Evidence for Positive Selection on *Drosophila melanogaster* Seminal Fluid Protease Homologs

Alex Wong, Michael C. Turchin, Mariana F. Wolfner, and Charles F. Aquadro

Department of Molecular Biology and Genetics, Cornell University

Proteins present in the seminal fluid of *Drosophila melanogaster* (accessory gland proteins Acps) contribute to female postmating behavioral changes, sperm storage, sperm competition, and immunity. Consequently, male–female coevolution and host–pathogen interactions are thought to underlie the rapid, adaptive evolution that characterizes several Acp-encoding genes. We propose that seminal fluid proteases are likely targets of selection due to their demonstrated or potential roles in between-sex interactions and immune processes. We use within- and between-species sequence data for 5 predicted protease-encoding Acp loci to test this hypothesis. Our polymorphism-based analyses find evidence for positive selection at 2 genes, both of which encode predicted serine protease homologs. One of these genes, *CG6069*, also shows evidence for consistent selection on a subset of codons over a deeper evolutionary time scale. The second gene, *CG9997*, was previously shown to be essential for normal sperm usage, suggesting that sexual selection may underlie its history of adaptation.

Introduction

Evolutionary biologists have long noted that morphological and behavioral traits involved in mating and reproduction diverge rapidly between species (e.g., Darwin 1871; Eberhard 1996). Recently, such observations have been extended to the molecular level, where studies in diverse taxa have found evidence for positive selection on numerous genes involved in reproduction (reviewed in Clark et al. 2006; Panhuis et al. 2006). Among reproductive molecules of animals with internal reproduction, proteins present in the male seminal fluid are of particular interest due to their influences on postmating behavior and physiology (reviewed in Wolfner 2002; Gillott 2003; Chapman and Davies 2004; Wolfner et al. 2005; Clark et al. 2006; Poiani 2006) and hence their importance to reproductive success.

In *Drosophila melanogaster*, seminal fluid includes over 100 proteins produced by the male's accessory glands (hereafter Acps for accessory gland proteins) as well as proteins expressed in the ejaculatory bulb and the ejaculatory duct (reviewed in Wolfner 2002; Chapman and Davies 2004; Wolfner et al. 2005; Ravi Ram and Wolfner 2007a). Studies using males that lack specific Acps, or that make no Acps at all, have demonstrated roles for these proteins in diverse postmating processes, for example, sperm storage, egg production and egg laying, increased female mortality, decreased female receptivity, and increased feeding (Kalb et al. 1993; Chapman et al. 1995; Tram and Wolfner 1998; Heifetz et al. 2000; Xue and Noll 2000; Chapman et al. 2003; Kubli 2003; Liu and Kubli 2003; Carvalho et al. 2006; Adams and Wolfner 2007). Moreover, mutational and knockdown analyses have ascribed specific roles to ~8 individual Acps (Aigaki et al. 1991; Herndon and Wolfner 1995; Neubaum and Wolfner 1999; Heifetz et al. 2000; Liu and Kubli 2003; Wigby and Chapman 2005; Ravi Ram et al. 2006; Ravi Ram and Wolfner, 2007b). Males lacking the prohormone ovulin (Acp26Aa), for example, induce less ovulation in their mates during the first 24 h postmating than do wild-type males (Herndon and Wolfner 1995; Heifetz et al. 2000), and the large glycoprotein Acp36DE is necessary for normal levels of sperm storage (Neubaum and Wolfner 1999; Bloch Qazi and Wolfner 2003). In addition to such knockout and knockdown approaches, association studies have suggested roles for several Acps (e.g., Acp29AB) in sperm competition (Clark et al. 1995; Fiumera et al. 2005, 2007).

Several studies have shown that genes encoding *Drosophila* Acps evolve differently on average than do other classes of genes. Acp genes show an elevated average level of amino acid divergence (d_N) compared with nonreproductive genes in comparisons between the closely related species *D. melanogaster* and *Drosophila simulans*, despite similar synonymous site divergence (d_S) (Swanson et al. 2001; Mueller et al. 2005). This observation holds true across a broader phylogenetic range as well: across the genomes of 6 fully sequenced species in the *melanogaster* species group, Haerty et al. (2007) found that mean d_N/d_S for 25 genes encoding seminal fluid proteins (including several Acps) is significantly higher than that of ~8,500 genes not encoding seminal fluid proteins. Acp genes also tend to have lower levels of codon bias than do non-Acps (Begun et al. 2000; Mueller et al. 2005) and show an overdispersal of amino acid substitutions within the *D. simulans* species complex (Kern et al. 2004).

Positive selection appears to account for at least some of the unusual patterns of Acp molecular evolution. For several Acp genes, patterns of polymorphism and divergence are consistent with positive selection in comparisons between *D. melanogaster* and *D. simulans* (Aguadé et al. 1992; Tsaur and Wu 1997; Tsaur et al. 1998; Aguadé 1999; Begun et al. 2000; Holloway and Begun 2004). Moreover, reduced variation at the Acp loci *Acp36DE* (Begun et al. 2000) and *Lectin29Ca* (Holloway and Begun 2004) in non-African populations of *D. melanogaster* has been interpreted as evidence for recent selective sweeps at these loci. Moreover, recent divergence analyses find evidence for positive selection on a subset of codons for each of 9 additional Acp loci within the *melanogaster* species group (Haerty et al. 2007). Finally, polymorphism and divergence analyses on several putative Acp-encoding loci also provide substantial evidence for positive selection in *Drosophila pseudoobscura* (Stevison et al. 2004; Wagstaff and Begun 2005a; Schully and Hellberg 2006)

Key words: seminal fluid, Acp, protease, positive selection, adaptation evolution, immunity.

E-mail: aw246@cornell.edu.

Mol. Biol. Evol. 25(3):497–506. 2008
doi:10.1093/molbev/msm270
Advance Access publication December 4, 2007

and in 2 cactophilic species of *Drosophila* (Wagstaff and Begun 2005b). Nonetheless, a relatively small proportion of Acp loci have been examined in most previous studies, making generalizations about selective regimes difficult.

Several hypotheses have been forwarded to explain the rapid amino acid evolution of, and increased incidence of positive selection among, *Drosophila* Acps and seminal fluid proteins more generally (reviewed in Swanson and Vacquier 2002; Clark et al. 2006; Panhuis et al. 2006). First, male–female and male–male interactions may underlie rapid Acp evolution. Sexual conflict, sperm competition, cryptic female choice, and other forms of sexual selection may exert strong selective pressures on some Acps, given the roles of these proteins in clearly relevant physiological processes (e.g., sperm storage, postmating life span reduction). Second, host–pathogen interactions may account for some rapid Acp evolution. *Drosophila* seminal fluid contains several components with antibacterial activity (Samakovlis et al. 1991; Lung et al. 2001; Mueller et al. 2007), which may help to prevent infection during mating. Thus, Acps with roles in immunity may evolve rapidly as a consequence of a host–pathogen arms race.

