Evolutionary History of the Sex-Peptide (Acp70A) Gene Region in Drosophila melanogaster

Susanna Cirera and Montserrat Aguadé

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

Manuscript received January 31, 1997
Accepted for publication May 30, 1997

ABSTRACT

In Drosophila the products of the seminal fluid stimulate oviposition and suppress remating in the female. Of all the accessory gland peptides (Acp’s) involved in these two responses, the sex-peptide (coded by the Acp70A gene) is among the best characterized at the functional level. A 1.2-kb fragment encompassing the Acp70A gene of nine lines from a natural population of D. melanogaster and one allele of D. sechellia was sequenced to study the forces shaping nucleotide variation within and between species. The coding region of D. simulans and D. mauritiana was also sequenced. A Ser to Ala replacement polymorphism at the last position of the signal peptide was detected in D. melanogaster. The Ser and Ala alleles are at intermediate frequencies. The level of nucleotide variation is lower for the derived Ala allele, which is compatible with a recent origin and an increase in frequency due to positive selection. Variation at the 5’ flanking region is structured in two major highly differentiated haplotypes, whose distribution does not conform to neutral expectations. Selective and/or historical factors could contribute to the observed overall patterning of nucleotide variation at the Acp70A region.

In Drosophila, the seminal fluid transferred from the male to the female during copulation induces physiological and behavioral changes in the female: both ovulation and oviposition are stimulated, and mated females reject courting males by extruding their ovipositor (Chen 1984). Electrophoretic analysis of soluble proteins in extracts from accessory glands of different species of Drosophila has shown high levels of interspecific variation (Chen 1976). Several peptides of the male accessory secretions have been identified (Acp’s for accessory gland proteins) and the corresponding genes (Acp’s) have been cloned and sequenced: the sexpeptide gene (or Acp70A) (Chen et al. 1988; Styger 1992), Acp26Aa and Acp26Ab (Monsma and Wolfner 1988), and Acp57D (Simmerl et al. 1995). At the functional level, the sex-peptide is among the most exhaustively studied in different species of the melanogaster group of Drosophila (Chen et al. 1988; Chen and Balmer 1989; Aigaki et al. 1991; Styger 1992; E. Kubli, personal communication).

In D. melanogaster the sex-peptide (SP) is synthesized as a 55-amino acid precursor containing a signal peptide 19 amino acids long that is cleaved off during secretion in the accessory glands (Chen et al. 1988). The biological activity of the peptide has been established in two kinds of experiments: injection of the purified peptide into virgin females (Chen et al. 1988) and ectopic expression in transgenic virgin females (Aigaki et al. 1991). In both experiments females showed increased ovulation/oviposition rate and reduced receptivity. The Acp70A is 266 bp long, from the transcription start site to the polyadenylation site, and has two exons and a small intron (Chen et al. 1988). In the melanogaster group the Acp70A gene is a single copy gene, in contrast to the obscura group where the gene is duplicated and both copies are transcribed (S. Cirera, Y. Choffat, E. Kubli and M. Aguadé, unpublished results). Expression studies in transgenic D. melanogaster flies with different putative Acp70A promoter fragments (830 and 210 bp immediately adjacent to the transcription initiation site) indicate that the 210-bp fragment is sufficient for correct tissue- and time-specific expression (Styger 1992). Although according to these results the upstream region present only in the 830-bp fragment should not be included in the Acp70A gene promoter, the high sequence conservation of part of this region between the distantly related species D. melanogaster and D. subobscura (S. Cirera, Y. Choffat, E. Kubli and M. Aguadé, unpublished results) points to its possible involvement in the regulation of this gene.

Despite the interesting biology of the Acp’s, only the genes coding for two of these proteins, Acp26Aa and Acp26Ab, have been studied at the population level. The Acp26Aa gene revealed a high level of amino acid replacement variation both within and between species (Aguadé et al. 1992); insertion/deletion differences contributed to this between-species variation. In contrast, the level of within and between species amino acid replacement variation was lower for the tightly linked Acp26Ab gene.

