Evidence of Adaptive Evolution of Accessory Gland Proteins in Closely Related Species of the *Drosophila repleta* Group

Francisca C. Almeida† and Rob DeSalle*†*

*Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY; and †Department of Biology, New York University

Accessory gland proteins (Acps) are part of the seminal fluid of *Drosophila* species. These proteins have important reproductive functions, being responsible for the proper functioning of several steps of the fertilization process. Acps also contribute indirectly for the reproductive success of males by modulating female behavior. Evidence that Acps participate in sperm competition and sexual conflict includes findings that, on average, Acps have fast evolutionary rates, suggestive of adaptive evolution. This is especially true in species of the *Drosophila repleta* group. Nevertheless, only in a few occasions have robust statistical tests been used to determine whether observed evolutionary rates are in fact due to positive selection on amino acid substitutions between related species. Here we apply maximum likelihood tests for positive selection on 14 Acps of the *D. repleta* group. To increase statistical robustness, we use at least 8 sequences, all belonging to species of the *Drosophila mulleri* complex, for each gene analyzed. We found significant evidence of adaptive evolution for 10 of the tested genes. Among these, the ones with a conserved protein domain had positively selected sites within the functional region of the sequence. We also detected one instance of lineage-specific adaptive evolution in a clade formed by 2 sister species.

Introduction

Accessory gland proteins (Acps) form part of the *Drosophila* and other dipetrans' seminal fluid. During insemination, Acps are transferred into the female reproductive tract where they perform several functions during the fertilization process, being essential for its achievement (for a review of Acp function, see Wolfner 2002; Chapman and Davies 2004). Acps are not only important for proper sperm storage and utilization but also modulate female physiology and behavior in order to increase fertilization success. The biochemical functions of certain Acps suggest that these proteins are likely subjects of postmating sexual selection either by sperm competition or sexual conflict (Chen et al. 1988; Harshman and Prout 1994; Chapman et al. 1995; Clark et al. 1995; Civetta and Clark 2000; Lung et al. 2002; Wigby and Chapman 2005; Mueller et al. 2007).

Interestingly, the findings that suggest that Acps are under sexual selection were met by evidence that many Acp genes have patterns of nucleotide variation compatible with evolution by positive selection. In fact, the phenomenon of rapid, adaptive evolution in reproductive molecules has been observed in a wide variety of animal taxa (Swanson and Vacquier 2002; Panhuis et al. 2006). In *Drosophila*, some of the Acps showing signs of adaptive evolution are the same that had been linked to functions involved in sperm competition and sexual conflict (Aguadé et al. 1992; Aguadé 1999; Begun et al. 2000; Swanson, Clark et al. 2001; Mueller et al. 2005; Wagstaff and Begun 2005; Schully and Hellberg 2006). The implication that coevolution by sexual conflict drives the adaptive evolution of Acps has been reinforced by evidence that some female reproductive tract proteins are under positive selection (Swanson et al. 2004; Kelleher et al. 2007). Similar evidence was observed in different *Drosophila* species groups, such as *melanogaster*, *pseudoobscura*, and *repleta*.

Key words: Acp, *Drosophila repleta* group, adaptive evolution.

E-mail: falmeida@amnh.org.

Mol. Biol. Evol. 25(9):2043–2053. 2008
doi:10.1093/molbev/msn155
Advance Access publication July 17, 2008

The main evidence used to suggest adaptive evolution in Acps is a high nonsynonymous substitution (d_N) rate as compared with synonymous substitution rates (d_S). High d_N/d_S ratios (>0.5) were observed in most Acps described so far (Swanson, Clark et al. 2001; Wagstaff and Begun 2005; Haerty et al. 2007). Nevertheless, although a high d_N/d_S suggests that the gene in question may be under positive selection, it is simply an estimate not a statistical test. In addition, relatively high rates of nonsynonymous substitutions ($d_N/d_S \sim 1$) can also be due to a neutral selection regime. To have a better assessment of evolutionary patterns of Acps, different statistical tests have been employed to determine whether Acp evolutionary rates depart from neutral expectations. Most Acps with high d_N/d_S tested so far, however, fail to show sequence variation patterns consistent with positive selection in neutrality tests based on population polymorphism data (Kern et al. 2004). Another way of testing for positive selection is to compare the relative number of nonsynonymous substitutions in fixed and polymorphic substitutions when populational data are available for 2 closely related species, using the McDonald–Kreitman test. This test, with a few exceptions, also failed to provide support for positive selection for a number of Acps with high d_N/d_S estimates (Begun et al. 2000; Kern et al. 2004; Wagstaff and Begun 2005).

Here we apply a likelihood approach to test for positive selection in several Acp loci identified in an accessory gland cDNA library of *Drosophila mayaguana* (FC Almeida and R DeSalle, in review). The likelihood tests for positive selection rely on site-by-site estimates of d_N/d_S (Yang 1998). Because selection pressure can be heterogeneous across codons of a gene, likelihood tests have increased power to detect adaptive evolution. This is specially true when most codons of a gene are under purifying selection, and only a few sites are under positive selection. However, the accuracy of these tests depends on the assembly of a large enough number of sequences for comparison. The power to detect sites under positive selection is decreased considerably if the data set consists of 6 or fewer sequences (Anisimova et al. 2001).

Drosophila mayaguana is part of the *Drosophila repleta* group, whose Acps show extraordinarily high d_N/d_S

ratios (Wagstaff and Begun 2005, 2007; Almeida and DeSalle, in review). In this group, McDonald-Kretiman (MK) tests showed significant results suggestive of positive selection for 3 Acp loci (Wagstaff and Begun 2005). We have taken advantage of the intermediate phylogenetic relationship between *D. mayaguana* and *Drosophila mojavensis* and the availability of a genome sequence for the latter species to design primers that are likely conserved across most species of the *Drosophila mulleri* complex, to which both species belong. This advance made possible the first tests for positive selection on Acps using a considerable number of species. We were able to include samples for all the known species of the *mulleri* and *mojavensis* clusters. These sister clusters include species that diverged possibly less than 10 MYA (Russo et al. 1995). Such phylogenetic sampling allows for a comprehensive understanding of the evolution of Acps in closely related species and provides clues of how selection may have influenced species establishment and divergence.

Our approach allowed us to use robust tests for positive selection in 14 Acps of the *D. repleta* group. We found a significant evidence of adaptive evolution in 10 of them. Among these, the 3 Acps with a conserved protein domain had positively selected amino acids sites identified within the functional domains. Three additional genes, for which tests were not applied due to a lack of adequate number of sequences, showed high d_N/d_S ratios suggestive of positive selection. We also identified one case of lineage-specific accelerated evolution in a protease gene.

