SEMINAL FLUID PROTEIN DIVERGENCE AMONG POPULATIONS EXHIBITING POSTMATING PREZYGOTIC REPRODUCTIVE ISOLATION

ARTICLE TYPE: LETTER

RUNNING TITLE: Proteomics of postmating prezygotic isolation

Martin D. Garlovsky[1*], Caroline Evans[2], Mathew A. Rosenow[3], Timothy L. Karr[4] & Rhonda R. Snook[5]

[1]Department of Animal and Plant Sciences, The University of Sheffield, UK

[*]Current address: Department of Biology, Syracuse University, Syracuse NY, USA

[2]Department of Chemical and Biological Engineering, The University of Sheffield, UK

[3]Caris Life Sciences, Phoenix, USA

[4]Centre for Mechanisms of Evolution, The Biodesign Institute, Arizona State University, USA

[5]Department of Zoology, Stockholm University, Sweden

Corresponding author: martingarlovsky@gmail.com, ORCiD: 0000-0002-3426-4341

KEYWORDS: Postcopulatory sexual selection, seminal fluid proteins, postmating prezygotic isolation, speciation, proteomics, tandem mass-spectrometry, accessory glands, Drosophila, reproductive isolation, ejaculate

Abstract word count: 270

Total word count: 5085

ABSTRACT

Despite holding a central role for fertilisation success, reproductive traits often show elevated rates of evolution and diversification. The rapid evolution of seminal fluid proteins (Sfps) within populations is predicted to cause mis-signalling between the male ejaculate and female reproductive tract between populations resulting in postmating prezygotic (PMPZ) isolation. Crosses between populations of *Drosophila montana* show PMPZ isolation in the form of reduced fertilisation success in both noncompetitive and competitive contexts. Here we test whether male ejaculate proteins deriving from either the accessory glands or the ejaculatory bulb differ between populations using liquid chromatography tandem mass spectrometry. We find more than 150 differentially abundant proteins between populations which may contribute to PMPZ isolation. These proteins include a number of proteases and peptidases, and several orthologs of *D. melanogaster* Sfps, all known to mediate fertilisation success and which mimic PMPZ isolation phenotypes. Males of one population typically produced greater quantities of Sfps and the strongest PMPZ isolation occurs in this direction. The accessory glands and ejaculatory bulb have different functions and the ejaculatory bulb contributes more to population differences than the accessory glands. Proteins with a secretory signal, but not Sfps, evolve faster than non-secretory proteins although the conservative criteria used to define Sfps may have impaired the ability to identify rapidly evolving proteins. We take advantage of quantitative proteomics data from three *Drosophila* species to determine shared and unique functional enrichments of Sfps that could be subject to selection between taxa and subsequently mediate PMPZ isolation. Our study provides the first high throughput quantitative proteomic evidence showing divergence of reproductive proteins implicated in the emergence of PMPZ isolation between populations.

IMPACT SUMMARY (word count: 300)

Identifying traits that prevent successful interbreeding is key to understanding early stages of the formation of new species, or speciation. Reproductive isolation arising prior to and during fertilisation frequently involves differences in how the sexes interact. In internally fertilising taxa, such interactions are mediated between the female reproductive tract where fertilisation occurs and the receipt of the ejaculate necessary for fertilisation. Because ejaculate proteins are at least partially responsible for these interactions, differences in male ejaculate protein composition could negatively impact fertilisation success, generating reproductive isolation. While the biological classes of ejaculate proteins are shared across all animal taxa, proteins that are secreted by males tend to show rapid evolution in gene expression and genetic sequence. Thus, reproductive proteins are suggested as prime targets facilitating reproductive isolation that arises after mating but before fertilisation (PostMating PreZygotic or PMPZ isolation). Most research on PMPZ isolation has focussed on differences between species for which it is not possible to determine the causative and temporal order of early speciation processes. Here, we test whether populations that exhibit few genetic differences but show strong PMPZ isolation also exhibit variation in ejaculate composition using quantitative high throughput proteomic analyses. We find a number of proteins are differentially abundant between populations including several known to impact fertilisation success in other species. We show that secreted proteins are evolving at an elevated rate, implicating their potential role in PMPZ isolation. We test divergence in ejaculate composition between species, finding a core set of functions that were conserved across species which last shared a common ancestor more than 40 million years ago along with species-specific investment. This work highlights the divergent evolution of reproductive proteins which may contribute to barriers between populations within a species early during speciation, extendable to similar analyses in other taxa in the future.

INTRODUCTION

For internally fertilising taxa the male ejaculate and female reproductive tract must interact during and after mating to ensure optimal fertility (Pitnick et al. 2009, 2020). In polyandrous species, ejaculate x female reproductive tract interactions are subject to rapid coevolution and diversification due to postcopulatory sexual selection (sperm competition and cryptic female choice) and sexually antagonistic coevolution (Birkhead and Pizzari 2002; Sirot et al. 2015; Firman et al. 2017; Meslin et al. 2017). Thus, despite holding a central role for fertilisation, ejaculate and female reproductive tract traits often show elevated rates of molecular and morphological evolution (Rowe et al. 2015; Ahmed-Braimah et al. 2017; Meslin et al. 2017; VanKuren and Long 2018; Simmons and Fitzpatrick 2019; McGeary and Findlay 2020). Divergence between populations in these traits is predicted to result in the early emergence of reproductive isolation that occurs after mating but before fertilisation (postmating prezygotic; PMPZ) (Lande 1981; Gavrilets 2000; Panhuis et al. 2001). Studies have increasingly documented PMPZ isolation, including prior to any postzygotic isolation (Howard et al. 2009; Bono et al. 2011; Sagga and Civetta 2011; Manier et al. 2013; Jennings et al. 2014; Cramer et al. 2016; Devigili et al. 2018; Garlovsky and Snook 2018; Turissini et al. 2018). In the *Drosophila melanogaster* subgroup, PMPZ isolation accumulates quickly measured by relative rates of evolution of different types of reproductive isolating mechanisms and suggests that PMPZ isolation is important in promoting new species and maintaining species barriers (Turissini et al. 2018). Despite the increasing recognition of the importance of PMPZ isolation, there is little understanding of the molecular basis of ejaculate x female reproductive tract interactions that may generate such barriers (McDonough et al. 2016).