The study of nucleotide variation within and/or between species has proved a powerful approach for de-
detecting the action of natural selection. Within species, natural selection acting on a particular polymorphism should affect variation in linked neutral sites. In fact, a positively selected variant on its way to fixation or to equilibrium reduces the level of intraspecific variation, while old balanced polymorphisms are expected to generate an excess of linked silent polymorphism relative to silent divergence. In this context, comparison of intra- and interspecific variation has allowed detection of selection acting within species: for example, balancing selection maintaining the Adh allozyme polymorphism in D. melanogaster (Kreitman and Hudson 1991), or directional selection as one of the main forces responsible for the reduced levels of variation observed in regions of low recombination in both D. melanogaster and D. simulans (Begun and Aquadro 1991; Berry et al. 1991; Martín-Campos et al. 1992). The within/between species approach has also revealed directional selection shaping protein evolution, for example, for Adh and G6pd in D. melanogaster and D. simulans (McDonald and Kreitman 1991; Eanes et al. 1993), which exhibit an excess of fixed replacement differences between species. In other cases, only the within species pattern of variation has revealed the action of natural selection, as in the red and white regions of D. melanogaster (Hudson et al. 1994; Kirby and Stephan 1995, 1996). To study the forces shaping variation in the Acp70A region, which is located in a region with normal levels of recombination (Hudson and Kaplan 1995), we have sequenced a 1.2-kb fragment in nine lines of a Spanish natural population of D. melanogaster, one line of D. sechellia, and also the coding region of their sibling species D. simulans and D. mauritiana. The region studied encompasses the Acp70A gene and its promoter as well as a further upstream region that might also be involved in regulating the expression of this gene (see above).

**MATERIALS AND METHODS**

**Fly stocks:** Nine isochromosomal lines for the third chromosome of D. melanogaster were obtained from a sample collected in Montemayor (Córdoba, Spain) in 1990. The balancer stock TM6/MK7S was used for extracting wild chromosomes, using one F1 virgin female from each isofemale line. One line each of D. simulans (from Barcelona), D. sechellia, and D. mauritiana (from the Umed Fly Stock Center) were also included in the present study.

**DNA preparation and sequencing:** Genomic DNA from the Montemayor lines was CsCl purified (Bingham et al. 1981), and that from D. simulans and D. mauritiana was extracted using protocol 48 in Ashburner (1989) with minor modifications. Based on the sequence from D. melanogaster (kindly provided by E. Kubli), synthetic oligonucleotides on average 220 nucleotides apart were designed for both strands according to Rozas (1991). For lines of D. melanogaster, a 1.2-kb fragment was PCR amplified, while for D. simulans and D. mauritiana a 0.5-kb fragment including the coding and the 5' flanking regions of the Acp70A gene was amplified. In each case two independent PCR amplifications were performed using one kinased primer. Single-stranded DNA of the amplified products was obtained by λ-exonuclease treatment (Hilorch and Ochman 1989). Both strands were sequenced by the dideoxy chain termination method (Sanger et al. 1977) using T7 DNA polymerase or Sequenase. One allele from D. sechellia was isolated by screening a random genomic library (Aquadro et al. 1992) using the 1.24-kb PCR-amplified fragment of D. melanogaster as a probe. A 2.9-kb BamHI-Cld fragment from a positive phage was cloned in pBluescript II in both orientations. A set of nested deletions was obtained for each clone (Henikoff 1984) and subclones covering 1.6 kb were sequenced by the dideoxy method using double-stranded DNA. Sequences were assembled using Staden's programs (1982) and multiply aligned with the Clustal V program (Higgins et al. 1992). The sequences reported in this paper have been deposited in EMBL sequence database library under accession numbers X99407-X99418.

**Sequence analysis:** The MacClade program version 3.0 (Maddison and Maddison 1992) was used to edit the sequences for further analyses, and the DnaSP version 2.08 program (Rozas and Rozas 1997) was used for most intraspecific and some interspecific analyses. The level of polymorphism in D. melanogaster was estimated as nucleotide diversity or π (Nei 1987), which is the average number of nucleotide differences per site.

