Molecular Population Genetics of Male Accessory Gland Proteins in Drosophila

David J. Begun,* Penn Whitley,* Bridget L. Todd,† Heidi M. Waldrip-Dail† and Andrew G. Clark†

**Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712 and †Department of Biology and Institute for Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802*

Manuscript received January 3, 2000
Accepted for publication August 31, 2000

ABSTRACT

Drosophila seminal proteins have an unusually high rate of molecular sequence evolution, suggesting either a high rate of neutral substitution or rapid adaptive evolution. To further quantify patterns of polymorphism and divergence in genes encoding seminal proteins, also called accessory gland proteins (*Acp*'s), we conducted a sequencing survey of 10 *Acp* genes in samples of *Drosophila melanogaster* and *D. simulans* (*Acp29AB*, *Acp32CD*, *Acp33A*, *Acp36DE*, *Acp53Ea*, *Acp62F*, *Acp63F*, *Acp76A*, *Acp95EF*, and *Acp98AB*). Mean heterozygosity at replacement sites in *D. simulans* was 0.0074 for *Acp* genes and 0.0013 for a set of 19 non-*Acp* genes, and mean *melanogaster-simulans* divergence at replacement sites was 0.0497 for *Acp* genes and 0.0107 at non-*Acp* genes. The elevated divergence of *Acp* genes is thus accompanied by elevated within-species polymorphism. In addition to the already-reported departures of *Acp26A*, *Acp29AB*, and *Acp70A* from neutrality, our data reject neutrality at *Acp29AB* and *Acp36DE* in the direction of excess replacements in interspecific comparisons.

MOLECULAR population genetic analysis of particular classes of proteins or genetic pathways may eventually allow us to make general inferences about the connection between the functional and evolutionary properties of genes. Sexual phenotypes involved in male-male and male-female interactions in Drosophila have attracted much recent attention (Chapman *et al.* 1995; Rice 1996; Clark *et al.* 1999; Holland and Rice 1999). A large proportion of wild-caught *Drosophila melanogaster* females carry sperm from two or more males in their reproductive tract (Cobbs 1977; Griffiths *et al.* 1982; Harshman and Clark 1998; Imhof *et al.* 1998). The presence of sperm from two or more males in the female reproductive tract promotes "competition" between males for access to fertilizations. As is true for most insects (Smith 1984; Simmons and Siva-Jothy 1998), if two *Drosophila melanogaster* males mate in succession with a single female, the sperm from the second male fertilizes the majority of eggs (Gromko and Pyle 1978; Gromko *et al.* 1984). Sperm competition is a complex phenomenon, the outcome of which may be affected by a large number of different gene products and contexts (Simmons and Siva-Jothy 1998). For example, Drosophila seminal fluid (as opposed to sperm alone) has been shown to play a role in sperm displacement (Harshman and Prout 1994). In addition to its role in sperm competition, seminal fluid stimulates oviposition and reduces female receptivity to future matings (Richmond *et al.* 1980; Chen 1984, 1996). There is abundant genetic variation affecting both male and female components of sperm use in Drosophila (Clark *et al.* 1995, 1999; Rice 1996; Clark and Begun 1998). Given our intuition that the outcome of sperm competition should be highly correlated with male fitness, the large amount of genetic variation affecting the trait is somewhat surprising.

Accessory gland proteins are an important component of Drosophila seminal fluid. Two-dimension gel electrophoresis indicates that there are large numbers of seminal proteins (Thomas and Singh 1992), the functions of most of which are unknown. Protein electrophoretic analysis of accessory gland proteins showed that on average they are more polymorphic than, and evolve faster than, the other proteins examined (Coulthart and Singh 1988; Thomas and Singh 1992; Civetta and Singh 1995). However, those data provide no means of distinguishing whether rapid evolution is best explained as a result of natural selection or rather as a result of reduced functional constraints. Molecular population genetic analysis of accessory gland protein encoding genes *Acp26Aa* and *Acp26Ab* suggested that natural selection has caused amino acid evolution of the former, but not the latter (Aguadé 1997, 1998; Tsaur and Wu 1998). Analysis of *Acp29AB* (Aguadé 1999) also provided weak evidence for adaptive protein evolution in *D. melanogaster* and *D. simulans*, yet analysis of *Acp70A* (Cirera and Aguadé 1997) provided no support for the notion that selection has caused interspecific divergence in this gene. Therefore, despite the broad appeal of the idea that amino acid evolution of seminal fluid proteins is driven primarily by selection, it is supported by little data. Here we present an analysis of variation in 13 accessory gland proteins in *D. melanogaster* and

Corresponding author: David Begun, Section of Evolution and Ecology, University of California, Davis, CA 95616.
E-mail: djbegun@ucdavis.edu

D. simulans. These data allow us to test the general hypothesis that interspecific divergence in this class of proteins is primarily driven by natural selection.

MATERIALS AND METHODS

D. simulans data for nine *Acp* genes (*Acp28AB*, *Acp32CD*, *Acp33A*, *Acp36DE*, *Acp53Ea*, *Acp62F*, *Acp63F*, *Acp76A*, *Acp95EF*, and *Acp98AB*) are from 10–14 highly inbred lines made from females captured at the Wolfskill Orchard in Winter, California. *D. melanogaster* data from the United States for these same nine genes are from 10–12 isogenic stocks derived from females also captured at Wolfskill. *D. melanogaster* data from Africa for these genes were from six to eight homozygous chromosomes isolated from a Zimbabwe sample or from single Zimbabwe chromosomes placed over appropriate deficiencies. Populations of *D. melanogaster* from Zimbabwe are often different from those in other locations and may represent populations that are closer to equilibrium (Begun and Aquadro 1993). DNA sequences were obtained directly from PCR products on ABI (Foster City, CA) 377 and Beckman (Fullerton, CA) CEQ-2000 automated sequencers. Data from *Acp26Aa* and *Acp26Ab* were from Aguadé *et al.* (1994) and Aguadé (1998). Data for *Acp70A* were from Cirera and Aguadé (1997). Sequences were analyzed using DnaSP (Rozas and Rozas 1999), SITES (Hey and Wakeley 1997), and the molecular evolutionary analysis package (E. Moriyama, unpublished data). Sequences can be found under GenBank accession nos. AY010527–AY010711.