The ejaculate consists of sperm and male seminal fluid proteins (Sfps) that impact fertilisation (e.g., Avila et al. 2011; South and Lewis 2011). Most Sfps are products of male secretory

glands (e.g. in mammals, the prostate gland and seminal vesicles; in arthropods, accessory glands and ejaculate ducts/bulb; for review, see Sirot et al. 2015; Fig. S1). Different secretory organs contribute distinct sets of proteins to the ejaculate allowing increased complexity and modulation or tailoring of the ejaculate (Bayram et al. 2019). The majority of work on insect Sfp evolution has been done on *Drosophila melanogaster* with over 200 Sfps identified (Mueller et al. 2005; Findlay et al. 2008, 2009). However, many of the biochemical classes of Sfps are similar across animals; for example, proteases and protease inhibitors, and those with antimicrobial/immune related functions (Avila et al. 2011; Sirot et al. 2015). Despite conserved protein classes observed in the seminal fluid of all animals, a large fraction of Sfps show rapid molecular evolution and, therefore, even Sfps of the same classes in different species are not orthologous (Avila et al. 2011; Perry et al. 2013; Sirot et al. 2015). Functional confirmation of Sfps, performed mostly in *D. melanogaster*, indicate they aid in sperm transfer and storage, influence the outcome of sperm competition, and/or alter female physiology, behaviour and reproductive tract morphology after mating (Ravi Ram and Wolfner 2007; Wong et al. 2008; Avila and Wolfner 2009; Holman 2009; Wigby et al. 2009; Wolfner 2009; Avila et al. 2011; Fedorka et al. 2011; Mattei et al. 2015). Moreover, some Sfps elicit PMPZ-like phenotypes when genetically manipulated (Ravi Ram and Wolfner 2007; LaFlamme et al. 2012). Differences in Sfp expression between species or abnormal gene expression profiles in the female reproductive tract after mating with hetero- vs. con- specific males are associated with PMPZ isolation (Bono et al. 2011; Ahmed-Braimah et al. 2017). These shared and divergent patterns support Sfps as putative causative agents of PMPZ isolation.

However, while studies showing divergence in gene expression are associated with disrupted ejaculate x female reproductive tract interactions, changes in gene expression may not correlate with changes in protein abundance (Wang et al. 2019), where the molecular

126 interactions causing PMPZ isolation take place. Divergence in protein identity or abundance
127 between taxa could disrupt ejaculate x female reproductive tract interactions leading to PMPZ
128 isolation (Goenaga et al. 2015). High-throughput proteomics using liquid chromatography
129 tandem mass spectrometry (LC-MS/MS) has revolutionised identification and quantification of
130 Sfps, revealing that the male ejaculate often contains hundreds of unique proteins
131 (McDonough et al. 2016; Bayram et al. 2019; Karr 2019; Rowe et al. 2019; Whittington et al.
132 2019). Using LC-MS/MS combined with genomics, Sfps can be predicted by identifying
133 ejaculate proteins with a signal peptide sequence, sometimes called the "secretome", and
134 those secretome proteins that have an extracellular signal sequence, sometimes called the
135 "exoproteome" (Avila et al. 2011; Ahmed-Braimah et al. 2017; Bayram et al. 2019; Karr et al.
136 2019; Sepil et al. 2019). Molecular evolution analyses show that proteins with secretory signal
137 and/or are secreted evolve faster than non-secreted ejaculate proteins (Mueller et al. 2005;
138 Wagstaff and Begun 2005; Ramm et al. 2009; Bono et al. 2015; Tsuda et al. 2015; Ahmed-
139 Braimah et al. 2017; Karr et al. 2019)
140
141 While these results suggest Sfps are prime candidates for generating PMPZ isolation, no study
142 using high throughput quantitative proteomics has tested the prediction that ejaculate
143 composition will vary between populations exhibiting PMPZ isolation, and that these proteins
144 will more rapidly evolve. Similarly, while previous work has suggested that the different
145 ejaculate secretory organs may perform different functions (Bayram et al. 2019), their potential
146 contribution to PMPZ isolation has not been examined. Here we use LC-MS/MS on the
147 accessory glands and the ejaculatory bulb/duct, followed by molecular evolutionary rates
148 analysis, to test these predictions using the malt fly, *Drosophila montana*. We have focused on
149 two populations (Crested Butte, Colorado; Vancouver, Canada) and found that reciprocal
150 crosses between populations result in PMPZ isolation both after a single mating as males

transfer sperm to females who store them, but many eggs are not fertilised (Jennings et al. 2014; Garlovsky and Snook 2018), and where within- and between- population males compete for fertilisation (frequently referred to as conspecific sperm precedence) (Garlovsky et al. in review). The similar results suggest a shared mechanism generating PMPZ isolation. Genomic analysis found no fixed SNPs between these populations, likely due to a history of gene flow during divergence (Parker et al. 2018; Garlovsky et al. in review), although genes enriched for reproductive function show divergence (Parker et al. 2018). These results support focussing on Sfps as potential causative agents of PMPZ isolation. We also leverage recent high throughput mass spectrophotometry data on ejaculate composition in two other *Drosophila* species (Karr et al. 2019; Sepil et al. 2019) to provide insights into shared and divergent Sfp functions that may contribute to PMPZ isolation.

METHODS

Fly stocks

Adult *Drosophila montana* were collected with malt bait buckets and mouth aspirators in Crested Butte, Colorado, USA (38°49'N, 107°04'W) in 2013 (referred to as Colorado), and Vancouver, British Columbia, Canada (48°55'N, 123°48'W) in 2008 (referred to as Vancouver) (Fig. S1). Stocks were established by combining 20 F3 males and females from 20 isofemale lines (800 flies total per population) and cultured on Lakovaara malt media (Lakovaara 1969) in overlapping generations in constant light at 19ºC. Flies were collected within 3 days of eclosion and housed in groups of between 10-20 single sex individuals in food vials until reproductively mature at 21 days old.

176 *Tissue collection and protein extraction*

177 Twenty-one-day old males were anaesthetised with ether and the accessory glands and
178 ejaculatory duct/bulb separated from nontarget tissues, and from each other. We collected
179 three biological replicates, two of which were separated into technical replicates (Fig. S1).
180 Following protein extraction and purification, we quantified protein concentration to standardize
181 loading 5µg of protein for each sample into the mass spectrometer (see supplementary
182 material; Fig. S2). Samples were reduced with TCEP, alkylated by addition of MMTS, and
183 digested with trypsin, followed by drying to completion using vacuum centrifugation. Samples
184 were resuspended in 20µl 3% v/v acetonitrile, 0.1% v/v trifluoroacetic acid prior to LC-MS/MS
185 analysis.

186

187 *LC-MS/MS analysis*

188 Detailed description of LC-MS/MS data acquisition and processing can be found in the
189 supplementary material.

190 LC-MS/MS was performed by nano-flow liquid chromatography (U3000 RSLCnano, Thermo
191 Fisher™) coupled to a hybrid quadrupole-orbitrap mass spectrometer (QExactive HF, Thermo
192 Scientific™). Peptides were separated on an Easy-Spray C_{18} column (75 µm x 50 cm) using a
193 2-step gradient from 97% solvent A (0.1% formic acid in water) to 10% solvent B (0.08% formic
194 acid in 80% acetonitrile) over 5 min then 10% to 50% B over 75 min at 300 nL/min. The full
195 105-minute MS data dependent acquisition was set up from 375-1500 m/z acquired in the
196 Orbitrap in profile mode, resolution 120,000. Subsequent fragmentation was Top 10 in the HCD
197 cell, with detection of ions in the Orbitrap using centroid mode, resolution 30,000. MS
198 parameters; MS1: Automatic Gain Control (AGC) target 1e6 with a maximum injection time (IT)
199 of 60 ms; MS2: AGC target 1e5, IT of 60 ms and isolation window 2 Da.

200

We performed label free quantitative proteomic analysis using MaxQuant to generate relative peptide and protein intensities (Cox et al. 2014; Tyanova et al. 2016) (see supplementary material). For protein identification we matched mass spectra to the *D. montana* predicted proteome (Parker et al. 2018), generated using gene predictions from the Maker2 pipeline (Holt and Yandell 2011) reciprocally blasted against *D. virilis* proteins (Parker et al. 2018).