Linkage disequilibrium, which measures the degree of association between nucleotide variants of different polymorphic sites, was estimated as R² (Hill and Robertson 1968) using only informative sites (for which the rarest variant is present more than once). The four-gamete test (Hudson and Kaplan 1985), based on the presence of all four gametic types in the sample for any pair of polymorphisms, was used to infer the minimum number of recombination events. The recombination parameter 4Nc (where N is the effective population size and c the recombination rate) was estimated by the method of Hudson (1987), which assumes a neutral equilibrium model and is based on the variance in the number of pairwise differences.

Interspecific divergence was estimated separately for coding and noncoding DNA regions (Nei and Gojobori 1986), and the genetic distances were corrected for multiple hits according to Jukes and Cantor (1969). Phylogenetic trees were built by the neighbor-joining method (Saitou and Nei 1987) using the MEGA program (Kumar et al. 1994).

**Statistical methods:** Possible departures from the mutation-drift equilibrium hypothesis were tested using the Tajima test (Tajima 1989) and the HKA test (Hudson et al. 1987). The Tajima test analyzes the allele frequency spectrum of sites segregating in a set of sequences and therefore requires data only from intraspecific variation. The HKA test is based on the prediction of the neutral hypothesis of a direct relationship between levels of polymorphism and divergence; it requires data from interspecific divergence of at least two loci and data from intraspecific polymorphism for the same regions from one species. In addition, the haplotype test (Hudson et al. 1994), which is based on the occurrence of subsets of a sample with low levels of variation, was used.

**RESULTS**

**Interspecific variation in the melanogaster subgroup**

**DNA sequence comparison:** One hundred nucleotide and 19 length differences were detected when comparing the 1.2-kb region between the nine lines of D. melanogaster and the allele of D. sechellia. Thirty-nine and seven of these changes, respectively, segregated in D. melanogaster (Figure 2). All length differences were in noncoding regions and their length varied between one
and 29 bp. There were six nonsynonymous differences between *D. melanogaster* and *D. sechellia* resulting in six amino acid replacements. One additional nonsynonymous difference segregated in *D. melanogaster*.

Estimated nonsynonymous divergence is higher between *D. melanogaster* and *D. sechellia* (0.053) than between *D. melanogaster* and either *D. simulans* (0.028) or *D. mauritiana* (0.028), due to three additional nonsynonymous substitutions in the *D. sechellia* lineage. Estimated silent divergence in the coding region is, however, similar in the three comparisons (0.112, 0.112 and 0.101, respectively). Table 1 summarizes nucleotide divergence at the Acp70A region between *D. melanogaster* and *D. sechellia*. For the whole region, the level of silent divergence (0.078) falls within the range of previous estimates of divergence between *D. melanogaster* and *D. simulans*, *D. mauritiana* or *D. sechellia*: from 0.052 for *Adh* (KREITMAN and HUSON 1991) to 0.150 for *Pgld* (BEGUN and AQUADRO 1993). Silent divergence for the coding region is in the high part of that range (0.112). Divergence at the 3′ flanking region (0.035) is, as in other genes, much lower.

**Protein sequence comparison:** Figure 1 shows the comparison of the predicted amino acid sequences of SP in the four species of the *melanogaster* subgroup. The sequence of *D. melanogaster* differs from that of the other three species of the *melanogaster* subgroup by three amino acid replacements (two in the signal peptide and one at the active peptide). *D. sechellia* presents three additional amino acid changes (one at the signal peptide and two at the active peptide). The changes in the active peptide are located in the N-terminal half, where most amino acid changes and length differences have been observed, not only relative to *D. suzukii* (SCHMIDT et al. 1993), a more distantly related species of the *melanogaster* group, but also relative to *D. subobscura*, a member of the *obscura* group (S. CIRERA, Y. CIOFFATTI, E. KUBLI and M. AGUADÉ, unpublished results). As in the case of the Acp26Ab protein, no length differences were found between species for the sex-peptide or Acp70A protein. The mean number of amino acid differences for the sex-peptide between the three species and the consensus sequence of *D. melanogaster* (3.0 out of 55 amino acids or 0.055 per amino acid) is of the same order as that for the Acp26Ab protein (4.3 out of 90 or 0.048) but lower than for the Acp26Aa protein (73.7 out of 250 or 0.295).