In this study, we focus on the molecular evolution and molecular population genetics of 5 predicted proteases and protease homologs (i.e., proteins resembling proteases in sequence and structure but with one or more catalytic site mutations; Ross et al. 2003) expressed in the *D. melanogaster* male accessory gland. Three lines of evidence suggest that reproductive tract proteases and protease homologs may be subject to strong selection. First, proteolysis regulators (used here to refer to proteases and their inhibitors as well as protease homologs) are likely to mediate male–female interactions during mating. Previous work has suggested roles for both male- and female-derived factors in the processing of at least one Acp prohormone (ovulin; Park and Wolfner 1995; Ravi Ram et al. 2006), and numerous proteolysis regulators are present in both male seminal fluid and in the female reproductive tract (Swanson et al. 2001, 2004; Mueller et al. 2004; Mack et al. 2006; Kelleher et al. 2007; Lawniczak and Begun 2007). As such, interactions between male- and female-derived proteolysis regulators may be the foci for sexual selection. Consistent with this hypothesis, several proteases expressed in the *Drosophila* female reproductive tract show evidence of positive selection (Swanson et al. 2004; Panhuis and Swanson 2006; Kelleher et al. 2007; Lawniczak and Begun 2007), and 1 Acp protease inhibitor (Acp76A) out of 2 that have been examined shows evidence for positive selection along the *D. simulans* lineage (Begun et al. 2000; Kern et al. 2004).

Second, proteases, protease homologs, and protease inhibitors are associated with fertility effects in several species, including *Drosophila* (Ravi Ram and Wolfner 2007b) and mice (Mbikay et al. 1997; Murer et al. 2001; Carpentier et al. 2004; Nie et al. 2005; Uhrin et al. 2007), again raising the possibility that proteolysis regulators are subject to sexual selection. Moreover, several predicted protease inhibitors present in male seminal fluid are toxic upon ectopic expression and may therefore contribute to the cost of mating (Lung et al. 2002; Mueller et al. 2007). Third, proteolytic cascades play important roles in immunity and defense in many organisms (Ligoxygakis, Pelte, Hoffmann, et al. 2002; Ligoxygakis, Pelte, Ji, et al. 2002; Sim and Tsiftsoglou 2004) and thus may experience selection pressure from pathogens.

Table 1
Genes Used in This Study

Gene	Codons	Gene Ontology
CG6069	283	SPH
CG6168	314	Metalloprotease
CG9997	330	SPH
CG11664	209	SPH
CG11864	251	Metalloprotease

NOTE.—An SPH resembles a catalytically active serine protease but bears one or more active site mutations, suggesting that catalytic function is likely absent. Such molecules have been proposed to regulate proteolytic cascades (Kwon et al. 2000; Lee et al. 2002; Asgari et al. 2003; Jiang, Wang, Yu, Kanost, 2003; Jiang, Wang, Yu, Zhu, et al. 2003; Yu et al. 2003; Gupta et al. 2005).

In this study, we report results from molecular population genetic surveys and divergence analyses of 5 Acp genes encoding predicted proteases or protease homologs (see table 1 for gene symbols, gene ontologies, and coding sequence lengths). These genes encode 5 of the 6 protease/protease homologs reported in an expressed sequence tag (EST) screen of the male accessory gland (Swanson et al. 2001), although more genes encoding predicted proteases or protease homologs with accessory gland–biased expression have since been identified (Chintapalli et al. 2007; Ravi Ram and Wolfner 2007a). Two of these 5 genes, *CG11864* and *CG6168*, are each predicted to encode a metalloprotease (Mueller et al. 2004), that is, a protease with a metal ion at its active site. Previous studies have assigned potential physiological roles to both metalloproteases: *CG11864* is essential for cleavage of at least 2 other Acps, Ovulin/Acp26Aa and Acp36DE (Ravi Ram et al. 2006), and ectopic expression of *CG6168* in a virgin female increases her ability to clear a bacterial infection (Mueller et al. 2007).

The other 3 genes examined here, *CG6069*, *CG9997*, and *CG11664*, are predicted to encode serine protease homolog (SPHs)—although they are predicted to resemble serine proteases in overall structure, mutations at one or more of the 3 canonical active site residues likely render them noncatalytic (Mueller et al. 2004). Of these 3 genes, a function has thus far only been assigned to 1: RNA interference (RNAi) knockdown of *CG9997* suggests that it is essential for normal sperm usage (Ravi Ram and Wolfner, 2007b). We note that the biochemical and physiological roles of protease homologs are currently not well understood. Although these proteins likely lack catalytic activity, several studies suggest roles for protease homologs in regulating the activity of catalytically active proteases, either as cofactors or as competitive inhibitors (Kwon et al. 2000; Lee et al. 2002; Asgari et al. 2003; Jiang, Wang, Yu, Kanost, 2003; Jiang, Wang, Yu, Zhu, et al. 2003; Gupta et al. 2005). Such roles would make proteases and protease homologs alike subject to the evolutionary pressures just described.

Using molecular population genetic surveys of African population samples of *D. melanogaster* (Pool and Aquadro 2006), we find evidence for strong directional selection at 2 loci out of the 5 examined, *CG6069* and *CG9997*. At a deeper evolutionary time scale, we also find evidence

Table 2
Summary Statistics

Gene	*n*	Length (bp)	S	θ	π_{Tot}	π_{Syn}	π_{Non}	d_S	d_N	*D*	*H*
CG6069	20	945	8	0.0025	0.0018	0.0060	0.0002	0.136	0.0161	−0.971	0.853
CG6168	18	905	80	0.0275	0.0228	0.0579	0.0128	0.178	0.0360	−0.560	0.732
CG9997	12	912	17	0.0062	0.0066	0.0228	0.0017	0.099	0.0251	0.262	2.242
CG11664	12	641	12	0.0041	0.0038	0.0136	0.0007	0.194	0.0247	−0.293	0.545
CG11864	12	694	11	0.0053	0.0053	0.0099	0.0029	0.119	0.033	−0.526	−2.27

NOTE.—Population summary statistics are for African populations of *Drosophila melanogaster*. *n*: alleles sequenced and S: number of segregating sites. d_N and d_S were calculated using *D. melanogaster* and *Drosophila simulans*. *D*: Tajima's *D* statistic (Tajima 1989). *H*: Fay and Wu's *H* statistic (Fay and Wu 2000). No locus rejects neutrality using *D* or *H*.

for recurrent positive selection on a subset of codons in *CG6069*. These findings, along with previous studies on reproductive tract proteolysis regulators in male (Kern et al. 2004) and female *Drosophila* (Swanson et al. 2004; Panhuis and Swanson 2006; Kelleher et al. 2007; Lawniczak and Begun 2007), support the hypothesis that interactions between males and females drive the rapid evolution of some reproductive genes.

