Positive Selection Drives the Evolution of the Acp29AB Accessory Gland Protein in Drosophila

Montserrat Aguadé

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

Manuscript received November 10, 1998
Accepted for publication February 22, 1999

ABSTRACT

Nucleotide sequence variation at the *Acp29AB* gene region has been surveyed in *Drosophila melanogaster* from Spain (12 lines), Ivory Coast (14 lines), and Malawi (13 lines) and in one line of *D. simulans*. The ~1.7-kb region studied encompasses the *Acp29AB* gene that codes for a male accessory gland protein and its flanking regions. Seventy-seven nucleotide and 8 length polymorphisms were detected. Nonsynonymous polymorphism was an order of magnitude lower than synonymous polymorphism, but still high relative to other non-sex-related genes. In *D. melanogaster* variation at this region revealed no major genetic differentiation between East and West African populations, while differentiation was highly significant between the European and the two African populations. Comparison of polymorphism and divergence at synonymous and nonsynonymous sites showed an excess of fixed nonsynonymous changes, which indicates that the evolution of the Acp29AB protein has been driven by directional selection at least after the split of the *D. melanogaster* and *D. simulans* lineages. The pattern of variation in extant populations of *D. melanogaster* favors a scenario where the fixation of advantageous replacement substitutions occurred in the early stages of speciation and balancing selection is maintaining variation in this species.

A ratio of nonsynonymous to synonymous divergence (K_a/K_s) significantly higher than one has been considered good evidence of positive selection driving protein evolution. The first genes for which such a pattern was found were those encoding proteins that mediate self *vs.* nonself recognition, such as some genes of the major histocompatibility complex (MHC) in mammals (Hughes and Nei 1988, 1990) and self-incompatibility alleles in Solanaceae (Clark and Kao 1991). Also some proteins involved in gamete recognition in marine invertebrates, such as sperm lysin in abalone (Lee and Vacquier 1995) and sperm bindin in sea urchins (Metz and Palumbi 1996), appeared to be driven by positive selection.

In Drosophila some proteins involved in sexual reproduction, such as those mediating gamete recognition or sperm competition, seem to evolve relatively rapidly. Civetta and Singh (1998) found that when closely related species of Drosophila (*D. melanogaster* and *D. simulans*) were compared, sex-related genes presented a high nonsynonymous to synonymous nucleotide substitution ratio relative to other genes. They argued that this high ratio was a result of directional selection and proposed that a burst of amino acid replacement substitutions accompanied the early phases of speciation. However, when the individual genes in that study are considered, in none of the comparisons between *D. melanogaster* and *D. simulans* is the K_a/K_s ratio significantly higher than one. For the *Acp26Aa* gene, which codes for a male accessory gland peptide that stimulates egg laying in mated *D. melanogaster* females during the first postmating day (Herndorn and Wolfner 1995), this ratio is highest and close to one (Aguadé *et al.* 1992). When the more distantly related species of the *melanogaster* group, *D. yakuba* and *D. teissieri*, were compared to *D. melanogaster* and *D. simulans*, this gene showed a K_a/K_s ratio significantly higher than one (Tsaur and Wu 1997).

Comparison of polymorphism and divergence at nonsynonymous and synonymous sites (McDonald and Kreitman 1991) has proved to be more powerful to detect adaptive protein evolution in closely related species than comparison of nonsynonymous (K_a) and synonymous (K_s) divergence. It has allowed detection of positive selection for the *Acp26Aa* gene in the *D. melanogaster-D. simulans* comparison (Aguadé 1998; Tsaur *et al.* 1998). However, this approach has not revealed a significant excess of nonsynonymous fixed changes for other genes that are also expressed in the male reproductive tract and whose products are transferred to the female as part of the seminal fluid: the *Acp26Ab* (Aguadé 1998; Tsaur *et al.* 1998), *Acp70A* (Cirera and Aguadé 1997), and *Est-6* (Karotam *et al.* 1993) genes.

In the present study nucleotide polymorphism and divergence have been analyzed at the *Acp29AB* gene (Wolfner *et al.* 1997), which is a male accessory gland protein gene encoding a 234-amino-acid-long protein. The mature peptide, composed of 213 amino acids, is

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

transferred to the female during mating (O. Lung, U. Tram and M. F. Wolfner, personal communication). Given that variation at this gene was associated with the ability of males to resist displacement by subsequent sperm, the protein could be involved in sperm competition (Clark *et al.* 1995). Unlike other *Acp* genes that showed a similar association with sperm competition in Clark *et al.* (1995), single strand conformation polymorphism (SSCP) analysis revealed only two variants at the *Acp29AB* region studied in the two samples from North America.

The ~1.7-kb region studied includes the previously sequenced *Acp29AB* gene (Wolfner *et al.* 1997) and a newly sequenced fragment of both its 5′ and 3′ flanking regions. Nucleotide variation at this region has been studied in samples from one European and two African populations of *D. melanogaster* and one line of the sibling species *D. simulans*. Different subsamples of the same *D. melanogaster* populations had been surveyed for variation at the *Acp26Aa* and *Acp26Ab* gene regions (Aguadé 1998), and they showed high levels of both synonymous and nonsynonymous variation in those genes. To test whether, as in the case of the *Acp26Aa* gene, the *Acp29AB* gene had been subject to strong positive selection, we compared nonsynonymous and synonymous variation in the coding region within and between species and also silent/synonymous variation across the region studied.