**Intraspecific variation in *D. melanogaster***

**Sequence comparison:** Figure 2 and Table 1 give a summary of polymorphism in the 1.2-kb sequenced region for the nine lines of *D. melanogaster*. Thirty-nine nucleotide and seven length polymorphisms were detected. We only found one nonsynonymous polymorphism (at position 835), which results in the replacement of a serine (neutral-polar amino acid) by an alanine (neutral-nonpolar amino acid) at the last position of the signal peptide; these two variants segregated at intermediate frequencies in our sample: we found four serine and five alanine lines. Six length polymorphisms (three deletions of 2 bp, one deletion of 10 bp, one deletion of 29 bp and one insertion of 7 bp) are located

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of polymorphism and divergence at the Acp70A region</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>5′ flanking</td>
</tr>
<tr>
<td>No. of sites</td>
</tr>
<tr>
<td>No. polymorphisms</td>
</tr>
<tr>
<td>π</td>
</tr>
<tr>
<td>Divergence</td>
</tr>
<tr>
<td>Div/π</td>
</tr>
</tbody>
</table>

s, synonymous in coding regions and silent in noncoding regions; ns, nonsynonymous; e1, exon 1; i, intron; e2, exon 2. Only the *D. sechellia* sequence was used to estimate divergence.
Figure 2.—Polymorphism at the Acp70A region of D. melanogaster. E1, exon 1; E2, exon 2; I, intron; i#, insertion of #bp; d#, deletion of #bp; anc, ancestral; *, the same nucleotide as in the ancestral sequence; +/-, presence or absence of insertions and deletions; ns, nonsilent change; Rc, recombination event; [ ]. polymorphic site included in d10. The ancestral sequence was established by comparison with the sequence of D. sechellia. The clear boxes indicate putative recombination or gene conversion tracts, detected by visual inspection and for the nine mates for other accessory gland genes (0.006 for Acj26Al)). These lines were named Ser2 lines. Polymorphisms both for nucleotide substitutions and insertions/deletions were used for estimating linkage disequilibrium. Polymorphic sites 383 (included in a deleted fragment in some of the lines) and 795 (segmenting for three nucleotides) were not considered for this analysis. Three hundred and forty-six out of 741 pairwise comparisons showed a significant linkage disequilibrium by the chi-square test (without correcting for multiple comparisons). Pairs of polymorphic sites with significant linkage disequilibrium are not distributed at random along the sequenced fragment. There are two clusters of linkage disequilibria (Figure 4) that define two regions or linkage disequilibria blocks (big shaded boxes in Figure 2). The first block, located at the 5' flanking region, extends between polymorphisms at sites 243 and 537. The second block, between polymorphisms at sites 693 and 1138, encompasses the Acp70A gene.

Despite the observed linkage disequilibria, there is evidence in our sample of some recombination. In the history of our sample, a minimum number of three recombination events was inferred by the four-gamete test (Hudson and Kaplan 1985), between positions (282, 611), (611, 693) and (905, 1143), respectively. Two of these events could reflect gene conversion events, but they could also be due to recombination. Each of them affects a single line: lines M40 and M36, respectively (see smaller lighter boxes inside the big shaded boxes in Figure 2). The third recombination
event between positions 611 and 693 clearly separates the two linkage disequilibria blocks, which could have different evolutionary histories.

For the whole sequenced region the 4Ne estimate (Hudson 1987) was 9.3, which results in a per nucleotide estimate of 0.008. This estimate is of the same order as reported for other genes also located in regions with normal levels of recombination: 0.004 for Sod (Hudson et al. 1994) and 0.022 for G6pd (Eanes et al. 1996), for example. The 4Ne estimate drops to zero for the left linkage disequilibria block (see above), concordant with the lack of recombination events detected in the history of this fragment.

