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. Within SP-responsive species, 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 specific phylogenetic lineages that correspond with increased expression of the SP receptor in female reproductive tracts or increased sperm length, respectively. 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

25 **Introduction**

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

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

48 the uterus (Yapici et al., 2008, Hasemeyer et al., 2009, Yang et al., 2009). This gradual "dosing"
49 of SP causes the persistence of the hormone's effects on female behavior and physiology. SPR
50 signaling is also required for the efficient release of sperm from the storage organs (Avila et al.,
51 2015).

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

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

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

93 While *Drosophila* species differ in reproductive traits for many reasons (Markow &
94 O'Grady, 2005), some of these differences relate directly to the SP network and could thus be
95 causes or consequences of SP network evolution. For example, a change in sperm length may
96 affect the amount of SP that can bind, and other structural changes to sperm could affect the
97 binding of SP and other network proteins that interact with sperm, such as CG1652 and CG1656
98 (Ravi Ram & Wolfner, 2009, Singh et al., 2018). The amount of SP bound to sperm, the rate and
99 efficacy of its release (Peng et al., 2005), and its ability to bind SPR (Yapici et al., 2008) could
100 affect re-mating rates, while the structure of female sperm storage organs could affect the ability
101 of the network proteins to bind SP to sperm or modulate SP's interaction with SPR. Changes in
102 these traits – sperm length, female remating rate, and female sperm storage structures – have
103 been well documented in the literature (Alpern et al., 2019, Joly & Bressac, 1994, Markow,
104 1996, Markow & O'Grady, 2005, Pitnick et al., 1999, Snook et al., 1994, Snook, 1995, Singh et
105 al., 2002), and we summarize them and infer their timing in Figure 1. Of particular relevance to
106 this study, the phenotypic and phylogenetic data are consistent with SPR evolving to be
107 expressed in female reproductive tracts along the lineage leading to the *melanogaster* group of
108 flies (Tsuda et al., 2015; branch 11) and an increase in sperm length on the lineage leading to *D.*
109 *ananassae* and *D. bipectinata* (Joly & Bressac, 1994, Markow, 1996, Pitnick et al., 1999; branch
110 15).

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

116 reproductive system in two such species from published data (Yang et al., 2018, Kelleher et al.,
117 2009). Using PAML (Yang, 2007) we determined that recurrent positive selection has acted on
118 specific sites in several of these proteins. We also detected marginal evidence that positive
119 selection has acted on certain network proteins on key phylogenetic lineages corresponding with
120 major changes in SP-related phenotypes. Finally, we used RT-PCR to investigate whether any
121 gross-scale changes in gene expression occurred surrounding the gene duplication event that
122 gave rise to one of the SP network proteins, seminase. Taken together, our results suggest that
123 the members of the SP network had the potential to influence reproductive success before the
124 onset of SP/SPR-mediated responses in the reproductive tract of mated females of the
125 *melanogaster* group of species. However, additional adaptive changes in these proteins occurred
126 concurrent with, and subsequent to, these critical changes in the fly reproductive system. These
127 results underscore the strength of sexual selection acting in *Drosophila* and illustrate potential
128 molecular changes that occur in the face of such selection.

129 **Methods**

130 *Identification of SP network proteins across* Drosophila *species*

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

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

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

152 and a sequence that could be translated conceptually to produce a polypeptide without premature
153 stops, were retained for study. In cases of duplicate genes (*seminase*, *CG1652* and *CG1656*), we
154 also used gene order and synteny to confirm correct ortholog identification.
155
156 *Sequence alignment*
157 For each SP network protein, we used MUSCLE as implemented in MEGA 6.06 (Tamura
158 et al., 2013) to align amino acid sequences, then visually checked and edited each alignment for
159 accuracy. Amino acid alignments were then back-translated in MEGA to obtain the cDNA
160 alignment.
161
162 *Phylogenetic analysis*
163 To infer a *Drosophila* consensus phylogeny based on all SP network proteins, we
164 concatenated the amino acid alignments of all SP network proteins within each of the 22 species.
165 We used PROML in Phylip (Felsenstein, 2005) to infer an unrooted maximum-likelihood
166 phylogeny (with random input order, slow analysis, and all other default parameters). Gaps in the
167 alignment were used in cases in which a protein was not present in a particular species. The
168 resulting phylogeny matched published *Drosophila* phylogenies, except for *D. virilis* and *D.*
169 *mojavensis* (Drosophila 12 Genomes et al., 2007, Markow & O'Grady, 2005, Seetharam &
170 Stuart, 2013). We then used this consensus tree for the PAML analyses, with species removed
171 on a gene-by-gene basis as described below.
172
173 *Detection of recombination*

174 Because recombination within a gene sequence can impact the results of analyses to
175 detect selection, we first used GARD with default parameters as implemented in DataMonkey
176 2.0 to check for evidence of recombination within each gene (Kosakovsky Pond et al., 2006,
177 Weaver et al., 2018). Genes were partitioned at breakpoints evaluated as significant by the
178 Kishino-Hasegawa test (p-value < 0.05 for both LH and RH; Table S2), and PAML was run on
179 each gene segment separately. We performed PAML analyses on sequence alignments spanning
180 two different ranges in the *Drosophila* phylogeny: the branch and branch-sites tests (see below)
181 were run on species from the entire genus, while the sites test (see below) was run on species
182 from only the *melanogaster* group. Thus, we generated a set of recombination breakpoints for
183 each set of species (i.e., all examined members of the genus or members of the *melanogaster*
184 group; Table S2). For each species set, six SP network genes showed evidence of
185 recombination, but the genes that showed recombination differed between the two sets of
186 species.
187
188 *PAML analyses*
189 For each protein, we used codeml of the PAML package to perform evolutionary
190 analyses on protein-coding DNA sequence (Yang, 2007). To test for heterogeneity in the rate of
191 each protein's evolution across the phylogeny, we utilized the PAML branch test, which uses a
192 likelihood ratio test (LRT) to compare the "free ratio" model, allowing for different ω values for
193 each branch, with model M0, which estimates a single ω for the whole phylogeny (Yang, 1998).
194 For these tests, and for the branch-sites tests below, we used the consensus tree described above
195 that covered the entire *Drosophila* phylogeny, but manually removed from it any species for
196 which: a) an ortholog could not be identified, or b) an ortholog was identified, but it could not be

197 confidently aligned due to ambiguity over an intron position or the end of the protein-coding
198 region. Table S1 shows the set of species used for the molecular evolutionary analyses for each
199 gene.

200 To test whether a subset of sites in a protein had evolved under recurrent positive
201 selection, we used LRTs to compare an evolutionary model (M8) that allows a class of sites to
202 have $\omega > 1$ to models M7 and M8a, which allow only purifying selection or neutral evolution
203 (Swanson et al., 2003, Yang et al., 2000). If model M8 can explain the observed sequence data
204 significantly better than models M7 and M8a, as assessed by a LRT, then positive selection
205 acting on a subset of sites can be inferred. For proteins for which model M8 was significantly
206 preferred to models M7 and M8a, we used the Bayes Empirical Bayes (BEB) approach to
207 identify at the 0.9 confidence level the specific residues that have evolved adaptively. These
208 comparisons were done only for species within the *melanogaster* group due to the possibility of
209 synonymous site saturation if more divergent species were included. (Such saturation occurs
210 when the rate of substitutions at synonymous sites, d_S, approaches 1, since it then becomes
211 impossible to distinguish whether one or multiple mutations have occurred at these sites. This,
212 in turn, reduces the accuracy of estimates of d_N/d_S.) To check for convergence in the 'free-ratio'
213 and the sites models, we ran codeml twice with the initial omega set at 0.4 and 2, respectively.

