

Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of *Drosophila melanogaster*

A. M. PROKUPEK*, S.-I. EYUN*, L. KO*, E. N. MORIYAMA*† & L. G. HARSHMAN*

*School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA

†Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE USA

Keywords:

conflicts;
genomics;
insects;
molecular evolution;
population genetics;
sexual selection.

Abstract

Sperm storage organs are common and broadly distributed among animal taxa. However, little is known about how these organs function at the molecular level. Additionally, there is a paucity of knowledge about the evolution of genes expressed in these organs. This investigation is an evolutionary expressed sequence tag (EST) study of genes expressed in the seminal receptacle, one of the sperm storage organs in *Drosophila*. The incidence of positive selection is higher for the seminal receptacle genes than *Drosophila* reproductive genes as a whole, but lower than genes associated with the spermatheca, a second type of *Drosophila* sperm storage organ. By identifying overrepresented classes of proteins and classes for which sperm storage function is suggested by the nature of the proteins, candidate genes were discovered. These candidates belong to protein classes such as muscle contraction, odorant binding and odorant receptor, protease inhibitor and immunity.

Introduction

An emerging trend in evolutionary biology is that many genes related to reproduction evolve rapidly. This phenomenon is observed in plants, invertebrates and vertebrates (Clark *et al.*, 2006). Rapidly evolving reproductive genes in *Drosophila* have been the subject of many studies. One such study looked at genes encoding male and female *Drosophila* reproductive proteins and found higher rates of evolution in these proteins compared to genes encoding nonreproductive proteins (Haerty *et al.*, 2007). Especially, intensely studied are male proteins passed to females at the time of mating (Findlay *et al.*, 2008). One class of transferred proteins, male accessory gland proteins (Acps), have been identified as rapidly evolving in a number of studies (Civetta & Singh, 1995; Begun *et al.*, 2000; Swanson *et al.*, 2001; Mueller *et al.*, 2005; Wagstaff & Begun, 2005). Additionally, studies of genes expressed in the lower reproductive tract of females show that genes in this region evolve relatively rapidly (Swanson *et al.*, 2004; Kelleher *et al.*, 2007). However, there has been little investigation into the

evolution of proteins expressed in specific female reproductive organs, though one previous study investigating molecular evolution in a specific sperm storage organ, the spermathecae, observed a high incidence of positive selection (Prokupek *et al.*, 2008).

The process of reproduction in taxa with internal fertilization, such as *Drosophila* species, presents opportunities for strong selection to operate on female reproductive proteins. One possible result of this selection is the rapid evolution of female reproductive proteins. Post-mating sexual selection could include processes such as the female mediation of sperm competition and female cryptic choice of sperm. A series of steps are taken before fertilization, including the processes involved in sperm storage (Adams & Wolfner, 2007), during which male and female reproductive proteins could interact. This interaction could be an underlying cause of sexually antagonistic coevolution (Rice, 1996; Holland & Rice, 1999; Arnqvist *et al.*, 2000; Miller & Pitnick, 2002; Rice & Holland, 2005; Rice *et al.*, 2006). Another consideration is that females must protect themselves from bacteria and fungi transferred with the seminal fluid (Ferrandon *et al.*, 1998; Wira & Fahey, 2004). Host–pathogen coevolution may play a role in the rapid evolution of female reproductive proteins. In both cases, proteins within the female reproductive tract interact with molecules introduced

Correspondence: Lawrence G. Harshman, School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588, USA.
Tel.: 1 402 472 0680; fax: 1 402 472 2083;
e-mail: lharsh@unlserve.unl.edu

during the mating process creating opportunities for positive selection driven by coevolutionary dynamics.

Sperm storage by females is widespread in animals including insects (Baer *et al.*, 2006; Collins *et al.*, 2006; Adams & Wolfner, 2007), reptiles (Gist & Congdon, 1998; Yamanouye *et al.*, 2004), birds (Das *et al.*, 2006; Zhang, 2006) and mammals (Racey, 1972, 1979). Sperm storage organs are likely to be important sites of interaction between male and female reproductive molecules. These interactions can be synergistic and ensure that sperm are properly directed, whether towards the fertilization site or into separate sections of the reproductive tract for storage (Bloch-Qazi & Wolfner, 2003; Gao *et al.*, 2003). Conversely, male and female proteins may also interact antagonistically with differential fitness consequences for each sex. Female sperm storage organs increase the length of time sperm spend within the female reproductive tract, thus maximizing the time of contact between male and female proteins. Studying these organs provides a unique opportunity to identify proteins important in male–female reproductive interactions.

Drosophila melanogaster females have two sperm storage organs: the seminal receptacle and paired spermathecae. The seminal receptacle stores the majority (65–80%) of the sperm, and the spermathecae is thought to be the long-term sperm storage organ (reviewed in Neubaum & Wolfner, 1999). In *Drosophila*, sperm enter the seminal receptacle first, approximately 10–15 min after mating, and do not begin entering the spermathecae until approximately 35 min after mating (Adams & Wolfner, 2007). The entrance of sperm into storage is correlated with distinct morphological changes and muscular contractions in the female reproductive tract (Adams & Wolfner, 2007). Within 3 h of mating, the female is laying fertilized eggs, the sperm used for fertilization at this point is suspected to be from the seminal receptacle. At 3-h post-mating, the seminal receptacle is performing dual functions: the storage of sperm and the release of sperm for fertilization.

Transcriptome and molecular evolution studies can be informative in terms of providing insight into how organs function at the molecular level. A study by Prokupek *et al.*, 2009 used transcriptome analysis to investigate expression differences in the two sperm storage organs of *Drosophila* females at two time points post-mating (Prokupek *et al.*, 2009). This study found substantial differences between the transcription profiles of these two organs. The spermathecae was enriched for genes involved in proteolysis and metabolism whereas the seminal receptacle was enriched for genes involved in localization, signalling and ion transport. At 3-h post-mating, the seminal receptacle had over ten times the number of differentially expressed genes as the ST. The large number of differentially expressed genes in the seminal receptacle may be related to the dual role this organ is playing (storing and releasing sperm) at three-hour post-mating.

The methods for inferring patterns of evolution from DNA sequence divergence are well-established. Specific DNA sequences are typically compared between species often in the context of an established phylogeny (Yang, 2007; Ellegren, 2008). The analysis is based on comparing the rate of amino acid coding changes (nonsynonymous) to the rate of silent changes (synonymous). The ratio for comparison is the ratio of nonsynonymous changes per nonsynonymous site (d_N) to synonymous changes per synonymous site (d_S). The null hypothesis is neutral evolution in contrast to inference about several modes of natural selection. Positive Darwinian selection is indicated by elevated d_N/d_S ratios (Yang & Bielawski, 2000; Bielawski & Yang, 2003). This inference is made by comparison with the actual sequence divergence patterns with the expectations under the null hypothesis of neutral evolution using maximum-likelihood statistics. This strategy allows for discernment of adaptive evolution of specific genes (Swanson, 2003). One of the generalities to emerge from these analyses is that reproductive proteins tend to evolve more rapidly in terms of the rate of amino acid substitutions (Swanson & Vacquier, 2002). To understand the basis for the observed pattern of rapid evolution, it is important to investigate genes that encode male and female reproductive proteins that could interact.

The present study examines genes encoding female reproductive proteins that are candidates to interact with male reproductive proteins. Sexual selection at the level of interacting male and female reproductive proteins is a likely cause of positive selection acting on sperm storage organ proteins. If sexual selection is operating, then one would expect high levels of amino acid replacement and the subsequent inference of positive selection. Molecular evolution studies provide tools to study sexual selection at the level of male and female reproductive proteins.

