Cross-Species Comparison of Drosophila Male Accessory Gland Protein Genes

J. L. Mueller,* K. Ravi Ram,* L. A. McGraw,* M. C. Bloch Qazi,*† E. D. Siggia,†
A. G. Clark,* C. F. Aquadro* and M. F. Wolfner*‡

*Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853 and †Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10021

Manuscript received March 29, 2005
Accepted for publication May 19, 2005

ABSTRACT

Drosophila melanogaster males transfer seminal fluid proteins along with sperm during mating. Among these proteins, ACPs (Accessory gland proteins) from the male’s accessory gland induce behavioral, physiological, and life span reduction in mated females and mediate sperm storage and utilization. A previous evolutionary EST screen in D. simulans identified partial cDNAs for 57 new candidate ACPs. Here we report the annotation and confirmation of the corresponding Acp genes in D. melanogaster. Of 57 new candidate Acp genes previously reported in D. melanogaster, 34 conform to our more stringent criteria for encoding putative male accessory gland extracellular proteins, thus bringing the total number of ACPs identified to 52 (34 plus 18 previously identified). This comprehensive set of Acp genes allows us to dissect the patterns of evolutionary change in a suite of proteins from a single male-specific reproductive tissue. We used sequence-based analysis to examine codon bias, gene duplications, and levels of divergence (via $d_S$/d$_S$ values and ortholog detection) of the 52 D. melanogaster ACPs in D. simulans, D. yakuba, and D. pseudoobscura. We show that 58% of the 52 D. melanogaster Acp genes are detectable in D. pseudoobscura. Sequence comparisons of ACPs shared and not shared between D. melanogaster and D. pseudoobscura show that there are separate classes undergoing distinctly dissimilar evolutionary dynamics.

ACCESSORY gland proteins (ACPs) induce a variety of physiological, behavioral, and reproductive changes when transferred to the female. Between 25 and 150 ACPs were initially thought to be transferred to the female during mating (INGMAN-BAKER and CANDIDO 1980; SCHMIDT et al. 1985; WHALEN and WILSON 1986; COULTHART and SINGH 1988; WOLFNER et al. 1997). Males lacking ACPs have impaired fertility, indicating that ACPs perform important reproductive functions (KALB et al. 1993; XUE and NOLL 2000). Specifically, ACPs cause females to increase their egg-production, egg-laying, and ovulation rates, decrease their propensity to remate, and store and utilize sperm (reviewed in WOLFNER 2002; CHAPMAN and DAVIES 2004). ACPs also participate in formation of the mating plug (LUNG and WOLFNER 2001) and mediate a decrease in the mated female’s life span (CHAPMAN et al. 1995). Genetic analyses have revealed the functions of four ACPs thus far. Acp26Aa (ovulin) is a prohormone that triggers an increase in ovulation rate (HERNDON and WOLFNER 1995; HEIFETZ et al. 2000). Acp36DE is a glycoprotein that is essential for sperm storage (NEUBAUM and WOLFNER 1999), by regulating sperm accumulation into storage (BLOCH QAZI and WOLFNER 2003). Acp70A (sex peptide) induces egg laying and decreases females’ receptivity to remating; it also contributes to the cost of mating to females (CHEN et al. 1988; AIGAKI et al. 1991; CHAPMAN et al. 2003; LIU and KUBLA 2003; WIGBY and CHAPMAN 2005). Acp62F is a trypsin protease inhibitor that localizes to the sperm storage organs of mated females and has been suggested to preserve sperm viability (LUNG et al. 2002). Acp62F also enters the female’s circulation and is toxic to flies upon repeated ectopic expression, suggesting a possible role in the life span cost of mating (LUNG et al. 2002). In addition, the transfer of antimicrobial ACPs to the female (LUNG et al. 2001) and the Acp-induced upregulation of antimicrobial peptides in mated females (LAWNICKAK and BEGUN 2004; MCGRAW et al. 2004) suggests that ACPs may contribute to a female’s immune defense. Altogether, ACPs appear to participate in a complex set of interactions by competing/cooperating with seminal fluid proteins of other males (CLARK et al. 1995; CLARK et al. 1999; PROUT and CLARK 1996; SNOOK and HOSKEN 2004), receptors present in the female or on sperm, and pathogens. To better understand this diverse set of interactions of ACPs it is important to

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ088689–DQ088699, DQ079991–DQ079998, and BK005692–BK005702.

†Present address: Department of Biology, Nobel Hall of Science, Gustavus Adolphus College, St. Peter, MN 56082.

‡Corresponding author: Department of Molecular Biology and Genetics, 425 Biotechnology Bldg., Cornell University, Ithaca, NY 14853.

E-mail: mfw5@cornell.edu

Genetics 171: 131–143 (September 2005)
fully characterize the ACPs involved and examine their evolutionary dynamics.

Initially, 18 *Drosophila melanogaster* ACPs had been identified from multiple screens (Chen et al. 1988; Simmerl et al. 1995; Wolfnner et al. 1997); however, this was far below the predicted 25–150 ACPs (Inghman-Baker and Candido 1980; Schmidt et al. 1985; Whalen and Wilson 1986; Civetta and Singh 1995; Wolfnner et al. 1997). In an extensive screen (Swanson et al. 2001a), 57 new candidate ACPs were identified from partial gene sequencing of ESTs obtained from a *D. simulans* accessory gland cDNA library. These 57 candidate ACPs, plus the 18 previously identified, led to 75 putative ACPs. Statistical analysis of the frequency of multiple isolates predicted that these genes represented ~90% of the total number of Acp genes (Swanson et al. 2001a). The Swanson et al. (2001a) EST screen identified ACPs from partial gene sequencing and from a species in which genetic analysis is not routine, *D. simulans*. Because it is important to obtain the complete sequence of these genes in a species in which genetic analysis is possible, we obtained and report here the *D. melanogaster* orthologs of the 57 *D. simulans* Acp candidates. Our RT-PCR and bioinformatic analyses determined that 34 of the candidate 57 ACPs identified by Swanson et al. (2001a) have sequences suggestive of encoding extracellular proteins and expression patterns suggestive of encoding ACPs. This resets the total number of *D. melanogaster* ACPs identified to 52 (34 plus 18 previously identified).

