DOI: 10.1534/genetics.103.020883

Molecular Population Genetics of Male Accessory Gland Proteins in the *Drosophila simulans* Complex

Andrew D. Kern,[1] Corbin D. Jones and David J. Begun

Center for Population Biology, University of California, Davis, California 95616

Manuscript received August 4, 2003
Accepted for publication February 8, 2004

ABSTRACT

Accessory gland proteins are a major component of Drosophila seminal fluid. These proteins have a variety of functions and may be subject to sexual selection and/or antagonistic evolution between the sexes. Most population genetic data from these proteins are from *D. melanogaster* and *D. simulans*. Here, we extend the population genetic analysis of *Acp* genes to the other *simulans* complex species, *D. mauritiana* and *D. sechellia*. We sequenced population samples of seven *Acp*'s from *D. mauritiana, D. sechellia*, and *D. simulans*. We investigated the population genetics of these genes on individual *simulans* complex lineages and compared *Acp* polymorphism and divergence to polymorphism and divergence from a set of non-*Acp* loci in the same species. Polymorphism and divergence data from the *simulans* complex revealed little evidence for adaptive protein evolution at individual loci. However, we observed a dramatically inflated index of dispersion for amino acid substitutions in the *simulans* complex at *Acp* genes, but not at non-*Acp* genes. This pattern of episodic bursts of protein evolution in *Acp*'s provides the strongest evidence to date that the population genetic mechanisms driving *Acp* divergence are different from the mechanisms driving evolution at most Drosophila genes.

THE evolution of proteins involved in reproduction has attracted considerable interest. One generality emerging from this research is that reproduction-related genes tend to evolve quickly (reviewed in Swanson and Vacquier 2002). However, the importance of directional selection in driving this rapid protein evolution in reproduction-related genes is still unclear. In a few proteins, the ratio of nonsynonymous to synonymous substitution (d_N/d_S) is significantly greater than one, suggesting that the rapid evolution is a result of selection (Lee *et al.* 1995; Metz and Palumbi 1996; Tsaur and Wu 1997; Aguadé 1998, 1999; Ting *et al.* 1998, 2000; Wyckoff *et al.* 2000). In most cases, however, there is little support from d_N/d_S ratios that adaptive protein divergence is a general property of genes functioning in reproduction. Similarly, molecular population genetic analysis suggests the action of directional selection in some reproduction-related genes (Aguadé *et al.* 1992; Aguadé 1998, 1999; Nurminsky *et al.* 1998; Tsaur *et al.* 1998, 2001; Begun *et al.* 2000), but for most such genes there is no compelling evidence for adaptive protein divergence.

One particular group of reproduction-related proteins, the male accessory gland proteins (*Acp*'s) of Drosophila, has been the object of much research. Acp's, a major component of seminal fluid, have several demonstrated effects on female physiology and sperm use (Clark *et al.* 1995; Herndon and Wolfner 1995; Neubaum and Wolfner 1999; Tram and Wolfner 1999; reviewed in Chen 1996; Wolfner 1997). Natural variation of *Acp*'s was first investigated by protein electrophoresis, which revealed that accessory gland proteins are more polymorphic and evolve faster than proteins from many other nonreproductive male tissues (Coulthart and Singh 1988; Thomas and Singh 1992; Civetta and Singh 1995). These inferences from protein gels were subsequently supported by DNA sequence analyses, which revealed that *Acp*'s are much more polymorphic and diverge more quickly than "typical" proteins in the *Drosophila melanogaster vs. D. simulans* comparison (Begun *et al.* 2000; Swanson *et al.* 2001). Data from *Acp26Aa*, *Acp29AB*, and *Acp36DE* provided some evidence for adaptive protein divergence (Aguadé *et al.* 1992; Tsaur and Wu 1997; Aguadé 1998, 1999; Tsaur *et al.* 1998, 2001; Begun *et al.* 2000). Data from most *Acp*'s, however, were not suggestive of adaptive protein evolution in *D. melanogaster* and *D. simulans* (Begun *et al.* 2000).

Despite the relatively detailed study of *Acp*'s in *D. melanogaster* and *D. simulans*, little attention has been paid to *Acp* variation in the close relatives of these species, *D. mauritiana* and *D. sechellia*. Studies of *Acp*'s in these species allow us to address several key population genetic issues. Are patterns of polymorphism and divergence in *Acp*'s different from patterns in "non-*Acp*" genes in all *melanogaster* subgroup species? Are patterns of substitution in *Acp*'s similar across individual *simulans*

[1]*Corresponding author:* Center for Population Biology, University of California, 1 Shields Ave., Davis, CA 95616.
E-mail: adkern@ucdavis.edu

complex lineages? Does analysis of *Acp*'s along individual lineages provide evidence for adaptive protein evolution? Do some lineages experience more directional selection than others? Here, we address these questions through an analysis of variation in seven *Acp*'s in *D. mauritiana*, *D. sechellia*, and *D. simulans*.

MATERIALS AND METHODS

Stocks/loci: New *D. simulans* sequences reported here were collected from inbred lines derived from flies collected in the Wolfskill Orchard in Winters, California (Begun and Whitley 2000). *D. sechellia* sequences were from lines obtained from the species stock center and J. Coyne. Line sech01, also known as "Robertson," was collected in 1980 on Cousin Island by L. Tsacas. Line sech77 (a.k.a. s77 25x) was collected by K. Kimura on Praslin Island in 1987. Lines sech33, sech34, and sech35 were collected by J. W. O. Ballard on Mahe in 1998. All remaining lines were collected by J. R. David in 1985 on Cousin island. *D. mauritiana* sequences were from lines kindly provided by J. Coyne and D. Barbash. The *D. mauritiana* lines were originally collected in Mauritius in 1981 by O. Kitagawa. All lines were originally established from single, inseminated females and show very low levels of residual heterozygosity. In all cases, direct sequencing of PCR products in both directions was determined on an ABI 3700 sequencer. Raw data were analyzed using phred (Ewing *et al.* 1998) and phrap (Ewing and Green 1998). Sequences were then examined by eye using Consed (Gordon *et al.* 1998). Polymorphic sites in multiple alignments were verified using the MACE package of scripts (B. Gilliland, unpublished results; http://ludwig.ucdavis.edu/MACE/). New sequences reported in this article can be found in GenBank under accession nos. AY505178–AY505293.