The *Drosophila* seminal receptacle is the focus of the current study. This organ is understudied in terms of the identification and function of genes expressed within, or the characterization of the evolutionary rates of these genes. In the present study, we investigated expressed sequence tags (ESTs) of genes enriched for expression in *D. melanogaster* seminal receptacle. We identified genes expressed in the seminal receptacle that are rapidly evolving and may be prime candidates to play an important role in female-ejaculate interactions. Functional annotation revealed a number of genes that provide clues into how the seminal receptacle operates.

Materials and methods

Stock maintenance and experimental design

The Canton-S stock of *D. melanogaster* was used in the present study. Flies were reared in larval density controlled vials, on a standard *Drosophila* diet in a temperature controlled environment (25 degrees Celsius) with a

standard 12/12 light/dark cycle. Adults were collected as virgins, and virginity was confirmed by observing that no larvae appeared in the vials used to hold the females. The flies were held as virgins until the fourth day of adult life when each female was paired with a single male for mating. Females were separated from males immediately after the observed mating had ended naturally. Prior to dissection, females were immobilized by placing vials containing females on ice for 5 min. Seminal receptacles were collected from virgin and 3-h post-mating females by dissection on a cold plate in RNase Later (Ambion).

cDNA library preparation and sequence generation

RNA was isolated from seminal receptacles dissected from 250 *D. melanogaster* females. Total RNA was isolated from seminal receptacles using the TRIzol reagent (Invitrogen). Total RNA was also purified from female whole bodies minus seminal receptacle to be used as the driver in subtractive hybridization. cDNA was generated from total RNA using the SMART approach (Zhu *et al.*, 2001). A cDNA library was generated (Evrogen Inc., Moscow) using the suppressive subtraction hybridization method in both directions (tester vs. driver and driver vs. tester) (Diatchenko *et al.*, 1996, 1999). An aliquot of the library was plated and resulting colonies were used for DNA template generation by rolling circle amplification using TempliPhi (Amersham Biosciences). Using a MegaBACE 400 automated DNA sequencer (Amersham Biosciences), 2304 DNA sequences were generated. For DNA sequence analysis, vector sequences were masked using the CAP3 program (Huang & Madan, 1999).

Identification of genes expressed in seminal receptacles

Each of the 2304 *D. melanogaster* ESTs was used as a query in a blastn DNA similarity search (Altschul *et al.*, 1990) conducted against the entire coding sequence (CDS) set of the *D. melanogaster* genome (Flybase Release 5.9; <http://flybase.org>). We excluded those sequences with identities lower than 95% or lower (with the exception of four ESTs with at least 92%) and those with expected (E) values greater than 10^{-6} (1168 sequences were excluded).

Ortholog identification and reconstruction of multiple alignments

Using each of the *D. melanogaster* CDSs as a query, a blastp protein similarity search (Altschul *et al.*, 1990) was performed to identify ortholog candidates from five additional *Drosophila* genomes (*D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta* and *D. ananassae*). The top hit from each species was then used as a query in a reciprocal blastp search against the entire *D. melanogaster* protein set to confirm the orthologous relationships. When multiple

sequences were identified with almost identical lowest E values, all were used as the queries for the reciprocal search.

To determine the presence or absence of possible distant orthologs in other species, reciprocal blast was performed against an additional six (more distantly related) *Drosophila* genomes (*D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis* and *D. grimshawi*). In addition to blastp, tblastn against the DNA scaffolds was also performed to identify ortholog candidates even if they are not included in the annotated CDS set. If two or more genes were identified as the top hits with almost identical E values, then these genes were analysed as possible duplicates by further investigation using the phylogenetic analysis to identify species-specific gene duplication events. All of the *Drosophila* complete genome sequences were downloaded from the FlyBase (Annotation Release 1.2; <http://flybase.org>).

Protein alignments were reconstructed using MAFFT with L-INS-i method (Katoh & Toh, 2008). The PAL2NAL script was used for converting sequence alignments of proteins into corresponding codon-based nucleotide alignments (Suyama *et al.*, 2006). Alignments were inspected at both DNA and protein-levels and adjusted manually when necessary.

Molecular evolution analyses

The relative contribution of nonsynonymous (d_N) to synonymous (d_S) nucleotide changes was compared using the codon-based maximum-likelihood framework described by Goldman and Yang (1994) implemented in PAML version 4 (Yang, 2007). For all genes having an identifiable *D. simulans* ortholog, pairwise comparisons were performed between *D. simulans* and *D. melanogaster*. A d_N/d_S ratio of 0.5 was found to be reasonable for the identification of candidate genes to undergo further investigation into the forces of selection (Swanson *et al.*, 2004).

The d_N/d_S ratios were also estimated by using the 'branch model' as well as the 'site model' of PAML based on six-species phylogenies. When six orthologs were not identified, a minimum of four orthologs were used to circumvent problems caused by model convergence. To ensure that possible incomplete lineage sorting reported in the melanogaster subgroup (Pollard *et al.*, 2006; Wong *et al.*, 2008) did not affect the outcome of our analysis, the analysis was performed using three alternative six-species trees varying the phylogenetic placement of *D. erecta* and *D. yakuba*. The 'branch model' allows the d_N/d_S ratios to vary among branches in a given phylogeny to detect positive selection acting on particular lineages (Yang, 1998). Variation in the d_N/d_S ratios among sites was investigated using the 'site model' described by Yang *et al.* (2000), which compares the fit of the data to different models of codon evolution (Yang & Nielsen, 2000). We examined the fit of data to

the one-ratio model (M0) against the model that classifies sites into 3 classes (M3). Two additional model comparisons were used to perform more direct tests of positive selection. Briefly, two likelihood ratio tests were used to compare null models that do not allow $d_N/d_S > 1$, M1a (nearly neutral) and M7 (beta), with alternative models that allow a class of sites to have $d_N/d_S > 1$, M2a (positive selection) and M8 (beta and omega) (Yang & Nielsen, 2002; Yang & Swanson, 2002). For a full description of the assumptions of the models and test statistics, see Yang *et al.* (2000).

Transmembrane and signal peptide prediction

Protein sequences from the *D. melanogaster* genes were used for motif prediction. Transmembrane region prediction was conducted using two programs: HMMTOP version 2.1 (Tusndy & Simon, 2001) and Phobius version 1.01 (Kll *et al.*, 2004). Both methods use hidden Markov models (HMMs) for predicting the transmembrane topology. Phobius combines TM prediction and signal peptide prediction to identify signal peptides from N-terminal regions, often misidentified as a TM region by these prediction methods. We list a protein as having transmembrane regions if both HMMTOP and Phobius predicted one or more TM regions, or if one program predicted multiple TM regions. For signal peptide prediction, we used TargetP version 1.1 (Emanuelsson *et al.*, 2007) in addition to Phobius. The TargetP program ranks support for the signal peptides. Only the genes in the highest class of support in TargetP, which were also identified as having signal peptides by Phobius, were listed as having signal peptides.

Functional categories

Protein function was inferred from a combination of information gained from the conserved domain searches, FlyBase, Gene Ontology database classification and literature searches. All genes were subject first to conserved domain searches by CD-Search at National Center for Biotechnology Information (Marchler-Bauer & Bryant, 2004).

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (<http://david.abcc.ncifcrf.gov>) was used to classify genes using the most relevant gene ontology (GO)-associated terms and to determine significant enrichment of known functional annotations within our gene list. DAVID calculates the probability of observed representation of genes within a given category compared to a predefined genome list corresponding to the same category.

The Protein ANALysis THrough Evolutionary Relationships (PANTHER) (<http://www.pantherdb.org>) classification system was also used to classify genes into broad functional categories. Genes were classified into families and subfamilies using profile HMMs, then further clas-

sified by molecular function and biological process ontology terms. The ontology of PANTHER uses a controlled vocabulary of molecular function and biological process arranged as directed acyclic graphs, similar to the GO, but abbreviated and simplified. Currently, the PANTHER database contains annotation data for 14,115 *D. melanogaster* genes.

