Molecular Evolution of the Sex Peptide Network in *Drosophila*

Meaghan K. McGeary[1,2] and Geoffrey D. Findlay[1,*]

[1]Department of Biology, College of the Holy Cross, Worcester, MA 01610

[2]Department of Pathology, Yale School of Medicine, New Haven, CT 06520

*Correspondence to: gfindlay@holycross.edu, 508-793-2655 (office), 508-793-2656 (fax)

Short title: *Drosophila* sex peptide network evolution

Abstract

Successful reproduction depends on interactions between numerous proteins beyond those involved directly in gamete fusion. While such reproductive proteins evolve in response to sexual selection pressures, how networks of interacting proteins arise and evolve as reproductive phenotypes change remains an open question. Here, we investigated the molecular evolution of the "sex peptide network" of *Drosophila melanogaster,* a functionally well-characterized reproductive protein network. In this species, the peptide hormone sex peptide (SP) and its interacting proteins cause major changes in female physiology and behavior after mating. In contrast, females of more distantly related *Drosophila* species do not respond to SP. In spite of these phenotypic differences, we detected orthologs of all network proteins across 22 diverse *Drosophila* species and found evidence that most orthologs likely function in reproduction throughout the genus. In *D. melanogaster* and closely related species that show similar responses to SP, we detected the recurrent, adaptive evolution of several network proteins, consistent with sexual selection acting to continually refine network function. We also found some evidence for adaptive evolution of several proteins along two key branches of the *Drosophila* phylogeny on which major changes in SP-related phenotypes likely occurred. Finally, we used gene expression profiling to examine the likely degree of functional conservation of the paralogs of an SP network protein that arose via gene duplication. Our results suggest a dynamic history for the SP network in which network members arose before the onset of robust SP-mediated responses and then were shaped by both purifying and positive selection.

Keywords: sex peptide, *Drosophila*, seminal fluid, sexual selection, molecular evolution

Introduction

Successful reproduction requires the fusion of egg and sperm cells, yet this fusion is often facilitated by proteins that are not part of the gametes. For example, non-gametic reproductive proteins provided in male seminal fluid or produced in the female reproductive tract can facilitate sperm motility, induce or manage sperm storage, or cause changes to female reproductive physiology (Wilburn & Swanson, 2016, Schnakenberg et al., 2011). While proteomic and comparative genomic methods have enabled the identification of hundreds of gametic and non-gametic reproductive proteins across diverse taxa (reviewed in McDonough et al., 2016), understanding how these proteins interact, and how such interactions evolve, remain areas of active research.

Some of the best-characterized reproductive protein interactions occur in the "sex peptide network" of *Drosophila melanogaster* that regulates female post-mating behavior and physiology. The network centers on the sex peptide (SP), a short peptide hormone transferred from males to females as a non-gametic component of seminal fluid (Chen et al., 1988). The presence of SP in the female reproductive tract stimulates egg production (Soller et al., 1999), reduces receptivity to remating (Liu & Kubli, 2003, Chapman et al., 2003), facilitates the release of sperm from storage prior to fertilization (Avila et al., 2010), and affects numerous other female behaviors, including feeding, defecation and sleep (Carvalho et al., 2006, Apger-McGlaughon & Wolfner, 2013, Isaac et al., 2010). SP-mediated effects on females persist for several days after mating because SP binds to sperm, which become stored in specialized sperm storage organs in the female tract (Peng et al., 2005). SP is then gradually cleaved from sperm and released from the storage organs into the female tract, where it interacts with the sex peptide receptor (SPR), a G-protein coupled receptor that is expressed in a subset of neurons innervating

the uterus (Yapici et al., 2008, Hasemeyer et al., 2009, Yang et al., 2009). This gradual "dosing" of SP causes the persistence of the hormone's effects on female behavior and physiology. SPR signaling is also required for the efficient release of sperm from the storage organs (Avila et al., 2015).

While the molecule(s) on sperm to which SP binds remain unknown, RNAi screens have identified several additional male seminal fluid proteins and female reproductive tract proteins required for robust SP responses (Ravi Ram & Wolfner, 2007, Ravi Ram & Wolfner, 2009, LaFlamme et al., 2012, Findlay et al., 2014, Singh et al., 2018). Together with SP and SPR, these proteins comprise the SP network. The male-derived proteins include: predicted C-type lectins CG1652 and CG1656; predicted proteases/protease homologs CG9997, seminase, aquarius and intrepid; and, predicted cysteine-rich secretory proteins CG17575 and antares. The female-derived proteins include fra mauro (a predicted metallopeptidase), Esp (a predicted anion transporter) and hadley (which lacks identifiable protein domains). The male-derived proteins act interdependently to facilitate SP binding to sperm (Ravi Ram & Wolfner, 2009, Findlay et al., 2014, Singh et al., 2018), while the female-derived proteins act downstream of SP binding to sperm, potentially by facilitating SP-SPR signaling (Findlay et al., 2014). Other genes expressed in the secondary cells of the male accessory gland are also required for SP-mediated responses, though it remains unclear whether these genes encode proteins that interact directly with the network proteins described above (Sitnik et al., 2016).

SP's functions and interactions have been well characterized in *D. melanogaster*, but comparative genomic and functional studies have shown that the SP response is not conserved throughout the *Drosophila* genus. Tsuda et al. (2015) found that only species of the *melanogaster* group of *Drosophila* (Fig. 1) show changes in female remating receptivity and egg

production upon injection with synthetic SP, even though SP and SPR orthologs can be detected outside of this group. Furthermore, by incubating GFP-labeled SP with female reproductive tracts from progressively more divergent species, Tsuda et al. (2015) discovered that SP could bind to the female tract only in *melanogaster* group species. This observation suggested that robust expression of SPR in the female tract evolved on the phylogenetic lineage leading to the *melanogaster* group, which the authors tested by comparing *SPR* gene expression between in-group and out-group species (Tsuda et al., 2015). Consistent with *D. melanogaster* expression patterns (Yapici et al., 2008), they found that *SPR* was expressed in non-reproductive areas in both sexes of all species examined. However, its expression in the female reproductive tract was largely limited to the *melanogaster* group. (The only outgroup species that showed expression in this location was *D. virilis*, but conspecific GFP-labeled SP did not bind to female reproductive tracts in this species). Intriguingly, the SP ortholog from *D. pseudoobscura* (a non-*melanogaster* group species) is expressed in *D. pseudoobscura* male reproductive tracts (Yang et al., 2018) and can trigger SP-mediated responses when injected into *D. melanogaster* females, but not when injected into conspecifics (Tsuda et al., 2015). This result suggests that the SP protein might have evolved the potential to affect female post-mating behavior before the emergence of the *melanogaster* group, but this function was not fully realized until the subsequent evolution of SPR expression in the female reproductive tract (and, perhaps, within specific neurons in the tract) (Hasemeyer et al., 2009, Yang et al., 2009, Yapici et al., 2008, Rezaval et al., 2012). It is also possible that the transition to high levels of SPR expression in the female reproductive tract created or increased an evolutionary selective pressure to bind higher levels of SP to stored sperm.

While *Drosophila* species differ in reproductive traits for many reasons (Markow & O'Grady, 2005), some of these differences relate directly to the SP network and could thus be causes or consequences of SP network evolution. For example, a change in sperm length may affect the amount of SP that can bind, and other structural changes to sperm could affect the binding of SP and other network proteins that interact with sperm, such as CG1652 and CG1656 (Ravi Ram & Wolfner, 2009, Singh et al., 2018). The amount of SP bound to sperm, the rate and efficacy of its release (Peng et al., 2005), and its ability to bind SPR (Yapici et al., 2008) could affect re-mating rates, while the structure of female sperm storage organs could affect the ability of the network proteins to bind SP to sperm or modulate SP's interaction with SPR. Changes in these traits – sperm length, female remating rate, and female sperm storage structures – have been well documented in the literature (Alpern et al., 2019, Joly & Bressac, 1994, Markow, 1996, Markow & O'Grady, 2005, Pitnick et al., 1999, Snook et al., 1994, Snook, 1995, Singh et al., 2002), and we summarize them and infer their timing in Figure 1. Of particular relevance to this study, the phenotypic and phylogenetic data are consistent with SPR evolving to be expressed in female reproductive tracts along the lineage leading to the *melanogaster* group of flies (Tsuda et al., 2015; branch 11), and an increase in sperm length on the lineage leading to *D. ananassae* and *D. bipectinata* (Joly & Bressac, 1994, Markow, 1996, Pitnick et al., 1999; branch 15).