214 Finally, we performed the branch-sites test for positive selection (Zhang et al., 2005) to
215 identify classes of sites that had evolved adaptively along either of two specific lineages in the
216 phylogeny that we identified *a priori* because they represent likely evolutionary transitions in
217 key SP-related traits. First, we tested for sites under selection on the branch leading to the
218 *melanogaster* group of species (Fig. 1, branch 11), since this branch corresponds with the
219 inferred timing of when the SP receptor became expressed in the female reproductive tract and,

220 consequently, when females became sensitive to the non-receptivity effect caused by SP (Tsuda
221 et al., 2015). Second, we tested for sites under selection in the lineage that leads to and separates
222 *D. ananassae* and *D. bipectinata* from the rest of the *melanogaster* group species (Fig. 1, branch
223 15), because these species are known to have somewhat longer sperm (Joly & Bressac, 1994,
224 Markow, 1996, Pitnick et al., 1999). Although we inferred other important evolutionary
225 transitions in reproductive traits on the broader *Drosophila* phylogeny (Fig. 1), we limited our
226 branch-sites analyses to these two lineages because of the greater number of available sequenced
227 species in the *melanogaster* group.

228 In the branch-sites test, we used a LRT to compare a null model allowing for only
229 purifying and neutral selection on the focal branch with an alternative model allowing for a class
230 of sites to evolve under positive selection (Yang, 2007, Yang & Dos Reis, 2011, Zhang et al.,
231 2005). Recently, Venkat et al. (2018) found that this branch-sites test can have a high rate of
232 false positives driven by multinucleotide mutations within codons (i.e., mutations at adjacent
233 sites). To control for this issue, we implemented the tests in the Venkat model, a version of
234 PAML developed by these authors that runs the analysis after masking these sites. PAML
235 analyses were implemented using custom batch scripts for GNU parallel (Tange, 2018) and
236 PAML version 4.8a or HyPhy version 2.5.1 (in the case of the Venkat model).

237

238 *Identification of* seminase *orthologs and paralogs*

239 We identified the predicted amino acid sequences for orthologs of seminase, CG11037
240 and CG10587 in *Drosophila* species using the methods described above. To confirm that calls of
241 orthology for seminase and its paralogs were accurate, we used Phylip's PROML program
242 (Felsenstein, 2005) to infer a maximum-likelihood rooted tree (using the single copies in *D.*

243 *pseudoobscura* and *D. persimillis* as the outgroup, and default PROML parameters). This was
244 consistent with the orthology assignments made using conserved gene order, except for *D.*
245 *ananassae* and *D. bipectinata*, which are likely confounded by their long branch length.
246
247 *Evaluation of gene expression*
248 *D. melanogaster, D. yakuba, D. ficusphila, D. bipectinata, D. annanassae, D.*
249 *pseudoobscura* and *D. willistoni* were raised in the lab as in Tsuda et al. (2015). We CO_2-
250 anesthetized 9-day-old flies of each species, separated them by sex, homogenized male or female
251 whole flies in TRIzol reagent, and purified RNA from samples and synthesized cDNA as
252 previously described (Gubala et al., 2017). We then used species-specific primers to amplify
253 *seminase*, *CG11037* or *CG10587*, with the non-reproductive, housekeeping gene, *ribosomal*
254 *protein L32* (*RpL32*) as a control that should be expressed equally in both sexes. Genomic DNA
255 was used as a positive control for PCR reactions, and water was used in place of template in
256 negative control reactions.

257 **Results and Discussion**

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

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

267 Figure 2 shows that all SP network protein-coding gene orthologs are detectable in a
268 large majority of the species surveyed, including those outside of the *melanogaster* group. For
269 example, we found all currently known network proteins in *D. pseudoobscura* and *D. willistoni*,
270 and all but one ortholog in *D. virilis*. To assess whether these orthologs were likely to function
271 in reproduction outside of the *melanogaster* group, we examined publicly availably RNAseq data
272 from male reproductive tracts in *D. pseudoobscura* (Yang et al., 2018) and proteomic data from
273 male accessory glands in *D. mojavensis* (Kelleher et al., 2009). Transcripts of orthologs of male-
274 derived network proteins were consistently enriched in (or entirely specific to) samples from
275 whole males, male testes, and male carcasses in *D. pseudoobscura*, while showing either no or
276 low expression in females or in male heads (Fig. S1). Expression enrichment in whole males and
277 male carcasses is consistent with expectations for reproductive proteins produced in the male
278 accessory gland. Some of these proteins may also be expressed in the testes, or the "testis"
279 expression could be due to contamination of testis dissections with accessory gland tissue. The

280 genes encoding female-derived proteins showed broader expression patterns (Fig. S1), including
281 in whole females and whole males, but this pattern is consistent with their *D. melanogaster*
282 orthologs, the expression of which is not limited to the female reproductive system (Brown et al.,
283 2014). Predicted orthologs of the male-derived network proteins CG1652, CG1656, CG9997,
284 CG17575, seminase, aquarius and antares were also identified in a proteomic analysis of the *D.*
285 *mojavensis* accessory gland (Kelleher et al., 2009). Subsequent work showed that males of this
286 species transfer transcripts of the *antares* ortholog to females during mating (Bono et al., 2011).
287 Thus, RNAseq and proteomic data from two outgroup species are consistent with many SP
288 network proteins functioning in reproduction outside of the *melanogaster* group.
289 It is likely that some SP network proteins function in other processes in certain species
290 that impact their evolutionary trajectories. For example, SPR is expressed in both sexes outside
291 of the reproductive tract (Tsuda et al., 2015, Yapici et al., 2008), and myoinhibitory peptides
292 (MIPs) are known ligands in addition to SP (Kim et al., 2010, Poels et al., 2010, Yamanaka et al.,
293 2010). SPR-MIP interactions outside of the reproductive tract affect sleep patterns in both sexes
294 and remating propensity in females (Jang et al., 2017, Oh et al., 2014). Such interactions, in
295 addition to the sexual selective pressures exerted by SP network-mediated interactions and
296 reproductive phenotypes, have likely contributed to the evolution of SP network proteins in
297 various *Drosophila* lineages. While most male-derived network proteins appear to have male-
298 specific or heavily male-biased expression (in species for which expression data are available),
299 the female-derived proteins show broader expression patterns. Understanding these proteins'
300 non-reproductive functions will shed additional light on evolutionary forces that may have
301 shaped them.
302

303 *SP network proteins demonstrate evolutionary rate heterogeneity across the* Drosophila
304 *phylogeny*