RESULTS

Summaries of polymorphism and divergence in *Acp* genes are presented in Tables 1–3 and Figures 1 and 2. Most *Acp* genes are short, with the consequence that for many genes there are too few segregating sites or interspecific differences to test evolutionary models or estimate model parameters with much confidence. Consequently, the most powerful analyses of these data ask questions about polymorphism and divergence across the whole set of genes. We present these analyses first. We then note interesting observations at individual genes.

Polymorphism and divergence in *Acp* genes: There are 9 *Acp* loci (*Acp28AB*, *Acp32CD*, *Acp33A*, *Acp36DE*, *Acp53Ea*, *Acp62F*, *Acp63F*, *Acp76A*, and *Acp98AB*) for which we have polymorphism data from *D. simulans* and *D. melanogaster* (Table 2). The mean replacement θ (Watterson 1975) for these 9 *D. simulans Acp* genes (weighted by length of surveyed region) is 0.0074, compared to a mean replacement θ of 0.0013 for 19 *D. simulans* genes scattered across chromosome arm 3R (Begun and Whitley 2000). None of these *Acp* genes or other genes on 3R are in regions of very low recombination in *D. simulans*, so levels of heterozygosity between *Acp* genes and other genes can be compared without accounting for effects of differences in recombination rate between classes of proteins. Mean replacement divergence for the 9 *Acp* genes is 0.0497, while the mean replacement divergence for 19 genes on 3R is 0.0107. Thus, replacement heterozygosity is about five to six times greater for *Acp*'s than for 19 genes on 3R (Begun and Whitley 2000), while the replacement divergence is about four to five times greater than for the 19 genes on 3R. Mean (weighted by length) silent heterozygosity for the nine *D. simulans Acp* genes is 0.028. This is close to the average silent heterozygosity, 0.035, for the 19 *D. simulans* genes surveyed on 3R. The average silent divergence for these 9 *Acp* genes is 0.117, also very similar to previous results from other genes compared in these two species (Bauer and Aquadro 1997; Begun and Whitley 2000).

We have polymorphism data from African *D. melanogaster* samples for these nine Acp genes; the mean replacement heterozygosity is 0.0025, compared to 0.0074 for *D. simulans*. The approximately threefold difference in heterozygosity is roughly similar to that previously observed between these two species (*e.g.*, Moriyama and Powell 1996). There are five *Acp* genes (*29AB*, *32CD*, *33A*, *36DE*, and *53Ea*) for which we have polymorphism from both African and non-African samples of *D. melanogaster*, as well as for *D. simulans*. The mean replacement heterozygosity is essentially the same in the African and non-African *D. melanogaster* samples for these five loci (0.0027 and 0.0026, respectively). Silent site heterozygosity is also very similar for these five genes in the African and non-African samples (Table 3) and for a slightly larger sample of seven *Acp* genes for which we have survey data for African and non-African samples (silent θ for the two samples are 0.018 and 0.014, respectively). Possible exceptions to this generalization can be found in the case of *Acp36DE* (for which the African sample is about twice as variable as the non-African sample) and *Acp95EF* (for which the non-African sample is completely lacking variation, while the African sample has more "typical" levels of silent heterozygosity).

Comparison of numbers of silent and replacement polymorphisms in *D. melanogaster* and *D. simulans* suggests that the ratio of replacement to silent polymorphism for autosomal genes is different in the two species (Begun 1996; Moriyama and Powell 1996). For nine *Acp* genes (all *Acp*'s are autosomal; Wolfner 1997) for which we have population samples for *D. simulans* and African *D. melanogaster* samples, the numbers of silent and replacement polymorphisms in *D. simulans* are 76 and 81, respectively; the numbers of silent and replacement polymorphisms in African *D. melanogaster* are 42 and 21, respectively. The 2×2 contingency table is significantly heterogeneous by a *G*-test ($P = 0.013$). This heterogeneity supports the earlier analysis of data from non-*Acp* genes (Begun 1996; Moriyama and Powell 1996). Importantly, however, data from *Acp*'s is *opposite* in direction from the data from other loci. As previously noted (Aquadro *et al.* 1988; Begun 1996), the general pattern is one of proportionally more replacement polymorphism in *D. melanogaster* than in *D. simulans* (relative to silent polymorphism). For *Acp* loci, it is *D. simulans*

TABLE 1

Silent and replacement variation in *Acp* genes of *D. melanogaster* and *D. simulans*

		Poly		Fixed		
		Silent	Replacement	Silent	Replacement	Prob.[d]
Acp26Aa[a]	*mel* (United States)	7	9	24	78	0.109
	mel (Malawi)	19	15	20	77	0.002
Acp26Ab[a]	*mel* (United States)	3	2	2	3	
	mel (Malawi)	7	3	2	3	
Acp29AB	*sim*	12	6			
	mel (United States)	8	3	24	33	0.017
	mel (Africa)	3	2	25	33	0.053
	mel (United States + Africa)	11	4	23	33	
	mel[b] (Europe)	10	3	33	37	
	mel[b] (Africa)	13	5	33	36	
Acp32CD	*sim*	2	10			
	mel (United States + Africa)	0	3	1	1	—
Acp33A	*sim*	1	2			
	mel (United States)	2	0	1	0	—
	mel (Africa)	2	0	0	0	—
	mel (United States + Africa)	2	0			
Acp36DE	*sim*	24	25			
	mel (United States)	15	15	55	72	0.509
	mel (Africa)	24	10	52	72	0.025
	mel (United States + Africa)	31	23	49	72	
Acp53Ea	*sim*	2	4			
	mel (United States)	7	1	10	7	0.756
	mel (Africa)	6	5	11	7	0.404
	mel (United States + Africa)	9	6	10	7	
Acp62F	*sim*	14	19			
	mel (Africa)	5	0	0	0	—
Acp63F	*sim*	4	10			
	mel (Africa)	1	1	5	14	1.0
Acp70[c]	*mel* (Europe)	3	1	—	—	—
Acp76A	*sim*	17	5			
	mel (Africa)	1	0	14	11	0.099
Acp95EF	*sim*					
	mel (United States)	0	0	—	—	
	mel (Africa)	1	1	—	—	
	mel (United States + Africa)	1	1	9	4	—
Acp98AB	*sim*	0	0			
	mel	0	0	0	8	—