Materials and Methods

Drosophila Strains, DNA Sequences, and Sequence Alignment

For polymorphism-based analyses, we used chromosome extraction lines derived from African populations of *D. melanogaster* (Uganda for chromosomes 2 and 3, Zimbabwe for the X Chromesome; population samples are described in Pool and Aquadro 2006), with *D. simulans* as an outgroup (*D. simulans* sequences reported in Mueller et al. 2005). Sample sizes for each gene are given in table 2. For divergence analyses, we used 6 species in the *D. melanogaster* subgroup. Sequences from *D. melanogaster*, *D. simulans*, and *Drosophila yakuba* are from Mueller et al. (2005). Additional sequences were collected from *Drosophila teissieri* (Tucson *Drosophila* stock center *D. teissieri* 257.0), *Drosophila erecta* (S-18, kindly donated by Andrew Clark; originally from the Ashburner laboratory), and *Drosophila santomea* (strain CAR1566-8, kindly donated by Peter Andolfatto).

DNA extractions were performed using the Puregene DNA purification kit (Gentra Systems, Minneapolis, MN), and genes were amplified by polymerase chain reaction; primer sequences and conditions are available upon request. Sequencing was carried out on an ABI 3730 automated sequencer using BigDye chemistry (Applied Biosystems, Foster City, CA). Sequence alignments were carried out using the ClustalW algorithm as implemented in MegAlign (DNASTAR, Inc., Madison, WI) on protein sequences. Sequences have been deposited in GenBank under accession numbers EU328540–EU328631. Introns in *CG11864* were identified in other species using the *D. melanogaster* annotation as a guide; the other genes contain no introns in their coding regions. Consensus sites for intron start (AG) and stop (GT) were conserved across all species.

Analysis of Polymorphism

Summary statistics (θ, π, and divergence) for each gene were calculated using DnaSP4.1 (Rozas et al. 2003). McDonald–Kreitman tests (McDonald and Kreitman 1991), as well as Tajima's *D* test (Tajima 1989) and Fay and Wu's *H* test (Fay and Wu 2000), were also performed using DnaSP. *D. simulans* was used for all analyses requiring an outgroup. In order to calculate statistical significance for Tajima's *D* and Fay and Wu's *H* test, we used the coalescent simulator with recombination implemented in DnaSP. The population recombination rate $R = 4N_e rm$ was estimated using $N_e = 1 \times 10^6$ (Kreitman 1983), where *m* is the size in base pairs of the gene under consideration (table 4), and with per base pair recombination rate estimates *r* obtained from Hey and Kliman (2002). Estimated values of *R* were 100.65 for *CG6069*, 138.82 for *CG6168*, 108.84 for *CG9997*, 9.83 for *CG11664*, and 12.29 for *CG11864*.

Hudson–Kreitman–Aguadé (HKA) tests were performed using the maximum likelihood method of Wright and Charlesworth (2004) (Hudson et al. 1987; Wright and Charlesworth 2004). This method uses loci specified a priori to generate a null model of sequence evolution and assesses the fit of one or more loci of interest to that null model. The parameter *k* measures the decrease or increase of polymorphism relative to divergence, with the neutral expectation that $k = 1$. We used 4 X-linked noncoding loci reported in Pool and Aquadro (2006) as representative "neutral" loci and tested each protease- or protease homolog–encoding gene individually.

Divergence-Based Analyses

Inferences of positive selection using comparisons between the number or rates of nonsynonymous and synonymous substitutions can be mislead if the latter varies across a sequence. If, for example, some sites have a particularly low rate of synonymous substitution, then $\omega > 1$ may be inferred even in the absence of positive selection, under the assumption of a single synonymous rate. As such, we used maximum likelihood methods implemented in HyPhy (Pond et al. 2005) to test for heterogeneity in the rate of synonymous substitution (d_S) at different codons in each multiple sequence alignment. The null model assumes no variation in d_S but allows variation in the rate of nonsynonymous substitution (d_N) in the form of 3 discrete rate classes (Pond and Muse 2005). The alternative model allows variation in d_S, with 2 discrete rate classes. The 2 models can be compared using either a likelihood ratio test (LRT), with twice the difference in −lnL between models following a χ^2_4 distribution, or Akaike information criterion scores.

Rejection of the null model provides evidence for variation in d_S at different codons.

Sequence alignments were analyzed for evidence of positive selection in the form of an elevated rate of nonsynonymous substitution compared with the rate of synonymous substitution (ω) using phylegenetic analysis by maximum likehood (PAML) (Yang 1997; Yang et al. 2000). Two model comparisons were performed. In the first comparison, the null model M1a allows for 2 classes of sites: one with $0 < \omega < 1$ and one with $\omega = 1$. The alternative model M2a adds a third site class with $\omega > 1$. In the second comparison, the null model M8A uses a beta distribution to describe sites with $0 < \omega < 1$, with an extra category of sites with $\omega = 1$. The alternative model M8 allows the extra category to undergo positive selection, that is, has $\omega > 1$. For both comparisons, the null and alternative models can be compared via LRT, with the difference in log likelihoods (δ) following a χ^2_2 distribution (M1a vs. M2a) or a χ^2_1 distribution (M8A vs. M8).

In order to evaluate the fit of δ to the appropriate χ^2 distribution, and as an independent estimate of the *P* value for each model comparison, we also implemented a parametric bootstrap. Parameter estimates from M1a were used to generate 250 simulated data sets using evolverNSsites (Anisimova et al. 2001), which were then analyzed under M1a, M2a, M8A, and M8. Values of δ from the simulated neutral data sets were then calculated and used to obtain the probability of obtaining the observed value of δ under the null hypothesis.

For both HyPhy and PAML analyses, a single tree – ((*D. melanogaster*, *D. simulans*), (*D. teissieri*, (*D. yakuba*, *D. santomea*)), *D. erecta*) – was assumed, following current understanding (e.g., LaChaise et al. 2000; Ko et al. 2003; Pollard et al. 2006; Wong et al. 2007). We note that several authors have suggested that lineage sorting has occurred in the common ancestor of the *melanogaster* subgroup and that this may introduce inferential problems (Pollard et al. 2006; Wong et al. 2007). Here, however, all analyses were conducted on an unrooted tree, with no species basal to the *melanogaster* subgroup; as such, our analyses should not be affected by lineage sorting.

