DOI: 10.1534/genetics.105.053611

Molecular Evolution and Population Genetic Analysis of Candidate Female Reproductive Genes in Drosophila

Tami M. Panhuis[1] and Willie J. Swanson

Department of Genome Sciences, University of Washington, Seattle, Washington 98195

Manuscript received November 18, 2005
Accepted for publication June 1, 2006

ABSTRACT

Molecular analyses in several taxa have consistently shown that genes involved in reproduction are rapidly evolving and subjected to positive selection. The mechanism behind this evolution is not clear, but several proposed hypotheses involve the coevolution between males and females. In Drosophila, several male reproductive proteins (Acps) involved in male–male and male–female interactions show evidence of rapid adaptive evolution. What has been missing from the Drosophila literature is the identification and analysis of female reproductive genes. Recently, an evolutionary expressed sequence tag analysis of Drosophila female reproductive tract genes identified 169 candidate female reproductive genes. Many of these candidate genes still await further molecular analysis and independent verification of positive selection. Our goal was to expand our understanding of the molecular evolution of Drosophila female reproductive genes with a detailed polymorphism and divergence study on seven additional candidate female reproductive genes and a reanalysis of two genes from the above study. We demonstrate that 6 candidate female genes of the 9 genes surveyed show evidence of positive selection using both polymorphism and divergence data. One of these proteins (CG17012) is modeled to reveal that the sites under selection fall around and within the active site of this protease, suggesting potential differences between species. We discuss our results in light of potential function as well as interaction with male reproductive proteins.

MOLECULAR analyses reveal that many proteins involved in reproduction are rapidly evolving. This phenomenon is observed for a number of different organisms, including marine invertebrates, mammals, plants, and insects (Swanson and Vacquier 2002; Clark *et al.* 2006). For example, one of the fastest evolving metazoan proteins is the abalone sperm protein lysin (Metz *et al.* 1998). In mammals, the egg coat protein ZP2 is among the most divergent molecules between human and rodent (Swanson *et al.* 2001a). In many cases, patterns of nucleotide variation within and between species suggest that natural selection has promoted the divergence of reproductive proteins (Swanson and Vacquier 2002; Yang 2005).

In Drosophila, patterns of nucleotide variation reveal rapid evolution by positive selection in a group of male reproductive proteins known as the accessory gland proteins (Acps) (Swanson *et al.* 2001b). Darwinian selection on Acps is potentially influenced by their role in mediating both male–male and male–female postmating interactions (Chapman *et al.* 2002). Acps are synthesized and secreted by cells in a paired reproductive structure called the accessory gland. These proteins are ejaculated into females during mating and affect sperm storage, sperm defense during sperm competition, female remating receptivity, egg-laying rate, and longevity (see review by Wolfner 1997). Additionally, several Acp proteins have been localized to specific regions in a mated female's reproductive tract (Bertram *et al.* 1996; Heifetz *et al.* 2000; Ram *et al.* 2005). The negative effect of male Acps on female fitness suggests females should evolve interacting proteins to Acps that might play a role in processes relating to male reproductive success. There is evidence that females affect male reproductive success, both in sperm competition and in the fecundity of a single mating (Price 1997; Clark and Begun 1998; Clark *et al.* 1999; T. M. Panhuis and L. Nunney, unpublished data). However, there is still little information on the evolution of female reproductive proteins (Swanson *et al.* 2004). Insight into the evolution of female reproductive proteins is the missing ingredient to a better understanding of coevolutionary processes involved in male–female postmating interactions (Panhuis *et al.* 2006), such as sexual selection and sexual conflict.

Recently, Swanson *et al.* (2004) identified candidate Drosophila female reproductive genes from an expressed sequence tag (EST) analysis. The screen for candidate genes was based on an evolutionary EST analysis and presence of a signal sequence (indicating a secreted protein). Candidate female genes may be those that show adaptive divergence and potential to be secreted.

[1]*Corresponding author:* Department of Biology, University of California, Riverside, CA 92521. E-mail: tpanhuis@jsd.claremont.edu

Adaptive divergence is the most common recurring observation for reproductive proteins and may be explained by selection pressures, such as sexual conflict, sexual selection, immune defense, and self *vs.* nonself recognition (Swanson and Vacquier 2002). Adaptive divergence was determined for female reproductive genes by generating ESTs from *Drosophila simulans* and aligning them to their putative orthologs in the completed *D. melanogaster* genome (Adams *et al.* 2000; Swanson *et al.* 2004). The overall analysis revealed that candidate genes, as a group, had a 50% increase in nonsynonymous sequence divergence compared to noncandidate reproductive proteins, and that synonymous sequence divergence was similar to the expected value for the species pair (Swanson *et al.* 2004). Swanson *et al.* (2004) also performed a polymorphism survey and divergence study on portions of nine candidate genes. This survey revealed positive selection in six of the nine genes independently surveyed.