[a] Aguadé (1998).
[b] Aguadé (1999).
[c] Cirera and Aguadé (1997).
[d] Probability determined by *G*-test. Tests were carried out on polymorphism data from *D. simulans* (when possible) and the *D. melanogaster* samples indicated. Tests were not carried out for loci with very few observations.

rather than *D. melanogaster* that appears to have proportionally more replacement polymorphism.

Joint analysis of polymorphism and divergence at silent and replacement sites can provide more powerful tests of evolutionary models than can analysis of polymorphism alone (*e.g.*, Hudson *et al.* 1987). The numbers of silent and replacement polymorphisms in the two species can be compared to the numbers of silent and replacement "fixed" differences between species at the *Acp* genes (McDonald and Kreitman 1991). The null hypothesis under neutrality is that the ratio of silent to replacement polymorphism will be similar to the ratio

TABLE 2

Silent and replacement site heterozygosity and divergence at *Acp* genes of *D. melanogaster* and *D. simulans*

Gene	No. of sites		Sample	$\theta_{sil.}$	$\theta_{repl.}$	$Div._{sil.}$	$Div._{repl.}$
	Sil.	Repl.					
Acp26Aa[a]	174	615	*mel* (United States, $n = 10$)	0.014	0.006	0.167	0.156
			mel (Africa, $n = 18$)	0.033	0.008		
Acp26Ab[a]	55	215	*mel* (United States, $n = 10$)	0.019	0.003	0.059	0.018
			mel (Africa, $n = 18$)	0.037	0.004		
Acp29AB	143	544	*sim* ($n = 8$)	0.0323	0.0043	0.2063	0.0740
			mel (United States, $n = 9$)	0.0235	0.0020		
			mel (Africa, $n = 8$)	0.0080	0.0014		
Acp29AB[b]	145	557	*mel* (Europe, $n = 12$)	0.0229	0.0018		
			mel (Africa, $n = 13$)	0.0290	0.0029		
Acp32CD	162	489	*sim* ($n = 6$)	0.0054	0.0063	0.0119	0.0140
			mel (United States, $n = 8$)	0.0000	0.0034		
			mel (Africa, $n = 9$)	0.0000	0.0022		
Acp33A	24	72	*sim* ($n = 8$)	0.0164	0.01065	0.0814	0.0069
			mel (United States, $n = 9$)	0.0308	0.0000		
			mel (Africa, $n = 9$)	0.0271	0.0000		
Acp36DE	494	1705	*sim* ($n = 6$)	0.0296	0.0041	0.1318	0.0489
			mel (United States, $n = 9$)	0.0110	0.0032		
			mel (Africa, $n = 7$)	0.0195	0.0024		
Acp53Ea	88	269	*sim* ($n = 8$)	0.0087	0.0057	0.1428	0.0393
			mel (United States, $n = 10$)	0.0277	0.0013		
			mel (Africa, $n = 5$)	0.0327	0.0089		
Acp62F	59	184	*sim* ($n = 8$)	0.0921	0.0398	0.1259	0.0502
			mel (Africa, $n = 10$)	0.0241	0.0000		
Acp63F	41	148	*sim* ($n = 7$)	0.0395	0.0276	0.1756	0.1321
			mel (Africa, $n = 9$)	0.0089	0.0024		
Acp70A[c]	38	127	*mel* (Europe, $n = 9$)	0.0389	0.0029	0.1235	0.0280
Acp76A	165	576	*sim* ($n = 6$)	0.0451	0.0038	0.1417	0.0252
			mel (Africa, $n = 10$)	0.0021	0.0000		
Acp95EF	41	115	*mel* (United States, $n = 9$)	0.0000	0.0000		
			mel (Africa, $n = 5$)	0.0116	0.0000	0.2226	0.0366
Acp98AB	20	67	*sim* ($n = 8$)	0.0000	0.0000	0.0000	0.1188
			mel (Africa, $n = 10$)	0.0000	0.0000		

Divergence is between all pairs of *D. melanogaster* and *D. simulans* genes (Jukes-Cantor corrected). Sil., silent; Repl., replacement; Div., divergence.

[a] Polymorphism data for *Acp26Aa* and *Acp26Ab* are from Aguadé (1998). Divergence data for *Acp26Aa* and *Acp26Ab* are from Aguadé *et al.* (1992).

[b] Polymorphism data for *Acp29AB* not from this article are from Aguadé (1999).

[c] Polymorphism and divergence data for *Acp70A* are from Cirera and Aguadé (1997).

of silent to replacement divergence. We include data from all the genes in Table 2 for our analysis of silent polymorphism and fixations. Where possible, we use polymorphism data from African *D. melanogaster* samples (Table 1). The total numbers of polymorphisms and fixed differences are shown in Table 3. The contingency table reflects significant heterogeneity ($G = 17.5$; $P < 0.001$). This is not surprising given the significant deviations from homogeneity observed at *Acp26Aa* and *Acp-36DE* individually. If we remove the data from these two genes the resulting 2×2 contingency table (Table 3) is not significantly heterogeneous. By this criterion there is no compelling evidence that directional selection is the predominant cause of amino acid substitutions in these genes (although the strength of such a conclusion is compromised by the reduction in statistical power associated with removing large numbers of sites from *Acp26Aa* and *Acp36DE* from the analysis).

