Molecular Characterization and Evolution of a Gene Family Encoding Both Female- and Male-Specific Reproductive Proteins in *Drosophila*

Laura K. Sirot,[†,1,2] Geoffrey D. Findlay,[†,1,3] Jessica L. Sitnik,[†,1] Dorina Frasheri,[‡,1] Frank W. Avila,[1] and Mariana F. Wolfner*[,1]

[1]Department of Molecular Biology and Genetics, Cornell University

[2]Department of Biology, College of Wooster

[3]Department of Biology, College of the Holy Cross

[†]These authors contributed equally to this work.

[‡]Present address: Neuroscience Center Zürich and Brain Research Institute, ETH Zürich and University of Zürich, Zürich, Switzerland

***Corresponding author:** mfw5@cornell.edu.

Associate editor: Patricia Wittkopp

Abstract

Gene duplication is an important mechanism for the evolution of new reproductive proteins. However, in most cases, each resulting paralog continues to function within the same sex. To investigate the possibility that seminal fluid proteins arise through duplicates of female reproductive genes that become "co-opted" by males, we screened female reproductive genes in *Drosophila melanogaster* for cases of duplication in which one of the resulting paralogs produces a protein in males that is transferred to females during mating. We identified a set of three tandemly duplicated genes that encode secreted serine-type endopeptidase homologs, two of which are expressed primarily in the female reproductive tract (RT), whereas the third is expressed specifically in the male RT and encodes a seminal fluid protein. Evolutionary and gene expression analyses across *Drosophila* species suggest that this family arose from a single-copy gene that was female-specific; after duplication, one paralog evolved male-specific expression. Functional tests of knockdowns of each gene in *D. melanogaster* show that one female-expressed gene is essential for full fecundity, and both female-expressed genes contribute singly or in combination to a female's propensity to remate. In contrast, knockdown of the male-expressed paralog had no significant effect on female fecundity or remating. These data are consistent with a model in which members of this gene family exert effects on females by acting on a common, female-expressed target. After duplication and male co-option of one paralog, the evolution of the interacting proteins could have resulted in differential strengths or effects of each paralog.

***Key words: Drosophila*, seminal proteins, protease, spermathecal proteins, gene duplication, sex-specific expression.**

Article

Introduction

Gene duplication provides an opportunity for the evolution of novel protein functions. New functions could result from mutations in the protein-coding region of one of the paralogs or through modifications to transcript splicing patterns (Ohno 1970; Lynch and Conery 2000; Zhang 2003; Roth et al. 2007). New functions also could result from mutations in regulatory regions that alter the tissue specificity, timing, and/or level of gene expression (Lynch and Force 2000; Ohta 2003; Chan et al. 2010; Makino and McLysaght 2010) or from differential susceptibility to epigenetic modifications (Rodin and Riggs 2003). Such changes in the expression of paralogs can have profound impacts on developmental patterns (e.g., Hox genes; Holland and García-Fernández 1996) and have been proposed to provide a resolution to intralocus sexual conflict through the evolution of differential expression patterns in the two sexes (Stewart et al. 2010; Connallon and Clark 2011; Gallach and Bertran 2011; but see Harano et al. 2010; Hosken 2011).

Across a range of taxa, gene duplication has played a major role in shaping a species' suite of reproductive proteins. In abalone (genus *Haliotis*), extant species have two major acrosomal proteins with distinct functions: lysin, which dissolves a hole in the egg's vitelline envelope, and Sp18, which mediates sperm–egg fusion (Lewis et al. 1982; Swanson and Vacquier 1995). Both proteins have evolved rapidly, but their similar sizes, molecular weights, three-dimensional structures, and patterns of exons and introns suggest an ancient duplication event, presumably followed by subfunctionalization (Vacquier et al. 1997; Kresge et al. 2001). In insects, a duplication predating the divergence of many orders gave rise to several isoforms of tubulin, including the highly conserved testis-specific isoform β2 (Smith et al. 2007; Nielsen et al. 2010) that is essential for sperm axoneme function (Raff

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

et al. 2000; Nielsen et al. 2006). Some insect lineages (including wasps, bees, pea aphids, and stalk-eyed flies) experienced subsequent duplications resulting in new paralogs that are more rapidly evolving (Nielsen et al. 2010; Baker et al. 2012). Other examples of reproductive gene duplication are seen in mammalian ovary Nod-like receptors (Tian et al. 2009), a set of tandemly duplicated serine proteases expressed in female *Anopheles gambiae* reproductive tracts (RTs; Mancini, Tammaro, et al. 2011), and a variety of testis-expressed *Drosophila* genes (e.g., Parsch et al. 2001; Torgerson and Singh 2004; Gao et al. 2011).

Functional consequences of reproductive protein duplication have been studied in several cases involving *Drosophila*. In one example, *Gld2* arose from the duplication of the ancestral *wispy* (*wisp*) gene (also found in *Caenorhabditis* and *Xenopus*; Sartain et al. 2011). The WISP and GLD2 proteins are each cytoplasmic regulators of mRNA poly(A) tail stability, but whereas WISP (and its worm and frog orthologs) act in the female germline (Sartain et al. 2011), GLD2 is instead expressed specifically in the male germline (Sartain et al. 2011). As such, WISP is required for female fertility (playing essential roles in oogenesis and egg activation; Benoit et al. 2008; Cui et al. 2008), and GLD2 is necessary for male fertility (playing essential roles in the production of mature sperm; Sartain et al. 2011). Duplication events may also provide evidence for coevolution of reproductive proteins between the sexes. For example, in a desert *Drosophila* species, *Drosophila arizonae*, female RTs express functional paralogs of several proteases (Kelleher et al. 2007), while male seminal fluid contains several duplicated protease inhibitors (Kelleher et al. 2009). Proteolytic enzymatic activity in female RTs decreases upon mating, suggesting interactions between proteases in the female and inhibitors from the male (Kelleher and Pennington 2009).

Gene duplication has been an important force for generating diversity among seminal fluid proteins (Sfps), a class of proteins that have dramatic effects on female physiology and behavior (Sirot et al. 2009; Avila et al. 2011). Rodent and primate lineages have each independently experienced several duplication events that have given rise to seminal vesicle secretion (Svs) genes (Clauss et al. 2005; Lin et al. 2005; Hurle et al. 2007). In several species of *Anopheline* mosquitoes, duplications have generated three paralogs (termed *AgAcp334A1-3*) of a Sfp hypothesized to regulate sperm motility (Rogers et al. 2009; Mancini, Baldini, et al. 2011). Proteomic analysis in *D. melanogaster* found that >30 (of 133) transferred Sfps were encoded by genes with apparent tandem paralogs encoding other transferred Sfps (Findlay et al. 2008). Finally, there are several cases of recently duplicated, adaptively evolving seminal protein pairs between the closely related *D. arizonae* and *Drosophila mojavensis* (Wagstaff and Begun 2007). Thus, duplication of existing Sfp-encoding genes is an important source of Sfp diversity across taxa.

Gene duplication also could act to incorporate new types of proteins into seminal fluid through the duplication and subsequent change in expression pattern of non-Sfp encoding genes (similar to β2 tubulin). Evidence for such reproductive "co-option" of proteins comes from observations that members of large gene families, which typically function outside of reproduction, are sometimes found in the seminal fluid. In *Drosophila*, serine proteases and protease homologs (Ross et al. 2003), odorant binding proteins (Graham and Davies 2002; Hekmat-Scafe et al. 2002), and acid lipases (Horne et al. 2009) have been detected in the seminal fluid (Findlay et al. 2008). In these cases, the reproduction-specific family members are found in specific clades or in single lineages on the protein phylogenetic tree surrounded by nonreproductive family members, suggesting that their co-option occurred relatively recently. Similar examples have been found in the seminal fluid of other insects (e.g., Sirot et al. 2008, 2011; Rogers et al. 2009) and of mammals (e.g., Clauss et al. 2005; Tian et al. 2009).

The duplication of a secreted female RT gene (analogous to the case of WISP/GLD2) could be an evolutionarily rapid means of creating an effective Sfp. Such a mechanism, previously undocumented to our knowledge, would allow the newly derived Sfp to have built-in functionality in the female and, because of its secretion signal sequence, would already be a prime candidate for transfer during mating. To search for such a case, we screened 20 secreted, female-specific reproductive proteins in *D. melanogaster* for paralogs that are known Sfps. We report the first case of the evolution of a novel Sfp through the duplication of a gene that has highly biased expression in the female RT. We examine the evolution and expression patterns of this gene family across *Drosophila* species and find evidence for at least two rounds of gene duplication followed, in one case, by a switch in expression from the female RT to the male RT. Each of the five species in the *melanogaster* subgroup has three paralogs, which have maintained consistent patterns of expression. RNA interference testing of the genes singly and in combination in *D. melanogaster* suggests that members of this gene family affect both egg-laying rate and the probability of remating.

