Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*

Mara K. N. Lawniczak and David J. Begun

Center for Population Biology, Section of Evolution and Ecology, University of California, Davis

Population genetic analyses have shown that directional selection causes amino acid substitution in several seminal fluid proteins (Acps) and that in general, Acps tend to diverge rapidly. If rapid, adaptive divergence of such male reproduction-related genes is driven by sexual conflict, we might also expect to observe rapid, adaptive evolution in female reproduction-related genes, especially those mediating conflicts between the sexes. Female expressed genes differentially expressed shortly after mating were recently identified using whole genome expression micro-arrays. Such genes may play roles in storing sperm and mediating effects of seminal fluid proteins. Here, we report the results of a molecular population genetic survey from five female reproductive tract expressed serine proteases that show increased transcription shortly after mating. These genes are evolving rapidly, in some cases under directional selection, consistent with models of conflict.

Introduction

Many species show patterns of rapid evolution in reproduction-related genes (reviewed in Birkhead and Pizzari 2002, Swanson and Vacquier 2002). Under sexual conflict scenarios, in which there is ongoing adaptation and counter-adaptation between the sexes, rapid adaptive evolution of both male and female reproduction-related genes is predicted (Civetta 2003). Several lines of evidence suggest that sexual conflict occurs in *Drosophila melanogaster* and that it might play a role in the evolution of reproduction-related genes. For example, experimental evolution studies suggest that males can evolve increased or decreased seminal fluid toxicity depending on the mating system and that females can evolve resistance to the toxic effects of males (Rice 1996; Holland and Rice 1999). The fact that seminal fluid toxicity is positively correlated with increased sperm defense ability also suggests the potential for conflict—fitness gains for males may occur at fitness costs to females (Civetta and Clark 2000). Such effects may be mediated through seminal fluid proteins (e.g., Chapman et al. 1995, Wigby and Chapman 2005). Population genetic analyses of some *melanogaster* subgroup seminal fluid genes (*Acp26Aa*, *Acp29AB*, *Lectin29Ca*, and *Acp36DE*) have detected evidence for adaptive protein evolution (Aguade 1992, 1998, 1999; Tsaur et al. 1998; Begun et al. 2000; Holloway and Begun 2004).

Given models of male-female coevolution, a basic empirical question is whether the high levels of protein variation and evidence of adaptive divergence observed at *Acps* are also characteristic of genes that might be particularly important in female responses to and interactions with seminal fluid proteins and sperm (Bloch Qazi et al. 2003, Civetta 2003). A collection of female-expressed genes plausibly involved in male–female postcopulatory interactions is required to begin investigating patterns of molecular evolution in such genes and contrasting these patterns to those observed in *Acps*.

Recent experimental approaches for identifying female-expressed genes mediating male–female interactions have used EST collections and micro-array analyses of recently mated female flies or tissues dissected from such flies (Lawniczak & Begun 2004; McGraw et al. 2004; Swanson et al. 2004; Mack et al. 2006). Molecular evolutionary analysis of genes identified from these experiments provide support for rapid protein evolution of these genes (Swanson et al. 2004; Panhuis and Swanson 2006); however, there are still too few data to make strong generalizations.

Here, we report the results of our molecular population genetic analyses of five female expressed mating-induced serine proteases and one mating-induced gene, which though not a protease, is physically located in a small cluster of mating-induced proteases. We focus on these genes for a number of reasons. First, serine proteases constitute a well studied enzyme family which is known to be involved in diverse biological roles including digestion, development, degradation, coagulation, immunity, and fertilization (Krem and Di Cera 2002; Ross et al. 2003). The catalytic domains and substrate binding regions of serine proteases are known; specific functions of several particular residues are understood, potentially allowing functional data to inform our population genetic analyses (Perona and Craik 1995; Krem et al. 2000; Krem and Di Cera 2001) and facilitating future functional investigation of interesting natural variants. Second, several Acps are cleaved subsequent to transfer to females (Monsma et al. 1990; Bertram et al. 1996; Neubaum and Wolfner 1999; Ravi Ram and Wolfner 2005). In at least one case, cleavage of the Acp requires female-derived proteins, likely proteases (Park and Wolfner 1995). More generally, Drosophila males transfer many serine protease and serine protease inhibitors (known as serpins) in the seminal fluid (Swanson et al. 2001). These data suggest the possibility of a male–female molecular arms race in which male-derived serpins inhibit female-derived proteases. Interestingly, proteolytic processing was the second most overrepresented gene ontology category in female somatic tissues (Parisi et al. 2004).