TABLE 3

Contingency tables of polymorphic and fixed variation at *Acp* genes

	Silent	Replacement	
	All *Acp* loci		
Polymorphic	158	142	
Fixed	132	230	$G = 17.5$, $P = 0.00003$
	Without *Acp26Aa* and *Acp36DE*		
Polymorphic	91	92	
Fixed	60	81	$G = 1.65$, $P = 0.20$

D. melanogaster data are from African samples whenever possible.

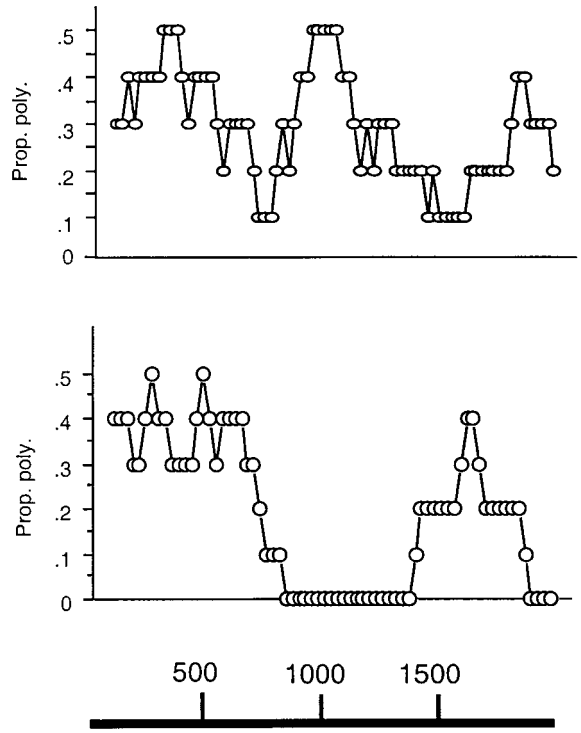

FIGURE 1.—Proportion of sites polymorphic [polymorphic/(polymorphic + fixed)] in the *Acp36DE* genes of *D. melanogaster* samples from Zimbabwe (top) and the United States (bottom), as determined from the runs test of McDONALD (1996).

Despite this statistical inference regarding amino acid evolution pooled across loci, we should not be too quick to conclude that polymorphism and divergence data provide no support for the notion that directional selection plays a role in *Acp* protein evolution. Under the neutral model one expects the ratio of replacement to silent fixations to equal the ratio of replacement to silent polymorphisms. However, data sets from genes evolving under the neutral model should show minor deviations from this expectation because of sampling and stochastic variances. This suggests that for a sample of several loci, a null hypothesis under the neutral model is that half the loci are expected to show a greater ratio of replacement to silent fixations than the ratio of replacement to silent polymorphisms, while half are expected to show a smaller ratio of replacement to silent fixations compared to the ratio of replacement to silent polymorphisms.

The data from *Acp*'s appear to be inconsistent with this prediction. Note that there is a consistent pattern across genes of greater ratios of replacement to silent fixations compared to ratios of replacement to silent polymorphisms. For the seven *Acp* genes for which there are at least five total polymorphisms and at least five total fixations, all seven show proportionally more replacement fixations than replacement polymorphisms. If, under the neutral model, a locus has a 50% probability of a greater ratio of replacement to silent divergence than replacement to silent polymorphism, our observation of seven loci with proportionally more replacement fixations than replacement polymorphisms is highly unlikely (binomial probability, $P = 0.008$). How do the *Acp*'s compare to other genes sequenced in these species? The data set from BEGUN and WHITLEY (2000) may be most appropriate, as there is no bias in the genes sampled toward those thought to be under natural selection on the basis of other types of data (*e.g.*, allozyme data). Among 35 genes that satisfy the criterion of at least five total polymorphisms and five total fixations, 25 show a greater ratio of replacement to silent fixations compared to the ratio of replacement to silent polymorphism; only 10 show a smaller ratio of replacement to silent fixations compared to the ratio of replacement to silent polymorphism. Comparison of the ratios in the two sets of loci (7:0 to 25:10) is not significant by Fisher's exact test. This comparison is somewhat problematic, however, in that the sample of 35 genes from BEGUN and WHITLEY (2000) includes no polymorphism data from this species. Unfortunately, only five *Acp*'s satisfy our criterion of at least five total polymorphisms and at least five total fixations when polymorphism data from *D. melanogaster* are removed. Comparison of replacement to silent fixations *vs.* replacement to silent polymorphisms for this subset of the *Acp*'s *vs.* other loci yields 3:2 to 25:10, which is not significant. However, given that only five genes are analyzed one cannot draw a strong conclusion. As noted earlier, the probability of observing 7 of 7 loci with greater ratios of replacement to silent fixations than replacement to silent polymorphisms is unlikely under the neutral model. However, our "random" sample of 35 loci shows 25 with a greater ratio of replacement to silent fixations than of replacement to silent polymorphism; this too is unlikely under the neutral model ($P = 0.01$). We should not conclude from this, however, that adaptive amino acid fixations are the cause of the pattern. The data are confounded by the pooling of substitutions that occurred along two lineages (*D. melanogaster* and *D. simulans*) with polymor-

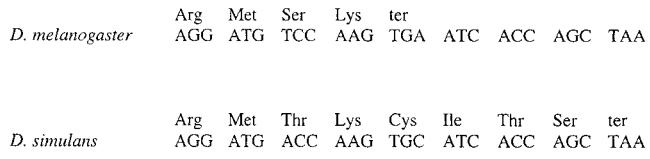

FIGURE 2.—Organization of the 3′ end of the *Acp98AB* gene in *D. melanogaster* and *D. simulans*.

phism data. AKASHI (1996) showed that relative to silent substitution rates, the amino acid substitution rate is higher in the *melanogaster* lineage than in the *simulans* lineage. The extent to which this pattern results from directional selection rather than genetic drift remains unclear.

In general, then, despite the large amount of data from *Acp*'s, we are still unable to make strong statements regarding the importance of directional selection in the evolution of these genes. The addition of data from outgroup species for *Acp*'s and other loci will help address this uncertainty.