An unusually high fraction of the genes encoding ACPs show signs of positive selection (Aguadé et al. 1992; Cirera and Aguadé 1997; Tsaúr and Wu 1997; Aguadé 1999; Begun et al. 2000; Panhuis et al. 2003; Kern et al. 2004; Kohns et al. 2004; Steverson et al. 2004). ACPs, as a class, evolve at about twice the rate of nonreproductive proteins (Whalen and Wilson 1986; Civetta and Singh 1995; Swanson et al. 2001a). Swanson et al. (2001a) found that ~11% of the partially sequenced ESTs that they identified have an excess of nonsynonymous over synonymous nucleotide changes, suggesting that divergence of these genes is being accelerated by positive selection. Three selective forces are predicted to drive the generation of sequence diversity of ACPs: female sperm preference (Eberhard and Cordero 1995), sperm competition (Clark et al. 1995), and sexual conflict (Rice 1996). Previous evolutionary analyses of ACPs focused on some of the initially identified 18 ACPs (Aguadé et al. 1992; Cirera and Aguadé 1997; Tsaúr and Wu 1997; Aguadé 1999; Begun et al. 2000; Kern et al. 2004). Here we present a detailed examination of the molecular evolution of the entire set of stringently selected and annotated 52 ACPs. We performed sequence-based comparisons of these *D. melanogaster* ACPs with their orthologs in three *Drosophila* species (*D. simulans*, *D. yakuba*, and *D. pseudoobscura*). This allowed us to determine levels of codon bias, rates of gene duplication, and levels of sequence divergence among three members of the *D. melanogaster* subgroup (*D. melanogaster*, *D. simulans*, and *D. yakuba*) and, via ortholog detection, which ACPs are conserved between *D. melanogaster* and *D. pseudoobscura*. These evolutionary analyses demonstrate that ACPs represent a combination of divergent and conserved proteins that undergo different patterns of sequence evolution.

**MATERIALS AND METHODS**

**Annotation of *D. melanogaster* orthologs of *D. simulans* ACP-ESTs:** We sequenced *D. simulans* Acp ESTs (Swanson et al. 2001a) from their 3'-ends to determine the translational stop position. This in combination with previously sequenced 5'-end sequences (Swanson et al. 2001a) provided each candidate ACP's complete ORF. The complete EST sequences can be found under GenBank accession nos. DQ088689–DQ088699 and DQ079991–DQ079998. These *D. simulans* EST sequences were subsequently aligned using Sequencher 4.0.5 (Gene Codes) to the *D. melanogaster* genome (Release 4.0) (Celiker et al. 2002) to identify their *D. melanogaster* orthologs. Each translational start was located by the presence of sequences encoding a predicted signal peptide, either from the Berkeley Drosophila Genome Project (BDGP) *D. melanogaster* annotation or via manual inspection if the BDGP annotation did not match the *D. simulans* EST. Manual searches for predicted signal peptides constituted scanning ~1.5 kb of noncoding upstream *D. melanogaster* sequence from the 5'-end of the *D. simulans* EST or BDGP predicted translational start site. Predicted signal peptides were identified using SignalP (Nielsen et al. 1997).

Candidate Acp genes were then examined for accessory gland-specific expression. Eighteen previously identified ACPs (Chen et al. 1988; Simmerl et al. 1995; Wolfnner et al. 1997) were already known to show accessory gland-predominant or -exclusive expression. We searched each of the 57 new candidate ACPs identified by Swanson et al. (2001a) against the *D. melanogaster* BDGP EST database (http://www.fruitfly.org/EST/EST.shtml) to see if the gene was expressed in other tissues (e.g., head, embryo, tissue culture). Occasionally adult testis ESTs (Rubin et al. 2000a) included our Acp candidates. For example, Acp36DE [a highly expressed Acp (Wolfnner et al. 1997)] has 32 testes EST hits, but has been shown by Western blots to be an accessory gland-specific protein (Bertram et al. 1996; Wolfnner et al. 1997). These differences may result from low-level contamination by accessory gland fragments or cells in the large-scale testes preparations for the EST project or may indicate that Acp36DE is transcribed, but not translated, in the testes. Consistent with such models, all 15 Acp antibodies thus far generated detect exclusively accessory gland-specific proteins, even though 9 of 15 genes [CG8982 (Acp26Aa), CG4605 (Acp32CD), CG7157 (Acp36DE), CG6289, CG8137, CG9334, CG17575, CG16566, CG9029] (Monsma et al. 1990; Coleman et al. 1995; Bertram et al. 1996; Lung et al. 2002; Ravi Ram et al. 2005) have testis EST hits (Andrews et al. 2000; Rubin et al. 2000a; Parisi et al. 2003).

Of the 57 Acp candidates previously selected by Swanson et al. (2001a), 7 were eliminated from further study because mutational analysis (per FlyBase, http://flybase.net/) indicates that their phenotypes affect nonreproductive processes. Sixteen further candidates were removed because they either had EST hits in multiple nonreproductive tissue types or could not be annotated, thus leaving 34 candidate ACPs (see http://www.genetics.org/supplemental/ for list of ACPs removed
from the previous 57 candidates). It is important to note that secreted proteins found in other tissues, or whose mutants have additional nonreproductive phenotypes, may be present in accessory gland secretions. However, we focused on accessory gland-specific candidates since the evolutionary pressures and functions of these genes should be more comprehensible than those of genes expressed in multiple tissues and thus likely having multiple functions.

This selection process resulted in a collection of 52 ACPs (Table 1). Twelve (CG1262 (Acp62F), CG4986, CG6069, CG10284, CG10956, CG11598, CG14034, CG17097, BG642378, BG642312, BG642167, and BG642163) either were not identified in or have different ORFs from those predicted in the D. melanogaster genome sequence (Release 4.0) (Celniker et al. 2002). We may have identified alternative splice forms of the predicted genes. An example is CG10956, whose Release 4.0 annotation predicts a single exon, while our annotation has identified a second exon at the 3′-end. Our annotation may have also revealed species-specific differences, since the EST library was constructed from D. simulans, or differences with the D. melanogaster annotation (Release 4.0) (Celniker et al. 2002).

We revised the current D. melanogaster (Release 4.0) annotation (Celniker et al. 2002) of the translational start sites for both CG4986 and CG10956; the splicing patterns for CG1262 (Acp62F), CG11598, CG6069, CG10284, and CG17097; and the translational start, translational stop, and splicing pattern for CG14034. Four Acp simulans ESTs (BG642378, BG642312, BG642167, and BG642163, Swanson et al. 2001a) likely represent real genes but remain unannotated in the current D. melanogaster genome Release 4.0 (Celniker et al. 2002). All genes unidentified and/or misannotated in the D. melanogaster (Release 4.0) genome annotation were submitted to GenBank under accession nos. BK005692–BK005702.