MATERIALS AND METHODS

Drosophila stocks: Twelve isofemale lines collected in Montblanc (Spain) in 1993 were isogenized for the second chromosome upon arrival in the laboratory by the corresponding series of crosses with a balancer stock. A subsample of 14 lines from Lamto (Ivory Coast) and 13 lines from Malawi, kindly provided by M. Veuille and V. Bénassi (see Bénassi *et al.* 1993; Bénassi and Veuille 1995), were used in the present study. African lines were used as previously described (Aguadé 1998) except that the deficiency *Df(2L)TE29Aa-11*, which covers cytological positions 28E4 to 29C1, was used to obtain individuals hemizygous for the wild *Acp29AB* locus. The *D. simulans* line used for the interspecific analysis was collected in Montblanc (Spain) in 1993 and was sibmated for 10 generations upon arrival in the laboratory. Individuals of the *D. melanogaster* and *D. simulans* lines from Montblanc were frozen one or two generations after isogenization.

DNA extraction, PCR amplification, and sequencing: DNA from the mst319.5 plasmid that includes the *Acp29AB* gene (Wolfner *et al.* 1997) was kindly provided by M. F. Wolfner. Genomic DNA from *D. melanogaster* and *D. simulans* was extracted from 1 or 10 adult flies by a modification of protocol 48 in Ashburner (1989).

Primers T3 and T7 from the pBluescript (Stratagene, La Jolla, CA) vector polylinker were used to amplify by PCR the insert of plasmid mst319.5 (Wolfner *et al.* 1997). The flanking regions (~650 and 550 bp in each direction) were sequenced by primer walking, which started in the coding region. The PCR product was purified with a Qiaquick column (QIAGEN, Chatsworth, CA) and cycle sequenced using fluorescent dideoxy terminators according to the manufacturer's instructions (Perkin-Elmer, Norwalk, CT; Amersham, Arlington Heights, IL). After excess dye-terminators were removed by ethanol precipitation, the sequencing product was separated with an ABI 377 automated DNA Sequencer (Perkin-Elmer).

For each of the *D. melanogaster* lines, 20-nucleotide-long primers (5′AAAGAAGATGCCCTGGGATA3′ and 5′GATGG CCGAGAGCAGAAGTT3′) were used to amplify by PCR an ~1.8-kb region encompassing the *Acp29AB* gene and its 5′ and 3′ flanking regions. Primers (17 nucleotides long) spaced on average 350 nucleotides were used to sequence the purified PCR products as described above. The *D. melanogaster* primers were used to amplify by PCR the homologous region of *D. simulans*; in this case, however, the ~1.8-kb region was amplified in two overlapping fragments. The sequences newly reported in this article have been deposited in the EMBL sequence database library under accession nos. AJ240513–AJ240552.

Sequence analysis: The SeqEd program (Perkin-Elmer) was used to assemble and align the sequences and also to check all variable sites. The MacClade version 3.0.6 program was used to edit the sequences for further analyses (Maddison and Maddison 1992). The DnaSP version 2.92 program (Rozas and Rozas 1997) was used for most intraspecific and some interspecific analyses.

RESULTS

Sequencing of the 5′ and 3′ flanking regions of the *Acp29AB* gene: Sequencing of the 5′ flanking region revealed that the *Acp29AB* gene lacks a consensus TATA box. Also, the promoter region of this gene did not present either the decamer (AATGCAAAAT) or the octamer (ATTGCAAT) motifs described for other genes expressed in male accessory glands (DiBenedetto *et al.* 1990; Simmerl *et al.* 1995). The 5′ flanking region had a high AT content (68%) and TTT was the most common trinucleotide (9.2%). No open reading frame longer than 100 residues was detected in this 5′ region, while one at least 131 residues long was detected at the 3′ flanking region. This open reading frame presumably extends beyond the region sequenced in this study, but no consensus promoter signals were associated with this putative reading frame. Both the 5′ and 3′ flanking regions are considered *a priori* as noncoding in the following analyses.

Nucleotide polymorphism: In the Ivory Coast population, unlike in the other two studied populations, the *In(2L)t* inversion is present at a very high frequency (73% as reported in Veuille *et al.* 1998). Although in the present study 11 of the 14 lines analyzed carried this inversion (M. Veuille, personal communication), lines were not separated according to gene arrangement. The reason is that the *Acp29AB* region is located rather centrally in the region affected by the inversion, and no genetic differentiation between inverted and noninverted chromosomes was found when different subsamples of this same population were surveyed for variation at other regions also located in central positions of the inversion (*P6* or *Fbp2*, Bénassi *et al.* 1993; *Acp26A*, Aguadé 1998).