Gene Ontology (GO) and functional analysis

We performed network analyses and GO enrichment for Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF) with the ClueGO plugin (Bindea et al. 2009) for Cytoscape (Shannon et al. 2003) using FlyBase gene numbers (FBgns) for *D. virilis* orthologs of *D. montana* genes retrieved from Parker et al. (2018) or *D. melanogaster* orthologs converted via FlyBase.org. Specific settings for network groups are provided in figure and table legends. For GO enrichment we used right-sided hypergeometric tests with Benjamini-Hochberg multiple test correction.

Differential abundance analysis between D. montana populations and functional differences between tissues

We performed differential abundance analysis of MaxLFQ ion intensities using the 'edgeR' (Robinson et al. 2010) and 'limma' (Ritchie et al. 2015) packages in R (v.3.5.1) (R Core Team 2018) (see supplementary material). Proteins were considered differentially abundant based on an adjusted p-value < 0.05. To identify differentially abundant proteins between populations, we analysed the accessory gland proteome and the ejaculatory bulb proteome separately. We only considered proteins that were present in all five replicates of each tissue for both populations. To identify differentially abundant proteins between tissues, we analysed each

population separately. Again, we only considered proteins that were present in all five replicates of each population for both tissues. (Table S1).

Characterising the male seminal fluid proteome across species

We compared Sfp functions for three *Drosophila* species for which proteomic data generated using LC-MS/MS is available for the male accessory gland and ejaculatory duct and bulb tissues: *D. montana* (this study), *D. melanogaster* (Sepil et al. 2019), and *D. pseudoobscura* (Karr et al. 2019). We retrieved FBgns for *D. melanogaster* genes identified by Sepil et al. (2019) and *D. melanogaster* orthologs for *D. pseudoobscura* genes identified by Karr et al. (2019) and downloaded the corresponding canonical protein sequences from uniprot.org. For proteins we identified in our analysis we retrieved *D. montana* protein sequences from Parker et al. (2018). We submitted protein sequences for each species to *SignalP* (Petersen et al. 2011) and *Phobius* (Käll et al. 2004) and combined the resulting lists of proteins containing a signal peptide to generate a list of secretome proteins for each species. For *D. montana* we converted the corresponding *D. virilis* FBgns for each protein to *D. melanogaster* orthologs via FlyBase.org (for 215/245, 88%). To identify Sfps for each species we submitted secretome lists to FlyBase.org to retrieve genes with GO terms containing “extracelluar” (Fig. S3; Table S2). To compare GO enrichment between species we adjusted network settings in ClueGO to reflect the different numbers of proteins identified in each species.

Evolutionary rates analysis

To obtain sequence divergence estimates for *D. montana* proteins we used a pipeline developed previously (Wright et al. 2015). We obtained protein coding sequences for *D. montana* from Parker et al. (2018) and for *D. pseudoobscura* (r3.04, September 2019) and *D. virilis* (r1.07, August 2019) from FlyBase.org. We identified the longest isoform of each gene

for each species and determined orthology with reciprocal BLASTN (Altschul et al. 1990), using a minimum percentage identity of 30% and an E-value cut-off of $1x10^{-10}$. We then identified reciprocal one-to-one orthologs across all three species using the highest BLAST score. We identified open reading frames using BLASTx and aligned orthologs using PRANK (Löytynoja and Goldman 2010). We calculated the ratio of non-synonymous (dN) to synonymous (dS) nucleotide substitutions, omega (ω), using the CODEML package in PAML (Yang 2007) (one-ratio estimates, model 0) with an unrooted phylogeny. Results were filtered to exclude orthologs with branch-specific $dS \geq 2$ (due to potential mutational saturation) or where $S*dS \leq 1$.

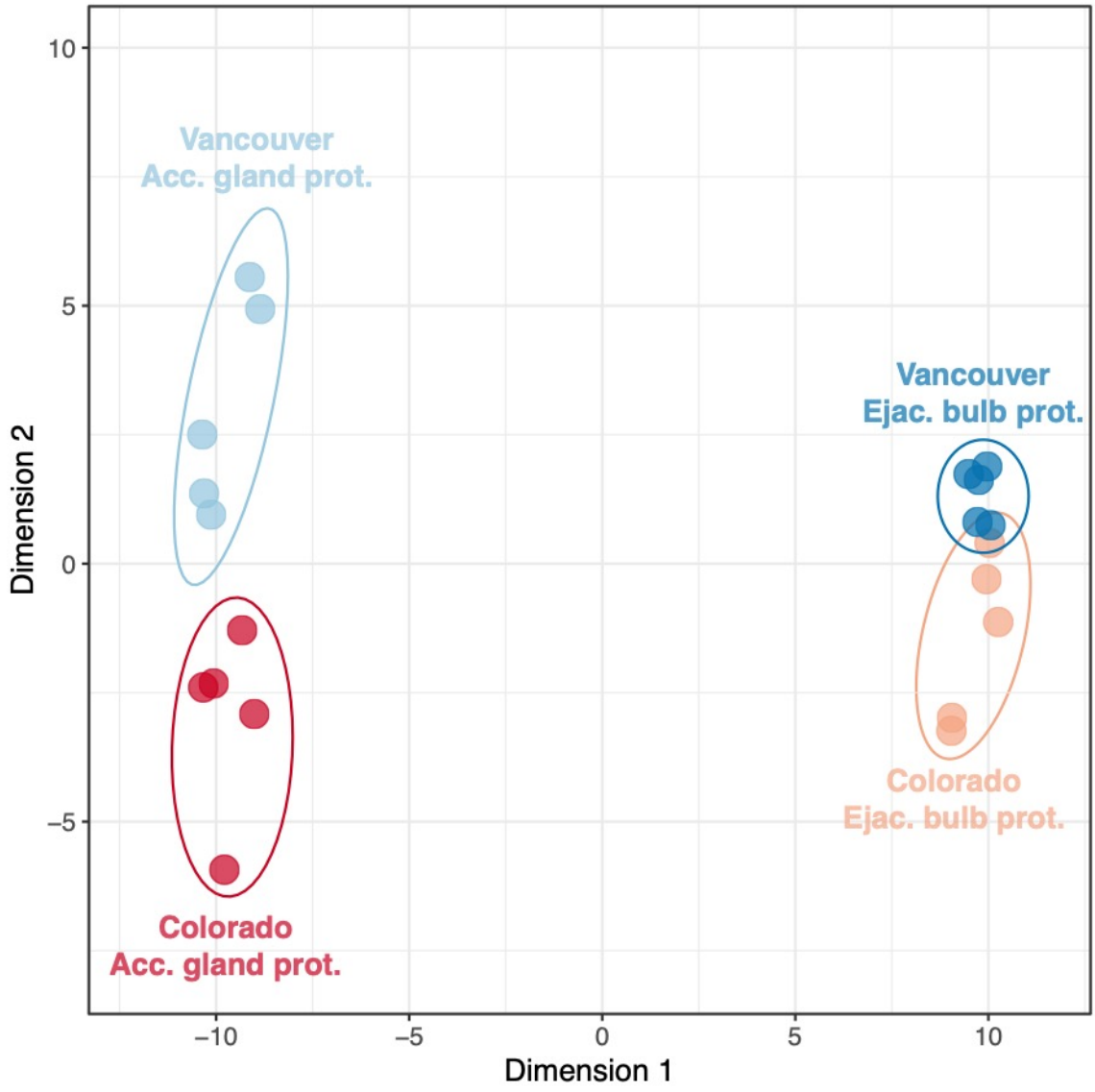
We then tested for differences in evolutionary rates between sets of proteins we identified in our LC-MS/MS analysis. We relaxed filtering criteria so that a protein need only be identified in a single replicate in a single population or tissue, but still had to be identified by two or more unique peptides. After filtering, we obtained ω values for 757/1474 (51%) proteins with a reciprocal one-to-one ortholog. We categorised genes as having higher abundance in either the accessory gland proteome or ejaculatory bulb proteome as those proteins showing concordant differential abundance between tissues across populations (Fig. 3). We classified proteins as belonging to the secretome based on presence of a signal peptide and as Sfps as secretome proteins with extracellular annotation (see above) plus *D. melanogaster* Sfps. We classified the remainder of proteins excluding those with higher abundance in the accessory gland proteome, ejaculatory bulb proteome, secretome, or Sfps, as background tissue. Each class consisted of a unique set of proteins, such that accessory gland proteins did not include ejaculatory bulb proteins, secretome, Sfps, or remaining proteins; secretome proteins did not include Sfps, etc. We tested for differences in evolutionary rates using the Kruskal-Wallis rank