**Gene genealogy and nucleotide variation:** The sequenced fragment was divided into two regions that encompass, respectively, the two linkage disequilibria blocks: 5' or left region (from site 1 to site 620) and 3' or right region (from site 621 to site 1160). Figure 5, A and B, shows the gene genealogy for the two regions of the nine lines of *D. melanogaster* using *D. sechellia* as the outgroup. The five Ala lines cluster together with the Ser2 lines when considering the left region, while they form a distinct group from all Ser lines (both Ser1 and Ser2) when considering the right region. Furthermore, the two additional recombination events are also reflected in these trees. In the left region tree (Figure 5A), the M40 line is the one closest to the Ser1 lineage; likewise, in the right region tree (Figure 5B) the M36 line is the Ala line closest to the Ser lines.

Table 2 summarizes nucleotide variation for the different lineages. As compared to the overall estimate, variation within the two variable lineages (Ser2 and Ala) is low both when the whole region and when only the left part are considered. In contrast, for the right part only the Ala lineage shows a lower level of variation than the complete sample of nine lines. Also, the pattern of polymorphism along the studied region is different when all the lines are considered (Figure 3A) than when the Ser1 lines are excluded (Figure 3B), and when only the alanine lines are considered (Figure 3C). In the left region, polymorphism decreases considerably in both cases B and C, while only in case C it drops to zero for most of the right region where the *Aep70A* gene is located. Table 2 also shows the average number of nucleotide differences between the three different haplotypes or lineages (Ser1, Ser2 and Ala).
The two major 5' haplotypes in our sample have little variation (Table 2), despite being highly differentiated. We estimated the probability that a subsample of seven lines has the observed number of polymorphisms for different fragments of this region (HUDSON et al. 1994). Given that estimates of 4Nc have large variances, we have performed Hudson's haplotype test using 10 as the recombination parameter, but also the more conservative values of 5 and 0. When considering the 5' linkage disequilibria block, the probability that between coordinates 243 and 537 a subsample of seven alleles has only three polymorphisms is lower than 5% for recombination parameters 10 (P = 0.017) and 5 (P = 0.032), and slightly higher than 5% (P = 0.08) for a recombination parameter equal to 0. However, when the fragment including the putative gene conversion tract is not considered, the probability of zero polymorphisms in a subsample of seven alleles (between polymorphisms at coordinates 308 and 537) is in all cases lower than 5% (0.001, 0.004 and 0.01, respectively). This departure from neutral expectations of the pattern of variation at the left region can also be observed when estimating that probability from sliding a window of 10 polymorphisms along the studied region (Figure 6A). We have also estimated the probability of the maximum subset of lines in the dataset with zero polymorphisms,

![Figure 5](image5.png)

**Figure 5.**—Gene genealogy of the nine *D. melanogaster* lines reconstructed by the neighbor-joining method using sequence information of the left region (from site 1 to site 620) that includes the left linkage disequilibria block (A) and of the right region (from site 621 to the end) that includes the right linkage disequilibria block (B). Black squares, serine; light squares, alanine.

![Figure 6](image6.png)

**Figure 6.**—Sliding window analysis of haplotype test. (A) Probability (P-value) of a subsample of seven lines with zero polymorphisms in a window of 10 polymorphisms. (B) Probability (P-value) of the maximal subset of lines with zero polymorphisms in a window of 10 polymorphisms.

---

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>Whole region</th>
<th>5' region</th>
<th>3' region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>16.1</td>
<td>10.9</td>
<td>4.9</td>
</tr>
<tr>
<td>Ser1 (n = 2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ser2 (n = 2)</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Ala (n = 5)</td>
<td>6</td>
<td>4.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Ser1 vs. Ser2</td>
<td>18.2</td>
<td>13.3</td>
<td>4.8</td>
</tr>
<tr>
<td>Ser1 vs. Ala</td>
<td>15.8</td>
<td>10.5</td>
<td>5.3</td>
</tr>
<tr>
<td>Ser2 vs. Ala</td>
<td>10.1</td>
<td>4.8</td>
<td>5.2</td>
</tr>
</tbody>
</table>
using 10 as the recombination parameter and sliding a window of 10 polymorphisms (see KIRBY and STEPHAN 1995). In this case, the same 5' fragment stands out for its low probability (Figure 6B) relative to the null distribution after randomization of sites (KIRBY and STEPHAN 1995). This 5' haplotype, which is found in rather high frequency in our sample, would therefore have too little variation to be explained by neutral causes if the population is in equilibrium.