The genes investigated here are the serine proteases *CG17240*, *CG17239*, *CG17234*, *CG17012* and *CG18125*, and an unknown gene, *CG31686*. Our comparison of gene expression in mated vs. unmated females revealed that *CG17012* and *CG18125* showed strong mating induction (Lawniczak and Begun 2004). In a subsequent, similar experiment (unpublished data) using the Affymetrix Dros2 GeneChip, which has better transcript coverage, the

Key words: reproduction, accessory gland proteins, mating, serine proteases.

E-mail: marakat@gmail.com.

Mol. Biol. Evol. 24(9):1944–1951. 2007
doi:10.1093/molbev/msm122
Advance Access publication June 14, 2007

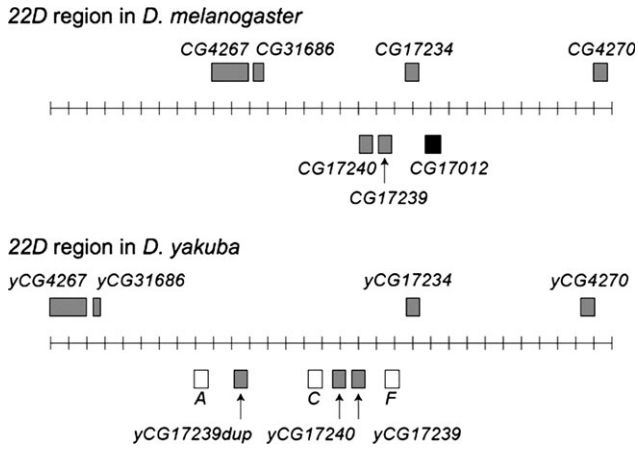

FIG. 1.—Comparison between *D. melanogaster* and *D. yakuba* of the *22D* region. The region is expanded in *D. yakuba* relative to *D. melanogaster*. CG numbers are present below genes. Three proteases (labeled *A*, *C*, *F*) are present only in *yakuba* and *CG17012* is present only in *D. melanogaster* (and *D. simulans*, not shown).

4 most significant and dramatically mating-induced genes included *CG18125*, 2 additional serine proteases, *CG17239* and *CG17234*, and a fourth gene *CG31686* which has no known functional domain. An independent micro-array study detected a fourth mating-induced serine protease, *CG17240* (McGraw et al. 2004). *CG31686*, *CG17240, CG17239*, *CG17234*, and *CG17012* are tandemly arranged in polytene chromosome band *22D* (figure 1). Physical clustering of serine proteases in the genome may reflect functional aspects of co-regulation of proteases involved in the same cascade (Yousef et al. 2003). Both genomic location and sequence similarity indicate evolutionary relationships among these serine proteases. Indeed, *CG17239* and *CG17240* (*Ser12*) are more similar to each other than either is to any other gene in the genome (54% amino acid identity). In addition, the predicted protein of *CG18125* (located at band *35A*) is more similar to that of *CG17234* (59% amino acid identity) than to any other gene in the genome. *CG18125* also shows 47% identity to *CG17012*.

The fact that all the proteases examined here show predicted signal sequences suggests they are likely secreted and thus are plausible candidates for participation in direct interactions with male-derived proteins. Here we present gene expression experiments and molecular population genetic analyses addressing the biology and evolution of these 6 genes.

Materials and Methods

Tissue *in situ* Hybridizations

Our initial micro-array results showing post-mating up-regulation of serine proteases, including *CG18125*, were from RNA isolated from whole females. Under the premise that more refined gene expression data could help narrow down the range of biological function, we used *in situ* hybridization to investigate tissue-specific expression of *CG18125*. Dissections of reproductive tracts of identically aged virgin and mated *D. melanogaster* and *D. simulans* females were performed 2 hours post-mating. For each species, a third *in situ* was completed on whole male and female flies with the abdomen and thorax slit. The detailed *in situ* protocol, which largely followed the Boehringer manual, is available upon request. Briefly, tissues were fixed overnight at 4°C in 4% paraformaldehyde in 1X PBS made with DEPC-H2O. *CG18125* PCR product (514 bp) was cloned into the PCR Blunt-II TOPO vector (Invitrogen). RNA sense and antisense probes were made using the MAXIscript kit (Ambion) and digoxigenin labeled ribonucleotides (Roche). Probe was hydrolyzed into ~75 base pair fragments. Anti-digoxigenin antibody attached to alkaline phosphatase was used to detect transcript.