sum test followed by pairwise Wilcox rank sum tests corrected for multiple testing using the Benjamini-Hochberg method.

RESULTS

The D. montana *accessory gland proteome and ejaculatory bulb proteome*

We identified 1711 proteins, of which 1474 (86%) were identified by two or more unique peptides. The majority of proteins (1013/1474; 69%) were shared across male secretory tissues, while 138 (9%) and 323 (22%) proteins were unique to the accessory glands and ejaculatory bulb, respectively (Fig. S4a). Proteins identified only in the accessory gland proteome showed a 3.2-fold lower mean abundance compared to the remainder of proteins whereas proteins identified only in the ejaculatory bulb proteome showed a 14.9-fold reduction. These proteins likely represent missed rather than truly unique proteins and are not considered further. We identified 79 *D. montana* Sfps, consisting of 38 orthologs of *D. melanogaster* Sfps identified by converting *D. virilis* FBgns on FlyBase.org, plus 55 secretome proteins with extracellular annotations identified by 2 or more unique peptides (14 of which overlapped) (Fig. S4a; Table S3) (Mueller et al. 2005; Findlay et al. 2008, 2009). A multidimensional scaling (MDS) plot of normalised intensities using all proteins (n = 1474) showed a clear separation of samples by tissue type (dimension 1), and separation by population (dimension 2) with clear separation of populations for the accessory gland proteome and marginal overlap between populations in the ejaculatory bulb proteome (Fig. 1).



294

295 Figure 1. Multidimensional scaling (MDS) plot of normalised intensities for proteins identified
296 by two or more unique peptides (n = 1474) in each replicate (points). Dimension 1 separates
297 the two tissue types (accessory glands and ejaculatory bulbs). Dimension 2 separates the two
298 populations (Colorado and Vancouver).

299

300 *Differential abundance of reproductive proteins between populations*

301 The majority of proteins were identified in both populations (1322/1474; 90%), while 45 (3%)
302 and 107 (7%) were only identified in Colorado, and Vancouver, respectively (Fig. S4b).
303 Proteins only identified in one population showed a 263-fold, and 171-fold, lower mean

304 abundance compared to the rest of proteins in Colorado, and Vancouver, respectively. As
305 above, these low abundance proteins are not considered further. For shared proteins, we then
306 tested for differential abundance. We identified 154 (out of 725) differentially abundant proteins
307 produced in the accessory glands between populations (Fig. 2a), including nine *D.*
308 *melanogaster* Sfps (Table 1). We identified 244 (out of 929) differentially abundant proteins
309 produced in the ejaculatory bulbs (Fig. 2b). Again, this included nine *D. melanogaster* Sfps,
310 two of which overlapped with those identified in the accessory gland proteome (Table 1). In the
311 accessory gland proteome, Sfps and proteins with a predicted secretory signal were not
312 overrepresented in the cohort of proteins showing differential abundance (Chi-squared test, X^2
313 = 1.57, df = 2, p = 0.456) but were overrepresented in the cohort of differentially abundant
314 proteins in the ejaculatory bulb proteome (X^2 = 44.56, df = 2, p < 0.001; Fig. S5). Out of 45
315 proteins that were differentially abundant between populations in both male reproductive
316 tissues, 36 showed higher abundance in one population in both tissues (Fig. S6). Significantly
317 enriched gene ontology (GO) categories for proteins showing differential abundance between
318 populations are in Tables S4-S6.
319

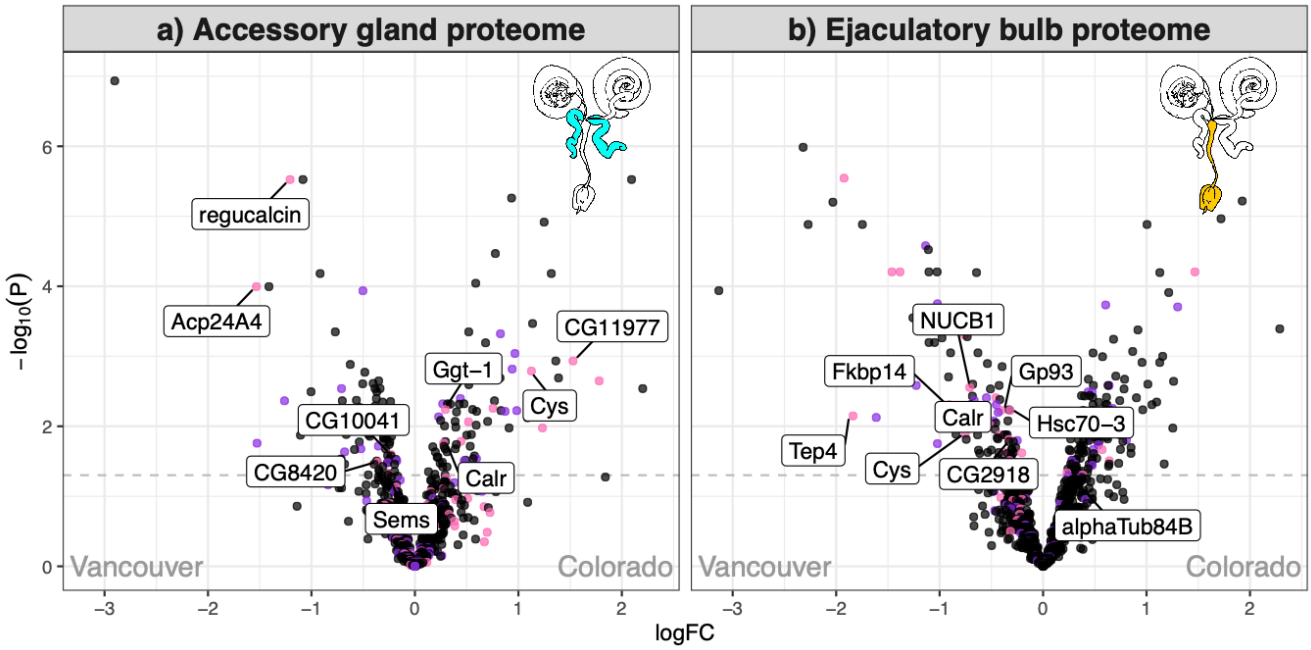


Figure 2. Differentially abundant proteins between Colorado and Vancouver in a) the accessory glands (n = 725) and b) the ejaculatory bulbs (n = 929). Secretome proteins are shown in purple and Sfps in pink. Significantly differentially abundant proteins with a known Sfp ortholog in *D. melanogaster* are labelled.