Confirmation of D. melanogaster annotations: RT-PCR of full coding regions in D. melanogaster was performed from RNA isolated from whole, 3-day-old adult virgin Canton-S males. Approximately 30 flies were homogenized in Trizol according to the manufacturer’s instructions (GIBCO, Bethesda, MD) and total RNA was prepared for RT-PCR as in Carninci and Hayashizaki (1999). Full-length coding regions were amplified using primers designed from our annotations, which verified the annotations and expression. All amplified products were PCR-purified, cloned into pENTR-DTopo or pDONR-201 vectors (Invitrogen), and sequenced by the Biotechnology Resource Center at Cornell using the vector’s internal primers. ACPs that could not be RT-PCRd from whole adult male Canton-S DNA were amplified from available EST clones (Rubin et al. 2000a) and subsequently cloned as above. Incomplete sequence information for Acp53Eb and a very short coding sequence for CG31056 (Acp98AB) (Wolnick et al. 1997) did not allow cloning into pENTR-DTopo or pDONR-201 vectors. Complete coding, amino acid, and primer sequences for each of the 34 new ACPs can be found in the supplemental materials (http://www.genetics.org/supplemental/).

D. melanogaster Acp sequence analysis: Codon bias: Codon bias was measured by both the frequency of optimal codons (Fop) and the percentage G/C content in the third codon position (G/C3rd) (Moriyama and Powell 1997). Fop values range from 0.33 to 1, where 0.33 indicates homogeneous codon usage and 1 indicates that only optimal codons are used. Fop, G/C3rd, and gene GC content calculations were performed using the codonw program (http://www.molbiol.ox.ac.uk/cu/). Codon bias values (see http://www.genetics.org/supplemental/) were calculated using the D. melanogaster codon frequency table settings of the codonw program. Previous codon bias analysis of CG32952 (Acp33A) in D. melanogaster to D. simulans comparisons have combined its two ORFs (CG32952-A and CG32952-B) (Begun et al. 2000); however, since each ORF contains its own predicted signal sequence we performed our analysis as two separate genes.

For comparison, we generated a random sample of 100 D. melanogaster genes showing twofold higher expression in testes vs. ovaries from the Parisi et al. (2004) microarray data set. Additionally, a random sample of 150 D. melanogaster genes with approximately the same gene lengths as ACPs (Acp mean gene nucleotide length, 994.7; random gene nucleotide length, 957.6) was obtained from BDGP (http://www.fruitfly.org/sequence/dlMisc.shtml).

Gene duplications: Sequence comparisons and chromosomal location were used together to identify gene duplicates. Individual Acp protein sequences were compared to the D. melanogaster genome using BlastP. Acp gene duplicate candidates were considered if they had a conservative E-value of 10−10 and a minimum of 30% sequence identity across ≥80% of the protein (Gu et al. 2002). Because many gene duplicates often are found in tandem (Friedman and Hughes 2003) we extended our search to locate significant matches falling within neighboring Acp genes that did not meet the >30% sequence identity cutoff. If such a hit was present, we checked for a similar protein domain prediction (http://www.crysl.bioc.cam.ac.uk/~fugue/phsearch.html) and conserved splicing pattern to support its being a possible duplicate. Candidates that both had a BlastP E-value of 10−10 or smaller and matched all three sequence search criteria were also considered gene duplicates, even though their sequence identity may be <30%. Gene duplication conservation in D. simulans and D. yakuba was searched via tBlastN to their whole-genome alignments (WashU-GSC, http://genome.wustl.edu/tools/blast/).

Calculation of the expected number of ACPs in the D. melanogaster genome: Two estimates of the total number of ACP genes in the D. melanogaster genome were performed as in Swanson et al. (2001a) by using maximum-likelihood fits to a truncated Poisson distribution. A second estimate was obtained by nonparametric maximum likelihood. The first two predictions differ with respect to how they deal with 5 ACPs (Acp26Aa, Acp26Ab, Acp32CD, Acp33A, Acp36DE) that were not adequately prescreened by Swanson et al. (and hence appeared in the postscreening library). In the first estimate, we ignore the 5 ACPs prescreened by Swanson et al. (2001a) and fit a truncated Poisson distribution to the frequency spectrum (counts of singleton hits, doubleton hits, etc.). This gives a maximum-likelihood count of 52 ACPs in addition to the 18 that were prescreened by Swanson et al. (2001a), for a total of 70. For the second estimate, we include the 5 Acp hit counts as though they were not prescreened at all, and we obtain a maximum-likelihood count of 59 ACPs. If the 13 ACPs that were successfully prescreened (or at least were not observed among the sequenced clones) are added back to the estimate of 59 ACPs, this yields a prediction of 72 ACP genes in the D. melanogaster genome. The third method was designed for an unscreened library, and fits the data to a Poisson mixture model by nonparametric maximum likelihood (Ji-Ping Wang, personal communication). The perl script eststat.pl (available at http://www.floralgenome.org/cgi-bin/eststat/eststat.cgi) took the frequency spectrum of EST hits and produced an estimate of the total count of distinct ACPs in the library at 106. This figure may be considered as an upper bound because of the prescreening that was applied to the library, leaving a more uniform frequency distribution than would be found in an unscreened library.

D. simulans and D. yakuba sequence comparisons to D. melanogaster ACPs: Nonsynonymous substitutions per nonsynonymous site (dN) and synonymous substitutions per synonymous site (dS) values for some previously characterized ACPs
Detection of Acp orthologs in *D. pseudoobscura*: The whole-genome alignment (WGA) of the *D. melanogaster* and *D. pseudoobscura* genome (Richards et al. 2005) was taken from (Emberly et al. 2003). The SMASH program (Zavolan et al. 2003) was used to find the strongest set of syntenic anchors between the *D. pseudoobscura* contigs and the *D. melanogaster* genome. Anchors were high-similarity regions from 10s to 100s of base pairs and covered ~30% of the genome. The LAGAN program (Brudno et al. 2005) gave similar alignments. Since the size of syntenic domains between the two species generally exceeds 10 kb (i.e., much larger than most repeat elements within the sequenced euchromatin), using synteny eliminated almost all ambiguities due to repeats. The SMASH blocks along with the contigs they matched were displayed on top of the Release 3 annotation (Celniker et al. 2002) using GBROWSE (http://www.gmod.org/ggb/index.shtml).