Several analyses used previously published sequences. For *Acp26Aa* in *D. mauritiana*, 22 published sequences were used (Tsaur *et al.* 2001). Previously published *D. melanogaster* and *D. simulans Acp* sequences used in this analysis, with the exception of *D. simulans Acp26Aa, Acp26Ab,* and *Acp62F* (which were collected for this paper), were from Begun *et al.* (2000) and references therein. Comparisons between *Acp*'s and non-*Acp*'s in *D. sechellia* and *D. mauritiana* were made using sequences from Hey and Kliman (1993), Kliman *et al.* (2000), Parsch *et al.* (2001), and references therein, as well as new sequence data from *Relish* (AY505188–AY505192; see supplementary table at http://www.genetics.org/supplemental/ for polymorphism data). This non-*Acp* data set is composed of 11 loci, *Adh*, *ase*, *ci*, *est-6*, *janA*, *janB*, *per*, *Rel*, *yp2*, *z*, and *Zw*, surveyed for variation in *simulans* complex species. For polymorphism comparisons across species, *ase* and *ci* were excluded as they are from regions of low recombination in *D. melanogaster*. Similarly, *janA* was omitted for comparisons of replacement and silent variation across species due to a lack of polymorphism data in *D. sechellia*. All analyses were performed on gene regions for which homologous sequence was available from the *simulans* complex species as well as for *D. melanogaster*. Note that the polymorphism statistics reported in Begun *et al.* (2000) for *D. simulans Acp62F* disagree with those reported here, which are from a separate population sample. We use the new data in this article. Additionally our analysis of silent and replacement mutations included sites for which a base could not be called in every sequence sampled, leading to some small differences in mutation counts between the present report and Begun *et al.* (2000).

Expression analysis: We used reverse transcription-PCR (RT-PCR) on cDNA from whole adult males and whole adult females to validate that putative *Acp*'s were male limited in expression in the *simulans* complex. Gene-specific primers were used in all cases (sequences available from the authors). Poly(A)+ RNA was prepared from whole flies using a Micro-Poly(A) kit (Ambion, Austin, TX). cDNA for reverse transcriptase-PCR and rapid amplification of cDNA ends (RACE) was prepared from this RNA using the SMART RACE cDNA amplification kit (CLONTECH, Palo Alto, CA). SuperScript II reverse transcriptase (GIBCO BRL, Rockville, MD) was used for all RT reactions. In all cases we found that *simulans* complex homologs of *D. melanogaster Acp*'s showed male-limited expression. This demonstrates that genes that are *Acp*'s in *D. melanogaster* show similar patterns of expression in *simulans* clade species.

Sequence analysis: Most population genetic analyses were performed using software developed by the authors. Source code in Smalltalk is available upon request. Throughout the article, "silent sites" refers to synonymous coding positions in exons. Polarized mutations are those changes that could be unambiguously assigned to individual lineages of the *simulans* complex under parsimony (using *D. melanogaster* as an outgroup). Divergence estimates were calculated using the maximum-likelihood method of Goldman and Yang (1994) as implemented in PAML (Yang 1997). For maximum-likelihood (ML) divergence estimation we allowed equilibrium codon frequencies to be free parameters of the model. In addition we estimated the transition/transversion ratio for each locus. Estimated substitution rates along individual lineages of the *simulans* complex and indices of dispersion [$R(t)$] for individual loci were calculated in a manner adapted from Gillespie (1989). Gillespie's (1989) analysis uses sequence comparisons of one allele from each of three species for each of several loci and then corrects for lineage effects by weighting each lineage by the mean divergence along that lineage. Explicitly, let N_i, $i = 1, 2, 3$, be the number of substitutions at a particular locus on the ith lineage. These N_i can be estimated using estimates of pairwise divergence between all possible pairs of species such that

$$N_1 = (D_{12} + D_{13} - D_{23})/2$$
$$N_2 = (D_{12} + D_{23} - D_{13})/2$$
$$N_3 = (D_{13} + D_{23} - D_{12})/2,$$

where D_{ij} is the estimate of divergence between species i and j. The N_i represent random variables drawn from at most three different distributions. As such, the moments of N_i can be written as

$$E\{N_i\} = w_i\mu$$
$$\mathrm{Var}\{N_i\} = w_i\sigma^2.$$

Here, w_i is the weighting factor that will be shared by all loci on a particular lineage. μ and σ^2 are the locus-specific contributions to the mean and variance in the number of substitutions. The w_i are defined such that

$$\frac{1}{3}\sum_{i=1}^{3} w_i = 1.$$

This correction adjusts the mean number of substitutions along a lineage without changing the average rate of evolution at a particular locus. In practice, lineage-specific weights are calculated by choosing values such that the mean number of substitutions per 100 sites is equal among lineages. We are now in a position to write down the index of dispersion in a manner that is free of lineage effects:

$$R(t) = \frac{\mathrm{Var}\{N_i\}}{E\{N_i\}} = \frac{w_i\sigma^2}{w_i\mu} = \frac{\sigma^2}{\mu}.$$

Lineage effects may be removed by dividing the number of substitutions along the ith lineage by the weight w_i. Thus the mean number of substitutions for a particular locus will be estimated as

$$M = \frac{1}{3}\sum_{i=1}^{3}\frac{N_i}{w_i}.$$

The expectation of this estimator is then

$$E\{M\} = \frac{1}{3}\sum_{i=1}^{3}\frac{E\{N_i\}}{w_i} = \mu,$$

which is an unbiased estimator of μ. The variance may be estimated in a straightforward manner by

$$S^2 = \frac{9}{2}\left(\frac{1}{\sum_{i=1}^{3}1/w_i}\right)\left(\frac{1}{3}\sum_{i=1}^{3}(N_i/w_i)^2 - M^2\right).$$

This estimator of variance is also unbiased. From the estimators of the mean and variance in the numbers of substitutions one can write an estimator for $R(t)$,

$$R = \frac{S^2}{M}.$$

As we had polymorphism data for multiple loci, we developed a computational procedure to use this information in our analyses of $R(t)$. Briefly, we iterated Gillespie's routine 1000 times, where each iteration consisted of choosing one allele at random from each species at every locus examined, estimating divergences, calculating lineage-specific weights, and then estimating $R(t)$ for each locus. From each iteration we recorded the values of $R(t)$, thereby generating an empirical distribution of the test statistic $R(t)$ given different configurations of the data. These distributions allow us to make statistical inferences about the index of dispersion at individual loci and across classes of loci. *Zeste* was omitted from this analysis because there were no interspecific differences. Tests for deviations of $R(t)$ from neutral, equilibrium expectations were determined through simulations of the substitution process. We assumed a Poisson molecular clock and conditioned on the mean number of estimated substitutions at a locus across our iterations. For each locus, 10^6 independent replicates of this procedure were performed. Negative branch lengths were set to zero for subsequent analyses.

Contingency table analyses for McDonald-Kreitman (MK; McDonald and Kreitman 1991) tests were done using 10^5 replicates of a Monte Carlo procedure (Engels 1988) because numerous cells contained small values or values of zero. A resampling procedure to detect small-scale hitchhiking effects associated with polarized fixations was implemented according to Kern *et al.* [2002; Kern-Jones-Begun (KJB) test].