Figure 1 summarizes the distribution of nucleotide

	-530	-482	-441	-436	-415	-414	-413	-318	-302	-295	-294	-281	-273	-271	-252	-215	-197	-176	-175	-169	-139	-136	-132	-123	-117	-115	-113	-104	-85	-71	-70	-69	[-69]	-40	-38	-26	-12	-2	16	94	106	115	145	187	196	204	341	366	367	388	486	502	550	562	568	574	580	607	625	670	753	755	762	794	819	822	826	852	960	973	989	990	1066	1078	1093	1123	1126	1129	1169
																																										n				n	n	n			n									n																			
Mo80b	C	T	A	A	A	A	C	C	G	T	T	G	G	T	G	C	T	A	C	G	C	A	G	T	T	A	d	C	G	G	T	T	d	T	G	G	C	C	T	A	G	G	A	T	C	G	G	T	G	G	A	C	G	C	A	C	C	T	G	G	T	G	G	T	T	T	G	T	T	T	C	C	C	C	G	A	T	T	C
Mo1b	G	.	C	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	G	C	T	.	.	G	A	A	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T
Mo2b	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo8b	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	G	A	A	T	.	.	.	C	T	A	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo15b	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	A	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo34a	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo36a	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo37a	.	.	.	G	C	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	G	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo40b	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	A	A	T	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo47a	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Mo52b	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	C	T	A	C	.	.	.	.	.	.	C	C	.	.	.	.	.	.	T	.	.	G	.	.	T
Mo58b	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
La1	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	G	A	G	C	.	.	C	T	.	.	G	A	A	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	.	G	C	C	.
La3	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	C	.	.	.	T	.	.	G	A	.	.	.	T	A	C	T	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.
La4	.	.	.	.	C	G	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	G	G	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	A	.	.	.	T	A	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T
La5	.	.	.	.	C	G	G	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	T	.	.	.	.	d	.	.	.	T	.	.	G	.	.	.	.	.	.	.	G	A	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.
La6	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	T	.	.	.	d	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	T	.	.	G	C	C	.
La9	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	T	.	.	.	d	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	T	.	.	G	C	C	.
La13	G	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	T	.	A	T	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	A	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T	.	.	.	.	C	.
La14	.	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	C	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	G	.	.	.
La15	.	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	C	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	G	.	.	.
La16	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	.	.	G	G	G	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	.	T	.	.	A	T	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
La21	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	A	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	A	.	.	.	T	A	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.	G	.	.	T
La25	.	.	.	.	C	G	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	G	G	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	A	.	.	.	T	A	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T
La34	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	G	A	G	C	.	.	C	T	.	.	G	A	A	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	.	G	C	C	.
La35	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	.	.	.	T	.	.	G	.	.	.	.	.	.	T	.	.	A	.	.	.	.	C	.	.	C	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Ma5	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	C	.	.	.	.	A	.	G	G	G	.	.	.	.	.	d	G	.	.	.	.	.	G	.	.	G	.	T	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	G	.	.	T
Ma7	.	.	.	.	C	G	G	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	G	.	.	.	.	.	G	.	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	C	.
Ma8	.	.	.	.	C	G	G	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	.	.	G	A	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	C	.
Ma13	.	.	.	G	C	G	G	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	T	A	C	T	A	C	.	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	.	G	.	.	.
Ma18	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	G	C	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Ma20	.	.	.	G	C	G	G	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	T	T	A	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	C	.
Ma21	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	G	G	G	.	.	.	.	.	d	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	C	T	A	C	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
Ma43	G	.	.	.	.	.	.	.	.	.	.	A	T	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	T	A	A	A	.	.	.	.	G	A	G	.	.	.	.	.	A	.	G	A	A	.	.	.	.	.	.	.	C	A	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.
Ma45	.	.	.	.	C	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	.	C	T	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ma48	.	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	G	C	T	.	.	.	.	.	.	.	.	.	C	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	C	.
Ma50	.	.	.	.	C	G	G	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	G	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ma52	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	G	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	C	T	A	C	.	.	.	.	.	.	C	C	.	.	.	.	.	.	T	.	.	G	.	.	T
Ma67	G	.	.	.	.	.	.	.	C	C	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	d	.	.	.	.	.	d	.	.	.	.	.	.	G	.	.	G	.	T	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
		*	*										*	*	*				*		*		*	*				*		*	*	*				*						*																	*	*		*	*						*		*	*							

Figure 1.—Nucleotide polymorphisms at the *Acp29AB* gene region in three natural populations of *D. melanogaster.* Polymorphisms were numbered according to the sequence in Wolfner *et al.* (1997). Lines from Montblanc (Mo), Lamto (La), and Malawi (Ma) are presented sequentially. The thick-bordered box above the sequence information represents the amino acid coding region and the thick line the leader and trailer regions. Dots represent the same nucleotide as in the first sequence and asterisks indicate singletons. n, nonsynonymous; d, deletion. Complex mutational events that include both nucleotide changes and indels are shaded.