We then examined the syntenic regions for each Acp individually at the sequence level. In 48 of 51 cases (51 ACPs instead of 52 were compared because Acp53Eb’s sequence information has yet to be determined), SMASH blocks from a single contig either bracketed or “hit” the annotated gene in *D. melanogaster*. SMASH blocks from a single *D. pseudoobscura* contig that span a given Acp locus indicate that the Acp genomic region in question can be aligned at the sequence level. For CG31872, a contiguous *D. pseudoobscura* sequence could not be aligned because the Acp gene fell into a gap between two contigs. Two other cases, CG14560 and CG9074, contained SMASH block hits to multiple contigs that differed from the contig spanning this region. The coding sequence of CG14560 and CG9074 were then submitted to Repeatmasker (http://www.repeatmasker.org/), which indicated that both ACPs contained repetitive regions, thus explaining the multiple SMASH block contiguous hits. After filtering out the repetitive regions for CG14560 and CG9074, we could generate a single contig that bracketed each gene. Upon verification of the *D. melanogaster* to *D. pseudoobscura* contiguous alignments of the ACPs, we retrieved the corresponding *D. pseudoobscura* sequence within the aligned contig and searched the *D. pseudoobscura* contig sequence via tBLASTN using the *D. melanogaster* protein sequence. If coding sequence alignments could not be identified, we used GENSCAN (Burge and Karlin 1997) and Genie (Reese et al. 2000) to locate possible ORFs. All ACPs for which coding sequence alignments could be generated with the corresponding *D. pseudoobscura* contiguous region are considered true orthologs (Table 1). The SMASH block-based coding sequence alignments were confirmed using another more recent *D. pseudoobscura* WGA (Karolchik et al. 2003). *D. pseudoobscura* coding sequences of conserved ACPs and *D. melanogaster* to *D. pseudoobscura* contig alignments for absent or undetectable ACPs can be found in the web supplement (http://www.genetics.org/supplemental/).

It is important to note that even though we define conserved ACPs between *D. melanogaster* and *D. pseudoobscura* as true orthologs, we have not determined whether these ACPs have maintained their accessory gland expression in *D. pseudoobscura*.

*RESULTS AND DISCUSSION*

*D. melanogaster* Acp genes: Secreted proteins synthesized by the *D. melanogaster* male accessory gland have important functions in reproduction (reviewed in Wolner 2002; Kubli 2003; Chapman and Davies 2004). To address more thoroughly the functions and evolution of these Acp proteins, we carried out a comprehensive identification and annotation of *D. melanogaster* Acp genes. Prior to 2001, 18 Acp genes had been reported in *D. melanogaster* (Chen et al. 1988; Simmerl et al. 1995; Wolner et al. 1997). In 2001, Swanson et al. (2001a) identified 57 additional candidate Acp genes in *D. simulans* via an evolutionary EST approach that was performed to permit a rapid scan to identify genes with features suggesting rapid evolution. However, the ESTs identified by Swanson et al. (2001a) were partial cDNAs and from a species, *D. simulans*, which is presently less amenable to genetic analyses than is *D. melanogaster*. We therefore full-length sequenced a select set of the *D. simulans* EST sequences identified by Swanson et al. (2001a). The full-length *D. simulans* Acp EST sequences allowed us to identify the complete ORF of their *D. melanogaster* orthologs. We then applied a more stringent set of criteria to identify those genes on which to focus, based on what is known of the initial 18 ACPs. We define *bona fide* ACPs here as genes that: (a) encode a protein with a predicted secretion signal sequence, (b) have a pattern of EST hits in other tissue- or cell-type-specific EST screens
TABLE 1
Cross-species comparisons of sequence divergence levels and ortholog detection analyses for individual *D. melanogaster* ACPs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Functional class</th>
<th>Conserved in <em>D. pseudoobscura</em></th>
<th><em>D. melanogaster</em> ACPs not identifiable via our WGA detection methods</th>
<th>Without a <em>D. pseudoobscura</em> true ortholog</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG1262 (Acp62F)</td>
<td>Trypsin protease inhibitor</td>
<td>0.050 0.126 0.399 0.219 0.359 0.611 + NA</td>
<td>- +</td>
<td>-</td>
</tr>
<tr>
<td>CG1462^*</td>
<td>Alkaline phosphatase</td>
<td>0.013 0.125 0.107 0.054 0.348 0.097 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG1652^*</td>
<td>C-type lectin</td>
<td>0.021 0.148 0.140 0.052 0.281 0.186 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG1656^*</td>
<td>C-type lectin</td>
<td>0.016 0.096 0.171 0.071 0.562 0.573 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG3359^*</td>
<td>Fasciclin</td>
<td>0.010 0.091 0.104 0.006 0.200 0.028 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG4605 (Acp32CD)</td>
<td>Cysteine protease</td>
<td>0.014 0.012 1.176 0.069 0.220 0.312 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG4847^*</td>
<td>Serine protease</td>
<td>0.020 0.114 0.177 0.043 0.282 0.151 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG6069^*</td>
<td>Serine protease</td>
<td>0.016 0.131 0.119 0.120 0.404 0.297 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG6168^*</td>
<td>Serine protease</td>
<td>0.036 0.176 0.203 0.085 0.365 0.232 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG8093^*</td>
<td>Acid lipase</td>
<td>0.005 0.119 0.039 0.016 0.399 0.040 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG8194^*</td>
<td>RNase</td>
<td>0.011 0.121 0.094 0.020 0.276 0.072 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG8622 (Acp53Ea)</td>
<td></td>
<td>0.039 0.143 0.275 0.120 0.282 0.425 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG9024 (Acp26Ab)</td>
<td></td>
<td>0.018 0.059 0.305 0.150 0.356 0.422 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG9029^*</td>
<td></td>
<td>0.077 0.156 0.494 0.309 0.444 0.695 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG9997^*</td>
<td>Serine protease</td>
<td>0.031 0.108 0.291 0.100 0.393 0.254 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG10284^*</td>
<td>CRISP</td>
<td>0.044 0.106 0.413 0.135 0.348 0.387 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG10365^*</td>
<td>α-Macroglobulin</td>
<td>0.015 0.071 0.210 0.036 0.291 0.123 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG10397 (1a8)^*</td>
<td>Defensin</td>
<td>0.061 0.027 0.317 0.057 0.105 0.547 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG11598^*</td>
<td>Acid lipase</td>
<td>0.026 0.217 0.121 0.466 1.421 0.328 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG11864^*</td>
<td>Metalloprotease</td>
<td>0.029 0.090 0.227 0.099 0.339 0.293 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG13399^*</td>
<td></td>
<td>0.029 0.117 0.246 0.085 0.301 0.382 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG16707^*</td>
<td></td>
<td>0.052 0.108 0.483 0.064 0.211 0.304 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17097^*</td>
<td>Acid lipase</td>
<td>0.010 0.125 0.082 0.064 0.208 0.390 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17575^*</td>
<td>CRISP</td>
<td>0.007 0.165 0.039 0.016 0.172 0.093 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17673 (Acp70A)</td>
<td></td>
<td>0.028 0.124 0.227 0.146 0.294 0.497 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17845^*</td>
<td>Thioredoxin</td>
<td>0.019 0.110 0.175 0.059 0.403 0.147 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17924 (Acp95EF)</td>
<td></td>
<td>0.037 0.223 0.164 0.259 0.411 0.630 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG18284^*</td>
<td>Acid lipase</td>
<td>0.034 0.190 0.179 0.104 0.424 0.244 + NA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG32952-A (Acp33A)</td>
<td></td>
<td>0.007 0.081 0.085 0.129 0.331 0.243 + NA</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* D. melanogaster ACPs conserved in *D. pseudoobscura* are listed first and *D. melanogaster* ACPs not identifiable via our WGA detection methods are listed second. c/b, contig breakpoint at site of an Acp, presence of ortholog undeterminable; TBN, TBLASTN hits against the *D. pseudoobscura* genome; WGA, SMASH blocks-based whole-genome alignment identification of true Acp orthologs in *D. pseudoobscura*; sim, *D. simulans*; yak, *D. yakuba*; δk, nonsynonymous nucleotide substitution value; δs, synonymous nucleotide substitution value; NA, not applicable.
* The 34 new ACPs selected from Swanson et al. (2001a) EST that fit our newly defined criteria.
consistent with accessory gland predominant expression, (c) have no previously characterized non-Acp function, and (d) show male and/or accessory gland predominant expression in *D. melanogaster*. Using these stringent criteria we utilized secretion signal prediction programs, EST databases, reports of mutant phenotypes, and RT-PCR to screen through the 57 candidate ACPs identified by Swanson et al. (2001a) (see MATERIALS AND METHODS for details). Thirty-four ACPs fit these new stringent criteria (see Table 1). The other 23 genes identified by Swanson et al. (2001a) could encode proteins made in accessory glands and potentially also transferred to females, but their additional tissues of expression and/or nonreproductive functions complicate genetic and functional analyses and evolutionary interpretations; thus, we do not consider them further. It is also formally possible that the expression characteristics of some of these 23 ACPs differ in *D. melanogaster* and *D. simulans*, resulting in their exclusion from the stringently selected *D. melanogaster* ACPs on which we focus. The 34 stringently selected *D. melanogaster* ACPs that fit the above criteria, in combination with the 18 previously known ACPs, make a total of 52 *D. melanogaster* ACPs (Table 1) whose gene boundaries and expression have been confirmed. This comprehensive and characterized set of 52 ACPs has also allowed a recalculating of the predicted number of ACPs in the genome. Fitting the frequency spectrum of the 52 ACPs with EST hits from the Swanson et al. (2001a) screen to a truncated Poisson distribution and to a Poisson mixture model gave maximum-likelihood estimates in the range of 70–106 ACPs in the *D. melanogaster* genome (see MATERIALS AND METHODS), respectively. Additionally, recently identified ACPs CG8626, CG15616, and CG17799 (Holloway and Begun 2004) suggest that the field is steadily approaching a complete list of ACPs in *D. melanogaster*.