Materials and Methods

Flies Samples

Ten species of the *D. repleta* group, all belonging to the *D. mulleri* complex (Durando et al. 2000), were used in this study: *D. mojavensis*, *Drosophila arizonae*, *Drosophila navojoa* (*D. mojavensis* cluster), *D. mulleri*, *Drosophila aldrichi*, *Drosophila huaylasi*, *Drosophila wheeleri*, *Drosophila nigrodumosa*, *D. mayaguana*, and *Drosophila parisiena* (*D. mulleri* cluster). We also tried to amplify Acps in 2 other *repleta* group species, *Drosophila propachuca* and *Drosophila spenceri*, which belong to the *Drosophila longicornis* cluster, sister to the *mulleri* complex (Durando et al. 2000). Except for specimens of *D. mayaguana* (15081-1397.03) and *D. parisiena* (15081-1392.01) that were obtained from the Tucson stock center, all others came from the collections of W. B. Heed from University of Arizonae, donated to the Ambrose Monnel Cryo Collection of the American Museum of Natural History (supplementary table 1, Supplementary Material online). DNA was extracted from 1 to 3 flies using the DNeasy Extraction Kit (QIAGEN, Valencia, CA). *Drosophila straubae*, another member of the *mulleri* complex, was not included in the analyses because genetic differentiation between this species and the sister taxon *D. parisiena* has yet to be proven (O'Grady et al. 2002).

Loci and DNA Sequencing

Forty-three Acp candidates were selected from an accessory gland cDNA library of *D. mayaguana*. The selection was based on the existence of good quality sequences longer than 300 bp and a negative dot blot result when probed with female cDNA (Almeida and DeSalle, in review). Six of these loci, however, were not classified as Acp based on the commonly used Acp criteria (Swanson, Clark et al. 2001; Mueller et al. 2004). *Drosophila mayaguana* cDNA sequences were aligned with their best Blast hit in the *D. mojavensis* genome, and potentially conserved primers sites were identified. Conserved primers that allow for the amplification of fragments longer than 250 bp were designed from the alignments of 26 loci (supplementary table 2, Supplementary Material online). Polymerase chain reaction (PCR) amplification protocols were optimized for each locus (conditions available upon request). Sequences were obtained using a 3730XL automated sequencer and edited with Sequencher 4.5 (Genecodes, Ann Arbor, MI). For genes classified as Acp in both *D. mayaguana* (Almeida and DeSalle, in review) and *D. mojavensis* (Wagstaff and Begun 2005), the species prefix was omitted from the gene name.

Sequence Analyses

Alignments were performed with MAFFT 6.236b (Katoh et al. 2005) and trimming of noncoding regions (introns and 3′ untranslated region [UTR]) was done in MacClade 4.08 (Maddison D and Maddison W 2000), using *D. mayaguana* cDNA sequence for determining introns. Manual codon alignment (gap placement) was performed in MacClade 4.08 on alignments obtained with MAFFT. Presence of a signal peptide in the amplified sequence was checked using the program SignalP 3.0 (Bendtsen et al. 2004). Codon bias was estimated by the effective number of codons (ENC) and proportion of G and C in the third codon position (G/C 3rd), both obtained with DNAsp (Rozas et al. 2003). Conserved domain alignments were obtained using the CD-search online program and CDD database (Marchler-Bauer and Bryant 2004; Marchler-Bauer et al. 2005). Relative rate tests were done using HyPhy (Pond et al. 2005) using the general reversible model. In these tests, only the species of the *mulleri* subcluster were included and *D. mayaguana* was used as outgroup.

Tests for Positive Selection

The role of positive selection in the evolution of the coding region of Acp genes was assessed in cross-species comparisons and statistical tests were applied when sequences of a minimum of 8 species were available. Tests for positive selection were carried out using the program codeml of the PAML 3.15 package (Yang 1997; Yang and Bielawski 2000; Yang and Nielsen 2002). The program codeml provides a number of nucleotide substitution models and the likelihood of these models given the data can be compared using the likelihood ratio test (LRT) with a chi-square distribution (Yang et al. 1998, 2005; Wong et al. 2004). Average d_N/d_S across codons was obtained for each gene using the maximum likelihood estimates (Yang et al. 1998), assuming homogeneous replacement rates across

sites (M0). For assessing whether some sites are under positive selection, we used 3 model comparisons: M1a × M2a, M7 × M8, and M8a × M8 (Swanson et al. 2003; Wong et al. 2004; Yang et al. 2005). The first and the second models compare at one side models that assume that site d_N/d_S ratios are distributed from 0 to 1 (M1a and M7) with their alternative hypothesis which assume that a few sites are outside of this distribution and have $d_N/d_S > 1$ (M2a and M8, respectively). M8a is another null hypothesis for model M8, in which d_N/d_S is fixed at 1 for the class of sites with $d_N/d_S > 1$, being a robust test for positive selection (Swanson et al. 2003). Whereas M1a and M2a assume a discrete distribution of d_N/d_S classes, M7 and M8 assume a beta distribution. Branch models, which allow for different d_N/d_S in different branches of the tree, were run using codeml (model = 2, *N*sites = 0), with the null hypothesis being that all branches have the same d_N/d_S (model = 0, *N*sites = 0).

The phylogenetic relationships among species are very important in the likelihood estimation of these models. Because some genes, especially Acps, may have multiple copies in a genome (paralogs), the species phylogeny may not represent the relationship of the sequences analyzed here. For this reason, we obtained gene trees for each one of the genes analyzed. With minor exceptions, the species relationships recovered were largely congruent among Acps and very similar to what has been obtained using other genes, such as *hunchback* and 16S (Durando et al. 2000). For the genes with tree topologies discordant from the most supported one (fig. 1), positive selection tests were rerun using the "species tree" (i.e., the most supported topology). All phylogenetic analyses were done using the maximum likelihood algorithm and the general time reversible + I + Γ model as implemented in PAUP* (Swofford 2003), with parameters estimated from the data based on an initial tree obtained by maximum parsimony. Detailed results of the phylogenetic analyses will be described elsewhere (FC Almeida, S-O Kolokotronis and R DeSalle in preparation).

As a comparison, tests for positive selection were also conducted on the gene *hunchback*, the only nuclear, protein-coding gene for which sequences are available for all the species included here. For the genes with evidence of positive selection (alternative hypothesis with significantly higher likelihood than the null hypothesis), the probability of each site being subject to positive selection was estimated using a Bayes Empirical Bayes (BEB) approach also using codeml (Yang et al. 2005). This approach estimates the probability of each site belonging to 3 main categories: class 0 with $d_N/d_S < 1$, class 1 with $d_N/d_S = 1$, and class 2 with $d_N/d_S > 1$.