Results

A Targeted Search for Female Reproductive Proteins with Sex-Switched Duplicates

Our targeted search for potential female RT proteins with Sfp paralogs yielded two sets of candidate genes that fit our search criteria (table 1). The first set includes the lipases Yp1, Yp2, and Yp3—expressed in the fat body and spermathecae (Chintapalli et al. 2007)—which share highest sequence similarity with each other but are next most closely related (as measured by sequence similarity) to the Sfp CG5162. However, based on percent protein sequence identity and phylogenetic clustering (Horne et al. 2009), CG5162 appears more closely related to two other proteins: CG5665 (expressed in embryos; St Pierre et al. 2013) and CG18258 (expressed in male accessory glands and female spermathecae; Chintapalli et al. 2007). The evolutionary relationship between CG5162 and the Yp proteins is thus unclear. In contrast, the second set of proteins that fit our criteria includes three members that all share highest sequence

Table 1. Genes Highly Expressed in *Drosophila melanogaster* Female Sperm Storage Organs with Information on the Presence and Identity of Seminal Fluid Protein Paralogs.

Family	Gene	Class[a]	Sfp Paralog	References
Spermathecal endopeptidases (SEND)	*CG17012 (Send1)*	Serine protease	None	Allen and Spradling (2008), Arbeitman et al. (2004), Chintapalli et al. (2007), Prokupek et al. (2009)
	CG17234			
	CG17239			
	CG17240 (ser12)			
	CG18125 (Send2)			
	CG31861			
Inactive spermathecal endopeptidases	*CG9897*	Inactive serine protease	*CG32833*	Allen and Spradling (2008), Chintapalli et al. (2007), Prokupek et al. (2009)
	CG32834			
Yolk protein	*CG2985 (Yp1)*	Lipase	*CG5162*	Allen and Spradling (2008), Prokupek et al. (2009)
	CG2979 (Yp2)			
	CG11129 (Yp3)			
Other	*CG6426*[b]	Destabilase	None	Allen and Spradling (2008), Prokupek et al. (2009)
	CG13318	Serine protease	None	Chintapalli et al. (2007)
	CG18067[b]	3′,5′-cyclic-nucleotide phosphodiesterase activity	None	Allen and Spradling (2008), Prokupek et al. (2009)
	CG18525	Serine protease inhibitor	None	Allen and Spradling (2008), Prokupek et al. (2009)
	CG18628[b]	No conserved domains	None	Prokupek et al. (2009)
	CG30371	Serine protease	None	Prokupek et al. (2009)
	CG31686	No conserved domains	None	Chintapalli et al. (2007)
	CG32277	Serine protease	None	Allen and Spradling (2008), Chintapalli et al. (2007), Prokupek et al. (2009)
	CG32751	Hydrolase	None	Chintapalli et al. (2007)

[a]Based on Flybase.org (Marygold et al. 2013), SMART (Letunic et al. 2012), and Pfam classifications (Punta et al. 2012).
[b]Seminal fluid protein-encoding genes (Swanson et al. 2001; Findlay et al. 2008).

similarity with each other. CG9897 and CG32834, which are reciprocal best BLAST hits and both highly expressed in the spermathecae, share highest sequence similarity with the Sfp CG32833. The three genes encoding these proteins are clustered in a 4-kb region of chromosome 2R. The next closest gene to this cluster (*CG9896*) is over 7 kb away from the 5′-end of *CG9897* and shows no protein sequence similarity to CG9897, CG32834, and CG32833. Because all of the genes in this cluster are located together in the same genomic region, and because the group had no confounding, nonreproductive member (such as in the first case), we focused on this second group of genes for subsequent evolutionary, expression, and functional analysis. Additionally, we found three genes (underlined in table 1) that are highly expressed in the female sperm storage organs (Chintapalli et al. 2007) and also encode Sfps (Findlay et al. 2008), demonstrating that secreted RT proteins can be shared between the two sexes.

Identification of Family Members in Non-*melanogaster* Species

Putative orthologs were identified based on reciprocal best BLASTP results. We identified the three members of this gene family in five other *Drosophila* species (*simulans*, *sechellia*, *yakuba*, *erecta*, and *ananassae*), and one in *Drosophila pseudoobscura*. We noticed a fourth paralog, GF11311, annotated in *Drosophila ananassae*, suggesting an extra round of gene duplication in this lineage. However, RT–PCR and sequencing showed that this gene is misannotated; as expressed, the gene contains a premature stop codon and thus is likely to be a nonfunctional pseudogene (Supplementary text S1 and fig. S1, Supplementary Material online). The putative orthologs of each *melanogaster* gene are listed in figure 1; as shown below, a combination of gene expression analysis, phylogenetic clustering, and conservation of gene order within each syntenic region strongly suggests that these genes are true orthologs, so we refer to them as such below.

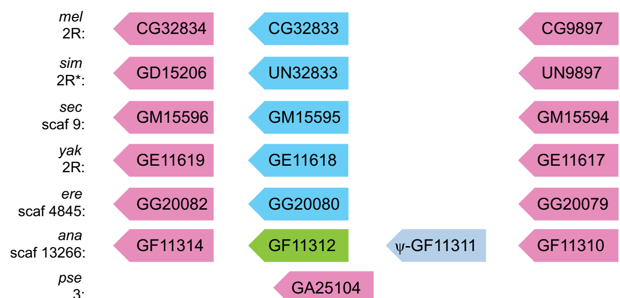

Fig. 1. Chromosomal locations and gene order of *CG9897*, *CG32833*, and *CG32834* and their orthologs in *Drosophila* species. In *D. simulans*, *GD15206* is found in an unassembled part of chromosome (chr) 2R (indicated by the asterisk), whereas *UN32833* and *UN9897* represent unannotated copies whose sequences we determined by sequencing or BLAST. The *Dsim\UN32833* sequence is only partially determined (the 142 codons at the start of the coding sequence). Color indicates gene expression pattern: pink is female-specific or female-biased; bright blue is male specific; light blue is expressed only at a low level in males; green is expressed approximately equally in both sexes. Gene order and conserved expression patterns were consistent with calls of orthology and with phylogenetic clustering (see fig. 4).

	Drosophila melanogaster gene names			*D. ananassae*/ *D. pseudoobscura* gene names	Positive control
	CG32834	*CG32833*	*CG9897*	*GF11310*/ *GA25104*[a]	*RpL32*
	F M g N	F M g N	F M g N	F M g N	F M g N
melanogaster					
simulans					
sechellia					
yakuba					
erecta					
ananassae					
pseudo-obscura					

Fig. 2. Whole animal expression patterns of *Drosophila melanogaster CG32834*, *CG32833*, and *CG9897* and their orthologs in other congeners. F: female; M: male; g: genomic; N: negative control (water used as template). *RpL32* is a ribosomal protein-encoding gene used as a control; primers for this gene were designed to span an intron to check for contamination of cDNA with genomic DNA. Gene names of the orthologs from conspecific species are given in figure 1. Note: (a) *GF11310* is a *D. ananassae* homolog to the gene family but is not definitively orthologous with any one gene. *GA25104* is a *D. pseudoobscura* homolog to the gene family but is not definitively orthologous with any one gene.

Gene Expression

The gene expression pattern for each of the three *D. melanogaster* genes and almost all of their orthologs is either sex-specific or strongly sex biased (figs. 2 and 3). Most orthologs of *CG32834* and *CG9897* are expressed only in females (fig. 2). The exceptions (*CG9897* and its orthologs in *Drosophila yakuba* and *Drosophila simulans*) all have strong female-biased expression. In contrast, *CG32833* and all of its orthologs in the *melanogaster* subgroup are expressed limited to or strongly biased in males, whereas the *D. ananassae* ortholog, *GF11312*, is expressed in both sexes at apparently equivalent levels. In *D. pseudoobscura*, the single copy of this gene family (*GA25104*) is expressed only in females.

Based on data from microarrays (Chintapalli et al. 2007), the three *D. melanogaster* genes were thought to be expressed primarily or exclusively in the reproductive accessory glands of males (*CG32833*) or the spermathecae of females (*CG32834* and *CG9897*). We confirmed this expression pattern using RT-PCR and further tested for tissue-specific expression patterns of the orthologs found outside of the *melanogaster* subgroup, in *D. ananassae* and *D. pseudoobscura*. In all cases, expression was either limited to or strongly biased in the somatic reproductive tissue (RT without gonads), when compared with the gonads and the carcass (fig. 3). Interestingly, although the *D. ananassae* gene *GF11312* (the CG32833 ortholog) shows no sex bias in expression, it is expressed solely in the somatic reproductive tissues of each sex (fig. 3).

