplotype II (hII) in O_{3+4} and corresponds to the single haplotype shared between both arrangements. Shaded boxes indicate the amino acids characteristic of the *Acph*^{054} and *Acph*$^{>100}$ electrophoretic alleles within each chromosomal arrangement.

1996) and no replacement polymorphism was detected in that exon. When only random samples were considered, 19 residues (15 singletons) were polymorphic within O_{st} and only 7 (3 singletons) were polymorphic within O_{3+4}. These polymorphic residues result in 17 different protein haplotypes in O_{st} and 9 in O_{3+4}. Only one of these haplotypes is shared between arrangements (Figure 4). The number of protein haplotypes in the random samples of each arrangement exceeds that expected under the neutral model with no recombination. The probabilities of observing 17 or more haplotypes in O_{st} and 9 or more in O_{3+4} are $P = 0.000$ and $P = 0.009$, respectively. In fact, the values of the F_s statistics are $F_s = -15.206$ for O_{st} and $F_s = -4.715$ for O_{3+4}. The critical values for these test statistics at the 0.005 level of significance are -5.01 and -3.76 for each arrangement,

respectively. Although these values may be affected by recombination, the large departure detected within O_{st} would probably still be significant.

None of the detected replacements differentiated all the *Acph-1*^{054} lines from all the *Acph-1*^{100} lines. Therefore, a single amino acid replacement cannot explain the difference in mobility between both electromorphs. However, when the *Acph-1*^{054} and *Acph-1*^{100} electrophoretic alleles were compared within chromosomal arrangement, all *Acph-1*^{054} O_{st} lines had a lysine (K) at position 255 instead of the glutamic acid (E) present in all *Acph-1*^{100} O_{st} lines. Thus, the E/K replacement may explain the difference in mobility between electromorphs but only within O_{st}. In contrast, all *Acph-1*^{054} O_{3+4} lines shared the presence of a lysine (K) at position 241 instead of the asparagine (N) present in the *Acph-1*^{100} O_{3+4} lines, indicating that the N/K replacement would be responsible for the *Acph-1*^{100} and *Acph-1*^{054} electrophoretic alleles within O_{3+4}. This result was further confirmed by the presence of a lysine at position 241 in one *Acph-1*^{054} O_{3+4} line from Barcelona. Therefore, the electrophoretic allele *Acph-1*^{054} is not a homogeneous class but it includes two protein classes characterized by the presence of a lysine at residue 255 or 241 in complete linkage disequilibrium with arrangements O_{st} and O_{3+4}, respectively. The presence at residues 255 and 241 of a glutamic acid (E) and an asparagine (N), respectively, both in *D. madeirensis* and *D. guanche*, indicates that the presence of a lysine (K) in these residues for the *Acph-1*^{054} electrophoretic alleles would be the derived state.

Different amino acid replacements within each arrangement seem also to be responsible for the higher mobility electromorph. The only O_{st} *Acph-1*$^{>100}$ line from El Pedroso presented a distinctive isoleucine (I) at site 83 instead of an arginine (R); however, O_{3+4} *Acph-1*$^{>100}$ lines presented an arginine (R) at that site, but they presented a distinctive serine (S) at site 205. An additional O_{3+4} line from Central Europe confirmed that within O_{3+4} the R/S replacement at residue 205 causes the higher mobility of the *Acph-1*$^{>100}$ lines.

The level of nucleotide polymorphism within chromosomal arrangement and also the level of genetic differentiation relative to the *Acph-1*^{100} electrophoretic variant was analyzed to get more information about the history of the different electrophoretic classes (Table 2). Within each arrangement, nucleotide diversity (π) was similar for the *Acph-1*^{054} and *Acph-1*^{100} electrophoretic classes. In contrast, the estimated π for the three O_{3+4} *Acph-1*$^{>100}$ lines was almost three times lower than the nucleotide diversity for O_{3+4} *Acph-1*^{100} lines.

Neutrality tests: Tajima's (1989) and Fu and Li (1993) tests were used to contrast whether the observed pattern of polymorphism deviated significantly from that expected under the neutral model (Tables 3 and 4). The different tests were performed for all sites or only noncoding, synonymous, or nonsynonymous sites. None of the tests were significant ($P > 0.1$) either for O_{st} or O_{3+4} when considering all sites. However, the negative sign of the different statistics within O_{st} indicates a higher than expected number of low frequency polymorphisms for all sites and also for noncoding and synonymous sites. This deviation was significant for some of the tests when only nonsynonymous sites were considered, which indicates an excess of singletons or unique nonsynonymous polymorphisms within O_{st}. In contrast, within O_{3+4} the frequency spectrum of polymorphisms was in good agreement with neutral predictions, although the different statistics for nonsynonymous polymorphisms also presented negative values.

Putative departure from the direct relationship between polymorphism and divergence predicted by the neutral theory was determined by the Hudson-Kreitman-Aguadé test (Hudson *et al.* 1987) using either *D. madeirensis* or *D. guanche* in the interspecific comparison. The test was applied after dividing the *Acph-1* gene region into two regions of equal length. None of the tests performed within O_{st} or within O_{3+4} random samples were significant (results not shown). Heterogeneity in the ratio of polymorphic to fixed differences across the *Acph-1* region was also tested by the runs test proposed by McDonald (1996). The number of runs detected within O_{st} was 64 or 42 when using *D. guanche* or *D. madeirensis* in the interspecific comparison. These numbers were not significantly smaller than those expected under the neutral model ($P = 0.12$ and $P = 0.55$, respectively, after 5000 replicates with a recombination parameter $R = 16$). Likewise, the 76 and 39 runs detected within O_{3+4} in the *D. guanche* and *D. madeirensis* interspecific comparisons did not depart from neutral expectations ($P > 0.5$).