Results

Coding sequences in the cDNA library

Of the 2,304 EST library clone sequences, 1,136 matched CDSs in the *D. melanogaster* genome representing 292 unique CDSs. The accepted sequences had an average E value of 4.5×10^{-6} .

Orthologs and functional categories

Sequence similarity and reciprocal blast were used to identify putative orthologs in six *Drosophila* genomes: *D. simulans*, *D. melanogaster*, *D. sechellia*, *D. yakuba*, *D. erecta* and *D. ananassae*. Of the 292 genes identified from the library, 248 had identifiable orthologous genes in all six species (268 in at least four).

Gene function was investigated for all genes (292 genes in total; Table 1, Supplementary Table S1). Genes expressed in the seminal receptacle represented a diverse range of functional categories. The largest category consisted of genes of unknown function (41 genes). Other gene categories included signalling (19 genes), lipid/carbohydrate metabolism (24 genes), ion transport (15 genes), proteases (17 genes) and muscle contraction (14 genes). Additional categories include protease inhibitors, structural, transcription, detoxification/antioxidants and electron transport.

Overrepresented gene categories

Genes were clustered into functional categories (using molecular function, biological process and cellular component), and statistical measurements of overrepresentation were calculated. Statistically significant gene categories are those with a P value of less than 0.001. These categories include cytoskeletal proteins, signalling, transport, protease inhibitor, muscle contraction, organ development, ion transport and lipid particle. Two pathways (based on the KEGG metabolic pathway database) were also identified as being overrepresented. These pathways are cell communication and amino-sugar metabolism.

Evolutionary analyses

A total of 261 genes had identifiable orthologs in the *D. simulans* genome. Nonsynonymous and synonymous substitution rates were determined for each of the 261

Table 1 Seminal receptacle genes grouped by functional annotation.

Function*	Number†	Sp‡	TM§	d_N/d_S ¶	PAML**
Unknown	41	19	3	4	12
Lipid/carb metabolism	24	8	3		2
Other transport	24	5	10	1	8
Signalling	19	5	6		2
Protease	17	9	2		5
Ion transport	15	2	8		4
Defence/immunity	14	8	1		2
Muscle contraction	14		2		3
Binding	12	5			3
Structural	12	4			2
Receptor	11	5	5	2	3
Detox/antioxidant	10	2	1		2
Ribosomal protein	10				3
Protease inhibitor	9	7			2
Transcription	9	1			3
Mitochondrial	8			8	
Hydrolase	6	1			1
Ligase	5	1			
Chaperone	4				
Kinase	4				1
Membrane	4	2	1		1
Isomerase	3	1			1
Apoptosis	3				
Cell adhesion	3	1			
Electron transport	2			1	1
Protein interaction	2				
Translation	2				
Reductase	2			1	
Nerve transmission	2				1
mRNA catabolism	1				

The number of genes from each category predicted to have secretion signal peptides and/or transmembrane domains is given. Evolutionary information including the number of genes having d_N/d_S values > 0.5 and predicted to be positively selected by site-model PAML analysis is listed.

*Predicted function of encoded proteins.

†Number of genes.

‡Number of genes predicted to encode proteins that have secretion signal peptides.

§Number of genes predicted to encode proteins that with transmembrane domains.

¶Number of genes for each functional category with d_N/d_S ratios > 0.5.

**Number of genes for each functional category predicted by PAML to be under positive selection.

D. melanogaster genes by comparing it with its ortholog in *D. simulans*. The average d_N/d_S ratio was 0.1357 ± 0.0195 , with an average d_N of 0.0304 and an average d_S of 0.2530. In this comparison, 9 of the 261 genes (Fig. 1) have d_N/d_S higher than the 0.5 threshold adopted by Swanson *et al.* (2004).

The 'site-model' analysis in PAML was used to explore heterogeneity in d_N/d_S along the gene and to test for positive Darwinian selection (Table 1, Supplementary Table S1). These comparisons were restricted to the 268

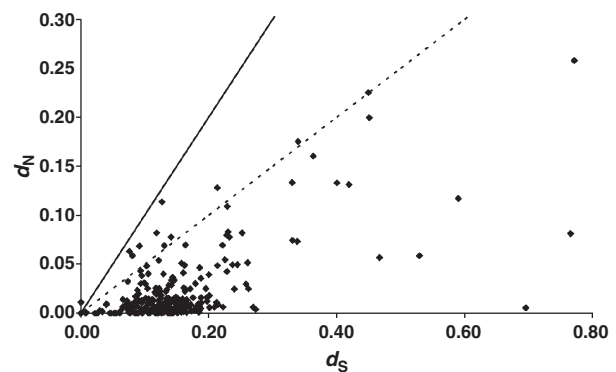


Fig. 1 The number of nonsynonymous substitutions per site (d_N) plotted against the number of synonymous substitutions per site (d_S). The solid line represents d_N/d_S of 1.0, the dashed line represents d_N/d_S of 0.5. Each square on the graph represents a gene from the seminal receptacle library.

genes for which sequences were available from at least four species. The first comparison examined the fit of data to M0 against M3. For 236 of the 268 genes, the fit of data to M3 was significantly better than M0, indicating heterogeneity of evolutionary rates along the gene for a high percentage of the genes analysed. Comparisons of M7 to M8 and M1 to M2 provided evidence for positive selection in 58 of 268 genes. Genes having elevated d_N/d_S from the pairwise comparison and/or support for positive selection from PAML models are listed in Supplementary Table 1.

The branch-model analysis using PAML shows little variation among branch lengths. For example, there was little difference in the overall d_N/d_S averages based on the three phylogenetic trees varying the placement of *D. yakuba* and *D. erecta*. As an exception to this consistency, the branch leading to *D. ananassae* has, on average, the lowest d_N/d_S values because of increased d_S (>1) for most genes. *D. ananassae* is the most distantly related species in the analysis. All data from PAML branch-model analysis is reported in Supplementary Table S2.

To obtain extended taxonomic insight into the evolution of spermathecal genes, the presence or absence of homologous genes was examined comparing *D. melanogaster* to 11 sequenced genomes (*D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis* and *D. grimshawi*) (Supplementary Table S1). Possible species-specific duplications were found for a total of eleven genes across seven species (Supplementary Table S3).

Secretion signal sequence, transmembrane region prediction and evolutionary rates

Of the 292 genes examined, 86 (29%) are predicted to have secretion signal peptides and 53 (18%) predicted to have transmembrane regions (Table 1). Twenty of the genes predicted to have secretion signal peptides and

seven of the genes predicted to have transmembrane regions show evidence of positive selection (Supplementary Table S1).

Discussion

Evolution

Comparative molecular evolution

Genes were identified in the seminal receptacle, which exhibit rapid evolution, through the proportion of these genes is much less than that found in a similar study of the spermathecae (Prokupek *et al.*, 2008). In a comparison of *D. melanogaster* and *D. simulans* sequences in the present study, there is an average d_N of 0.0304 and an average d_S of 0.2530. In this comparison, 9 of the 261 genes (Fig. 1) have d_N/d_S higher than the 0.5. In an analogous pairwise comparison, genes in the spermatheca had an average d_N/d_S ratio of 0.269 ± 0.2932 , with an average d_N of 0.032 and an average d_S of 0.119. The level of d_S in the seminal receptacle is about twice that of the spermatheca, but this value is well within the 10-fold range of d_S (0.2–2.0) observed in genes that have diverged between *D. melanogaster* and *D. simulans* (Powell & Moriyama, 1997). Of 42 genes, 10 had a d_N/d_S higher than the 0.5 (Prokupek *et al.*, 2008), a significantly higher proportion ($Z = 4.7$, $p < 0.01$) than seen in the seminal receptacle. The distribution of d_N/d_S for the seminal receptacle genes compared to the spermathecae genes (Fig. 2) shows a marked difference. Specifically, the spermathecae distribution appears to be bimodal, with a small second mode comprised of rapidly evolving genes.