Table 1. Differentially abundant proteins between *D. montana* populations with a known *D. melanogaster* seminal fluid protein (Sfp) ortholog. Gene names were retrieved from FlyBase.org using the corresponding *D. virilis* FBgns for *D. montana* proteins we identified via LC-MS/MS. The population for which each protein showed higher abundance is given. Proteins found in both tissue comparisons indicate in which population there was higher abundance. Abbreviations: Acgs, accessory glands; Ebs, ejaculatory bulbs.

Tissue comparison	Gene name	Higher abundance
Accessory glands	CG11977	Colorado
	gamma-glutamyl transpeptidase	Colorado
	Acp24A4	Vancouver
	CG10041	Vancouver
	CG8420	Vancouver
	Regucalcin	Vancouver
	Seminase	Vancouver
Ejaculatory bulbs	alpha-Tubulin at 84B	Colorado
	CG2918	Vancouver
	FK506-binding protein 14	Vancouver
	Glycoprotein 93	Vancouver
	Heat shock 70-kDa protein cognate 3	Vancouver
	NUCB1	Vancouver
	Thioester-containing protein 4	Vancouver
Both	Calreticulin	Colorado^{Acgs}/Vancouver^{Ebs}
	Cystatin-like	Vancouver^{Acgs}/Colorado^{Ebs}

The accessory gland and ejaculatory bulb proteomes differ in function

To test whether the accessory glands and ejaculatory bulb provide different functions we performed differential abundance analysis between tissues for Colorado and Vancouver separately. We found 524 (out of 652) differentially abundant proteins between tissues in Colorado. Similarly, in Vancouver we found 557 (out of 676) differentially abundant proteins. The majority of these proteins were found in both populations (609 proteins). To identify consistently differentially abundant proteins between tissues, we compared the log2-fold change in abundance in each population of these 609 proteins. Proteins with higher abundance

in the accessory gland proteome or ejaculatory bulb proteome in Colorado generally also showed higher abundance in Vancouver (Spearman's rank correlation, $\rho = 0.945$, $p < 0.001$, n = 609) (Fig. 3a). GO analyses identified both tissues as having functions expected for highly metabolically active secretory organs (Table S7). Different GO terms were enriched in each tissue highlighting that the two secretory organs provide distinct reproductive functions to the ejaculate (Fig. 3b; Table S7).

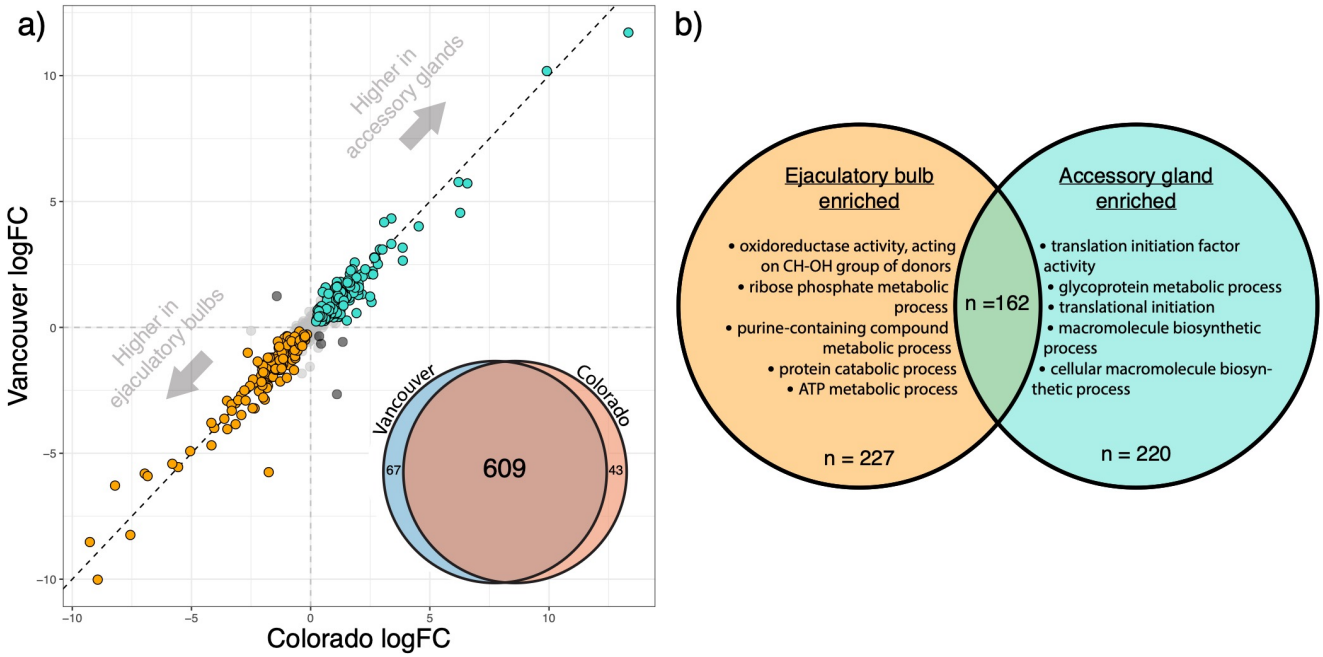
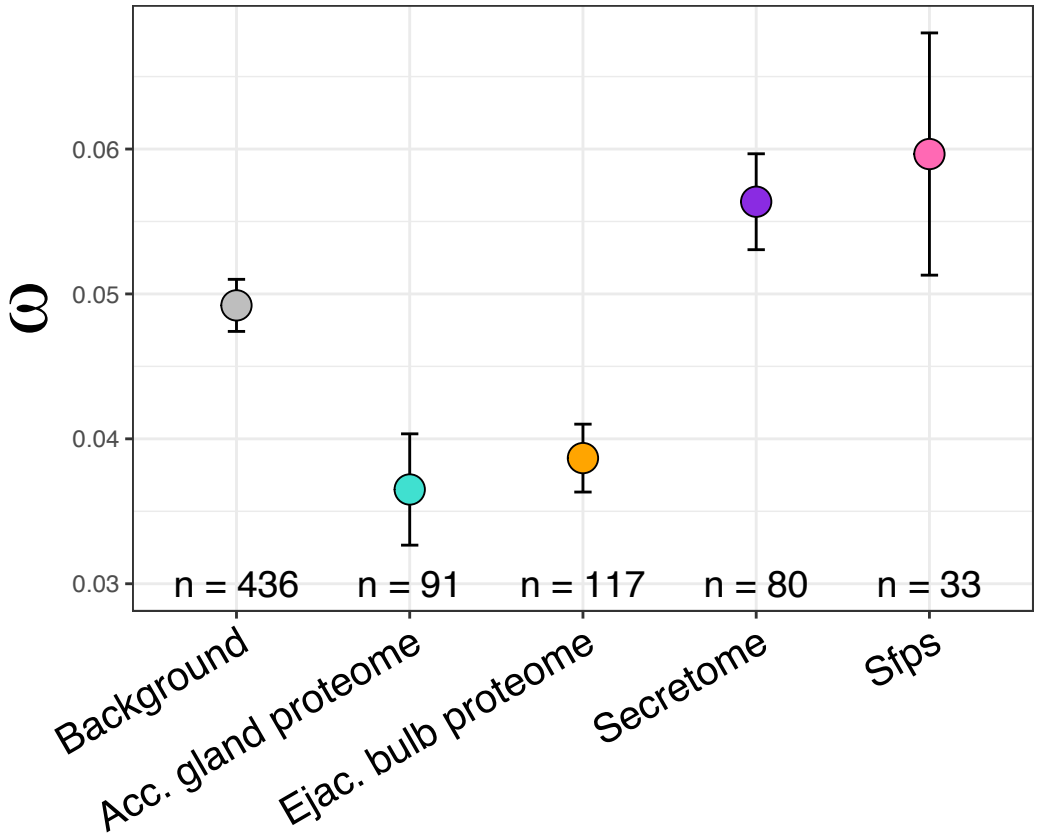