Results and Discussion

PCR Results and Sequence Characterization

PCR Amplifications

Amplification was successful in at least 8 species for 16 genes (table 1). Two of these genes did not meet the Acp criteria, although they are expressed in the accessory glands of *D. mayaguana*. Among the remaining genes for which conserved primers were designed (9 loci), some could be amplified in different numbers of species (6 loci, table 1) and 3 of them did not work well even for *D. mayaguana* and/or *D. mojavensis*. Almost all genes that could be amplified in *D. mayaguana* and *D. mojavensis* were also amplified in their sister species, *D. parisiena* and *D. arizonae*, respectively (table 1). Considerably fewer genes could be amplified in *D. propachuca* (3) and *D. spenceri* (4), as expected due to their more distant relationships with the other species used in this study (Durando et al. 2000). Seven of the genes amplified for 8 or more species were also found to be Acps in *D. mojavensis* (Wagstaff and Begun 2005). This expression pattern conservation between *D. mayaguana* and *D. mojavensis* suggests that these genes are probably also expressed in the accessory glands of the other species of the *mulleri* cluster.

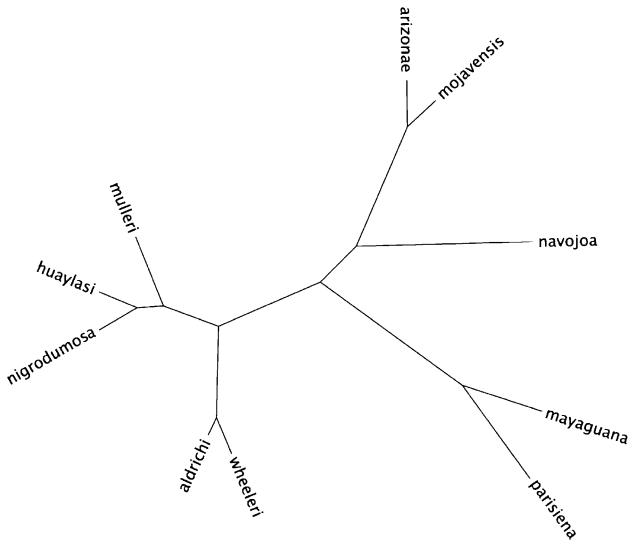

FIG. 1.—Unrooted gene tree based on *Acp7* sequences. This tree illustrates the relationships among species of the *mulleri* complex most frequently recovered with the Acp genes analyzed here.

Drosophila navojoa, among the species analyzed here, had a particularly high number of failures to amplify (7/16), as compared with other species similarly related to *D. mayaguana* and *D. mojavensis* (species used for primer design): *D. wheeleri* (2/16), *D. aldrichi* (1/16), and *D. mulleri* (1/16). This result could be related to particularly high evolutionary rates in *D. navojoa*, resulting in nonconserved primer sites or gene loss. We tested this hypothesis by comparing rates between species with the relative rate test in 9 loci for which *D. navojoa* sequences were available. The results did not support a general higher evolutionary rate in the lineage of *D. navojoa*. This species showed a significantly higher substitution rate in only 1 gene, *may97*, in 4 (out of 7) pairwise comparisons with other species. On the other hand, in *Acp25*, *D. navojoa* showed significantly lower rates in all possible comparisons.

Sequences

High-quality sequences were obtained for most PCR products, but some PCR products resulted in double sequences that suggest lineage-specific gene duplications.

Table 1
Gene Information and PCR Amplification Results

Locus	Function	may	par	moj	ari	nav	mul	ald	nig	whe	hua	spe	pro	# sp
Acp1	*Acp53C*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			10
Acp2a	Acp53C	✓	✓	✓	✓									4
Acp7	Unknown	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			11
Acp11	Unknown	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	11
Acp19	CRISP	✓	✓	✓	✓		✓	✓	✓	✓	✓			9
Acp25	*Acp53C*	✓	✓	✓	✓	✓	✓	✓	✓	✓				9
Acp42	Unknown	✓	✓	✓	✓		✓	✓	✓	✓	✓			9
Acp44	Lectin	✓		d	✓	✓	✓		✓		✓			7
Acp45	Unknown	✓	✓	✓	✓	✓	✓	✓	✓	✓				9
mayAcp56	CRISP	✓	✓		✓									3
mayAcp57	CRISP	✓	✓	✓	✓		✓	✓	✓	✓	✓			9
mayAcp58	Thiol reductase	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			10
mayAcp62	Lipase	✓	✓	✓	✓		✓							5
mayAcp63	Unknown	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			10
mayAcp64	Fibrinogen	✓	✓	d	✓							✓	✓	4
mayAcp65	Unknown	✓	✓	✓	✓	✓		✓	✓	✓	✓			9
mayAcp68b	Unknown	✓	d	✓				✓						4
mayAcp69a	Unknown	✓	✓	✓	✓		✓	✓	✓		✓			8
mayAcp73	Protease	✓	✓	d	✓		✓	✓						6
mayAcp74	Protease	✓	✓	✓	✓		✓		✓		✓	✓		8
mayAcp75	Fibrinogen	✓	✓	✓	✓		✓	✓	✓	✓	✓			9
mayAcp77	CRISP	✓	✓	✓										3
may82	Unknown	✓	✓	✓	✓									4
may83	Unknown	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
may97	ER protein	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			10

NOTE.—may, *Drosophila mayaguana*; par, *Drosophila parisiena*; moj, *Drosophila mojavensis*; ari, *Drosophila arizonae*; mul, *Drosophila mulleri*; ald, *Drosophila aldrichi*; whe, *Drosophila wheeleri*; hua, *Drosophila huaylasi*; spe, *Drosophila spenceri*; pro, *Drosophila propachuca*; and # sp, total number of species with successful amplification results. √, successful amplification; d, double sequence.

We cloned the PCR products of *D. mulleri* for the gene *mayAcp74* and sequenced 20 colonies. The sequences revealed that, in fact, 2 sequences, with 82.5% similarity, were being amplified in that species. A phylogenetic analysis of all the sequences obtained for this gene did not support a species-specific duplication in *D. mulleri* (fig. 2). Instead, it suggests that the duplication occurred before the split between *D. mulleri* and *D. nigrodumosa* + *D. huaylasi*. If in fact the latter 2 species have only one copy of the gene, the results of the phylogenetic analysis imply gene loss in those species.

For one of the non-Acp genes, *may83*, even though amplification was successful for 12 species (table 1), it was not possible to find a common open reading frame (ORF) for all the species in the alignment. No ORF was found in *D. propachuca*, and the ORF found in *D. mayaguana* and *D. parisiena* was in reverse orientation and not overlapping with the one found in the remaining species. Nevertheless, the alignment of the ORF found the largest number of species (9) revealed that *D. arizonae* and *D. nigrodumosa* had frameshift mutations a few bases downstream to the start codon, which did not allow for an accurate alignment of codons. The fragments sequenced probably represent pseudogenes or unstranslated mRNA. This gene was not included in further analyses.