The Swanson *et al.* (2004) study is an important first step in identifying potential female reproductive genes, but only 9 of the 169 candidate genes had been surveyed independently for positive selection. Thus, many of the candidate genes still await further molecular analysis and independent verification of positive selection. In this study, our goal was to expand our understanding of the molecular evolution of Drosophila female reproductive genes with a detailed polymorphism and divergence study on 7 additional candidate female reproductive genes and a reanalysis of 2 genes from Swanson *et al.* (2004). Unlike Swanson *et al.* (2004), whose polymorphism survey focused on a portion of a gene, we present nucleotide polymorphism surveys on the entire gene (coding and noncoding) for all 9 genes. We also assessed DNA sequence divergence among a number of increasingly divergent species of Drosophila by comparing rates of nonsynonymous and synonymous substitutions using PAML (Yang 2000). Our polymorphism and divergence results identified 6 of 9 genes evolving by positive selection. In combination with Swanson *et al.* (2004), this work identifies interesting candidate female reproductive genes to evaluate further in experimental and functional analyses from those identified in the original EST analysis. Future analysis biochemically identifying interacting male–female reproductive genes will be important.

MATERIALS AND METHODS

Fly stocks and DNA preparation: *D. melanogaster* flies used in this study came from 42 isofemale lines established by E. Warren and L. Nunney in 2000 from orange groves at the University of California, Riverside. Genomic DNA was extracted from 10–15 adults/line using Purgene DNA isolation kit from Gentra Systems (Minneapolis). Nine candidate female genes (CG10200, CG5106, CG5273, CG9897, CG13004, CG17108, CG17012, CG8453, CG5976) were sequenced. All genes, except CG17012, were picked from Swanson *et al.* (2004) on the basis of the presence of a signal sequence or on an overall ratio of nonsynonymous (d_N) to synonymous (d_S) changes >0.5 (a value predicted by the authors to be indicative of adaptive evolution). CG17012 was chosen as a female candidate gene on the basis of work by Arbeitman *et al.* (2004), who show this gene to be strongly expressed in the female spermatheca and parovaria, two regions of the female reproductive tract likely to interact with male reproductive proteins or sperm. For all genes, Table 1 shows the chromosome of each locus, the final number of isofemale lines sequenced per gene (N), the potential gene ontology function from FlyBase, the estimated EST d_N/d_S ratio from Swanson *et al.* (2004), and the presence or absence of a signal sequence. Polymerase chain reaction products were diluted fivefold with water, sequenced directly using ABI big dye terminator sequencing chemistry, and analyzed on an ABI 3100 automated sequencer. A consensus sequence for each gene was generated from visualizing both PCR product sequence strands using Sequencer 4.1 for Mac (Genecodes, Ann Arbor, MI). PCR primers and conditions are available from the authors upon request. Sequences are deposited in GenBank under accession nos. DQ539048–DQ539337.

Polymorphism survey and divergence study: We performed tests of neutrality for each gene including Tajima's D (Tajima 1989), Fu and Li's D (with an outgroup) and D^* (without an outgroup) (Fu and Li 1993), and Fay and Wu's H (Fay and Wu 2000) using DnaSP4.0 (Rozas and Rozas 1999). We used a combination of test statistics to reveal more about the pattern of selection (Otto 2000). Significance for all tests was determined by coalescent simulation with R (recombination) estimated from the data using Hudson (1987) and putative orthologous *D. simulans* sequences for an outgroup sequence in Fu and Li's D and Fay and Wu's H. Orthologous *D. simulans* sequences were determined from NCBI BLAST (Altschul *et al.* 1990) and aligned to the *D. melanogaster* sequences with Clustal W (Thompson *et al.* 1994). Each of the three tests analyze sequences on the basis of a site-frequency spectrum (proportion of alleles at high or intermediate *vs.* low frequencies) and determine if loci depart from the expectation of a neutral equilibrium model (Aquadro 1997). An excess of rare alleles is consistent with positive selection and is indicated by a negative Tajima's D and/or Fu and Li's D (Tajima 1989, Fu and Li 1993). An excess of high-frequency variation is consistent with balancing selection and is indicated by a positive Tajima's D and/or Fu and Li's D (Tajima 1989, Fu and Li 1993). Fay and Wu's H predicts positive selection on the basis of a site-frequency spectrum comparing the proportion of alleles at intermediate *vs.* high frequency and is indicated by a negative Fay and Wu's H (Fay and Wu 2000). Departures from neutrality are predicted to be associated with recent selection acting at or near the locus. To control for demographic effects, we also sequenced and analyzed five randomly chosen loci (CG12191, CG1124, CG11105, CG33554, CG3022) focusing on intronic regions. Intronic regions were the focus to increase the number of polymorphic sites. Sequence protocol was similar to that used for the candidate female genes.

A Hudson–Kreitman–Aquade test (Hudson *et al.* 1987) was used to compare the levels of polymorphism to divergence, as implemented in the HKA program for multiple loci by Jody Hey (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA). Under neutrality, the levels of polymorphism within a species and divergence between species should be proportional to the neutral mutation rate. A McDonald–Kreitman test (McDonald and Kreitman 1991) was performed for each gene individually and for all loci combined using DnaSP 4.0 (Rozas and Rozas 1999). The McDonald–Kreitman test tests the prediction that if both synonymous (silent, d_S) and

TABLE 1

General information on female candidate genes examined in this study

Gene	Position	*N*	GO function	EST d_N/d_S	Signal sequence
CG10200[a]	2R	25	Unknown	1.265	Yes
CG5106	3R	34	Unknown	2.2792	Yes
CG5273	X	28	Unknown	2.225	Yes
CG9897	2R	36	Protease	0.356	Yes
CG13004	X	36	Unknown	0.5358	No
CG17108[a]	2L	29	Unknown	0.3583	Yes
CG17012	2L	37	Protease	0.765[b]	Yes[b]
CG8453 (Cyp6g1)	2R	28	Oxidative enzyme	0.4767	No
CG5976	3L	37	Glutaminyl cyclase	1.0594	Yes

Position refers to the chromosomal position of each locus. EST d_N/d_S ratio is from Swanson *et al.* 2004. Signal sequence indicates the presence of a signal sequence. *N*, number of isofemale lines sequenced per gene; GO function, gene ontology function (Ashburner *et al.* 2000).