Results

Various modes of adaptation can leave different signatures in sequence data, such that different kinds of sample are suitable for their detection. For example, repeated episodes of positive selection on a few codons of a coding sequence can be inferred using multispecies divergence data, while polymorphism data from a single population are suited to the detection of a recent selective sweep or ongoing balancing selection. We have collected both divergence and polymorphism data in order to gain a comprehensive view of the patterns of molecular evolution at 5 putative protease-or protease homdog-encoding Acp genes.

Polymorphism Analyses

We collected polymorphism data for each gene from African populations of *D. melanogaster*. We used an African population rather than a North American, European, or Asian population in order to avoid, as best as possible, inferential problems stemming from nonequilibrium demographic histories (e.g., Jensen et al. 2005; Thornton et al. 2007). In all 12–20 alleles were sequenced for each gene; sample sizes and summary statistics are given in table 2. Neither Tajima's *D* (Tajima 1989) nor Fay and Wu's *H* test (Fay and Wu 2000) deviates from the neutral expectation for any gene (table 2).

We used 2 additional tests of neutrality to assess the fit of the polymorphism data to the standard neutral model. The first, the McDonald–Kreitman (MK) test (McDonald and Kreitman 1991), tests the neutral prediction that the ratio of nonsynonymous to synonymous substitutions between species should equal the ratio of nonsynonymous to synonymous polymorphisms within species (table 3). For 4 genes, *CG6069*, *CG6168*, *CG11664*, and *CG11864*, we fail to find any deviation from the null hypothesis. However, for *CG9997*, the null hypothesis of equal ratios of nonsynonymous to synonymous changes within and between species is rejected ($P = 0.008$). Rejection of the null hypotheses could in theory result from deviations from the neutral expectation in any cell of the MK table; we suggest that an excess of nonsynonymous fixations is the most likely explanation. Nonsynonymous divergence is high at *CG9997* ($d_N = 0.025$ for the *melanogaster/simulans* comparison; table 2), relative to an average d_N of 0.0124 (95% confidence interval [CI]: 0.0121–0.0128) compiled from ~8500 genes (data from Larracuente et al (forthcoming)). This does not appear to be the result of a high mutation rate at *CG9997* because synonymous divergence ($d_s = 0.099$) is slightly lower than average (0.128; 95% CI: 0.126–0.129). In addition, levels of polymorphism at *CG9997* do not appear to differ substantially from average. Thus, it is likely that an excess of nonsynonymous substitution due to positive selection in the lineages leading to *D. melanogaster* and/or *D. simulans* accounts for this result.

We also used the HKA test (Hudson et al. 1987) to assess the neutral prediction that the ratio of polymorphism to divergence should be the same for different loci (table 4). This test is particularly useful for detecting a deficit or an excess of polymorphism due to recent directional selection or balancing selection, respectively. We used 4 X-linked noncoding loci (named after their cytological locations: 4F2, 8A4, 11A5, and 12F1) reported in Pool and Aquadro (2006) as representative neutral loci and tested each protease- or protease homolog–encoding gene against the neutral prediction, using the maximum likelihood HKA test of Wright and Charlesworth (2004). We found that 4 protease/protease homolog genes conformed to the neutral prediction, with 1 protease homolog gene, *CG6069*, rejecting neutrality ($P = 0.016$). The latter rejection of neutrality could be the result of either elevated silent site divergence or a deficit of silent polymorphism. Given that d_S is about average for *CG6069* (0.139 for CG6069 vs. 0.128 genome wide), although very few polymorphisms were observed ($\theta_S = 0.00694$), we suggest that *CG6069* is depauperate for variation, consistent with the action of recent selection at or near this locus. The rate of recombination in this region of the genome is moderate ($r = 2.6 \times 10^{-6}$ cm/m; Hey and Kliman 2002), consistent with *CG6069* (rather than a linked locus) being the target of selection. Tests of the frequency

Table 3
McDonald–Kreitman Tests

			Polymorphic		Fixed		
Gene	Population	*n*	Silent	Replacement	Silent	Replacement	Probability
CG6069	Uganda	20	5	1	25	10	0.665
CG6168	Uganda	18	45	36	25	16	0.699
CG9997	Uganda	12	14	3	16	23	0.008*
CG11664	Zimbabwe	12	6	2	18	12	0.684
CG11864	Uganda	12	5	5	15	11	0.722

NOTE.—*P* values were obtained using a 2-tailed Fisher's exact test. *$p<0.05$.

spectrum (Tajima's *D* and Fay and Wu's *H*) do not reject neutrality; we suspect that low variation (perhaps due to very recent selection) reduces the power of these tests.

CG6168 presents an interesting case. Polymorphism at this gene is extremely high (table 2), with 80 segregating sites in the coding region and $\pi_{Syn} = 0.0579$ (vs. an average of ~0.029 genome wide; Andolfatto 2005), yet the HKA test does not reject neutrality. Similarly, tests of neutrality based on site frequency spectra do not find deviations from the neutral expectation, either using all polymorphisms (table 2) or synonymous and nonsynonymous polymorphisms separately (table 5). Polymorphism is also very high in a population sample collected from Pennsylvania, despite a recent bottleneck for non-African populations of *D. melanogaster* (Fiumera A, personal communication). High silent site divergence at *CG6168* ($d_S = 0.178$; table 2) may account for the failure to reject neutrality using the HKA test. There is no evidence that selection drives high synonymous site divergence as patterns of unpreferred and preferred differences within and between species are not significantly different using the frequency distribution divergence Mann-Whitney U (fddMWU) test (Akashi 1999). We suspect that balancing selection may operate to maintain high levels of polymorphism at this locus, but more data will be required to rigorously evaluate this hypothesis.

Divergence Analyses

Variation in d_S within a gene can mislead commonly used individual locus divergence-based tests for positive selection (Pond and Muse 2005). Using model comparisons implemented in HyPhy (Pond et al. 2005), we fail to find evidence of variation in d_S at any gene examined in this study. For all 5 genes, the data do not fit a model incorporating variation in d_S significantly better than they fit a null model with no such variation (table 6). Although failure to reject the null hypothesis does not warrant its acceptance, this result suggests that use of models that assume a single synonymous substitution rate to infer positive selection should not be misled by variation in d_S.