Most sequences had a high probability of carrying a signal peptide ($P > 0.9$), indicating that the coding regions analyzed were complete or almost complete at their 5′ end (table 2). Most of the sequences were also complete in their 3′ end of the coding regions as inferred by the presence of a stop codon. The presence of a signal peptide can also be interpreted as evidence that these genes are Acps in the other species besides *D. mayaguana*, although this could only be confirmed by mRNA analyses.

The average codon bias for Acp genes was low, with ENC = 54.4 (ENC varies from 20 to 61, where 61 is no bias) and C/G 3rd = 0.48 (table 2). The codon bias, as measured by both ENC and C/G 3rd, of the only 2 non-Acp genes analyzed here, *may97* and *hunchback*, was higher

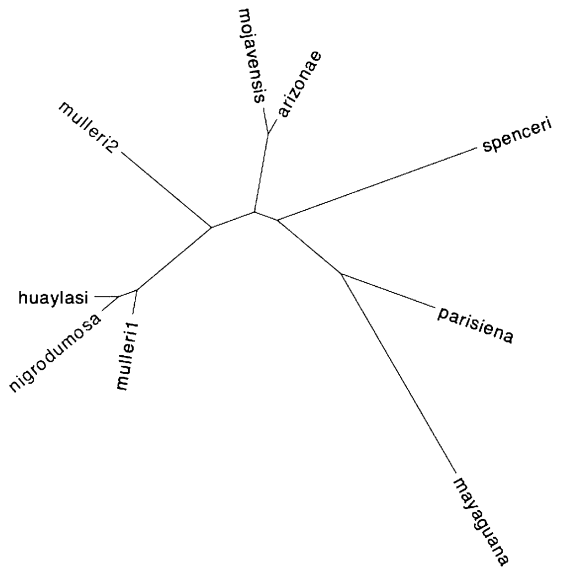

FIG. 2.—Unrooted gene tree of *mayAcp74* orthologs showing that a duplication event occurred likely before the split of *Drosophila mulleri* and the clade formed by *Drosophila nigrodumosa* + *Drosophila huaylasi*.

Table 2
Characterization of Genes with 8 or More Sequences in the *Drosophila repleta* Group

Locus	*N* Seqs	bp	Signal Pep	ENC	G + C 3rd	ω	d_N	d_S
Acp1	10	318	Y	58.04	0.491	0.686	0.528	0.770
Acp7	10	609	Y	57.94	0.407	1.275	0.699	0.549
Acp11	11	252	Y	58.37	0.593	0.848	0.916	1.080
Acp19	9	690	Y	59.24	0.402	0.896	0.334	0.373
Acp25	9	363	Y	55.80	0.475	0.640	0.473	0.739
Acp42	8	609	Y	54.01	0.358	1.144	0.763	0.667
Acp45	9	465	Y	51.69	0.359	0.800	0.615	0.769
mayAcp57	8	479	Y	45.65	0.626	0.179	0.079	0.440
mayAcp58	10	479	N	56.91	0.512	0.635	0.597	0.940
mayAcp63	10	360	N	52.34	0.573	0.163	0.106	0.648
mayAcp65	8	255	Y	56.49	0.460	0.786	0.482	0.613
mayAcp69a	8	282	Y	46.17	0.628	0.585	0.654	1.118
mayAcp74	9	792	N	56.89	0.339	0.770	0.866	1.125
mayAcp75	9	576	N	51.47	0.438	0.463	0.213	0.461
may97	10	705	N	40.12	0.741	0.111	0.071	0.639
hunchback	10	501	N	45.269	0.687	0.236	0.134	0.569

NOTE.—*N* seqs, number of sequences analyzed; bp, length of the alignment in base pairs; signal pep, presence of signal peptide in the sequences; ENC, effective number of codons; G + C 3rd, proportion of C and G in the third codon position; y, yes; and n, no.

than the most biased Acp. Little is known about codon bias in the species of the *repleta* group. The only other gene for which there are data on codon bias for several species of the *repleta* group, including 4 of the species studied here, is the *xanthine dehydrogenase* locus (Begun and Whitley 2002). In that study, codon bias was assessed for several species of 4 *Drosophila* groups, among which the *repleta* group had the highest levels of codon bias, especially the species of the *mulleri* cluster. Among the genes analyzed here, there was significant heterogeneity in ENC (Kruskal–Wallis chi-squared = 97.40, $P = 4.044 \times 10^{-14}$). On the other hand, codon bias was relatively homogeneous across the species of the *mulleri* complex (*D. spenceri* and *D. propachuca* were excluded because very few sequences were available) as measured by ENC (Kruskal–Wallis chi-squared = 4.04, $P = 0.91$).

Extensive Positive Selection in Acps across Species Boundaries in the *repleta* Group

Substitution Rates

Synonymous (d_S) and nonsynonymous (d_N) substitution rates and the d_N/d_S ratios for genes with at least 8 sequences are shown in table 2. Overall, across sites, only *Acp7* and *Acp42* had $d_N/d_S > 1$. Another 11 Acps had $d_N/d_S > 0.5$, which is relatively high as compared with the average d_N/d_S of non-Acp genes in *Drosophila* (Mueller et al. 2005; Wagstaff and Begun 2005). High d_N/d_S values observed could not be attributed to extraordinarily low d_S because we found a positive correlation between d_N and d_S (Spearman rank correlation, $\rho = 0.63$, $P = 0.01$). A possible explanation for this correlation is related to codon bias; if genes with high d_N/d_S have low codon bias as previously found (Akashi 1996; Kim 2004), then low pressure for preferred codons could lead to high d_S in these loci. In fact, we found a highly negative correlation between d_N/d_S and codon bias as measured by both ENC and C + G 3rd (Spearman rank correlation, $\rho = 0.67$, $P = 0.006$, and $\rho = -0.75$, $P = 0.001$, respectively).

Even though we chose not to apply tests for positive selection in genes with less than 8 sequences, we calculated the average overall d_N/d_S for 6 of them. Among these, only *mayAcp68b* showed $d_N/d_S > 1$ but 2 (*mayAcp73* and *may82*) other had $d_N/d_S > 0.7$ (supplementary table 2, Supplementary Material online).

Tests for Positive Selection

Table 3 shows the results of the LRTs obtained in comparisons between M7 and M8 and between M8 and M8a conducted to examine positive selection in 17 genes. Results obtained in the comparison between M1a and M2a (data not shown) were in general agreement with the other tests' results. The LRTs comparing M7 and M8 support the presence of sites under positive selection in 10 Acps with $P < 0.01$ ($P < 0.001$ was found in 7 genes). The same conclusion was reached by LRT in comparisons between M8a and M8 (table 3). The results strongly support the notion that positive selection is a cause for the inflated d_N/d_S values, rather than simply relaxed selection. The occurrence of positive selection on amino acid substitutions was significant at the $P < 0.01$ level for all the genes classified as Acps in both *D. mayaguana* and *D. mojavensis*. These genes are those that are most likely Acps in the other species of the *mulleri* cluster. Sites under positive selection were detected in genes with average d_N/d_S as low as 0.635 (*mayAcp58*).