[a] Candidate genes reanalyzed from Swanson *et al.* 2004.

[b] Ratio determined from entire gene using DnaSP and signal sequence predicted from SignalP (http://www.cbs.dtu.dk/services/SignalP-2.0; Nielsen *et al.* 1997).

nonsynonymous (replacement, d_N) mutations are neutral, then the ratio of synonymous to nonsynonymous polymorphism within a species will be similar to the ratio of synonymous to nonsynonymous divergence between species (fixed differences). Statistical departure from neutrality was tested with a G-test on a 2×2 contingency table of silent and replacement fixed differences between species and silent and replacement polymorphic changes within *D. melanogaster*. Polymorphism data is from the *D. melanogaster* sequences after alignment with the putative orthologous *D. simulans* sequence. We also report the neutrality index (NI) (Rand and Kann 1996), which shows the directionality of the McDonald–Kreitman test. An NI value >1 is consistent with negative selection, while an NI value <1 is consistent with positive selection. McDonald–Kreitman tests were not performed on the neutral loci because coding regions were not sequenced.

For the divergence study, we used BLAST to determine putative orthologous sequences from several Drosophila species (*D. simulans*, *D. yakuba*, *D. anannasse*, *D. pseudoobscura*, *D. virilis*, and *D. mojavensis*) for all female genes except CG17108 (resulting in a total of eight genes analyzed in the divergence study). CG17108 was not used due to its biased amino acid and codon usage, which may induce errors in parameter estimates that use codon models (Swanson *et al.* 2004). *D. pseudoobscura* putative orthologs could not be identified for CG5273 and CG5106, and *D. virilis* putative orthologs could not be identified for CG9897, CG10200, and CG13004. Sequences were aligned using MEGA 3.1 (Kumar *et al.* 2004) and Se-Al v2.0 (Rambaut 1996) and analyzed in the phylogenetic analysis by maximum likelihood program (PAML, Yang 2000). PAML tests for positive selection using a likelihood ratio test that compares a null (neutral) model where no codons could have a d_N/d_S ratio >1 (L_0) with the likelihood of a model in which a subset of sites could have a d_N/d_S ratio >1 (L_1) (Nielsen and Yang 1998; Yang and Bielawski 2000; Yang *et al.* 2000). Statistical significance is calculated by the negative of twice the difference in the log-likelihood obtained from these two models ($-2[\log(L_0) - \log(L_1)]$) compared to a chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters (Swanson *et al.* 2004). We examined variation in the d_N/d_S ratio between sites using both discrete (PAML models M0 and M3) and β- (PAML models M7 and M8) distributions. The comparison of M0 and M3 is not a robust test of adaptive evolution, but tests for variation in the d_N/d_S ratio between sites (Swanson *et al.* 2004). The M7 and M8 comparison is a robust test of adaptive evolution. Details on the test statistics and distributions can be found at Yang *et al.* (2000). Sites predicted to be subjected to positive selection, using a Bayes empirical approach (Yang *et al.* 2005), were mapped onto a three-dimensional structure for CG17012 predicted using SwissModel (Schwede *et al.* 2003).

RESULTS

Tajima's *D*, Fu and Li's *D and *D*, and Fay and Wu's *H*:** To test if any of the nine candidate genes have been subjected to positive selection, we performed several different tests of neutrality. Table 2 shows that levels of polymorphism in five of the nine genes surveyed depart significantly from equilibrium neutral expectations. Genes CG5273 and CG9897 show significant departure from neutral expectations on the basis of Tajima's *D* (Tajima 1989) and/or Fu and Li's *D** (Fu and Li 1993). A significant negative Tajima's *D*, Fu and Li's *D** or *D* value (as in the case of CG5273) indicates an excess of rare alleles, which may result from a recent selective sweep. When a selective sweep occurs, in the presence of recombination, linked variation is dragged toward fixation, which results in an excess of high-frequency-derived mutations in regions near the site of selection. Polymorphism at sites near the selected site is eliminated as the favored variant is fixed. After the sweep, an excess of rare alleles is observed as new mutations occur in the targeted region and drift upward in frequency, since every new mutation produces a new allele. The time to return to levels expected under neutral equilibrium will depend on the population size and can be slow to occur for large populations. A significantly positive Fu and Li's *D** (without an outgroup) for CG9897 suggest that this gene may be experiencing balancing selection, since balancing selection maintains mutations at intermediate frequencies.