Phylogenetic Analysis

We constructed a protein sequence tree illustrating the degree of amino acid sequence similarity between the core protease domain (corresponding to residues 34–267 of CG32833; for the complete alignment, see supplementary file S1, Supplementary Material online) of each protein in this family across the *melanogaster* group (*D. melanogaster*–*D. ananassae*), using the single *D. pseudoobscura* copy as the outgroup (fig. 4). Phylogenetic clustering was consistent with both the chromosomal order of the genes found in each species and the patterns of expression, supporting our above determinations of orthology. For example, CG32834 and its orthologs are all found at the downstream end of the gene cluster in the genomes of their respective species (fig. 1), and all show female-specific/biased expression. The consensus tree formed from 100 bootstrap replicates generally supported the observed tree topology. However, bootstrap support for the most ancestral nodes was low, making it difficult to infer the order of duplication events that gave rise to the extant gene families. We also observed lower bootstrap support for more recent nodes showing the divergence of *D. erecta* and *D. yakuba*, but this result is commonly observed for genes from this pair of species (*Drosophila* 12 Genomes Consortium 2007). The more important point is that the tree shows distinct, well-supported clades for each group of orthologs, with the only major ambiguity relating to the assignment of *D. ananassae* GF11310. GF11310 does not cluster with any

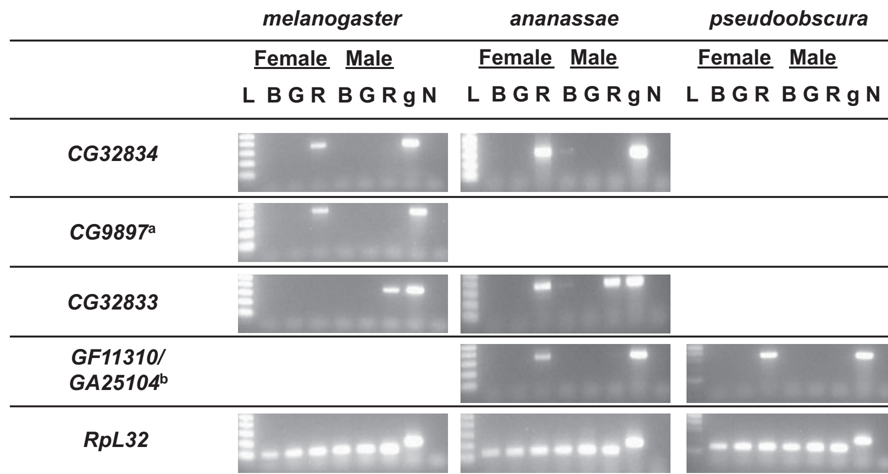

FIG. 3. Tissue-specific expression patterns of *Drosophila melanogaster CG32834*, *CG32833*, and *CG9897* and their orthologs in *D. ananassae*. Expression patterns are also shown for *GA25104*, the single *D. pseudoobscura* member to the gene family. L: ladder; B: body without RT; G: gonads; R: RT without gonads; g: genomic DNA; and N: negative control (water used as template). *RpL32* expression was used as a control. Notes: (a) At 35 cycles, a light band appears in the male gonads and RTs without gonads of Dmel for *CG9897*. (b) *GF11310* is a *D. ananassae* homolog to the gene family but is not definitively orthologous with any one gene. *GA25104* is a *D. pseudoobscura* homolog to the gene family but is not definitively orthologous with any one gene.

of the three groups of orthologs, but its female-specific expression pattern and its position at the upstream end of the cluster (fig. 1) suggest that it is orthologous to the CG9897 group of genes. The lack of clustering on the tree could be explained by the high level of divergence between GF11310 and other members of the gene family and/or the poor resolution of ancestral nodes.

Phenotypes in *Drosophila melanogaster*

We examined the phenotypic effects of the three *D. melanogaster* genes using RNAi-mediated knockdown. Specifically, we tested for the effect of knockdown on two postmating phenotypic responses in females known to be influenced by Sfps (Ravi Ram and Wolfner 2007b; LaFlamme et al. 2012): probability of remating and the number of eggs laid.

Probability of Remating

We evaluated the effects of knocking down each gene in the cluster on female remating; results are summarized in table 2 and fig. 5 (all χ^2 and *P*-values result from logit loglinear tests). In the first set of experiments (Trial 1; fig. 5*A*), we tested for remating after an initial mating between control females and females knocked down for *CG9897* and *CG32834* (individually and in combination) with either control males or males knocked down for *CG32833*. Knockdown of the female-expressed genes tended to cause females to be less receptive to remating, but different genes had effects at different timepoints and that depended on whether they were knocked down individually or in combination. Four days after an initial mating, knockdown of *CG9897* significantly reduced the probability of remating ($\chi^2_1 = 6.0$; $P = 0.01$). Ten days after an initial mating, there was an interaction effect of knockdown of *CG9897* and *CG32834* ($\chi^2_1 = 4.9$; $P = 0.02$) such that when either gene was knocked down individually, females tended to be more likely to remate than controls, but when the two genes were knocked down simultaneously, females were less likely to remate than controls. There was also a nonsignificant trend for females to be more likely to remate after mating to males knocked down for *CG32833* than after mating with control males ($\chi^2_1 = 2.9$; $P = 0.09$).

To check for replication of the effects we observed in the Trial 1 matings, we performed a second set of experiments (Trial 2; fig. 5*B*) in which we tested for remating in females knocked down for *CG9897* and *CG32834* (individually and in combination) after an initial mating to wild-type (Canton S) males. We observed the same general pattern as in Trial 1 (table 2 and fig. 5). At 4 days after the initial mating, knockdown of *CG9897* significantly reduced the probability of remating ($\chi^2_1 = 6.5$; $P = 0.01$). At 10 days after the initial mating, there was again a significant interaction effect ($\chi^2_1 = 7.6$; $P = 0.006$) of *CG32834* and *CG9897* such that when *CG32834* was knocked down individually, females were more likely to remate than controls, but when the two genes were knocked down simultaneously, females were less likely to remate than controls. However, in contrast to Trial 1, in Trial 2 we found females were less likely to remate 1 day after the initial mating when *CG32834* was knocked down ($\chi^2_1 = 3.7$; $P = 0.05$) and also at 10 days after the initial mating when *CG9897* was knocked down ($\chi^2_1 = 27.2$; $P < 0.0001$). Thus, knockdown of the female-expressed genes generally causes reduced levels of remating.

Number of Eggs Laid

We conducted two trials to assess the effects of knocking down each gene individually and in combination on the number of eggs laid; results are summarized in table 3 and fig. 6. In the first trial (Trial 1; fig. 6*A*), we compared the number of eggs laid for 5 days between control and *CG9897* and *CG32834* knockdown females (individually and in combination) after mating to either control males or males knocked down for *CG32833*. We found that knockdown of *CG32834* significantly reduced the total number of eggs laid