Simulation of sperm competition: Clark (2002) proposed a scramble competition model of sperm displacement. We used computer simulations of this model to shed light on patterns of molecular evolution that might be expected for genes functioning in sperm competition. In this haploid model, the fitness of an allele is defined by its ability to displace other genotypes after competitively mating. Consider two alleles, x_i and x_j, which mate in succession such that x_j mates first, followed by x_i. Let s_{ij} represent the proportion of offspring sired by x_i (also known as the P_2 score of x_i), where $s_{ij} = \frac{1}{2}$ for $(i = j)$. Assuming random mating, the frequency of x_i after a generation of selection is

$$x_i' = \frac{\sum_j x_i x_j s_{ij}}{\sum_i \sum_j x_i x_j s_{ij}}.$$

The distribution of s_{ij} is an important parameter because distributions with greater variances will produce weaker frequency dependence. Fortunately several experiments provide some information about the variation in P_2 scores (s_{ij} values) in *D. melanogaster* (Clark *et al.* 1995). P_2 has a modal value near 1.0 and a tail of smaller values that disappears at ~0.5 in laboratory experiments. We simulated Clark's model using three different distributions of s_{ij} values: a Uniform (0, 1), which is probably inappropriate, a Uniform (0.5, 1), which is slightly closer to empirical data, and a Beta (5, 1), which is likely to be the most biologically relevant distribution examined here. These simulations used an infinite-site, no-recombination representation of a single gene (Watterson 1975). Allele frequencies change each generation according to the effects of selection, drift, and mutation, in that order. The procedure for specifying allelic fitness means that many, but certainly not all, new mutations will have different fitness values from that of their parent allele. Further, allelic fitness can change purely as a function of the other alleles segregating in the population at any point during the simulation. Thus, the mutation parameter reflects both mutation rates to neutral and selected alleles. We employed a simple allelic genealogy (Takahata 1990) to maintain a tree of alleles in the population at any time. This structure keeps track of the relationships among alleles, the generation in which individual alleles arose by mutation, and, for alleles that ultimately fix, their fixation times. Thus, the allelic genealogy provides statistics of polymorphism and divergence.

RESULTS

Polymorphism: Table 1 gives summary statistics of DNA polymorphism from seven *Acp*'s from *D. sechellia*, *D. simulans*, and *D. mauritiana*. *Acp*'s harbor more replacement polymorphism than non-*Acp*'s in all species. In *D. mauritiana*, average silent heterozygosity is not significantly greater for *Acp*'s than for non-*Acp*'s. However, *D. mauritiana Acp*'s are significantly more polymorphic at replacement sites than are non-*Acp*'s (Mann-Whitney *U*-test: $P = 0.0017$ for θ_w; $P = 0.0013$ for θ_π). Surprisingly, one allele in our population sample from *D. mauritiana* (mau205) contained a premature stop codon at *Acp29*. This premature termination codon was a change from a lysine at residue 110 (base 325 in our alignment) of our sampled region, which is about two-thirds of the length of the full *Acp29* protein and was supported by a phred quality score of 90 (*i.e.*, the probability this call is in error is 10^{-9}). No other polymorphisms occurred on this allele downstream of this premature termination codon. *D. sechellia Acp* replacement polymorphism is almost eightfold higher than non-*Acp*'s replacement polymorphism, although this difference is not statistically significant (Mann-Whitney *U*-test: $P = 0.0851$ for θ_w; $P = 0.0851$ for θ_π). Finally, as previously reported (Begun *et al.* 2000), *D. simulans Acp* replacement polymorphism is much greater than replacement polymorphism at non-*Acp*'s (Mann-Whitney *U*-test: $P = 0.0013$ for θ_w; $P = 0.0013$ for θ_π).

Divergence: Consistent with previous observations (Begun *et al.* 2000; Swanson *et al.* 2001), *Acp*'s show high replacement substitution rates. Table 1 presents maximum-likelihood estimates (Goldman and Yang 1994) of silent and replacement divergence of *D. sim-*

TABLE 1

Summary statistics of DNA polymorphism for seven *Acp*'s from *D. mauritiana*, *D. sechellia*, and *D. simulans*

Species	Locus	n	S	Sil. θ_W	Sil. θ_π	Repl. θ_W	Repl. θ_π	Sil. div.	Repl. div.	Sil. Taj. D	Rep. Taj. D
D. mauritiana	*Acp26Aa*	22	58	0.0472	0.0309	0.0132	0.0095	0.1435	0.1901	−1.3251	−1.2367
	Acp26Ab	6	14	0.0242	0.0184	0.0142	0.0108	0.1401	0.0337	−1.2331	−1.3903
	Acp29AB	6	15	0.0255	0.0255	0.0142	0.0125	0.1719	0.0967	NA	−0.8945
	Acp36DE	6	40	0.0159	0.0167	0.0094	0.0098	0.1147	0.0622	0.3292	0.2360
	Acp53Ea	6	14	0.0169	0.0189	0.0036	0.0027	0.1851	0.0175	0.6003	NA
	Acp62F	6	3	0.0181	0.0248	0.0030	0.0036	0.1848	0.0367	NA	NA
	Acp76A	6	47	0.0587	0.0642	0.0065	0.0062	0.2486	0.0291	0.5920	−0.2731
D. sechellia	*Acp26Aa*	10	12	0.0022	0.0044	0.0032	0.0029	0.1272	0.1824	NA	−1.6360
	Acp26Ab	9	3	0.0000	0.0000	0.0034	0.0036	0.2214	0.0237	NA	NA
	Acp29AB	12	2	0.0035	0.0052	0.0000	0.0000	0.2134	0.0881	NA	NA
	Acp36DE	11	17	0.0048	0.0043	0.0033	0.0021	0.1051	0.0580	−0.4036	−1.3448
	Acp53Ea	12	1	0.0000	0.0000	0.0000	0.0000	0.1568	0.0178	NA	NA
	Acp62F	12	0	0.0000	0.0000	0.0000	0.0000	0.1599	0.0467	NA	NA
	Acp76A	12	17	0.0049	0.0062	0.0026	0.0019	0.2044	0.0317	0.8722	−0.4478
D. simulans	*Acp26Aa*	8	25	0.0233	0.0246	0.0055	0.0062	0.1305	0.1785	0.2191	0.0922
	Acp26Ab	8	15	0.0578	0.0743	0.0054	0.0066	0.1954	0.0347	1.3775	0.9657
	Acp29AB	8	9	0.0245	0.0235	0.0034	0.0050	0.2440	0.0829	−0.2010	1.8548
	Acp36DE	6	40	0.0250	0.0266	0.0065	0.0064	0.1045	0.0681	0.4104	−0.1407
	Acp53Ea	8	12	0.0064	0.0064	0.0062	0.0055	0.1758	0.0363	NA	−0.5247
	Acp62F	8	5	0.0271	0.0218	0.0057	0.0061	0.2660	0.0176	−0.8125	NA
	Acp76A	6	29	0.0469	0.0413	0.0031	0.0023	0.1130	0.0212	−0.7240	−1.3370