In light of the differences between species in reproductive phenotypes, we used comparative genomics and molecular evolutionary analysis to gain insights into the evolution of the SP network. While robust, long-lasting changes in female behavior and physiology due to SP are found only in the *melanogaster* group of *Drosophila*, we identified orthologs of each SP network protein in numerous outgroup species and verified their expression in the male

reproductive system in two such species using published data (Yang et al., 2018, Kelleher et al., 2009). Using PAML (Yang, 2007) we determined that recurrent positive selection has acted on specific sites in several of these proteins. We also detected marginal evidence that positive selection has acted on certain network proteins on key phylogenetic lineages corresponding with major changes in SP-related phenotypes. Finally, we used gene expression analysis to investigate the functional consequences of a gene duplication event that gave rise to one of the SP network proteins, seminase. Taken together, our results suggest that the members of the SP network had the potential to influence reproductive success before the onset of SP/SPR-mediated responses in the reproductive tract of mated females of the *melanogaster* group of species. However, additional adaptive changes in these proteins occurred concurrent with, and subsequent to, these critical changes in the fly reproductive system. These results underscore the strength of sexual selection acting in *Drosophila* and illustrate potential molecular changes that occur in the face of such selection.

Methods

Identification of SP network proteins across Drosophila *species*

We obtained the protein sequence for each SP network protein in *D. melanogaster* from FlyBase. For species for which protein annotations were available on FlyBase (Drosophila 12 Genomes et al., 2007), we obtained orthologous protein-coding DNA sequences using the FlyBase Orthologs feature. These species included *Drosophila simulans, sechellia, yakuba, erecta, ananassae, pseudoobscura, persimillis, willistoni, mojavensis, virilis* and *grimshawi.* For species with sequenced genomes that lacked FlyBase protein annotations (Chen et al., 2014), we manually searched for gene orthologs using tBLASTn and the *D. melanogaster* protein sequence as the query. These species include *Drosophila ficusphila, eugracilis, takahashii, elegans, rhopoloa, kikkawai, bipectinata, miranda* and *albomicans.* For genes expected to have introns based on the *D. melanogaster* gene structure, we looked in the unannotated species for the approximate location of the *D. melanogaster* intron, and used known intron border consensus sequences and six-frame translation, implemented in EMBOSS SixPack (Madeira et al., 2019), to identify predicted intron borders and remove intronic sequences prior to the analyses below.

To study the gene duplication events that gave rise to *seminase* and its tandem gene duplicates (*CG10587* and *CG11037* in *D. melanogaster*), we identified the genes flanking these three genes and used them to identify the syntenic region of the other *Drosophila* genomes. We assumed conservation of gene order within this syntenic region in assigning orthologs for this gene family (Figure S1).

For all putative orthologs identified by bioinformatic methods, we verified that the ortholog was the reciprocal best BLAST hit to the expected SP network member of *D. melanogaster.* Inferred orthologs with a high degree of similarity, successful reciprocal best hits,

and a sequence that could be translated conceptually to produce a polypeptide without premature stops, were retained for study. In cases of duplicate genes (*seminase*, *CG1652* and *CG1656*), we also used gene order and synteny to confirm correct ortholog identification.

Sequence alignment

For each SP network protein, we used MUSCLE as implemented in MEGA 6.06 (Tamura et al., 2013) to align amino acid sequences, then visually checked and edited each alignment for accuracy. Amino acid alignments were then back-translated in MEGA to obtain the cDNA alignment.

Phylogenetic analysis

To infer a *Drosophila* consensus phylogeny based on all SP network proteins, we concatenated the amino acid alignments of all SP network proteins within each of the 22 species. We used PROML in Phylip (Felsenstein, 2005) to infer an unrooted maximum-likelihood phylogeny (with random input order, slow analysis, and all other default parameters). Gaps in the alignment were used in cases in which a protein was not present in a particular species. The resulting phylogeny matched published *Drosophila* phylogenies, expect for *D. virilis* and *D. mojavensis* (Drosophila 12 Genomes et al., 2007, Markow & O'Grady, 2005, Seetharam & Stuart, 2013). We then used this consensus tree for the PAML analyses, with species removed on a gene-by-gene basis as described below.

Detection of recombination

Because recombination within a gene sequence can impact the results of analyses to detect selection, we first used GARD with default parameters as implemented in DataMonkey 2.0 to check for evidence of recombination within each gene (Kosakovsky Pond et al., 2006, Weaver et al., 2018). Genes were partitioned at breakpoints evaluated as significant by the Kishino-Hasegawa test (p-value < 0.05 for both LH and RH; Table S2), and PAML was run on each gene segment separately. We performed PAML analyses on sequence alignments spanning two different ranges in the *Drosophila* phylogeny: the branch and branch-sites tests (see below) were run on species from the entire genus, while the sites test (see below) was run on species from only the *melanogaster* group. Thus, we generated a set of recombination breakpoints for each set of species (Table S2). For each species set, six SP network genes showed evidence of recombination, but the sets of genes that showed recombination differed between the two sets of species.

PAML analyses

For each protein, we used codeml of the PAML package to perform evolutionary analyses on protein-coding DNA sequence (Yang, 2007). To test for heterogeneity in the rate of each protein's evolution across the phylogeny, we utilized the PAML branch test, which uses a likelihood ratio test (LRT) to compare the "free ratio" model, allowing for different ω values for each branch, with model M0, which estimates a single ω for the whole phylogeny (Yang, 1998). For these tests, and for the branch-sites tests below, we used the consensus tree described above that covered the entire *Drosophila* phylogeny, but manually removed from it any species for which: a) an ortholog could not be identified, or b) an ortholog was identified, but it could not be confidently aligned due to ambiguity over an intron position or the end of the protein-coding

197 region. Table S1 shows the set of species used for the molecular evolutionary analyses for each
198 gene.

199 To test whether a subset of sites in a protein had evolved under recurrent positive
200 selection, we used LRTs to compare an evolutionary model (M8) that allows a class of sites to
201 have $\omega > 1$ to models M7 and M8a, which do not (Swanson et al., 2003, Yang et al., 2000). For
202 proteins for which model M8 was significantly preferred to models M7 and M8a, we used the
203 Bayes Empirical Bayes (BEB) approach to identify at the 0.9 confidence level the specific
204 residues that have evolved adaptively. These comparisons were done only for species within the
205 *melanogaster* group due to the possibility of synonymous site saturation if more divergent
206 species were included. To check for convergence in the 'free-ratio' and the sites models, we ran
207 codeml twice with the initial omega set at 0.4 and 2, respectively.

208 Finally, we performed the branch-sites test for positive selection (Zhang et al., 2005) to
209 identify classes of sites that had evolved adaptively along either of two specific lineages in the
210 phylogeny that we identified *a priori* because they represent likely evolutionary transitions in
211 key SP-related traits. First, we tested for sites under selection on the branch leading to the
212 *melanogaster* group of species (Fig. 1, branch 11), since this branch corresponds with the
213 inferred timing of when the SP receptor became expressed in the female reproductive tract and,
214 consequently, when females became sensitive to the non-receptivity effect caused by SP (Tsuda
215 et al., 2015). Second, we tested for sites under selection in the lineage that leads to and separates
216 *D. ananassae* and *D. bipectinata* from the rest of the *melanogaster* group species (Fig. 1, branch
217 15), because these species are known to have somewhat longer sperm (Joly & Bressac, 1994,
218 Markow, 1996, Pitnick et al., 1999). Although we inferred other important evolutionary
219 transitions in reproductive traits on the broader *Drosophila* phylogeny (Fig. 1), we limited our

220 branch-sites analyses to these two lineages because of the greater number of available sequenced
221 species in the *melanogaster* group.

222 In the branch-sites test, we used a LRT to compare a null model allowing for only
223 purifying and neutral selection on the focal branch with an alternative model allowing for a class
224 of sites to evolve under positive selection (Yang, 2007, Yang & Dos Reis, 2011, Zhang et al.,
225 2005). Recently, Venkat et al. (2018) found that this branch-sites test can have a high rate of
226 false positives driven by multinucleotide mutations within codons (i.e., mutations at adjacent
227 sites). To control for this issue, we implemented the tests in the Venkat model, a version of
228 PAML developed by these authors that runs the analysis after masking these sites. PAML
229 analyses were implemented using custom batch scripts for GNU parallel (Tange, 2018) and
230 PAML version 4.8a, or HyPhy version 2.5.1 (in the case of the Venkat model).

231

232 *Identification of* seminase *orthologs and paralogs*

233 We identified the predicted amino acid sequences for orthologs of seminase, CG11037
234 and CG10587 in *Drosophila* species using the methods described above. To confirm that calls of
235 orthology for seminase and its paralogs were accurate, we used Phylip's PROML program
236 (Felsenstein, 2005) to infer a maximum-likelihood rooted tree (using the single copies in *D.*
237 *pseudoobscura* and *D. persimillis* as the outgroup, and default PROML parameters). This was
238 consistent with the orthology assignments made using conserved gene order, except for *D.*
239 *ananassae* and *D. bipectinata*, which are likely confounded by their long branch length.