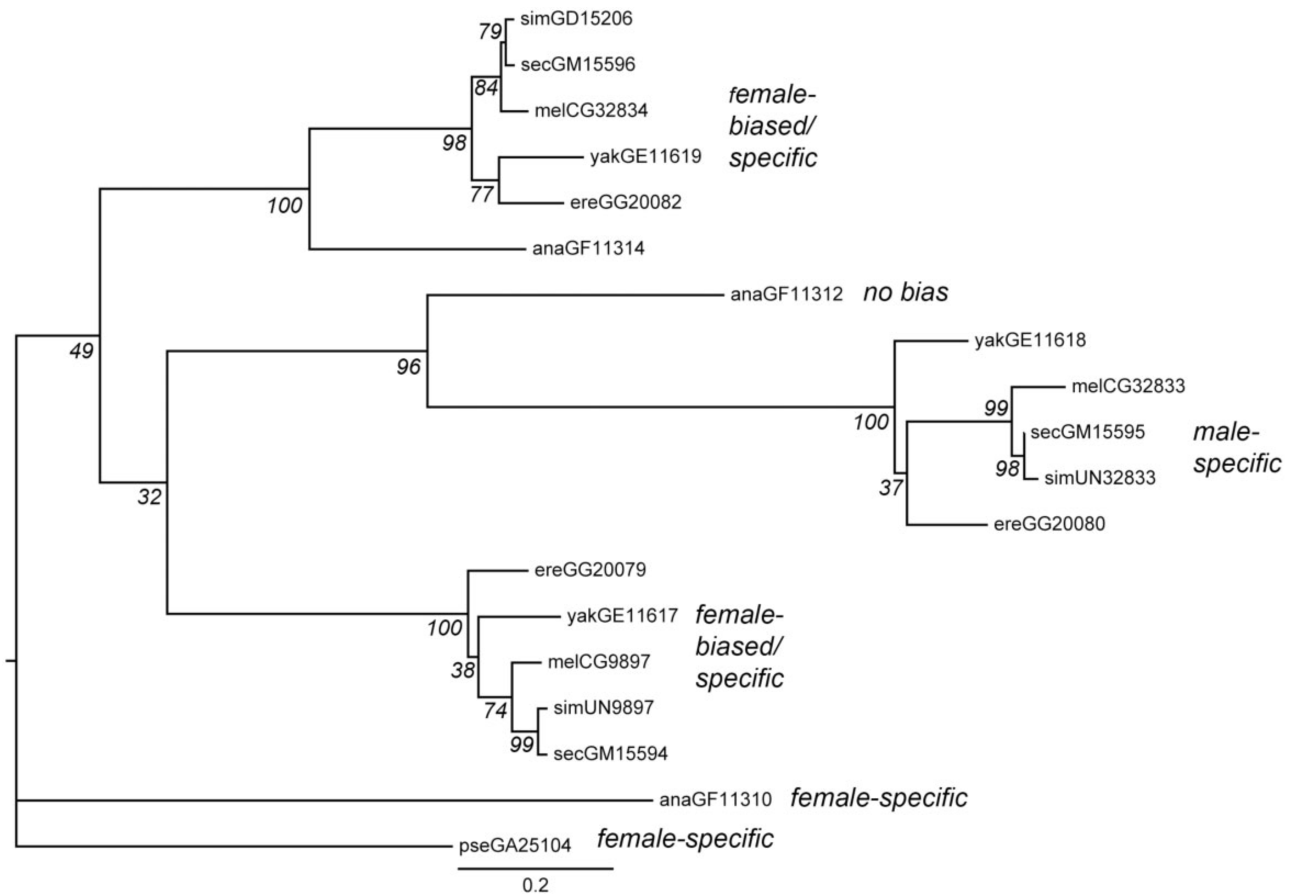

Fig. 4. Maximum-likelihood phylogeny of protein sequences for each member of the gene family. Bootstrap support based on 100 replicates is shown in italics at each node. Tip labels indicate protein names; the first three letters indicate the *Drosophila* species (mel: *melanogaster*; sim: *simulans*; sec: *sechellia*; yak: *yakuba*; ere: *erecta*; ana: *ananassae*; pse: *pseudoobscura*), and the following characters indicate the FlyBase gene name. "UN" in the gene name indicates a previously unannotated copy of the gene in *D. simulans*. Scale bar indicates the number of substitutions per site. Calls of orthology are consistent with phylogenetic clustering and gene order (see fig. 1): the six genes shown at the top of the figure (*GD15206*–*GF11314*) are one set of orthologs, *GF11312*–*GG20080* are another set, and *GG20079*–*GF11310* are the third set. The tree is rooted on the single *D. pseudoobscura* copy of this gene family, *GA25104*. Gene expression patterns from figure 2 are indicated in italicized text.

Table 2. Results of Analyses of Remating Patterns by Females after Initial Matings between Females and Males Knocked Down for Various Combinations of *CG32834*, *CG9897*, and *CG32833*.

	Trial 1[a]			Trial 2[b]		
Gene that was knocked down	**1 day**	**4 days**	**10 days**	**1 day**	**4 days**	**10 days**
CG32833	n.s.[c]	n.s.	n.s.	No data[c]	No data	No data
CG9897	n.s.	$\chi^2_1 = 6.0$*	n.s.	n.s.	$\chi^2_1 = 6.5$**	$\chi^2_1 = 27.2$***
CG32834	n.s.	n.s.	$\chi^2_1 = 3.9$*	$\chi^2_1 = 3.7$*	n.s.	$\chi^2_1 = 4.7$*
CG9897 and *CG32834*	n.s.	n.s.	$\chi^2_1 = 4.9$*	n.s.	n.s.	$\chi^2_1 = 7.6$**
CG9897 and *CG32833*	n.s.	n.s.	n.s.	No data	No data	No data
CG32834 and *CG32833*	n.s.	n.s.	n.s.	No data	No data	No data
CG9897, *CG32834*, and *CG32833*	n.s.	n.s.	n.s.	No data	No data	No data

Note.—Data were analyzed using a generalized linear model with a logit link in JMP. Variables were eliminated using backward iteration until only variables with P values ≤ 0.15 remained.

[a]In Trial 1, control females or females knocked down for each female gene (individually and in combination) were mated to control males or males knocked down for *CG32833* for the initial mating (on Day 0) and then tested for remating with a Canton S male on the indicated day. Day 1: $N = 17$–37 females per treatment; Day 4: $N = 15$–37 females per treatment; Day 10: $N = 73$–88 females per treatment (three replicates combined).

[b]In Trial 2, control females or females knocked down for each female gene (individually and in combination) were mated to wild-type males (Canton S) for the initial mating (on Day 0) and then tested for remating with a Canton S male on the indicated day. Day 1: $N = 37$–45 females per treatment (two replicates combined); Day 4: $N = 35$–53 females per treatment (two replicates combined); Day 10: $N = 88$–108 females per treatment (four replicates combined).

[c]n.s.: not significant; no data: not tested.

*$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$.

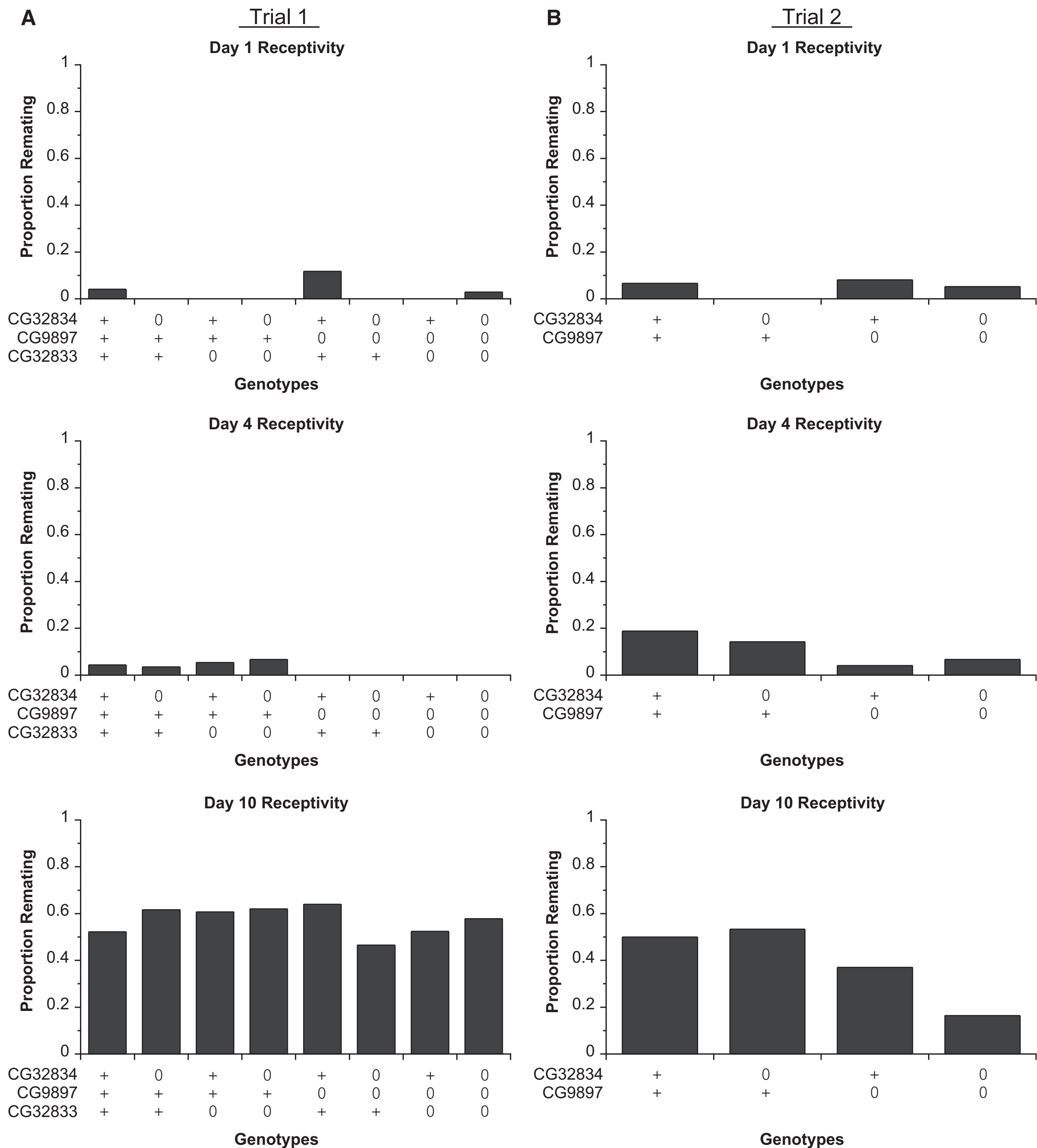

Fig. 5. Probability of remating by *Drosophila melanogaster* females after initial matings between females and males knocked down for various combinations of *CG32834*, *CG9897*, and *CG32833*. (*A*) Remating probabilities at 1, 4, and 10 days after an initial mating for Trial 1. Trial 1 includes all eight possible combinations of gene knockdown for female genes *CG32834* and *CG9897* and male gene *CG32833*. Gene presence is indicated on the *x*-axis of each graph with a " + ," whereas knockdown is indicated by a "0." Day 1: $N = 17–37$ females per treatment; Day 4: $N = 15–37$ females per treatment; Day 10: $N = 73–88$ females per treatment (three replicates combined). (*B*) Remating probabilities at 1, 4, and 10 days after an initial mating for Trial 2. Trial 2 included all four possible combinations of gene knockdown for female genes *CG32834* and *CG9897*. All males used in this trial were the wild-type Canton S stock. Day 1: $N = 37–45$ females per treatment (two replicates combined); Day 4: $N = 35–53$ females per treatment (two replicates combined); Day 10: $N = 88–108$ females per treatment (four replicates combined).

after a single mating (fig. 6A, table 3; CG32834: $F_{1,333} = 189.5$, $P = 0.046$). None of the other gene knockdowns individually or in combination significantly affected the number of eggs laid.

To check for replication of the egg-laying patterns in Trial 1, we conducted a follow-up experiment (Trial 2; fig. 6B) in which we mated *CG32834* and *CG9897* knockdown females (individually and in combination) to wild-type males (Canton S). Again, the only significant effect was that knockdown of *CG32834* reduced the number of eggs laid ($F_{1,127} = 4.06$, $P = 0.046$; fig. 6B and table 3).