TABLE 2

Five candidate genes show an indication of positive selection on the basis of polymorphism surveys

Gene	Chromosome	*N*	bp Riv	π	θ	Taj. *D*	F&L *D**	bp sim&Riv	F&L *D*	F&W *H*
Candidate genes										
CG10200	2R	25	1390	0.0029	0.0044	−1.19	−0.43	847	−0.99	−2.64
CG5106	3R	34	1071	0.0042	0.0069	−1.37	−1.26	857	−0.19	−18.52***
CG5273	X	28	1765	0.0005	0.0016	−2.14**	−2.98*	1181	−2.09*	0.58
CG9897	2R	36	840	0.0103	0.0089	0.56	1.54*	840	1.73	−3.35
CG13004	X	36	975	0.0023	0.0035	−1.05	0.21	624	−0.76	−0.83
CG17108	2L	29	1166	0.0011	0.0018	−1.16	−0.56	1123	−0.66	−3.45*
CG17012	2L	37	902	0.0041	0.0043	−0.14	−0.02	869	−0.19	−2.86
CG8453	2R	28	1974	0.0017	0.0017	−0.09	−0.66	1951	−0.89	−5.93**
CG5976	3L	37	1530	0.0009	0.0014	−1.17	−1.71	1528	−1.22	−1.44
Neutral loci										
CG12191	3L	30	446	0.0047	0.0057	−0.75	−0.75	443	−0.68	0.48
CG1124	3R	29	488	0.0016	0.0008	−1.18	−1.46	488	−1.55	−1.16
CG11105	X	42	679	0.0044	0.0041	0.22	−1.11	679	−1.23	−0.91
CG33554	2R	18	499	0.0054	0.0041	1.12	1.31	463	1.14	1.01
CG3022	2L	32	517	0.0017	0.0029	−1.41	−1.14	516	−1.26	0.63

N, number of isofemale lines analyzed for each gene; bp Riv, number of base pairs from Riverside sequences used for Taj. *D* and F&L *D** analyses (excludes sites with gaps or missing data); π and θ, nucleotide diversity; Taj. *D*, Tajima's *D*; F&L *D**, Fu and Li's *D** (without an outgroup sequence); bp sim&Riv, number of base pairs from data set with *D. melanogaster* Riverside sequences and *D. simulans* sequence used for F&L *D* and F&W *H* (excludes sites with gaps or missing data); F&L *D*, Fu and Li's *D* (with *D. simulans* sequence as an outgroup); F&W *H*, Fay and Wu's *H* (with *D. simulans* sequence as an outgroup). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Three additional genes show positive selection on the basis of Fay and Wu's *H* when a putative orthologous *D. simulans* sequence is used as an outgroup. An outgroup sequence is required to infer the ancestral and derived states at segregating sites. These genes are CG5106, CG17108, and CG8453. A negative *H* value is indicative of genetic hitchhiking. Genetic hitchhiking occurs when a neutral mutation is tightly linked to a locus under positive selection. This linkage allows the mutation to rise to high frequency, which is indicated by a significantly negative *H* value (Fay and Wu 2000).

Statistical tests of neutrality on the basis of the site frequency spectrum are known to be confounded by demographic processes. For example, Tajima's *D* and Fu and Li's *D* can be affected by population bottlenecks and migration (Tajima 1989; Fu and Li 1993). Additionally, population admixture can result in significant Fay and Wu's *H* values (Przeworski 2002). Therefore, care needs to be taken to account for demographic effects, particularly in derived (non-African) populations of *D. melanogaster* (as studied herein) where a bottleneck model is the appropriate demographic null model (Haddrill *et al.* 2005; Thornton and Andolfatto 2006). We took two approaches to control for demographic effects. First, under a simple neutral model, selection is often thought to be a locus-specific effect while demographics may affect the entire genome; therefore, we sequenced five randomly chosen loci (Table 2) from the same individuals used for the candidate female genes. Consistent with a lack of demographic effects, none of these randomly chosen genes showed a departure from neutrality (Table 2). However, it should be noted that certain demographic scenarios may increase the variance between loci (Nielsen 2005). Therefore, our second approach (see below) compares synonymous and nonsynonymous changes using the McDonald–Kreitman test (McDonald and Kreitman 1991) and divergence between species (Nielsen and Yang 1998; Yang *et al.* 2000), both of which are not confounded by demographic effects (Nielsen 2005).

Hudson–Kreitman–Aquadé and McDonald–Kreitman test: We used two tests that compare the ratio of polymorphisms within species to divergence between species. First, we used a multilocus Hudson–Kreitman–Aquadé (HKA) (Hudson *et al.* 1987) test to assess departures from a neutral model (Hammer *et al.* 2004). We found significant departure from a neutral model ($P < 0.001$), with most loci showing a deficit of polymorphism. One exception was CG9897, in which we observed twice as many polymorphisms as expected. Along with the positive Fu and Li *D** test (Table 2), this observation from the HKA test is consistent with CG9897 being subjected to some form of balancing selection.