An excess of silent polymorphism is expected to accumulate around old polymorphisms maintained by balancing selection (Hudson and Kaplan 1988). The distribution of silent polymorphism was therefore further analyzed to determine whether such an excess had accumulated differentially around the nonsynonymous sites that cause the difference in mobility of the electrophoretic alleles *Acph-1*^{100} and *Acph-1*^{054} in each arrangement. The average number of silent nucleotide differences per site, silent d_{xy}, between all *Acph-1*^{100} and all *Acph-1*^{054} lines was analyzed across the studied region for each chromosomal arrangement using the sliding window approach (Figure 5). The graph also includes the distribution of the mean silent divergence between the *Acph-1*^{100} lines and *D. guanche*. The highest peak in the silent d_{xy} values present in both arrangements is caused by those polymorphisms in linkage disequilibrium at the beginning of exon 3 that are shared by both arrangements (Figure 3). The region including these polymorphisms corresponds, however, to a region with a rather high divergence. No peak in the silent d_{xy} values was detected either around site 1241 (residue 255 of the protein) within O_{st} (Figure 5a) or around site 1201 (resi-

TABLE 2

Nucleotide polymorphism and differentiation of the different *Acph-1* electrophoretic alleles

Arrangement	Electrophoretic allele	No. of lines	k	π	d_{xy}
O_{st}	*Acph-1*^{100}	19	20.152	0.0094	
	Acph-1^{054}	6	21.667	0.0101	
	Acph-1^{100}/*Acph-1*^{054}				0.0107
	Acph-1^{100}/*Acph-1*$^{>100}$				0.0136
	Acph-1^{054}/*Acph-1*$^{>100}$				0.0145
O_{3+4}	*Acph-1*^{100}	18	26.366	0.0124	
	Acph-1^{054}	4	22.333	0.0105	
	Acph-1$^{>100}$	3	9.333	0.0044	
	Acph-1^{100}/*Acph-1*^{054}				0.0122
	Acph-1^{100}/*Acph-1*$^{>100}$				0.0111
	Acph-1^{054}/*Acph-1*$^{>100}$				0.0109

due 241 of the protein) within O_{3+4} (Figure 5b). Despite no detected excess in silent d_{xy} values around the sites responsible for the *Acph-1*^{054} electrophoretic alleles, linkage disequilibrium was reanalyzed considering all O_{st} and all O_{3+4} lines although they did not correspond to a random sample. Within O_{st}, significant linkage disequilibrium was detected by the χ^2 test between site 1241 (responsible for the E/K replacement at residue 255 of the protein) and sites 985, 991 ($0.001 < P < 0.01$), and 1288 ($P < 0.001$). Interestingly, the less frequent variant in these sites is present only in the *Acph-1*^{054}/O_{st} lines, which would indicate that polymorphisms at sites 985, 991, and 1288 arose in the *Acph-1*^{054} allele. Strong linkage disequilibrium was also detected between sites 1126 and 1241 ($P < 0.001$); however, in this case a recombination event was detected by the four-gamete test (Hudson and Kaplan 1985) and the polymorphism at site 1126 is shared by both arrangements. Therefore, there is some evidence that within O_{st}, the *Acph-1*^{054} lines have accumulated variation differentially from the *Acph-1*^{100} lines.

The previous analysis was extended to the other amino acid replacements that do not cause a change in electrophoretic mobility and that are segregating within the random samples of each arrangement. The most notable result of these analyses refers to the replacement N/S at residue 296 of the protein. As shown in Figure 6, site 1365, the nucleotide site that causes the N/S

TABLE 3

Statistical tests of neutrality for the O_{st} random sample ($n = 21$)

			Coding region	
	Total	Noncoding	Synonymous	Nonsynonymous
Tests without outgroup				
No. of sites analyzed	2145	804	331	1010
S	92	34	39	19
Singletons (%)	40 (43)	16 (47)	9 (23)	15 (79)
Tajima's test				
D	−0.7517	−1.1038	0.1618	−1.7765
Fu and Li tests				
D	−0.8752	−1.0311	0.3565	−2.6880*
F	−0.9790	−1.2285	0.3474	−2.8143*
Tests with outgroup				
No. of sites analyzed	2125	784	331	1010
S	90	32	39	19
Singletons (%)	39 (43)	15 (47)	9 (23)	15 (79)
Fu and Li tests				
D (*D. guanche*)	−1.0353	−0.9006	0.3401	−3.2832**[a]
F (*D. guanche*)	−1.1532	−1.1931	0.3421	−3.3868**[a]
D (*D. madeirensis*)	−1.0353	−1.1200	0.3401	−3.2832**[a]
F (*D. madeirensis*)	−1.1532	−1.3714	0.3421	−3.3868**[a]

* $0.01 < P < 0.05$; ** $0.001 < P < 0.01$.

[a] Significant by the Bonferroni procedure.

TABLE 4

Statistical tests of neutrality for the O_{3+4} random sample (n = 20)

			Coding region	
	Total	Noncoding	Synonymous	Nonsynonymous
Test without outgroup				
No. of sites analyzed	2145	804	331	1010
S	91	37	47	7
Singletons (%)	27 (30)	12 (32)	12 (25)	3 (43)
Tajima's test				
D	0.1089	−0.2854	0.5867	−0.8410
Fu and Li tests				
D	0.0000	−0.1570	0.2393	−0.5777
F	0.0384	−0.2275	0.4002	−0.7540
Tests with outgroup				
No. of sites analyzed	2125	784	331	1010
S	89	35	47	7
Singletons (%)	25 (28)	10 (29)	12 (25)	3 (43)
Fu and Li tests				
D (*D. guanche*)	0.1787	0.1729	0.3462	−0.7170
F (*D. guanche*)	0.2164	0.0780	0.5232	−0.8971
D (*D. madeirensis*)	0.3501	0.3727	0.5002	−0.7170
F (*D. madeirensis*)	0.3546	0.2408	0.6482	−0.8971

replacement, lies in a region with a peak of silent differences between O_{3+4} lines carrying the alternative amino acids. Actually, this peak corresponds to the second peak detected in the distribution of nucleotide diversity within O_{3+4} (Figure 2). However, in this region divergence is also high.

McDonald and Kreitman (1991) proposed a test of neutrality to determine whether the ratio of nonsynonymous to synonymous polymorphisms within species is equal to the ratio of nonsynonymous to synonymous substitutions between species as predicted by the neutral theory. The McDonald-Kreitman test was applied using *D. guanche* as the outgroup species, as previous interspecific analysis had detected deviation in the *D. madeirensis* lineage toward an excess of fixed nonsynonymous substitutions (Navarro-Sabaté *et al.* 1999). As shown in Table 5, the McDonald-Kreitman test was highly significant within the O_{st} random sample (G = 12.2, 1 d.f., P = 0.0004). In fact, 33% of the polymorphisms within species were nonsynonymous, but only 5% of the fixed differences between species were nonsynonymous. These percentages were more similar within O_{3+4} (13% and 8%, respectively), and no significant departure from neutral expectations was detected for this arrangement (G = 0.61, 1 d.f., P = 0.43). There was, therefore, an excess of nonsynonymous polymorphisms within O_{st}, but not within O_{3+4}. This was further confirmed by applying the McDonald-Kreitman test to the random sample of lines collected in autumn 1989. The test was again significant for O_{st} (G = 7.3, 1 d.f., P = 0.007), indicating that pooling data from different seasons had not caused the detected excess of nonsynonymous polymorphisms within O_{st}.