We therefore used PAML, which assumes a single value of d_S for each gene, to infer the action of recurrent positive selection on individual codons (table 7). We find strong evidence for positive selection on 1 gene, *CG6069*. Using both the M2a versus M1a and the M8 versus M8A comparisons, the data for *CG6069* fit the alternative (selection) model significantly better than they do the null model. About 4–5% of codons are estimated to belong to the selected class, with $\omega = 3.44$ under M8 ($\omega = 3.98$ under M2a). Because the predicted 3-dimensional structure of the protein encoded by *CG6069* was previously modeled (Mueller et al. 2004), we could locate the putative positively selected residues on its predicted structure (fig. 1). The sites whose mean $\omega \pm 1$ standard error is greater than 1 (corresponding to posterior probabilities >0.774 of belonging to the selected class) that fell within the modeled domain (5 out of 6 total) are predicted to be on the protein's surface, although none lies within the predicted substrate binding cleft. For a second gene, *CG11664*, the M8 versus M8A comparison is marginally significant using a χ^2 test ($P = 0.048$), with other tests being marginally nonsignificant.

Use of a parametric bootstrap to evaluate the significance of model comparisons was consistent with the results obtained from LRT (table 7). In most cases, the LRT and the bootstrap resulted in rejections, or failures to reject, for the same comparisons. The one exception is for the M8A versus M8 comparison for *CG11664*, where the LRT result is marginally significant ($P = 0.048$) and the bootstrap result is nonsignificant ($P = 0.071$).

Because neutrality was rejected for *CG6069* using both divergence and polymorphism-based tests (tables 4 and 7), we were interested in determining whether recent selection on this gene has targeted the same residues as those identified as under positive selection by PAML. Of the 6 codons

Table 4
Maximum Likelihood HKA-Tests

Gene	Silent S	Silent Divergence	2 × ΔlnL	*P*	k
CG6069	5	27	5.83	0.016*	0.22
CG6168	45	35	0.83	0.36	1.57
CG9997	14	24	0.04	0.85	0.8
CG11664	8	20	0.26	0.61	0.71
CG11864	5	18.5	1.79	0.17	0.41

NOTE.—Silent S: synonymous segregating sites, Silent Divergence: synonymous divergence between *Drosophila melanogaster* and *Drosophila simulans*, 2 × ΔlnL: twice the difference in log likelihood between the null and selection models, *P*: *P* value obtained from a χ^2 test (degrees of freedom [df] = 1), and k: estimated ratio of variation at the given locus to the neutral expectation. *$p<0.05$.

Table 5
Comparisons of Frequency Spectra for Synonymous and Nonsynonymous Polymorphisms at *CG6168*

	Tajima's *D*	Fu and Li's *D*	Fay and Wu's *H*
Synonymous	−0.500	−0.851	−1.935
Nonsynonymous	−0.756	−1.044	−0.418

identified by PAML as having $\omega > 1$, 2 —^{190}Ile and ^{268}Ser—appear to have changed along the *melanogaster* species lineage, although the high variability of both codons makes polarization of changes uncertain. These 2 codons are particularly good candidates for having been recent targets of selection in *D. melanogaster*. An additional 5 sites, ^{124}Ile, ^{152}Ser, ^{208}Ile, ^{230}Gly, and ^{285}Thr, appear to have fixed along the *melanogaster* lineage but do not have high posterior probabilities of $\omega > 1$.

Discussion

A number of genes encoding seminal fluid proteins show evidence for positive selection in diverse taxa, for example, *Drosophila* (reviewed in Clark et al. 2006; Panhuis et al. 2006), crickets (Andrès et al. 2006), and primates (Clark and Swanson 2005). Several explanations have been proposed for the rapid, adaptive evolution of genes encoding seminal fluid proteins, including postmating male–female or male–male interactions and immune pressures. We hypothesized that some Acp proteases would be the targets of adaptive evolution in *D. melanogaster* and its close relatives, given the potential role of proteolysis regulators in mediating male–female interactions, and known or suspected roles for several such proteins in immunity, sperm usage, and proteolytic processing of other rapidly evolving Acps. Using polymorphism-based tests, we find evidence for positive selection on 2 protease homolog genes out of 5 genes examined: *CG9997* appears to have undergone an excess of amino acid substitutions between *D. melanogaster* and *D. simulans*, whereas patterns of polymorphism at *CG6069* are consistent with a recent selective sweep. Furthermore, between-species analyses suggest that *CG6069* has experienced pervasive positive selection on a subset of codons in the *melanogaster* subgroup.

RNAi knockdown studies on *CG9997* suggest a role for this gene's product in regulating the release of sperm from storage in females (Ravi Ram and Wolfner, 2007b). Because sperm storage is potentially involved in cryptic female choice and sperm competition (Eberhard 1996; Simmons 2001), it is likely that sexual selection of some variety underlies the molecular evolution of this gene. However, it should be noted that *CG9997*'s role in other systems potentially subject to strong selection, for example, the immune response, has not been fully investigated. Ectopic expression of *CG9997* in females does not affect systemic clearance of the gram-negative bacterium *Serratia marcescens* (Mueller et al. 2007), but its activity against gram-positive bacteria or fungi, or any localized activity in the reproductive tract, has not been examined.

Knockdown and ectopic expression studies have not yet uncovered any potential role for *CG6069*, the second positively selected gene identified here, in the regulation of postmating responses (Ravi Ram and Wolfner 2007b), seminal fluid toxicity (Mueller et al. 2007), or in immunity (Mueller et al. 2007). Moreover, no data currently exist with respect to the localization of *CG6069*'s protein product in the female reproductive tract. As such, it is not currently possible to ascribe this gene's rapid molecular evolution to a particular physiological process.

CG9997 and *CG6069*, the 2 genes inferred in this study to have experienced positive selection, are predicted to encode SPHs, that is, their protein products are predicted to resemble serine proteases but bear mutations in one or more of the 3 canonical active site residues (Ross et al. 2003; Mueller et al. 2004). As such, these proteins are probably not proteolytically active. However, noncatalytic roles have been assigned to, or suggested for, SPHs in several systems. For example, studies on the cleavage of prophenoloxidase (proPO) to phenoloxidase, which is involved in the melanization of pathogens, have suggested a role for SPHs in modulating the activity of proPO-activating proteases (PAPs). In tobacco hornworm (Jiang, Wang, Yu, Kanost, et al. 2003; Jiang, Wang, Yu, Zhu, et al. 2003; Yu et al. 2003; Gupta et al. 2005) and several beetles (Kwon et al. 2000; Lee et al. 2002), SPHs are required for full proteolytic activity of PAPs. Conversely, an SPH present in the venom of a parasitic wasp is capable of interfering