Tests of molecular evolution in a phylogenetic framework further support the pattern of differences in evolutionary dynamics between the two organs. Molecular evolution analyses using PAML indicates that 58 of 268 (21.6%) of seminal receptacle genes are evolving by positive selection in a subset of the codons (Supplementary table 1) by the acceptance of M8 over M7 and/or M2 over M1. Similar to what was observed using d_N/d_S pairwise comparisons, a significantly higher proportion ($Z = 2.67$, $p < 0.01$) of spermathecae genes (17 of 40; 42.5%) are evolving by positive selection using PAML analysis compared to genes of the seminal receptacle. Comparison of the genes in the seminal receptacle to other relevant studies shows that the proportion of genes evolving by positive selection in the seminal receptacle is high. From a sample of 25 seminal fluid genes, four (16%) exhibited positive selection by the acceptance of M8 over M7 (Haerty *et al.*, 2007). The same study analysed 679 genes expressed in the reproductive tract of *Drosophila* and found 6.2% exhibited positive selection by PAML analysis.

Overall, the incidence of positive selection in the spermathecae is significantly higher than the seminal receptacle. Of the genes undergoing positive selection, there is a higher proportion encoding proteins with secretion signals and transmembrane domains in the

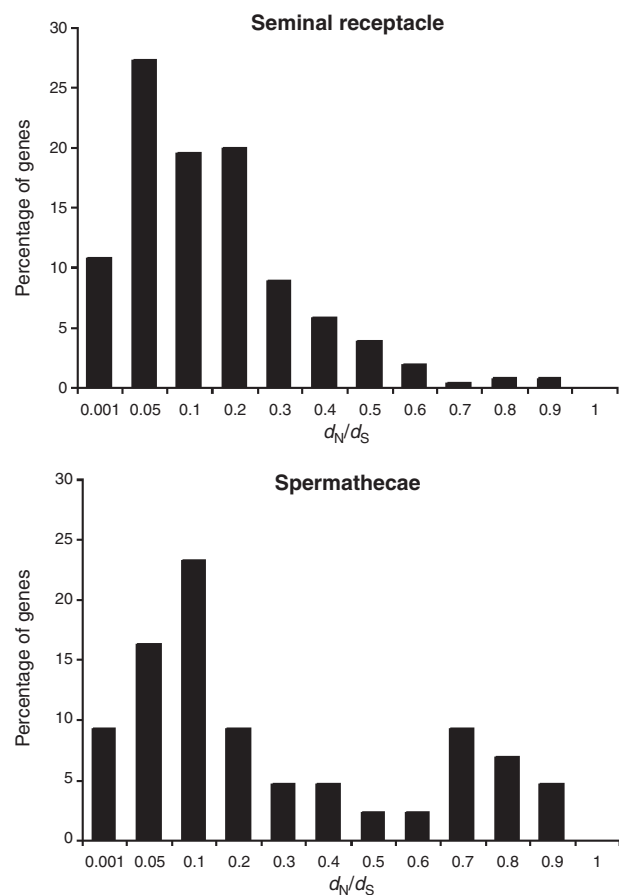


Fig. 2 The distribution of d_N/d_S for genes present in the seminal receptacle expressed sequence tag (EST) library compared to the distribution of d_N/d_S for genes present in the spermathecae EST library.

spermatheca. This is potentially important because interaction between female and male reproductive protein interactions are hypothesized to underlie positive selection (Pitnick *et al.*, 2009). Secreted female proteins, or those with transmembrane domains, are most likely to interact with male reproductive proteins transferred to the female.

Given the relatively small sample size of the spermathecae EST library study compared to that of the seminal receptacle, there is some concern about sampling bias. One issue is whether the sample from the spermathecae library is representative of the transcript level distributions, although the use of normalized libraries should address this concern. To further address this issue, we compared the gene lists from the spermathecae EST library and the seminal receptacle EST library to gene expression data generated as part of a microarray study of the sperm storage organs of *Drosophila melanogaster* (Prokupek *et al.*, 2009). Both libraries contain transcripts from a range of expression levels with uncommon transcripts represented in similar proportions.

Seminal receptacle genes expressed in other organs or tissues

Seminal receptacle genes could be expressed in other reproductive tissues and somatic tissues. Thus, the net effect of selection of genes found in the seminal receptacle library may not be due solely to expression in this organ. It is expected that the process of subtractive hybridization, as used in this study, will enrich the library with transcripts specifically expressed in the seminal receptacle rather than totally exclude genes that are expressed elsewhere. We found that genes expressed in the seminal receptacle overlap with genes expressed in the spermatheca. Our subtractive hybridization did not remove messages from male-specific genes unless they were present in the sperm. Overlap between female genes and those previously presumed to be male specific is interesting in terms of both function and evolution.

Seminal receptacle expression compared with other studies

In this study, a hybrid selected library was used and thus the genes identified are enriched for expression in this sperm storage organ. However, the genes can be expressed in other tissues and selection of them need not be due only to the role of the encoded protein in the seminal receptacle. In this section, the expression of these genes in other places in the body is considered, but a comprehensive analysis of expression in other tissues will be made in a different study.

Comparison with the spermathecae EST library

Seven seminal receptacle genes (CG8331, Treh, Act57B, CG10469, Drs, TspEl and Qm) overlap with genes in the spermathecae EST library (Prokupek *et al.*, 2008). Four of these genes show evidence for positive selection in both studies (CG8331, Act57B, Treh and Qm) based on the analysis using the PAML site model; none show elevated d_N/d_S in the pairwise comparison between *D. melanogaster* and *D. simulans*. Different splice forms of genes have been observed in the two sperm storage organs (Treh-PF and Zm-PA in the seminal receptacle as well as Treh-PC and Om-PA in the spermatheca).

Comparison with the sperm proteome

Although it is generally accepted that sperm are in an inactive transcriptional state, studies involving human have found a small number of mRNAs to be present in ejaculated sperm (reviewed in Miller, 2000). The suggestion is that the mRNAs are not from ongoing transcription within the mature sperm cells, but remnants of stored mRNA from spermatogenesis. The mRNA could be a nonfunctional remnant, but one argument in place is that stored mRNA may be used during the initial steps of fertilization and may contribute to paternal imprinting (Braude *et al.*, 1988; Siffroi & Dadoune,

2001). Additionally, during the process of capacitation and acrosome reaction, low levels of transcription have been reported in mature sperm cells (Miteva *et al.*, 1995; Naz, 1998; Lambard *et al.*, 2004).

Four genes (Acp5C, dp, sesB and Cp-P60A) identified in the seminal receptacle library overlap with those identified as part of the sperm proteome (Karr, 2007). Act5C is related to the cytoskeleton and has been noted to play a role in sperm individualization; dp is a transcription factor; sesB is involved in ADP transport (ion transport); and Ca-P60A is an ATPase involved in calcium ion homeostasis (ion transport). Three of these genes (dp, Act5C and sesB) are predicted to be evolving under positive selection by PAML analysis. The fourth gene Ca-P60A had no predicted orthologs in other *Drosophila* species thus preventing evolutionary analysis. The lack of orthologs in other species, however, indirectly suggests a gene evolving by positive selection. It is likely that these genes are expressed in the seminal receptacle, but it is possible that the mRNA came from stored sperm. Sperm are also present in the ST, so the process of subtractive hybridization used to create the seminal receptacle library should have eliminated the cDNAs from mRNA found in sperm. One hypothesis is that proteins essential to sperm function, which are produced in male structures such as the testis or epididymis where sperm are held, are also produced by the female in the seminal receptacle. As such, proteins expressed in the seminal receptacle may overlap those identified as part of the sperm proteome and seminal fluid/accessory gland proteome. For example, Ca-P60A is an ATPase involved in calcium ion homeostasis and could play a role in preventing premature capacitation of the sperm. The lipid and carbohydrate metabolism genes could be providing an eternal energy source.