TABLE 1

Nucleotide polymorphism

		Coding			
	5′	Syn	Nsyn	3′	Total silent
No. of sites	533	145	557	429	1107
Total					
S	37	15	6	19	64
π	0.0135	0.0319	0.00292	0.0085	0.0130
θ	0.0147	0.0245	0.00255	0.0099	0.0137
K	0.1084	0.2386	0.07815	0.0769	0.1132
Montblanc					
S	15	10	3	6	36
π	0.0093	0.0199	0.0023	0.0038	0.0093
θ	0.0093	0.0229	0.0018	0.0046	0.0100
K	0.1099	0.2376	0.0791	0.0787	0.1145
Lamto					
S	25	12	4	14	51
π	0.0115	0.0347	0.0027	0.0100	0.0138
θ	0.0132	0.0261	0.0023	0.0096	0.0134
K	0.1194	0.2853	0.0773	0.0837	0.1246
Malawi					
S	25	13	5	13	51
π	0.0134	0.0358	0.0027	0.0079	0.0141
θ	0.0135	0.0290	0.0029	0.0092	0.0138
K	0.1146	0.2881	0.0784	0.0782	0.1204

Syn, synonymous; nsyn, nonsynonymous; silent, silent in noncoding regions and synonymous in coding regions; *S*, π, and θ refer to the number of segregating sites, nucleotide diversity (Nei 1987), and the Watterson parameter or expected nucleotide heterozygosity (Watterson 1975), respectively; *K* is the number of nucleotide substitutions per site, or nucleotide divergence between *D. melanogaster* and *D. simulans*, corrected according to Jukes and Cantor (1969).

sequence variation in the 1664-nucleotide region studied (excluding alignment gaps), including only 244 nucleotides of the detected 3′ open reading frame. A total of 77 nucleotide and 8 length polymorphisms were detected. All length polymorphisms (not shown) were in noncoding regions. Four of these polymorphisms were single-base indels associated with runs of at least eight T's or eight A's and could be an artefact of the Taq polymerase; however, no such variation was detected in two similar runs of A's present in the coding region. Two other single-base indels can be considered part of complex mutations: TAAT to GAGTG and GTT to TAAA, starting at positions −117 and −71, respectively. The remaining two indels were 11 (see discussion) and 3 bp long.

Table 1 gives the estimates of nucleotide variation for the whole region studied and for its different functional parts. The level of synonymous variation was generally higher than that of silent variation in the flanking regions. Nonsynonymous variation was an order of magnitude lower than synonymous variation, but still rather high as compared to that in other genes whose expression is not sex-related (Civetta and Singh 1998). In the sample from Montblanc, estimates of nucleotide variation were slightly lower than in the two African samples. Neither the Tajima test (Tajima 1989) nor the Fu and Li tests (Fu and Li 1993) detected any departure from neutral expectations in any of the populations studied when the complete region was considered (results not shown).

Population differentiation as revealed by variation at the *Acp29AB* region was analyzed, considering the complete region and its different functional parts (Table 2). The comparison between the two African populations showed the lowest F_{st} estimates, and although there was generally marginal statistical significance, in one case there was greater statistical support for differentiation ($P = 0.014$). However, the sample from Montblanc showed in all cases a significant differentiation from the African samples.

The *Acp29AB* gene is located in a region of high recombination (Kliman and Hey 1993). In the ~1.7-kb region studied, the four-gamete test (Hudson and Kaplan 1985) detected 7, 11, and 11 recombination events in the history of the samples from Montblanc, Lamto, and Malawi, respectively. Linkage disequilibrium between informative sites (for which the rarest variant is present more than once in the sample) was estimated separately for each population, as significant pairwise association between polymorphic sites can result from admixture of differentiated populations. The percentages of comparisons that showed a significant

TABLE 2

Genetic differentiation

	Mo-La	Mo-Ma	La-Ma
Total			
F_{st}	0.221	0.127	0.062
Prob	0.000	0.003	0.055
5′ region			
F_{st}	0.276	0.114	0.093
Prob	0.000	0.016	0.014
Coding			
F_{st}	0.146	0.062	0.066
Prob	0.001	0.011	0.049
3′ region			
F_{st}	0.212	0.255	0.033
Prob	0.003	0.002	0.049

Mo, Montblanc; La, Lamto; Ma, Malawi. F_{st}, which measures the proportion of nucleotide diversity attributable to variation between populations, was estimated from the average number of differences between alleles according to Hudson *et al.* (1992a). Prob, probability assessed by the Hudson *et al.*, (1992b) permutation procedure, using K_s^* as the test statistic.

association using the χ^2 test were 28, 16, and 8% for Montblanc, Lamto, and Malawi, respectively. Only in the sample from Montblanc was there some clustering of linkage disequilibria.

Silent nucleotide polymorphism and divergence: Table 1 gives a summary of silent and/or synonymous nucleotide divergence. As in the case of polymorphism, the highest estimate corresponds to synonymous divergence and the lowest to the 3′ flanking region. As shown in Figure 2, levels of silent polymorphism and divergence seem to vary concordantly along the region studied. The Hudson, Kreitman, and Aguadé (HKA) test (Hudson *et al.* 1987), which compares polymorphism and divergence in two regions, was applied to silent variation. The region studied was divided into 5′ flanking and rest, 5′ flanking and coding, and coding and 3′ flanking; in none of the population samples did the different HKA tests reveal any departure from the neutral expectation of a direct relationship between levels of polymorphism and divergence (results not shown). Unlike the HKA test, the different tests of heterogeneity in the ratio of polymorphism to divergence across a given DNA region developed by McDonald (1996, 1998) do not require any *a priori* partition of the region studied. When these tests were applied to the *Acp29AB* region, no heterogeneity was detected (results not shown).