Table 6
Tests for Variation between Sites in the Rate of Synonymous Substitution Using HyPhy

	logL			
Gene	MG94 × REV Nonsynonymous GDD 3	MG94 × REV Dual GDD 2 × 3	*P* value	ΔAIC
CG6069	−2479.46	−2475.53	0.097	−0.16
CG6168	−2582.97	−2580.94	0.398	−3.93
CG9997	−2588.96	−2588.14	0.803	−6.37
CG11664	−1626.42	−1626.29	0.992	−7.74
CG11864	−1943.62	−1943.43	0.989	−7.62

NOTE.—MG94 × REV Nonsynonymous GDD 3 is a model incorporating variation in the rate of nonsynonymous substitution (3 rate classes) but not in the rate of synonymous substitution. The dual model incorporates variation in both the rate of nonsynonymous substitution (3 rate classes) and the rate of synonymous substitution (2 rate classes). The indicated *P* value is for the LRT between the nonsynonymous and dual models, using the asymptotic distribution of χ^2_4. No tests were significant with $\alpha = 0.05$, indicating no evidence for variation in the rate of synonymous substitution. ΔAIC: difference in Akaike information criterion (AIC) scores between the nonsynonymous and dual models; negative ΔAIC values indicate that the dual GDD 2 × 3 model does not outperform the GDD 3 model, given the extra parameters used by GDD 2 × 3.

Table 7
Tests for Positive Selection Using PAML

	−lnL				*P* value			
Gene	M1a	M2a	M8A	M8	M2a versus M1a	M8 versus M8A	ω M2a (proportion)	ω M8 (proportion)
CG6069	2424.71	2420.67	2424.72	2420.42	0.018*; 0.004*	0.003*; 0.004*	3.98 (0.038)	3.44 (0.055)
CG6168	2328.43	2328.43	2328.32	2328.32	1; 1	1; 1	1	1
CG9997	2348.63	2348.32	2348.64	2348.33	0.738; 0.280	0.432; 0.308	2.9	1.60
CG11664	1574.68	1573.33	1574.69	1573.06	0.06; 0.259	0.048*; 0.071	3.15	3.41
CG11864	1690.86	1690.86	1690.86	1690.86	1; 1	1; 1	1	1

NOTE.—lnL: negative log likelihood for the indicated model. *P* values: the first *P* value reported is for the LRT between the selection (M2a or M8) and neutral (M1a or M8A) models, using the asymptotic distribution of χ^2_2(M2a vs. M1a) or χ^2_1(M8 vs. M8A). The second *P* value was obtained by parametric bootstrapping under the maximum likelihood parameter estimates from model M1a. In all 250 bootstrap replicates were generated using evolverNSsites. The last 2 columns give estimated values of ω under the indicated model. Numbers in parentheses indicate the proportion of codons estimated to belong to the selected class, for comparisons where positive selection was inferred. *$p<0.05$.

with proPO cleavage, perhaps by competing with host SPHs for binding to PAP and/or proPO (Asgari et al. 2003). If positively selected *Drosophila* SPHs present in the seminal fluid function either as agonists or antagonists of catalytically active proteases, then coevolution with proteases, protease substrates, inhibitors, or other binding partners may underlie their adaptive evolution. Other documented molecular functions for SPHs include glycoprotein binding (Watorek 2003) and cell adhesion (Huang et al. 2000; Lin et al. 2006); *Drosophila* seminal fluid SPHs could also be involved in any of these functions as a number of other Acps are glycosylated (Monsma and Wolfner 1988; Bertram et al. 1996; Saudan et al. 2002), and cell adhesion may be important for sperm storage and/or fertilization.

We found no evidence for positive selection on 3 other protease-encoding Acp genes, 2 of which, *CG11864* and *CG6168*, have been ascribed functions using genetic methods. Knockdown of *CG11864* shows that this putative metalloprotease is necessary for the proteolytic cleavage of 2 Acps, the egg-laying prohormone ovulin and the sperm storage protein Acp36DE (Ravi Ram et al. 2006). Although both ovulin and Acp36DE appear to have experienced positive selection (Aguadé et al. 1992; Begun et al. 2000; Fay and Wu 2000), we found no evidence of a similar history for *CG11864*. This is not, we suggest, a surprising result: If proteolytic cleavage of ovulin and/or Acp36DE is necessary for some aspect of their functions (although this is, as y unknown), then both the cleavage sites and the responsible proteases should be well conserved. We note that other regions of ovulin thought to be structurally important are highly conserved between species (Wong et al. 2006) and suspect that the same will be true of cleavage sites.

FIG. 1.—Structural model of the predicted protease homolog encoded by *CG6069*. Sites whose inferred ω ± 1 standard error is greater than 1 are shown in white. All 5 selected residues that fall within the modeled domain are predicted to lie on the protein's surface, although none lies in the predicted substrate-binding cleft. The model was generated by Mueller et al. (2004).

Ectopic expression of the predicted metalloprotease *CG6168* in females aids in the clearance of systemic *S. marcescens* infection (Mueller et al. 2007), suggesting that this protein may participate in immune regulatory cascades. Polymorphism at *CG6168* is high, although several tests find no deviations from the neutral expectation. Classic studies attribute extremely high levels of polymorphism at major histocompatibility complex (MHC) genes to balancing selection arising from host–pathogen interactions (e.g., Hughes and Nei 1988; McConnell et al. 1988). It is possible that a similar explanation underlies high polymorphism at *CG6168*, but further statistical and functional analyses are required.

Conclusions

An understanding of the rapid evolution of an elevated proportion of *Drosophila* Acps and reproduction-related genes more generally requires both extensive sequence data and functional characterization. Full genome sequences from multiple species of *Drosophila* have allowed a comprehensive examination of sex- and reproduction-related genes on a deep phylogenetic scale (Haerty et al. 2007). Population genetic analyses, however, have been narrower in scope, with most studies focusing on a limited set of genes. We conducted divergence and polymorphism analyses at 5 male Acp-encoding loci that have not been previously examined at the population level and found evidence for positive selection at 2 predicted protease homolog–encoding genes. Adaptive evolution of protease, protease homolog, or protease inhibitor genes has now been documented in genes expressed in either the male accessory gland (this study; Kern et al. 2004) or the female reproductive tract (Swanson et al. 2004; Panhuis and Swanson 2006;

Kelleher et al. 2007; Lawniczak and Begun 2007). Although definitive interpretation of these results must await functional characterization of positively selected genes (data from females are particularly lacking), the finding of positive selection on both male and female reproductive tract genes suggests that between-sex interactions, rather than simply male–male competition, drive the rapid evolution of some reproductive genes.

Acknowledgments

We would like to thank the Wolfner and Aquadro laboratories, as well as 2 anonymous reviewers, for helpful comments and suggestions on the manuscript. This work was funded by National Institutes of Health grants to M.F.W. (HD38921) and C.F.A. (GM36431) and a National Science Foundation DDIG (0508152) to A.W. A.W. is a Howard Hughes Medical Institute predoctoral fellow.