As we might expect given the results of AKASHI (1996), average codon bias (WRIGHT 1990; Table 4) in *D. simulans Acp* genes is greater than the average codon bias in *D. melanogaster Acp* genes [effective number of codons (ENC) = 50.6 and 53.3, respectively]. The absolute levels of codon bias in *Acp* genes are low compared to the average degree of codon bias in Drosophila genes (POWELL and MORIYAMA 1997; MCVEAN and VIEIRA 1999). The accessory gland protein-coding genes were isolated in such a way that favors overrepresentation of highly expressed genes (DIBENEDETTO *et al.* 1990; BERTRAM *et al.* 1996; WOLFNER *et al.* 1997). Furthermore, accessory gland proteins are fairly short. Both factors should favor higher, rather than lower degrees of codon bias (reviewed in POWELL and MORIYAMA 1997). One possible explanation for the relatively low levels of codon bias in *Acp*'s is the notion that rapid rates of protein evolution (as we see in *Acp*'s) should be correlated with lower levels of codon bias (*e.g.*, AKASHI 1994). An unusual aspect of codon bias of *Acp* genes is that *Acp63F* and *Acp95EF* have quite different ENC values in the two species (with *D. melanogaster* showing much less bias than *D. simulans*). Data from outgroup species for these genes would be required to understand how evolution of silent sites has proceeded differently along the two lineages.

Patterns at individual genes: *Acp36DE:* We observed 24 silent and 25 replacement polymorphisms in our sample of 6 *D. simulans* alleles, and we observed 54 silent and 23 replacement polymorphisms in our sample of 16 *D. melanogaster* alleles (Table 1). Levels of silent polymorphism as measured by θ and π are about the same in the *D. simulans* sample and the Zimbabwe *D. melanogaster* sample. As has been observed in several genes (*e.g.*, BEGUN and AQUADRO 1993), the amount of silent variation is less in U.S. samples than in Zimbabwe samples of *D. melanogaster*. Comparison of replacement and silent polymorphism in the three samples reveals that the ratio of replacement to silent polymorphism is the same in the *D. simulans* sample and the U.S. *D. melanogaster* sample. However, the ratio of replacement to silent polymorphism is lower in the Zimbabwe *D. melanogaster* sample. In fact, although the U.S. sample is about half as polymorphic as the Zimbabwe sample for silent variation, the two samples are about equally variable for replacement variation.

TABLE 4

Codon bias and base composition at fourfold degenerate sites in *Acp* genes of *D. melanogaster* and *D. simulans*

	No. codons		ENC		% GC	
Gene	*mel*	*sim*	*mel*	*sim*	*mel*	*sim*
Acp26Aa	257	256	61.0	61.0	49.0	51.9
Acp26Ab	90	91	54.7	49.1	58.6	63.3
Acp29AB	234	229	59.9	61.0	46.8	50.6
Acp32CD	277	275	58.7	58.8	51.0	50.7
Acp33A	44	44	46.7	41.7	23.8	31.8
Acp36DE	576	573	52.8	53.4	34.5	34.9
Acp53Ea	120	120	58.9	61.0	42.8	45.5
Acp62F	90	91	61.0	56.0	49.0	49.9
Acp63F	67	68	61.0	45.7	32.0	28.0
Acp70A	55	55	59.9	61.0	37.0	37.0
Acp76A	336	336	56.6	55.9	51.5	58.3
Acp95EF	51	51	48.9	30.7	36.7	51.6
Acp98AB	33	33	25.8	28.1	66.7	69.2

Tests of homogeneity of the silent and replacement polymorphic and fixed variants (MCDONALD and KREITMAN 1991) show that silent and replacement polymorphism and divergence of the *D. simulans* and Zimbabwe *D. melanogaster* samples are significantly heterogeneous ($P = 0.03$). Under the assumption that amino acid mutations are likely to be under stronger selection than are silent mutations, the data support the idea that there is an excess of amino acid fixations at *Acp36DE*. However, the polymorphism and divergence from the *D. simulans* and U.S. *D. melanogaster* sample (Table 1) are not significantly heterogeneous ($P = 0.51$). In other words, the results of the two tests are indicative of significant differences in the configuration of silent and replacement variation in the Zimbabwe and U.S. *D. melanogaster* samples. A sliding window analysis of θ for silent sites in the two *D. melanogaster* samples (Figure 1) shows that the distribution of variation in the two samples is roughly similar, the only exception being a region around nucleotides 750–1500, where there is a large drop in the amount of variation in the U.S. sample. We performed the "runs" tests of MCDONALD (1996) on the two *D. melanogaster* samples. There was a highly significant deviation from the null hypothesis in the U.S. sample, but no deviation in the Zimbabwe sample (Table 1). Inspection of the silent and replacement variation in the U.S. and Zimbabwe *D. melanogaster* samples over

this region shows that there are eight silent and two replacement polymorphisms in Zimbabwe and zero silent and six replacement polymorphisms in the United States. These ratios of silent to replacement polymorphisms are significantly different by Fisher's exact test ($P = 0.009$), although the P value cannot be taken literally because the test was performed on a subset of the data that appeared "unusual." Nevertheless, the results suggest different types of deviations from the null hypothesis in the Zimbabwe and U.S. *D. melanogaster* samples. The comparison of Zimbabwe *D. melanogaster* and *D. simulans* provides evidence of amino acid differences between species caused by natural selection. The U.S. *D. melanogaster* sample provides evidence for a very recent perturbation of the polymorphism by natural selection. Specifically, a small region within the Acp36DE protein in the U.S. *D. melanogaster* sample appears to be depauperate of silent variation and to have "extra" amino acid variation. We cannot rule out the possibility that selection on silent sites contributes to rejection of the null model; however, the fact that selection acting on codon bias at *Acp36DE* is weak supports the idea that selection on amino acids is a likely cause of rejection of the neutral model.