Species and Locus are the species and loci surveyed; n is the number of alleles sampled; S is the number of segregating sites in this sample, including those sites that contain ambiguous bases; Sil. θ_W and Sil. θ_π are estimates of nucleotide heterozygosity at silent sites, and sites containing ambiguous bases have been removed for these estimates; Repl. θ_W and Repl. θ_π are estimates of nucleotide heterozygosity at replacement sites and sites containing ambiguous bases have been removed for these estimates; Sil. div. and Repl. div. are estimates of nucleotide divergence at silent and replacement sites, respectively; and Sil. Taj. D and Repl. Taj. D are Tajima's measure of the site frequency spectrum at silent and replacement sites, respectively.

ulans, *D. mauritiana*, and *D. sechellia vs. D. melanogaster*. Silent divergence between *D. melanogaster* and each of the three *simulans* complex species is not significantly different for *Acp*'s than for non-*Acp*'s. However, rates of amino acid evolution are significantly greater from *D. melanogaster* to *D. mauritiana*, *D. simulans*, and *D. sechellia* (*D. mauritiana*, Mann-Whitney *U*-test, $P = 0.0049$; *D. simulans*, Mann-Whitney *U*-test, $P = 0.0083$; *D. sechellia*, Mann-Whitney *U*-test, $P = 0.011$). This effect is approximately threefold for each species.

Divergences of *D. mauritiana*, *D. sechellia*, and *D. simulans vs. D. melanogaster* are nonindependent as a result of the shared evolutionary history of these species. We used unrooted three-taxon trees to investigate substitution rates along recently separated, individual *simulans*-complex lineages (Table 2). Replacement divergence since the speciation event(s) is significantly greater at *Acp*'s than at non-*Acp*'s for all lineages (*D. mauritiana*, Mann-Whitney *U*-test, $P = 0.0063$; *D. simulans*, Mann-Whitney *U*-test, $P = 0.0027$; *D. sechellia*, Mann-Whitney *U*-test, $P = 0.0192$). Silent divergence, however, is comparable between *Acp*'s and non-*Acp*'s. Thus, rapid rates of *Acp* protein evolution are characteristic of both the more recent and more ancient histories of the *melanogaster* subgroup.

In addition to characterizing mean replacement divergence among *Acp*'s and non-*Acp*'s in the *simulans* complex, we estimated the index of dispersion [$R(t)$] for replacement sites (Table 3). The mean $R(t)$ across loci, across iterations of our computational procedure (see MATERIALS AND METHODS) for all *Acp*'s is significantly overdispersed [$R(t) = 6.50$; $P < 0.001$] as are values of $R(t)$ for several loci individually. Non-*Acp*'s do not reject a Poisson clock [mean $R(t) = 2.56$]. Additionally estimates of $R(t)$ are significantly greater for *Acp*'s than for non-*Acp*'s using a nonparametric test (Mann-Whitney *U*-test: $P = 0.0096$). It is worth noting that the only two non-*Acp* loci that showed significant overdispersion, *janA* and *janB*, are both reproductive proteins.

A potential caveat associated with these $R(t)$ analyses is that the *simulans* complex species have only recently diverged and thus may segregate ancestral polymorphism (*e.g.*, KLIMAN *et al.* 2000). For example, although sequence comparisons of a single randomly selected *Acp* allele from each of the three *simulans* complex species may reveal a number of amino acid differences, few of these differences will be fixed differences between species. This stands in contrast to typical analyses of $R(t)$, which are applied to situations in which polymorphism would be expected to have negligible effects on

TABLE 2

Lineage-specific divergence for *Acp* genes and non-*Acp* genes from *D. mauritiana*, *D. sechellia*, and *D. simulans*

	Silent divergence			Replacement divergence		
Locus	*D. simulans*	*D. mauritiana*	*D. sechellia*	*D. simulans*	*D. mauritiana*	*D. sechellia*
Acp genes						
Acp26Aa	0.0000	0.0178	0.0124	0.0048	0.0200	0.0094
Acp26Ab	0.0423	0.0492	0.0974	0.0056	0.0040	0.0150
Acp29	0.0421	0.0306	0.0443	0.0060	0.0161	0.0070
Acp36	0.0164	0.0187	0.0136	0.0171	0.0146	0.0042
Acp53	0.0000	0.0288	0.0176	0.0134	0.0045	0.0039
Acp76A	0.0894	0.0332	0.0685	0.0298	0.0034	0.0056
Acp62F	0.0333	0.0378	0.0000	0.0104	0.0076	0.0185
Non-*Acp* genes						
Adh	0.0149	0.0544	0.0486	0.0149	0.0544	0.0486
janA	0.0838	0.0000	0.0751	0.0000	0.0039	0.0002
janB	0.0001	0.0000	0.0001	0.0074	0.0000	0.0114
ase	0.0003	0.0063	0.0020	0.0000	0.0014	0.0017
ci	0.0012	0.0182	0.0078	0.0020	0.0005	0.0007
est6	0.0314	0.0173	0.0464	0.0030	0.0067	0.0070
per	0.0472	0.0937	0.0763	0.0007	0.0000	0.0014
rel	0.0474	0.0000	0.0183	0.0001	0.0000	0.0001
yp2	0.0122	0.0105	0.0098	0.0000	0.0023	0.0012
z	0.0075	0.0077	0.0075	0.0000	0.0000	0.0000
Zw	0.0380	0.0000	0.0565	0.0038	0.0000	0.0010

estimates of substitution rates (*e.g.*, Gillespie 1989). Further, the method used to correct for lineage effects in this analysis weights each gene along a lineage equally, thus ignoring any lineage-by-population size contributions to the variance in standing levels of polymorphism among loci. There are three reasons to think that polymorphism will not bias either our rejection of the neutral model for *Acp*'s or our *Acp*'s *vs.* non-*Acp*'s comparison. First, under the strictly neutral model, the expectation that $R(t) = 1$ should not depend on the particular time during the substitution process that the populations are sampled. That is, the number of mutations occurring on each lineage is Poisson distributed regardless of the "length" of that lineage. Second, our analysis shows that the inflated $R(t)$ for *Acp*'s is not dependent on the particular alleles sampled. Finally, and perhaps most importantly, both *Acp*'s and non-*Acp*'s are segregating ancestral polymorphism, yet only *Acp*'s show an inflated $R(t)$.