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

310 To begin investigating these proteins' molecular evolution across the genus, we used
311 PAML model M0 to estimate for each gene a single d_N/d_S ratio (ω) across all species for either
312 the full-length protein-coding sequence or for each segment identified by the GARD analysis.
313 We then performed the branch test (Yang, 1998) to assess whether ω varied significantly across
314 different branches of the phylogeny. Most network proteins (and segments of proteins) had full-
315 length ω estimates around 0.2 across the full-genus tree (Table S3). While these rates are
316 generally indicative of purifying selection, they are nonetheless faster than those observed for
317 most proteins in *D. melanogaster* (Drosophila 12 Genomes Consortium, 2007), consistent with
318 the general trend for reproductive proteins (Wilburn & Swanson, 2016). Three proteins showed
319 notably slower evolutionary rates: CG17575 (most of protein found in segments with $\omega < 0.1$), a
320 male-expressed cysteine-rich secretory protein required for binding of SP to sperm (Ravi Ram &
321 Wolfner, 2009); Esp ($\omega = 0.03$), a female-expressed, predicted sulfate membrane transporter also
322 required for long-term fertility (Findlay et al., 2014); and SPR (most of the protein contained in a
323 segment with $\omega = 0.05$), the female-expressed G-protein coupled receptor for SP required for
324 female post-mating changes including egg-laying, resistance to remating and release of sperm
325 from sperm-storage organs (Avila et al., 2015, Yapici et al., 2008). While these proteins' slow

326 rates of evolution could indicate that they play highly conserved roles in reproduction, it is also
327 possible that they have evolved adaptively at only a few sites or on a few lineages (see below).
328 We next ran the "free ratio" model in which PAML estimates an ω value for each branch
329 of the phylogeny. We found significant evidence of evolutionary rate heterogeneity for all but
330 one network protein, intrepid (Table S3). Additionally, all proteins but intrepid had at least one
331 phylogenetic branch for which ω was estimated to be > 1. While the branch test is not a rigorous
332 test of positive selection acting on specific branches, the results indicate that the evolutionary
333 rates of most SP network proteins have varied significantly across their evolutionary histories.
334 Indeed, several network proteins had branch-specific ω estimates > 1 in some of the more
335 ancestral branches of the tree, highlighting the potential for more detailed studies of these
336 proteins' evolutionary histories as the genomes of more divergent *Drosophila* species become
337 available. In contrast, the constant, slow rate of evolution for intrepid implies that this protein
338 has likely played a conserved and important role since the origin of the genus. Intrepid has
339 undergone less functional characterization than other male-expressed male network proteins, so
340 we cannot speculate further about its specific role(s) in reproduction.
341
342 *Several SP network proteins have undergone recurrent positive selection at specific sites since*
343 *the evolution of SPR expression in female reproductive tracts*
344 To determine the extent to which positive selection has shaped the evolution of the SP
345 network proteins, we used the PAML sites test to ask whether any protein had a particular subset
346 of sites that had undergone recurrent positive selection. Because of the likelihood of
347 synonymous site saturation over longer phylogenetic distances, we limited the sequences used in
348 this analysis to those from the *melanogaster* group. This set of species also represents the likely

349 extent of major SP/SPR-mediated post-mating responses, as only these species express SPR at
350 high levels in the female reproductive tract and respond to injection of synthetic SP (Tsuda et al.,
351 2015). Thus, our analyses identify proteins that might have evolved adaptively to further
352 improve/refine network function in the past ~15 million years (Seetharam & Stuart, 2013).
353 The results of the sites analyses are shown in Table 1. Four proteins – CG9997, fra
354 mauro, CG1652 and hadley – show significant evidence for having a class of amino acid sites
355 that have evolved under recurrent positive selection across the *melanogaster* group of species.
356 Three other proteins (antares, intrepid and CG17575) each have a class of sites found to be under
357 positive selection in the Model M7/M8 comparison, but these results are no longer significant
358 when comparing Models M8 and M8a, suggesting that the class of more quickly evolving sites
359 identified for each protein in Model M8 may be evolving neutrally rather than under positive
360 selection.
361 While all members of a protein network have a degree of functional co-dependence, the
362 male-expressed network proteins that have evolved adaptively are known to have specific,
363 genetic interdependencies. CG9997, a serine protease homolog predicted to be catalytically
364 inactive, must be produced in the male accessory glands for CG1652, a C-type lectin, to be
365 transferred to mated females (Ravi Ram & Wolfner, 2009). Likewise, in the absence of CG1652,
366 CG9997 is not efficiently "processed" from its 45-kDa form to its 36-kDa form in mated females
367 (Ravi Ram & Wolfner, 2009, Singh et al., 2018). The loss of either protein prevents SP from
368 accumulating on stored sperm in females. Recent work has shown that both CG9997 and
369 CG1652 also bind to sperm, though their sperm-binding is detectable only in the hours after
370 mating, while SP binding lasts for several days (Peng et al., 2005, Singh et al., 2018). CG9997
371 and CG1652 also show significant evidence of evolutionary rate covariation (Findlay et al.,

372 2014). These results suggest that pressure to maintain their functional interactions may be a
373 factor driving the adaptive evolution of CG9997 and CG1652, as has been observed for pairs of
374 interacting reproductive proteins in other systems (Clark et al., 2009, Grayson, 2015).
375 Other work on CG9997 is consistent with its adaptive evolution. Wong et al. (2008)
376 found evidence for recent positive selection acting on this gene by examining patterns of
377 polymorphism and divergence between populations of *D. melanogaster* and *D. simulans*. They
378 hypothesized that non-catalytically active serine protease homologs like CG9997 function as
379 agonists or antagonists for active proteases. Another possibility, which is not mutually
380 exclusive, is that protease homologs bind to other proteins or molecules in the female tract to
381 slow their rate of digestion by active, female-derived proteases (LaFlamme & Wolfner, 2013).
382 Under either scenario, protease homologs like CG9997 may need to continually coevolve with
383 their interacting partners, providing the impetus for the recurrent, adaptive evolution detected
384 here. Additionally, knockdown of *CG9997* diminishes male sperm competitive ability (Castillo
385 & Moyle, 2014), suggesting another potential factor in its adaptive evolution.
386 Less functional information exists for the adaptively evolving, female-expressed proteins.
387 Both fra mauro and hadley were identified in a screen for female-expressed proteins that
388 coevolved with a male-expressed SP network protein; in each case, the coevolutionary signal
389 was with CG17575 (Findlay et al., 2014). RNAi knockdown of either gene reduced female
390 fertility, though knockdown females could receive SP and store it properly on sperm (Findlay et
391 al., 2014). These data suggested that the proteins could be involved in maintaining the female
392 long-term response to SP, though *fra mauro* knockdown females also showed a significant
393 fertility defect in the 24 hrs after mating (Findlay et al., 2014). The fra mauro protein encodes a
394 predicted neprilysin protease, which may coevolve with its as yet unknown molecular targets or

395 antagonists (LaFlamme & Wolfner, 2013). As noted above, functional domains have not been
396 identified for the hadley protein, so it is difficult to speculate on potential forces driving its
397 adaptive evolution.

398 Notably, several proteins in the SP network showed no evidence of recurrent adaptive
399 evolution within the *melanogaster* group, while others had subsets of sites with evolutionary
400 rates that were elevated, but approximated neutrality. These data suggest that while some
401 network proteins may contain regions that are under relaxed constraint, much of the functionality
402 and interdependence of the network might have already existed at the origin of the *melanogaster*
403 group.