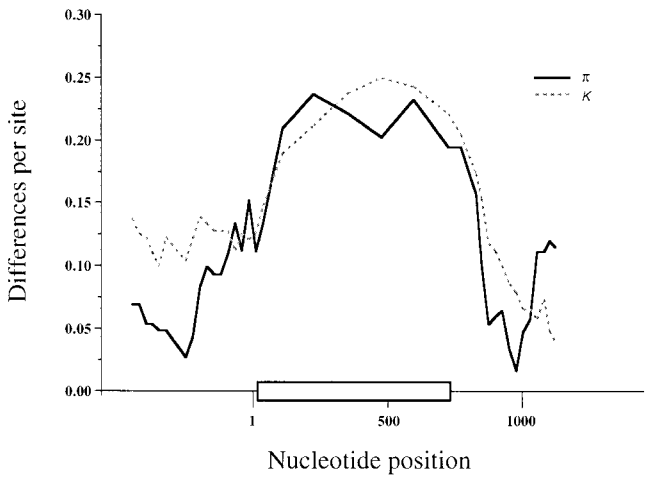

Figure 2.—Sliding window plot of silent polymorphism (π) in the sample from Lamto and divergence (K) between this population and *D. simulans*, plotted using a window of 200 silent nucleotides. Silent polymorphism was multiplied by the time of divergence as estimated by the HKA test. The box on the *x*-axis represents the coding region. Nucleotide positions were numbered according to the sequence in Wolfner *et al.* (1997).

Amino acid replacement polymorphism and divergence: Six replacement polymorphisms were detected in the Acp26AB protein (Figure 3). Three polymorphisms (at residues 59, 113, and 153) segregated in the three populations. Although they were all present at intermediate frequencies in the combined sample, only the two variants at residue 153 (Met/Lys) were present at similar frequencies in the three population samples: 0.33, 0.29, and 0.38 for the less frequent variant in Montblanc, Lamto, and Malawi, respectively. The two putative glycosylation sites (Wolfner *et al.* 1997) were monomorphic in *D. melanogaster* and conserved between species. In contrast, of the eight putative peptidase cleavage sites described in *D. melanogaster* (Wolfner *et al.* 1997), two were polymorphic in this species (at residues 29 and 113), and no putative cleavage sites were present in *D. simulans* at residues 29, 97, and 112.

According to neutral predictions the ratio of nonsynonymous to synonymous changes should be the same

29	59	105	113	153	214	Mo	La	Ma	mel
Lys	Arg	Ala	Arg	Met	Glu	1	0	0	1
.	Lys	.	Leu	Lys	.	7	0	1	8
.	Lys	.	.	.	.	2	1	0	3
.	.	.	Leu	.	.	1	2	5	8
.	.	.	Leu	Lys	.	1	0	4	5
.	.	.	.	Lys	.	0	7	1	8
.	Lys	.	.	Lys	.	0	2	1	3
.	.	Ser	Leu	.	.	0	1	0	1
.	.	Ser	Leu	Lys	.	0	1	0	1
Asn	Lys	.	Leu	Lys	Asp	0	0	1	1
Asn	.	.	.	Lys	.				

Figure 3.—Amino acid replacement haplotypes and their absolute frequencies in three natural populations of *D. melanogaster*. Numbers in the first row indicate the amino acid residue in the *D. melanogaster* protein. Dots indicate same amino acid as in the first sequence. The bottom line shows the amino acids present in *D. simulans* in those sites. Mo, Montblanc; La, Lamto; Ma, Malawi; mel, *D. melanogaster*.

within and between species. The McDonald and Kreitman (MK) test (McDonald and Kreitman 1991) contrasts this prediction by comparing the number of nonsynonymous and synonymous changes that are polymorphic within species and fixed between species. This test was applied separately to each population and to the combined data set (Table 3). In all cases there was an excess of fixed nonsynonymous changes, which was significant (for Montblanc and Lamto) or marginally significant (for Malawi and the complete data set).

DISCUSSION

Nucleotide polymorphism and divergence: As in previous studies of nucleotide variation that included African samples of *D. melanogaster* (Begun and Aquadro 1993, 1995; Aguadé 1998), variation at the *Acp29AB* region was higher in the African populations than in the European population, and this latter population showed a higher level of linkage disequilibrium. Also, genetic differentiation at this region was highly significant between the European population and any of the two African populations. In a similar way to the *Acp26A* region (Aguadé 1998) but unlike that previously reported for the *Adh* region (Bénassi and Veuille 1995; see, however, Veuille *et al.* 1998), the African populations were either marginally or slightly differentiated.

The levels of silent polymorphism and divergence vary along the region studied (Figure 2). As reported for other genes, variation at synonymous sites of the *Acp29AB* gene was higher than in its flanking regions, which indicates stronger selective constraints in those regions. The 3′ flanking region showed the lowest estimates even when only the 185 nucleotides between the *Acp29AB* coding region and the detected open reading frame were considered: $\pi = 0.0072$ and $K = 0.103$. If this open reading frame was a functional gene, the intergenic region would not only contain the *Acp29AB* trailer but also some regulatory sequences of this new gene, which would also contribute to constraining variation at this region.