Of the 16 genes analyzed, *mayAcp57*, *mayAcp63*, *mayAcp69a*, *mayAcp75*, *may97*, and *hunchback* were the only genes for which the LRTs did not reject the null hypothesis. *may97* belongs to a family of endoplasmatic reticulum proteins (ERp29) that are highly expressed in secretory cells. Its likely ortholog in *Drosophila melanogaster*, *windbeutel*, is a gene involved in dorsal–ventral patterning and whose sequence is conserved across distantly related *Drosophila* species. A low d_N indicates that strong purifying selection is acting on this gene (table 2). Three of the 4 Acps analyzed here that had no codon under selection according to the likelihood tests, also showed relatively low d_N. *mayAcp57* is part of a gene family that

Table 3
Results of Tests for Positive Selection Based on LRT in M7 × M8 and M8 × M8a Comparisons

Locus	ω	M7 × M8	M8 × M8a	ω2	P2	BEB Sites
Acp1	0.69	20.40***	19.42***	2.41	0.292	9(3)
Acp7	1.28	26.74***	26.71***	3.00	0.372	8(1)
Acp11	0.85	21.14***	18.51***	4.71	0.158	5(3)
Acp19	0.90	16.47***	15.94***	2.92	0.218	7(1)
Acp25	0.64	9.51**	9.38**	4.87	0.062	3(0)
Acp42	1.14	10.60**	10.46**	2.03	0.575	7(0)
Acp45	0.80	28.09***	27.18**	5.67	0.062	4(2)
mayAcp57	0.18	0.66	n.a.	1.38	0.091	0
mayAcp58	0.64	17.93**	16.55**	3.04	0.134	7(2)
mayAcp63	0.16	0.35	n.a.	1.17	0.134	0
mayAcp65	0.79	10.03**	9.77**	3.46	0.135	4(1)
mayAcp69a	0.59	4.74	n.a.	3.86	0.076	1(0)
mayAcp74 (mul1)	0.77	59.42***	51.79***	4.69	0.116	17(5)
mayAcp74 (mul2)	0.70	53.81***	44.75***	3.95	0.120	13(5)
mayAcp75	0.65	5.37	n.a.	6.07	0.020	2(0)
may97	0.46	1.00	n.a.	1.07	0.065	0
hunchback	0.24	5.91	n.a.	2.38	0.089	1

NOTE.—ω2, the average d_N/d_S for the class of sites with $d_N/d_S > 1$ (class 2); p2, frequency of sites in class 2; and sites, number of sites with BEB $P > 0.95$ of being in class 2 (in parenthesis, sites with $P > 0.99$). *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

encodes for proteins with a CRISP domain, frequently found among *Drosophila* Acps (Mueller et al. 2004). Only in *D. mayaguana*, 3 other expressed sequence tags with the same domain were identified, including *Acp19*, also analyzed here. The function of *mayAcp63* is unknown. It has a likely ortholog in *D. melanogaster* (CG13585) that was recovered in a testes cDNA library, suggesting a conserved function in reproduction. *mayAcp75* encodes for a protein containing a fibrinogen-like domain. Although it was identified in a cDNA library of accessory glands (Almeida and DeSalle, in review), its domain has not been found in any gene expressed in the same organ of any other *Drosophila* species. It is possible that its function as an Acp is restricted to *D. mayaguana*.

mayAcp69a, however, did show a high d_N estimate. Nevertheless, d_S estimates were also quite high for this gene, and we believe that this might reflect a paralogy problem. The *D. mayaguana* and *D. parisiena* sequences were very divergent from the rest as they contained many indels. For *D. parisiena*, the *mayAcp69a* sequence was considerably shorter than that of the remaining species analyzed due to a premature stop codon and, for this reason, was not included in the tests for positive selection. Although the gene tree obtained with the sequences of *mayAcp69a* was congruent with the most accepted relationships among species, the node leading to *D. parisiena* and *D. mayaguana* was exceptionally long. One explanation for this result is that we actually amplified a paralog of *mayAcp69a* in the remaining species. In fact, a paralog of this gene, *mayAcp69b*, was also found to be expressed in the accessory glands of *D. mayaguana* (Almeida and DeSalle, in review). The relative rate test showed that the evolutionary rate of *D. mayaguana* is significantly ($P < 0.00001$) higher than that of other species in all pairwise comparisons.

Indel Substitutions

The alignments of some genes analyzed here revealed many indels in their coding sequences. Particularly, *Acp7*, *Acp42*, *Acp45*, and *mayAcp69a* had large numbers of indels (table 4). Although, in theory, indels can be under positive selection, tests to demonstrate it are complicated. Indels are ignored in the likelihood models used here and in most other tests available for positive selection. One way to test whether there is positive selection for indels is to compare the rate of indel appearance in a coding region with the same rate in noncoding sequences that are likely to be neutral, weighting for divergence time (Podlaha and Zhang 2003). This test has been used to show that indels are likely under positive selection in *Acp26Aa* in the *Drosophila pseudoobscura* group (Schully and Hellberg 2006). This approach, however, is not available for the species studied here because the rate of indels in noncoding sequences is unknown for the *repleta* group and there is no reliable estimate of divergence time for the species of the *mulleri* complex. Nevertheless, some of the genes used here have introns and 3′ UTR sequences that allow for some rough comparisons. Although introns and 3′ UTRs may have regulatory function and therefore cannot be assumed to be neutral (Healy et al. 1996; Rodriguez-Trelles et al. 2002, 2003), these regions are definitely less constrained than coding sequences. Indels in noncoding sequences are naturally less constrained because the survival of a new mutation does not dependent on the number of sites involved, whereas in coding sequences, they have to be in multiples of 3 to survive.