Quantitative Real-Time PCR

Using Trizol reagent (GIBCO), total RNA was extracted from 2 independent replicates of mated whole flies, virgin whole flies, mated female carcasses with ovaries but without the remainder of the reproductive tract, and mated female sperm storage organs and parovaria. RNA was purified and DNased using the RNeasy kit and Rnase-free DNase set (Qiagen). cDNA was made using first-strand Taqman RT reagents (ABI). SYBR green PCR mix (ABI) was used for cDNA detection using primers designed on ABI's Primer Express 2.0 software (primer sequences available upon request). Each reaction was run at a 25 ul volume. For each RNA prep/primer pair combination, 2 or 3 replicate SYBR green reactions were completed. DNA contamination and primer-dimer were controlled for by carrying out 2 replicates of minus reverse transcriptase (−RT) reactions and 2 replicates of no template control (NTC). Additionally, melting curve analysis was used at the end of each run to verify product specificity. ABI SDS software was used for visualization and quantification. Baseline and threshold values were appropriate at the software default levels and were not adjusted. Estimates of transcript abundances for genes of interest were normalized relative to estimates from the housekeeping gene *Gpdh*, on a per plate basis, in order to account for variation in absolute amounts of RNA. This was accomplished by subtracting the housekeeping gene average C_t (across replicate wells) from the gene-of-interest average C_t. We then averaged the RNA preps together, subtracted this mated female average C_t from the average C_t of the virgin RNA preps, and used the delta-delta Ct method to estimate fold change (Livak and Schmittgen 2001).

Sequencing and DNA Analysis

D. melanogaster alleles from *CG31686, CG17240*, *CG17239*, *CG17234*, *CG17012* and *CG18125* were sequenced from up to 11 inbred stocks originating from Malawi, 3 from Zimbabwe, 7 from California, and 3 from North Carolina. *D. simulans* sequence data are from 8 highly inbred lines collected at Wolfskill Orchard in Winters, California (Begun and Whitley 2000). *D. yakuba* data are from genome assembly version 1.0 (Washington University Medical School Genome Sequencing Center), or from our own sequencing of strain Tai18E2, the strain used in the *D. yakuba* genome project. PCR products were

purified using Qiagen's PCR purification kit and sequenced at the UC Davis Division of Biological Sciences sequencing facility on an ABI 3700 machine. Sequence editing and assembly was completed using Phred, Phrap, and Consed Software (http://www.phrap.org). Sequences can be found under GenBank accession numbers AM765857-AM765991. DNAsp version 4.0 (Rozas and Rozas 1999) was used to analyze polymorphism and divergence. Only bases alignable within and between species were analyzed. *CG18125* had a segregating early termination codon that was 15 bp upstream of the more common termination codon. Although this termination codon was found in only two alleles (one from North America and one from Africa), all polymorphism and divergence data 3′ of this codon were excluded. HKA tests of serine protease datasets were compared to data from *vermilion*, which was sampled from N. American and African populations of *D. melanogaster* and *D. simulans* and showed no evidence of departures from the neutral, equilibrium model (Begun and Aquadro 1995).

Investigation of the polytene band *22D* region in both *D. yakuba* and *D. melanogaster*, was carried out by reciprocal BLAST analysis using *D. melanogaster* nucleotide coding sequences from the 5 genes in this region, plus the flanking genes *CG4267* and *CG4270*; these 7 genes constitute a *D. melanogaster* syntenic block of approximately 22kb. Putative orthologs in *D. yakuba* were confirmed by syntenic alignments and reciprocal BLAST analysis of coding + flanking sequence from *D. melanogaster* and *D. yakuba*. *D. yakuba* orthologs were aligned to *D. melanogaster* genes using ClustalW (default parameters).