A G test of independence was used to determine whether the number of synonymous and nonsynonymous polymorphisms was significantly different in both arrangements. Significant departure from neutral expectations was detected (G = 6.14, 1 d.f., P = 0.013), confirming the excess of nonsynonymous polymorphism within the O_{st} arrangement.

Gene genealogy: Figure 7 shows the genealogy of all lines studied, which was reconstructed by the neighbor-joining method (Saitou and Nei 1987) using *D. guanche* as the outgroup. Genetic distances were estimated for all sites with the complete deletion option and corrected according to Jukes and Cantor (1969). All O_{st} lines cluster together in the gene tree and all O_{3+4} lines do as well (percentage bootstrap values of 36 and 78, respectively, after 1000 replicates). Therefore, despite evidence of gene conversion between arrangements, the gene genealogy still reflects the unique origin of each arrangement. However, line A4ST, with a rather long gene conversion tract from O_{3+4}, shows an anomalous branching within the O_{st} cluster. When this line is subtracted from the analysis, bootstrap values increase to 93 for O_{st} lines and 84 for O_{3+4} lines. The clustering of lines according to their gene arrangement was also obtained when using Kimura's (1980) two-parameter distance or Tamura's (1992) distance. In addition, within both O_{st} and O_{3+4} there were two subclusters in the gene tree. This subclustering within gene arrangement was caused by synonymous polymorphisms in linkage disequilibrium at the beginning of exon 3, as no subclustering was detected when those sites were removed from the analysis (result not shown).

Finally, no clustering of the *Acph-1*^{054} lines was de-

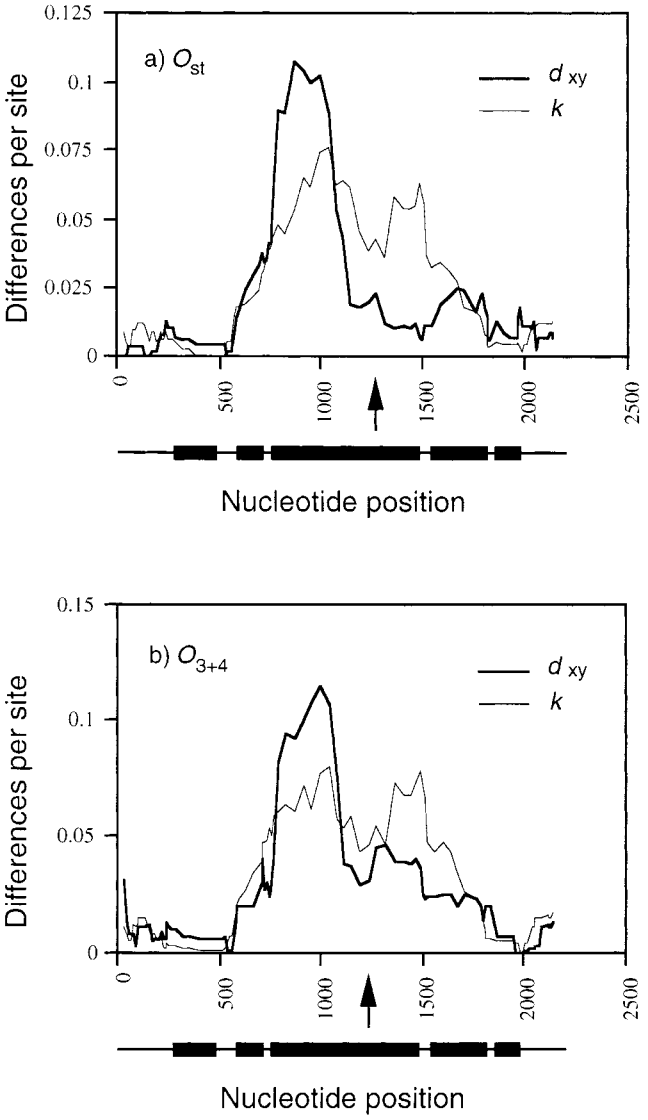


Figure 5.—Sliding window analysis of the distribution of the average number of silent differences per site (silent d_{xy}) between *Acph-1*^{100} and *Acph-1*^{054} lines within (a) O_{st} and (b) O_{3+4} arrangements. Windows include 75 silent sites with successive displacements of 10 sites. The graph also shows the distribution of silent divergence (k) between *D. guanche* and lines included in the random samples of the (a) O_{st} and (b) O_{3+4} arrangements. Silent divergence was scaled by the divergence time $T + 1$ as estimated when applying the Hudson-Kreitman-Aguadé test for each chromosomal arrangement. The *x*-axis values indicate the nucleotide position across the *Acph-1* gene region and those in the *y*-axis indicate silent d_{xy} or divergence. The approximate locations of nucleotide sites 1245 and 1205 are indicated by an arrow in a and b, respectively; the nucleotide polymorphism at these sites is responsible for the amino acid replacement causing the difference in electrophoretic mobility between the *Acph-1*^{100} and *Acph-1*^{054} electrophoretic alleles within each gene arrangement. Solid boxes in the lower part of the figures indicate the coding exons; a thin line shows flanking regions and introns.

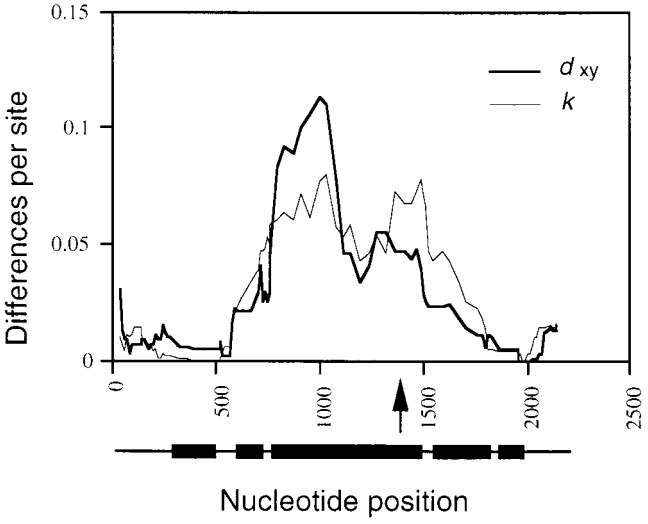


Figure 6.—Sliding window analysis of the distribution of the average number of silent differences per site (silent d_{xy}) between the O_{3+4} lines that differ at residue 296 (nucleotide site 1365) of the ACPH-1 preprotein. Windows include 75 silent sites with successive displacements of 10 sites. The graph also shows the distribution of silent divergence (k) between lines included in the O_{3+4} random sample and *D. guanche*. Silent divergence was scaled by $T + 1$ (see legend of Figure 5). The values in the *x*-axis indicate the nucleotide position across the *Acph-1* gene region and those in the *y*-axis indicate silent d_{xy} or divergence. The approximate location of nucleotide site 1365 is indicated by an arrow. Solid boxes in the lower part of the figure indicate the coding exons; a thin line shows flanking regions and introns.

tected within either the O_{st} or O_{3+4} gene arrangements. Assuming that the amino acid replacement causing the *Acph-1*^{054} electrophoretic mobility occurred only once in each arrangement, this result indicates that recombination within arrangement has obscured the evolutionary history of these lines. Actually, recombination within arrangements may have hidden the real relationships among lines both within the O_{st} and within the O_{3+4} clusters. However, the three O_{3+4} *Acph-1*$^{>100}$ lines (J21, A11, and A12) clustered together in the gene tree, which could be consistent with their more recent origin as indicated by their low level of nucleotide diversity (Table 2).