We compared the rates of indel substitutions per base pair between coding and noncoding sequences for 12 genes (table 4). Numbers of indels were calculated by taking into account the phylogeny and disregarding the number of base pairs included in the indel. In all comparisons, the frequency of indels is larger in noncoding sequences, suggesting that indels are mostly under negative selection in coding sequences. This comparison is very conservative, and it is not an appropriate test for selection on indels. Even genes with high d_N/d_S have regions of the coding sequence under strong purifying selection, which usually contain very few or no indels. To minimize the effect of these highly constrained regions, ratios of indels per base pair in coding

Table 4
Number of Indels in Coding and Noncoding Sequences in the Alignment of Sequences of Species of the *Drosophila mulleri* Complex

	Coding Region		Introns			3′ UTR		Coding	Noncoding	c/nc
Locus	bp	Indels	*N*	bp	Indels	bp	Indels	Rate	Rate	
Acp1	318	0	1	87	11	79	5	0.000	0.096	0.000
Acp7	609	14	1	66	0	173	8	0.023	0.033	0.687
Acp11	252	2	0	—	—	116	2	0.008	0.017	0.460
Acp19	690	2	1	61	3	130	6	0.003	0.047	0.062
Acp25	363	0	1	87	4	66	1	0.000	0.033	0.000
Acp42	609	19	1	67	3	0	0	0.031	0.045	0.697
Acp45	552	16	1	59	3	117	5	0.029	0.045	0.638
mayAcp57	479	2	2	64/84	4	—	—	0.004	0.027	0.154
mayAcp58	479	1	2	55/69	2	—	—	0.002	0.016	0.129
mayAcp63	360	0	3	89/56	15	224	23	0.000	0.103	0.000
mayAcp65	255	2	0	—	—	70	2	0.008	0.029	0.275
mayAcp69a	282	9	1	137	7	105	2	0.032	0.037	0.858
mayAcp74	792	5	0	—	—	—	—	0.006	—	—
mayAcp75	576	0	1	58	1	—	—	—	0.017	—
may97	705	0	0	—	—	—	—	—	—	—

NOTE.—bp, sequence length in base pairs; *N*, the number of introns; and c/nc, ratio between indel substitution rates in coding and noncoding sequences.

sequences were recalculated for the 4 genes with the largest number of indels (*Acp7*, *Acp42*, *Acp45*, and *mayAcp69a*) using the number of base pairs (sequence length) equivalent to the proportion of codons in either class 1 ($d_N/d_S = 1$) or class 2 ($d_N/d_S > 1$) (table 5). In this way, gene regions under purifying selection are excluded, making comparisons with noncoding sequences more realistic. With this approach, the rate of indels in the coding sequence is higher than in noncoding sequences in *Acp7* and *mayAcp69a*, and very similar to that of noncoding sequences in *Acp42* and *Acp45*, suggesting that some indels in these loci are likely to be under selection.

Evolutionary and Functional Trends in Acps of the *D. mulleri* Complex

Positive Selection and Functional Categories

Among the 14 Acp genes analyzed in tests for positive selection, only 5 had a conserved protein domain (table 1). This sample is too small to assess whether adaptive evolution in Acps is related to function. Among these 5 Acps, 3 had highly significant results in the LRTs for positive selection: a serine protease (*mayAcp74*), a thiol reductase (*mayAcp58*), and cysteine-rich secreted protein (CRISP, *Acp19*). Among the 9 Acps with unknown function, 7 showed highly significant results in the tests for positive selection. Acps without a known conserved domains many times have hormonal activity and are involved in sperm competition as has been shown for *Acp70A*, *Acp26Aa*, and *Acp53Ea* of the *D. melanogaster* group (Chen et al. 1988; Clark et al. 1995; Heifetz et al. 2000). Given that genes with a d_N/d_S ratio as low as 0.64 had significant test results, it is likely that *mayAcp68b*, *may82*, and *mayAcp73* (genes for which not enough sequences were available for the tests) also have some amino acid sites under selection (supplementary table 3, Supplementary Material online). Among these, only the latter gene had a conserved protein domain, zinc-dependent metalloprotease.

The results obtained here suggest that positive selection in Acps is not directly related to function. The 2 genes containing a CRISP domain, *Acp19* and *mayAcp57*, showed very different evolutionary patterns. These genes had similar synonymous substitution rates (d_S), but the amino acid substitution rates were considerably different, leading to very different overall d_N/d_S ratios (table 2). Also, whereas *Acp19* had highly significant results in the tests for positive selection, *mayAcp57* had no support for the presence of positively selected amino acid substitutions. It is possible that the Acp function of *mayAcp57* is restricted to *D. mayaguana* because we do not have cDNA evidence for the other species. On the other hand, there is no indication that *mayAcp57* shows lineage-specific patterns in *D. mayaguana*. Codon bias of *D. mayaguana* is similar to those of the other species, and nonsignificant results were obtained when we tested for different d_N/d_S in *D. mayaguana* (using the branch model). Interestingly, *Acp19* and

Table 5
Estimates of Rates of Indels in Coding and Noncoding Sequences Taking into Account Only the Sites under Neutral or Positive Selection

Locus	bp	p1 + p2	bp (p1 + p2)	Coding Rate	Noncoding Rate	c/nc
Acp7	609	0.544	331.30	0.042	0.033	1.262
Acp42	609	0.724	440.92	0.043	0.045	0.962
Acp45	552	0.689	380.33	0.042	0.045	0.926
mayAcp69a	282	0.661	186.40	0.048	0.037	1.298

NOTE.—bp, length of the coding sequence in base pairs; p1 + p2, proportion of sites in classes 1 and 2; bp (p1 + p2), number of base pair in classes 1 and 2; and c/nc, ratio between indel substitution rates in coding and noncoding sequences.

mayAcp57 also diverge in the amount of codon bias. *Acp19* has almost null bias, whereas *mayAcp57* has the third highest codon bias among the genes analyzed here. This result raises the question of whether optimal codon selection can restrict adaptive evolution of a gene. Although the function of Acps containing the CRISP domain is not clear, this gene family includes genes involved in defense response in other organisms from plants to humans. Immune defense genes are often found to be under positive selection (Schlenke and Begun 2003; Vallender and Lahn 2004). In vertebrates, some proteins of the CRISP family are related to sperm binding at different stages of reproduction (Olson et al. 2001; Voight et al. 2006).

In order to examine the importance of positive selection in modulating changes in protein function, we analyzed the position of positively selected sites (class 2 BEB $P > 0.90$) in relation to the functional domain and active sites of 3 Acps. These were the Acp genes with a conserved functional domain and a significant result in tests for positive selection (*Acp19*, *mayAcp58*, and *mayAcp74*). In *Acp19*, the conserved CRISP domain encompassed 135 of the 230 amino acids in the coding region, starting at amino acid position 80. Five positively selected sites were found before the beginning and one after the ending of the CRISP domain. The remaining 5 positively selected sites were within the domain, 2 of which in more or less conserved sites in the domain alignments. Similar results were obtained in a study on mammalian "fertilin," a reproductive protein that also carries a CRISP domain (Civetta 2003).

In *mayAcp58* and *mayAcp74*, almost all the sites with BEB $P \geq 0.90$ of being in class 2 were within the conserved protein domains. These included 9 out of 10 sites in *mayAcp58* and all 22 sites in *mayAcp74*. In *D. mayaguana*, the gene *mayAcp74* had nonsynonymous substitutions in 1 of the 3 active sites and 1 of the 3 binding sites. It is possible that *D. mayaguana* carries a nonactive pseudogene, although the coding region was intact without early stop codons or frameshift mutations. These mutated sites were not among the ones with high probability of being under positive selection.