Results

Tissue *in situ* Hybridizations

Results of the tissue *in situ* hybridization of *CG18125* probe to virgin and mated reproductive tracts indicate that the gene is highly induced in the spermathecae of recently mated females. No transcript was detected in whole males or whole females (without the reproductive tract) from either species, although this was a fairly crude assay. The control probe showed no staining to the glandular cells of the spermathecae. Experimental probe showed very light staining in virgin spermathecae and intense staining in the mated spermathecae of both *D. melanogaster* and *D. simulans* females (Supplementary figure 1). Both microarray and qPCR experiments on whole females indicate rapid induction of this gene in whole females shortly after mating (Lawniczak and Begun 2004; McGraw et al. 2004). The tissue *in situ* data presented here show that upregulation is of *CG18125* shortly after mating is spermatheca-specific.

Quantitative Real-Time PCR

Each of the genes examined here shows between 1.4 and 76-fold induction of gene expression just 3–6 hours post-mating in whole females (*CG17012* 1.4-fold: *CG17240* 2-fold; *CG31686* 6-fold; *CG17239* 9-fold; *CG17234* 53-fold; *CG18125* 76-fold). In addition, dissected reproductive tracts excluding ovaries of mated females were examined using qPCR and compared to the rest of the body without the reproductive tract. qPCR shows that the expression of all genes is either restricted to the reproductive tract or much more abundantly expressed in the sperm storage organs and/or paraovaria than the remainder of the body (data not presented).

Sequencing and DNA analysis

I. Polymorphism

Summaries of polymorphism are presented in table 1. Average levels of silent heterozygosity for these genes, which are located in regions of normal crossing-over on autosomes, are consistent with previous estimates in the 2 species (Andolfatto 2001). However, levels of replacement polymorphism in these genes are much greater than previous estimates from random samples of genes. For *D. melanogaster*, the average replacement heterozygosity for the genes investigated here is 0.005, compared to an average of 0.001 for a haphazard sample of autosomal genes (Andolfatto 2001). For *D. simulans*, the average replacement heterozygosity for the genes investigated here is 0.007 compared to a previous estimate of 0.002 (Andolfatto 2001). The high protein polymorphism observed in these mating-induced genes is similar in magnitude to that observed at *Acps* in *D. melanogaster* and *D. simulans* (Begun et al. 2000).

II. Divergence

Table 2 shows the *Ka* and *Ks* estimates for the comparison between *D. melanogaster* and *D. simulans* for the genes examined here. Silent divergence between *D. melanogaster* and *D. simulans* is slightly higher than the genome average ($Ks = 0.11$–0.13), though within the range previously observed in these species (Begun and Whitley 2000). However, the genes examined here show amino acid divergence that is several-fold higher than the average for these species ($Ka = 0.011$, Begun 2002).

Though rapid evolution of these female-expressed serine proteases might be expected under a sexual conflict scenario, the high rate of evolution observed for these genes might simply reflect a general property of serine protease evolution rather than a property of their putative role in reproduction. To address this possibility, mean *Ka*, *Ks*, and *Ka*/*Ks* for 166 serine proteases were compared to mean *Ka*, *Ks*, and *Ka*/*Ks* for all *D. melanogaster* genes using the whole genome sequences of *D. melanogaster* and *D. simulans*. Permutation analyses were carried out by randomly drawing 166 genes from the set of all genes (10,000 times) and determining how frequently the averages of *Ka*, *Ks*, and *Ka*/*Ks* of the drawn genes were greater than those of the serine proteases. Both silent sites and replacement sites show faster rates of evolution ($Ks = 0.1279$, $p = 0.0006$; $Ka = 0.0258$, $p < 0.0001$; 10,000 permuations); however, average *Ka*/*Ks* (0.203) is not significantly different from the average *Ka*/*Ks* across all genes ($p = 0.3722$). We also compared the 5 mating-induced serine proteases to