Model of sperm competition: The highly significant replacement $R(t)$ rejects the simple neutral model for *Acp* protein evolution. However, it remains unclear which models of evolution might fit the data. Several models with different flavors of selection produce an overdispersed molecular clock (*e.g.*, Gillespie 1993, 1994a,b; Cutler 2000). Clark (2002) proposed a scramble competition model for sperm competition, but did not characterize the properties of this model in a stochastic framework. Here we present a preliminary analysis of whether scramble competition can produce values of $R(t)$ as high as those observed for *Acp*'s. Figure 1 shows $R(t)$ as a function of $4Nu$ (where N is held constant at 10^6 and u is allowed to vary) for three different distributions of s_{ij} (see materials and methods). It is clear that $R(t) \geq 1$ for only a small portion of the parameter space for each distribution of s_{ij}. To inform our expectations about actual data, we drew samples of size $n = 8$ (the typical sample size for our *D. simulans* data) from simulated populations at independent intervals and then recorded polymorphism statistics. This procedure allows us to determine the approximate values of $4Nu$, which, under scramble competition, give the average number of segregating sites observed among *D. simulans Acp*'s (assuming $N = 10^6$). For the Beta(5, 1) case, this value corresponded to $4Nu \approx 1$. For the Uniform(0.5, 1) case, this value was $4Nu \approx 3$. Notably, the Uniform(0, 1) did not produce the high levels of polymorphism observed within the parameter space examined here. Thus, irrespective of the distribution of selection coefficients, observed values of $R(t)$ appear to be incompatible with Clark's scramble competition model.

Frequency distribution: We used Tajima's (1989) D to test for deviations of the site frequency distribution from the neutral equilibrium expectation (Table 1) at silent and replacement sites independently. Statistical significance was assessed using 10^5 standard coalescent simulations conditioning on the observed number of segregating sites. D values were nonsignificant in all cases.

Polymorphic *vs.* fixed mutations: Tables 4 and 5 show

TABLE 3

Index of dispersion for replacement divergence among the *simulans* complex species

Locus	Mean $R(t)$
Acp26Aa	7.73**
Acp26Ab	9.43**
Acp29AB	2.39
Acp36DE	5.40*
Acp53Ea	2.19
Acp62F	11.14***
Acp76A	7.25**
	Average = 6.50**
Adh	2.55
janA	4.6*
janB	12.30***
ase	1.01
ci	0.29
est6	1.21
per	0.60
rel	0.96
yp2	0.87
Zw	1.22
	Average = 2.56

Mean $R(t)$ is the mean value of the index of dispersion over 1000 iterations. The average is the mean of means across all loci and iterations (see MATERIALS AND METHODS for details). *$P < 0.01$; **$P < 0.001$; ***$P < 0.0001$.

results of several MK tests. Unpolarized MK tests from each of the *simulans* complex species to *D. melanogaster* (Table 4) are nonsignificant, with the exception of *Acp26Aa* in the *D. mauritiana* and *D. simulans* lineages (the MK test of *Acp26Aa* variation in *D. sechellia* is not significant after a conservative Bonferroni correction for multiple tests although the data show the same pattern observed in the other species). These results, as well as previously reported significant deviations for *Acp26Aa*, which were obtained without polymorphism data from *D. simulans* or *D. sechellia* (AGUADÉ 1998; TSAUR *et al.* 1998), are consistent with an excess of amino acid fixations at this locus. Polarized MK tests (Table 5) provide no support for adaptive evolution at individual loci after speciation of the three sister taxa, with the notable exception of *D. simulans Acp76A*, which is significant even after corrections for multiple tests. Interestingly, this pattern is not nearly as strong in the unpolarized comparison to *D. melanogaster*. Thus, it may represent a lineage-specific phenomenon in *D. simulans*.

Table 6 shows polarized *Acp* mutations summed across loci for each *simulans* complex lineage. Two patterns are of note. First, if we restrict our attention to the two highly variable species *D. simulans* and *D. mauritiana*, *D. mauritiana* is much more polymorphic than *D. simulans* at replacement sites. Second, the contingency table for *D. sechellia* is individually significant ($P = 0.037$), most plausibly as the result of an excess of silent fixations (see below).

Table 7 shows a contingency table of polarized silent mutations categorized as preferred (a mutation from an unpreferred to a preferred codon), unpreferred (a mutation from a preferred codon to unpreferred codon), and no change (mutations from preferred to preferred or from unpreferred to unpreferred) mutations. No single-locus contingency table is significantly heterogeneous, nor is the contingency table of pooled mutations significantly heterogeneous for any species. Never-

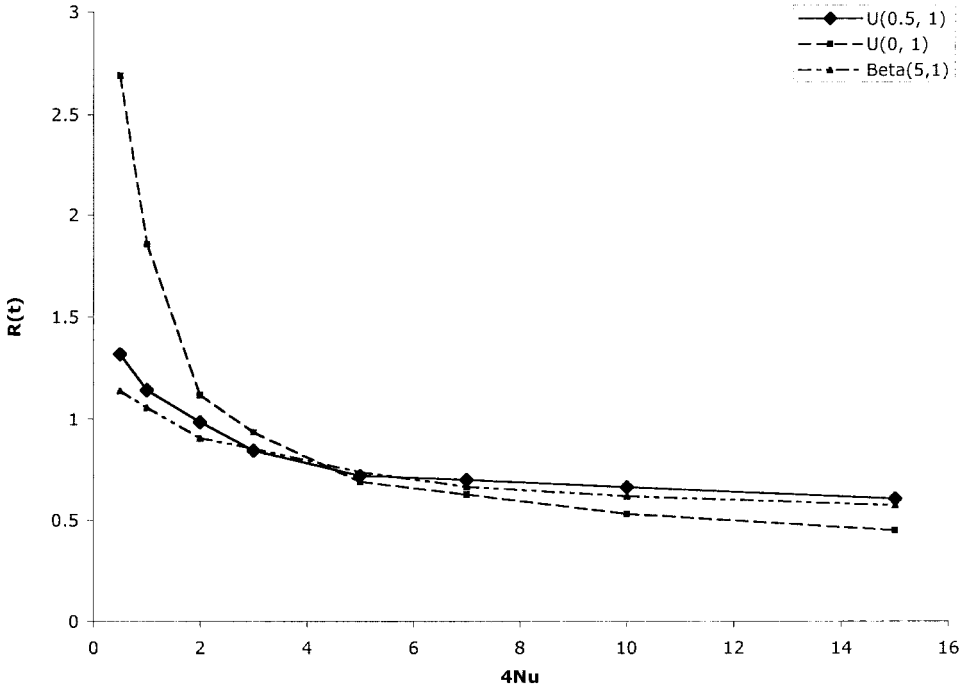

FIGURE 1.—Simulation results from Clark's scramble model. Plotted is the index of dispersion $R(t)$ vs. $4Nu$, where $N = 10^6$ and u is allowed to vary. For each simulation $N = 10^6$ haploid individuals. For each parameter set, the simulation was burnt in for the first 2000 fixations, and then results were recorded for the next 100,000 substitutions. $R(t)$ was estimated from C_0 in the simulation (see GILLESPIE 1993).