Figure 3. Populations show consistent differential abundance between secretory organs. a) Concordance between populations in log2-fold change in abundance of proteins found in either the accessory gland proteome or ejaculatory bulb proteome. Positive values indicate proteins with higher abundance in the accessory glands in both populations (top right), negative values indicate proteins with higher abundance in the ejaculatory bulb (bottom left). Proteins are coloured based on whether they showed a concordant pattern of significantly higher abundance in the accessory gland proteome (turquoise, n = 220), the ejaculatory bulb proteome (orange, n = 227), were discordant (black, n = 5), or were not significantly differentially abundant between tissues (grey). Dashed black line shows 1:1. Inset: Venn

Diagram showing numbers of proteins included in separate differential abundance analysis between tissues in each population and overlap. b) Top 5 significantly enriched Biological Process GO terms ranked by percent identity of proteins to each tissue (see Table S7 for full list).

Evolutionary rates analysis

We tested whether genes with higher protein abundance in either the accessory gland proteome, ejaculatory bulb proteome (excluding the secretome and Sfps), secretome (excluding Sfps), or Sfps (secretome proteins with extracellular annotation), were evolving at different rates compared to background proteins, defined as those proteins that do not differ in protein abundance between the accessory glands and ejaculatory bulb and excluding secretome and Sfps. There was a significant difference between protein groups in evolutionary rates (Kruskal-Wallis test, X^2 = 40.51, df = 4, $p < 0.001$) (Fig. 4). The secretome and Sfps were evolving at similar rates (pairwise Wilcox rank sum test with Benjamini-Hochberg adjustment, p = 0.668). The secretome was evolving faster than proteins with higher abundance in the accessory gland proteome, ejaculatory bulb proteome, or background (all $p < 0.007$). Sfps were also evolving faster than proteins with higher abundance in the accessory gland proteome (p = 0.001) and ejaculatory bulb proteome (p = 0.007) but at a similar rate to background (p = 0.171). Proteins with higher abundance in the accessory gland proteome and ejaculatory bulb proteome were evolving at similar rates (p = 0.089), and slower than the remaining background proteome (accessory gland proteome vs. background, $p < 0.001$; ejaculatory bulb proteome vs. background, p = 0.011).



391

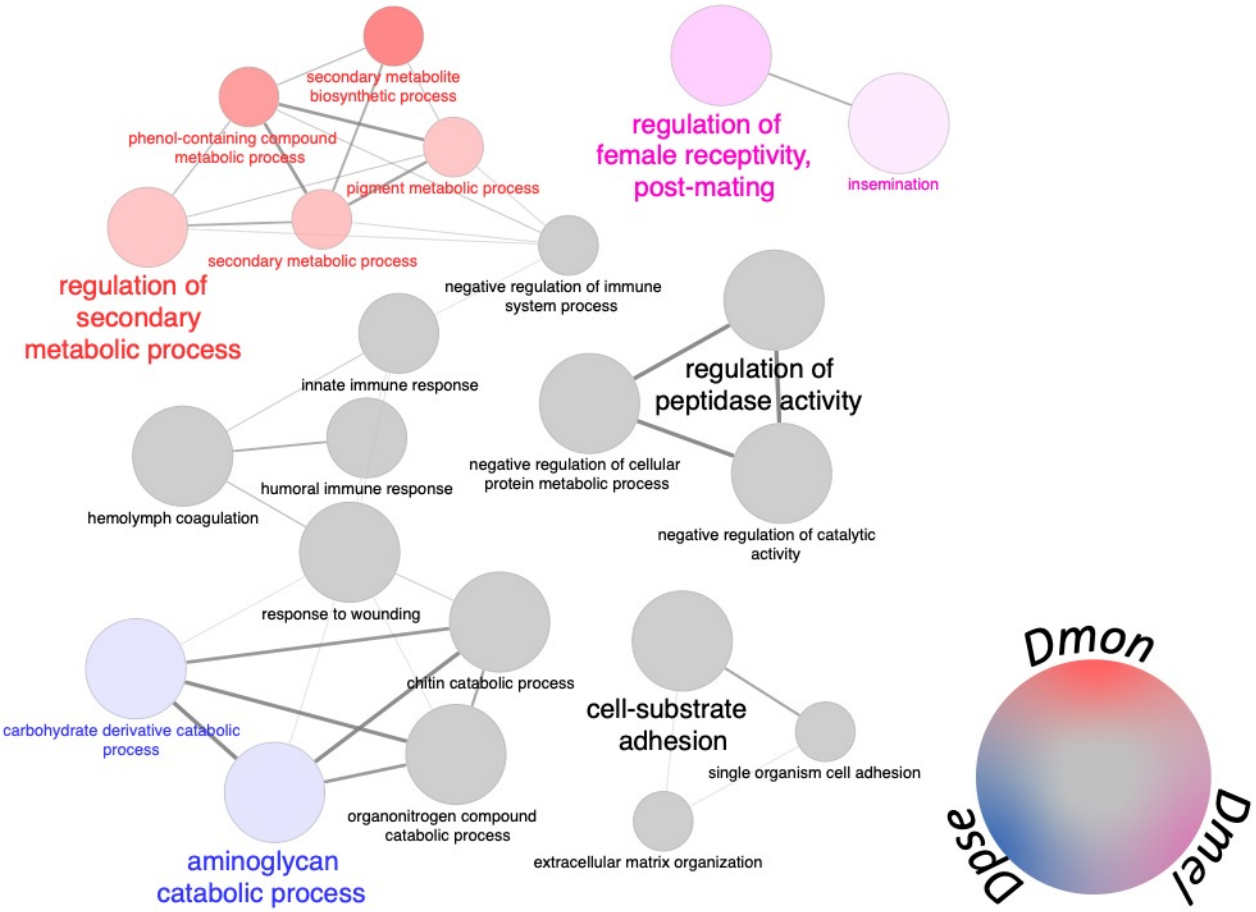
392 Figure 4. Mean non-synonymous (dN) to synonymous (dS) base substitution rate (ω) estimates
393 (± standard error) for *D. montana* genes identified in our LC-MS/MS analysis with equal
394 abundance in the accessory gland proteome and ejaculatory bulb proteome ('background';
395 grey), higher abundance in the accessory gland proteome (turquoise), ejaculatory bulb
396 proteome (orange), or found in the secretome (purple) or Sfps (pink). See Fig. S7 for separate
397 dN and dS plots.