For our second test comparing polymorphism to divergence, we used a McDonald–Kreitman test (McDonald and Kreitman 1991), which uses information from both silent (synonymous) and replacement (nonsynonymous) sites to test for positive selection (whereas the neutrality tests used in the frequency spectrum and HKA test do not distinguish between silent and replacement sites). Unlike the frequency spectrum-based tests and HKA test above, the results of a McDonald–Kreitman test are not confounded by demographic effects (Nielsen 2005). This analysis depends on the

TABLE 3

Contingency tables and G-test results for silent and replacement polymorphic and fixed variation in candidate female genes using *D. melanogaster* (Riverside) sequences and a putative orthologous *D. simulans* sequence

	Polymorphic		Fixed			
Gene	Silent	Replacement	Silent	Replacement	NI	*P*-value[a]
CG10200	8	2	50	47	0.266	0.074
CG5106	21	5	28	4	1.667	0.483
CG5273	1	3	5	20	0.750	0.822
CG9897	19	12	22	17	0.817	0.680
CG13004	6	1	19	20	0.158	0.056
CG17108	5	1	64	45	0.284	0.205
CG17012	9	5	27	59	0.254	*0.020*
CG8453	8	1	54	19	0.355	0.289
CG5976	1	2	22	8	5.500	0.171
All	78	32	291	239	0.500	*0.002*
All but CG17102	69	27	264	180	0.574	*0.021*

NI refers to the neutrality index (Rand and Kann 1996). "All" is a combined G-test on all nine loci and "All but CG17012" is an analysis with all loci except CG17012, which showed the greatest departures from homogeneity in the individual tests (see materials and methods).

[a] Probability determined by a G-test. Values in italics indicate significance at $P < 0.05$.

prediction that under neutrality, the number of substitutions between two species and the number of polymorphic changes within a species will both be proportional to the mutation rate, and this will be true for both silent and replacement nucleotide changes (Otto 2000). Table 3 shows silent and replacement variation and the NI (Rand and Kann 1996) in all female genes for polymorphisms from *D. melanogaster* sequences and fixed differences between *D. melanogaster* and a putative orthologous *D. simulans* sequence. A G-test revealed that CG17012 deviates significantly from homogeneity and has an NI value less than one, which is consistent with positive selection. When all loci are combined there is significant heterogeneity (Table 3). This result may be due to the significant deviation from homogeneity in CG17012. When this locus is removed from the analysis, the analysis remains significant (Table 3). This indicates that overall these female genes have been subjected to positive selection (as indicated by the NI value less than one and significant G-test).

Divergence study: Unlike the neutrality tests on the basis of polymorphism data, which detect recent selection in a single species, the divergence analysis detects recurrent positive selection events on the same codons of several species. We found a significant amount of variation in the d_N/d_S ratio between sites for all eight genes analyzed with the discrete model M3 (Table 4, M0 *vs.* M3). Three of these genes have a class of sites with a $d_N/d_S > 1$ (Table 4). This discrete model, however, is only an indication of variable d_N/d_S ratios between sites and should not be used as a robust test of adaptive evolution (Swanson *et al.* 2001a). A more robust test of adaptive evolution is one with a beta distribution of d_N/d_S for "neutral" or functionally constrained codons that covers the interval 0–1 and permits a class of sites to vary freely, including $d_N/d_S > 1$ (model M8, Table 4). Evidence of selection acting on a subset of codons was found for four of the eight genes using this model (Table 4, M7 *vs.* M8). For one of the genes, CG17012, we were able to map residues predicted to be under positive selection onto a three-dimensional model (Figure 1). The sites under positive selection fall around the active site of this protease, suggesting potential functional differences between species. For example, these methods have been used to identify sites under positive selection that effectively predict the location of the binding sites in MHC (Yang and Swanson 2002), abalone lysin (Yang and Swanson 2002), and mammalian egg coat proteins (Swanson *et al.* 2001a).

DISCUSSION

Rapid reproductive protein evolution is emerging as a common phenomenon among many taxa. Much of this evolution appears to be adaptive. Sperm–egg binding partners in mammals and abalone, for instance, show rapid adaptive evolution in both the male and the female proteins involved in the interaction (Swanson *et al.* 2001a; Jansa *et al.* 2003; Panhuis *et al.* 2006), suggesting the selective pressure relates to a coevolutionary process between the male and the female. In Drosophila, the outcome of male–female postmating interactions reveals that male Acps are influential in altering female physiology and behavior and that female genotype contributes to the reproductive success of male gametes (Wolfner 1997, 2002). Male proteins have been studied in great detail and several Acps are rapidly evolving by adaptive evolution. One explanation for this rapid evolution is coevolution between the sexes; however, several other evolutionary hypotheses

TABLE 4

Positive selection detected for female candidate reproductive genes by maximum likelihood analysis

			M0 *vs.* M3		M7 *vs.* M8	
Gene	Species	d_N/d_S	p_s	ω	p_s	ω
CG10200	mel, sim, yak, ana, pseudo, moj	0.28	0.36***	1.08	0.32***	1.2
CG5106	mel, sim, yak, ana, vir, moj	0.03	0.03***	0.43	0.005	1.0
CG5273	mel, sim, yak, ana, vir, moj	0.11	0.22***	0.5	0.000	18.4
CG9897	mel, sim, yak, ana, pseudo, moj	0.16	0.15***	2.03	0.19***	1.6
CG13004	mel, sim, yak, ana, pseudo, moj	0.10	0.16***	0.67	0.13***	1.0
CG17012	mel, sim, yak, ana, pseudo, vir, moj	0.22	0.11***	4.5	0.09***	5.6
CG8453	mel, sim, yak, ana, pseudo, vir, moj	0.05	0.16***	0.25	0.010	1.0
CG5976	mel, sim, yak, ana, pseudo, vir, moj	0.05	0.02***	0.51	0.000	1.0