Table 3. Results of Analyses of Number of Eggs Laid after Matings between Females and Males Knocked Down for Various Combinations of *CG32834*, *CG9897*, and *CG32833*.

Gene that was knocked down	Trial 1	Trial 2
CG32833	n.s.	No data
CG9897	n.s	n.s
CG32834	$F_{1,333} = 189.5$*	$F_{1,127} = 4.06$*
CG9897 and *CG32834*	n.s	n.s.
CG9897 and *CG32833*	n.s	No data
CG32834 and *CG32833*	n.s	No data
CG9897, *CG32834*, and *CG32833*	n.s.	No data

NOTE.—Data were analyzed using an ANOVA in SPSS after testing for normality with the Kolmogorv–Smirnov test.

[a]In Trial 1, control females or females knocked down for each female gene (individually and in combination) were mated to control males or males knocked down for *CG32833* (on Day 0). $N = 32–58$ females per treatment.

[b]In Trial 2, control females or females knocked down for each female gene (individually and in combination) were mated to wild-type males (on Day 0). $N = 15–40$ females per treatment.

[c]n.s.: not significant; no data: not tested.

*$P \leq 0.05$.

Discussion

Gene duplication is an important evolutionary mechanism for diversifying the suite of reproductive proteins expressed within a sex (e.g., Kresge et al. 2001; Clark et al. 2007; Kelleher et al. 2007; Nielsen et al. 2010; Mancini, Baldini, et al. 2011; Mancini, Tammaro, et al. 2011). Instances in which reproductive gene duplication results in one paralog switching its sex specificity of expression are less common. Analogous to the case of WISP/GLD2 discussed in the Introduction, we have discovered a case in which a reproductive protein with ancestral female expression underwent gene duplication, with one resulting paralog becoming male specific. However, two novel features distinguish the present gene family. First, this gene family experienced a second round of gene duplication, such that all *melanogaster* subgroup species have three functional copies. Second, although the paralog *CG32833* is expressed in males, its protein product is transferred to females during mating (Findlay et al. 2008). Thus, though its primary sequence and pattern of expression are different, its site of action may be conserved.

RNAi analyses of these proteins reveal that each female paralog is required for normal postmating responses in females. Females knocked down for one of the female genes (*CG32834*) consistently showed significantly reduced egg laying in the first 24 h after mating, suggesting that this protein is required for early, maximal fertility. Furthermore, knockdown of the two female genes together consistently caused females to be significantly less likely to remate at 10 days after the initial mating, when females are often receptive to remating. Knockdown of *CG9897* individually also consistently decreased remating at 4 days after the initial mating. Because few *D. melanogaster* female somatic reproductive proteins have been functionally characterized (Yapici et al. 2008), these experiments provide an important contribution to the understanding of the influence of female RT proteins on postmating responses. Our results also suggest that in spite of millions of years of evolution and changes in sex specificity of expression, three members of this family likely remain functional in each *melanogaster* group species.

Evolutionary History

Our evolutionary data suggest that this gene family appears to have had a dynamic evolutionary history. Because the single copy of this gene family in *D. pseudoobscura* is expressed exclusively in the female RT, we hypothesize that the ancestral single copy of this gene was female specific. After the divergence of *D. pseudoobscura* and the *melanogaster* species group, two duplication events occurred, giving rise to the three functional copies observed in *D. melanogaster*–*D. ananassae*. A fourth copy of this gene family is annotated as *GF11311* in *D. ananassae*. Our RT–PCR and sequencing data (supplementary fig. S1, Supplementary Material online) suggest that this is a pseudogene. However, we found the *GF11311* transcript to be expressed specifically in male RTs (data not shown), and the corresponding amino acid sequence showed greatest identity to *GF11312*, which is expressed in both sexes. Thus, it is likely that *D. ananassae* experienced a lineage-specific duplication of its copy of the paralog that would eventually become *CG32833* in *D. melanogaster*. However, one duplicate copy has since become a pseudogene, so only one paralog remains functional today. Why and how this paralog (*GF11312*) is expressed in both sexes remains an open question, as does the exact timing of when the *CG32833* orthologs in the *melanogaster* subgroup became specifically expressed in males.

The three proteins we have described show sequence similarity to two serine-type endopeptidases, SEND1 and SEND2 (Schnakenberg et al. 2011). Like CG9897 and CG32834, these SEND proteins are expressed specifically in the female spermathecae. Cloning of the *Send1* and *Send2* regulatory sequences showed that they have different patterns of expression: *Send1* is expressed in both virgin and mated females, whereas *Send2* is upregulated after mating (Schnakenberg et al. 2011). *Send1* is also located in a gene cluster of serine-type endopeptidases with spermatheca-specific expression. These results, combined with our finding that evolutionarily related serine endopeptidases have undergone changes in sex-specific expression, suggest that the regulation of this family of reproductive serine endopeptidases may be evolutionarily labile and/or may require relatively few evolutionary steps. Cloning and functional analysis of the regulatory

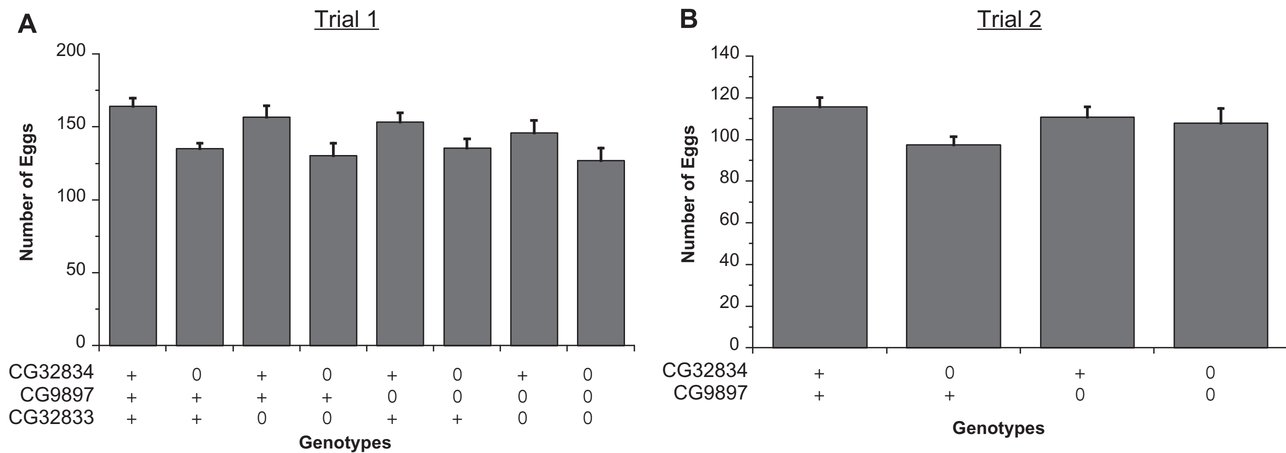

Fig. 6. Egg laying by *Drosophila melanogaster* females resulting from matings between females and males knocked down for various combinations of *CG32834*, *CG9897*, and *CG32833*. (*A*) Trial 1 used females knocked down for *CG32834*, *CG9897*, both genes, or neither gene, and males knocked down for *CG32833* or not. (*N* = 32–58 females per treatment; total eggs laid for 5 days after mating.) (*B*) Trial 2 used females knocked down for *CG32834*, *CG9897*, both genes, or neither gene, mated to wild-type Canton S males. (N = 15–40 females per treatment; total eggs laid for 4 days after mating). Gene presence is indicated on the *x*-axis of each graph with a " + ," whereas knockdown is indicated by a "0." Error bars indicate one standard error of the mean.

sequences of *CG9897*, *CG32833*, and *CG32834* may lead to insights into how the expression of these genes is regulated and which *cis* regulatory elements are required for their transcription.

Our finding that a Sfp, CG32833, arose from the duplication of a female-specific reproductive protein adds to a growing body of work on the evolutionary origin of Sfps. Although this mechanism—duplication of a female gene followed by male "co-option"—is intriguing, our screen for similar cases suggests it is also rare. In contrast, tandem duplication of existing Sfps is widespread in *D. melanogaster* and related species (Findlay et al. 2008) and has been reported in other species (e.g., *Anopheles* mosquitoes: Mancini, Baldini, et al. 2011; primates: Clark and Swanson 2005; rodents: Karn et al. 2008). This process can lead to lineage-specific gene gains and losses (Wagstaff and Begun 2005, 2007; Findlay et al. 2008). However, other Sfps appear to have arisen de novo from noncoding regions of the genome (Begun et al. 2006; Findlay et al. 2009). Sfps in the latter class tend to be short and are often lineage restricted, although tandem duplication of an existing protein affords the opportunity for the paralogous protein to begin with complex functional domains.

Tissue-Specific Gene Expression

Although tandem gene duplication is important for generating Sfp diversity in *Drosophila* (Findlay et al. 2008), there has been little investigation of how the patterns of gene expression change or remain conserved after such duplication. In the case of *CG32833*, a tandem duplicate expressed in the male accessory glands but located in the genome between two female-expressed genes (see fig. 1), a change in *cis* regulatory elements could have occurred that allowed it to become expressed in the opposite sex. If both the coding sequence and the *cis* regulatory sequence were duplicated, subsequent mutations in the regulatory region of the paralog could have altered its expression. Alternatively, only the coding sequence of *CG32833* could have duplicated, but this sequence could have been inserted downstream of an element that permitted a change in its expression (e.g., Rebeiz et al. 2011).