DISCUSSION

Level and distribution of nucleotide polymorphism: The O_{st} and O_{3+4} arrangements arose independently from the O_3 arrangement (Ramos-Onsins *et al.* 1998). The gene genealogy based on the *Acph-1* gene region (Figure 7), where all O_{st} lines cluster together as all O_{3+4} lines do, clearly supports their monophyletic character, which was also inferred from variation at the *rp49* gene region (Rozas and Aguadé 1994). Therefore, both arrangements were affected at some time in the past by the extreme bottleneck implied by their origin. Conse-

TABLE 5

McDonald and Kreitman tests using *D. guanche* as an outgroup

	O_{st}		O_{3+4}	
	Fixed	Polymorphic	Fixed	Polymorphic
Synonymous	37	39	35	47
Nonsynonymous	2	19	3	7
	$G = 12.210$***		$G = 0.611$	

*** $P < 0.001$.

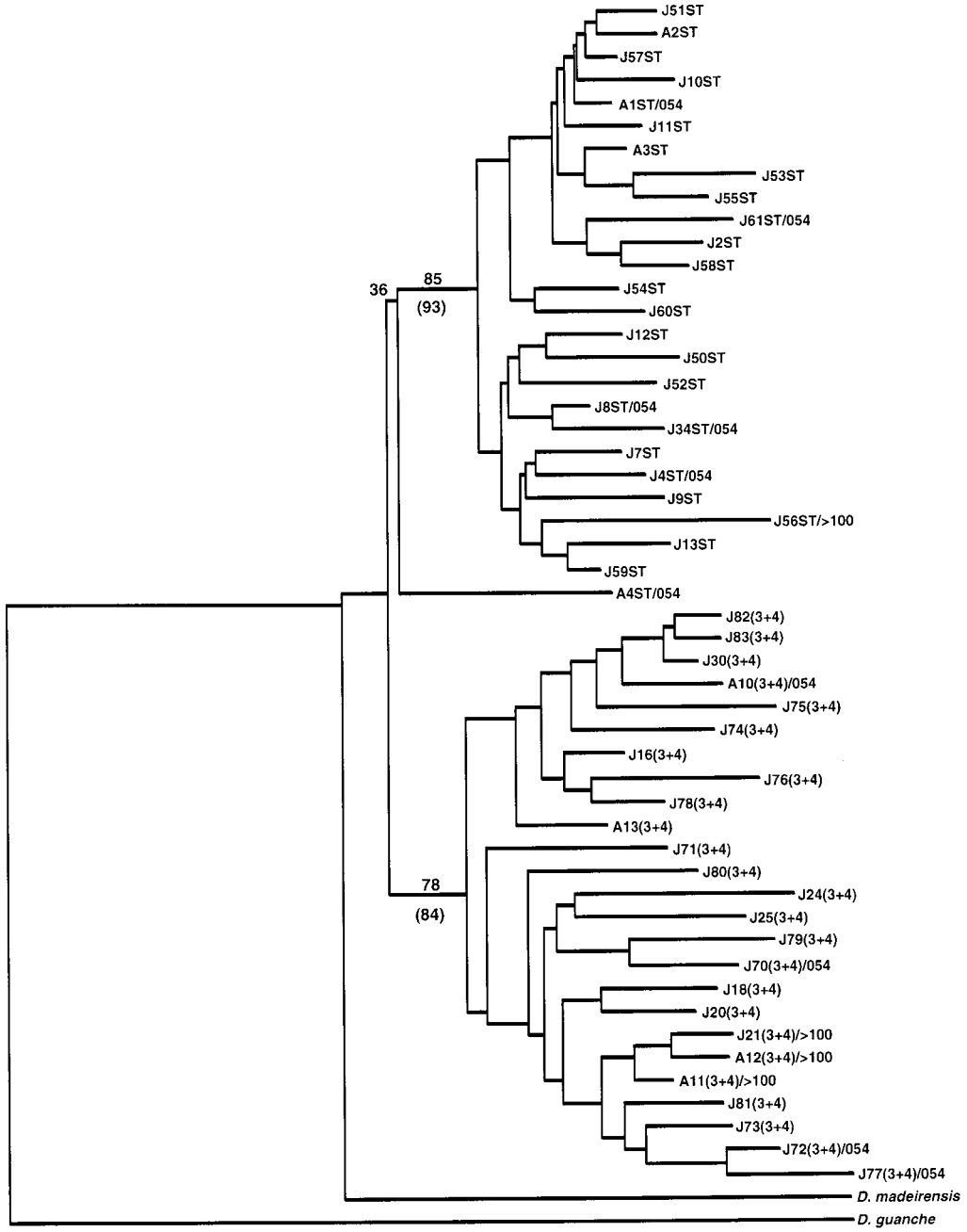


Figure 7.—Gene genealogy of the 51 lines of *D. subobscura* from El Pedroso reconstructed by the neighbor-joining method (Saitou and Nei 1987) with the complete deletion option. The number of substitutions per site corrected according to Jukes and Cantor (1969) in the whole *Acph-1* gene region studied was used as the genetic distance. Numbers indicate bootstrap percentages for the main nodes on the tree; the values in brackets show bootstrap percentages for the nodes grouping all O_{st} and all O_{3+4} lines when line A4ST/054 is not included in the analysis. Lines are named as in Figure 1.

quently, nucleotide polymorphisms present in each arrangement either have originated independently by mutation or have been incorporated by genetic exchange, most likely by gene conversion (Rozas and Aguadé 1994; Rozas *et al.* 1999; Navarro *et al.* 1997), between O_{st} and O_{3+4}, and even between any of them and O_3.