404

405 *Several network proteins underwent adaptive evolution on specific lineages correlating with*
406 *changes in reproductive phenotypes*

407 While the PAML sites test described above detects recurrent adaptive evolution, protein
408 networks can also be shaped by bursts of episodic positive selection acting on specific
409 phylogenetic lineages. One important evolutionary transition for the SP network occurred at the
410 base of the *melanogaster* group, when *SPR* evolved expression in the lower female reproductive
411 tract (Tsuda et al., 2015). This change likely created (or exacerbated) a selective pressure for
412 higher SP levels in this location, as prolonged SP-SPR signaling could promote continued egg
413 production and prolong female non-receptivity to re-mating. Because a primary purpose of the
414 male-expressed SP network proteins in *D. melanogaster* is to bind SP to sperm to prolong the
415 post-mating response, we hypothesized that some of these proteins might have experienced a
416 burst of adaptive evolution on the same phylogenetic branch on which female reproductive *SPR*
417 expression is inferred to have evolved. Likewise, the increase in *SPR* expression in females

418 could have created a selective pressure for other female-expressed members of the network to
419 evolve. To test these ideas, we used the Venkat model, a modified PAML branch-sites test
420 (Venkat et al., 2018, Zhang et al., 2005), to ask whether any network protein had a subset of sites
421 under selection on the branch leading to the *melanogaster* group (i.e., branch 11 in Fig. 1).
422 Table 2 (left columns) shows the results of these tests. Two proteins show marginal
423 evidence for adaptive evolution on branch 11, leading to the *melanogaster* group: CG1656 and
424 SPR. As originally formulated (Zhang et al., 2005), the LRT for the branch-sites test follows a
425 null distribution described as an equal mixture of point mass 0 and a chi-square distribution with
426 1 degree of freedom (df). Under this null distribution, the test statistic corresponding with a *p*-
427 value of 0.05 is 2.71, a value exceeded in tests of both CG1656 and SPR. However, the test is
428 typically conducted conservatively (Venkat et al., 2018, Zhang et al., 2005), following only a
429 chi-square distribution with 1 df. The p-values listed in Table 2 are calculated based on this
430 latter distribution, and they are marginally non-significant ($0.05 < p < 0.1$) for CG1656 and SPR.
431 The potential adaptive evolution of sites in the SPR protein along branch 11 is
432 interesting, because this lineage also represents the time during which the protein became
433 expressed in the female reproductive tract (Tsuda et al., 2015). Thus, it is possible that the SPR
434 gene underwent both regulatory and protein-coding adaptations that altered how the female post-
435 mating response is controlled. The other protein that potentially underwent adaptive evolution
436 along this lineage is the predicted C-type lectin CG1656, which functions similarly to its
437 recurrently rapidly evolving paralog described above, CG1652. Both lectins are required for
438 SP's long-term binding to stored sperm, and both proteins themselves bind sperm temporarily in
439 the hours after mating (Singh et al., 2018). Given the potential selective pressure to bind more
440 SP to stored sperm in female tracts expressing *SPR*, it is possible that the adaptive evolution of

441 CG1656 on this key phylogenetic branch could have helped to improve the efficiency of SP's
442 binding to sperm. This idea could be tested in future experiments by either identifying and
443 mutating the residues likely to have changed along branch 11 and/or by substituting an outgroup
444 ortholog of CG1656 (and potentially its duplicate, CG1652) into *D. melanogaster* and examining
445 the effects on SP's sperm binding and on the female long-term post-mating response.

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

457 To test for whether any SP network proteins experienced adaptive evolution concurrent
458 with this change in sperm length, we again used the modified branch-sites test. Two network
459 proteins, antares and CG17575, show evidence of positive selection acting on specific sites on
460 the lineage leading to *D. ananassae* and *D. bipectinata* (Table 2, right columns). Antares' signal
461 of selection is significant under both null distributions described above, while CG17575's signal
462 is significant under the mixed null distribution and approached significance ($p = 0.0504$) under
463 the conservative test. In addition to facilitating SP's long-term binding to sperm, antares also

464 binds to sperm itself for a shorter period (Findlay et al., 2014, Singh et al., 2018). Thus, antares
465 might have evolved adaptively to facilitate greater or more efficient binding of either itself or SP
466 to sperm as sperm tails lengthened. Interestingly, the antares ortholog in outgroup species *D.*
467 *mojavensis* and *D. arizonae* was also found to evolve under diversifying selection (Bono et al.,
468 2015), even though *D. mojavensis* does not have a currently detectable SP ortholog (Tsuda et al.,
469 2015) (Fig. 2). Heterospecific matings between these species fail due to post-mating, pre-zygotic
470 isolating barriers, which include problems with sperm storage in the female reproductive tract
471 (Kelleher & Markow, 2007). It is thus possible that antares plays an essential role in binding
472 molecules to sperm and/or facilitating sperm storage, and that the male reproductive activity of
473 antares has been refined by different selective pressures in different lineages.

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

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

487

488 *Seminase gene duplicates retain male-specific expression patterns across* melanogaster *group*
489 *species*

490 In addition to CG17575, the male-expressed serine protease seminase is required for the
491 localization of SP and other male-expressed proteins to the SR after mating (LaFlamme et al.,
492 2012, Singh et al., 2018). Seminase arose through gene duplication in the lineage leading to the
493 *melanogaster* group of flies. The genomes of *D. pseudoobscura* and other outgroup species
494 contain only one detectable copy of the gene, but in *D. melanogaster* and its fellow *melanogaster*
495 group members, there are three tandemly arrayed, intron-containing copies, suggesting two
496 distinct DNA-based duplication events. The other genes are *CG10587* and *CG11037*. Like
497 *seminase*, both are expressed specifically in the male accessory gland in *D. melanogaster*
498 (Brown et al., 2014, Leader et al., 2018). While we detected no recurrent or episodic positive
499 selection acting on seminase after these duplications (Tables 1-2), we were curious whether it or
500 its paralogs might have evolved different expression patterns (and, thus, potential functions) after
501 duplication. We thus performed RT-PCR to amplify each paralog from cDNA isolated from
502 males or females of a variety of species from the *melanogaster* group. We also assessed the
503 expression of the single-copy parent gene from *D. pseudoobscura* and *D. willistoni*. Our results
504 (Figure 3) show that both the single-copy genes from the outgroup species, as well as all of
505 paralogs from all *melanogaster* group species tested, are expressed specifically in adult males.
506 This result is consistent with the ancestral single copy of seminase also functioning in male
507 reproduction (and potentially with other SP network proteins).

508 Given that seminase itself has additional reproductive functions beyond its role in the SP
509 network (LaFlamme et al., 2012), it is possible that the paralogs have sub- or neo-functionalized

510 to have unique roles, in spite of their conserved expression patterns. Future studies should
511 evaluate how the paralogs contribute to reproduction, which may suggest possible evolutionary
512 forces that affected their evolution after the gene duplication events.