Literature Cited

Adams EM, Wolfner MF. 2007. Seminal proteins but not sperm induce morphological changes in the *Drosophila melanogaster* female reproductive tract during sperm storage. J Insect Physiol. 53:319–331.

Aguadé M. 1999. Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. Genetics. 152:543–551.

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

Aigaki T, Fleischmann I, Chen PS, Kubli E. 1991. Ectopic expression of sex peptide alters reproductive behavior of female *D. melanogaster*. Neuron. 7:557–563.

Akashi H. 1999. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics. 151:221–238.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature. 437:1149–1152.

Andrès JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. 2006. Molecular evolution of seminal proteins in field crickets. Mol Biol Evol. 23:1574–1584.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol. 18:1585–1592.

Asgari S, Zhang G, Zareie R, Schmidt O. 2003. A serine proteinase homolog venom protein from an endoparasitoid wasp inhibits melanization of the host hemolymph. Insect Biochem Mol Biol. 33:1017–1024.

Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics. 156:1879–1888.

Bertram MJ, Neubaum DM, Wolfner MF. 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.

Bloch Qazi MC, Wolfner MF. 2003. An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. J Exp Biol. 206:3521–3528.

Carpentier M, Guillemette C, Bailey JL, Boileau G, Jeannotte L, DesGroseillers L, Charron J. 2004. Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. Mol Cell Biol. 24:4428–4437.

Carvalho GB, Kapahi P, Anderson J, Benzer S. 2006. Allocrine modulation of feeding behavior by the sex peptide of *Drosophila*. Curr Biol. 16:692–696.

Chapman T, Bangham J, Vinti G, Seifried B, Lung O, Wolfner MF, Smith HK, Partridge L. 2003. The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci USA. 100:9923–9928.

Chapman T, Davies SJ. 2004. Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies. Peptides. 25:1477–1490.

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

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet. 39:715–720.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. Reproduction. 131:11–22.

Clark AG, Aguadé M, Prout T, Harshman LG, Langley CH. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics. 139:189–201.

Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet. 1:e35.

Darwin C. 1871. The descent of man, and selection in relation to sex. London: John Murray.

Eberhard WG. 1996. Female control: sexual selection by cryptic female choice. Princeton (NJ): Princeton University Press.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413.

Fiumera AC, Dumont BL, Clark AG. 2005. Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. Genetics. 169:243–257.

Fiumera AC, Dumont BL, Clark AG. 2007. Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. Genetics. 176:1245–1260.

Gillott C. 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. Annu Rev Entomol. 48:163–184.

Gupta S, Wang Y, Jiang H. 2005. *Manduca sexta* prophenoloxidase (proPO) activation requires proPO-activating proteinase (PAP) and serine proteinase homologs (SPHs) simultaneously. Insect Biochem Mol Biol. 35:241–248.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al (11 co-authors). 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics. 177:1321–1335.

Heifetz Y, Lung O, Frongillo EA Jr, Wolfner MF. 2000. The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Curr Biol. 10:99–102.

Herndon LA, Wolfner MF. 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, Kliman RM. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. Genetics. 160:595–608.

Holloway AK, Begun DJ. 2004. Molecular evolution and population genetics of duplicated accessory gland protein genes in *Drosophila*. Mol Biol Evol. 21:1625–1628.

Huang TS, Wang H, Lee SY, Johansson MW, Soderhall K, Cerenius L. 2000. A cell adhesion protein from the crayfish *Pacifastacus leniusculus*, a serine proteinase homologue similar to *Drosophila* masquerade. J Biol Chem. 275:9996–10001.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116: 153–159.
Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 335:167–170.
Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics. 170: 1401–1410.
Jiang H, Wang Y, Yu XQ, Kanost MR. 2003. Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*. A bacteria-inducible serine proteinase containing two clip domains. J Biol Chem. 278:3552–3561.
Jiang H, Wang Y, Yu XQ, Zhu Y, Kanost M. 2003. Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. Insect Biochem Mol Biol. 33:1049–1060.
Kalb JM, DiBenedetto AJ, Wolfner MF. 1993. Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. Proc Natl Acad Sci USA. 90:8093–8097.
Kelleher ES, Swanson WJ, Markow TA. 2007. Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. PLoS Genet. 3: 1541–1549.
Kern AD, Jones CD, Begun DJ. 2004. Molecular population genetics of male accessory gland proteins in the *Drosophila simulans* complex. Genetics. 167:725–735.
Ko WY, David RM, Akashi H. 2003. Molecular phylogeny of the *Drosophila melanogaster* species subgroup. J Mol Evol. 57:562–573.
Kreitman M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature. 304:412–417.
Kubli E. 2003. Sex-peptides: seminal peptides of the *Drosophila* male. Cell Mol Life Sci. 60:1689–1704.
Kwon TH, Kim MS, Choi HW, Joo CH, Cho MY, Lee BL. 2000. A masquerade-like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, *Holotrichia diomphalia* larvae. Eur J Biochem. 267:6188–6196.
Lachaise D, Harry M, Solignac M, Lemeunier F, Bénassi V, Cariou ML. 2000. Evolutionary novelties in islands: *Drosophila santomea* a new melanogaster sister species from São Toḿ. Proc Biol Sci. 267:1487–1495.
Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. Forthcoming. Evolution of protein-coding genes in *Drosophila*. Trends genet.
Lawniczak MK, Begun DJ. 2007. Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. Mol Biol Evol. 24:1944–1951.
Lee KY, Zhang R, Kim MS, Park JW, Park HY, Kawabata S, Lee BL. 2002. A zymogen form of masquerade-like serine proteinase homologue is cleaved during pro-phenoloxidase activation by Ca2+ in coleopteran and *Tenebrio molitor* larvae. Eur J Biochem. 269:4375–4383.
Ligoxygakis P, Pelte N, Hoffmann JA, Reichhart JM. 2002. Activation of Drosophila Toll during fungal infection by a blood serine protease. Science. 297:114–116.
Ligoxygakis P, Pelte N, Ji C, Leclerc V, Duvic B, Belvin M, Jiang H, Hoffmann JA, Reichhart JM. 2002. A serpin mutant links Toll activation to melanization in the host defence of Drosophila. EMBO J. 21:6330–6337.
Lin CY, Hu KY, Ho SH, Song YL. 2006. Cloning and characterization of a shrimp clip domain serine protease homolog (c-SPH) as a cell adhesion molecule. Dev Comp Immunol. 30:1132–1144.
Liu H, Kubli E. 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 100:9929–9933.
Lung O, Kuo L, Wolfner MF. 2001. Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. J Insect Physiol. 47:617–622.
Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, Wolfner MF. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics. 160:211–224.
Mack PD, Kapelnikov A, Heifetz Y, Bender M. 2006. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. Proc Natl Acad Sci USA. 103:10358–10363.
Mbikay M, Tadros H, Ishida N, et al. 1997. Impaired fertility in mice deficient for the testicular germ-cell protease PC4. Proc Natl Acad Sci USA. 94:6842–6846.
McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK. 1988. The origin of MHC class II gene polymorphism within the genus Mus. Nature. 332:651–654.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 351:652–654.
Monsma SA, Wolfner MF. 1988. Structure and expression of a Drosophila male accessory gland gene whose product resembles a peptide pheromone precursor. Genes Dev. 2:1063–1073.
Mueller JL, Page JL, Wolfner MF. 2007. An ectopic expression screen reveals the protective and toxic effects of *Drosophila* seminal fluid proteins. Genetics. 175:777–783.
Mueller JL, Ram KR, McGraw LA, Bloch Qazi MC, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005. Cross-species comparison of Drosophila male accessory gland protein genes. Genetics. 171:131–143.
Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF. 2004. Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. Proc Natl Acad Sci USA. 101:13542–13547.
Murer V, Spetz JF, Hengst U, Altrogge LM, de Agostini A, Monard D. 2001. Male fertility defects in mice lacking the serine protease inhibitor protease nexin-1. Proc Natl Acad Sci USA. 98:3029–3033.
Neubaum DM, Wolfner MF. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics. 153:845–857.
Nie G, Li Y, Wang M, Liu YX, Findlay JK, Salamonsen LA. 2005. Inhibiting uterine PC6 blocks embryo implantation: an obligatory role for a proprotein convertase in fertility. Biol Reprod. 72:1029–1036.
Panhuis TM, Clark NL, Swanson WJ. 2006. Rapid evolution of reproductive proteins in abalone and *Drosophila*. Philos Trans R Soc Lond B Biol Sci. 361:261–268.
Panhuis TM, Swanson WJ. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. Genetics. 173:2039–2047.
Park M, Wolfner MF. 1995. Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. Dev Biol. 171:694–702.
Poiani A. 2006. Complexity of seminal fluid: a review. Behav Ecol Sociobiol. 60:289–310.
Pollard DA, Iyer VN, Moses AM, Eisen MB. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2:e173.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.
Pond SL, Muse SV. 2005. Site-to-site variation of synonymous substitution rates. Mol Biol Evol. 22:2375–2385.