These 52 *D. melanogaster* ACPs are expected to be extracellular and thus transferred to the female upon mating and to be produced primarily or exclusively in the male’s accessory gland. Indeed, all 16 Acp genes tested so far encode seminal proteins detectable only in the male’s accessory gland and transferred to the female during mating (Chien et al. 1988; Monsma et al. 1990; Coleman et al. 1995; Bertram et al. 1996; Lung et al. 2002; Albright 2003; Ravi Ram et al. 2005). Additional support that this set of 52 *D. melanogaster* ACPs truly represents accessory-gland predominant genes stems from the finding that 29 of 46 tested Acp genes showed twofold or higher expression values in germlineless male vs. germlineless female comparisons (6 ACPs were not present on the microarrays) (Parist et al. 2004).

**Presence of multiple Acp gene duplicates across the *D. melanogaster* genome:** About 40% of the *D. melanogaster* genome (5536 of 13601 genes) appears to be gene duplicates (Rubin et al. 2000b). Similarly, 16 (31%) of the 52 ACPs appear to have gene duplicates (Table 2) within the *D. melanogaster* genome. CG8137 and CG9334 are the only gene duplicates not in tandem, although they share the same intron splice positions. Percent identities of the Acp gene duplicates range from 25 [CG3801 (Acp76A) and BG642378] to 92% (CG6289 and CG6663), indicating that a range of recent and ancient gene duplicates have been identified.

Nine of these cases of gene duplication are within the 52 Acp collection (3 duplicate pairs plus 1 triplicate) (Table 2B), indicating that these duplicates have similar expression profiles. This is consistent with the observation that gene duplication events often lead to coexpressed genes that cluster together (Boutanaev et al. 2002). These 9 Acp gene duplicates are found in tandem clusters of paired (or triplicate) genes, and they share the same relative splice site positions, which are also conserved in *D. simulans* and *D. yakuba*.

For seven additional ACPs we detect duplicates in the genome (Table 2A). Again, tandem arrangements are seen in *D. simulans* and *D. yakuba*, and the *D. melanogaster* duplicates share the same splice site positions. However, in these seven cases, only one member of each duplicate pair is a member of our 52-Acp collection. This could be because the collection is incomplete (only 52 of the predicted 70–106 ACPs are described here), because a given duplicate’s expression might not fit our stringent criteria of accessory gland-predominant expression, or because a given duplicate has an entirely different expression pattern. An example of the first is CG17799. This gene duplicate of CG17797 (Acp29AB) has recently been shown to also be expressed in the *D. melanogaster* accessory gland (Holloway and Begun 2004), but is not among the 52 genes we focused on here, simply because it was not detected in the Swanson et al.
Possible that these low levels of codon bias could be due to rapid rates of protein evolution of ACPs (Akashi 1994). Drosophila codon bias can also be influenced by sequence length (Duret and Mouchiroud 1999), expression level, and local GC content. Because short Drosophila genes tend to exhibit high levels of codon bias (Duret and Mouchiroud 1999), and because Acp genes also tend to be short, our control set was selected to be genes of similar length to avoid the contribution of gene length. The unusual levels of codon bias seen for both ACPs and testis-genes (Table 3) suggest that male-reproductive proteins in general may exhibit lower levels of codon bias. Low levels of codon bias for D. melanogaster tests genes is consistent with their poorly conserved sequence and sex-specific expression pattern when compared to Anopheles gambiae (Paris et al. 2003) or D. simulans (Ranz et al. 2003), respectively. That male-biased genes evolve more rapidly at the sequence (Singh and Kullathinal 2000) and expression pattern levels (Meiklejohn et al. 2003) suggests that their rapid evolution may not allow adaptation to high levels of codon bias.