Current levels of silent polymorphism at the *Acph-1* gene region ($\pi = 0.0159$ and $\pi = 0.0219$ in O_{st} and O_{3+4}, respectively) are higher than those previously reported (Rozas *et al.* 1999) at the *rp49* gene region for the same population ($\pi = 0.0080$ and $\pi = 0.0101$, respectively). Therefore, lower constraints against the accumulation of silent variation seem to act at the *Acph-1* rather than at the *rp49* gene region as already inferred from silent divergence in these regions (Navarro-Sabaté *et al.* 1999). Nevertheless, silent nucleotide diversity at *Acph-1* in O_{st} is very similar to that detected at the *Acp70A* gene region ($\pi = 0.014$) in *D. subobscura* (Cirera and Aguadé 1998).

The distribution of polymorphism along the studied region shows a pronounced peak of nucleotide diversity at the beginning of exon 3 both in the O_{st} and in the O_{3+4} samples (Figure 2). This region includes different shared synonymous polymorphisms, which are in strong linkage disequilibrium within each arrangement (Figure 3). These sites are, in fact, combined in two main haplotypes that segregate at intermediate frequencies in both arrangements. It is the information in these sites that causes both O_{st} and O_{3+4} lines to form two subclusters in the neighbor-joining tree (Figure 7). Although the easiest explanation for this clustering of shared polymorphisms may be a gene conversion event between arrangements, none of the detected tracts include these sites. However, this possibility cannot be completely discarded since the algorithm proposed by Betrán *et al.* (1997) detects only part of the putative gene conversion tracts.

The sequences of *D. madeirensis* and *D. guanche* differ at most of the sites that define the two major haplotypes at the beginning of exon 3 in *D. subosbcura* (Figure 3). This would suggest that these sites already segregated in the O_3 arrangement at least before the split of the *D. madeirensis* lineage, which occurred 0.6–1 mya (Ramos-Onsins *et al.* 1998). If these polymorphisms were ancestral, O_{st} and O_{3+4} may have captured different haplotypes in their origin. Subsequently, the characteristic haplotype of each arrangement may have been transferred to the other arrangement by independent gene conversion events. Within each arrangement the transferred haplotype would have later increased in frequency. Alternatively, the haplotype not captured in the origin of O_{3+4} (or O_{st}) might have been transferred from O_3 to this arrangement by gene conversion, and thereafter it would have attained an intermediate frequency. At least two gene conversion events would also be required under this scenario, either from O_3 to O_{3+4} and O_{st}, or sequentially from O_3 to one of the two arangements and from this to the other one. The observed clustering of linkage disequilibria can, therefore, have a historical origin as gene conversion events including highly differentiated sites between arrangements may cause strong linkage disequilibrium between adjacent sites that may persist until recombination breaks the association. An example of this situation is the linkage disequilibrium detected between sites 1358 and 1360 (Figure 1). However, in both arrangements the newly transferred haplotype would have attained intermediate frequencies, which would be more compatible with selection. Although linkage with a replacement polymorphism maintained by balancing selection in both arrangements can be discarded at least for the *Acph-1* gene region, the possibility of epistatic selection acting on these sites to maintain mRNA structure (Kirby *et al.* 1995) cannot be rejected.

Age of the O_{st} and O_{3+4} arrangements: Silent nucleotide diversity at *Acph-1* is higher within O_{3+4} than within O_{st}. This result is in agreement with that previously found at the *rp49* gene region and is consistent with a more distant origin of the O_{3+4} than of the O_{st} arrangement from O_3. In fact, levels of silent variation in O_{3+4} and O_{st} arrangements can be used to date the origin of these arrangements. Rozas *et al.* (1999) proposed estimates of these events according to the expansion model (Slatkin and Hudson 1991; Rogers 1995) and to variation at the *rp49* gene region. These estimates can be contrasted with present data at the *Acph-1* gene region.

When the expansion model is applied to estimate the age of an inversion, it is assumed that nucleotide variation within each arrangement has not yet reached equilibrium and that it has accumulated independently. Therefore, polymorphic sites included in gene conversion tracts between arrangements have to be excluded from the analysis. Two criteria may confirm that nucleotide variation within the arrangement is still in the transient phase to equilibrium: first, the negative sign of Tajima's *D* statistics, which indicates an excess of rare variants, and second, the shape of the pairwise nucleotide difference distribution or mismatch distribution that is Poisson-like in expanding populations and that can be characterized by the raggedness (*r*) statistics (Harpending *et al.* 1993; Harpending 1994).

Variation at the *Acph-1* noncoding regions was used to estimate the age of the O_{st} and O_{3+4} arrangements. The coding region was not considered for this purpose since the significant excess of nonsynonymous polymorphism within O_{st} and the clustering of synonymous polymorphism at the beginning of exon 3 in both arrangements suggest that the pattern of polymorphism in this region was not neutral. Table 6 shows Tajima's *D* and the raggedness statistics for the noncoding region within the random samples of each arrangement. Tajima's *D* is negative in both arrangements and the raggedness statistic is significant for both O_{3+4} and O_{st}. These results together with the Poisson-like distribution in

TABLE 6

Tajima's *D* and raggedness statistics in noncoding regions

	O_{st}	O_{3+4}
No. of sites	797	797
S	32	30
Singletons	16	11
k	6.133	7.358
π	0.0077	0.0092
Tajima's *D*	−1.2087	−0.5090
P value	0.1042	0.3395
Raggedness	0.0181	0.0119
P value	0.0400*	0.0116*

These estimates were obtained from the random samples of each arrangement after excluding the noncoding polymorphic sites that are included in gene conversion tracts within O_{st}. *$0.01 < P < 0.05$.

both arrangements (results not shown) are consistent with the expansion model. According to this model, $\tau = 2\mu t$, where μ is the mutation rate. When variation is null at the moment of the expansion, as is the case when an inversion originates, τ corresponds to the average number of nucleotide differences (π in Table 6). The neutral mutation rate can be estimated from the rate of nucleotide substitutions in interspecific comparisons. Average divergence at *Acph-1* noncoding sites between *D. subobscura* random samples and *D. guanche* is 0.0536, which results in a rate of 14.89×10^{-9} substitutions per site and per year when the time of divergence between both species is assumed to be 1.8 million years (Ramos-Onsins *et al.* 1998). Therefore, O_{st} arrangement arose some 0.26 mya and O_{3+4} arose 0.31 mya. These estimates are very similar to those previously obtained for the *rp49* gene region (0.24 and 0.33 mya, respectively; Rozas *et al.* 1999).