TABLE 4

Unpolarized McDonald-Kreitman tests between *D. melanogaster* and each of the *simulans* clade species

	FAA	PAA	FS	PS	*P* value
D. mauritiana					
Acp26Aa	69	43	23	44	0.00068*
Acp26Ab	6	9	4	9	0.710
Acp29AB	20	12	19	8	0.589
Acp36DE	48	39	37	31	0.900
Acp53Ea	4	8	9	12	0.720
Acp62F	5	1	5	7	0.152
Acp76A	22	11	28	29	0.127
D. sechellia					
Acp26Aa	76	23	26	21	0.0159*
Acp26Ab	5	4	8	6	0.900
Acp29AB	22	2	21	4	0.667
Acp36DE	50	40	33	34	0.515
Acp53Ea	4	5	8	9	0.900
Acp62F	5	0	6	5	0.120
Acp76A	22	7	31	5	0.345
D. simulans					
Acp26Aa	74	28	24	29	0.002*
Acp26Ab	6	5	3	14	0.094
Acp29AB	22	5	21	9	0.360
Acp36DE	58	32	34	36	0.054
Acp53Ea	7	10	10	11	0.752
Acp62F	2	2	6	8	0.990
Acp76A	11	5	14	23	0.069

FAA and PAA are the number of fixed and polymorphic amino acid changes, respectively. FS and PS are the number of fixed and polymorphic silent changes, respectively. *P* values were determined using 10^5 replicates of a Monte Carlo procedure. *$P < 0.05$.

TABLE 5

Polarized McDonald-Kreitman tests for each of the *simulans* complex species

	FAA	PAA	FS	PS	*P* value
D. mauritiana					
Acp26Aa	0	18	0	21	0.999
Acp26Ab	0	6	1	2	0.33
Acp29AB	1	8	0	2	0.99
Acp36DE	6	21	2	10	0.99
Acp53Ea	0	1	0	2	0.99
Acp62F	0	1	0	1	0.99
Acp76A	1	7	1	10	0.99
D. sechellia					
Acp26Aa	2	4	0	2	0.57
Acp26Ab	1	2	3	0	0.4
Acp29AB	1	0	2	1	0.99
Acp36DE	2	8	5	3	0.14
Acp53Ea	0	0	0	0	0.99
Acp62F	1	0	0	0	0.99
Acp76A	1	2	4	1	0.45
D. simulans					
Acp26Aa	1	3	0	5	0.45
Acp26Ab	0	3	1	5	0.99
Acp29AB	2	2	2	3	0.99
Acp36DE	6	15	4	10	0.99
Acp53Ea	0	4	0	2	0.99
Acp62F	0	1	0	3	0.99
Acp76A	6	2	0	9	0.0025*

See Table 4 for explanation.

theless, there is a strong lineage effect for *D. sechellia* silent sites. In contrast to *D. simulans* and *D. mauritiana*, which have fixed approximately equal numbers of preferred and unpreferred mutations, *D. sechellia* has fixed significantly more unpreferred codons than preferred codons (10:4; binomial test, $P = 0.028$). This supports the conclusion from Kliman *et al.* (2000) that *D. sechellia* shows a genomic trend toward reduced codon bias.

DISCUSSION

An interesting new result from our analyses is that the index of dispersion [$R(t)$] for *Acp*'s, in contrast to that for non-*Acp*'s, is overdispersed. While $R(t)$ has been shown to have limitations (Goldman 1994; Nielsen 1997), our results strengthen the idea that *Acp* protein evolution is generally incompatible with simple neutrality (Table 3). Overdispersion does not appear to be a general property of Drosophila proteins (this report; Zeng *et al.* 1998). Thus the evolution of the amino acid sequences of *Acp* proteins requires a special explanation. The overdispersed $R(t)$ for *Acp*'s rejects not only most neutral models of molecular evolution (but see Takahata 1987), but also many models of selection. For example, recurrent genetic hitchhiking models (such as the normal shift model of Gillespie 1994b) produce $R(t) < 1$. Overdominance can also produce $R(t) < 1$, as can random environment models, such as the SAS-CFF model (Gillespie 1978) with a low environmental autocorrelation. However, random environment models with fluctuations on the timescale of molecular evolution can produce $R(t) > 1$ (Gillespie 1993), as can the house of cards model (Ohta and Tachida 1990; Gillespie 1994b) and other deleterious mutation models (Cutler 2000).

Clark (2002) proposed a simple model of scramble

TABLE 6

Polarized mutations summed across loci for individual lineages of the *simulans* complex species

	FAA	PAA	FS	PS	*P* value
D. mauritiana	8	62	4	48	0.550
D. sechellia	8	16	14	7	0.037
D. simulans	15	30	7	37	0.084

See Table 4 for explanation.

TABLE 7

Polarized silent fixations and polymorphisms

	Fixed *P*	Poly. *P*	FU	PU	FNC	PNC	*P* value
D. mauritiana							
Acp26Aa	0	5	0	7	0	9	0.99
Acp26Ab	1	1	0	1	0	0	0.99
Acp29AB	0	1	0	0	0	1	0.99
Acp36DE	1	3	1	1	0	6	0.22
Acp53Ea	0	1	0	0	0	1	0.99
Acp62F	0	0	0	1	0	0	0.99
Acp76A	0	2	0	3	1	4	0.99
Total	2	13	1	13	1	21	0.80
D. sechellia							
Acp26Aa	0	0	0	2	0	0	0.99
Acp26Ab	0	0	3	0	0	0	0.99
Acp29AB	0	0	2	1	0	0	0.99
Acp36DE	3	1	2	1	0	1	0.67
Acp53Ea	0	0	0	0	0	0	0.99
Acp62F	0	0	0	0	0	0	0.99
Acp76A	1	0	3	1	0	0	0.99
Total	4	1	10	5	0	1	0.29
D. simulans							
Acp26Aa	0	3	0	1	0	1	0.99
Acp26Ab	0	1	0	3	1	1	0.99
Acp29AB	0	2	1	0	1	1	0.60
Acp36DE	2	4	0	2	2	4	0.64
Acp53Ea	0	2	0	0	0	0	0.99
Acp62F	0	1	0	1	0	1	0.99
Acp76A	0	1	0	3	0	5	0.99
Total	2	14	1	10	4	13	0.65