Pool JE, Aquadro CF. 2006. History and structure of sub-Saharan populations of *Drosophila melanogaster*. Genetics. 174: 915–929.

Ravi Ram K, Sirot LK, Wolfner MF. 2006. Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 103:18674–18679.

Ravi Ram K, Wolfner MF. 2007a. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. Integr Comp Biol. 47:427–445.

Ravi Ram K, Wolfner MF. 2007b. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. PLoS Genet. 3:e238.

Ross J, Jiang H, Kanost MR, Wang Y. 2003. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene. 304:117–131.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Samakovlis C, Kylsten P, Kimbrell DA, Engstrom A, Hultmark D. 1991. The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. EMBO J. 10:163–169.

Saudan P, Hauck K, Soller M, et al. 2002. Ductus ejaculatorius peptide 99B (DUP99B), a novel *Drosophila melanogaster* sex-peptide pheromone. Eur J Biochem. 269:989–997.

Schully SD, Hellberg ME. 2006. Positive selection on nucleotide substitutions and indels in accessory gland proteins of the *Drosophila pseudoobscura* subgroup. J Mol Evol. 62:793–802.

Sim RB, Tsiftsoglou SA. 2004. Proteases of the complement system. Biochem Soc Trans. 32:21–27.

Simmons L. 2001. Sperm competition and its evolutionary consequences. Princeton (NJ): Princeton University Press.

Stevison LS, Counterman BA, Noor MA. 2004. Molecular evolution of X-linked accessory gland proteins in *Drosophila pseudoobscura*. J Hered. 95:114–118.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci USA. 98:7375–7379.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3:137–144.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics. 168:1457–1465.

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

Thornton KR, Jensen JD, Becquet C, Andolfatto P. 2007. Progress and prospects in mapping recent selection in the genome. Heredity. 98:340–348.

Tram U, Wolfner MF. 1998. Seminal fluid regulation of female sexual attractiveness in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 95:4051–4054.

Tsaur SC, Ting CT, Wu CI. 1998. Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of *Drosophila*: II. Divergence versus polymorphism. Mol Biol Evol. 15:1040–1046.

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

Uhrin P, Schofer C, Zaujec J, et al. 2007. Male fertility and protein C inhibitor/plasminogen activator inhibitor-3 (PCI): localization of PCI in mouse testis and failure of single plasminogen activator knockout to restore spermatogenesis in PCI-deficient mice. Fertil Steril. 88:1049–1057.

Wagstaff BJ, Begun DJ. 2005a. Comparative genomics of accessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. Mol Biol Evol. 22:818–832.

Wagstaff BJ, Begun DJ. 2005b. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. Genetics. 171:1083–1101.

Watorek W. 2003. Azurocidin—inactive serine proteinase homolog acting as a multifunctional inflammatory mediator. Acta Biochim Pol. 50:743–752.

Wigby S, Chapman T. 2005. Sex peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol. 15:316–321.

Wolfner MF. 2002. The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. Heredity. 88:85–93.

Wolfner MF, Heifetz Y, Applebaum SW. 2005. Gonadal glands and their gene products. In: Gilbert LI, Iatrou K, Gill SS, editors. Comprehensive molecular insect science: reproduction and development. Oxford: Elsevier Ltd. p.179–212.

Wong A, Albright SN, Wolfner MF. 2006. Evidence for structural constraint on ovulin, a rapidly evolving *Drosophila melanogaster* seminal protein. Proc Natl Acad Sci USA. 103:18644–18649.

Wong A, Jensen AD, Pool JE, Aquadro CF. 2007. Phylogenetic incongruence in the *Drosophila melanogaster* species group. Mol Phylogenet Evol. 43:1138–1150.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics. 168:1071–1076.

Xue L, Noll M. 2000. *Drosophila* female sexual behavior induced by sterile males showing copulation complementation. Proc Natl Acad Sci USA. 97:3272–3275.

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

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 155:431–449.

Yu XQ, Jiang H, Wang Y, Kanost MR. 2003. Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. Insect Biochem Mol Biol. 33:197–208.

Diethard Tautz, Associate Editor

Accepted November 19, 2007