Levels of divergence: A high $d_{SN}/d_{SV}$ ratio can identify genes for which amino acid replacement is being driven by a selective pressure. Acp genes have already been reported to demonstrate higher levels of sequence divergence than non-Acp genes between D. simulans and D. melanogaster (Swanson et al. 2001a; Kern et al. 2004; Stevison et al. 2004). However, those analyses used only partial sequences or included genes that our present analyses have shown not to fit the stringent definition of ACPs in D. melanogaster and thus could be subject to additional or different selection pressures.

Here we compare our complete sequences of a set of stringently selected D. melanogaster ACPs with their
D. simulans and D. yakuba orthologs. ACPs exhibit high levels of sequence divergence with average $d_N$ values for D. simulans of 0.045 (Table 1), similar to previously reported $d_N$ values for D. simulans ACPs of 0.052 (Swanson et al. 2001a) and 0.050 (Begun et al. 2000). The average level of $d_S$ for this set of ACPs in D. simulans is 0.13 (Table 1), similar to the known average D. simulans $d_S$ value of 0.11 (Bauer and Aquadro 1997; Moriyama and Powell 1997; Begun and Whitley 2000; Betancourt et al. 2002). We also compared Acp to non-ACP levels of sequence divergence between D. melanogaster and D. yakuba (Table 4). In this comparison as well, ACPs have significantly higher $d_S$ (0.161) and $d_S/d_N$ (0.407) values than non-ACP ($d_S$ and $d_S/d_N$ values of 0.026 and 0.082, respectively) (Mann-Whitney test, both $d_N$ and $d_S/d_N$, $P < 0.001$).

Using levels of $d_N$ and $d_S$ as a metric to identify rapidly evolving genes, which have a $d_S/d_N$ value $>1$, Swanson et al. (2001a) identified 19 genes whose partial sequence had $d_S/d_N > 1$ in D. melanogaster/D. simulans comparisons. However, our reanalysis of the 52 ACPs using complete gene sequences yields only 3 ACPs from both D. melanogaster/D. simulans and D. melanogaster/D. yakuba comparisons with $d_S/d_N > 1$ (Table 1). We believe this discrepancy between the Swanson et al. (2001a) results and those reported here is because we analyzed full-length coding regions from an accurately annotated list of genes instead of partially sequenced cDNAs, which in some cases were misaligned. In addition, for many rapidly evolving genes only part of the gene is under positive selection (Hughes and Nei 1988; Swanson et al. 2001b). Thus, some partial cDNAs analyzed by Swanson et al. (2001a) may have fortuitously contained regions under positive selection giving a higher $d_S/d_N$ than when the entire gene is tested. For this reason a $d_S/d_N > 0.5$ was recently proposed as a more practical cutoff when using full-length sequences, to identify candidate genes that may be driven by positive selection (Swanson et al. 2004). Applying this cutoff value of 0.5 to the 52 ACPs we find that 9 ACPs (but not the same 9 as in Swanson et al. 2001a) in both D. melanogaster/D. simulans and D. melanogaster/D. yakuba have $d_S/d_N > 0.5$. This proportion of ACPs (9/52, 17%) is similar to the percentage of ACPs identified in the Swanson et al. (2001a) male accessory gland EST screen (19%) with a $d_N/d_S > 0.5$. Comparable percentages of ACPs with a $d_N/d_S > 0.5$ described here to those ACPs identified in Swanson et al. (2001a) support the idea that $d_N/d_S > 0.5$ may serve as a good indicator for candidate rapidly evolving genes (Swanson et al. 2004). Further analysis of the role of natural selection in shaping Acp sequence evolution using codon-substitution models will be presented elsewhere.

Detection of D. melanogaster Acp orthologs in D. pseudoobscura: The complete genome sequence of D. pseudoobscura (Richards et al. 2005) allowed us to search for conserved D. melanogaster ACPs in a distantly related species outside of the D. melanogaster subgroup. A whole-genome alignment (WGA) approach was used to determine which of the 52 D. melanogaster ACPs can be identified in D. pseudoobscura. Syntenic regions covering each Acp were generated for 50 ACPs. Limited sequence information for the other 2 ACPs (Acp53Eb and CG31872) prevented generation of accurate comparative genome sequence alignments. We verified all the D. melanogaster to D. pseudoobscura contig alignments and identified the corresponding D. pseudoobscura Acp, to generate coding sequence alignments between the two species. All D. melanogaster ACPs for which coding sequence alignments could be generated with the corresponding D. pseudoobscura contig are considered true orthologs (see Table 1). We found that, via WGA, 58% (29/50) of the D. melanogaster ACPs have true orthologs in D. pseudoobscura (Table 1). For the 21 D. melanogaster ACPs for which true orthologs could not be identified in D. pseudoobscura we used tBlastN against all D. pseudoobscura contigs and orphan sequences to ensure that we had not missed D. melanogaster ACPs that had moved to non-syntenic chromosomal locations in D. pseudoobscura. In 10 cases, tBlastN comparisons gave significant D. pseudoobscura hits. However, each hit was interpreted as a false positive because it matched either repetitive sequence in the Acp or a different D. pseudoobscura gene with a respective non-Acp D. melanogaster ortholog (see MATERIALS AND METHODS). Our inability to detect a D. pseudoobscura ortholog for a D. melanogaster Acp gene via this method does not mean that a D. pseudoobscura ortholog does not exist, but only that our searches were negative. D. melanogaster ACPs undetectable in D. pseudoobscura via our methods could be highly diverged, located in an unsequenced region of the D. pseudoobscura genome, or potential D. melanogaster lineage-specific proteins. A recent study (Wagstaff and Begun 2005) uncovered a D. pseudoobscura gene with 18.5% amino

<table>
<thead>
<tr>
<th>Divergence level averages</th>
<th>ACPs</th>
<th>Non-ACPs</th>
<th>Mann-Whitney test ($P$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster/D. yakuba $d_N$</td>
<td>0.161</td>
<td>0.026</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>D. melanogaster/D. yakuba $d_S$</td>
<td>0.397</td>
<td>0.306</td>
<td>0.002</td>
</tr>
<tr>
<td>D. melanogaster/D. yakuba $d_S/d_N$</td>
<td>0.407</td>
<td>0.082</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Nonsynonymous ($d_N$) and synonymous ($d_S$) nucleotide substitution rates.
TABLE 5
Distinct sequence evolution patterns for D. melanogaster ACPs present and undetectable in D. pseudoobscura

A.