Species are from *D. melanogaster* (mel), *D. simulans* (sim), *D. yakuba* (yak), *D. anannasse* (ana), *D. pseudoobscura* (pseudo), *D. virilis* (vir), *D. mojavensis* (moj); M0 *vs.* M3, M3 parameter estimates of d_N/d_S in the highest class (ω) and proportion of sites (p_s) estimated to belong to that class; M7 *vs.* M8, M8 free parameter estimate of d_N/d_S (ω) and proportion of sites (p_s) estimated to belong to that class. ***$P < 0.001$.

have been proposed (Parker 1970; Eberhard 1996; Rice 1996; Gavrilets 2000; Swanson and Vacquier 2002; Swanson *et al.* 2004). To fully evaluate a process that facilitates coevolution between interacting proteins we need information on both male and female protein evolution (Chapman *et al.* 2003; Swanson *et al.* 2004; Panhuis *et al.* 2006). This study has identified six (including CG17018 previously identified by Swanson *et al.* 2004) female reproductive genes that have experienced adaptive evolution. Furthermore, these six genes make excellent candidates for further functional studies on their role in male–female postmating interactions.

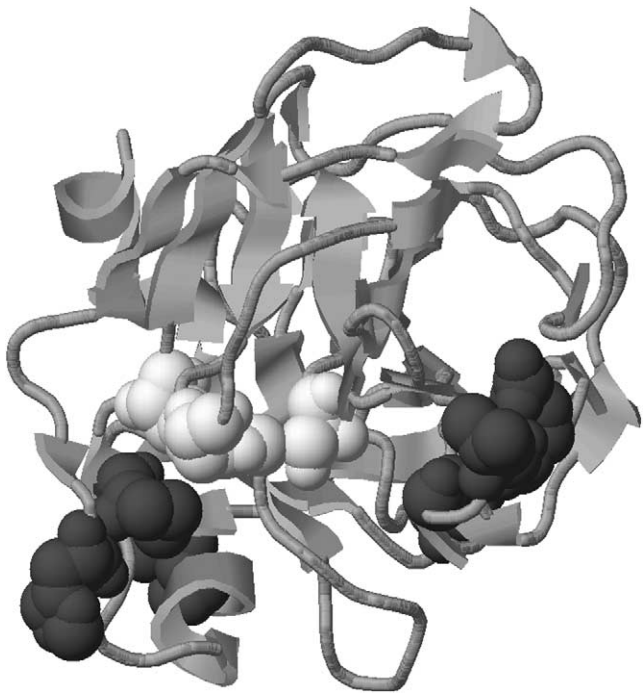

Figure 1.—Three-dimensional model of the protease CG17012. The active site catalytic triad is shown in white and the sites predicted to be subjected to positive selection are shown in dark gray. Note that the sites under selection fall around and within the active site of the protease, suggesting potential functional differences between species.

The two approaches we used to detect positive selection in these candidate female genes, polymorphism surveys, and divergence analyses, detect different kinds of adaptive evolution. The strength of the polymorphism survey lies in its ability to detect signs of recent selection acting on a locus. The divergence analysis, however, detects recurrent selection that may act upon codons in most lineages studied (Anisimova *et al.* 2001; Swanson *et al.* 2004). A significant result from either method is good evidence of adaptive evolution (Swanson *et al.* 2004; Nielsen 2005). Importantly, by using two methods (McDonald–Kreitman test and d_N/d_S analyses) that are robust to demographic effects (Nielsen 2005) we have increased confidence that the signatures we have documented are not the result of demography.

Our polymorphism survey used tests of neutrality to look for signs of recent selective events in each locus. These tests revealed that five of the nine genes surveyed exhibit a recent selective event. Significant negative values for four of the five genes (CG5106, CG17108, CG8453, and CG5273) suggest a role of directional selection, while CG9897 may be influenced by balancing selection, as indicated by a positive Fu and Li's *D** (Table 2). Consistent with balancing selection is the high level of replacement polymorphism seen at this locus in the 36 *D. melanogaster* sequences surveyed (Table 3). A McDonald–Kreitman test on all loci combined revealed a significant departure from homogeneity; even when CG17012 is removed from this analysis (due to its significant departure from homogeneity alone) there is a significant departure from neutrality (Table 3). Our divergence analyses indicated positive selection, a d_N/d_S ratio >1, for three of the nine loci (Table 4, M7 *vs.* M8): CG9897, CG10200, and CG17012. Several of the genes surveyed that show signs of positive selection (CG5106,

CG5273, CG10200, and CG17108) (Table 1) have no documented function. Three of the genes surveyed, which also show signs of positive selection, have a putative ontology function (Table 1) (Ashburner *et al.* 2000). We discuss these positively selected genes in light of their potential function and interaction with male reproductive proteins. We stress, however, that experiments to demonstrate functionality are needed before we can conclude their role in male–female interactions.

Positive selection detected for two trypsin-like serine proteases—CG17012 and CG9897: Two loci with signs of positive selection, CG9897 and CG17012, contain a conserved trypsin-like serine protease domain (Blastp, 100% aligned, expect 7e-33 and 3e-46, respectively; 232 and 231 residues, respectively) (Arbeitman *et al.* 2004). Proteases are enzymes that cleave other proteins and may be important in interactions with male reproductive proteins (Swanson *et al.* 2004).