**DISCUSSION**

The interspecific comparison indicates that serine at the last position of the signal peptide is the ancestral residue in the *melanogaster* subgroup. The alanine allele would therefore have originated from a serine allele in the *D. melanogaster* lineage. In our sample from this species, the serine and alanine alleles are present at intermediate frequencies. Although the Ser/Ala replacement polymorphism does not affect the active peptide, it could affect the processing of the preprotein. Alanine is the most frequent amino acid at the position preceding cleavage in eukaryotes (in 80 out of 161 peptides reported by VON HEIJNE (1986)), although serine has also been found quite often (in 20 out of those 161 peptides). Using VON HEIJNE’s method (1986) Ser and Ala alleles have the same most likely cleavage site for the signal peptide, although the score is slightly higher for Ala than for Ser alleles (63.9 vs. 58.2). In prokaryotic *in vitro* systems the percentage of processing at 90 min is higher if alanine is the last amino acid of the signal peptide (FIRES et al. 1990); however, nothing is known about processing *in vivo* in prokaryotic and eukaryotic systems.

Even if the Ser and Ala alleles showed any functional differences that would not in itself be an indication that they affect fitness. The Ser/Ala polymorphism could be a transient neutral polymorphism; alternatively, the alanine allele could be subject to positive selection, either directional or balancing, and be on its way to fixation or to equilibrium. Neither comparison of within and between species variation nor analysis of intraspecific variation at the *Acp70A* region have revealed any departure from neutral expectations. However, if the rather high frequency of the Ala allele is not a peculiarity of the population sampled, its overall lower level of nucleotide variation (Table 2 and Figure 3C) would be compatible with a recent origin of the Ala allele and its rapid increase in frequency due to positive selection. However, the number of fixed or nearly fixed silent differences between the Ser2 and Ala alleles in the present sample (six) seems high for a recent origin of the Ala allele. The Oregon R strain sequenced by STYGER (1992) has a serine at the last position of the signal peptide; however, when the amino acid replacement change is not considered, it has the typical Ala haplotype. Although this line could reflect a gene conversion event between Ala and Ser2 haplotypes, it could also be an old Ser2 derived haplotype segregating in natural populations. In this case, the mutation causing the Ser vs. Ala replacement could have occurred in such a divergent line. However, the presence of such a divergent line in a region with normal levels of recombination might require some additional explanation (as discussed below).

In the region 5' of the *Acp70A* transcription unit, there are two highly differentiated 5' haplotypes (Ser1 and Ser2 + Ala) represented by two and seven lines in our sample. In the fragment delimited by polymorphisms located at coordinates 243 and 537, these haplotypes differ by 16 fixed or nearly fixed nucleotide differences (depending on whether or not the three variants included in a putative gene conversion tract are considered, see Figure 2) and six length differences. However, both haplotypes are nearly equally diverged from the sequence of *D. sechellia*, indicating that neither originated from the other but are rather old lineages. Although most variants in this 5' flanking region are in linkage disequilibrium (Figure 4), they are in linkage equilibrium with variants at the *Acp70A* gene, including the amino acid replacement change Ser/Ala. This situation is reminiscent of that observed by ODGERS et al. (1995) at the Est-6 region of *D. melanogaster*. They also found two predominant 5' promoter haplotypes and a lack of correlation between them and the different Est-6 allozyme alleles. However, in the present case all lines with alanine at the last position of the signal peptide have the same 5' haplotype, pointing to a historical association due to its proposed recent origin.

The observed linkage disequilibrium between variants in this ~300-bp fragment at the *Acp70A* 5' region has probably originated by mutation as different mutations seem to have accumulated in the two old branches of the genealogy (Figure 5A). However, these two 5' haplotypes have persisted for a long time, as indicated by the number of fixed differences, without the expected decay of linkage disequilibrium in a region with normal levels of recombination. Recombination could be reduced if one of the haplotypes were associated with an inversion. In fact, the inversion *In(3L)P* extends between bands 63C and 72E1-2 and therefore includes the *Acp70A* region. Although we could not check the karyotypes of the nine sequenced lines from Montemayor, we studied an additional Spanish sample (Montblanc, northeast Spain) for chromosomal and nucleotide variation (SSCP analysis) at the *Acp70A* region (S. CIRERA and M. AGUADE, unpublished results). In a sample of 50 chromosomes we only detected one chromosome bearing the *In(3L)P* inversion but we detected the 5' Ser1 haplotype with a frequency of 15%.