240

241 *Evaluation of gene expression*

242 *D. melanogaster, D. yakuba, D. ficusphila, D. bipectinata, D. annanassae, D.*
243 *pseudoobscura* and *D. willistoni* were raised in the lab as in Tsuda et al. (2015). We CO_2-
244 anesthetized 9-day-old flies of each species, separated them by sex, homogenized male or female
245 whole flies in TRIzol reagent, and purified RNA from samples and synthesized cDNA as
246 previously described (Gubala et al., 2017). We then used species-specific primers to amplify
247 *seminase*, *CG11037*, *CG10587* and *RpL32* as a housekeeping gene control. Genomic DNA was
248 used as a positive control for PCR reactions, and water was used in place of template in negative
249 control reactions.

Results and Discussion

SP network proteins are present in species outside of the melanogaster *group*

While SP orthologs have been found in species outside of the *melanogaster* group, only females of species within this group appear to show large-scale SP-mediated reproductive responses (Tsuda et al., 2015). One likely factor for this change is the evolution of *SPR* expression in the female reproductive tract in the last common ancestor of the *melanogaster* group (Tsuda et al., 2015). This evolutionary history raises the question of whether the remaining members of the SP network – all of which are critical for SP responses in *D. melanogaster* – are present outside of the group. We addressed this question bioinformatically by searching for intact orthologs across 22 *Drosophila* species with sequenced genomes.

Figure 2 shows that all SP network protein-coding gene orthologs are detectable in a large majority of the species surveyed, including those outside of the *melanogaster* group. For example, we found all currently known network proteins in *D. pseudoobscura* and *D. willistoni*, and all but one ortholog in *D. virilis*. To assess whether these orthologs were likely to function in reproduction outside of the *melanogaster* group, we examined publicly availably RNAseq data from male reproductive tracts in *D. pseudoobscura* (Yang et al., 2018) and proteomic data from male accessory glands in *D. mojavensis* (Kelleher et al., 2009). Transcripts of orthologs of male-derived network proteins were consistently enriched in (or entirely specific to) samples from whole males, male testes, and male carcasses in *D. pseudoobscura*, while showing either no or low expression in females or in male heads (Fig. S2). This pattern is consistent with expectations for reproductive proteins produced in the male accessory gland, with the "testis" expression likely reflecting contamination of testis dissections with accessory gland tissue. The genes encoding female-derived proteins showed broader expression patterns (Fig. S2), including

in whole females and whole males, but this pattern is consistent with their *D. melanogaster* orthologs, the expression of which is not limited to the female reproductive system (Brown et al., 2014). Predicted orthologs of the male-derived network proteins CG1652, CG1656, CG9997, CG17575, seminase, aquarius and antares were also identified in a proteomic analysis of the *D. mojavensis* accessory gland (Kelleher et al., 2009). Subsequent work showed that males of this species transfer transcripts of the *antares* ortholog to females during mating (Bono et al., 2011). Thus, RNAseq and proteomic data from two outgroup species are consistent with many SP network proteins functioning in reproduction outside of the *melanogaster* group.

It is likely that some SP network proteins function in other processes in certain species that impact their evolutionary trajectories. For example, SPR is expressed in both sexes outside of the reproductive tract (Tsuda et al., 2015, Yapici et al., 2008), and myoinhibitory peptides (MIPs) are known ligands in addition to SP (Kim et al., 2010, Poels et al., 2010, Yamanaka et al., 2010). SPR-MIP interactions outside of the reproductive tract affect sleep patterns in both sexes and remating propensity in females (Jang et al., 2017, Oh et al., 2014). Such interactions, in addition to the sexual selective pressures exerted by SP network-mediated interactions and reproductive phenotypes, have likely contributed to the evolution of SP network proteins in various *Drosophila* lineages. While most male-derived network proteins appear to have male-specific or heavily male-biased expression (in species for which expression data are available), the female-derived proteins show broader expression patterns. Understanding these proteins' non-reproductive functions will shed additional light on evolutionary forces that may have shaped them.

295 *SP network proteins demonstrate evolutionary rate heterogeneity across the* Drosophila
296 *phylogeny*

297 Because recombination within a gene can cause false positive results in the PAML
298 analyses, we first analyzed each set of orthologs using GARD (Kosakovsky Pond et al., 2006) to
299 identify high-confidence recombination sites, which were detected for six of the proteins (Table
300 S2). These six proteins were thus split into segments corresponding to the regions between
301 recombination breakpoints, which we analyzed independently.

302 To begin investigating these proteins' molecular evolution across the genus, we used
303 PAML model M0 to estimate a single d_N/d_S ratio (ω) for each full-length protein-coding
304 sequence across all species. For the six genes for which we detected evidence of recombination
305 (Table S2), we also estimated ω across the full length of each segment. We then performed the
306 branch test (Yang, 1998) to assess whether ω varied significantly across different branches of the
307 phylogeny. Most network proteins (and segments of proteins) had full-length ω estimates
308 between 0.2 and 0.3 across the full-genus tree (Table S3). Three proteins showed notably slower
309 evolutionary rates: CG17575 (ω = 0.07), a male-expressed cysteine-rich secretory protein
310 required for binding of SP to sperm (Ravi Ram & Wolfner, 2009); Esp (ω = 0.03), a female-
311 expressed, predicted sulfate membrane transporter also required for long-term fertility (Findlay
312 et al., 2014); and SPR (ω = 0.06), the female-expressed G-protein coupled receptor for SP
313 required for female post-mating changes including egg-laying, resistance to remating and release
314 of sperm from sperm-storage organs (Avila et al., 2015, Yapici et al., 2008). While these
315 proteins' slow rates of evolution could indicate that they play highly conserved roles in
316 reproduction, it is also possible that they have evolved adaptively at only a few sites or on a few
317 lineages (see below).

We next ran the "free ratio" model in which PAML estimates an ω value for each branch of the phylogeny. We found significant evidence of evolutionary rate heterogeneity for all but one network protein, intrepid (Table S3). Additionally, all proteins but intrepid had at least one phylogenetic branch for which ω was estimated to be > 1. While the branch test is not a rigorous test of positive selection acting on specific branches, the results indicate that the evolutionary rates of most SP network proteins have varied significantly across their evolutionary histories. In contrast, the constant, slow rate of evolution for intrepid implies that this protein has likely played a conserved and important role since the origin of the genus. Intrepid has undergone less functional characterization than other male-expressed male network proteins, so we cannot speculate further about its specific role(s) in reproduction.

Several SP network proteins have undergone recurrent positive selection at specific sites since the evolution of SPR expression in female reproductive tracts

To determine the extent to which positive selection has shaped the evolution of the SP network proteins, we used the PAML sites test to ask whether any protein had a particular subset of sites that had undergone recurrent positive selection. Because of the likelihood of synonymous site saturation over longer phylogenetic distances, we limited the sequences used in this analysis to those from the *melanogaster* group. This set of species also represents the likely extent of major SP/SPR-mediated post-mating responses, as only these species express SPR at high levels in the female reproductive tract and respond to injection of synthetic SP (Tsuda et al., 2015). Thus, our analyses identify proteins that might have evolved adaptively to further improve/refine network function in the past ~15 million years (Seetharam & Stuart, 2013).

340 The results of the sites analyses are shown in Table 1. Four proteins – CG9997, fra
341 mauro, CG1652 and hadley – show significant evidence for having a class of amino acid sites
342 that have evolved under recurrent positive selection across the *melanogaster* group of species.
343 Three other proteins (antares, intrepid and CG17575) each have a class of sites found to be under
344 positive selection in the Model M7/M8 comparison, but these results are no longer significant
345 when comparing Models M8 and M8a, suggesting that the class of more quickly evolving sites
346 identified for each protein in Model M8 may be evolving neutrally rather than under positive
347 selection.

348 The male-expressed network proteins that have evolved adaptively are functionally co-
349 dependent. CG9997, a serine protease homolog predicted to be catalytically inactive, must be
350 produced in the male accessory glands for CG1652, a C-type lectin, to be transferred to mated
351 females (Ravi Ram & Wolfner, 2009). Likewise, in the absence of CG1652, CG9997 is not
352 efficiently "processed" from its 45-kDa form to its 36-kDa form in mated females (Ravi Ram &
353 Wolfner, 2009, Singh et al., 2018). The loss of either protein prevents SP from accumulating on
354 stored sperm in females. Recent work has shown that both CG9997 and CG1652 also bind to
355 sperm, though their sperm-binding is detectable only in the hours after mating, while SP binding
356 lasts for several days (Peng et al., 2005, Singh et al., 2018). CG9997 and CG1652 also show
357 significant evidence of evolutionary rate covariation (Findlay et al., 2014). These results suggest
358 that pressure to maintain their functional interactions may be a factor driving the adaptive
359 evolution of CG9997 and CG1652, as has been observed for pairs of interacting reproductive
360 proteins in other systems (Clark et al., 2009, Grayson, 2015).