Acp29AB: A test of the contingency table of polymorphic/fixed and silent/replacement variation at *Acp29AB* results is a significant rejection of homogeneity (Table 1). A previous analysis of Acp29AB yielded similar results (Aguadé 1999). As was the case for *Acp36DE*, the configuration of the contingency table is most readily interpreted in terms of an excess of amino acid fixations between species. Our estimate of silent heterozygosity for a U.S. sample of *D. melanogaster* was similar to that reported by Aguadé (1998) for a French sample (0.024 *vs.* 0.023). Aguadé (1999) reported very similar levels of silent heterozygosity in French and African (Lamto and Malawi) samples. Our African sample from Zimbabwe seems to be different from the African samples used by Aguadé (1999), as our silent heterozygosity is only about one-third that of samples from other locations. Comparison of the Zimbabwe *D. melanogaster* to the African data from Aguadé (1999) confirms that the two samples are in fact significantly different ($K_{st} = 0.07$, $P < 0.001$ by a permutation test of Hudson *et al.* 1992). The Zimbabwe *D. melanogaster* sample was polymorphic at four amino acid residues; these four residues were also polymorphic in the *D. melanogaster* sample analyzed by Aguadé (1999). Aguadé (1999) reported two polymorphic residues that were monomorphic in our sample; 105 (present twice in Aguadé 1999) and 215 (present once in Aguadé 1999).

Acp32CD: Acp32CD appears to be an unusual *Acp* gene in that it is evolving slowly at both silent and replacement sites. *Acp32CD* also harbors relatively little polymorphism. This is especially true of silent sites. No silent polymorphisms were observed in *D. melanogaster* and only two silent polymorphisms were seen in *D. simulans*. Replacement polymorphism was also low in both species, although not atypically low relative to levels of replacement polymorphism at other *Acp* genes (Table 2). Under the neutral model these observations would be interpreted as a result of unusually high functional constraints in this gene compared to functional constraints on other *Acp*'s (Kimura 1983). An alternative viewpoint (Gillespie 1991) is that directional selection impinges on this locus to a lesser degree than it does at other *Acp* loci. As a consequence of low polymorphism and low divergence in a short gene region of just 415 bp, the contingency table of polymorphisms and fixed differences is not significantly heterogeneous. It should be noted, however, that with a contingency table containing only two fixed differences (one silent and one replacement), it is impossible to reject the null hypothesis.

Acp62F: Acp62F shows an unusual pattern of polymorphism and divergence not previously observed in samples of genes from *D. simulans* and *D. melanogaster*. The silent heterozygosity in *D. simulans* (0.092) is very high compared to the average silent heterozygosity of autosomal genes (Begun and Whitley 2000) in this species (0.035), although the 95% confidence interval (C.I.; Kreitman and Hudson 1991) for *D. simulans Acp62F* (0.020–0.202) does overlap the average silent θ for autosomal loci in *D. simulans* (Begun and Whitley 2000). Perhaps more surprising than the high heterozygosity in *D. simulans* is the fact that two of the *D. simulans* alleles, DS6C and DS7C, cluster with *D. melanogaster* alleles rather than other *D. simulans* alleles in both neighbor-joining and maximum-likelihood trees (data not shown), although sequence data from an outgroup species will be required to root such a tree. Balancing selection is one explanation for high levels of polymorphism of the *D. simulans Acp62F*. However, a Hudson-Kreitman-Aguadé (HKA) test (Hudson *et al.* 1987) of *Acp62F vs. AATS-GluPro* (a "typical" *D. simulans* autosomal locus in terms of polymorphism and divergence; Begun and Whitley 2000) provides no evidence for deviations from the neutral model ($\chi^2 = 1.10$, $P = 0.29$).

Acp98AB: Acp98AB is unusual in that a mutation in a termination codon has created a "fixed" length difference between species in the protein. The final base of the TGA termination codon in *D. melanogaster* is a C in *D. simulans*, with the result that the codon homologous to the *D. melanogaster* stop codon codes for cysteine in *D. simulans* (Figure 2). The two species have the same sequence for the next 12 bases; the last triplet among these 12 bases codes for a termination codon in *D. simulans*. This unusual mutation at a stop codon was independently confirmed with a PCR-restriction fragment length polymorphism analysis of the mutation. *Acp98AB* is also unusual in the patterns of polymorphisms and divergence in the amino acid residues that are shared by both species. There are no polymorphisms of any type within either species and no differences

between species at silent sites. However, there are eight amino acid differences between species. This pattern is suggestive of adaptive protein evolution. However, the small size of the locus (160 bp) diminishes our ability to carry out hypothesis tests of variation at this gene.

DISCUSSION

There is abundant evidence that male sexual traits often evolve more quickly than other kinds of traits (*e.g.*, Eberhard 1985). Whether this pattern extends in a very general sense to genes required for male reproduction is an important issue of evolutionary genetics. A subset of the proteins of male reproduction in Drosophila are the accessory gland proteins. Although the functions of most *Acp*'s are unknown, it is clear that seminal fluid of Drosophila affects female physiology (Chapman *et al.* 1995; Rice 1996), behavior (Chen *et al.* 1988), patterns of sperm storage by females (Tram and Wolfner 1999), and sperm competition between males (Harshman and Prout 1994). Accessory gland proteins are candidates for mediating at least some of these phenomena.

Evidence from protein gel electrophoresis has suggested that the accessory gland proteins tend to evolve more quickly and tend to be more polymorphic than most other proteins in Drosophila (Coulthart and Singh 1988; Thomas and Singh 1992; Civetta and Singh 1995). However, the nature of the electrophoretic data precluded accurate descriptions of the underlying genetic changes within and between species. The data presented here provide our first general picture of polymorphism and divergence of accessory gland proteins in two closely related Drosophila species.