FP and PP are the numbers of silent changes from an unpreferred codon to a preferred codon that are fixed and polymorphic, respectively. FU and PU are the numbers of changes from a preferred codon to an unpreferred codon that are fixed and polymorphic, respectively. FNC and PNC are the numbers of changes that did not change the category of codon (*e.g.*, changes from preferred to preferred or from unpreferred to unpreferred) that are fixed and polymorphic, respectively. *P* values were determined using 10^5 replicates of a Monte Carlo procedure. Note that in no case was there significant heterogeneity among the classes of codon fixations.

competition to describe the molecular evolution of *Acp*'s. Given that this model has a balancing component in the absence of fluctuating parameters, we might expect $R(t) < 1$ (see Cutler 2000). Indeed, simulation results (Figure 1) of the scramble competition model clearly demonstrate that $R(t) \leq 1$ for most of the parameter space examined. Interestingly, in these simulations $R(t)$ decreases with increasing mutation rate. An intuitive explanation for this result is that the strength of the balancing component (*i.e.*, frequency-dependent selection) in the scramble competition model increases with the number of unique alleles in the population. Thus, scramble competition cannot explain the observed patterns of divergence for the levels of heterozygosity observed in our samples. Nor can the *Acp* data be explained by shift models of hitchhiking (*e.g.*, Gillespie 1997). Instead, deleterious allele models and random environment models with high environmental autocorrelations seem to provide a better fit to the observations. Overdispersed molecular clocks are a signature of episodic bursts of substitution. If *Acp* evolution is driven by sexual conflict (*e.g.*, Rice 1996; Holland and Rice 1999), perhaps these data suggest that there are periods of escalated conflict interspersed with periods of relative quiescence.

Despite the evidence for nonneutral evolution from estimates of $R(t)$, neither contrasts of polymorphism and divergence nor analyses of the frequency spectrum show convincing evidence of selection on *Acp*'s. Certainly there is little evidence for adaptive evolution at individual loci. For example, no gene/species examined here deviates from the neutral site frequency spectrum (Table 1). Furthermore, most genes/species are consistent with the neutral model in McDonald-Kreitman tests. There are, however, exceptions. The unpolarized test of *Acp26Aa* between *D. melanogaster* and each *simulans* complex species rejects neutrality in the direction of an excess of amino acid fixations (this report; Aguadé

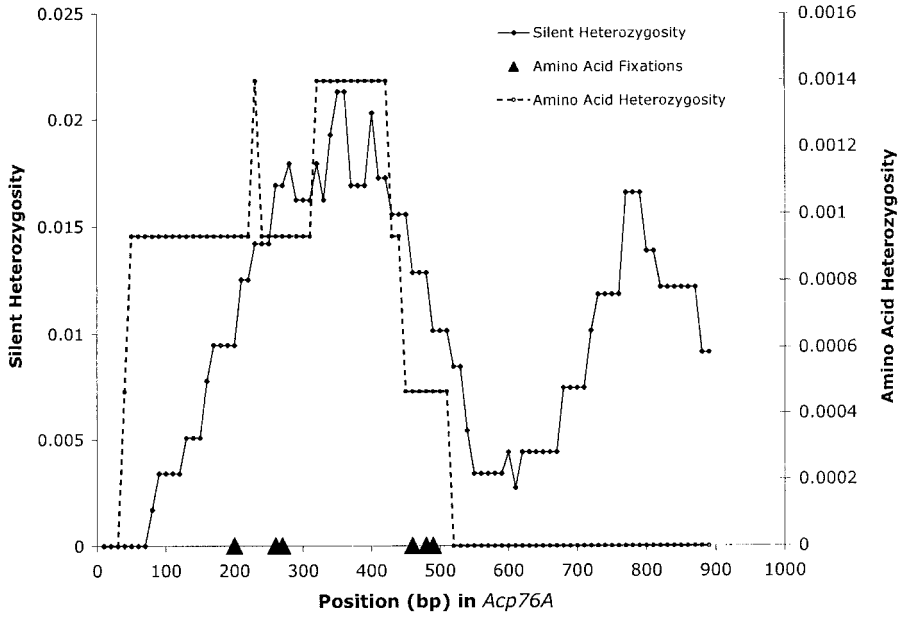

FIGURE 2.—All six replacement fixations in *D. simulans Acp76A* occurred in a region that is highly polymorphic at silent and replacement sites. The heterozygosity statistic is θ_π. Position refers to the location within the sampled region of *Acp76A*. This sampled region includes almost all of this gene. The window size shown is 200 bp with a displacement of 10 bp. There are 26 segregating sites within this coding region.

et al. 1992; AGUADÉ 1998; TSAUR and WU 1997; TSAUR *et al.* 1998, 2001). AGUADÉ (1999) reported evidence for directional selection in *Acp29AB* in an unpolarized comparison between *D. melanogaster* and *D. simulans*. Data from the slightly smaller region of the gene surveyed here do not deviate from neutrality. Furthermore, polarized analyses of polymorphism and divergence provide little evidence for adaptive protein divergence between *simulans* complex species. Even *Acp26Aa* data are compatible with neutrality. This contrast between the results for polarized and unpolarized MK tests at *Acp26Aa* is consistent with a burst of adaptive protein evolution on the lineage connecting the most recent common ancestor of the *simulans* complex to *D. melanogaster*. However, further investigation of this hypothesis is problematic because of uncertain alignments with the outgroup, *D. yakuba*.

A notable result from analyses of polymorphism and divergence at the single-locus level comes from *Acp76A*. A polarized MK test rejects the neutral model in a manner consistent with an excess of amino acid fixations in *D. simulans*. If amino acid mutations fix under directional selection, one might predict that heterozygosity would be reduced near such fixations (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; STEPHAN *et al.* 1992; GILLESPIE 1997, 2000). We used a KJB test (KERN *et al.* 2002) to test this prediction. This test uses a resampling method to determine whether levels of variation near amino acid fixations (or any defined fixation type) are inconsistent with the average level of variation in a gene. Surprisingly, *Acp76A* amino acid fixations are associated with a significant *excess* of polymorphism (two-tailed test and a 200-bp window; $P = 0.025$ for θ_w; $P = 0.013$ for θ_π) rather than with reduced polymorphism. In fact, all six fixations occurred in a region of *Acp76A* that is highly polymorphic at silent and replacement sites (Figure 2). Thus, two patterns require explanation. First, there is an apparent excess of amino acid fixations. Second, there is a nonrandom physical organization of polymorphism—the region of the protein that presumably shows an excess of amino acid fixations has more polymorphism than other regions of the same gene. This is certainly not the expectation under hitchhiking models. Note that the fact that the region showing rapid protein divergence also is highly polymorphic does not negate the conclusion that there is an excess of amino acid fixation in the gene as a whole. A possible explanation for this pattern is that one region of *Acp76A* is a "hotspot" for both adaptive fixations and balanced polymorphisms (*cf.* GILLESPIE 1994a). Perhaps mutations under balancing selection rapidly fix when the environment changes, and the biology of the *Acp76A* protein is such that these selected sites are physically clustered. If this hypothesis is correct then one might be able to detect functional variation at this locus in *D. simulans* populations. Alternatively, it remains possible given the relatively few observed mutations that the significant MK test is a case of falsely rejecting the null hypothesis of neutral evolution just by chance (type I error).