361 Other work on CG9997 is consistent with its adaptive evolution. Wong et al. (2008)
362 found evidence for recent positive selection acting on this gene by examining patterns of

363 polymorphism and divergence between populations of *D. melanogaster* and *D. simulans*. They
364 hypothesized that non-catalytically active serine protease homologs like CG9997 function as
365 agonists or antagonists for active proteases, while others have speculated that protease homologs
366 bind to other proteins or molecules in the female tract to slow their rate of digestion by active,
367 female-derived proteases (Laflamme & Wolfner, 2013). Under either scenario, protease
368 homologs like CG9997 may need to continually coevolve with their interacting partners,
369 providing the impetus for the recurrent, adaptive evolution detected here. Additionally,
370 knockdown of *CG9997* diminishes male sperm competitive ability (Castillo & Moyle, 2014),
371 suggesting another potential factor in its adaptive evolution.

372 Less functional information exists for the adaptively evolving, female-expressed proteins.
373 Both fra mauro and hadley were identified in a screen for female-expressed proteins that
374 coevolved with a male-expressed SP network protein; in each case, the coevolutionary signal
375 was with CG17575 (Findlay et al., 2014). RNAi knockdown of either gene reduced female
376 fertility, though knockdown females could receive SP and store it properly on sperm (Findlay et
377 al., 2014). These data suggested that the proteins could be involved in maintaining the female
378 long-term response to SP, though *fra mauro* knockdown females also showed a significant
379 fertility defect in the 24 hrs after mating (Findlay et al., 2014). The fra mauro protein encodes a
380 predicted neprilysin protease, which may coevolve with its as yet unknown molecular targets or
381 antagonists (Laflamme & Wolfner, 2013). As noted above, functional domains have not been
382 identified for the hadley protein, so it is difficult to speculate on potential forces driving its
383 adaptive evolution.

384 Notably, several proteins in the SP network showed no evidence of recurrent adaptive
385 evolution within the *melanogaster* group, while others had subsets of sites with evolutionary

386 rates that were elevated, but approximated neutrality. These data suggest that while some
387 network proteins may contain regions that are under relaxed constraint, much of the functionality
388 and interdependence of the network might have already existed at the origin of the *melanogaster*
389 group.
390
391 *Several network proteins underwent adaptive evolution on specific lineages correlating with*
392 *changes in reproductive phenotypes*
393 While the PAML sites test described above detects recurrent adaptive evolution, protein
394 networks can also be shaped by bursts of episodic positive selection acting on specific
395 phylogenetic lineages. One important evolutionary transition for the SP network occurred at the
396 base of the *melanogaster* group, when *SPR* evolved expression in the lower female reproductive
397 tract (Tsuda et al., 2015). This change likely created (or exacerbated) a selective pressure for
398 higher SP levels in this location, as prolonged SP-SPR signaling could promote continued egg
399 production and prolong female non-receptivity to re-mating. Because a primary purpose of the
400 male-expressed SP network proteins in *D. melanogaster* is to bind SP to sperm to prolong the
401 post-mating response, we hypothesized that some of these proteins might have experienced a
402 burst of adaptive evolution on the same phylogenetic branch on which female reproductive *SPR*
403 expression is inferred to have evolved. Likewise, the increase in *SPR* expression in females
404 could have created a selective pressure for other female-expressed members of the network to
405 evolve. To test these ideas, we used the Venkat model, a modified PAML branch-sites test
406 (Venkat et al., 2018, Zhang et al., 2005), to ask whether any network protein had a subset of sites
407 under selection on the branch leading to the *melanogaster* group (i.e., branch 11 in Fig. 1).

Table 2 shows the results of these tests. Two proteins show marginal evidence for adaptive evolution on branch 11: CG1656 and SPR. As originally formulated (Zhang et al., 2005), the LRT for the branch-sites test follows a null distribution described as an equal mixture of point mass 0 and a chi-square distribution with 1 degree of freedom (df). Under this null distribution, the test statistic corresponding with a *p*-value of 0.05 is 2.71, a value exceeded by each gene. However, the test is typically conducted conservatively (Venkat et al., 2018, Zhang et al., 2005), following only a chi-square distribution with 1 df. The p-values listed in Table 2 are calculated based on this latter distribution, and they are marginally significant ($0.05 < p < 0.1$) for CG1656 and SPR.

The potential adaptive evolution of sites in the SPR protein along branch 11 is interesting, because this lineage also represents the time during which the protein became expressed in the female reproductive tract (Tsuda et al., 2015). Thus, it is possible that the SPR gene underwent both regulatory and protein-coding adaptations that altered how the female post-mating response is controlled. The other protein that potentially underwent adaptive evolution along this lineage is the predicted C-type lectin CG1656, which functions similarly to its recurrently rapidly evolving paralog described above, CG1652. Both lectins are required for SP's long-term binding to stored sperm, and both proteins themselves bind sperm temporarily in the hours after mating (Singh et al., 2018). Given the potential selective pressure to bind more SP to stored sperm in female tracts expressing *SPR*, it is possible that the adaptive evolution of CG1656 on this key phylogenetic branch could have helped to improve the efficiency of SP's binding to sperm. This idea could be tested in future experiments by either identifying and mutating the residues likely to have changed along branch 11 and/or by substituting an outgroup

ortholog of CG1656 (and potentially its duplicate, CG1652) into *D. melanogaster* and examining the effects on SP's sperm binding and on the female long-term post-mating response.

Prior work demonstrated that SP binds to the full length of *D. melanogaster* sperm (Peng et al., 2005, Ravi Ram & Wolfner, 2009, Singh et al., 2018). Indeed, the ability of SP (and potentially other molecules) to bind sperm and then influence post-mating responses is one hypothesis for why sperm tails have evolved to be so long in many *Drosophila* species. Within the *melanogaster* group species that experience SP-mediated post-mating responses, one notable change in reproductive physiology is that the sperm of *D. ananassae* and its closely related species are considerably longer than those of *D. melanogaster* (*D. ananassae* sperm length: 3.3 mm; *D. melanogaster* and other *melanogaster* group species sperm length: just under 2 mm (Pitnick et al., 1999, Joly & Bressac, 1994, Markow, 1996)). We thus infer that a major (>50%) increase in sperm length occurred on the branch of the phylogeny leading to *D. ananassae* and its close sister species *D. bipectinata* (branch 15 in Fig. 1).

To test for whether any SP network proteins experienced adaptive evolution concurrent with this change in sperm length, we again used the modified branch-sites test. Two network proteins, antares and CG17575, show evidence of positive selection acting on specific sites on the lineage leading to *D. ananassae* and *D. bipectinata* (Table 3). Antares' signal of selection is significant under both null distributions described above, while CG17575's signal is significant under the mixed null distribution and marginally significant ($p = 0.0504$) under the conservative test. In addition to facilitating SP's long-term binding to sperm, antares also binds to sperm itself for a shorter period (Findlay et al., 2014, Singh et al., 2018). Thus, antares might have evolved adaptively to facilitate greater or more efficient binding of either itself or SP to sperm as sperm tails lengthened. Interestingly, the antares ortholog in outgroup species *D. mojavensis* and *D.*

453 *arizonae* was also found to evolve under diversifying selection (Bono et al., 2015), even though
454 *D. mojavensis* does not have a currently detectable SP ortholog (Tsuda et al., 2015) (Fig. 2).
455 Heterospecific matings between these species fail due to post-mating, pre-zygotic isolating
456 barriers, which include problems with sperm storage in the female reproductive tract (Kelleher &
457 Markow, 2007). It is thus possible that antares plays an essential role in binding molecules to
458 sperm and/or facilitating sperm storage, and that the male reproductive activity of antares has
459 been refined by different selective pressures in different lineages.

460 CG17575 is a male-expressed, cysteine-rich secretory protein required for SP and other
461 sperm-binding network proteins to localize from the female uterus, where seminal proteins and
462 sperm are first deposited, into the seminal receptacle (SR), the primary site of sperm storage in
463 *D. melanogaster* (Ravi Ram & Wolfner, 2009, Singh et al., 2018). Since CG17575 does not
464 itself bind sperm (Singh et al., 2018), further details of how CG17575 provides for proper
465 localization of other seminal proteins to the seminal receptacle are needed before we can
466 speculate on the selective forces that might have contributed to its evolution in this lineage.

467 The branch-sites tests for branches 11 and 15 reported above were conducted using full-
468 length gene sequences, since the test has limited power. However, we repeated this analysis on
469 all segments of the six genes for which recombination was detected. These results (Table S4)
470 found marginal evidence for selection for antares on branch 11 and for a segment of CG1652 on
471 branch 15. CG1656 was not among the genes for which recombination was detected (Table S2),
472 so its results above are unaltered.