Table 1
Polymorphism Data

Gene	Spp	Pop	size	*n*	S	π_A	π_S	π_A/π_S
CG31686	*mel*		522	19	21	0.011	0.014	0.805
		Afr. *mel*		10	19	0.012	0.015	0.844
		N.A. *mel*		9	10	0.005	0.005	0.888
	sim		480	9	14	0.011	0.027	0.413
CG17240	*mel*		738	7	17	0.003	0.018	0.143
		Afr. *mel*		2	16	0.009	0.060	0.153
		N.A. *mel*		5	1	0	0.002	0
	sim		738	7	5	0.002	0.002	1.054
CG17239	*mel*		420	12	7	0.004	0.013	0.273
		Afr. *mel*		6	5	0.004	0.008	0.499
		N.A. *mel*		6	4	0.001	0.009	0.113
	sim		420	7	6	0.002	0.011	0.170
CG17234	*mel*		714	20	13	0.002	0.008	0.262
		Afr. *mel*		11	11	0.002	0.011	0.215
		N.A. *mel*		9	2	0.001	0	—
	sim		714	9	19	0.008	0.015	0.522
CG17012	*mel*		768	20	27	0.005	0.018	0.264
		Afr. *mel*		12	25	0.006	0.018	0.301
		N.A. *mel*		8	12	0.003	0.017	0.187
	sim		768	8	10	0.007	0.005	1.330
CG18125	*mel*		693	16	11	0.003	0.007	0.445
		Afr. *mel*		9	10	0.003	0.008	0.409
		N.A. *mel*		7	5	0.003	0.006	0.524
	sim		693	8	32	0.014	0.050	0.271
Gene average	*mel*					0.028	0.016	0.438
		Afr. *mel*				0.036	0.0240	0.484
		N.A. *mel*				0.013	0.0106	0.428
	sim					0.044	0.0220	0.752
Genomic average	*mel*					0.001	0.0090	0.111
	sim					0.002	0.0280	0.071

NOTE.—Size refers to the amount of sequence data gathered for each of *n* alleles. S is the number of segregating sites.

all 166 serine proteases using the same method and found that they are evolving significantly faster at replacement sites, but not silent sites ($Ka = 0.0841$, $p = 0.0009$; $Ks = 0.1200$, $p = 0.7077$) resulting in a significantly higher Ka/Ks (0.6945, $p < 0.0001$). Thus, although proteases tend to evolve more rapidly than other genes, mating-induced proteases are outliers even within the proteases, evolving particularly rapidly at amino acid sites.

More than 60% of the amino acid residues that determine substrate specificity in serine proteases are present in the 50 C-terminal residues (Krem et al. 2000, Rose and Di Cera 2002). In 3 of the 5 proteases examined here, the Ka/Ks ratio in the substrate specificity region (50 C-terminal residues) of the protein is near or greater than 1 and considerably higher than the ratio in the remainder of the protein (table 2) suggesting that rapid evolution may be related to altered substrate recognition for these genes. Additional support for this idea comes from the examination of particular residues known to be crucial in determining the substrate bond cleaved (e.g., trypsin vs. chymotrypsin specificity) as well as substrate specificity (Perona and Craik 1995). Residues known to be crucial for determining the type of substrate bond cleaved are highly variable between the three species for all genes except *CG17239* (figure 2).

Table 2
Divergence Between *D. melanogaster* and *D. simulans* for the Genes Examined Here

Gene	Whole gene			Substrate recognition region				Rest of protein			
	Ka	*Ks*	*Ka/Ks*	Coord	*Ka*	*Ks*	*Ka/ks*	Coord	*Ka*	*Ks*	*Ka/Ks*
CG31686	0.11	0.2	0.58	—	—	—		—	—	—	
CG17240	0.07	0.12	0.53	589–738	0.02	0.13	0.13	1–588	0.08	0.11	0.66
CG17239	0.05	0.14	0.33	271–420	0.04	0.18	0.19	1–270	0.05	0.10	0.47
CG17234	0.07	0.11	0.64	550–714	0.04	0.03	1.68	1–549	0.08	0.13	0.58
CG17012	0.12	0.16	0.79	619–768	0.12	0.12	0.95	1–618	0.11	0.15	0.76
CG18125	0.09	0.11	0.77	544–693	0.09	0.07	1.29	1–543	0.08	0.11	0.68
Gene average	0.10	0.17	0.73		0.08	0.13	1.06		0.10	0.15	0.79
Genomic average	0.02	0.11	0.14								

NOTE.—The proteases are further broken down into substrate recognition regions and the remainder of the protein. "Coord" refers to sequence position in the Genbank accession.

```
                          *
CG17240  mel  DSCHGDSGGP
         sim  DACRGDSGGP
         yak  DSCQGDSGGP

CG17239  all  DACSGDSGGP

CG17234  mel  TACHGDSGGP
         sim  TACLGDSGGP
         yak  TVCEGDSGGP

CG17012  mel  GGCYGDSGGP
         sim  AGCDGDSGGP

CG18125  mel  AACKGDSGGP
         sim  TACNGDSGGP
         yak  SACNGDSGGP
```

FIG. 2.—Amino acid residues surrounding the active site serine. Residues in bold are critical for determining the substrate amino acid at which the protease cleaves. Note how variable these residues are across the species for most of the proteases. * indicates the active site serine. Underlined residue in *CG17234* is polymorphic (L/H) in *D. simulans*.