Another possible regulatory mechanism for this family of genes could be noncoding RNAs. The current annotation of the *D. melanogaster* genome shows two noncoding RNAs encoded in the same genomic location as the genes studied here. One is a predicted antisense RNA, *CR42742*; its sequence completely overlaps the *CG9897* gene but is transcribed in the opposite direction. ModENCODE data (*Drosophila* modENCODE Consortium 2010) show that this transcript is expressed specifically in the male RT, where it could potentially decrease *CG9897* expression, consistent with our observation that *CG9897* is expressed in a female-biased, but not female-specific, manner. The other is an annotated microRNA, mir-4939, which overlaps the 5′-end of *CG32833* and is transcribed in the same direction as that gene (Berezikov 2011); data about its expression pattern, which could hint at a potential regulatory role, are presently unavailable.

Function

Although a common outcome of gene duplication is pseudogenization of the resulting paralog, three copies of these genes have been retained in diverse *Drosophila* species, and their expression patterns are largely conserved in *D. melanogaster* through *D. erecta*. Further, the Sfp status of the CG32833 orthologs appears to be maintained within the *melanogaster* subgroup, because the *D. yakuba* ortholog is also a transferred seminal fluid protein (Findlay et al. 2008). Thus, there has presumably been a selective benefit to retaining the protein sequences and expression patterns of *CG32834* and

CG9897 (and their orthologs) in females and of *CG32833* (and its orthologs) in males.

Our RNAi studies of the female-expressed genes allow us to infer the reproductive benefits they may provide. These two genes promote reproductive success in two ways. First, CG32834 boosts female fecundity in the first 24 h after mating by ~20%. Second, the two female-expressed genes appear to work together to promote remating: at 10 days after an initial mating, females remated less frequently when both genes were knocked down than when either gene was at its normal level. Through this latter mechanism, these proteins could play roles in mediating intersexual conflict. After a single mating, females are typically unreceptive to remating for several days (Chapman et al. 2003; Liu and Kubli 2003). However, at 10 days postmating, female egg production and stored sperm are at relatively low levels (Ravi Ram and Wolfner 2007a; LaFlamme et al. 2012), and an additional mating would provide females with a fresh supply of sperm and Sfps, which would in turn increase fertility. Remating may also benefit females indirectly by increasing the genetic diversity of her offspring or the genetic compatibility with her mate. Female remating may be detrimental to the first male, however, as his remaining stored sperm would then face competition. Thus, the interests of a mating pair are likely to be in conflict over whether (and when) the female remates. Interestingly, we found that CG9897 also promotes remating at 4 days after mating but that CG32384 inhibits remating at 10 days after mating when adjusted independently of CG9897. Thus, females may control their receptivity to remating by adjusting the relative amounts of these proteins produced, whereas the male protein may act, albeit weakly, to inhibit female remating. Notably, wild-caught female *D. melanogaster* are typically found to contain sperm from several males (Imhof et al. 1998), suggesting that maintaining or regaining some level of remating receptivity after a prior mating is advantageous.

The RNAi experiments on CG32833 resemble previous findings that many Sfps have small or no effects on fertility or remating when knocked down (e.g., Herndon and Wolfner 1995; Heifetz et al. 2000; Ravi Ram and Wolfner 2007b; Mueller et al. 2008; LaFlamme et al. 2012; Findlay et al. 2014). If relatively few Sfps have dramatic effects, why does the male produce so many? One possibility is that having many versions of a particular type of protein is selectively advantageous. Another idea, suggested by Chapman (2008), is that the large number of Sfps could reflect an ongoing, coevolutionary process between males (in competition when each mates with the same female) and/or between the sexes (in conflict over female reproductive investment and postmating behavior). In this scenario, only one or a few members of each Sfp functional class would be most relevant at a given time, the others representing "evolutionary relics" rendered less important by counteradaptations or the emergence of more effective members of their class. Regarding the genes considered in our study, the female target(s) for the original gene may have evolved to become less sensitive to the male-expressed paralog while maintaining sensitivity to the female-expressed paralogs.

To further complicate matters, the genetic background of the male may impact the effectiveness of some reproductive genes. This is suggested by the variable effects on remating observed between our two trials (which used different males). Thus, the proteins in our study likely have interactions with other proteins that mediate their effect(s) on remating. It should also be noted that the proteins with little or no effect in the laboratory may retain effects on fitness that remain relevant (i.e., visible to selection) in wild populations, as typical laboratory assays used to measure Sfp function cannot measure all possible fitness-related phenotypes. Proteins like CG32833 may play roles not detected by our assays. Further, we achieved variable degrees of knockdown for each gene in this study (supplementary fig. S2, Supplementary Material online), and phenotypes are likely to be more pronounced and/or consistent in null mutants.

Co-Option as a Mechanism for the Evolution of "Influential" Proteins

Together, our results suggest that a Sfp evolved through a process of co-option in which an existing female RT protein was duplicated and subsequently changed its sex-specific pattern of expression. Co-option may be a mechanism for the evolution of other types of "influential proteins," that is, proteins from one individual that influence the phenotype of another. Other classes of influential proteins include those involved in various stages of sexual reproduction (e.g., courtship progression; Houck and Reagan 1990; Rollmann et al. 1999) and in maternal–fetal interactions (Stewart and Allen 1995). The effects of influential proteins can be either beneficial or detrimental to the affected individual (Wolfner 2009). Future research should investigate whether other classes of influential proteins can also evolve through the conspecific co-option mechanisms suggested by our results for the evolution of Sfps.

Materials and Methods

Searching for Female Reproductive Proteins with Sex-Switched Duplicates

To investigate whether female-expressed reproductive proteins might have paralogs present in male seminal fluid, we selected 20 predicted-secreted proteins expressed in sperm storage organs (Arbeitman et al. 2004; Chintapalli et al. 2007; Allen and Spradling 2008; Prokupek et al. 2009; table 1). We used BLASTP to compare each protein to all other annotated proteins in *D. melanogaster*. Up to five hits per protein that showed evidence of homology (alignment score >80; $e < 10^{-3}$, identity $>30\%$) were checked against published data (Chintapalli et al. 2007; Ravi Ram and Wolfner 2007a; Findlay et al. 2008) for whether they were transferred in male seminal fluid or predicted to be expressed in the male accessory glands. If any candidate met either criterion, we then examined whether the proteins showed evidence of paralogy, as judged to a first approximation by reciprocal BLAST comparisons.

Identification of Orthologs

Our search identified one pair of female-derived proteins (CG9897 and CG32834) that show sequence similarity to an Sfp, CG32833. The three genes encoding these proteins occur in a tightly linked cluster on chromosome 2R. Therefore, for the rest of our study, we focused on analyzing the evolutionary history and reproductive phenotypes of these three genes. We searched for orthologs of the *D. melanogaster* genes *CG9897*, *CG32833*, and *CG32834* in 11 other *Drosophila* species that have had their genomes sequenced (Drosophila 12 Genomes Consortium 2007). First, we used BLASTP to compare each gene's protein sequence from *D. melanogaster* against all predicted protein sequences from each of the other species. We detected reciprocal best BLAST proteins for one or more of the *D. melanogaster* proteins in each species from *Drosophila simulans* to *D. pseudoobscura* and *Drosophila persimilis*; clear orthologs were not identifiable in more distantly related species. We found just one copy from this gene family in *D. pseudoobscura/persimilis*, four total copies in *D. ananassae*, and one ortholog for each *D. melanogaster* gene in each of the species from *D. simulans* to *D. erecta*.

In several instances, we found that the predicted gene models for the orthologs were incomplete or incorrect. We combined experimental approaches (PCR and sequencing) with bioinformatic alignment methods (Findlay et al. 2009) to determine consistent gene structures across all species studied. For example, we found that in our laboratory wild-type strain of *D. melanogaster* (Canton S), the annotated intron in the CG9897 gene is, instead, coding DNA sequence. Furthermore, this gene region contains a 1-bp deletion relative to the FlyBase annotation (supplementary fig. S3, Supplementary Material online). The effect of this deletion is to maintain a single open reading frame throughout the length of the coding DNA sequence, suggesting that the transcript produces a functional protein. Interestingly, although we confirmed the lack of splicing of the annotated intronic region in the strain of *D. melanogaster* used for genome sequence (*y*; *cn bw sp*), we did not observe the 1-bp deletion, suggesting that this strain carries an allele of CG9897 that has a premature stop codon that truncates the predicted protein by ~40%. Other examples of gene annotation corrections are given in supplementary text S1 and figure S1, Supplementary Material online, and the final protein sequences used for each ortholog are given in supplementary file S2, Supplementary Material online. We have submitted to FlyBase the corrected gene annotations for all species examined.