Electrophoresis reveals only a minor fraction of variation at the ACPH-1 protein: Levels of variation at the ACPH-1 protein are much higher than those previously reported by electrophoresis. For instance, 15 and 7 amino acid variants have been detected for the *Acph-1*^{100} electrophoretic class within O_{st} and O_{3+4}, respectively (Figure 4). Some of these variants differ in the net charge of the mature peptide and thus they would be expected to present a different electrophoretic mobility. In addition, the different mobilities of the detected electromorphs are only partially explained by present results. In O_{3+4} the N/K replacement responsible for the *Acph-1*^{100}/*Acph-1*^{054} electrophoretic alleles implies one charge unit change, which can explain the lower mobility toward the anode of the latter electromorph. However, in O_{st} the two electromorphs differ by two charge units (E/K) and, consequently, the *Acph-1*^{054} electrophoretic allele would be expected to present a lower mobility in O_{st} than in O_{3+4}. Nevertheless, the electrophoretic mobility of the *Acph-1*^{054} alleles may also be affected by the number of glycosylation sites, since the N residue at site 241, which is lost in the *Acph-1*^{054}/O_{3+4} electrophoretic allele, forms part of a putative glycosylation site (Chung *et al.* 1996). In contrast, within both O_{st} and O_{3+4} the *Acph-1*$^{>100}$ electrophoretic alleles are due to the change of a basic amino acid (R) to a noncharged amino acid (I and S, respectively), which can explain their higher mobility toward the anode.

***Acph-1*^{054} and *Acph-1*$^{>100}$ electrophoretic alleles differ in O_{st} and O_{3+4} arrangements:** Our analysis at the nucleotide level has shown that the amino acid replacement responsible for the *Acph-1*^{054} and *Acph-1*$^{>100}$ electrophoretic variants is different within O_{st} and O_{3+4} arrangements. This result invalidates, therefore, all previous studies on linkage disequilibrium between chromosomal and allozymic polymorphisms at the *Acph-1* locus. Present results can explain the observed lack of consistent associations, which had been interpreted as indicative of a high genetic exchange between arrangements at *Acph-1*. Actually, each of the two different *Acph-1*^{054} alleles is in complete linkage disequilibrium with the corresponding chromosomal arrangement. The same argument holds for each of the *Acph-1*$^{>100}$ alleles, although in this case only one line within O_{st} was studied.

Forces responsible for allozyme polymorphism: It has long been discussed whether the presence of allozymic polymorphisms is due to the action of selection. Maintenance of old amino acid polymorphisms by balancing selection causes an excess of silent variation at adjacent sites (Hudson and Kaplan 1988). In contrast, a rapid increase in frequency of a recent allozyme by directional selection causes a lack of polymorphism in that allele (Hudson *et al.* 1994). As discussed by Hasson *et al.* (1998), data on intraspecific nucleotide variation in allozyme loci of *D. melanogaster* do not seem to support the notion that allozymes are old polymorphisms maintained by balancing selection, except for the *Adh* locus (Kreitman and Hudson 1991), and are generally more consistent with the action of directional selection. As the ACPH-1 protein was classically known to be polymorphic in *D. subobscura*, this enzymatic system seemed especially suitable for analyzing this aspect.

Acph-1$^{>100}$ lines within O_{3+4} cluster together in the gene genealogy and their nucleotide diversity is much lower than that present in *Acph-1*^{100}/O_{3+4} lines. These results point to a rather recent origin of the *Acph-1*$^{>100}$ electrophoretic allele within O_{3+4}. However, as this allele is present at very low frequencies in natural populations, it cannot be argued that positive selection is acting on this electrophoretic variant.

The results are completely different for the *Acph-1*^{054} electrophoretic allele. *Acph-1*^{054} lines do not cluster together in the gene genealogy within either O_{st} or O_{3+4}. In addition, levels of nucleotide diversity within each arrangement are similar for the *Acph-1*^{054} and *Acph-1*^{100} lines. These data are not consistent with a recent and rapid increase in frequency of the *Acph-1*^{054} electropho-

retic allele within each chromosomal class and indicate that both *Acph-1*^{054}/*Acph-1*^{100} polymorphisms are rather old although likely not older than each arrangement. On the other hand, the lack in each case of a peak of silent d_{xy} values around the site responsible for the replacement polymorphism does not favor the hypothesis of balancing selection maintaining the corresponding allozyme variants, at least for a long period of time. However, there is some evidence that O_{st}/*Acph-1*^{054} lines have accumulated variation differentially from O_{st}/*Acph-1*^{100} lines, which may be consistent with the *Acph-1*^{054} allele within O_{st} being older than the *Acph-1*^{054} allele within O_{3+4}. Alternatively, both alleles could be equally old if recombination were higher within O_{3+4}, since it could have shuffled variation between the *Acph-1*^{054} and *Acph-1*^{100} alleles in this chromosomal class.

Excess amino acid polymorphism in the O_{st} arrangement: The number of nonsynonymous polymorphisms is higher within O_{st} than within O_{3+4}, which causes the nucleotide diversity at nonsynonymous sites to be nearly two times larger in O_{st} than in O_{3+4}. In O_{st} most of the nonsynonymous polymorphisms are singletons as indicated by the marginally significant ($0.1 > P > 0.05$) Tajima's D value and significant ($P < 0.05$) Fu and Li's statistics for nonsynonymous polymorphisms. The large negative F_s statistic (Fu 1997) in O_{st} also supports an excess of rare protein haplotypes and, thus, of single or recent nonsynonymous mutations within this arrangement. A negative value of these statistics was also detected in O_{3+4}, although the presence of only seven polymorphisms in this class makes the application of these tests more questionable. In addition, the McDonald-Kreitman test revealed an excess of nonsynonymous polymorphisms within the O_{st} but not within the O_{3+4} arrangement; also, a G test of independence showed that the number of nonsynonymous polymorphisms within O_{st} was significantly higher than the corresponding number within O_{3+4}. Although it would be tempting to argue that selection may account for the detected excess of nonsynonymous polymorphism within O_{st}, the fact that most of them are singletons does not favor this argument.

An excess of nonsynonymous or replacement polymorphisms has been previously reported for mitochondrial genes in Drosophila (Kaneko *et al.* 1993; Ballard and Kreitman 1994; Rand *et al.* 1994; Rand and Kann 1996) as well as in mice (Nachman *et al.* 1994), man, and chimpanzee (Nachman *et al.* 1996; Wise *et al.* 1998). In nuclear genes this deviation from neutrality has been detected only at the *Adh* (alcohol dehydrogenase; Miyashita *et al.* 1996) and *Pgi* (phosphoglucose isomerase; Terauchi *et al.* 1997) genes in plants, and at the *Gld* (glucose dehydrogenase; Hamblin and Aquadro 1997) and *pn* (prune; Simmons *et al.* 1994) genes in Drosophila.