513 **Conclusions**

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

541 **References**
542

543 Alpern, J. H. M., Asselin, M. M. & Moehring, A. J. 2019. Identification of a novel sperm class
544 and its role in fertilization in Drosophila. *J Evol Biol* **32**: 259-266.
545 Apger-McGlaughon, J. & Wolfner, M. F. 2013. Post-mating change in excretion by mated
546 Drosophila melanogaster females is a long-term response that depends on sex peptide and
547 sperm. *J Insect Physiol* **59**: 1024-30.
548 Avila, F. W., Mattei, A. L. & Wolfner, M. F. 2015. Sex peptide receptor is required for the
549 release of stored sperm by mated Drosophila melanogaster females. *Journal of Insect*
550 *Physiology* **76**: 1-6.
551 Avila, F. W., Ravi Ram, K., Bloch Qazi, M. C. & Wolfner, M. F. 2010. Sex peptide is required
552 for the efficient release of stored sperm in mated Drosophila females. *Genetics* **186**: 595-
553 600.
554 Bono, J. M., Matzkin, L. M., Hoang, K. & Brandsmeier, L. 2015. Molecular evolution of
555 candidate genes involved in post-mating-prezygotic reproductive isolation. *J Evol Biol*
556 **28**: 403-14.
557 Bono, J. M., Matzkin, L. M., Kelleher, E. S. & Markow, T. A. 2011. Postmating transcriptional
558 changes in reproductive tracts of con- and heterospecifically mated Drosophila
559 mojavensis females. *Proc Natl Acad Sci U S A* **108**: 7878-83.
560 Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W.,
561 Wen, J., Park, S., Suzuki, A. M., et al. 2014. Diversity and dynamics of the Drosophila
562 transcriptome. *Nature* **512**: 393-9.
563 Carvalho, G. B., Kapahi, P., Anderson, D. J. & Benzer, S. 2006. Allocrine modulation of feeding
564 behavior by the Sex Peptide of Drosophila. *Curr Biol* **16**: 692-6.
565 Castillo, D. M. & Moyle, L. C. 2014. Intraspecific sperm competition genes enforce post-mating
566 species barriers in Drosophila. *Proc Biol Sci* **281**: 20142050.
567 Chapman, T., Bangham, J., Vinti, G., Seifried, B., Lung, O., Wolfner, M. F., Smith, H. K. &
568 Partridge, L. 2003. The sex peptide of Drosophila melanogaster: female post-mating
569 responses analyzed by using RNA interference. *Proc Natl Acad Sci U S A* **100**: 9923-8.
570 Chen, P. S., Stumm-Zollinger, E., Aigaki, T., Balmer, J., Bienz, M. & Bohlen, P. 1988. A male
571 accessory gland peptide that regulates reproductive behavior of female D. melanogaster.
572 *Cell* **54**: 291-8.
573 Chen, Z. X., Sturgill, D., Qu, J., Jiang, H., Park, S., Boley, N., Suzuki, A. M., Fletcher, A. R.,
574 Plachetzki, D. C., FitzGerald, P. C., et al. 2014. Comparative validation of the D.
575 melanogaster modENCODE transcriptome annotation. *Genome Res* **24**: 1209-23.
576 Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F. & Swanson, W. J. 2009.
577 Coevolution of interacting fertilization proteins. *PLoS Genetics* **5**: e1000570.
578 Drosophila 12 Genomes Consortium, Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M.,
579 Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., et al. 2007.
580 Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203-18.
581 Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. pp. Distributed by
582 Author, Department of Genome Sciences, University of Washington, Seattle.
583 Findlay, G. D., Sitnik, J. L., Wang, W., Aquadro, C. F., Clark, N. L. & Wolfner, M. F. 2014.
584 Evolutionary Rate Covariation Identifies New Members of a Protein Network Required
585 for Drosophila melanogaster Female Post-Mating Responses. *PLoS Genetics* **10**:
586 e1004108.

587 Grayson, P. 2015. Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm-
588 egg binding partners. *R Soc Open Sci* **2**: 150296.
589 Gubala, A. M., Schmitz, J. F., Kearns, M. J., Vinh, T. T., Bornberg-Bauer, E., Wolfner, M. F. &
590 Findlay, G. D. 2017. The Goddard and Saturn Genes Are Essential for Drosophila Male
591 Fertility and May Have Arisen De Novo. *Mol Biol Evol* **34**: 1066-1082.
592 Hasemeyer, M., Yapici, N., Heberlein, U. & Dickson, B. J. 2009. Sensory neurons in the
593 Drosophila genital tract regulate female reproductive behavior. *Neuron* **61**: 511-8.
594 Isaac, R. E., Li, C., Leedale, A. E. & Shirras, A. D. 2010. Drosophila male sex peptide inhibits
595 siesta sleep and promotes locomotor activity in the post-mated female. *Proc Biol Sci* **277**:
596 65-70.
597 Jang, Y. H., Chae, H. S. & Kim, Y. J. 2017. Female-specific myoinhibitory peptide neurons
598 regulate mating receptivity in Drosophila melanogaster. *Nat Commun* **8**: 1630.
599 Joly, D. & Bressac, C. 1994. Sperm Length in Drososphilidae (Diptera): Estimation by Testis
600 and Receptacle Lengths. *International Journal of Insect Morphology & Embryology* **23**:
601 85-92.
602 Kelleher, E. S. & Markow, T. A. 2007. Reproductive tract interactions contribute to isolation in
603 Drosophila. *Fly (Austin)* **1**: 33-7.
604 Kelleher, E. S., Watts, T. D., LaFlamme, B. A., Haynes, P. A. & Markow, T. A. 2009. Proteomic
605 analysis of Drosophila mojavensis male accessory glands suggests novel classes of
606 seminal fluid proteins. *Insect Biochem Mol Biol* **39**: 366-71.
607 Kim, Y. J., Bartalska, K., Audsley, N., Yamanaka, N., Yapici, N., Lee, J. Y., Kim, Y. C.,
608 Markovic, M., Isaac, E., Tanaka, Y., et al. 2010. MIPs are ancestral ligands for the sex
609 peptide receptor. *Proc Natl Acad Sci U S A* **107**: 6520-5.
610 Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. 2006.
611 GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096-8.
612 LaFlamme, B. A., Ravi Ram, K. & Wolfner, M. F. 2012. The Drosophila melanogaster seminal
613 fluid protease "Seminase" regulates proteolytic and post-mating reproductive processes.
614 *PLoS Genetics* **8**: 30-32.
615 LaFlamme, B. A. & Wolfner, M. F. 2013. Identification and function of proteolysis regulators in
616 seminal fluid. *Mol Reprod Dev* **80**: 80-101.
617 Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A. & Dow, J. A. T. 2018. FlyAtlas 2: a new
618 version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and
619 sex-specific data. *Nucleic Acids Res* **46**: D809-D815.
620 Liu, H. & Kubli, E. 2003. Sex-peptide is the molecular basis of the sperm effect in Drosophila
621 melanogaster. *Proc Natl Acad Sci U S A* **100**: 9929-33.
622 Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.
623 R. N., Potter, S. C., Finn, R. D., et al. 2019. The EMBL-EBI search and sequence
624 analysis tools APIs in 2019. *Nucleic Acids Res* **47**: W636-W641.
625 Markow, T. A. (1996) Evolution of Drosophila mating systems. In: *Evolutionary Biology*, Vol.
626 29. pp. 73-106.
627 Markow, T. A. & O'Grady, P. M. 2005. Evolutionary genetics of reproductive behavior in
628 Drosophila: connecting the dots. *Annual Review of Genetics* **39**: 263-291.
629 McDonough, C. E., Whittington, E., Pitnick, S. & Dorus, S. 2016. Proteomics of reproductive
630 systems: Towards a molecular understanding of postmating, prezygotic reproductive
631 barriers. *J Proteomics* **135**: 26-37.