<table>
<thead>
<tr>
<th>Averages</th>
<th>D. pseudoobscura orthologs present</th>
<th>D. pseudoobscura orthologs undetectable</th>
<th>Mann-Whitney test (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fop(^d) D. melanogaster</td>
<td>0.439</td>
<td>0.351</td>
<td>0.001</td>
</tr>
<tr>
<td>G/C3rd(^d) D. melanogaster</td>
<td>0.559</td>
<td>0.456</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fop(^d) D. yakuba</td>
<td>0.467</td>
<td>0.371</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>G/C3rd(^d) D. yakuba</td>
<td>0.598</td>
<td>0.474</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(d_{S}/d_{S}) D. simulans</td>
<td>0.240</td>
<td>0.525</td>
<td>0.002</td>
</tr>
<tr>
<td>(d_{S}/d_{S}) D. yakuba</td>
<td>0.287</td>
<td>0.515</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th>Averages</th>
<th>All other predicted functional classes</th>
<th>Protease inhibitors</th>
<th>Mann-Whitney test (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fop(^d) D. melanogaster</td>
<td>0.476</td>
<td>0.365</td>
<td>0.002</td>
</tr>
<tr>
<td>G/C3rd(^d) D. melanogaster</td>
<td>0.596</td>
<td>0.477</td>
<td>0.009</td>
</tr>
<tr>
<td>Fop(^d) D. yakuba</td>
<td>0.503</td>
<td>0.395</td>
<td>0.007</td>
</tr>
<tr>
<td>G/C3rd(^d) D. yakuba</td>
<td>0.602</td>
<td>0.498</td>
<td>0.005</td>
</tr>
<tr>
<td>(d_{S}/d_{S}) D. simulans</td>
<td>0.173</td>
<td>0.496</td>
<td>0.001</td>
</tr>
<tr>
<td>(d_{S}/d_{S}) D. yakuba</td>
<td>0.235</td>
<td>0.477</td>
<td>0.004</td>
</tr>
</tbody>
</table>

\(^d\)High values are associated with codon bias for both the frequency of optimal codons (Fop) and the percentage of GC bases in the third position (G/C3rd).

\(^d\)Nonsynonymous (\(d_{S}\)) to synonymous (\(d_{S}\)) nucleotide substitution ratios.

acid sequence identity to D. melanogaster Acp26Aa. This is below the similarity level detectable in our search for D. pseudoobscura orthologs. For another gene, Acp95EF, our analysis revealed its D. pseudoobscura ortholog, which was undetected by Wagstaff and Begun (2005). Differences in methodologies and the limited alignability of the D. pseudoobscura genome (only \(\sim 48\%\); Richards et al. 2005) likely account for these two differences in Acp ortholog detection.

Of the 29 ACPs we found conserved between D. melanogaster and D. pseudoobscura it had been possible to generate comparative structural models to known protein classes for 20 (Table 1) (Mueller et al. 2004). This represents a greater fraction (20/29, 69\%) than is seen for those D. melanogaster ACPs that do not have D. pseudoobscura counterparts (9/21, 43\%). That more proteins within predicted protein functional classes are conserved between D. melanogaster and D. pseudoobscura suggests that these proteins may mediate reproductive strategies that are conserved across Drosophila. Interestingly, the protease inhibitor class is not well conserved between the two species (Table 1): only one (Acp62F) of seven predicted or known Acp protease inhibitors is identifiable between the two species (Table 1). The lack of conservation of protease inhibitors between D. melanogaster and D. pseudoobscura is significantly greater than the percentage of ACPs not shared in all other protein classes (chi-square = 12.28, d.f. = 1, \(P < 0.001\)). ACPs that are predicted protease inhibitors have been suggested to participate in sperm storage, cost of mating [specifically Acp62F (Lung et al. 2002)], and/or immune regulation (Khush and Lemaitre 2000; McGraw et al. 2004), which may contribute to their evolution between D. melanogaster and D. pseudoobscura lineages.

**Comparative sequence analysis within the D. melanogaster subgroup of ACPs shared or not shared with D. pseudoobscura:** Within the set of ACPs conserved between D. melanogaster and D. pseudoobscura, we examined levels of codon bias and \(d_{S}/d_{S}\) with two other species in the D. melanogaster subgroup. We tested whether codon bias and \(d_{S}/d_{S}\) values could distinguish those D. melanogaster ACPs that share or do not share true orthologs in D. pseudoobscura. We find that D. melanogaster ACPs without detectable D. pseudoobscura true orthologs have significantly lower levels of codon bias in D. melanogaster (Fop and G/C3rd Mann-Whitney test, \(P = 0.001\) and \(P < 0.001\), respectively) and D. yakuba than ACPs conserved between D. melanogaster and D. pseudoobscura (Fop and G/C3rd Mann-Whitney test, \(P < 0.001\) and \(P < 0.001\), respectively, Table 5A). Additionally, levels of \(d_{S}/d_{S}\) are significantly higher for D. melanogaster/D. simulans and D. melanogaster/D. yakuba comparisons of ACPs without true orthologs in D. pseudoobscura compared to ACPs conserved between D. melanogaster and D. pseudoobscura (D. simulans and D. yakuba, Mann-Whitney test, \(P = 0.002\) and \(P < 0.001\), respectively, Table 5A). This subgroup divergence analysis can be extended to the case of the D. melanogaster predicted protease inhibitor ACPs that do not have counterparts in D. pseudoobscura (Table 1). We find that the seven predicted or known Acp protease inhibitors
have both significantly lower levels of codon bias and higher levels of sequence divergence \((d_K/d_S)\) than ACPs in other predicted functional classes (Table 5B). Together, these results suggest that *D. melanogaster* ACPs without a true *D. pseudoobscura* ortholog have greater levels of sequence divergence \((d_K/d_S)\) within the *D. melanogaster* subgroup than *D. melanogaster* ACPs with a detectable *D. pseudoobscura* ortholog. Those *D. melanogaster* ACPs with higher sequence divergence levels that do not have a true ortholog in *D. pseudoobscura* thus serve as good candidates for mediating reproductive functions in close relatives of *D. melanogaster*.