This might suggest that the proteases are evolving rapidly to track changes in substrates.

McDonald-Kreitman Tests

Under the neutral model (Kimura 1983), the ratio of silent vs. replacement polymorphism should be similar to the ratio of silent vs. replacement divergence. We used the McDonald-Kreitman test (McDonald and Kreitman 1991) to test for deviations from neutral expectations. Polarized tests would be preferred, however, the high divergence between *D. yakuba* and *D. melanogaster*/*D. simulans* results in a large reduction of sites for which the ancestral state can be confidently assigned. Thus, unpolarized tests were performed. Three of 6 genes show evidence of directional selection (table 3).

CG17012 was recently reported to show evidence of directional selection using polymorphism data from a Riverside CA sample of *D. melanogaster* and a single *D. simulans* allele (Panhuis and Swanson 2006). Here, using polymorphism data from both species, we reject the neutral model in *D. melanogaster* but not *D. simulans*. However, because we cannot polarize the fixations (due to the absence of the gene in outgroup species), we cannot make incisive statements about the population genetic explanations for the data. *CG17240* shows evidence of directional selection only along the *D. melanogaster* lineage and *CG18125* shows evidence of directional selection only along the *D. simulans* lineage, but the polymorphism ratios in the non-significant tests for these loci show the same direction as the ratios in the significant tests, with fewer replacement than silent polymorphisms (table 3). These patterns suggest that these loci might be experiencing different evolutionary processes in these species, such as lineage-specific directional selection.

D. yakuba vs. *D. melanogaster* polytene band *22D* region

The region flanked by *CG4267* and *CG4270* in *D. melanogaster* contains 4 serine proteases plus *CG31686* over approximately 22 kb. The syntenic region in the *D. yakuba* genome is expanded relative to *D. melanogaster*, containing 7 rather than 4 serine proteases plus *CG31686*, over approximately 30 kb (figure 1). Of the four serine proteases of interest contained in this region in *D. melanogaster*, only *CG17240* has a single clearly orthologous copy in each species. *CG17239* shows an orthologous copy in *D. yakuba* (*yCG17239* in figure 1) but also appears to have a closely related duplicate (*yCG17239dup*) upstream. *D. yakuba* also contains 3 closely related copies of a serine protease that do not have a clear ortholog in *D. melanogaster* (A, C, and F in figure 1). These genes show the highest sequence similarity to *CG18125* and *CG17234*. However, orthologs of *CG18125* and *CG17234* are clearly present in the *D. yakuba* genome (although the *D. yakuba* ortholog of *CG17234* does not share the *D. melanogaster* canonical start codon). These data suggest that the three genes (A, C, and F) are absent from the *D. melanogaster* genome.

Interestingly, *CG17012* has no *D. yakuba* ortholog as it appears to be absent from *D. yakuba* as inferred from syntenic alignments to the genome assembly. Our PCR results support this conclusion, ruling out a *D. yakuba* genome assembly error as the explanation. PCR primers that in *D. melanogaster* amplify a 2.7kb product spanning the *CG17234*/*17012* region instead amplify in *D. yakuba* a 1.3kb product containing *CG17234* and the upstream flanking sequence of *CG17012*, but not *CG17012* itself.

Discussion

Though considerable evidence supports the idea that Drosophila seminal fluid protein genes may often evolve rapidly under the effects of directional selection (Begun et al. 2000; Swanson et al. 2001; Wagstaff and Begun 2005a; Wagstaff and Begun 2005b), female components of male-female interactions are more poorly described, functionally and evolutionarily (Civetta 2003; Swanson et al. 2004). Our results from *D. melanogaster* and *D. simulans*, which include elevated levels of replacement site heterozygosity, significant MK tests, increased rates of protein evolution in functionally important domains, and gene gains and losses, support the idea that directional selection plays a role in the divergence of genes up-regulated in female flies in response to mating.