473

474 *Seminase gene duplicates retain male-specific expression patterns across* melanogaster *group*
475 *species*

476 In addition to CG17575, the male-expressed serine protease seminase is required for the
477 localization of SP and other male-expressed proteins to the SR after mating (LaFlamme et al.,
478 2012, Singh et al., 2018). Seminase arose through gene duplication in the lineage leading to the
479 *melanogaster* group of flies. The genomes of *D. pseudoobscura* and other outgroup species
480 contain only one detectable copy of the gene, but in *D. melanogaster* and its fellow *melanogaster*
481 group members, there are three tandemly arrayed, intron-containing copies, suggesting two
482 distinct DNA-based duplication events (Figure S1). The other genes are *CG10587* and
483 *CG11037*. Like *seminase*, both are expressed specifically in the male accessory gland in *D.*
484 *melanogaster* (Brown et al., 2014, Leader et al., 2018). While we detected no recurrent or
485 episodic positive selection acting on seminase after these duplications (Tables 1-2), we were
486 curious whether it or its paralogs might have evolved different expression patterns (and, thus,
487 potential functions) after duplication. We thus performed RT-PCR to amplify each paralog from
488 cDNA isolated from males or females of a variety of species from the *melanogaster* group. We
489 also assessed the expression of the single-copy parent gene from *D. pseudoobscura* and *D.*
490 *willistoni*. Our results (Figure 3) show that both the single-copy genes from the outgroup
491 species, as well as all of paralogs from all *melanogaster* group species tested, are expressed
492 specifically in adult males. This result is consistent with the ancestral single copy of seminase
493 also functioning in male reproduction (and potentially with other SP network proteins).
494 Given that seminase itself has additional reproductive functions beyond its role in the SP
495 network (LaFlamme et al., 2012), it is possible that the paralogs have sub- or neo-functionalized
496 to have unique roles, in spite of their conserved expression patterns. Future studies should
497 evaluate how the paralogs contribute to reproduction, which may suggest possible evolutionary
498 forces that affected their evolution after the gene duplication events.

Conclusions

Sex peptide is directly responsible for major changes in female post-mating behavior and physiology and is therefore one of the best characterized reproductive proteins to date. SP-mediated responses appear to have arisen specifically in the *melanogaster* group of *Drosophila*, and they manifest in full only with the help of a suite of male- and female-derived proteins, the SP network. We have shown that these proteins are present and expressed in species outside of the *melanogaster* group, suggesting they likely function in reproduction in these species and that they did so in a common ancestor. Within the *melanogaster* group, several network proteins (CG9997, CG1652, fra mauro, and hadley) have experienced recurrent positive selection, suggesting that continued, adaptive evolution refined SP network function. A non-overlapping set of proteins, including CG1656, SPR, antares, and CG17575, showed some evidence of bursts of adaptive evolution on specific phylogenetic lineages corresponding with major changes in SP network reproductive phenotypes. Taken together, these data suggest that SP network proteins may have interacted to affect reproduction before the evolution of major SP-mediated changes in the *melanogaster* group. However, once SPR became expressed at high levels in the female reproductive tract in the common ancestor of this group (Tsuda et al., 2015), a combination of both quick bursts of adaptation on specific lineages and recurrent changes at specific protein sites helped the network evolve into the present form observed in *D. melanogaster*. This study demonstrates how changes in both regulatory and protein-coding regions can affect the evolution of protein networks and motivates future functional studies of the SP network proteins in *Drosophila* species both within and outside of the *melanogaster* group.

Acknowledgements

We thank L. Moyle, M. Hahn, A. Venkat and M. Wu for assistance in implementing the Venkat model, D. Thakral for help developing batch PAML scripts, and K. Ober and M. Wolfner for helpful discussions and feedback on the manuscript. This work was supported by NSF CAREER Award 1652013 and a summer research fellowship generously provided by Dr. William Crowley. The authors report no conflicts of interest.

References

Alpern, J. H. M., Asselin, M. M. & Moehring, A. J. 2019. Identification of a novel sperm class and its role in fertilization in Drosophila. *J Evol Biol* **32**: 259-266.

Apger-McGlaughon, J. & Wolfner, M. F. 2013. Post-mating change in excretion by mated Drosophila melanogaster females is a long-term response that depends on sex peptide and sperm. *J Insect Physiol* **59**: 1024-30.

Avila, F. W., Mattei, A. L. & Wolfner, M. F. 2015. Sex peptide receptor is required for the release of stored sperm by mated Drosophila melanogaster females. *Journal of Insect Physiology* **76**: 1-6.

Avila, F. W., Ravi Ram, K., Bloch Qazi, M. C. & Wolfner, M. F. 2010. Sex peptide is required for the efficient release of stored sperm in mated Drosophila females. *Genetics* **186**: 595-600.

Bono, J. M., Matzkin, L. M., Hoang, K. & Brandsmeier, L. 2015. Molecular evolution of candidate genes involved in post-mating-prezygotic reproductive isolation. *J Evol Biol* **28**: 403-14.

Bono, J. M., Matzkin, L. M., Kelleher, E. S. & Markow, T. A. 2011. Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated Drosophila mojavensis females. *Proc Natl Acad Sci U S A* **108**: 7878-83.

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., et al. 2014. Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**: 393-9.

Carvalho, G. B., Kapahi, P., Anderson, D. J. & Benzer, S. 2006. Allocrine modulation of feeding behavior by the Sex Peptide of Drosophila. *Curr Biol* **16**: 692-6.

Castillo, D. M. & Moyle, L. C. 2014. Intraspecific sperm competition genes enforce post-mating species barriers in Drosophila. *Proc Biol Sci* **281**: 20142050.

Chapman, T., Bangham, J., Vinti, G., Seifried, B., Lung, O., Wolfner, M. F., Smith, H. K. & 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**: 9923-8.

Chen, P. S., Stumm-Zollinger, E., Aigaki, T., Balmer, J., Bienz, M. & Bohlen, P. 1988. A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. *Cell* **54**: 291-8.

Chen, Z. X., Sturgill, D., Qu, J., Jiang, H., Park, S., Boley, N., Suzuki, A. M., Fletcher, A. R., Plachetzki, D. C., FitzGerald, P. C., et al. 2014. Comparative validation of the D. melanogaster modENCODE transcriptome annotation. *Genome Res* **24**: 1209-23.

Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F. & Swanson, W. J. 2009. Coevolution of interacting fertilization proteins. *PLoS Genetics* **5**: e1000570.

Drosophila 12 Genomes, C., Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203-18.

Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. pp. Distributed by Author, Department of Genome Sciences, University of Washington, Seattle.

Findlay, G. D., Sitnik, J. L., Wang, W., Aquadro, C. F., Clark, N. L. & Wolfner, M. F. 2014. Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for Drosophila melanogaster Female Post-Mating Responses. *PLoS Genetics* **10**: e1004108.

Grayson, P. 2015. Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm-egg binding partners. *R Soc Open Sci* **2**: 150296.

Gubala, A. M., Schmitz, J. F., Kearns, M. J., Vinh, T. T., Bornberg-Bauer, E., Wolfner, M. F. & Findlay, G. D. 2017. The Goddard and Saturn Genes Are Essential for Drosophila Male Fertility and May Have Arisen De Novo. *Mol Biol Evol* **34**: 1066-1082.

Hasemeyer, M., Yapici, N., Heberlein, U. & Dickson, B. J. 2009. Sensory neurons in the Drosophila genital tract regulate female reproductive behavior. *Neuron* **61**: 511-8.

Isaac, R. E., Li, C., Leedale, A. E. & Shirras, A. D. 2010. Drosophila male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. *Proc Biol Sci* **277**: 65-70.

Jang, Y. H., Chae, H. S. & Kim, Y. J. 2017. Female-specific myoinhibitory peptide neurons regulate mating receptivity in Drosophila melanogaster. *Nat Commun* **8**: 1630.

Joly, D. & Bressac, C. 1994. Sperm Length in Drososphilidae (Diptera): Estimation by Testis and Receptacle Lengths. *International Journal of Insect Morphology & Embryology* **23**: 85-92.

Kelleher, E. S. & Markow, T. A. 2007. Reproductive tract interactions contribute to isolation in Drosophila. *Fly (Austin)* **1**: 33-7.

Kelleher, E. S., Watts, T. D., LaFlamme, B. A., Haynes, P. A. & Markow, T. A. 2009. Proteomic analysis of Drosophila mojavensis male accessory glands suggests novel classes of seminal fluid proteins. *Insect Biochem Mol Biol* **39**: 366-71.

Kim, Y. J., Bartalska, K., Audsley, N., Yamanaka, N., Yapici, N., Lee, J. Y., Kim, Y. C., Markovic, M., Isaac, E., Tanaka, Y., et al. 2010. MIPs are ancestral ligands for the sex peptide receptor. *Proc Natl Acad Sci U S A* **107**: 6520-5.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096-8.

LaFlamme, B. A., Ravi Ram, K. & Wolfner, M. F. 2012. The Drosophila melanogaster seminal fluid protease "Seminase" regulates proteolytic and post-mating reproductive processes. *PLoS Genetics* **8**: 30-32.