Comparison with seminal fluid proteins/accessory gland proteins

Genes present in the seminal receptacle cDNA library can be compared to seminal proteins including the well-studied accessory gland proteins (Acps). In *Drosophila*, males pass to the female proteins at the time of mating a number of proteins which elicit a series of post-mating responses. Acps are responsible for triggering changes in female reproductive morphology and muscle contractions necessary for sperm to enter storage organs (Adams & Wolfner, 2007). Many of these proteins are produced in the male accessory gland. Approximately 112 male accessory gland proteins (Acps) have now been identified. A subset of these Acps have been characterized in terms of their localization within the female, effects they have on the post-mated female, as well as their evolutionary rates. Nine Acps localize in the sperm storage organs (Peng *et al.*, 2005; Ravi-Ram & Wolfner, 2005). Acps have been implicated in the displacement of previous male's sperm (5 Acps), as well as the defence of stored sperm against displacement (7 Acps) (Clark

et al., 1995; Fiumera *et al.*, 2005; Ravi-Ram & Wolfner, 2007). Four Acps are suggested to decrease the longevity of mated females (Acp62F, Sex Peptide, CG8137 and CG10433), two of which are protease inhibitors (reviewed in Ravi-Ram & Wolfner, 2007) and three of which enter the sperm storage organs (Ravi-Ram & Wolfner, 2005). Male seminal fluid proteins (SFPs) that are found in sperm storage organs and proteins on stored sperm are prime candidates to interact with female proteins secreted by the sperm storage organs or those female proteins on sperm storage organ membranes. A high percentage of Acps have evolved rapidly (Swanson & Vacquier, 2002; Haerty *et al.*, 2007; Wong *et al.*, 2008), and this elevated rate of evolution may be attributed to interactions with female reproductive proteins (Rice, 1996; Pitnick *et al.*, 2003).

In the present study, the list of seminal receptacle genes was compared to a list of SFPs which included accessory gland proteins (Acps). The male protein list was synthesized from the supplementary data from Findlay *et al.* (2008) and Ravi-Ram & Wolfner (2007), and included 112 accessory gland protein genes, and an additional 82 SFPs. Comparing the seminal receptacle to the SFP lists, there was a total of 12 common genes (5 Acps). The genes included two of unknown function, two involved in lipid/carbohydrate metabolism, two involved in defence and immunity (a protease inhibitor and a signalling gene), two hydrolases, a gene involved in cell adhesion, an isomerase and a structural gene involved in chitin binding. Esterase-6 (Est-6), a serine hydrolase, is recognized as a SFP with specific expression in the male ejaculatory bulb. Est-6 has been reported in previous studies as affecting both sperm storage and use in female *Drosophila* (Richmond *et al.*, 1980; Gilbert, 1981; Bloch-Qazi & Wolfner, 2003).

Two genes were suggested to play a role in spermatogenesis using GO classifications that were not identified in the sperm proteome or recognized as a seminal fluid protein. Poe is a ligase involved in spermatogenesis and motility and Yuri plays a role in actin and tubulin structures of spermatogenesis and locomotion. The overlap observed with the seminal fluid or accessory gland proteins is an interesting observation especially given that Sfps/Acps are transferred to the female after translation. One possibility is the SFPs have co-opted female sperm storage organ functions to manipulate sperm storage and use. The idea is that female proteins regulate female aspects of reproduction. However, if males express and transfer these proteins to females at the time of mating, then males can manipulate female reproduction in the interest of males.

Functional categories of genes in the seminal receptacle

In addition to evolutionary considerations, the present study provides insight into genes enriched for expression

in the seminal receptacle that could be playing an important role in how this organ functions.

Immune response

Immune response genes are an important category of genes in the seminal receptacle, not only because of the functional role they are likely to play during the reproductive process, but also because host–pathogen co-evolution may be one of the forces leading to the rapid evolution of reproductive proteins. De Gregorio *et al.* (2001) and Irving *et al.* (2001) performed microarray analyses on *D. melanogaster* after immune challenge (bacterial or fungal) and generated lists of genes classified as immune response genes. By comparing the seminal receptacle library genes with the immune response genes identified in the microarray studies we were able to classify 14 seminal receptacle genes as immunity/defence related. Six additional genes, not identified in the DIRG comparison, were classified as “involved in immunity and defence” using GO searches. In total, 20 putative defence/immunity genes (7 per cent of the identified seminal receptacle genes) were identified in the seminal receptacle.

Lipid/Carbohydrate metabolism

Lipid and carbohydrate metabolism genes made up the largest functional group in the seminal receptacle. Genes encoding two glycosyltransferases (CG17323 and CG11289) and two glycosidases (Trehalose (Treh) and hexoaminidase 2 (Hexo2) were present in the seminal receptacle library. These enzymes hydrolyse terminal sugar residues in different glycoconjugates. Glycosidases and glycosyltransferases have been found in several animal studies and have predicted roles in sperm–egg interactions. They form complexes with oligosaccharide substrates on the egg coats in noncatalytic conditions (Shur, 1989). In mammals, these enzymes facilitate the surface modification of spermatocytes to enhance sperm–egg interactions (Tulsiani *et al.*, 1998; Zitta *et al.*, 2006). N-Acetyl-D-glucosaminidase (NAG), also known as hexosaminidase, is found in fluids surrounding sperm and may prevent the initiation of premature acrosomal reaction by binding to glucosamine which triggers the acrosome reaction process (Brandelli *et al.*, 1994).

Secreted glycosidases have been discovered in male (including *Drosophila*) reproductive tracts (Hall *et al.*, 1992; Barbieri *et al.*, 1995; Cattaneo *et al.*, 2002; Intra *et al.*, 2006) and have suggested involvement in the modification of sperm surface glycoproteins and sperm maturation. Three distinct glycosidases (two B-N-acetylhexoaminidase isoforms HEX1 and HEX2 and a-mannosidase) have been characterized in male *D. melanogaster* (Cattaneo *et al.*, 2002). Both HEX1 and HEX2 are heterodimers orthologous to mammalian isoenzyme A and B, respectively (Cattaneo *et al.*, 2002).

One carbohydrate metabolism gene found in the seminal receptacle library has activity which is associated

with sperm function in humans. Gene CG14478 found in the seminal receptacle EST library has alpha-glucosidase activity. In humans, Alpha-glucosidase activity is positively correlated with motility and sperm forward progression (Viljoen *et al.*, 1990; Elzanaty *et al.*, 2002).

Additional roles for carbohydrate/lipid metabolism genes could include breaking down complex energy sources to be utilized by sperm in storage. Lipid metabolism genes could participate in the remodelling of the phospholipids membrane of the sperm for quiescence or capacitation.

Protease inhibitors

A total of nine protease inhibitors were identified in the seminal receptacle library. This category of genes was found to be significantly over-represented in the genes expressed in the seminal receptacle. Protease inhibitors are known to be important molecules in reproduction. For example, the absence or mutation of protease inhibitors in the seminal fluid impairs fertility in mammals such as mice (lacking protease nexin-1) (Murer *et al.*, 2001) and humans (lacking protein C inhibitor) (He *et al.*, 1999). Inhibitors in the seminal fluid have been proposed to: (i) protect and regulate the reproductive tract and SFPs (including sperm) from proteolysis by damaged or prematurely acrosome-reacted sperm (Monsma *et al.*, 1990; Park & Wolfner, 1995), (ii) regulate processing or proteolytic degradation of seminal fluid or sperm surface proteins (Veselsky & Cechova, 1980) and (iii) regulate the coagulation of seminal fluid (Kise *et al.*, 1996; Lung *et al.*, 2002). Protease inhibitors expressed in the seminal receptacle may be performing similar roles as those from the seminal fluid or may be interacting with male proteases transferred in the seminal fluid.