CG17012 is an interesting candidate locus for its potential interaction with male proteins. Our study has revealed that this gene shows signs of rapid adaptive evolution (Table 2–4) using both polymorphism and divergence data. Sites predicted under selection fall right around the active site of the protease, suggesting functional differentiation between species (Figure 1). Furthermore, this locus shows increased expression in the female sperm storage organ (spermatheca) and the lumen of the parovaria (Arbeitman *et al.* 2004). CG17012's high expression level in the spermatheca suggests that it may be involved in sperm storage or sperm motility, as tyrpsins are in Lepidoptera (Friedlander *et al.* 2001; Arbeitman *et al.* 2004). As a putative protease this protein may play a role in the proteolytic cleavage of a male prohormone inside the mated female (Park and Wolfner 1995; Arbeitman *et al.* 2004). Additionally, it may offset some of the harmful effects of male reproductive proteins after mating (reviewed in Wolfner 2002; Arbeitman *et al.* 2004). Evolutionary models of genes involved in male–female coevolution, such as sexual conflict and sexual selection, predict these genes to be rapidly evolving by adaptive evolution (Rice 1996; Gavrilets 2000). A role in any of these above scenarios would be consistent with coevolution between the sexes; our result of rapid adaptive evolution supports this prediction. Further functional studies on this gene are needed to confirm a role for this protein in male–female postmating interactions.

Like CG17012, locus CG9897 also has a conserved protease domain and is rapidly evolving by positive selection. This locus has a significant, positive Fu and Li's D^* (Table 3) indicating balancing selection may contribute to the evolution of this gene. Balancing selection tends to maintain mutations at intermediate frequencies, and may evolve in response to male–female interactions if a rare variant is favored, as seen in self *vs.* nonself recognition systems of several plant species (Charlesworth 2002). Sperm competition and single mating studies in Drosophila show that male fertilization success is partly determined by the female genotype (Clark and Begun 1998; Clark *et al.* 1999). This male × female interaction has been suggested to contribute to the maintenance of allelic variation and balancing selection in several male reproductive proteins and may also explain the degree of polymorphism and potential balancing selection seen for CG9897 (Clark *et al.* 1999).

Positive selection in a cytochrome P450 gene—CG8453: CG8453 is a cytochrome P450 gene, *Cyp6g1*. Cytochrome P450 genes are involved in oxidative degradation of various compounds, such as endogenous and exogenous toxins. This locus has been studied for its potential role in response to insecticide resistance, such as DDT and other harmful molecules (Daborn *et al.* 2002). Lines in *D. melanogaster* with high levels of *Cyp6g1* transcript appear to be resistant to DDT compared to susceptible lines, which do not show high expression level (Daborn *et al.* 2001, 2002; Schlenke and Begun 2004). A transposable element, *Accord*, inserted several hundred base pairs upstream of the transcription start site appears to be present only in DDT-resistant strains suggesting a role for this element in the upregulation of *Cyp6g1* and insecticide resistance (Daborn *et al.* 2002). A significant Fay and Wu's H seen at this locus (Table 2) is consistent with a recent hitchhiking event due to selection at or near this locus. If the *Accord* transposable element is indeed favored by selection due to its putative role in the upregulation of *Cyp6g1* we might expect a reduction in heterozygosity at this locus (Schlenke and Begun 2004). This hitchhiking event has been found at this locus in *D. simulans* (Schlenke and Begun 2004). Schlenke and Begun (2004) also found a low level of variation in a region near *Cyp6g1* for a small sample of *D. melanogaster*. This level of variation is similar to what we observe at this locus in our larger sample of *D. melanogaster* ($\theta = 0.00169$; Table 2). We do not have sequence data for the *Accord* region and cannot comment on whether or not our *D. melanogaster* sample from Riverside has this transposable element.

The potential detoxification effects of this protein may be important in male–female postmating interactions by reducing harmful molecules introduced to the female upon mating. Interestingly, toxic effects of male reproductive proteins on females have been shown (Lung *et al.* 2002) and a greater expression level of this gene in adult female tissue compared to male tissue (Arbeitman *et al.* 2004) may aid in the detoxification of male introduced toxins. A specific role of CG8453 in the detoxification of toxic male proteins awaits further study.

Conclusion: We have shown several genes expressed in the female reproductive tract that have been targets of positive selection. This includes a broad class of genes with a variety of functions. While identification of

positive selection in genes encoding reproductive proteins is a recurrent observation (Swanson and Vacquier 2002), there are still very few female genes for which adaptive evolution has been documented. Since reproduction, including sperm competition, is a dynamic process involving both male and female components we believe it is necessary to study the dynamics of reproductive genes from both sexes.

We thank David Rand, Jennifer Calkins, and two anonymous reviewers for comments on the manuscript. T.M.P was supported on National Institutes of Health Genome Training Grant HG-000035 and W.J.S. was supported by National Institutes of Health grant HD-42563 and National Science Foundation grant DEB-0410112.

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gogayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

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

Aquadro, C. F., 1997 Insights into the evolutionary process from patterns of DNA sequence variability. Curr. Opin. Genet. Dev. **7:** 835–840.