632 Oh, Y., Yoon, S. E., Zhang, Q., Chae, H. S., Daubnerova, I., Shafer, O. T., Choe, J. & Kim, Y. J.
633 2014. A homeostatic sleep-stabilizing pathway in Drosophila composed of the sex
634 peptide receptor and its ligand, the myoinhibitory peptide. *PLoS Biol* **12**: e1001974.
635 Peng, J., Zipperlen, P. & Kubli, E. 2005. Drosophila sex-peptide stimulates female innate
636 immune system after mating via the toll and Imd pathways. *Current Biology* **15**: 1690-
637 1694.
638 Pitnick, S., Markow, T. A. & Spicer, G. S. 1999. Evolution of Multiple Kinds of Female Sperm-
639 Storage Organis in Drosophila. *Evolution* **53**: 1804-1822.
640 Poels, J., Van Loy, T., Vandersmissen, H. P., Van Hiel, B., Van Soest, S., Nachman, R. J. &
641 Vanden Broeck, J. 2010. Myoinhibiting peptides are the ancestral ligands of the
642 promiscuous Drosophila sex peptide receptor. *Cell Mol Life Sci* **67**: 3511-22.
643 Ravi Ram, K. & Wolfner, M. F. 2007. Sustained post-mating response in Drosophila
644 melanogaster requires multiple seminal fluid proteins. *PLoS Genetics* **3**: 2428-2438.
645 Ravi Ram, K. & Wolfner, M. F. 2009. A network of interactions among seminal proteins
646 underlies the long-term postmating response in Drosophila. *Proceedings of the National*
647 *Academy of Sciences of the United States of America* **106**: 15384-15389.
648 Rezaval, C., Pavlou, H. J., Dornan, A. J., Chan, Y. B., Kravitz, E. A. & Goodwin, S. F. 2012.
649 Neural circuitry underlying Drosophila female postmating behavioral responses. *Curr*
650 *Biol* **22**: 1155-65.
651 Schnakenberg, S. L., Matias, W. R. & Siegal, M. L. 2011. Sperm-storage defects and live birth in
652 Drosophila females lacking spermathecal secretory cells. *PLoS Biol* **9**: e1001192.
653 Seetharam, A. S. & Stuart, G. W. 2013. Whole genome phylogeny for 21 Drosophila species
654 using predicted 2b-RAD fragments. *PeerJ* **1**: e226.
655 Singh, A., Buehner, N. A., Lin, H., Baranowski, K. J., Findlay, G. D. & Wolfner, M. F. 2018.
656 Long-term interaction between Drosophila sperm and sex peptide is mediated by other
657 seminal proteins that bind only transiently to sperm. *Insect Biochem Mol Biol* **102**: 43-51.
658 Singh, S. R., Singh, B. N. & Hoenigsberg, H. F. 2002. Female remating, sperm competition and
659 sexual selection in Drosophila. *Genetics and Molecular Research* **1**: 178-215.
660 Sitnik, J. L., Gligorov, D., Maeda, R. K., Karch, F. & Wolfner, M. F. 2016. The Female Post-
661 Mating Response Requires Genes Expressed in the Secondary Cells of the Male
662 Accessory Gland in Drosophila melanogaster. *Genetics* **202**: 1029-41.
663 Snook, R., Markow, T. & Karr, T. 1994. Functional nonequivalence of sperm in *Drosophila*
664 *pseudoobscura*. *Proceedings of the National Academy of Sciences of the United States of*
665 *America* **91**: 11222-11226.
666 Snook, R. R. 1995. The Evolution of Sperm Polymorphism in the Drosophila Obscura Group.
667 957.
668 Soller, M., Bownes, M. & Kubli, E. 1999. Control of oocyte maturation in sexually mature
669 Drosophila females. *Dev Biol* **208**: 337-51.
670 Swanson, W. J., Nielsen, R. & Yang, Q. 2003. Pervasive adaptive evolution in mammalian
671 fertilization proteins. *Molecular biology and evolution* **20**: 18-20.
672 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular
673 evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-
674 2729.
675 Tange, O. 2018. *GNU Parallel 2018*.
676 Tsuda, M., Peyre, J. B., Asano, T. & Aigaki, T. 2015. Visualizing Molecular Functions and
677 Cross-Species Activity of Sex-Peptide in Drosophila. *Genetics*: 1-10.

678 Venkat, A., Hahn, M. W. & Thornton, J. W. 2018. Multinucleotide mutations cause false
679 inferences of lineage-specific positive selection. *Nat Ecol Evol*.
680 Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V. & Kosakovsky Pond, S. L. 2018.
681 Datamonkey 2.0: a modern web application for characterizing selective and other
682 evolutionary processes. *Mol Biol Evol* **35**: 773-777.
683 Wilburn, D. B. & Swanson, W. J. 2016. From molecules to mating: Rapid evolution and
684 biochemical studies of reproductive proteins. *J Proteomics* **135**: 12-25.
685 Wong, A., Turchin, M. C., Wolfner, M. F. & Aquadro, C. F. 2008. Evidence for positive
686 selection on Drosophila melanogaster seminal fluid protease homologs. *Mol Biol Evol* **25**:
687 497-506.
688 Yamanaka, N., Hua, Y. J., Roller, L., Spalovska-Valachova, I., Mizoguchi, A., Kataoka, H. &
689 Tanaka, Y. 2010. Bombyx prothoracicostatic peptides activate the sex peptide receptor to
690 regulate ecdysteroid biosynthesis. *Proc Natl Acad Sci U S A* **107**: 2060-5.
691 Yang, C. H., Rumpf, S., Xiang, Y., Gordon, M. D., Song, W., Jan, L. Y. & Jan, Y. N. 2009.
692 Control of the postmating behavioral switch in Drosophila females by internal sensory
693 neurons. *Neuron* **61**: 519-26.
694 Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K. & Oliver, B.
695 2018. Re-annotation of eight Drosophila genomes. *Life Sci Alliance* **1**: e201800156.
696 Yang, Z. 1998. Likelihood Ratio Tests for Detecting Positive Selection and Application to
697 Primate Lysozyme Evolution. *Molecular Biology and Evolution* **15**: 568-573.
698 Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and*
699 *Evolution* **24**: 1586-1591.
700 Yang, Z. & Dos Reis, M. 2011. Statistical properties of the branch-site test of positive selection.
701 *Molecular Biology and Evolution* **28**: 1217-1228.
702 Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. 2000. Codon-substitution models to
703 detect adaptive evolution that account for heterogeneous selective pressures among site
704 classes. *Molecular biology and evolution* **19**: 49-57.
705 Yapici, N., Kim, Y.-J., Ribeiro, C. & Dickson, B. J. 2008. A receptor that mediates the post-
706 mating switch in Drosophila reproductive behaviour. *Nature* **451**: 33-37.
707 Zhang, J., Nielsen, R. & Yang, Z. 2005. Evaluation of an improved branch-site likelihood
708 method for detecting positive selection at the molecular level. *Molecular Biology and*
709 *Evolution* **22**: 2472-2479.
710
711