Our data suggest that the patterns revealed by protein electrophoresis and by previous studies of DNA variation at *Acp* genes are in fact general properties of accessory gland proteins. As a group, *Acp* genes exhibit greater protein polymorphism and greater protein divergence than the "average" gene in *D. simulans* and *D. melanogaster*. The small size of the *Acp* genes reveals the limitations of molecular population genetics to test the null model of neutral variation within and between species. The three largest *Acp* genes, *Acp26Aa*, *Acp29AB*, and *Acp36DE*, each show patterns of polymorphism and divergence consistent with adaptive protein divergence. None of the remaining genes individually exhibit statistically significant deviations from the neutral model. A contingency table of polymorphic and fixed, silent, and replacement mutations of the *Acp* genes (omitting *Acp26Aa*, *Acp29AB*, and *Acp36DE*) does not reject the neutral model, even though there are large numbers of sites surveyed in the pooled data. The results provide no compelling evidence that a large fraction of amino acid substitutions in these proteins result from directional selection. This inference about replacement sites, however, does not speak strongly to the issue of whether most *Acp* loci have been targets of directional selection at relatively few amino acid positions. Statements regarding the ubiquity of directional selection across loci are even more difficult for this set of genes because many of the genes are quite small (leading to low statistical power).

One prediction from theoretical studies of genes affecting sperm displacement is that polymorphism in such genes can be maintained by balancing selection (Prout and Clark 1996). These theoretical results along with the empirical observations of abundant genetic variation in male components of sperm competition and the potential role of *Acp*'s in such phenomena raise the issue of whether there is any evidence for balancing selection in the DNA data from these genes. One "signature" of balancing selection is excess polymorphism among sites that are tightly linked to a balanced polymorphism (Hudson and Kaplan 1988; Kreitman and Hudson 1991). The only potential candidate for balancing selection from the molecular data is the *Acp62F* gene in *D. simulans*, which shows an unusually high level of silent polymorphism and an unprecedented topology in which some *D. simulans* alleles show less divergence from *D. melanogaster* alleles than from other *D. simulans* alleles. However, given the small size of the regions surveyed for this article, additional data from flanking regions of these species as well as outgroup species will be required to determine if there is evidence for old, balanced polymorphisms in the *Acp* genes.

Another possibility is that rather than harboring old balanced polymorphisms, *Acp* genes may be subject to transient, yet strongly selected polymorphism (*cf.* Gillespie 1991). In such circumstances it is by no means clear that there is an expectation for excess polymorphism. The properties of samples of genes experiencing such selective histories are not well understood (but see Gillespie 1997). *Acp36DE* is particularly interesting with regard to models of the maintenance of variation. Haplotype variation at *Acp36DE* in U.S. samples of *D. melanogaster* was found to be correlated with sperm displacement phenotypes (Clark *et al.* 1995). Subsequent evidence from direct experiments has confirmed a role in sperm storage for this protein (Bertram *et al.* 1996; Neubaum and Wolfner 1999). Here we present evidence that natural selection has contributed to protein divergence between species. Our analysis also suggests that there has been a very recent episode of natural selection at this locus in U.S. populations of *D. melanogaster*. An important task for the future will be to determine the specific polymorphisms in *Acp36DE* having measurable effects on patterns of sperm use and to determine the history of these mutations. Such analyses will considerably improve our understanding of both the biology and the evolutionary dynamics of fitness variation associated with sperm use in Drosophila.

Even at this early stage, however, it is worth taking note of the fact that of the four *Acp* genes that appear to harbor alleles of large effect in *D. melanogaster* populations (Clark *et al.* 1995), three of them, *Acp26Aa*, *Acp-*

29AB, and *Acp36DE*, also show evidence of adaptive protein evolution between species. Perhaps the appropriate model for *Acp* genes is one in which there is a succession of short-term balanced polymorphisms that give rise to fixations as the selective environment changes over time.

Perhaps the most interesting general observation regarding *Acp* protein variation in *D. simulans* and *D. melanogaster* is the observation of proportionally more replacement polymorphism in *D. simulans*. This pattern is in stark contrast to several observations in the species pair (Aquadro *et al.* 1988; Moriyama and Powell 1996). There are two sorts of explanations for such a pattern. First, under the neutral model we would expect such an observation if the neutral mutation rate for amino acids at *Acp*'s were higher in *D. simulans* than in *D. melanogaster*. Alternatively, if balancing selection contributes to the amino acid polymorphism in these species, then one could explain the data by invoking stronger selection in *D. simulans* than in *D. melanogaster*. If this latter hypothesis were true, then one might predict that more of the *Acp* protein polymorphism would have measurable phenotypic effects in *D. simulans* than in *D. melanogaster*.

In a similar vein, one benefit of the data presented here is increased ability to identify polymorphisms or haplotypes that might have effects on patterns of sperm storage or sperm use. Clark *et al.* (1995) demonstrated that there were associations between sperm displacement phenotypes in *D. melanogaster* and single-strand conformation polymorphism (SSCP) alleles at *Acp* genes *26Aa*, *29AB*, *36DE*, and *53Ea*. For example, they reported on phenotypic effects associated with six SSCP alleles at *Acp53Ea* in *D. melanogaster* (in a sample of 36 chromosomes). However, the SSCP data did not allow for identification of individual mutations that might be responsible for the phenotypic variation and did not even permit determination of the number of polymorphic sites in the sample. Our sequence analysis of *Acp53Ea* reveals only a single amino acid polymorphism in the non-African sample of *D. melanogaster*; this polymorphism is now an excellent candidate for further analysis of phenotypic effects. The continued function characterization of the *Acp* genes will be important for determining which phenotypes should be examined in such association studies. Given the small size of most *Acp* genes, direct experimental confirmation of intermediate frequency polymorphisms in *Acp* genes that contribute to variation in sperm storage and sperm use may be the most direct strategy for determining whether positive selection often acts on *Acp* polymorphisms in Drosophila.

LITERATURE CITED

Aguadé, M., 1997 Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of Drosophila. Mol. Biol. Evol. **14:** 544–549.

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

Aguadé, M., 1999 Positive selection drives the evolution of the Acp-29AB accessory gland protein in Drosophila. Genetics **152:** 543–551.

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.

Akashi, H., 1994 Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics **136:** 927–935.