Both those genes have domains related to protein catalysis. *mayAcp58* contains a GILT domain, which is present in gamma interferon–inducible lysosomal thiol reductase, whose function is to catalyze thiol bond reduction, denaturating proteins and facilitating the action of proteases (West et al. 1994). *mayAcp74* is a trypsin-like serine protease, a functional category often found among Acps (Mueller et al. 2004). A third Acp protease analyzed here (*mayAcp73*) also showed a relatively high d_N/d_S, suggestive of adaptive evolution. The importance of proteolysis regulation in fertilization is ubiquitous. Proteases, reductases, and protease inhibitors are often present in the seminal fluid of a diversity of organisms. In *Drosophila*, proteolysis is implicated in the processing of other Acps, both before and after insemination (Ram et al. 2006). Acps involved in proteolysis regulation show pleiotropic and epistatic effects in sperm competition, and at least one Acp protease (CG6168) is involved in immune defense (Fiumera et al. 2007; Mueller et al. 2007). At least 2 Acp proteases have adaptive evolution in the *D. melanogaster* group (Wong et al. 2008).

Lineage-Specific Adaptive Evolution

So far, the models used here in tests for positive selection assume that the different lineages are affected by the same evolutionary forces, that is, d_N/d_S is assumed to be the same in all branches of the tree. Nevertheless, positive selection can be restricted to a node, a clade, or even to a single species. This hypothesis can be tested by selecting one or more branches (foreground) to have different d_N/d_S from that of the remaining branches (background) on a tree (using the branch model as implemented in codeml), obtaining the likelihood of this model, and comparing it to the likelihood of the null hypothesis, which is the model that assumes homogeneous d_N/d_S across branches. Although the branch model used here does not allow for testing positive selection, it can be used to show whether a certain branch has significantly faster evolution as compared with other branches of the tree. We used this approach in 2 cases where some of the results already discussed pointed to the possibility that a certain lineage might have particularly high amino acid substitution rate.

As evidenced by the relative rate tests, the low success in amplifying *D. navojoa* genes cannot be attributed to higher evolutionary rates in this species. One alternative explanation for the PCR results is related to the demographics of this species. *Drosophila navojoa* has a very limited geographic distribution and breeds exclusively in one species of cactus, which could lead—but not necessarily—to small population sizes (Ruiz et al. 1990). It has been proposed that selection is less efficient in small populations, leading to the accumulation of slightly deleterious mutations (Ohta 1973, 1993; Lynch and Conery 2003). This would lead to higher rates not only of gene loss but also of amino acid substitution. We used the branch model to test the hypothesis of a higher rate of amino acid substitution in *D. navojoa*. *Drosophila navojoa* had higher d_N/d_S than the average of the other species in only 2 genes, *Acp7* and *Acp25*, but the difference was not statistically significant (supplementary table 4, Supplementary Material online).

The second case where we used the branch model was that of the gene *mayAcp74*. The presence of amino acid substitutions in the active and binding sites of the protein exclusively in *D. mayaguana* raised the question of whether this lineage experiences relatively high rates of nonsynonymous substitution. When allowed to vary, estimates of d_N/d_S for *D. mayaguana* (1.712) were twice as large as the background d_N/d_S averaged across all the other branches (background $d_N/d_S = 0.678$). The likelihood of the model that allows for a different d_N/d_S in *D. mayaguana* in relation to that of the remaining species was significantly higher than the likelihood of the model that assumes homogeneity of d_N/d_S ratios across all species (table 6). It is possible, however, that the pattern observed was caused by positive selection acting on the ancestor of *D. mayaguana* and its sister species, *D. parisiena*. We tested this hypothesis by allowing for a different d_N/d_S ratio in the branch leading to the clade *D. mayaguana* + *D. parisiena*. The branch d_N/d_S (0.812) was not significantly different from the average across the other branches of the tree (0.767, table 6). Nevertheless, in a third model, where *D. parisiena* and *D. mayaguana* d_N/d_S ratios are set to be equal but

Table 6
Branch Model Tests Applied to *mayAcp74* to Test for Differential Selection Pressure in *Drosophila mayaguana* and *Drosophila parisiena*

Alternative Hypothesis	Null Hypothesis	ω_b	ω_f	lnL	lnL Null	LRT
may $\neq$ others	may = par = others	0.68	1.71	−3839.48	−3844.03	9.10*
(may + par) $\neq$ others	may = par = others	0.77	0.81	−3844.02	−3844.03	0.01
may = par $\neq$ others	may = par = others	0.64	1.44	−3838.51	−3844.03	11.05**
may $\neq$ par $\neq$ others	may = par $\neq$ others	0.64	1.66/1.15	−3838.25	−3838.51	0.51

NOTE.—may, *Drosophila mayaguana*; par, *Drosophila parisiena*; ω_b, background d_N/d_S; ω_f, foreground d_N/d_S; lnL, log likelihood of alternative hypothesis; and lnL, log likelihood of null hypothesis. *$P < 0.05$, **$P < 0.01$.

independent from the d_N/d_S of the remaining branches, the average d_N/d_S (1.440) of these 2 species was higher than the background ratio (0.639, $P < 0.01$). Allowing *D. mayaguana* and *D. parisiena* to vary independently from each other and from the remaining species did not increase the model likelihood, as reflected by the high d_N/d_S ratios obtained for both species (*D. mayaguana* $d_N/d_S = 1.657$ and *D. parisiena* $d_N/d_S = 1.151$; in this model, background $d_N/d_S = 0.640$). The branch models discussed here are summarized in table 6. These results suggest that *mayAcp74* could be involved in the development of reproductive isolation between the 2 sister species. Intraspecific patterns of nucleotide substitution in *D. mayaguana* and *D. parisiena* as compared with substitution patterns between species may provide further insight on the involvement of this gene in reproductive isolation.

One final point concerns the possibility that positive selection in *mayAcp74* might be restricted to *D. mayaguana* and *D. parisiena*. To address this hypothesis, we reran the tests for positive selection on this gene without the sequences of the 2 most divergent species. The results showed that positive selection is not restricted to *D. mayaguana* and *D. parisiena* (LRT = 27.030, $P < 0.001$), although considerably fewer sites were found to be under positive selection than in the tests including all the species (4 as compared with 14 with class 2 BEB $P > 0.95$).