398

399 *Comparison of male Sfps across species*

400 We identified 61 Sfps (secretome proteins with extracellular annotations) for *D. montana*, 249
401 Sfps for *D. melanogaster*, and 131 Sfps for *D. pseudoobscura* (Fig. S3; Table S2). Comparing
402 functional enrichment of Sfps across species identified a number of shared and unique GO
403 categories. Shared Biological Processes included chitin catabolic process, innate immune
404 response, cell-substrate adhesion, and regulation of peptidase activity (Fig. 5; see Table S8

405 for CC and MF terms). Uniquely enriched BP functions included regulation of secondary
406 metabolic process (*D. montana*); postmating regulation of female receptivity (*D. melanogaster*)
407 and amino glycan catabolic processes (*D. pseudoobscura*) (Fig. 5; see Table S8 for CC and
408 MF terms).
409



410

411 Figure 5. Seminal fluid protein (Sfps) comparisons for GO BP terms across species. Circle size
412 is associated with level of significance with increasing size indicating increasing significance.
413 Node colour indicates proportion of genes from each species associated with a term: *D.*
414 *montana* (red. Dmon), *D. melanogaster* (pink; Dmel), and *D. pseudoobscura* (blue; Dpse),
415 shared terms are shown in grey. Min. GO level = 3, max. GO level = 8. Number of genes/%
416 genes per group: *D. montana* 3/3%, *D. pseudoobscura* 6/6%, *D. melanogaster* 12/12%.
417 Percentage significance = 55%, kappa-score threshold = 0.25.

DISCUSSION

The molecular basis of mechanisms underlying PMPZ isolation are poorly understood. Seminal fluid proteins are likely to contribute to PMPZ isolation due to their effects on sperm use, fertilisation success, and rapid divergent evolution. We used quantitative proteomics to identify proteins produced in the accessory glands and ejaculatory duct and bulb in populations exhibiting PMPZ isolation and found a number of differentially abundant proteins between populations including several orthologs of *D. melanogaster* Sfps. The accessory glands and ejaculatory bulb showed different functions and there were more differentially abundant proteins found in the ejaculatory bulb than the accessory glands. For proteins found in both populations, but in separate tissues, there was strong concordance in abundance between populations. We found secretome proteins evolved at a faster rate than non-secretome proteins, both those differentially abundant between male secretory organs and those showing similar abundance between male tissues. Sfps did not show elevated rates of molecular evolution, likely because identification by extracellular annotation required orthology in *D. melanogaster*. Despite shared Sfps, and a core set of shared Sfp biological processes across three *Drosophila* species, there was species-specific enrichment of Sfp function.

D. montana from Colorado and Vancouver show low genome-wide divergence and a history of gene flow (Parker et al. 2018; Garlovsky et al. in review), yet show enrichment of genes with reproductive function that are divergent between populations (Parker et al. 2018). Crosses between Colorado and Vancouver show reduced fertilisation success after a single mating (Jennings et al. 2014; Garlovsky and Snook 2018) and exhibit conpopulation sperm precedence (Garlovsky et al. in review). Between different *Drosophila* species, mechanisms causing PMPZ isolation include abnormal sperm transfer and displacement, or mismatches between sperm length and female tract morphology (Price et al. 2001; Manier et al. 2013).

443 Females receiving a foreign ejaculate comprising an abnormal Sfp complement might also
444 result in mismatched ejaculate x female reproductive tract interactions measured by gene
445 expression differences (Bono et al. 2011; Plakke et al. 2015), although variation between
446 species in ejaculate composition has not been quantified in those studies.
447
448 We identified a number of differentially abundant proteins between populations exhibiting
449 PMPZ isolation, including several orthologs of *D. melanogaster* Sfps. Intriguingly, 11 of 14 of
450 these proteins were more abundant in Vancouver males than Colorado males. PMPZ isolation
451 between *D. montana* populations is asymmetric, with matings between Vancouver males and
452 Colorado females having lower fertilisation success compared to the reciprocal cross
453 (Jennings et al. 2014; Garlovsky and Snook 2018). If Vancouver males transfer more of these
454 Sfps to their mates, then the chemical environment in the reproductive tract of Colorado
455 females may be mismatched, more so than the reciprocal cross.
456
457 Additionally, a number of proteases and peptidases (or inhibitors) were differentially abundant
458 between populations. Proteases and peptidases are central to reproduction across taxa,
459 regulating proteolytic activity and initiating cascades of interactions among downstream
460 proteins (LaFlamme et al. 2012; LaFlamme and Wolfner 2013; Plakke et al. 2015, 2019;
461 Bayram et al. 2017, 2019). Divergence in proteases has been implicated in PMPZ isolation
462 between other insect species in both the male ejaculate and female reproductive tract
463 secretions (Kelleher et al. 2007; Kelleher and Pennington 2009; Marshall et al. 2009, 2011;
464 Meslin et al. 2017; Al-Wathiqui et al. 2018; Plakke et al. 2019).
465
466 In *D. montana* females receive and store motile sperm from incompatible males, but fertilisation
467 success is reduced (Jennings et al. 2014). Failure to either properly orient sperm in storage

(Manier et al. 2013), release sperm from storage, or have sperm release coincide with ovulation (Mattei et al. 2015) could explain PMPZ in this system (Jennings et al. 2014). Some notable differentially abundant Sfps and proteases we identified, and their potential relationship to PMPZ isolation in *D. montana* are seminase, γ-glutamyl transpeptidase, and regucalcin. Seminase (CG10586), is a serine protease and a member of the Sex Peptide (SP) network (Singh et al. 2018). Seminase acts early in the SP network and is required to process other Sfps in the mated female essential for proper sperm storage (Acp36DE) and ovulation (ovulin) (LaFlamme et al. 2012; Singh et al. 2018). RNAi knockdown of seminase in male *D. melanogaster* results in failure of mated females to release sperm from the seminal receptacle (LaFlamme et al. 2012). γ-glutamyl transpeptidase (CG6461) functions to maintain a protective redox environment for sperm (Walker et al. 2006). Mismatches between the male ejaculate and the redox environment of the female reproductive tract in which sperm are stored could reduce fertilisation success as sperm subject to increased oxidative stress are less fertilisation competent (Reinhardt and Ribou 2013). Regucalcin (CG1803), a Ca^{+2} binding protein, may also play an anti-oxidative role and, in mammals, is hypothesized to have an anti-capacitation role for sperm (Pillai et al. 2017). One aspect of capacitation, hyperactivation, increases sperm motility which is important for sperm storage in *Drosophila* (Köttgen et al. 2011). Sperm motility behaviour and how this may affect release from storage is unknown. Regucalcin gene expression varies between *D. montana* populations and has been suggested as a cold tolerance gene in diapausing females (Vesala et al. 2012) although it's expression in males has not been studied. These examples provide strong candidates for eliciting PMPZ isolation and will be subject to future studies, for instance using CRISPR/Cas9 gene editing, to further understand the molecular interactions causing PMPZ isolation in *D. montana*.