Laflamme, B. A. & Wolfner, M. F. 2013. Identification and function of proteolysis regulators in seminal fluid. *Mol Reprod Dev* **80**: 80-101.

Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A. & Dow, J. A. T. 2018. FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res* **46**: D809-D815.

Liu, H. & Kubli, E. 2003. Sex-peptide is the molecular basis of the sperm effect in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **100**: 9929-33.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**: W636-W641.

Markow, T. A. (1996) Evolution of Drosophila mating systems. In: *Evolutionary Biology*, Vol. 29. pp. 73-106.

Markow, T. A. & O'Grady, P. M. 2005. Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. *Annual Review of Genetics* **39**: 263-291.

McDonough, C. E., Whittington, E., Pitnick, S. & Dorus, S. 2016. Proteomics of reproductive systems: Towards a molecular understanding of postmating, prezygotic reproductive barriers. *J Proteomics* **135**: 26-37.

618 Oh, Y., Yoon, S. E., Zhang, Q., Chae, H. S., Daubnerova, I., Shafer, O. T., Choe, J. & Kim, Y. J.
619 2014. A homeostatic sleep-stabilizing pathway in Drosophila composed of the sex
620 peptide receptor and its ligand, the myoinhibitory peptide. *PLoS Biol* **12**: e1001974.
621 Peng, J., Zipperlen, P. & Kubli, E. 2005. Drosophila sex-peptide stimulates female innate
622 immune system after mating via the toll and Imd pathways. *Current Biology* **15**: 1690-
623 1694.
624 Pitnick, S., Markow, T. A. & Spicer, G. S. 1999. Evolution of Multiple Kinds of Female Sperm-
625 Storage Organis in Drosophila. *Evolution* **53**: 1804-1822.
626 Poels, J., Van Loy, T., Vandersmissen, H. P., Van Hiel, B., Van Soest, S., Nachman, R. J. &
627 Vanden Broeck, J. 2010. Myoinhibiting peptides are the ancestral ligands of the
628 promiscuous Drosophila sex peptide receptor. *Cell Mol Life Sci* **67**: 3511-22.
629 Ravi Ram, K. & Wolfner, M. F. 2007. Sustained post-mating response in Drosophila
630 melanogaster requires multiple seminal fluid proteins. *PLoS Genetics* **3**: 2428-2438.
631 Ravi Ram, K. & Wolfner, M. F. 2009. A network of interactions among seminal proteins
632 underlies the long-term postmating response in Drosophila. *Proceedings of the National*
633 *Academy of Sciences of the United States of America* **106**: 15384-15389.
634 Rezaval, C., Pavlou, H. J., Dornan, A. J., Chan, Y. B., Kravitz, E. A. & Goodwin, S. F. 2012.
635 Neural circuitry underlying Drosophila female postmating behavioral responses. *Curr*
636 *Biol* **22**: 1155-65.
637 Schnakenberg, S. L., Matias, W. R. & Siegal, M. L. 2011. Sperm-storage defects and live birth in
638 Drosophila females lacking spermathecal secretory cells. *PLoS Biol* **9**: e1001192.
639 Seetharam, A. S. & Stuart, G. W. 2013. Whole genome phylogeny for 21 Drosophila species
640 using predicted 2b-RAD fragments. *PeerJ* **1**: e226.
641 Singh, A., Buehner, N. A., Lin, H., Baranowski, K. J., Findlay, G. D. & Wolfner, M. F. 2018.
642 Long-term interaction between Drosophila sperm and sex peptide is mediated by other
643 seminal proteins that bind only transiently to sperm. *Insect Biochem Mol Biol* **102**: 43-51.
644 Singh, S. R., Singh, B. N. & Hoenigsberg, H. F. 2002. Female remating, sperm competition and
645 sexual selection in Drosophila. *Genetics and Molecular Research* **1**: 178-215.
646 Sitnik, J. L., Gligorov, D., Maeda, R. K., Karch, F. & Wolfner, M. F. 2016. The Female Post-
647 Mating Response Requires Genes Expressed in the Secondary Cells of the Male
648 Accessory Gland in Drosophila melanogaster. *Genetics* **202**: 1029-41.
649 Snook, R., Markow, T. & Karr, T. 1994. Functional nonequivalence of sperm in *Drosophila*
650 *pseudoobscura*. *Proceedings of the National Academy of Sciences of the United States of*
651 *America* **91**: 11222-11226.
652 Snook, R. R. 1995. The Evolution of Sperm Polymorphism in the Drosophila Obscura Group.
653 957.
654 Soller, M., Bownes, M. & Kubli, E. 1999. Control of oocyte maturation in sexually mature
655 Drosophila females. *Dev Biol* **208**: 337-51.
656 Swanson, W. J., Nielsen, R. & Yang, Q. 2003. Pervasive adaptive evolution in mammalian
657 fertilization proteins. *Molecular biology and evolution* **20**: 18-20.
658 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular
659 evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-
660 2729.
661 Tange, O. 2018. *GNU Parallel 2018*.
662 Tsuda, M., Peyre, J. B., Asano, T. & Aigaki, T. 2015. Visualizing Molecular Functions and
663 Cross-Species Activity of Sex-Peptide in Drosophila. *Genetics*: 1-10.

Venkat, A., Hahn, M. W. & Thornton, J. W. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nat Ecol Evol*.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V. & Kosakovsky Pond, S. L. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* **35**: 773-777.

Wilburn, D. B. & Swanson, W. J. 2016. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J Proteomics* **135**: 12-25.

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.

Yamanaka, N., Hua, Y. J., Roller, L., Spalovska-Valachova, I., Mizoguchi, A., Kataoka, H. & Tanaka, Y. 2010. Bombyx prothoracicostatic peptides activate the sex peptide receptor to regulate ecdysteroid biosynthesis. *Proc Natl Acad Sci U S A* **107**: 2060-5.

Yang, C. H., Rumpf, S., Xiang, Y., Gordon, M. D., Song, W., Jan, L. Y. & Jan, Y. N. 2009. Control of the postmating behavioral switch in Drosophila females by internal sensory neurons. *Neuron* **61**: 519-26.

Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K. & Oliver, B. 2018. Re-annotation of eight Drosophila genomes. *Life Sci Alliance* **1**: e201800156.

Yang, Z. 1998. Likelihood Ratio Tests for Detecting Positive Selection and Application to Primate Lysozyme Evolution. *Molecular Biology and Evolution* **15**: 568-573.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.

Yang, Z. & Dos Reis, M. 2011. Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution* **28**: 1217-1228.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. 2000. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular biology and evolution* **19**: 49-57.

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

Zhang, J., Nielsen, R. & Yang, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* **22**: 2472-2479.

Figure Legends

Figure 1. Phylogeny of *Drosophila* species examined in this study. The gray box indicates the *melanogaster* group. Each branch is numbered for reference in the main text. Key changes in reproductive tracts are indicated by letters a-d and are based on an examination of the literature (references cited in main text). The PAML branch-sites tests (see Results) were conducted on branches 11 and 15. Branch lengths are not proportional to evolutionary distances.

Figure 2. Bioinformatic identification of SP network proteins across 22 *Drosophila* species. Identified orthologs that were also reciprocal best BLAST hits are noted with a + sign, while a – sign indicates no ortholog could be identified.

Figure 3. RT-PCR on *seminase* and paralogs shows conserved, male-biased expression after duplication. Orthologs of *seminase*, *CG10586* and *CG11037* show male-specific expression in various *melanogaster* group species, though the level of expression between paralogs and species is somewhat variable. The single-copy parent gene in *D. pseudoobscura* and *D. willistoni* is also expressed in a male-specific manner.

Table 1. PAML sites tests for positive selection acting on SP network genes. Asterisks indicate genes for which recombination was detected, which were split into numbered segments as indicated. Specific codons that were inferred to be under selection by PAML's Bayes Empirical Bayes (BEB) analysis with Pr > 0.9 are shown for genes or segments for which positive selection was detected (i.e., in which model M8 was a significantly better fit to the data than models M7 and M8a). Amino acid site positions and identities refer to the *D. melanogaster* protein sequence.