Although the biological functions of the serine proteases examined here are not yet known, their strong induction in the reproductive tract shortly after mating as well as the highly spermatheca-specific expression documented here for *CG18125*, and the spermatheca and parovarium-specific expression documented previously (Arbeitman et al. 2004) for *CG17012* is suggestive of a role in female-ejaculate interaction. Seven Acps are known to enter the sperm storage organs (Bertram et al. 1996; Lung et al. 2002; Liu and Kubli 2003; Ravi Ram et al. 2005).

Table 3
McDonald–Kreitman Tests for Each of the Genes

Gene	Spp	bp	AA Fix	Sil Fix	AA Poly	Sil Poly	*p* value
CG31686	*mel*	564	30	15	14	3	ns
	sim	564			8	6	ns
CG17240	*mel*	738	30	15	5	12	0.009
	sim	738			3	1	ns
CG17239	*mel*	420	12	10	3	4	ns
	sim	420			2	4	ns
CG17234	*mel*	714	30	14	7	6	ns
	sim	714			13	7	ns
CG17012	*mel*	768	55	25	10	14	0.016
	sim	768			8	2	ns
CG18125	*mel*	693	31	11	7	5	ns
	sim	693			15	18	0.017

NOTE.—Fisher's exact test used to calculate *p*-value.

Additionally, four Acps are modified in the female, perhaps by proteolysis (Monsma et al. 1990; Bertram et al. 1996; Peng et al. 2005; Ravi Ram et al. 2005). Thus, these female-expressed proteases are good candidates for playing a role in cleavage of Acps.

Experimental support for the idea that female-expressed proteases act on Acps can be found in the biology of Sex Peptide. Sex Peptide is cleaved by a trypsin member of the serine protease family thereby maintaining its known effects for days after mating (Peng et al. 2005). Although it is not yet known which particular trypsin cleaves the C-terminal end of Sex Peptide off of stored sperm, the protease is likely to be female-derived because most Acps are not detectable longer than 6 hours post mating, yet cleavage of Sex Peptide from sperm appears to continue for as long as sperm are stored (up to several weeks) (Peng et al. 2005). Intriguingly, 4 of the 7 Acps known to enter the female sperm storage organs are serine protease inhibitors (serpins) and these 4 genes are also rapidly evolving (Mueller et al. 2005). Serpins irreversibly bind to serine proteases destroying their proteolytic function. Over evolutionary time, proteases may evolve to track the evolution of their substrate while the substrate evolves to escape proteolysis. Additionally or alternatively, proteases may evolve to escape serpin inhibition while serpins evolve to paralyze proteases. This type of antagonistic dynamic is consistent with rapid evolution at reproduction-related protease and serpin genes.

We note, however, that proteases are also known to function in triggering the immune signaling cascade that leads to antimicrobial expression (Ross et al. 2003). Thus, the serine proteases investigated here could play a role in suppressing pathogens or initiating a local female immune response. Indeed, several antimicrobials are mating induced, specifically by Sex Peptide, although the function and location of induction has not yet been determined (Peng et al. 2005; Lawniczak et al. 2007). Thus, we cannot rule out the possibility that fly-pathogen interactions play a role in the evolution of these genes. Functional data bearing on these proteases and their substrates would clarify evolutionary inferences and provide valuable guidance for designing functional investigation of natural variation. It is tempting to speculate that interactions between female derived proteases and male derived serpins might be a molecular component of an arms race between males and females, as each sex attempts to gain control over different aspects of fertilization.

Additional evidence for unusual evolution of these genes can be found in patterns of gene duplication and loss in *D. yakuba*. We found that one of the mating-induced *D. melanogaster* proteases (*CG17012*) is absent from the *D. yakuba* genome. This is in conflict with results from Panhuis and Swanson (2006). They reported discovering orthologs of *CG17012* in *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. virilis*, and *D. mojavensis*. However, we found that sequence divergence is so great between *D. melanogaster* and *D. ananassae* that producing a convincing alignment of the region is not possible. Although *D. ananassae* has 5 protease open reading frames in the homologous region (flanked by *CG4267* and *CG4270*), 2 show mutations in the active site sequences that likely render them unable to cleave; *CG17239* is the only gene in the region that is convincingly orthologous. Thus, *CG17012* might be a new gene in the *D. melanogaster* subgroup, it might have diverged so much as to be unrecognizable, or it might have been lost from *D. yakuba*. Intriguingly, Begun and Lindfors (2005) reported a case of an Acp that showed adaptive protein evolution in one lineage but that was lost from a closely related lineage. This provides another case of similar evolutionary dynamics of male seminal fluid proteins and female expressed mating-induced genes in Drosophila.