Acp's show greater amounts of amino acid polymorphism and divergence than do a set of non-*Acp* genes in all species of the *simulans* complex (as well as *D. melanogaster*; BEGUN *et al.* 2000). Not surprisingly (*e.g.*, KLIMAN *et al.* 2000), *D. sechellia Acp*'s were less polymorphic than *D. mauritiana* and *D. simulans Acp*'s. Interestingly, *D. mauritiana Acp*'s harbor more amino acid polymorphism than do *D. simulans Acp*'s (Mann-Whitney *U*-test: $P = 0.032$ for θ_w; $P = 0.044$ for θ_π; see Tables 5 and 6), in spite of the fact that *D. simulans* is more polymorphic than *D. mauritiana* at silent sites (HEY and KLIMAN 1993; KLIMAN *et al.* 2000; this report) overall.

TABLE 8

Ratio of replacement to silent heterozygosity for *Acp*'s

Species	$R/S\,\theta_\pi$
D. simulans	0.1666
D. mauritiana	0.3196
D. sechellia	0.4854

$R/S\,\theta_\pi$ is the ratio of replacement θ_π to silent θ_π averaged across all *Acp*'s.

More generally, the rank order of replacement to silent heterozygosity at *Acp*'s (Table 8) is negatively correlated with the presumed rank order effective population size (as inferred from silent heterozygosity). Given that *D. mauritiana* and *D. sechellia* are island endemics, a possible explanation for the relative excess of amino acid polymorphism is that reduced population sizes of *D. mauritiana* and *D. sechellia* result in a greater contribution of mildly deleterious amino acid mutations to polymorphism in these species relative to *D. simulans*. Nevertheless, the pattern does not appear to hold at non-*Acp* loci in *D. mauritiana*. Thus, one would have to explain why the biology of *Acp*'s might result in a distribution of fitness effects for segregating variation in the ancestral population that would be more nearly neutral than the distribution at other loci.

We thank J. Gillespie, P. Awadalla, A. Holloway, and the Nuzhdin lab for their input on this work. J. Coyne and D. Barbash provided stocks. S. Shih, M. Kerber, and R. Repking gave technical assistance. A.D.K. is a Howard Hughes Medical Institute predoctoral fellow. C.D.J. and D.J.B. were funded by the National Science Foundation.

LITERATURE CITED

Aguadé, M., 1998 Different forces drive the evolution of *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 *mst 355* male accessory gland gene region. Genetics **132:** 755–770.

Begun, D. J., and P. Whitley, 2000 Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **97** (11): 5960–5965.

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail and A. G. Clark, 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics **156:** 1879–1888.

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

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

Clark, A. G., 2002 Sperm competition and the maintenance of polymorphism. Heredity **88:** 148–153.

Clark, A. G., M. Agaudé, 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.

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.

Cutler, D. J., 2000 Understanding the overdispersed molecular clock. Genetics **154:** 1403–1417.

Engels, B., 1988 *Monte Carlo Contingency Table Test*, Version 2.1. University of Wisconsin, Madison.

Ewing, B., and P. Green, 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:** 186–194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

Gillespie, J. H., 1978 A general model to account for enzyme variation in natural population. V. The SAS-CFF model. Theor. Popul. Biol. **14:** 1–45.

Gillespie, J. H., 1989 Lineage effects and the index of dispersion of molecular evolution. Mol. Biol. Evol. **6:** 636–647.

Gillespie, J. H., 1993 Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. Genetics **134:** 971–981.

Gillespie, J. H., 1994a Substitution processes in molecular evolution. II. Exchangeable models from population genetics. Evolution **48:** 1101–1113.

Gillespie, J. H., 1994b Substitution processes in molecular evolution. III. Deleterious alleles. Genetics **138:** 943–952.

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

Gillespie, J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model. Genetics **155:** 909–919.

Goldman, N., 1994 Variance to mean ratio, R(t), for Poisson processes on phylogenetic trees. Mol. Phylogenet. Evol. **3:** 230–239.

Goldman, N., and Z. Yang, 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11:** 725–736.

Gordon, D., C. Abajian and P. Green, 1998 Consed: a graphical tool for sequence finishing. Genome Res. **8:** 195–202.

Herndon, L. A., and M. F. Wolfner, 1995 A *Drosophila* seminal fluid protein, *Acp26Aa*, stimulates egg laying in females for 1 day after mating. Proc. Natl. Acad. Sci. USA **92:** 10114–10118.

Hey, J., and R. N. Kliman, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

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:** 5083–5088.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kern, A. D., C. D. Jones and D. J. Begun, 2002 Genomic effects of nucleotide substitutions in *Drosophila simulans*. Genetics **162:** 1753–1761.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex of species. Genetics **156:** 1913–1931.

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

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive 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.

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

Nielsen, R., 1997 Robustness of the estimator of the index of dispersion for DNA sequences. Mol. Phylogenet. Evol. **7:** 346–351.

Nurminsky, D. I., M. V. Nurminskaya, D. D. Aguiar and D. L. Hartl, 1998 Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396:** 572–575.

Ohta, T., and H. Tachida, 1990 Theoretical study of near neutral-

ity. I. Heterozygosity and rate of mutant substitution. Genetics **126:** 219–229.

Parsch, J., C. D. Meiklejohn and D. L. Hartl, 2001 Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of *Drosophila simulans*. Genetics **159:** 647–657.

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

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254.

Swanson, W. J., and V. D. Vacquier, 2002 The rapid evolution of reproductive proteins. Nat. Rev. Genet. **3:** 137–144.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner and C. F. Aquadro, 2001 Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. USA **98:** 7375–7379.

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

Takahata, N., 1987 On the overdispersed molecular clock. Genetics **116:** 169–179.

Takahata, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87:** 2419–2423.

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.

Ting, C.-T., S. C. Tsaur, M.-L. Wu and C.-I Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504.

Ting, C.-T., S. C. Tsaur and C.-I Wu, 2000 The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. Proc. Natl. Acad. Sci. USA **97:** 5313–5316.

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

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

Tsaur, S. C., C.-T. Ting and C.-I Wu, 2001 Sex in *Drosophila mauritiana*: a very high level of amino acid polymorphism in a male reproductive gene, *Acp26Aa*. Mol. Biol. Evol. **18:** 22–26.

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.

Wyckoff, G. J., W. Wang and C.-I Wu, 2000 Rapid evolution of male reproductive proteins in the descent of man. Nature **403:** 304–309.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

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

Communicating editor: M. Aguadé