Gene	Segment Sites	M0 ω estimate	M7 vs. M8		M8 vs. M8a		% sites in ω > 1	Sites with BEB Pr > 0.90
			2*ΔlnL	*p*-value	2*ΔlnL	*p*-value		
aqrs		0.21	3.950	0.1388	0.773	0.3792		
antr		0.22	10.194	0.0061	1.935	0.1643		
intr		0.21	6.719	0.0348	0.547	0.4597		
CG9997		**0.25**	**21.001**	**2.75E-05**	**5.974**	**0.0145**	**10%**	**152S**
CG1652*		0.19	0.000	1.0000	0	1.0000		
CG1652_1	1-70	0.08	5.877	0.0529	0.797	0.3721		
CG1652_2	71-106	0.01	0	1.0000	0	1.0000		
CG1652_3	107-144	0.01	4.822	0.0897	0	1.0000		
CG1652_4	145-187	0.03	0	1.0000	0	1.0000		
CG1652_5	**188-322**	**0.67**	**6.953**	**0.0309**	**4.398**	**0.0360**	**38%**	**233P, 234G, 250V**
CG1656*		0.26	3.829	0.1474	3.221	0.0727		
CG1656_1	1-69	0.19	0.816	0.6651	0	1.0000		
CG1656_2	70-328	0.08	2.981	0.2252	0	1.0000		
CG17575*		0.06	24.088	5.88E-06	2.154	0.1422		
CG17575_1	1-139	0.03	0	1.0000	0	1.0000		
CG17575_2	140-298	0.08	16.612	0.0002	1.373	0.2413		
SP		0.22	1.463	0.4813	0	1.0000		
SPR		0.04	1.817	0.4031	8.828	0.0030		
Esp		0.03	0	1.0000	2.096	0.1477		
frma*		**0.24**	**15.986**	**0.0003**	**4.812**	**0.0283**	**7%**	**392A**
frma_1	**1-347**	**0.26**	**13.421**	**0.0012**	**4.515**	**0.0336**	**11%**	**none**
frma_2	**348-611**	**0.21**	**8.188**	**0.0167**	**6.106**	**0.0135**	**2%**	**392A**
hdly*		**0.29**	**36.156**	**1.41E-08**	**31.463**	**2.03E-08**	**6%**	**173V, 201I, 229S, 239I, 304A**
hdly_1	**1-364**	**0.30**	**42.194**	**0.0000**	**37.383**	**0.0000**	**7%**	**173V, 201I, 229S, 239I, 304A**
hdly_2	365-445	0.12	0.494	0.7813	0.463	0.4964		
sems		0.20	0	1.0000	0	1.0000		
sems_1	1-93	0.28	0.640	0.7262	0.409	0.5227		
sems_2	94-275	0.18	0.899	0.6379	0.320	0.5715		

726 **Table 2. Venkat model branch-sites tests for positive selection acting on specific sites of SP**
727 **network proteins on the lineage (Branch 11) leading to the *melanogaster* group of**
728 ***Drosophila.*** *P*-values are calculated based on a χ_1^2 distribution. Asterisks indicate likelihood
729 ratio test statistics that reach the $p < 0.05$ significance threshold for a null distribution derived
730 from a 50:50 ratio of point mass 0 and the χ_1^2 distribution.
731

Gene	Whole-Gene ω estimate	Venkat Model	
		2*ΔlnL	*p*-value
antr	298.77	2.464	0.116
aqrs	1.00	0	1.000
CG1652	7.62	1.428	0.232
CG1656	122.47	3.238*	0.072
CG9997	1.04	0	1.000
CG17575	424.52	2.425	0.119
Esp	1.08	0	1.000
frma	1.08	0.034	0.854
hdly	1.08	0	1.000
intr	1.00	0.039	0.843
sems	1.00	0	1.000
SP	1.00	0	1.000
SPR	9999.99	3.064*	0.080

732
733
734

735 **Table 3. Venkat model branch-sites tests for positive selection acting on specific sites of SP**
736 **network proteins on a lineage (Branch 15) corresponding to increased sperm length in *D.***
737 ***ananassae.*** *P*-values are calculated based on a χ_1^2 distribution. Asterisks indicate likelihood
738 ratio test statistics that reach the $p < 0.05$ significance threshold for a null distribution derived
739 from a 50:50 ratio of point mass 0 and the χ_1^2 distribution.
740

Gene	Whole-Gene ω estimate	Venkat Model	
		2*ΔlnL	*p*-value
antr	**6.19**	**3.948**	**0.047**
aqrs	2.53	0.388	0.533
CG1652	1.05	0.006	0.938
CG1656	1.04	0	1.000
CG9997	1.05	0.022	0.882
CG17575	422.98	3.827*	0.050
Esp	1.08	0	1.000
frma	1.04	0.012	0.913
hdly	1.64	0.008	0.929
intr	1.02	0	1.000
sems	403.51	0.962	0.327
SP	1.04	0.002	0.963
SPR	1.08	0	1.000

741
742

743 **Supplemental Materials**
744
745 **Table S1. Orthologs used for each gene in PAML analyses**. Some orthologs that were
746 identified in Table 2 were excluded from PAML analysis due to unresolved intron borders and/or
747 poor alignment quality. Only species above the dotted line (the *melanogaster* group) were
748 analyzed in the sites tests.
749
750 **Table S2. GARD results showing inferred recombination breakpoints.** Breakpoint positions
751 refer to nucleotide positions in the alignment files used. However, since alignments include
752 gaps, these positions do not necessarily have a 3:1 correspondence with the *D. melanogaster*
753 amino acid positions reported in Tables 1-3. The first table shows recombination breakpoints
754 detected for aligned sequences from the entire *Drosophila* genus, which were used for the branch
755 and branch-sites tests. The second table shows recombination breakpoints detected for aligned
756 sequences from only the *melanogaster* group, which were used for the sites tests.
757
758 **Figure S1. RNAseq data from *D. pseudoobscura* support reproductive functions for SP**
759 **network proteins in a species that lacks full-scale SP responses.** A) *D. pseudoobscura*
760 expression patterns for each member of the SP network. Dark shading indicates high expression
761 levels, stripes indicate low (but detectable) expression, and no shading indicates no expression
762 detected in a given sample. Male-derived network proteins show male-biased or male-specific
763 expression, consistent with reproductive functions. B) Examples of *D. pseudoobscura*
764 expression data for several SP network genes; shading in part (A) is based on these data. The
765 RNAseq data were accessed via FlyBase and generated by Yang et al. (Yang et al., 2018).
766
767 **Table S3. Branch tests for rate heterogeneity.** Partitions were implemented in PAML
768 analyses if they were significant in both the LH and RH tests.
769
770 **Table S4. Venkat model branch-sites tests for positive selection acting on specific sites of**
771 **SP network proteins detected by GARD to have multiple recombination segments.** The
772 table shows results for both branch 11 and branch 15 tests. Only the six genes for which
773 recombination was detected in the relevant trees are shown in the table. *P*-values are calculated
774 based on a χ_1^2 distribution. Asterisks indicate likelihood ratio test statistics that reach the $p <$
775 0.05 significance threshold for a null distribution derived from a 50:50 ratio of point mass 0 and
776 the χ_1^2 distribution.