Our expression and population genetic analyses are an important first step. However, further progress in disentangling the forces controlling variation and divergence in female-expressed proteins interacting with male products will require a functional understanding of polymorphic and diverged mutations because different processes, such as cryptic female choice and sexual conflict, may result in similar population genetic patterns (i.e., rapid evolution in both sexes).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgments

We thank Mark Siegal for help with the *in situ* hybridization and Alisha Holloway for running the permutation tests. This work was funded by NIH Rol GM071926 to DJB.

Literature Cited

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

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

Aguade M, Miyashita N, Langley CH. 1992. Polymorphism and divergence in the Mst26a male accessory gland gene region in *Drosophila*. Genetics. 132:755–770.

Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol. 18:279–290.

Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS. 2004. A genomic analysis of *Drosophila* somatic sexual differentiation and its regulation. Devel. 131:2007–2021.

Begun DJ. 2002. Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. Mol Biol Evol. 19:201–203.

Begun DJ, Aquadro CF. 1995. Evolution at the tip and base of the X-chromosome in an African population of *Drosophila melanogaster*. Mol Biol Evol. 12:382–390.

Begun DJ, Lindfors HA. 2005. Rapid evolution of genomic Acp complement in the *melanogaster* subgroup of *Drosophila*. Mol Biol Evol. 22:201–221.

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

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.

Birkhead TR, Pizzari T. 2002. Postcopulatory sexual selection. Nat Rev Genet. 3:262–273.

Bloch Qazi MC, Heifetz Y, Wolfner MF. 2003. The developments between gametogenesis and fertilization: ovulation and female sperm storage in *Drosophila melanogaster*. Dev Biol. 256:195–211.

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.

Civetta A. 2003. Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. Genome. 46:925–929.

Civetta A, Clark AG. 2000. Correlated effects of sperm competition and postmating female mortality. Proc Natl Acad Sci USA. 97:13162–13165.

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

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

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Krem MM, Di Cera E. 2001. Molecular markers of serine protease evolution. Embo J. 20:3036–3045.

Krem MM, Di Cera E. 2002. Evolution of enzyme cascades from embryonic development to blood coagulation. Trends Biochem Sci. 27:67–74.

Krem MM, Rose T, Di Cera E. 2000. Sequence determinants of function and evolution in serine proteases. Trends Cardiovas Med. 10:171–176.

Lawniczak MKN, Barnes AI, Linklater JR, Boone JM, Wigby S, Chapman T. 2007. Mating and immunity in invertebrates. Trends Ecol Evol. 22:48–55.

Lawniczak MKN, Begun DJ. 2004. A genome-wide analysis of courting and mating responses in *Drosophila* melanogaster females. Genome. 47:900–910.

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.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25:402–408.

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.

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

McGraw LA, Gibson G, Clark AG, Wolfner MF. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. Curr Biol. 14:1509–1514.

Monsma SA, Harada HA, Wolfner MF. 1990. Synthesis of 2 *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. Dev Biol. 142:465–475.

Mueller JL, Ram KR, McGraw LA, Qazi MCB, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005. Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics. 171:131–143.

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

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.

Parisi M, Nuttal R, Edwards P, Minor J, Naiman D, Lu J, Doctolero M, Vainer M, Chan C, Malley J, Eastman S, Oliver B. 2004. A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. Genome Biol. 5:R40.

Peng J, Zipperlen P, Kubli E. 2005. *Drosophila* sex-peptide stimulates female innate immune system after mating via the Toll and Imd pathways. Curr Biol. 15:1690–1694.

Perona JJ, Craik CS. 1995. Structural basis of substrate-specificity in the serine proteases. Prot Sci. 4:337–360.

Ravi Ram K, Ji S, Wolfner MF. 2005. Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. Insect Biochem Mol Biol. 35:1059–1071.

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

Rose T, Di Cera E. 2002. Substrate recognition drives the evolution of serine proteases. J Biol Chem. 277:19243–19246.

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

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.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA. 98:2509–2514.

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.

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

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

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

Yousef GM, Kopolovic AD, Elliott MB, Diamandis EP. 2003. Genomic overview of serine proteases. Biochem Biophys Res Commun. 305:28–36.

Koichiro Tamura, Associate Editor

Accepted June 11, 2007