Directional cues (chemical and mechanical)

After deposition in the female, sperm move away from the mating plug (formed during copulation) and congregate at the anterior uterus. The sperm must then navigate through the narrow openings which connect to the spermathecae and seminal receptacle, avoiding the large common oviduct (Gao *et al.*, 2003). The movement of the sperm within the female reproductive tract is likely to be accomplished by directional (mechanical or chemical) cues. When male sperm is unable to respond to directional cues in the female reproductive tract, the result is male sterility (Gao *et al.*, 2003; Venglarik *et al.*, 2004). Additionally, the individual release of sperm from storage and direction towards the egg for fertilization would likely require a second set of cues. Seminal receptacle overrepresented gene sets included those involved in muscle contraction as well as olfaction, functions that plausibly could play roles as directional cues for sperm.

Chemical cues (odorant receptor and odorant-binding)

Odorant receptors and signalling mechanisms have been discovered in mammalian sperm (Spehr *et al.*, 2006), and

odorant-binding genes have been found in the sperm proteome and SFPs of *Drosophila* (Findlay *et al.*, 2008; Takemori & Yamamoto, 2009). Both human and mouse sperm respond to odorant (or olfactory) receptor activation, including directed movement, hyperactive flagellar beating and elevated swimming speeds (Spehr *et al.*, 2006). Odorant receptors may play a role in direct sperm–egg communication or participate in other pre-fertilization process, which ‘prime’ sperm for fertilization (induction of capacitation, release from oviductal storage).

Six odorant-binding proteins are among the SFPs transferred to the female (Findlay *et al.*, 2008). The genes could play a role in male–female communication, similar to mammalian odorant receptors. Four genes expressed in the seminal receptacle had odorant-related functions, one odorant receptor (Or47a) and three odorant binding (Obp57a, CG1124 and CG13027). These female odorant proteins could present a mechanism for chemical guidance cues, ensuring the proper direction of sperm post-insemination. Further, odorant binding/receptors may play a role in the recognition of sperm and in this way may function as mechanisms for cryptic female choice.

Mechanical cues (Muscle contraction)

The *Drosophila* female reproductive tract, including the sperm storage organs, is integrated with muscle tissue. Rhythmic contractions have been observed in the ovarioles, peritoneal sheath, oviduct and spermathecae in correlation with egg movement and ovipositioning (Middleton *et al.*, 2006). The seminal receptacle is surrounded by a layer of visceral muscle (Blaney, 1970; Filosi & Perotti, 1975). Implications for muscular contraction have been suggested in female-mediated sperm storage as well as displacement in a variety of invertebrate taxa (Haase & Baer, 1995; Arthur *et al.*, 1998; Hellriegel & Bernasconi, 2000). Flour beetles require female muscle contraction for sperm storage to be successful ((Bloch-Qazi *et al.*, 1998; Fedina & Lewis, 2004). Acps may stimulate muscles in the female’s lower reproductive tract to contract and relax, facilitating the movement of the sperm mass towards storage sites, as well as exposing the openings of these sites (Adams & Wolfner, 2007). Vesicle release in lower reproductive tract nerve termini modulated by Acps could also mediate muscular contraction/release (Heifetz & Wolfner, 2004).

A total of 16 genes were found in the seminal receptacle library which play a role in muscle contraction. One of these genes (Prm) is predicted to play a role in defence/immunity, and a second (Mp20) is predicted to play a role in detoxification. Muscle contraction genes may play a role in the entrance into and the release from the seminal receptacle.

NADPH oxidase (dNOX) was found in the current study, which has been previously identified as a regulator of smooth muscle contraction in *Drosophila* (Ritsick *et al.*,

2007). RNAi knockdowns of these gene result in deficient female egg-laying. The decrease in egg-laying is attributed to a failure of muscle contraction in the movement of the egg from the ovary to the oviduct (Ritsick *et al.*, 2007). The *Drosophila* Nox gene is an ortholog to the human NOX5 and dDuox genes. In humans, NOX5 is expressed in lymphoid organs and testis. NOX5 generates superoxide in response to intracellular Ca^{2+} elevations (Banfi *et al.*, 2004) and is suggested to regulate redox-dependent processes in lymphocytes and spermatozoa.

Conclusions

This study focused on sperm storage organs, which have been understudied in terms of evolution and categorization of protein function. The genes expressed in the seminal receptacle are evolving at rates higher than that of general reproductive proteins, but lower than that of the spermatheca. Thus, the evolutionary pressures operating in these two organs appear to be quite different, despite their similar role in reproduction (sperm storage). Overrepresented gene categories in the seminal receptacle include muscle contraction, immunity, odorant binding/receptors and protease inhibitors. Muscle contraction and odorant binding/receptors may play a major role in the proper direction of sperm within the female. Although further research needs to be carried out to characterize male–female interactions, it is plausible that protease inhibitors interact directly with male proteases transferred in the seminal fluid and that these interactions may result in sexually antagonistic co-evolution. In general, this study provides unique information about the evolution of genes expressed in sperm storage organs and about categories of genes that could play an important role in the function of the seminal receptacle.

Acknowledgments

This research was supported by a National Science Foundation grant (DEB-ESP0346476). We thank Dr. Federico Hoffmann for contributions to the Introduction section. Appreciation is expressed to the reviewers for insightful comments.

References

- Adams, E.M. & Wolfner, M.F. 2007. Seminal proteins but not sperm induce morphological changes in the *Drosophila melanogaster* female reproductive tract during sperm storage. *J. Insect Physiol.* **53**: 319–331.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Arnqvist, G., Edvardsson, M., Friberg, U. & Nilsson, T. 2000. Sexual conflict promotes speciation in insects. *Proc. Natl Acad. Sci. USA* **97**: 10460–10464.
- Arthur, B.I., Hauschteck-Jungen, E., Nothiger, R. & Ward, P.I. 1998. A female nervous system is necessary for normal sperm storage in *Drosophila melanogaster*: a masculinized nervous system is as good as none. *Proc. R. Soc. Lond. B* **265**: 1749–1753.
- Baer, B., Armitage, S.A.O. & Boomsma, J.J. 2006. Sperm storage induces an immunity cost in ants. *Nature* **441**: 872–876.
- Banfi, B., Tirone, F., Durussel, I., Knisz, J., Moskwa, P., Molnar, G.Z., Krause, K.H. & Cox, J.A. 2004. Mechanism of Ca^{2+} activation of the NADPH oxidase 5 (NOX5). *J. Biol. Chem.* **279**: 18583–18591.
- Barbieri, A.M., Sosa, M.A., Grimalt, P., Mayorga, L.S. & Bertini, F. 1995. Phosphomannosyl receptors on the surface of spermatozoa from the caudal epididymis of the rat. *Int. J. Androl.* **18**: 113–119.
- Begun, D.J., Whitley, P., Todd, B.L., Waldrip-Dail, H.M. & Clark, A.G. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879–1888.
- Bielawski, J.P. & Yang, Z. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Funct. Genomics* **3**: 201–212.
- Blaney, W.M. 1970. Some observations on the sperm tail of *D. melanogaster*. *Drosoph. Inf. Serv.* **45**: 125–127.
- Bloch-Qazi, M.C. & Wolfner, M.F. 2003. An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. *J. Exp. Biol.* **206**: 3521–3528.
- Bloch-Qazi, M.C., Aprile, J.R. & Lewis, S.M. 1998. Female role in sperm storage in the red flour beetle, *Tribolium castaneum*. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* **120**: 641–648.
- Brandelli, A., Miranda, P.V. & Tezon, J.G. 1994. Participation of glycosylated residues in the human sperm acrosome reaction: possible role of N-acetylglucosaminidase. *Biochim. Biophys. Acta* **1220**: 299–304.
- Braude, P., Bolton, V. & Moore, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**: 459–461.
- Cattaneo, F., Ogiso, M., Hoshi, M., Perotti, M.E. & Pasini, M.E. 2002. Purification and characterization of the plasma membrane glycosidases of *Drosophila melanogaster* spermatozoa. *Insect Biochem. Mol. Biol.* **32**: 929–941.
- Civetta, A. & Singh, R.S. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J. Mol. Evol.* **41**: 1085–1095.
- Clark, A.G., Aguade, M., Prout, T., Harshman, L.G. & Langley, C.H. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**: 189–201.
- Clark, N.L., Aagaard, J.E. & Swanson, W.J. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**: 11–22.
- Collins, A.M., Caperna, T.J., Williams, V., Garrett, W.M. & Evans, J.D. 2006. Proteomic analysis of male contributions to honeybee sperm storage and mating. *Insect Mol. Biol.* **15**: 541–549.
- Das, S.C., Isobe, N., Nishibori, M. & Yoshimura, Y. 2006. Expression of transforming growth factor-beta isoforms and their receptors in utero-vaginal junction of hen oviduct in presence or absence of resident sperm with reference to sperm storage. *Reproduction* **132**: 781–790.
- De Gregorio, E., Spellman, P.T., Rubin, G.M. & Lemaitre, B. 2001. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl Acad. Sci. USA* **98**: 12590–12595.