Phylogenetic Analysis

Once protein sequences were determined from the corrected gene structures, we constructed a phylogenetic tree from the core serine endopeptidase domain of these sequences. This domain captured almost the entire length of the protein for the orthologs of CG9897 and CG32833, but excluded the repetitive C-terminal region of CG32834, which was difficult to align, as well as uncertain 5′-ends of a few other identified copies. Sequences were aligned using CLUSTAL Omega (Sievers et al. 2011) and checked by eye in the MEGA 5.05 program (Tamura et al. 2011). The alignment is included in a supplementary file S1, Supplementary Material online. We then used maximum likelihood to estimate the tree from the aligned protein sequences, using the *proml* program in PHYLIP v3.69 (Felsenstein 2011). We visualized the tree using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/, last accessed April 5, 2014). Bootstrap values, based on 100 replicates, were calculated using the PHYLIP programs *seqboot*, *proml*, and *consense*.

Gene Expression

We used RT-PCR to analyze the expression of each identified ortholog in its cognate species. We used the wild-type Canton S strain for *D. melanogaster* expression and those strains from other species that were used for genome sequencing (Drosophila 12 Genomes Consortium 2007). Using cDNA made from whole flies, we tested for sex-biased or sex-specific expression of each ortholog in its cognate species; we examined five species in the *melanogaster* subgroup (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*), a representative of the *melanogaster* group (*D. ananassae*), and an outgroup species, *D. pseudoobscura*. We tested for tissue-specific expression in *D. ananassae*, *D. pseudoobscura*, and *D. melanogaster* as a representative of the *melanogaster* subgroup. (In this subgroup, gene order and whole-fly expression patterns are conserved.) To test for sex-biased or sex-specific gene expression in whole adult flies, we extracted RNA from ten males or ten females using the TRIzol reagent (Invitrogen). To test for tissue-specific expression, we extracted RNA from the following tissues from each sex: the gonads (10/sample), the RT without the gonads (50/sample), and the remaining carcass without the RT (10/sample). To remove genomic DNA (gDNA) remaining after RNA extraction, we incubated 1 μg of the extracted RNA with 1–2 units RQ1 DNase (Promega, Madison, WI). We then used ~0.35 μg DNase-treated RNA to synthesize cDNA, using SmartScribe reverse transcriptase (Clontech, Mountain View, CA). The resulting cDNA was diluted 10-fold, and 1 μl was used in subsequent PCR reactions to test for gene expression.

PCR primers were designed with the Primer3 program v0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/, last accessed April 6, 2014) to produce ~350 bp amplicons from cDNA (supplementary table S1). As needed, individual primers were used to sequence PCR products to confirm sequences or to identify sequences from an incorrectly annotated species. As a positive control for RT-PCR, we amplified an intron-containing region of the *RpL32* gene. To ensure that cDNA preparations were free of gDNA contamination, *RpL32* primers were designed so that the product would span an intron, thus allowing gDNA amplicons to be detected as larger fragments. In no case was gDNA contamination observed. In general, PCR products and expression patterns could be readily discerned with 30 cycles of amplification.

Functional Characterization in *D. melanogaster*

We tested the function of each protein from the cluster in *D. melanogaster*, because this species is the most tractable for such analysis. We used the UAS-GAL4 system to knock down genes singly or in combination. For female-expressed genes *CG9897* and *CG32834*, we mated control or knockdown females to wild-type (Canton S) males; for male-expressed *CG32833*, we mated knockdown or control males to wild-type females. We also tested the effects of mating knockdown males to knockdown females. To achieve knockdown, we used UAS RNAi-lines from the Vienna Drosophila RNAi Center (Dietzl et al. 2007; VDRC Transformant IDs: *CG9897*: 104987, *CG32834*: 46434, and *CG32833*: 102866). Female genes were knocked down in the spermathecae with the spermathecal-specific *Send1*-GAL4 driver (*Send1*-GAL4, *CyO*/*Gla*; Schnakenberg et al. 2011). *CG32833* knockdown was accomplished using the GAL4-GAL80 system. CG32833 UAS-RNAi flies were crossed to *tubulin*-GAL80^{ts}; *tubulin*-GAL4/TM3,Sb, raised at room temperature, and shifted to the nonpermissible temperature (30 °C) 4 days before eclosion (Duffy 2002). For control flies, we crossed the background stock for the RNAi lines (*y,w[1118]; P{attP,y[+],w[3′]}*) to the appropriate driver. We mated the flies in all combinations (control or knockdown males mated to control females and to females knocked down for each gene individually and in combination). Knockdown was assessed by RT–PCR (Ravi Ram and Wolfner 2007b) and quantified using integrated optical density as measured by the ImageJ program (Schneider et al. 2012). In most cases, knockdown of >75–80% was achieved (supplementary fig. S2, Supplementary Material online).

We used standard assays (Ravi Ram and Wolfner 2007b; LaFlamme et al. 2012) to measure the probability of female remating and female fecundity (number of eggs laid) after a single mating in knockdown and control flies. We determined the probability of remating by mating females once and then testing to determine whether they would remate with a wild-type (Canton S strain) male within a 1-h time period at 1, 4, or 10 days after the initial mating. We conducted two replicates of the assay at 10 days after the initial mating. The replicates differed in that one replicate had females that were maintained individually and transferred to new vials every day, whereas in the other replicate individually maintained females were only transferred to new vials on day 5. We measured the number of eggs laid over a 4- to 5-day period after mating. Data from females that died before the end of the trial or that produced no live progeny were excluded from analyses.

Statistical Analysis

We used a generalized linear model with a logit link in JMP to test for the effect of male or female genotype (individually and in combination), replicate, and the interaction of replicate and each genotype on probability of remating. Variables were eliminated using backward iteration until only variables with *P* values ≤ 0.15 (or ones with higher-order level effects) remained. We used an analysis of variance (ANOVA) in SPSS to test for an effect of knocking down each gene (individually and in combination) and replicate on the number of eggs produced, after testing for normality with the Kolmogorov–Smirnov test.

Supplementary Material

Supplementary text S1, files S1 and S2, table S1, and figures S1–S3 are available at Molecular Biology and Evolution online (http://www.mbeoxfordjournals.org/).

Acknowledgments

We thank A. Wong for initially suggesting study of CG9897, and B. Bleakley, T. Connallon, and D. Fraga for useful discussions of earlier versions of this article. We are grateful to M. Siegal and S. Schnakenberg for generously providing us with the Send1-GAL4 driver line and to A. Clark and the University of California San Diego *Drosophila* Species Stock Center for providing the non-*melanogaster* *Drosophila* species. Sequencing was performed by Cornell University's Life Science Core Laboratories Center. This work was supported by the National Institutes of Health (grant numbers: R01HD038921 to M.F.W., F32GM097789 to G.D.F., and F32GM074361 to L.K.S.).

References

Allen AK, Spradling AC. 2008. The Sf1-related nuclear hormone receptor Hr39 regulates *Drosophila* female reproductive tract development and function. *Development* 135:311–321.

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

Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. 2011. Insect seminal fluid proteins: identification and function. *Annu Rev Entomol.* 56:21–40.

Baker RH, Narechania A, Johns PM, Wilkinson GS. 2012. Gene duplication, tissue-specific gene expression and sexual conflict in stalk-eyed flies (Diopsidae). *Phil Trans R Soc Lond B Biol Sci.* 367: 2357–2375.

Begun DJ, Lindfors HA, Thompson ME, Holloway AK. 2006. Recently evolved genes identified from *Drosophila yakuba* and *Drosophila erecta* accessory gland expressed sequence tags. *Genetics* 172: 1675–1681.

Benoit P, Papin C, Kwak JE, Wickens M, Simonelig M. 2008. PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. *Development* 135:1969–1979.

Berezikov E. 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet.* 12:846–860.

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302–305.

Chapman T. 2008. The soup in my fly: evolution, form and function of seminal fluid proteins. *PLoS Biol.* 6:1379–1382.

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

Chintapalli V, Wang J, Dow J. 2007. Using Flyatlas to identify better *Drosophila* models of human disease. *Comp Biochem Physiol A Mol Integr Physiol.* 150:S136–S137.

Cirera S, Aguade M. 1998. The sex-peptide gene (Acp70A) is duplicated in *Drosophila subobscura*. *Gene* 210:247–254.

Clark NL, Findlay GD, Yi XH, MacCoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. *Mol Biol Evol.* 24:2081–2090.

Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet.* 1:335–342.

Clauss A, Lilja H, Lundwall A. 2005. The evolution of a genetic locus encoding small serine proteinase inhibitors. *Biochem Biophys Res Commun.* 333:383–389.

Connallon T, Clark AG. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* 187:919–937.

Cui J, Sackton KL, Horner VL, Kumar KE, Wolfner MF. 2008. Wispy, the *Drosophila* homolog of GLD-2, is required during oogenesis and egg activation. *Genetics* 178:2017–2029.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448:151–156.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.

Drosophila modENCODE Consortium. 2010. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330:1787–1797.

Duffy JB. 2002. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* 34:1–15.

Felsenstein J. 2011. PHYLIP (Phylogeny Inference Package) version 3.69. Seattle: University of Washington.

Findlay GD, MacCoss MJ, Swanson WJ. 2009. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res.* 19:886–896.

Findlay GD, Sitnik JL, Wang W, Aquadro CA, Clark NL, Wolfner MF. 2014. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. *PLoS Genet.* 10:e1004108.