Conclusions

Our results suggest that positive selection is very likely acting on 10 of the 15 accessory gland–expressed genes tested here, all of them likely Acps (*Acp1*, *Acp7*, *Acp11*, *Acp19*, *Acp25*, *Acp42*, *Acp45*, *mayAcp58*, *mayAcp65*, and *mayAcp74*). These genes had overall (averaged across sites and lineages) d_N/d_S between 0.635 and 1.275. Genes with $d_N/d_S < 0.60$ had negative results in the test for positive selection, suggesting that 0.6 is perhaps a more precise and conservative cutoff than 0.5, when tests are not available or accurate (e.g., when too few sequences are available). We found that positive selection on Acps is acting, many times, within the conserved protein domains and may therefore cause divergence in activity and/or substrate specificity among species. The fact that all the species analyzed here belong to a recently diverged clade shows that adaptive evolution is responsible for gene sequence divergence in a relatively short time frame. As shown for *mayAcp74*, adaptive evolution maybe involved in sister species divergence.

The results of this study confirm previous suggestions of a faster evolutionary rate in Acps of the *D. repleta* group as compared with the *D. melanogaster* and *D. pseudoobscura* groups (Wagstaff and Begun 2005; Almeida and DeSalle, in review). They show significant statistical evidence of positive selection for an additional 8 Acps of the *repleta* group. Together with the 3 Acps with positive selection detected by the MK test in Wagstaff and Begun (2005), 2 of which (*Acp1* and *Acp25*) were confirmed here, the total number of *repleta* group Acps with adaptive evolution is 11. If we include evidence of positive selection in Acp gene families (in paralog divergence; Wagstaff and Begun 2007), this number is raised to 15. Similar results have been obtained for 3 protease gene families expressed in the female reproductive tract of *D. arizonae*, one of the species analyzed here (Kelleher et al. 2007). Positive selection on reproductive molecules of both sexes in species of the *D. repleta* group suggests a role of male × female antagonistic coevolution. Such selective pressure is expected given the high remating rates observed in species in the group (Markow 1996; Dorus et al. 2004). Another explanation is increased sperm competition due to multiple inseminations in a short period of time (Dorus et al. 2004).

Supplementary Material

Supplementary tables 1–4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Cymone Speed for her help in the laboratory and 2 anonymous reviewers for their helpful comments on an earlier version of the manuscript. Funds for this research were provided by the Sackler Institute for Comparative Genomics (American Museum of Natural History), the Cullman Program in Molecular Systematics, and National Science Foundation (DEB 0129105 to R.D.). F.C.A. was supported by the Henry McCraken Fellowship (New York University).

Literature Cited

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.

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

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.

Begun DJ, Whitley P. 2002. Molecular population genetics of *Xdh* and the evolution of base composition in *Drosophila*. Genetics. 162:1725–1735.

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.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: signalP 3.0. J Mol Biol. 340:783–795.

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.

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

Civetta A. 2003. Positive selection within sperm-egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. Mol Biol Evol. 20:21–29.

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

Clark AG, Aguade 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.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat Genet. 36:1326–1329.

Durando CM, Baker RH, Etges WJ, Heed WB, Wasserman M, DeSalle R. 2000. Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. Mol Phylogenet Evol. 16:296–307.

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.

Haerty W, Jagadeeshan S, Kulathinal RJ, et al. (11 co-authors). 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics. 177:1321–1335.

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

Healy MJ, Dumancic MM, Cao A, Oakeshott JG. 1996. Localization of sequences regulating ancestral and acquired sites of esterase6 activity in *Drosophila melanogaster*. Mol Biol Evol. 13:784–797.

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

Katoh K, Misawa K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Kelleher ES, Swanson WJ, Markow TA. 2007. Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. PLoS Genet. 3:e148.

Kern AD, Jones CD, Begun DJ. 2004. Molecular population genetics of male accessory gland proteins in the *Drosophila simulans* complex. Genetics. 167:725–735.

Kim Y. 2004. Effect of strong directional selection on weakly selected mutations at linked sites: implication for synonymous codon usage. Mol Biol Evol. 21:286–294.

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.

Lynch M, Conery JS. 2003. The origins of genome complexity. Science. 302:1401–1404.

Maddison D, Maddison W. 2000. MacClade 4: analysis of phylogeny and character evolution. Sunderland (MA): Sinauer Associates.

Marchler-Bauer A, Anderson JB, Cherukuri PF, et al. (24 co-authors). 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res. 33:D192–D196.

Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32:W327–W331.

Markow TA. 1996. Evolution of *Drosophila* mating systems. Evol Biol. 29:73–106.

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.

O'Grady PM, Durando CM, Heed WB, Wasserman M, Etges W, DeSalle R. 2002. Genetic divergence within the *Drosophila mayaguana* subcluster, a closely related triad of Caribbean species in the *repleta* species group. Hereditas. 136:240–245.

Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. Nature. 246:96–98.

Ohta T. 1993. Amino acid substitution at the Adh locus of *Drosophila* is facilitated by small population size. Proc Natl Acad Sci USA. 90:4548–4551.

Olson JH, Xiang X, Ziegert T, Kittelson A, Rawls A, Bieber AL, Chandler DE. 2001. Allurin, a 21-kDa sperm chemoattractant from *Xenopus* egg jelly, is related to mammalian sperm-binding proteins. Proc Natl Acad Sci USA. 98:11205–11210.

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.

Podlaha O, Zhang J. 2003. Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc Natl Acad Sci USA. 100:12241–12246.

Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.

Ram KR, 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.

Rodriguez-Trelles F, Tarrio R, Ayala FJ. 2002. A methodological bias toward overestimation of molecular evolutionary time scales. Proc Natl Acad Sci USA. 99:8112–8115.

Rodriguez-Trelles F, Tarrio R, Ayala FJ. 2003. Evolution of cis-regulatory regions versus codifying regions. Int J Dev Biol. 47:665–673.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Ruiz A, Heed WB, Wasserman M. 1990. Evolution of the *mojavensis* cluster of cactophilic *Drosophila* with descriptions of two new species. J Hered. 81:30–42.
Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol. 12:391–404.
Schlenke TA, Begun DJ. 2003. Natural selection drives *Drosophila* immune system evolution. Genetics. 164:1471–1480.
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.
Swanson WJ, Aquadro CF, Vacquier VD. 2001. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. Mol Biol Evol. 18: 376–383.
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, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 20:18–20.
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.
Swofford DL. 2003. PAUP*. Phylogenetic anlysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.
Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. Hum Mol Genet. 13:R245–R254.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.
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.
Wagstaff BJ, Begun DJ. 2007. Adaptive evolution of recently duplicated accessory gland protein genes in desert *Drosophila*. Genetics. 177:1023–1030.
West MA, Lucocq JM, Watts C. 1994. Antigen processing and class II MHC peptide-loading compartments in human B-lymphoblastoid cells. Nature. 369:147–151.
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.
Wong A, Turchin MC, Wolfner MF, Aquadro CF. 2008. Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. Mol Biol Evol. 25:497–506.
Wong WS, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 168:1041–1051.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13: 555–556.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15:568–573.
Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 19:908–917.
Yang Z, Nielsen R, Hasegawa M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol. 15:1600–1611.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.

Jody Hey, Associate Editor

Accepted July 3, 2008