TABLE 3

McDonald and Kreitman tests

	F_s	F_{ns}	P_s	P_{ns}	Prob
Montblanc	33	37	10	3	0.047*
Lamto	33	37	12	4	0.043*
Malawi	33	36	13	5	0.064
mel	33	36	15	6	0.057

F_s, number of synonymous fixed differences between species; F_{ns}, number of nonsynonymous fixed differences between species; P_s, number of synonymous segregating sites; P_{ns}, number of nonsynonymous segregating sites. mel, all *D. melanogaster* lines. Probability established with a *G*-test with Williams' correction for continuity. * $0.01 < P < 0.05$.

Nucleotide sequence variation and function: Although function can only be established experimentally, studies of nucleotide sequence variation have been often used to implicate selection and functional constraint. Generally, those studies have involved comparison of distantly related species, and the presence of highly conserved sequences has been considered an indication of functional constraint and, therefore, of purifying selection acting on those regions. Here, we have compared sequences from the same species and from a very closely related species. Consequently, our approach was different, and we looked specifically for variation at putative functional sequences. Figure 4 shows three different sequences of the *Acp29AB* trailer region found in our survey. *D. melanogaster* populations segregate for an 11-bp deletion that includes the previously described polyadenylation signal (Wolfner *et al.* 1997). However, a new poly(A) addition signal results from this deletion (Figure 4), which is displaced three nucleotides from the polyadenylation signal described by Wolfner *et al.* (1997). Although *D. simulans* also presents an 8-bp deletion in that region relative to *D. melanogaster*, this deletion does not affect the polyadenylation signal.

Some amino acid replacement changes affect putative peptidase cleavage sites both within and between species. Corresponding to the sites segregating in *D. melanogaster* (at residues 29 and 113), *D. simulans* has the less frequent variant in *D. melanogaster*, which at residue 29 is characterized by the loss of a peptidase cleavage sequence. In the case of residue 29, the loss-of-site variant is present only once in the sample from Malawi. In contrast, at residue 113 it segregates in all three populations and in the complete sample it is present at high frequency (61.5%). As *D. simulans* has lost the putative cleavage site at residue 112, it is tempting to speculate that in this species the cleavage site at residue 113 might be functional. On the other hand, in *D. melanogaster* the gain of function at residue 112 might compensate for

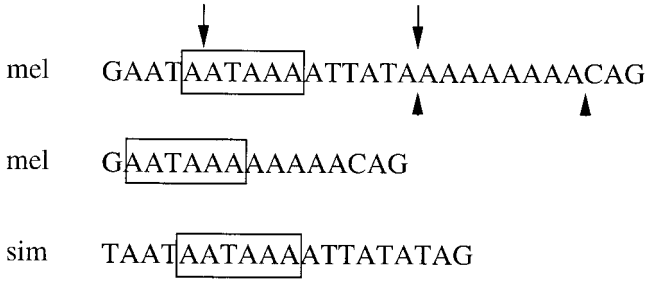

Figure 4.—Nucleotide and length changes affecting the polyadenylation signal in *D. melanogaster* (first two rows) and *D. simulans* (last row). Boxed nucleotides indicate the polyadenylation signal. Arrows on top of the first sequence indicate the extent of an 11-bp deletion segregating in *D. melanogaster* that causes a 3-bp displacement of the polyadenylation signal. Arrowheads below the first sequence indicate a deletion that differentiates *D. simulans* from *D. melanogaster*.

the loss-of-function variant segregating at residue 113. Also, two of the changes fixed between species might be compensatory. In fact, in both *D. melanogaster* and *D. simulans* there is one additional but different putative cleavage site close to the conserved sites at residues 202 and 203: at residue 204 in *D. melanogaster* and at residue 201 in *D. simulans*. If these additional sites were functional and sites 202 and 203 were not, the loss in *D. simulans* of site 204 might compensate the gain of site 201.

Nonsynonymous variation and selection: Unlike the case of the *Acp26Aa* gene where the K_a/K_s ratio was close to one in the *D. melanogaster vs. D. simulans* comparison, for the *Acp29AB* gene this ratio is lower than one. This would indicate that *Acp29AB*, like *Acp26Ab* and *Acp70A*, has more constraints on its variation than *Acp26Aa*. However, in the case of the *Acp29AB* gene the K_a/K_s ratio was threefold higher than the π_a/π_s ratio for the complete data set 0.327 and 0.09, respectively. As previously mentioned (see results), if variation in this region were neutral, one would expect this ratio to be the same. Meanwhile, the fixation of advantageous amino acid replacement changes would cause a deviation of this ratio from constancy due to a higher than expected nonsynonymous divergence. Similarly to synonymous polymorphism, the estimated synonymous divergence (K_s) between *D. melanogaster* and *D. simulans* lies on the upper part of the range of estimates for other genes, which indicates that the deviation from a constant ratio is due to a higher than expected nonsynonymous divergence (K_a). In fact, application of the MK test to the different populations surveyed and to the complete *D. melanogaster* sample revealed a significant or marginally significant excess of nonsynonymous fixed differences between *D. melanogaster* and *D. simulans* (Table 3). This indicates that at least after the split of the *D. melanogaster* and *D. simulans* lineages the evolution of the Acp29AB protein was driven by directional selection.