712 **Figure Legends**
713
714 **Figure 1. Phylogeny of *Drosophila* species examined in this study.** The gray box indicates the
715 *melanogaster* group. Each branch is numbered for reference in the main text. Key changes in
716 reproductive tracts are indicated by letters a-d and are based on an examination of the literature
717 (references cited in main text). The PAML branch-sites tests (see Results) were conducted on
718 branches 11 and 15. Branch lengths are not proportional to evolutionary distances.
719
720 **Figure 2. Bioinformatic identification of SP network proteins across 22 *Drosophila* species.**
721 Identified orthologs that were also reciprocal best BLAST hits are noted with a + sign, while a –
722 sign indicates no ortholog could be identified.
723
724 **Figure 3. RT-PCR on *seminase* and paralogs shows conserved, male-biased expression after**
725 **duplication.** Orthologs of *seminase*, *CG10586* and *CG11037* show male-specific expression in
726 various *melanogaster* group species, though the level of expression between paralogs and species
727 is somewhat variable. The single-copy parent gene in *D. pseudoobscura* and *D. willistoni* is also
728 expressed in a male-specific manner. The non-reproductive housekeeping gene, *RpL32*, was
729 included as a positive control for successful cDNA synthesis from both sexes.
730

Table 1. PAML sites tests for positive selection acting on SP network genes within the *melanogaster* group. Genes for which recombination was detected are 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		
antr		0.22	10.194	0.0061	1.935	0.1643		
aqrs		0.21	3.950	0.1388	0.773	0.3792		
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_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		
CG9997		**0.25**	**21.001**	**2.75E-05**	**5.974**	**0.0145**	**10%**	**152S**
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		
Esp		0.03	0	1.0000	2.096	0.1477		
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_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		
intr		0.21	6.719	0.0348	0.547	0.4597		
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		
SP		0.22	1.463	0.4813	0	1.0000		
SPR		0.04	1.817	0.4031	8.828	0.0030		

741 **Table 2. Venkat model branch-sites tests for positive selection acting on specific sites of SP**
742 **network proteins on two specific lineages.** *P*-values are calculated based on a χ_1^2 distribution.
743 Asterisks indicate likelihood ratio test statistics that reach the $p < 0.05$ significance threshold for
744 a null distribution derived from a 50:50 ratio of point mass 0 and the χ_1^2 distribution.
745

	Branch 11: to *melanogaster* group (*SPR* expressed in female tract)			**Branch 15: to *D. ananassae/bipictenata*** (increased sperm length)		
Gene	**Whole-Gene ω estimate**	**2*ΔlnL**	***p*-value**	**Whole-Gene ω estimate**	**2*ΔlnL**	***p*-value**
antr	298.77	2.464	0.116	**6.19**	**3.948**	**0.047**
aqrs	1.00	0	1.000	2.53	0.388	0.533
CG1652	7.62	1.428	0.232	1.05	0.006	0.938
CG1656	122.47	3.238*	0.072	1.04	0	1.000
CG9997	1.04	0	1.000	1.05	0.022	0.882
CG17575	424.52	2.425	0.119	422.98	3.827*	0.050
Esp	1.08	0	1.000	1.08	0	1.000
frma	1.08	0.034	0.854	1.04	0.012	0.913
hdly	1.08	0	1.000	1.64	0.008	0.929
intr	1.00	0.039	0.843	1.02	0	1.000
sems	1.00	0	1.000	403.51	0.962	0.327
SP	1.00	0	1.000	1.04	0.002	0.963
SPR	9999.99	3.064*	0.080	1.08	0	1.000

746
747
748

749 **Supplemental Materials**
750
751 **Table S1. Orthologs used for each gene in PAML analyses**. Some orthologs that were
752 identified in Table 2 were excluded from PAML analysis due to unresolved intron borders and/or
753 poor alignment quality. Only species above the dotted line (the *melanogaster* group) were
754 analyzed in the sites tests.
755
756 **Table S2. GARD results showing inferred recombination breakpoints.** Breakpoint positions
757 refer to nucleotide positions in the alignment files used. However, since alignments include
758 gaps, these positions do not necessarily have a 3:1 correspondence with the *D. melanogaster*
759 amino acid positions reported in Tables 1-3. The first table shows recombination breakpoints
760 detected for aligned sequences from the entire *Drosophila* genus, which were used for the branch
761 and branch-sites tests. The second table shows recombination breakpoints detected for aligned
762 sequences from only the *melanogaster* group, which were used for the sites tests.
763
764 **Figure S1. RNAseq data from *D. pseudoobscura* support reproductive functions for SP**
765 **network proteins in a species that lacks full-scale SP responses.** A) *D. pseudoobscura*
766 expression patterns for each member of the SP network. Dark shading indicates high expression
767 levels, stripes indicate low (but detectable) expression, and no shading indicates no expression
768 detected in a given sample. Male-derived network proteins show male-biased or male-specific
769 expression, consistent with reproductive functions. B) Examples of *D. pseudoobscura*
770 expression data for several SP network genes; shading in part (A) is based on these data. The
771 RNAseq data were accessed via FlyBase and generated by Yang et al. (Yang et al., 2018).
772
773 **Table S3. Branch tests for rate heterogeneity.** Partitions were implemented in PAML
774 analyses if they were significant in both the LH and RH tests.
775
776 **Table S4. Venkat model branch-sites tests for positive selection acting on specific sites of**
777 **SP network proteins detected by GARD to have multiple recombination segments.** The
778 table shows results for both branch 11 and branch 15 tests. Only the six genes for which
779 recombination was detected in the relevant trees are shown in the table. *P*-values are calculated
780 based on a χ_1^2 distribution. Asterisks indicate likelihood ratio test statistics that reach the $p <$
781 0.05 significance threshold for a null distribution derived from a 50:50 ratio of point mass 0 and
782 the χ_1^2 distribution.