Fig. 1

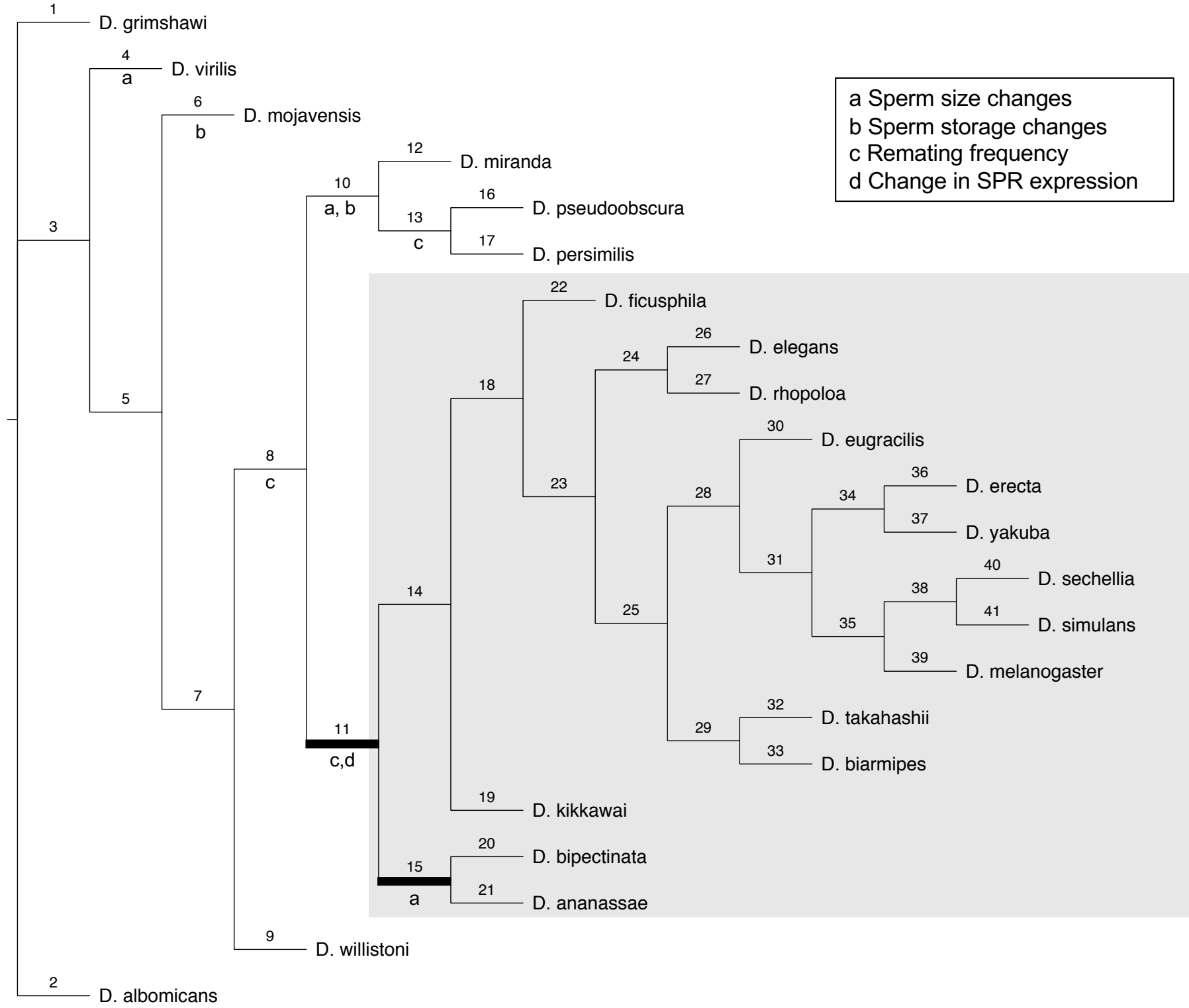
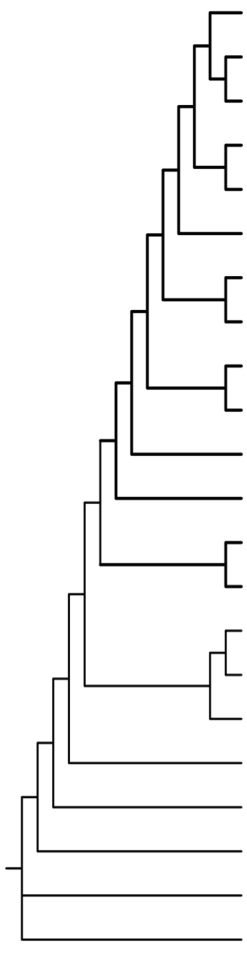
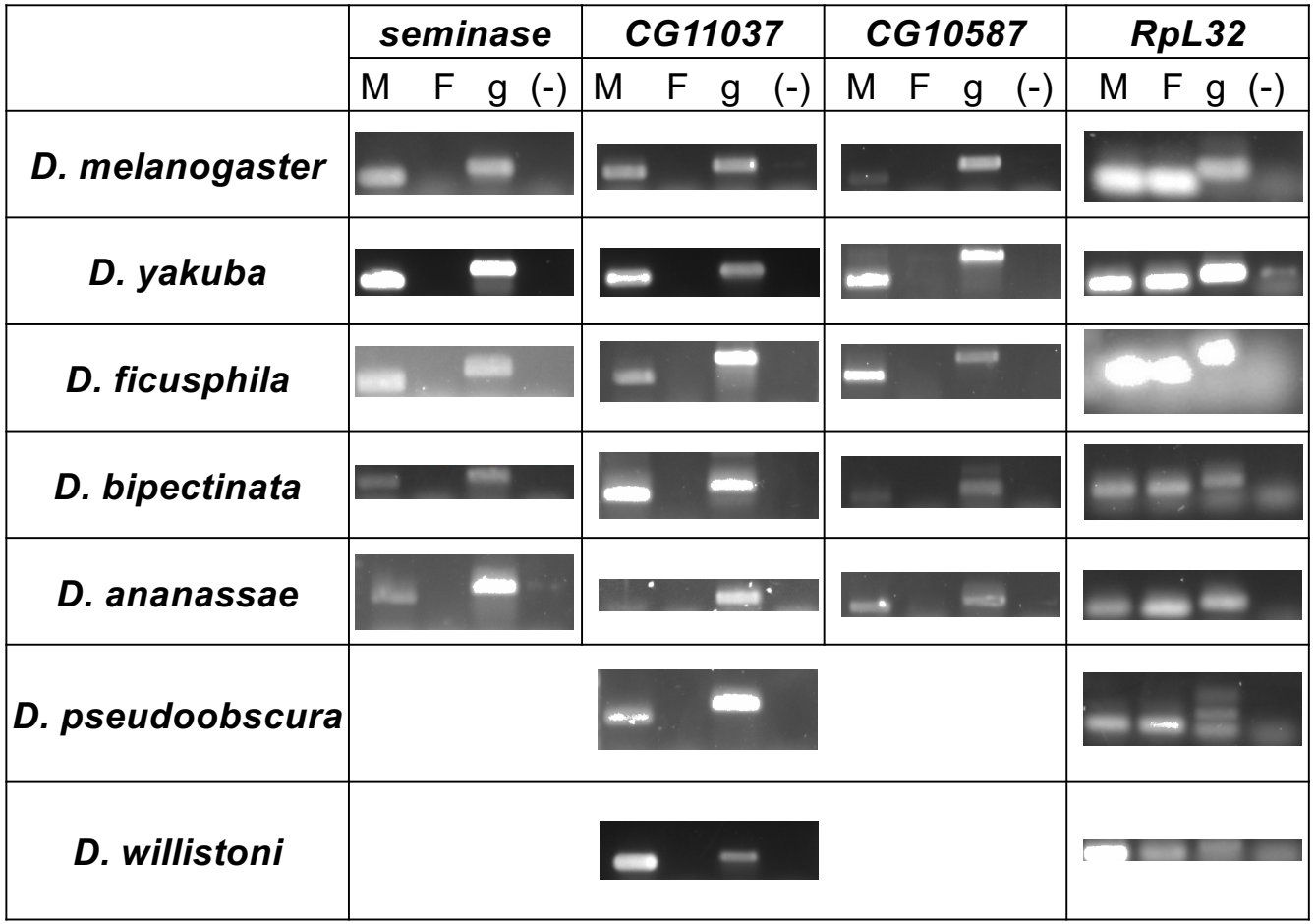

1
D. grimshawi
4
a
D. virilis
6
b
D. mojavensis
12
D. miranda
10
a, b
16
D. pseudoobscura
13
c
17
D. persimilis
3
22
D. ficusphila
26
D. elegans
24
27
D. rhopoloa
18
5
30
D. eugracilis
8
c
23
36
D. erecta
28
34
37
D. yakuba
31
40
D. sechellia
14
38
41
25
D. simulans
35
39
D. melanogaster
7
32
D. takahashii
11
c,d
29
33
D. biarmipes
19
D. kikkawai
20
D. bipectinata
15
a
21
D. ananassae
9
D. willistoni
2
D. albomicans
a Sperm size changes
b Sperm storage changes
c Remating frequency
d Change in SPR expression

Fig. 2



	intrepid	CG1652	CG1656	aquarius	antares	CG9997	CG17575	seminase	sex peptide	fra mauro	hadley	Esp	SPR
D.melanogaster	+	+	+	+	+	+	+	+	+	+	+	+	+
D.sechellia	+	+	+	+	+	+	+	+	+	+	+	+	+
D.simulans	+	+	+	+	+	+	+	+	+	+	+	+	+
D.erecta	+	+	+	+	+	+	+	+	+	+	+	+	+
D.yakuba	+	+	+	+	+	+	+	+	+	+	+	+	+
D.eugracilis	+	+	+	+	+	+	+	+	+	+	+	+	+
D.takahashii	+	+	+	+	+	+	+	+	+	+	+	+	+
D.biarmipes	+	+	+	+	+	+	+	+	+	+	+	+	+
D.elegans	+	+	+	+	+	+	+	+	+	+	+	+	+
D.rhopoloa	+	+	+	+	+	+	+	+	+	+	+	+	+
D.ficusphila	-	+	+	-	+	+	+	+	+	+	+	+	+
D.kikkawai	+	+	+	+	+	+	+	-	+	+	+	+	+
D.bipectinata	+	+	+	+	+	+	+	+	+	+	+	+	+
D.ananassae	+	+	+	+	+	+	+	+	+	+	+	+	+
D.pseudoobscura	+	+	+	+	+	+	+	+	+	+	+	+	+
D.persimilis	+	+	+	+	+	+	+	+	+	+	+	+	+
D.miranda	+	+	+	+	+	+	+	+	+	+	+	+	+
D.willistoni	+	+	+	+	+	+	+	+	+	+	+	+	+
D.mojavensis	-	+	+	+	+	+	+	+	-	+	+	+	+
D.virilis	+	+	+	+	+	+	+	+	+	+	+	-	+
D.albomicans	-	-	-	-	-	-	-	+	-	-	-	-	-
D.grimshawi	-	-	-	-	-	-	-	+	-	+	+	-	+

Fig. 3


seminase
CG11037
CG10587
RpL32
M F g (-)
M F g (-)
M F g (-)
M F g (-)
D. melanogaster
D. yakuba
D. ficusphila
D. bipectinata
D. ananassae
D. pseudoobscura
D. willistoni