The *Acp29AB* gene is not, however, depauperate of genetic variation as would be expected *a priori* for a region having suffered the fixation of advantageous amino acid replacements. Different scenarios would be compatible with this observation. If as proposed by Civetta and Singh (1998) the burst of amino acid replacements in sex-related genes occurred very close to speciation (2.5 mya), the time back to those selective sweeps would be longer than the expected time of coalescence of extant variation in *D. melanogaster* (Eanes *et al.* 1996; Aguadé 1998). On the other hand, if as suggested by Tsaur *et al.* (1998) the fixation of amino acid replacement mutations were due to a weak advantage of such mutations, the extent of the region affected by the corresponding selective sweep might be rather short and its footprint on nucleotide diversity difficult to detect. Lack of the sequences of *D. mauritiana* and *D. sechellia* for the *Acp29AB* region has precluded unambiguous establishment of the ancestral state of the detected polymorphisms and, therefore, has prevented the test proposed by these authors on the frequency spectrum of new mutations. However, a preliminary analysis based on only the *D. melanogaster-D. simulans* comparison did not reveal any excess of sites with the new nucleotide in high frequency (analysis not shown).

There is one last scenario to be discussed, which complements the scenario proposed by Civetta and Singh (1998): the possibility of directional selection having driven the amino acid replacement fixations between species in the early stages of speciation and some sort of balancing selection maintaining variation in *D. melanogaster* populations. This would be a somewhat similar scenario to that for the *Adh* locus, where positive selection has driven the evolution of the Adh protein in the *melanogaster* group (McDonald and Kreitman 1991), and balancing selection is maintaining the Adh^F/Adh^S allozyme polymorphism in *D. melanogaster* populations (Kreitman and Aguadé 1986; Kreitman and Hudson 1991; Berry and Kreitman 1993). In the case of the *Acp29AB* gene, there is no clear evidence of its polymorphism being adaptive. There is, however, some indication in that sense from the study by Clark *et al.* (1995) that showed a significant difference between males homozygous for the two SSCP morphs detected for this gene in their ability to defend against sperm displacement.

The possible involvement of Acp29AB in sperm displacement will have to be assessed by functional studies. If variation at this protein affected this or some other component of fitness, any of the three amino acid replacement polymorphisms that segregate in the three populations surveyed might be candidates for being the targets of selection. It is easy to speculate how changes in those residues might affect the function of the Acp29AB protein and, therefore, contribute to the detected differences of fitness. For example, residues 131–214 in the C-terminal part of the Acp29AB protein show similarity at the primary sequence and predicted tertiary structure levels with carbohydrate recognition domains (CRDs; W. Swanson, personal communication). Two of the amino acids reported here as variable, residues 153 and 214, would map within this CRD. Perhaps changes at these positions are functionally important, modifying the CRD's carbohydrate recognition specificity. Interestingly, variants at residue 153 segregate at rather similar frequencies in the three populations, which might also suggest that this residue is a candidate for contributing to putative fitness differences.

If selection were acting to maintain an amino acid polymorphism, one would expect that variants surrounding that site were at linkage disequilibrium. Also, if selection were maintaining the same alleles in the whole distribution area of the species, the same clustering of linkage disequilibria would be expected in the different populations surveyed. However, the extent of the region exhibiting linkage disequilibrium might be

rather short in a region of high recombination like the *Acp29AB* region (see above). This, together with the distribution of polymorphisms, might have hindered the detection of linkage disequilibrium in all three populations. Balancing selection leaves a characteristic footprint at linked neutral sites, causing a deviation in the frequency spectrum toward an excess of polymorphic sites with variants at intermediate frequencies. Also, an excess of synonymous polymorphism is expected to result from an old balanced polymorphism. Neither the HKA test nor the McDonald heterogeneity tests detected any excess of synonymous polymorphism in the *Acp29AB* coding region. The Tajima test and the Fu and Li tests, which contrast the frequency and distribution of variants, respectively, did not detect any deviation from neutral expectations when the whole region studied was considered. However, the test statistics presented negative nonsignificant values for the two flanking regions, while values for the coding region (both in Lamto and Malawi) were positive and in some cases significantly different from zero, an indication of variants maintained at intermediate frequencies. The scenario that combines positive selection at the early stages of speciation and balancing selection (either overdominant selection or frequency-dependent selection) would better explain most of the characteristics revealed by the present data on nucleotide variation at the *Acp29AB* region in conjunction with those by Clark *et al.* (1995).

All *Acp* genes surveyed to date for nucleotide variation in populations of *D. melanogaster* (*Acp26Aa* and *Acp26Ab*, Aguadé *et al.* 1992; Aguadé 1998; Tsaur *et al.* 1998; *Acp29AB*, present work; and *Acp70A*, Cirera and Aguadé 1997) are located in this species in regions of high recombination. The level of silent/synonymous variation in those gene regions is rather high, but comparable to that of other non-sex-related genes also located in regions of high recombination. On the other hand, their level of nonsynonymous variation varies among genes, although in general it is higher than for other non-sex-related genes (Civetta and Singh 1998). For only the two longer genes (*Acp26Aa* and *Acp29AB*) has comparison of intra- and interspecific variation revealed that positive selection has driven the evolution of the corresponding proteins, *i.e.*, that at least some of the amino acid replacements fixed after the split of the *D. melanogaster* and *D. simulans* lineages were adaptive and, therefore, fixed by directional positive selection. On the other hand, the level of nonsynonymous polymorphism is higher for the *Acp26Aa* than for the *Acp29AB* gene. We propose that the different pattern of amino acid replacement variation present in those genes in extant populations might be due to both different selective constraints on amino acid changes and balancing selection maintaining variation only in *Acp29AB*.