Fig. 1

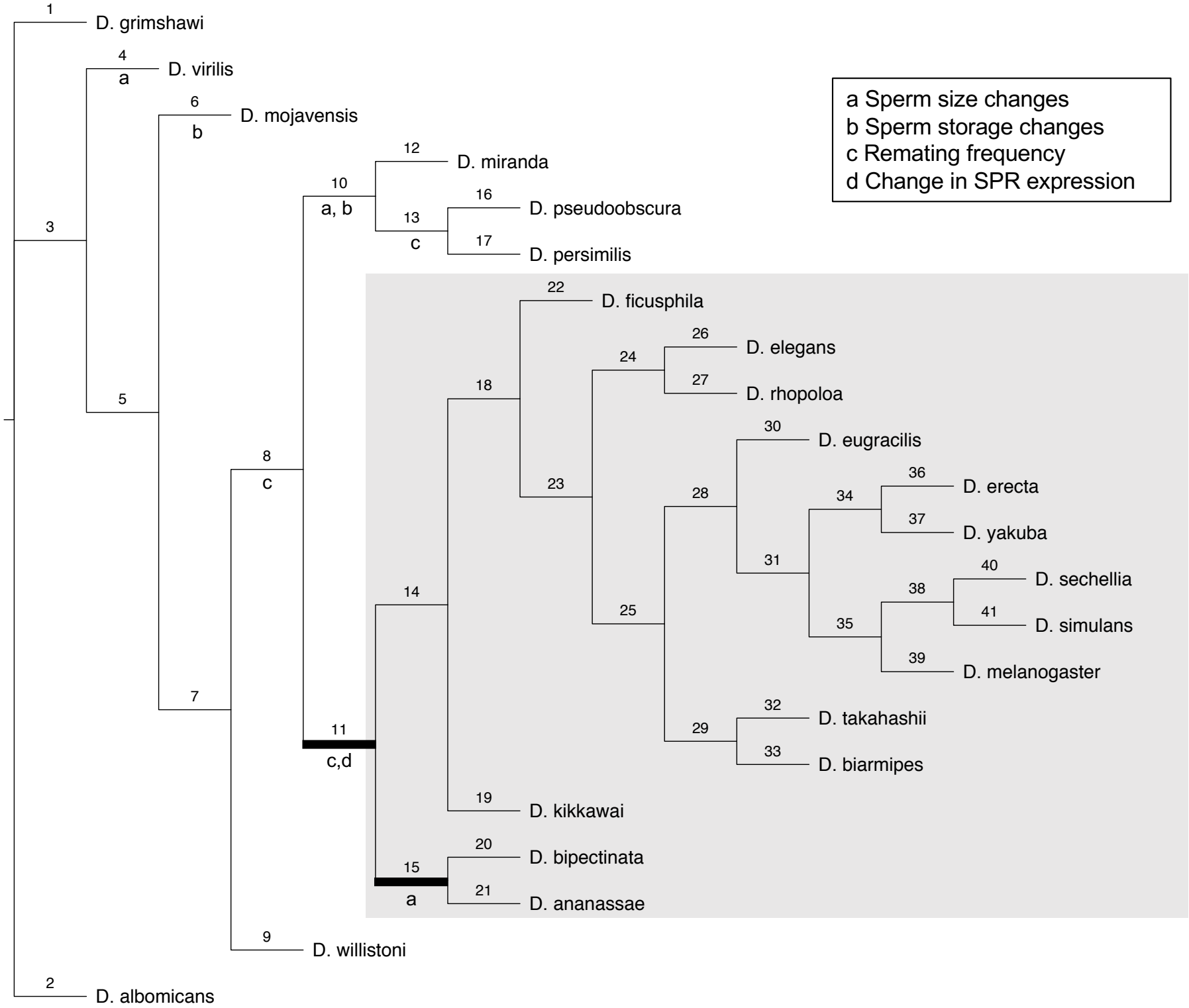
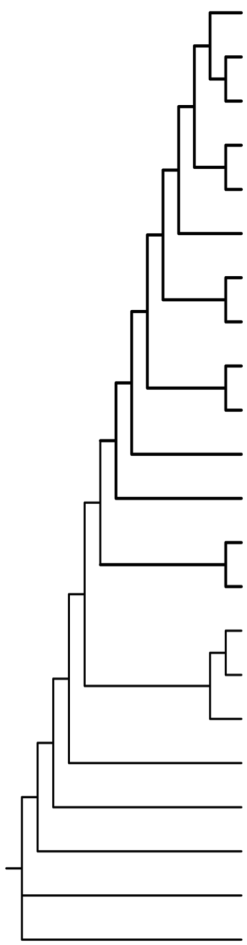
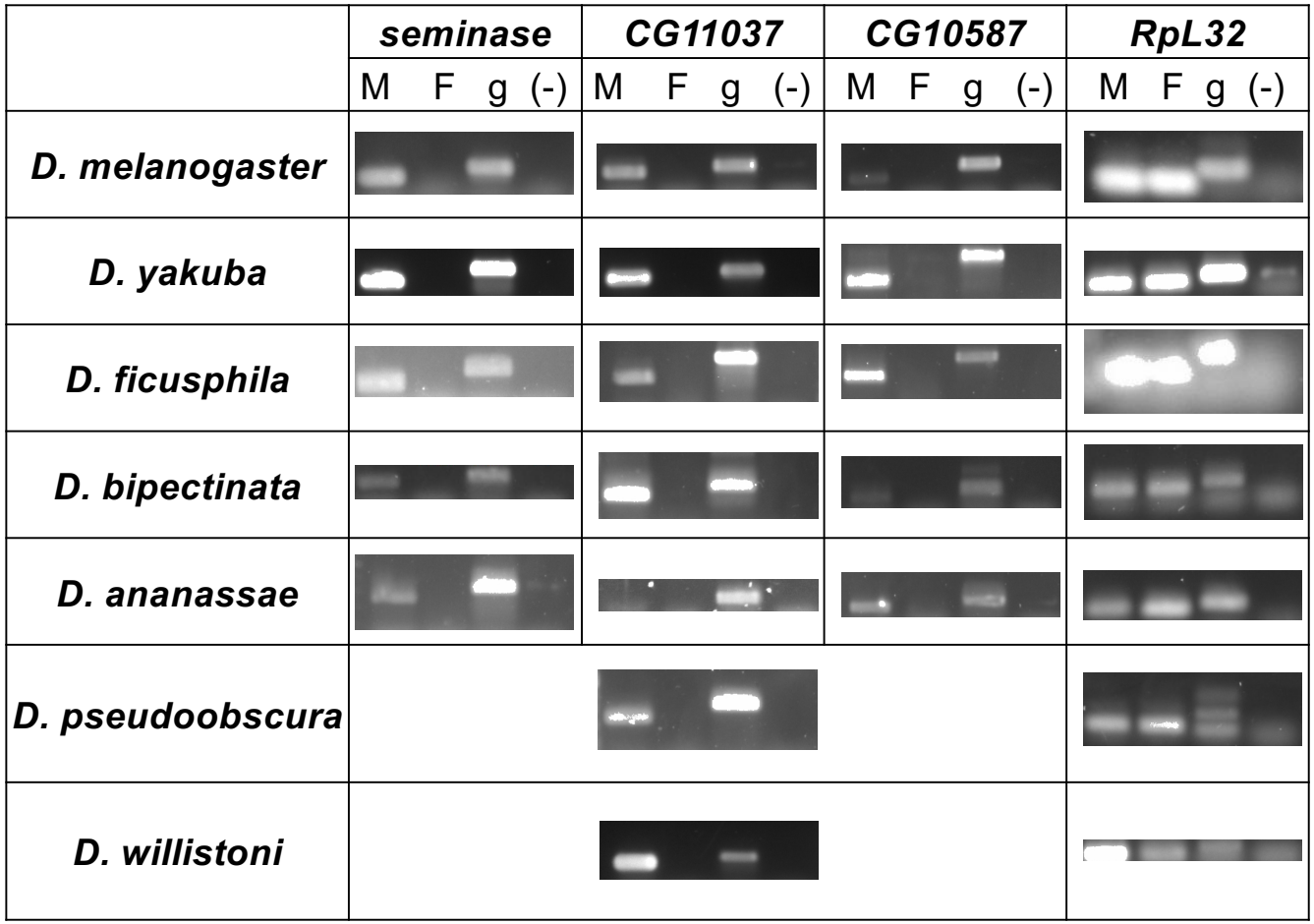

a Sperm size changes
b Sperm storage changes
c Remating frequency
d Change in SPR expression
1
D. grimshawi
4
a
D. virilis
6
b
D. mojavensis
12
D. miranda
10
a, b
16
D. pseudoobscura
13
c
17
D. persimilis
3
22
D. ficusphila
26
D. elegans
24
27
D. rhopoloa
18
5
30
D. eugracilis
8
c
36
D. erecta
23
28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

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