**Underrepresentation of ACPs on the *D. melanogaster* X chromosome:** As previously reported (Wolfner et al. 1997; Swanson et al. 2001a), ACPs’ chromosomal locations are biased to autosomes in *D. melanogaster*. Only 1 of the 52 ACPs, CG11664, falls on the X chromosome at cytological band 1D2 in *D. melanogaster*. The remaining 51 ACPs are evenly distributed across the second (27 ACPs) and third (24 ACPs) chromosomes. Given that the X chromosome contains \(\sim 17\%\) of the total *D. melanogaster* genome (Celniker et al. 2002), if the 52 ACPs were randomly distributed across the genome we would expect \(\sim 9\) of the 52 ACPs to fall on the X chromosome and 43 on autosomes. The presence of only a single X-linked Acp is highly unlikely to have occurred by chance \((G_{\text{corr}} = 7.908, \text{d.f.} = 1, P = 0.005)\), supporting reports that the *D. melanogaster* X chromosome is deficient in male-biased genes (Wolfner et al. 1997; Andrews et al. 2000; Swanson et al. 2001a; Parisi et al. 2003; Ranz et al. 2003).

An alternative approach to understanding the chromosomal bias of sex-specific genes is to focus on the region that contains the single X-linked *D. melanogaster* Acp. The 50-kb region flanking CG11664 is unusual in several respects. First, CG11664 lies in an apparently gene-poor region, with only six other genes within the surrounding 100 kb. On average there are \(\sim 11\) genes/100 kb in the *D. melanogaster* genome (= 13792 genes/120 Mb) (Adams et al. 2000; Celniker et al. 2002). Second, of the 6 neighboring genes, 4 (CG3713, CG11663, CG14634, and CG14635) appear to be testis biased in their expression (Andrews et al. 2000; Parisi et al. 2004); no expression data could be found for CG14632 and CG14633); thus, perhaps this region is a “hotspot” for harboring male-biased genes on the X chromosome. Third, more than half of the genes in this region do not appear to be conserved between *D. pseudoobscura* and *D. melanogaster*, consistent with the report that male-biased genes tend to evolve more rapidly at both expression (Ranz et al. 2003) and sequence (Parisi et al. 2003) levels. Fourth, five of the six neighboring ORFs, in addition to CG11664, are intronless, suggesting they may be retrogenes. Additionally, this region appears to also be a hotspot for transposable elements. In the recent transposable element (piggyBac and P element) insertion mutagenesis collection release of 16,500 fly lines (Thibault et al. 2004), the 100-kb region surrounding CG11664 contained 34 insertions, which is more than the average of \(\sim 14\) transposable elements/100 kb (= 16,500 elements/120 Mb). Altogether, the region surrounding CG11664 contains a number of unique features that may help determine what pressures are driving the evolution of sex-specific genes on the X chromosome in *D. melanogaster*.

Multiple hypotheses including sexual antagonism, dosage compensation, and X inactivation may explain the paucity of male-biased genes on the *D. melanogaster* X chromosome (reviewed in Oliver and Parisi 2004). The ability to help distinguish the importance of these phenomena could be assisted by looking at *D. pseudoobscura*. In *D. pseudoobscura*, the X chromosome consists primarily of a region largely syntenic to the left arm of the third chromosome in *D. melanogaster* (3L) that fused more recently in the *D. pseudoobscura* lineage to a region syntenic to the X chromosome of *D. melanogaster* (Segarra and Aguade 1992). Thus, all ACPs with *D. pseudoobscura* orthologs that are located on 3L in *D. melanogaster* [CG1262 (Acp62F), CG10852 (Acp63F), CG17673 (Acp70A), CG3801 (Acp76A), CG6289, CG13309, CG14560, BG642312, CG16707, CG8194, BG642378, and CG6168] would now be on the right arm of the *D. pseudoobscura* X chromosome (XR). If there is selection against X linkage for ACPs, we would expect a higher “loss” of ACPs from the “new” (3L) homolog X-linked genes in the *D. pseudoobscura* lineage than for ACPs on autosomes in *D. pseudoobscura*. We find that a larger proportion of new ACPs on the *D. pseudoobscura* X chromosome are not shared between the two species (as compared to autosomal ACPs in *D. pseudoobscura*), although this difference is not statistically significant [*D. pseudoobscura* X chromosome (7/13 = 54% absent or undetected) vs. autosomes (13/36 = 36% absent or undetected); chi-square = 1.01, d.f. = 1, \(P = 0.322\)]. That fewer X-linked *D. pseudoobscura* ACPs are conserved than autosomal ACPs is consistent with selection against X-linked Acp’s. However, the *D. melanogaster* 3L's base chromosome and its *D. pseudoobscura* XR counterpart show the second lowest level of genome sequence alignability between species: 46.5% of *D. melanogaster* 3L’s base pairs are alignable with *D. pseudoobscura* XR as compared to an average across all chromosomes of 48%. Therefore, the relatively low sequence conservation of the *D. pseudoobscura* XR arm suggests that loss or translocation of ACPs from this arm may have resulted from the particular X-chromosomal evolutionary dynamics in the *D. pseudoobscura* lineage rather than from any sex-specific selection acting differentially on X chromosomes vs. autosomes.

**Conclusions:** Genes with increased rates of evolution increase the frequency with which incompatibilities evolve between closely related species. Since some ACPs in Drosophila evolve faster than other genes, these
rapidly evolving ACPs serve as good candidates for examining the selection pressures associated with reproductive functions. We have characterized here such divergent ACPs, whose divergence may be attributable to sexually antagonistic evolution with proteins from the female or male (Swanson et al. 2001b; Swanson and Vacquier 2002). The female’s genotype has been shown to play an active role in spermatogenesis (Clark and Begun 1998) and a recent EST screen identified a number of candidate receptors/sexually antagonistic genes for ACPs (Swanson et al. 2004). Candidate receptors would likely serve as the most upstream female genes in signaling pathways for the numerous biological processes/pathways regulated by ACPs, sperm, and the act of mating (McGraw et al. 2004). The comprehensive set of ACPs described here thus provides a basis for understanding both the evolutionary dynamics and function of specific ACPs. This, in turn, may help tease apart the functional importance of male-female interactions during the evolution of reproductive isolation.

We thank all members of the Wolfner lab for comments on the manuscript, S. Albright and S. Ji for assistance with the annotations, and V. DuMont, A. Fiumera, G. Reeves, and A. Wong for helpful advice. This work was supported by National Institutes of Health Grants HD38921 to M.C.B.Q. was supported by an American Cancer Society Postdoctoral Fellowship. M.C.B.Q. was supported by an American Cancer Society Postdoctoral Fellowship. We appreciate permission (from Richard K. Wilson) to use the D. yakuba genomic sequences generated by the Genome Sequencing Center, Washington University-St. Louis School of Medicine, prior to their publication.

LITERATURE CITED


Communicating editor: L. Harshman