Findlay GD, Yi XH, MacCoss MJ, Swanson WJ. 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol.* 6:1417–1426.

Gallach M, Betran E. 2011. Intralocus sexual conflict resolved through gene duplication. *Trends Ecol Evol.* 26:222–228.

Gao GJ, Cheng Y, Wesolowska N, Rong YKS. 2011. Paternal imprint essential for the inheritance of telomere identity in *Drosophila*. *Proc Natl Acad Sci U S A.* 108:4932–4937.

Graham LA, Davies PL. 2002. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene* 292:43–55.

Harano T, Okada K, Nakayama S, Miyatake T, Hosken DJ. 2010. Intralocus sexual conflict unresolved by sex-limited trait expression. *Curr Biol.* 20:2036–2039.

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.

Hekmat-Scafe DS, Scafe CR, McKinney AJ, Tanouye MA. 2002. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* 12:1357–1369.

Herndon LA, Wolfner MF. 1995. A Drosophila seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc Natl Acad Sci U S A.* 92:10114–10118.

Holland PWH, García-Fernández J. 1996. Hox genes and chordate evolution. *Dev Biol.* 173:382–395.

Horne I, Haritos VS, Oakeshott JG. 2009. Comparative and functional genomics of lipases in holometabolous insects. *Insect Biochem Mol Biol.* 39:547–567.

Hosken DJ. 2011. Gene duplication might not resolve intralocus sexual conflict. *Trends Ecol Evol.* 26:558–559.

Houck LD, Reagan NL. 1990. Male courtship pheromones increase female receptivity in a plethodontid salamander. *Anim Behav.* 39: 729–734.

Hurle B, Swanson W, Green ED, Sequencing NC. 2007. Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Res.* 17:276–286.

Imhof M, Harr B, Brem G, Schlotterer C. 1998. Multiple mating in wild *Drosophila melanogaster* revisited by microsatellite analysis. *Mol Ecol.* 7:915–917.

Karn RC, Clark NL, Nguyen ED, Swanson WJ. 2008. Adaptive evolution in rodent seminal vesicle secretion proteins. *Mol Biol Evol.* 25: 2301–2310.

Kelleher ES, Markow TA. 2009. Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* 181:1451–1465.

Kelleher ES, Pennington JE. 2009. Protease gene duplication and proteolytic activity in *Drosophila* female reproductive tracts. *Mol Biol Evol.* 26:2125–2134.

Kelleher ES, Swanson WJ, Markow TA. 2007. Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. *PLoS Genet.* 3:1541–1549.

Kelleher ES, Watts TD, LaFlamme BA, Haynes PD, Markow TA. 2009. Proteomic analysis of Drosophila mojavensis male accessory glands suggests novel classes of seminal fluid proteins. *Insect Biochem Mol Biol.* 39:366–371.

Kresge N, Vacquier VD, Stout CD. 2001. Abalone lysin: the dissolving and evolving sperm protein. *Bioessays* 23:95–103.

LaFlamme BA, Ravi Ram K, Wolfner MF. 2012. The *Drosophila melanogaster* seminal fluid protease "Seminase" regulates proteolytic and post-mating reproductive processes. *PLoS Genet.* 8: e1002435.

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40:D302–D305.

Lewis CA, Talbot CF, Vacquier VD. 1982. A protein from abalone sperm dissolves the vitelline layer by a non-enzymatic mechanism. *Dev Biol.* 92:227–239.

Lin HJ, Lee CM, Luo CW, Chen YH. 2005. Functional preservation of duplicated pair for RSVS III gene in the REST locus of rat 3q42. *Biochem Biophys Res Commun.* 326: 355–363.

Liu HF, Kubli E. 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 100: 9929–9933.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.

Makino T, McLysaght A. 2010. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A.* 107:9270–9274.

Mancini E, Baldini F, Tammaro F, Calzetta M, Serrao A, George P, Morlais I, Masiga D, Sharakhov IV, Rogers DW, et al. 2011. Molecular characterization and evolution of a gene family encoding male-specific reproductive proteins in the African malaria vector *Anopheles gambiae*. *BMC Evol Biol.* 11:292.

Mancini E, Tammaro F, Baldini F, Via A, Raimondo D, George P, Audisio P, Sharakhov IV, Tramontano A, Catteruccia F, et al. 2011. Molecular evolution of a gene cluster of serine proteases expressed in the *Anopheles gambiae* female reproductive tract. *BMC Evol Biol.* 11:72.

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ, FlyBase C. 2013. FlyBase: improvements to the bibliography. *Nucleic Acids Res.* 41:D751–D757.

Mueller JL, Linklater JR, Ram KR, Chapman T, Wolfner MR. 2008. Targeted gene deletion and phenotypic analysis of the *Drosophila melanogaster* seminal fluid protease inhibitor Acp62F. *Genetics* 178: 1605–1614.

Nielsen MG, Caserta JM, Kidd SJ, Phillips CM. 2006. Functional constraint underlies 60 million year stasis of Dipteran testis-specific beta-tubulin. *Evol Dev.* 8:23–29.

Nielsen MG, Gadagkar SR, Gutzwiller L. 2010. Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. *BMC Evol Biol.* 10: 113.

Ohno S. 1970. Evolution by gene duplication. New York: Springer–Verlag.

Ohta T. 2003. Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. *Genetica* 118:209–216.

Parsch J, Meiklejohn CD, Hauschteck-Jungen E, Hunziker P, Hartl DL. 2001. Molecular evolution of the ocnus and janus genes in the *Drosophila melanogaster* species subgroup. *Mol Biol Evol.* 18: 801–811.

Prokupek AM, Kachman SD, Ladunga I, Harshman LG. 2009. Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*. *Insect Mol Biol.* 18:465–475.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.

Raff EC, Hutchens JA, Hoyle HD, Nielsen MG, Turner FR. 2000. Conserved axoneme symmetry altered by a component beta-tubulin. *Curr Biol.* 10:1391–1394.

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

Ravi Ram K, Wolfner MF. 2007b. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. *PLoS Genet.* 3:2428–2438.

Rebeiz M, Jikomes N, Kassner VA, Carroll SB. 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A.* 108:10036–10043.

Rodin SN, Riggs AD. 2003. Epigenetic silencing may aid evolution by gene duplication. *J Mol Evol.* 56:718–729.

Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, Morris HR, Catteruccia F. 2009. Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLoS Biol.* 7: e1000272.

Rollmann SM, Houck LD, Feldhoff RC. 1999. Proteinaceous pheromone affecting female receptivity in a terrestrial salamander. *Science* 285: 1907–1909.

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

Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zool Part B.* 308B: 58–73.

Sartain CV, Cui J, Meisel RP, Wolfner MF. 2011. The poly(A) polymerase GLD2 is required for spermatogenesis in *Drosophila melanogaster*. *Development* 138:1619–1629.

Schnakenberg SL, Matias WR, Siegal ML. 2011. Sperm-storage defects and live birth in *Drosophila* females lacking spermathecal secretory cells. *PLoS Biol.* 9:e1001192.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Soding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7:539.

Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M, Deewatthanawong P, Wolfner MF, Harrington LC. 2011. Towards a semen proteome of the Dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis.* 5: e989.

Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, Chow CY, Wolfner MF. 2009. Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. *Adv Genet.* 68: 23–56.

Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC. 2008. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. *Insect Biochem Mol Biol.* 38:176–189.

Smith RC, Walter MF, Hice RH, O'Brochta DA, Atkinson PW. 2007. Testis-specific expression of the beta 2 tubulin promoter of *Aedes aegypti* and its application as a genetic sex-separation marker. *Insect Mol Biol.* 16:61–71.

St Pierre SE, Ponting L, Stefancsik R, McQuilton P, the FlyBase Consortium. 2013. FlyBase 102-advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* 42:D780–D788.

Stewart AD, Pischedda A, Rice WR. 2010. Resolving intralocus sexual conflict: genetic mechanisms and time frame. *J Hered.* 101: S94–S99.

Stewart F, Allen WR. 1995. Comparative aspects of the evolution and function of the chorionic gonadotrophins. *Reprod Domest Anim.* 30: 231–239.

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 U S A.* 98: 7375–7379.

Swanson WJ, Vacquier VD. 1995. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. *Proc Natl Acad Sci U S A.* 92: 4957–4961.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Tian X, Pascal G, Monget P. 2009. Evolution and functional divergence of NLRP genes in mammalian reproductive systems. *BMC Evol Biol.* 9:202.

Torgerson DG, Singh RS. 2004. Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha 4 proteasome subunits in *Drosophila*. *Genetics* 168: 1421–1432.

Vacquier VD, Swanson WJ, Lee YH. 1997. Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? *J Mol Evol.* 44: S15–S22.

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

Wagstaff BJ, Begun DJ. 2007. Adaptive evolution of recently duplicated accessory gland protein genes in desert *Drosophila*. *Genetics* 177: 1023–1030.

Wolfner MF. 2009. Battle and ballet: molecular interactions between the sexes in *Drosophila*. *J Hered.* 100:399–410.

Yapici N, Kim YJ, Ribeiro C, Dickson BJ. 2008. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* 451:33–37.

Zhang JZ. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.