- Diatchenko, L., Lau, Y., Campbell, A., Chenchik, A., Mogadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. & Siebert, P. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA* **93**: 6025–6030.
- Diatchenko, L., Lukyanov, K., Lau, Y. & Siebert, P. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Meth. Enzymol.* **303**: 349–380.
- Ellegren, H. 2008. Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* **17**: 4586–4596.
- Elzanaty, S., Richthoff, J., Malm, J. & Giwercman, A. 2002. The impact of epididymal and accessory sex gland function on sperm motility. *Hum. Reprod.* **17**: 2904–2911.
- Emanuelsson, O., Ren-Brunak, S., vonHeijne, G. & Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**: 953–971.
- Fedina, T.Y. & Lewis, S.M. 2004. Female influence over offspring paternity in the red flour beetle *Tribolium castaneum*. *Proc. Biol. Sci.* **271**: 1393–1399.
- Ferrandon, D., Jung, A.C., Criqui, M., Lemaitre, B., Uttenweiler-Joseph, S., Michaut, L., Reichhart, J. & Hoffmann, J.A. 1998. A drosomycin-GFP reporter transgene reveals a local immune response in *Drosophila* that is not dependent on the Toll pathway. *EMBO J.* **17**: 1217–1227.
- Filosi, M. & Perotti, M. 1975. Fine structure of spermatheca of *Drosophila melanogaster* Meig. *J. Submicrosc. Cytol.* **7**: 259–270.
- Findlay, G.D., Yi, X., Maccoss, M.J. & Swanson, W.J. 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol.* **6**: e178.
- Fiumera, A.C., Dumont, B.L. & Clark, A.G. 2005. Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. *Genetics* **169**: 243–257.
- Gao, Z., Ruden, D.M. & Lu, X. 2003. PKD2 cation channel is required for directional sperm movement and male fertility. *Curr. Biol.* **13**: 2175–2178.
- Gilbert, D.G. 1981. Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster*. *J. Insect Physiol.* **27**: 641–650.
- Gist, D.H. & Congdon, J.D. 1998. Oviductal sperm storage as a reproductive tactic of turtles. *J. Exp. Zool.* **282**: 526–534.
- Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**: 725–736.
- Haase, M. & Baer, B. 1995. Variation in spermathecal morphology and storage of spermatozoa in the simultaneously hermaphroditic land snail *Ariata arbustorum* (Gastropoda: Pulmonata: Stylommatophora). *Invertebr. Reprod. Dev.* **28**: 33–42.
- Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ravi Ram, K., Sirot, L.K., Levesque, L., Artieri, C.G., Wolfner, M.F., Civetta, A. & Singh, R.S. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**: 1321–1335.
- Hall, J.C., Kochins, J.G. & Perez, F.M. 1992. Purification and characterization of N-acetyl-beta-D-hexosaminidase in different anatomical regions of the adult rat epididymis. *Biochem. Int.* **28**: 613–620.
- He, S., Lin, Y.L. & Liu, Y.X. 1999. Functionally inactive protein C inhibitor in seminal plasma may be associated with infertility. *Mol. Hum. Reprod.* **5**: 513–519.
- Heifetz, Y. & Wolfner, M.F. 2004. Mating, seminal fluid components, and sperm cause changes in vesicle release in the *Drosophila* female reproductive tract. *Proc. Natl Acad. Sci. USA* **101**: 6261–6266.
- Hellriegel, B. & Bernasconi, G. 2000. Female-mediated differential sperm storage in a fly with complex spermathecae, *Scatophaga stercoraria*. *Anim. Behav.* **59**: 311–317.
- Holland, B. & Rice, W.R. 1999. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc. Natl Acad. Sci. USA* **96**: 5083–5088.
- Huang, X. & Madan, A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* **9**: 868–877.
- Intra, J., Cenni, F. & Perotti, M.E. 2006. An alpha-L-fucosidase potentially involved in fertilization is present on *Drosophila* spermatozoa surface. *Mol. Reprod. Dev.* **73**: 1149–1158.
- Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., Reichhart, J.M., Hoffmann, J.A. & Hetru, C. 2001. A genome-wide analysis of immune responses in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**: 15119–15124.
- Käll, L., Krogh, A. & Sonnhammer, E.L.L. 2004. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**: 1027–1036.
- Karr, T.L. 2007. Fruit flies and the sperm proteome. *Hum. Mol. Genet.* 16 Spec No. 2: R124–R133.
- Katoh, K. & Toh, H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformatics* **9**: 286–298.
- Kelleher, E.S., Swanson, W.J. & Markow, T.A. 2007. Gene duplication and adaptive evolution of digestive proteases in *Drosophila arizonae* female reproductive tracts. *PLoS Genet.* **3**: e148.
- Kise, H., Nishioka, J., Kawamura, J. & Suzuki, K. 1996. Characterization of semenogelin II and its molecular interaction with prostate-specific antigen and protein C inhibitor. *Eur. J. Biochem.* **238**: 88–96.
- Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A. & Carreau, S. 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. *Mol. Hum. Reprod.* **10**: 535–541.
- Lung, O., Tram, U., Finnerty, C.M., Eipper-Mains, M.A., Kalb, J.M. & Wolfner, M.F. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics* **160**: 211–224.
- Marchler-Bauer, A. & Bryant, S. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**: 327–331.
- Middleton, C.A., Nongthomba, U., Parry, K., Sweeney, S.T., Sparrow, J.C. & Elliott, C.J. 2006. Neuromuscular organization and aminergic modulation of contractions in the *Drosophila* ovary. *BMC Biol.* **4**: 17.
- Miller, D. 2000. Analysis and significance of messenger RNA in human ejaculated spermatozoa. *Mol. Reprod. Dev.* **56**: 259–264.
- Miller, G.T. & Pitnick, S. 2002. Sperm-female coevolution in *Drosophila*. *Science* **298**: 1230–1233.
- Miteva, K., Valkov, N., Goncharova-Peinoval, J., Kovachev, K., Zlatarev, S., Pironcheva, G. & Russev, G. 1995. Electron microscopic demonstration of transcription of ram sperm chromatin. *Microbios* **84**: 91–96.
- Monsma, S., Harada, H. & Wolfner, M. 1990. Synthesis of two *Drosophila* male accessory gland proteins and their fate after transfer to the female during mating. *Dev. Biol.* **142**: 465–475.
- Mueller, J.L., Ravi Ram, K., McGraw, L.A., Bloch Qazi, M.C., Siggia, E.D., Clark, A.G., Aquadro, C.F. & Wolfner, M.F. 2005.

- Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**: 131–143.
- Murer, V., Spetz, J.F., Hengst, U., Altrogge, L.M., de Agostini, A. & Monard, D. 2001. Male fertility defects in mice lacking the serine protease inhibitor protease nexin-1. *Proc. Natl Acad. Sci. USA* **98**: 3029–3033.
- Naz, R.K. 1998. c-Abl proto-oncoprotein is expressed and tyrosine phosphorylated in human sperm cell. *Mol. Reprod. Dev.* **51**: 210–217.
- Neubaum, D.M. & Wolfner, M.F. 1999. Wise, winsome, or weird? Mechanisms of sperm storage in female animals. *Curr. Top. Dev. Biol.* **41**: 67–97.
- Park, M. & Wolfner, M.F. 1995. Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. *Dev. Biol.* **171**: 694–702.
- Peng, J., Chen, S., Busser, S., Liu, H., Honegger, T. & Kubli, E. 2005. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Curr. Biol.* **15**: 207–213.
- Pitnick, S., Miller, G.T., Schneider, K. & Markow, T.A. 2003. Ejaculate-female coevolution in *Drosophila mojavensis*. *Proc. Biol. Sci.* **270**: 1507–1512.
- Pitnick, S., Wolfner, M.F. & Suarez, S. 2009. Ejaculate- and sperm- female interactions. In: *Sperm Biology: An Evolutionary Perspective* (T.R. Birkhead, D.J. Hosken & S. Pitnick, eds), pp. 247–304. Elsevier Press, Amsterdam.
- Pollard, D.A., Iyer, V.N., Moses, A.M. & Eisen, M.B. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet.* **2**: e173.
- Powell, J.R. & Moriyama, E.N. 1997. Evolution of codon usage bias in *Drosophila*. *Proc. Natl Acad. Sci. USA* **94**: 7784–7790.
- Prokupek, A., Hoffmann, F., Eyun, S.I., Moriyama, E., Zhou, M. & Harshman, L. 2008. An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution* **62**: 2936–2947.
- Prokupek, A.M., Kachman, S.D., Ladunga, I. & Harshman, L.G. 2009. Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*. *Insect Mol. Biol.* **18**: 465–475.
- Racey, P.A. 1972. Viability of bat spermatozoa after prolonged storage in the epididymis. *J. Reprod. Fertil.* **28**: 309–311.
- Racey, P.A. 1979. The prolonged storage and survival of spermatozoa in Chiroptera. *J. Reprod. Fertil.* **56**: 391–402.
- Ravi-Ram, K. & Wolfner, M. 2005. Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **35**: 1059–1071.
- Ravi-Ram, K. & Wolfner, M.F. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr. Comp. Biol.* **47**: 427–445.
- Rice, W.R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**: 232–234.
- Rice, W.R. & Holland, B. 2005. Experimentally enforced monogamy: inadvertent selection, inbreeding, or evidence for sexually antagonistic coevolution? *Evolution* **59**: 682–685.
- Rice, W.R., Stewart, A.D., Morrow, E.H., Linder, J.E., Orteiza, N. & Byrne, P.G. 2006. Assessing sexual conflict in the *Drosophila melanogaster* laboratory model system. *Philos. Trans. R. Soc. Lond. B* **361**: 287–299.
- Richmond, R.C., Gilbert, D.G., Sheehan, K.B., Gromko, M.H. & Butterworth, F.M. 1980. Esterase 6 and reproduction in *Drosophila melanogaster*. *Science* **207**: 1483–1485.
- Ritsick, D.R., Edens, W.A., Finnerty, V. & Lambeth, J.D. 2007. Nox regulation of smooth muscle contraction. *Free Radic. Biol. Med.* **43**: 31–38.
- Shur, B.D. 1989. Expression and function of cell surface galactosyltransferase. *Biochim. Biophys. Acta* **988**: 389–409.
- Siffroi, J.P. & Dadoune, J.P. 2001. Accumulation of transcripts in the mature human sperm nucleus: implication of the haploid genome in a functional role. *Ital. J. Anat. Embryol.* **106**: 189–197.
- Spehr, M., Schwane, K., Riffell, J.A., Zimmer, R.K. & Hatt, H. 2006. Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. *Mol. Cell. Endocrinol.* **250**: 128–136.
- Suyama, M., Torrents, D. & Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl. Acids Res.* **34**: W609–W612.
- Swanson, W.J. 2003. Adaptive evolution of genes and gene families. *Curr. Opin. Genet. Dev.* **13**: 617–622.
- Swanson, W.J. & Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137–144.
- Swanson, W.J., Yang, Z., Wolfner, M.F. & Aquadro, C.F. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl Acad. Sci. USA* **98**: 2509–2514.
- Swanson, W.J., Wong, A., Wolfner, M.F. & Aquadro, C.F. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**: 1457–1465.
- Takemori, N. & Yamamoto, M.T. 2009. Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* **9**: 2484–2493.
- Tulsiani, D.R., Orgebin-Crist, M.C. & Skudlarek, M.D. 1998. Role of luminal fluid glycosyltransferases and glycosidases in the modification of rat sperm plasma membrane glycoproteins during epididymal maturation. *J. Reprod. Fertil. Suppl.* **53**: 85–97.
- Tusnady, G.E. & Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849–850.
- Venglarik, C.J., Gao, Z. & Lu, X. 2004. Evolutionary conservation of *Drosophila* polycystin-2 as a calcium-activated cation channel. *J. Am. Soc. Nephrol.* **15**: 1168–1177.
- Veselsky, L. & Cechova, D. 1980. Distribution of acrosin inhibitors in bull reproductive tissues and spermatozoa. *Hoppe-Seyler's Z. Physiol. Chem.* **361**: 715–722.
- Viljoen, M.H., Bornman, M.S., van der Merwe, M.P. & du Plessis, D.J. 1990. Alpha-glucosidase activity and sperm motility. *Andrologia* **22**: 205–208.
- Wagstaff, B.J. & Begun, D.J. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* **171**: 1083–1101.
- Wira, C.R. & Fahey, J.V. 2004. The innate immune system: gatekeeper to the female reproductive tract. *Immunology* **111**: 13–15.
- Wong, A., Turchin, M.C., Wolfner, M.F. & Aquadro, C.F. 2008. Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. *Mol. Biol. Evol.* **25**: 497–506.
- Yamanouye, N., Silveira, P.F., Abdalla, F.M., Almeida-Santos, S.M., Breno, M.C. & Salomao, M.G. 2004. Reproductive cycle of the Neotropical *Crotalus durissus terrificus*: II. Establishment and maintenance of the uterine muscular twisting, a

- strategy for long-term sperm storage. *Gen. Comp. Endocrinol.* **139**: 151–157.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**: 568–573.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.
- Yang, Z. & Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.
- Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *J. Mol. Evol.* **46**: 409–418.
- Yang, Z. & Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908–917.
- Yang, Z. & Swanson, W.J. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**: 49–57.
- Yang, Z., Swanson, W.J. & Vacquier, V.D. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* **17**: 1446–1454.
- Zhang, Y.Y. 2006. Semen characterization and sperm storage in Cabot's Tragopan. *Poult. Sci.* **85**: 892–898.
- Zhu, Y., Machleder, E., Chenchik, A. & Siebert, P. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**: 892–897.
- Zitta, K., Wertheimer, E.V. & Miranda, P.V. 2006. Sperm N-acetylglucosaminidase is involved in primary binding to the zona pellucida. *Mol. Hum. Reprod.* **12**: 557–563.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Gene identification numbers for *D. melanogaster*. *D. melanogaster* CG and FB gene identification numbers; Function of predicted proteins; Secretion Signal prediction; Transmembrane region prediction; Pairwise d_N/d_S ; PAML data; Presence of orthologs in 12 genomes.

Table S2 Gene identification numbers for *D. simulans* and *D. melanogaster*; Pairwise d_N/d_S ; PAML branch-model data.

Table S3 Species-specific gene duplications.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 12 March 2010; accepted 17 March 2010