Akashi, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics **144:** 1297–1307.

Aquadro, C. F., K. M. Lado and W. A. Noon, 1988 The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. Genetics **119:** 875–888.

Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **14:** 1252–1257.

Begun, D. J., 1996 Population genetics of silent and replacement variation in *Drosophila simulans* and *D. melanogaster*: X/autosome differences? Mol. Biol. Evol. **13:** 1405–1407.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Begun, D. J., and P. Whitley, 2000 Reduced X-linked nucleotide polymorphism in Drosophila simulans. Proc. Natl. Acad. Sci. USA **97:** 5960–5965.

Bertram, M. J., D. M. Neubaum and M. F. Wolfner, 1996 Localization of the Drosophila male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. Insect Biochem. Mol. Biol. **26:** 971–980.

Chapman, T., L. F. Liddle, J. M. Kalb, M. F. Wolfner and L. Partridge, 1995 Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature **373:** 241–244.

Chen, P. S., 1984 The functional morphology and biochemistry of insect male accessory glands and their secretions. Annu. Rev. Entomol. **20:** 233–255.

Chen, P. S., 1996 The accessory gland proteins in male Drosophila: structural, reproductive, and evolutionary aspects. Experientia **52:** 503–510.

Chen, P. S., E. Stumm-Zollinger, T. Aigaki, J. Balmer, M. Bienz *et al.*, 1988 A male accessory gland peptide that regulates reproductive behaviour of female *D. melanogaster*. Cell **54:** 291–298.

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, 1995 High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. J. Mol. Evol. **41:** 1085–1095.

Clark, A. G., and D. J. Begun, 1998 Female genotypes affect sperm displacement in Drosophila. Genetics **149:** 1487–1493.

Clark, A. G., M. Aguadé, T. Prout, L. G. 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.

Clark, A. G., D. J. Begun and T. Prout, 1999 Female × male interactions for sperm displacement in *Drosophila melanogaster*. Science **283:** 217–220.

Cobbs, G., 1977 Multiple insemination and male sexual selection in natural populations of *Drosophila pseudoobscura*. Am. Nat. **111:** 641–656.

Coulthart, M. B., and R. S. Singh, 1988 Differing amounts of genetic polymorphism in testes and male accessory glands of *Drosophila melanogaster* and *D. simulans*. Biochem. Genet. **26:** 153–164.

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.

Eberhard, W. G., 1985 *Sexual Selection and Animal Genitalia*. Harvard University Press, Cambridge, MA.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution.* Cambridge University Press, Cambridge, UK.

Gillespie, J. H., 1997 Junk ain't what junk does: neutral alleles in a selected context. Gene **205:** 291–299.

Griffiths, R. C., S. W. McKechnie and J. A. McKenzie, 1982 Multiple mating and sperm displacement in natural populations of *Drosophila melanogaster.* Proc. R. Soc. Lond. Ser. B Biol. Sci. **264:** 1271–1275.

Gromko, M. H., and D. W. Pyle, 1978 Sperm competition, male fitness, and repeated mating by female *Drosophila melanogaster.* Evolution **32:** 588–593.

Gromko, M. H., D. G. Gilbert and R. C. Richmond, 1984 Sperm transfer and use in the multiple mating system of Drosophila, pp. 372–427 in *Sperm Competition and the Evolution of Animal Mating Systems*, edited by R. L. Smith. Academic Press, New York.

Harshman, L. G., and A. G. Clark, 1998 Inference of sperm competition from broods of field-caught Drosophila. Evolution **52:** 1334–1341.

Harshman, L. G., and T. Prout, 1994 Sperm displacement without sperm transfer in *Drosophila melanogaster.* Evolution **48:** 758–766.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Holland, B., and W. R. Rice, 1999 Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. Proc. Natl. Acad. Sci. USA **96** (9): 5083–5088.

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 molecular 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.

Imhof, M., B. Harr, G. Brem and C. Schlötterer, 1998 Multiple mating in wild *Drosophila melanogaster* revisited by microsatellite analysis. Mol. Ecol. **7:** 915–917.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, UK.

Kreitman, M. E., 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.

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

McVean, G. A. T., and J. Vieira, 1999 The evolution of codon preferences in Drosophila: a maximum-likelihood approach to parameter estimation and hypothesis testing. J. Mol. Evol. **49:** 63–75.

Moriyama, E., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Neubaum, D. M., and M. F. Wolfner, 1999 Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics **153:** 845–857.

Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. USA **94:** 7784–7790.

Prout, T., and A. G. Clark, 1996 Polymorphism in genes that influence sperm displacement. Genetics **144:** 401–408.

Rice, W. R., 1996 Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature **381:** 232–234.

Richmond, R. C., D. G. Gilbert and K. B. Sheehan, 1980 Esterase 6 and reproduction in *Drosophila melanogaster.* Science **207:** 1483–1485.

Rozas, J., and R. Rozas, 1999 DnaSP 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Simmons, L. W., and M. T. Siva-Jothy, 1998 Sperm competition in insects: mechanisms and the potential for selection, pp. 341–434 in *Sperm Competition and Sexual Selection*, edited by T. R. Birkhead and A. P. Moller. Academic Press, New York.

Smith, R. L., 1984 *Sperm Competition and the Evolution of Animal Mating Systems.* Academic Press, New York.

Thomas, S., and R. S. Singh, 1992 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster.* VII. Varying rates of genic divergence as revealed by two-dimensional electrophoresis. Mol. Biol. Evol. **9:** 507–525.

Tram, U., and M. F. Wolfner, 1999 Male seminal fluid proteins are essential for sperm storage in *Drosophila melanogaster.* Genetics **153:** 837–844.

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

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

Wolfner, M. F., 1997 Tokens of love: functions and regulation of Drosophila male accessory gland products. Insect Biochem. Mol. Biol. **27:** 179–192.

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.

Wright, F., 1990 The "effective number of codons" used in a gene. Gene **87:** 23–29.

Communicating editor: W. Stephan