These data may be consistent with the nearly neutral model of molecular evolution (Ohta 1992), which proposes that mutations causing amino acid replacements are slightly deleterious. Slightly deleterious mutations may persist as polymorphic but they are unlikely to become fixed. The fate of these slightly deleterious mutations, however, will be affected by the effective size (N_e), since they will behave as neutral in small populations but will be efficiently eliminated by negative selection in large populations. The effectiveness of selection acting on weakly selected mutations is also affected by the recombination rate. In regions with a drastic reduction of recombination, levels of polymorphism of mildly selected mutations are closer to those of neutral variants. Therefore, an excess of slightly deleterious polymorphisms is expected in such regions (Charlesworth 1994).

Protein evolution in Drosophila seems to conform with neutral predictions (Zeng *et al.* 1998), suggesting that the effective size of natural populations is large enough for selection acting against slightly deleterious nonsynonymous mutations to be efficient. Interestingly, the above-described examples showing an excess of nonsynonymous polymorphisms in Drosophila are detected in mtDNA genes with a smaller effective size relative to nuclear genes and with no recombination, or in nuclear genes (*pn* and *Gld*) located in regions with a somewhat reduced recombination rate (Kliman and Hey 1993). A reduction of recombination is also expected in genes located near breakpoints of inversions, where recombination is highly suppressed in heterokaryotypes. Therefore, present results may be consistent with a smaller effective size of the O_{st} *vs.* the O_{3+4} arrangement assuming that nonsynonymous mutations are slightly deleterious. There are three not mutually exclusive factors that may contribute to a small effective size of the O_{st} arrangement.

First, in the sampled population from Galicia the average frequencies of O_{3+4} and O_{st} in 1989 were estimated as 0.767 and 0.147, respectively (Rodríguez-Trelles *et al.* 1996). The difference in the frequency of both arrangements may contribute by itself to a smaller effective size of O_{st}. Second, this difference causes the frequency of O_{st} homokaryotypes in the population to be much lower than that of O_{3+4} homokaryotypes. As recombination near breakpoints will only be free in homokaryotypes, recombination at the *Acph-1* gene region, and thus the effectiveness of selection, would be lower within O_{st} than within O_{3+4}. Therefore, the putative lower effective size of the O_{st} arrangement may be due both to its lower frequency and to the consequently lower recombination in this arrangement. However, both arguments are based on the current lower frequency of O_{st} in Galicia, but the frequency of these arrangements varies latitudinally in Europe. An RFLP survey at the *rp49* gene region (Rozas *et al.* 1995) failed to detect genetic differentiation within gene arrangement between European populations, suggesting that migration is large enough to homogenize the genetic content of

a particular arrangement. If sequencing studies for different regions confirmed that observation, for each gene arrangement European populations could be considered as a unit and the mean frequencies of O_{st} and O_{3+4} arrangements would be much more alike.

A third factor that may contribute to a smaller effective size of O_{st} relative to O_{3+4} is the more recent origin of the former gene arrangement. In fact, the origin of an inversion implies an extreme bottleneck with an effect on the effective size (N_e) of the new arrangement that persists over generations. After the initial bottleneck, the frequency of the new arrangement starts increasing until it reaches its equilibrium frequency, which can be envisaged as a selective sweep. Although the increase in frequency of the new arrangement should be very rapid, the reduced N_e does not recover so rapidly. Only when the effective size of the new arrangement had increased would selection be more efficient in eliminating slightly deleterious mutations. Therefore, only O_{st}, a younger arrangement than O_{3+4}, might be still reflecting the decreased efficiency of selection linked to the origin and establishment of a new arrangement.

However, if the proposed argument is true, an excess of replacement polymorphisms would be expected at other genes located near inversion breakpoints. The only available data on this aspect are those reported for the *rp49* region of *D. subobscura* (Rozas *et al.* 1999) and for the *hsp83* gene in *standard* and *In(3L)Payne* arrangements of *D. melanogaster* (Hasson and Eanes 1996). Although no replacement polymorphism was detected in either of these genes, this does not constitute evidence against the proposed argument. In fact, the proteins encoded by these genes have been highly conserved during evolution and they are, therefore, subject to strong purifying selection against amino acid replacement substitutions. In addition, the age of inversion *In(3L)Payne* was estimated to be around 0.36 million years (Hasson and Eanes 1996), an estimate more similar to that of O_{3+4} than to that of O_{st}. Only data on nucleotide variation at a large number of genes located near breakpoints of inversions with different ages will show whether the excess of replacement polymorphism detected at the *Acph-1* gene in O_{st} is a general situation for rather young inversions and reflects therefore the accumulation of slightly deleterious mutations during the phase of establishment of the new inversion in the population.

We thank J. Rozas for critical comments on the manuscript and S. Ramos-Onsins for sharing his computer program SignTestLD. We also thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. This work was supported by a predoctoral fellowship from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain, to À.N.-S. and by grants PB94-923 from Comisión Interdepartamental de Ciencia y Tecnología, Spain and 1995SGR-577 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain, to M.A.

LITERATURE CITED

Aguadé, M., 1988 Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and noninverted chromosomes. Genetics **119:** 135–140.

Aquadro, C. F., A. L. Weaver, S. W. Schaeffer and W. W. Anderson, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura*: the amylase gene region. Proc. Natl. Acad. Sci. USA **88:** 305–309.

Babcock, C. S., and W. W. Anderson, 1996 Molecular evolution of the Sex-ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. Mol. Biol. Evol. **13:** 297–308.

Ballard, J. W. O., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

Bénassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

Betrán, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146:** 89–99.

Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

Chung, H.-J., C. Shaffer and R. J. MacIntyre, 1996 Molecular characterization of the lysosomal acid phosphatase from *Drosophila melanogaster*. Mol. Gen. Genet. **250:** 635–646.

Cirera, S., and M. Aguadé, 1998 Molecular evolution of a duplication: the sex-peptide (*Acp70A*) gene region of *Drosophila subobscura* and *D. madeirensis*. Mol. Biol. Evol. **15:** 988–996.

Fontdevila, A., C. Zapata, G. Alvarez, L. Sanchez, J. Méndez *et al.*, 1983 Genetic coadaptation in the chromosomal polymorphism of *Drosophila subobscura*. Genetics **105:** 935–955.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Hamblin, M. T., and C. F. Aquadro, 1997 Contrasting patterns of nucleotide sequence variation at the glucose dehydrogenase (*Gld*) locus in different populations of *Drosophila melanogaster*. Genetics **145:** 1053–1062.

Harpending, H. C., 1994 Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. **66:** 591–600.

Harpending, H. C., S. T. Sherry, A. R. Rogers and M. Stoneking, 1993 The genetic structure of ancient human populations. Curr. Anthropol. **34:** 483–496.

Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144:** 1565–1575.