I thank Michel Veuille and Véronique Bénassi for the African lines and for sharing unpublished information on the chromosomal polymorphism of these lines, Julio Rozas for sharing version 2.92 of the DnaSP program, W. Swanson for comments on the manuscript and for sharing unpublished results, the Bowling Green Stock Center for the *Df(2L)TE29Aa-11/CyO* line, and Serveis Científico-Tècnics from Universitat de Barcelona for automated sequencing facilities. Special thanks are given to M. F. Wolfner for the mst319.5 plasmid and for sharing the *Acp29AB* gene sequence before publication, but mainly for her insightful comments on the manuscript; and to D. Salguero for his excellent technical assistance. This work was supported by grants PB94-0923 from Dirección General de Investigación Científica y Técnica, Spain, and 1997SGR-59 from Comissió Interdepartamental de Recerca i Tecnologia, Generalitat de Catalunya.

LITERATURE CITED

Aguadé, M., 1998 Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* genes in the melanogaster species complex. Genetics **150:** 1079–1089.

Aguadé, M., N. Miyashita and C. H. Langley, 1992 Polymorphism and divergence in the *Mst26A* male accessory gland gene region in Drosophila. Genetics **132:** 755–770.

Ashburner, M., 1989 *Drosophila, A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **356:** 519–520.

Begun, D. J., and C. F. Aquadro, 1995 Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

Bénassi, V., and M. Veuille, 1995 Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, West Africa and East Africa. Genet. Res. **65:** 95–103.

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

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics **134:** 869–893.

Cirera, S., and M. Aguadé, 1997 Evolutionary history of the sex-peptide (*Acp70A*) gene region in *Drosophila melanogaster*. Genetics **147:** 189–197.

Civetta, A., and R. S. Singh, 1998 Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. **15:** 901–909.

Clark, A. G., and T.-U. Kao, 1991 Excess nonsynonymous substitutions at shared sites among self-incompatibility alleles of Solanaceae. Proc. Natl. Acad. Sci. USA **88:** 9823–9827.

Clark, A. G., M. Aguadé, T. Prout, L. Harshman and C. H. Langley, 1995 Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics **139:** 189–201.

DiBenedetto, A. J., H. A. Harada and M. F. Wolfner, 1990 Structure, cell specific expression, and mating-induced regulation of a *Drosophila melanogaster* male accessory gland gene. Dev. Biol. **139:** 134–148.

Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, I-N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the G6pd locus in *Drosophila melanogaster* and *D. simulans*. Genetics **144:** 1027–1041.

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

Herndorn, L. A., and M. F. Wolfner, 1995 A Drosophila seminal fluid protein, Acp26Aa, stimulates egg-laying in females for one day after mating. Proc. Natl. Acad. Sci. USA **92:** 10114–10118.

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

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992a Estimation

of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

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

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

Hughes, A. L., and M. Nei, 1990 Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of Class I major-histocompatibility-complex molecules. Mol. Biol. Evol. **7:** 515–524.

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

Karotam, J., A. C. Delves and J. G. Oakeshott, 1993 Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling species of Drosophila. Genetica **88:** 11–28.

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 Excess polymorphism at the *Adh* locus in *Drosophila melanogaster*. Genetics **114:** 93–110.

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.

Lee, Y., and V. D. Vacquier, 1995 Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol. Biol. Evol. **12:** 231–238.

Maddison, W. P., and D. R. Maddison, 1992 *MacClade: Analysis of Phylogeny and Character Evolution, version 3.0*. Sinauer Associates, Sunderland, MA.

McDonald, J. H., 1996 Detecting nonneutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **13:** 253–260.

McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **15:** 377–384.

McDonald, J. H., and M. Kreitman, 1991 Adaptative protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Metz, E. C., and S. R. Palumbi, 1996 Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol. Biol. Evol. **13:** 397–406.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

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.

Simmerl, E., M. Schafer and U. Schafer, 1995 Structure and regulation of a gene cluster for male accessory gland transcripts in *Drosophila melanogaster*. Insect. Biochem. Mol. Biol. **25:** 127–137.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tsaur, S.-H., and C.-I. Wu, 1997 Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa*, of Drosophila. Mol. Biol. Evol. **14:** 544–549.

Tsaur, S.-H., C.-T. Ting and C.-I. Wu, 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of Drosophila. II. Divergence versus polymorphism. Mol. Biol. Evol. **15:** 1040–1046.

Veuille, M., V. Bénassi, S. Aulard and F. Depaulis, 1998 Allele-specific population structure of *Drosophila melanogaster* at the molecular level. Genetics **149:** 971–981.

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

Wolfner, M. F., H. A. Harada, M. J. Bertram, T. J. Stelick, K. W. Kraus *et al.*, 1997 New genes for male accessory gland proteins in *Drosophila melanogaster*. Insect. Biochem. Mol. Biol. **27:** 825–834.

Communicating editor: A. G. Clark