The pattern of nucleotide variation in the 5' region of the *Acp70A* gene does not conform to the predictions of the mutation-drift equilibrium. Under that assumption, the existence of the two highly differentiated hap-
lypes in a region with normal levels of recombination would point to epistatic selection maintaining them and removing lower fitness recombinants. Although no function has been described so far for this region, it could be part of a putative regulatory region of the Acp70A gene, as mentioned above, or of some other transcription unit located further upstream. The molecular mechanism causing the proposed epistatic fitness interactions could be the maintenance of the secondary structure of putative DNA signals necessary for DNA-protein interactions in a similar way as maintenance of pre-mRNA secondary structure would explain the clustering of linkage disequilibria in the intron of Adh in D. pseudoobscura (Stephan and Kirby 1993; Kirby et al. 1995).

Alternatively, the population studied may not be in equilibrium. The Ser1 haplotype could have recently entered the population from another population. Also it could have recently escaped from an inverted In(3L)P chromosome, in a similar way as the Adh Maine haplotype has been proposed to have recently escaped inversion In(2L)I (Kreitman and Hudson 1991). Oggers et al. (1993) also consider this possibility for the 5' region of Est6. Both if the Ser1 haplotype had arrived from another differentiated population or had escaped from an inverted chromosome, this old haplotype would have been incorporated into the population relatively recently. The linkage disequilibria generated by the admixture would still persist, as recombination would not have had enough time to break down the initial associations. Although the patterning of nucleotide variation in the Acp70A region could be due to different selective forces or to the recent incorporation in the population of differentiated haplotypes, it is reminiscent of the expansion of a few haplotypes after a founder event. In fact, in the sample from Montemayor there are three major haplotypes (Ser1, Ser2 and Ala) in relatively high frequencies, the average number of differences between haplotypes being rather high compared to the average number of differences within haplotypes (Table 2).

Comparison of nucleotide variation in African and non-African populations of D. melanogaster (Begun and Aquadé 1993, 1995) indicate that only East African populations, where D. melanogaster is supposed to have originated, might be in equilibrium. Consequently, a survey of large samples of both non-African and African populations, as well as of a large sample of lines carrying the In(3L)P inversion, may shed some light on the forces maintaining variation at the Acp70A region. If epistatic selection was causing the observed pattern of linkage disequilibrium at the 5' region, one would expect the same pattern in different populations; also recombinants in this 5' region should be more scarce than in the region where the Acp70A gene is located. On the other hand, the presence in high frequency of the Ser1 and related haplotypes in In(3L)P carrying lines would suggest, as in the case of Adh (Kreitman and Hudson 1991; Aquadé 1988), that it had recently escaped from an inverted chromosome. The study of a larger sample and of samples from different geographical origins may also help to ascertain whether the Ala allele is of recent origin and has increased in frequency due to positive selection, as in this case the same pattern of low variation should be expected in all populations and they should all present the same most common haplotype.

We thank A. Moragas for fly collection, G. Ribo for cytological and fly work, J. Rozas for sharing isochromosomal lines and the unpublished version 2.08 of the DnaSP program, R. R. Hudson, D. Kirby and J. Braverman for computer programs, and E. Kubi for sharing unpublished results. We also thank J. Rozas and J. Braverman for critical comments and discussion. This work was supported by a predoctoral fellowship FPI91 from Comission Interdepartamental de Recerca i Tecnologia to S.C., and by Grants PB91-245 from Dirección General de Investigación Científica y Técnica (Spain) and 1993GR-1037 from Comission Interdepartamental de Recerca i Tecnologia to M.A.

LITERATURE CITED


Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, E. Wang et