Hasson, E., I.-N. Wang, L.-W. Zeng, M. Kreitman and W. F. Eanes, 1998 Nucleotide variation in the triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **15:** 756–769.

Hudson, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral evolution based on nucleotide data. Genetics **116:** 153–159.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–120 in *Mammalian Protein Metabolism*, edited by H. W. Munro. Academic Press, New York.

Kaneko, M., Y. Satta, E. T. Matsuura and S. I. Chigusa, 1993 Evolution of the mitochondrial ATPase 6 gene in Drosophila: unusu-

ally high level of polymorphism in *D. melanogaster*. Genet. Res. **61:** 195–204.
Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.
Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA **92:** 9047–9051.
Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.
Kreitman, M., and M. Aguadé, 1986 Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA **83:** 3562–3566.
Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.
Krimbas, C. B., 1992 The inversion polymorphism of *Drosophila subobscura*, pp. 127–220 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.
Kumar, S., K. Tamura and M. Nei, 1994 MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput. Appl. Biosci. **12:** 231–238.
Kunze-Mühl, E., and E. Müller, 1958 Weitere Untersuchungen über die chromosomale Struktur und natürlichen Strukturtypen von *Drosophila subobscura*. Chromosoma **9:** 559–570.
Larruga, J. M., and W. Pinsker, 1984 The importance of geographic isolation, structural rearrangement and speciation for the genetic divergence of chromosome *O* in *Drosophila subobscura*, *D. guanche* and *D. madeirensis*. Z. Zool. Syst. Evolutions-forsch. **22:** 103–113.
Lewontin, R. C., 1995 The detection of linkage disequilibrium in molecular sequence data. Genetics **140:** 377–388.
Loukas, M., C. B. Krimbas and Y. Vergini, 1979 The genetics of *Drosophila subobscura* populations. IX. Studies on linkage disequilibrium in four natural populations. Genetics **93:** 497–523.
Loukas, M., C. B. Krimbas and K. Morgan, 1980 The genetics of *Drosophila subobscura* populations. XIV. Further data on linkage disequilibrium. Genetics **95:** 757–768.
MacIntyre, R. J., M. R. Dean and G. Batt, 1978 Evolution of acid-phosphatase-1 in the genus *Drosophila*. Immunological studies. J. Mol. Evol. **12:** 121–142.
Maddison, W. P., and D. R. Maddison, 1992 MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, MA.
McDonald, J. H., 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **13:** 253–260.
McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.
Miyashita, N. T., H. Innan and R. Terauchi, 1996 Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. Mol. Biol. Evol. **13:** 433–436.
Muller, H. J., 1940 Bearings of the Drosophila work on systematics, pp. 185–268 in *New Systematics*, edited by J. Huxley. Clarendon Press, Oxford.
Nachman, M. W., S. N. Boyer and C. F. Aquadro, 1994 Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.
Nachman, M. W., W. M. Drown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.
Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics **146:** 695–709.
Navarro-Sabaté, À., M. Aguadé and C. Segarra, 1999 The *Acid phosphatase-1* gene region in the *Drosophila* species of the *subobscura* cluster. Hereditas **130:** 65–75.
Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.
Popadić, A., and W. W. Anderson, 1994 The history of a genetic system. Proc. Natl. Acad. Sci. USA **91:** 6819–6823.
Popadić, A., and W. W. Anderson, 1995 Evidence for gene conversion in the Amylase multigene family of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12:** 564–572.
Popadić, A., D. Popadić and W. W. Anderson, 1995 Interchromosomal exchange of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12:** 938–943.
Popadić, A., R. A. Norman, W. W. Doanet and W. W. Anderson, 1996 The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. Mol. Biol. Evol. **13:** 883–888.
Prevosti, A., G. Ribó, L. Serra, M. Aguadé, J. Balaña *et al.*, 1988 Colonization of America by *Drosophila subobscura*: experiment in natural populations that support the adaptive role of chromosomal-inversion polymorphism. Proc. Natl. Acad. Sci. USA **85:** 5597–5600.
Ramos-Onsins, S., C. Segarra, J. Rozas and M. Aguadé, 1998 Molecular and chromosomal phylogeny in the obscura group of Drosophila inferred from sequences of the *rp49* gene region. Mol. Phylogenet. Evol. **9:** 33–41.
Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitocohondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. **13:** 735–748.
Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.
Rodríguez-Trelles, F., G. Alvarez and C. Zapata, 1996 Time-series analysis of seasonal changes of the *O* inversion polymorphism of *Drosophila subobscura*. Genetics **142:** 179–187.
Rogers, A. R., 1995 Genetic evidence for a pleistocene population explosion. Evolution **49:** 608–615.
Rozas, J., and M. Aguadé, 1990 Evidence of extensive genetic exchange in the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. Genetics **126:** 417–426.
Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. Proc. Natl. Acad. Sci. USA **90:** 8083–8087.
Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. Proc. Natl. Acad. Sci. USA **91:** 11517–11521.
Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.
Rozas, J., C. Segarra, C. Zapata, G. Alvarez and M. Aguadé, 1995 Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: lack of geographic subdivision within chromosomal arrangements in Europe. J. Evol. Biol. **8:** 355–367.
Rozas, J., C. Segarra, G. Ribó and M. Aguadé, 1999 Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. Genetics **151:** 189–202.
Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.
Segarra, C., G. Ribó and M. Aguadé, 1996 Differentiation of Muller's elements *D* and *E* in the obscura group of Drosophila. Genetics **144:** 139–146.
Simmons, G. M., W. Kwok, P. Matulonis and T. Venkatesh, 1994 Polymorphism and divergence at the *prune* locus in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **11:** 666–671.
Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.
Sperlich, D., H. Feuerbach-Mravlag, P. Lange, A. Michaelidis and A. Pentzos-Daponte, 1977 Genetic load and viability distribution in central and marginal populations of *Drosophila subobscura*. Genetics **86:** 835–848.
Staden, R., 1982 Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10:** 4731–4751.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.
Tamura, K., 1992 Estimation of the number of nucleotide substitu-

tions when there are strong transition-transversion and G+C-content biases. Mol. Biol. Evol. **9:** 678–687.

Terauchi, R., T. Terachi and N. T. Miyashita, 1997 DNA polymorphism at the *Pgi* locus of a wild yam, *Dioscorea tokoro*. Genetics **147:** 1899–1914.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

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

Wise, C. A., M. Sraml and S. Easteal, 1998 Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics **148:** 409–421.

Zeng, L.-W., J. M. Comeron, B. Chen and M. Kreitman, 1998 The molecular clock revisited: the rate of synonymous vs. replacement change in *Drosophila*. Genetica **102/103:** 369–382.

Communicating editor: A. G. Clark