Arbeitman, M. N., A. A. Fleming, M. L. Siegal, B. H. Null and B. S. Baker, 2004 A genomic analysis of *Drosophila* somatic sexual differentiation and its regulation. Development **131:** 2007–2021.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botsein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25:** 25–29.

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

Chapman, T., G. Arnqvist, J. Bangham and L. Rowe, 2002 Sexual conflict. TREE **18:** 41–46.

Chapman, T, J. Bangham, G. Vinti, B. Seifried, O. Lung *et al.*, 2003 The sex peptide of *Drosophila melanogaster*: female postmating responses analyzed by using RNA interference. Proc. Natl. Acad. Sci. USA **100:** 9923–9928.

Charlesworth, D., 2002 Self-incompatibility: how to stay incompatible. Curr. Biol. **12:** R424–R426.

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

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

Clark, N. L., J. E. Aagaard and W. J. Swanson, 2006 Evolution of reproductive proteins from animals and plants. Reproduction **131:** 11–22.

Daborn, P., S. Boundy, J. Yen, B. Pittendrigh, and R. ffrench-Constant, 2001 DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. Mol. Genet. Genomics **266:** 556–563.

Daborn, P., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single P450 allele associated with insecticide resistance in *Drosophila*. Science **297:** 2253–2256.

Eberhard, W. G., 1996 *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton University Press, Princeton, NJ.

Fay, J. C., and C.-I Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Friedlander, M., A. Jeshtadi and S. Reynolds, 2001 The structural mechanism of trypsin-induced intrinsic motility in Manduca sexta spermatozoa in vitro. J. Insect Physiol. **47:** 245–255.

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

Gavrilets, S., 2000 Rapid evolution of reproductive barriers driven by sexual conflict. Nature **403:** 886–889.

Haddrill, P. R., K. R. Thornton, B. Charlesworth and P. Andolfatto, 2005 Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. Genome Res. **15:** 790–799.

Hammer, M. F., D. Garrigan, E. Wood, J. A. Wilder, Z. Mobasher *et al.*, 2004 Heterogeneous patterns of variation among multiple human x linked loci: the possible role of diversity-reducing selection in non-Africans. Genetics **167:** 1841–1853.

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

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

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

Jansa, S. A., B. L. Lundrigan and P. K. Tucker, 2003 Tests for positive selection on immune and reproductive genes in closely related species of the murine genus mus. J. Mol. Evol. **56:** 294–307.

Kumar, S., K. Tamura and M. Nei, 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinformatics **5:** 150–163.

Lung, O., U. Tram, C. M. Finnerty, M. A. Eipper-Mains, J. M. Kalb *et al.*, 2002 The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. Genetics **160:** 211–224.

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

Metz, E. C., R. Robles-Sikisaka and V. D. Vacquier, 1998 Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc. Natl. Acad. Sci. USA **95:** 10676–10681.

Nielsen, H., J. Engelbrecht, S. Brunak and G. Von Heigne, 1997 A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int. J. Neural Syst. **8:** 581–599.

Nielsen, R., 2005 Molecular signatures of natural selection. Annu. Rev. Genet. **39:** 197–218.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.

Otto, S. P., 2000 Detecting the form of selection from DNA sequence data. Trends Genet. **16:** 526–529.

Panhuis, T. M., N. L. Clark and W. J. Swanson, 2006 Rapid evolution of reproductive proteins in Abalone and Drosophila. Phil. Trans. R. Soc. **361:** 261–268.

Park, M., and W. F. Wolfner, 1995 Male and female cooperate in the prohormone-like processing of a Drosophila melanogaster seminal fluid protein. Dev. Biol. **171:** 694–702.

Parker, G. A., 1970 Sperm competition and its evolutionary consequences in the insects. Biol. Rev. **45:** 525–567.

Price, C. S., 1997 Conspecific sperm precedence in Drosophila. Nature **388:** 663–666.

Przeworski, M., 2002 The signature of positive selection at randomly chosen loci. Genetics **160:** 1179–1189.

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

Rambaut, A., 1996 *Se-Al: Sequence Alignment Editor* (http://evolve.zoo.ox.ac.uk/).

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

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

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

Schlenke, T. A., and D. J. Begun, 2004 Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **101:** 1626–1631.

Schwede, T., J. Kopp, N. Guex and M. C. Peitsch, 2003 SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. **31:** 3381–3385.

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

Swanson, W. J., Z. Yang, M. F. Wolfner and C. F. Aquadro, 2001a Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. USA **98:** 2509–2514.

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

Swanson, W. J., A. Wong, M. F. Wolfner and C. F. Aquadro, 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 methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Thornton, K., and P. Andolfatto, 2006 Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. Genetics **172:** 1607–1619.

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

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

Yang, Z., 2000 *Phylogenetic Analysis by Maximum Likelihood (PAML).* University College London, London.

Yang, Z., 2005 The power of phylogenetic comparison in revealing protein function. Proc. Natl. Acad. Sci. USA. **102:** 3179–3180.

Yang, Z., and J. P. Bielawski, 2000 Statistical methods for detecting molecular adaptation. Trends. Ecol. Evol. **15:** 496–503.

Yang, Z., and W. J. Swanson, 2002 Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. **19:** 49–57.

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

Yang, Z., W. S. Wong and R. Nielsen, 2005 Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22:** 1107–1118.

Communicating editor: D. M. Rand