Reproductive proteins are predicted to evolve rapidly, driven by postcopulatory sexual selection and sexual conflict (Sirot et al. 2015; Ahmed-Braimah et al. 2017; Firman et al. 2017; Meslin et al. 2017). We found that proteins showing secretory signals (secretome) evolve faster than proteins without this signal. Sfps and secretome proteins were evolving at a similar rate. However, Sfps were not evolving faster than proteins with similar abundance between male tissues, possibly due to showing greater variation than other categories despite having a higher mean rate. The requirement to have extracelluar annotation determined from work in *D. melanogaster* limits our ability to identify rapidly evolving Sfps in *D. montana*. Thus, the putative number of Sfps in *D. montana*, 79, is a conservative estimate. Isotopically labelling males to identify proteins transferred to females increased the number of identified *D. melanogaster* Sfps (Findlay et al. 2008, 2009). Future work on *D. montana* can use this technique to identify additional Sfps.

One goal of this work was to assess whether the different male reproductive secretory organs had different functions, which would not be possible using the heavy labelling technique. Our work provides one of the first proteomic descriptions of both major Sfp secretory organs in *Drosophila*. Previous research in seed beetles has shown that division of labour across secretory organs enables increased complexity to the ejaculate and potential for ejaculate tailoring (Bayram et al. 2019). Most proteins we identified were found in both tissues but showed higher abundance in either the accessory glands or ejaculatory duct and bulb, suggesting these organs provide different functions to the ejaculate. The accessory gland proteome was enriched for translation and biosynthetic processes, whereas the ejaculatory bulb proteome showed enrichment for mainly metabolic processes. We found secretome proteins and Sfps were significantly overrepresented in the set of differentially abundant proteins in the ejaculatory bulb proteome but not the accessory gland proteome, suggesting

517 the two male secretory organs may contribute differently to PMPZ isolation. Most past work on
518 Sfp evolution has focused on the accessory glands, which could skew understanding of not
519 only molecules involved in reproduction but those reproductive molecules that may elicit PMPZ
520 isolation.
521
522 We took advantage of recent accessory gland proteomes for three *Drosophila* species
523 generated using high throughput LC-MS/MS to characterise shared and enriched functions of
524 Sfps between species. Using the same identification criteria for all species (secretory sequence
525 and extracellular annotation), we identified a set of shared GO categories between species
526 that last shared a common ancestor 40 million years ago. This core set included immune-
527 related genes, which are associated with sexual conflict in *D. melanogaster* (Innocenti and
528 Morrow 2009). We also found species-specific GO-enrichment of Sfps, suggesting divergence
529 in how they contribute to male ejaculate function between species. Differences may reflect how
530 selection has targeted particular ejaculate traits in different mating systems (Markow 2002).
531 Differences will also reflect the use of *D. melanogaster* as the reference for GO annotation. For
532 instance, *D. melanogaster* showed enrichment for reproductive genes but Sfps in the other
533 species clearly have reproductive functions. It is likely that reproductive genes which
534 experience strong divergent selection may no longer resemble *D. melanogaster* genes. Our
535 work offers a first insight into the proteomic composition of male ejaculate characteristics
536 across species. As understanding of the molecular interactions between the sexes matures, it
537 will be important to determine whether shared or divergent functions between species are more
538 likely to contribute to PMPZ isolation and when during speciation such divergence occurs. Are
539 shared functions more likely to diverge within populations early during speciation or are Sfps
540 that already show some species specificity more likely to contribute to early PMPZ isolation?
541

542 Here we have tested whether reproductive proteins show differential abundance between
543 populations that exhibit PMPZ isolation. Our study has focussed on *Drosophila*, a model
544 system for studying the evolution of reproductive processes and the evolution of reproductive
545 isolation in metazoans. However, reproductive processes, classes of reproductive proteins,
546 and the action of PMPZ isolation across animals show similarities. For example, differentially
547 abundant proteins between *D. montana* populations we found included a number of proteases
548 or peptidases which are common and important mediators of reproductive processes in all
549 animals. Differentially abundant proteins also included several orthologs of *D. melanogaster*
550 Sfps with functions that may be similar to altered reproductive processes generating PMPZ
551 isolation in *D. montana*, such as noncompetitive gametic isolation and conspecific sperm
552 precedence (Jennings et al., 2014; Garlovsky et al. in review). These reproductive isolating
553 mechanisms are found in many other metazoan taxa (for brief review, see Turissini et al. 2018).
554 We also showed secretome proteins are evolving faster than other proteins found in the
555 accessory glands or ejaculatory duct and bulb, and at the same rate as Sfps. Such rapid
556 evolution is frequently attributed to sexual selection and sexual conflict, and these dynamic
557 processes may contribute to speciation (Gavrilets 2000; Panhuis et al. 2001). Male
558 reproductive secretory tissues had divergent functions with the ejaculatory bulb contributing
559 more differentially abundant proteins than the accessory glands, and the direction and severity
560 of asymmetrical PMPZ isolation mirrors differential abundance. We also identified shared and
561 species-specific GO enrichment of male reproductive proteins that influence reproductive
562 processes, although whether PMPZ isolation more likely arises due to divergence in one or the
563 other of these categories requires additional data. Democratisation of high throughput
564 proteomics will facilitate understanding the evolution of male reproductive proteins, their
565 influence on reproductive processes per se, and their contribution to reproductive isolation.
566

ACKNOWLEDGEMENTS

Thanks to Priyanka Prajapati and Alexander Charles for assistance with proteomics data analysis, and Irem Sepil and the organisers and attendees of the Joint Wellcome Trust/EMBL-EBI Proteomics Bioinformatics course 2017 for discussion about analyses. Alison Wright, Daniela Palmer, Henry Barton and Toni Gossman provided helpful discussion and scripts for performing evolutionary rates analysis. We are grateful to Anneli Hoikkala for providing fly stocks, Michael G. Ritchie and Darren J. Parker for access to genomic and proteomic resources, and Roger Butlin for valuable feedback throughout the project. MDG was able to attend the Wellcome Trust/EMBL-EBI Proteomics Bioinformatics course thanks to a University of Sheffield Postgraduate Research Experience Programme (PREP) grant and was supported by the Adapting to the Challenges of a Changing Environment (ACCE) Doctoral Training Partnership grant NE/L002450/1, funded by the Natural Environment Research Council (NERC). Costs for proteomics was funded by a Royal Society Leverhulme Trust Senior Research Fellowship to RRS. CE acknowledges financial support from the Engineering and Physical Sciences Research Council, the ChELSI initiative (EP/E036252/1).

AUTHOR CONTRIBUTIONS

RRS, TLK and CE conceived the study. RRS and MDG received funds for the work. MDG and CE collected the data. MDG and TLK analysed the data. MAR provided additional analysis tools. MDG and RRS wrote the manuscript with contributions from all authors. All authors agreed on the final version of the manuscript.

DATA ACCESSIBILTY

Supplementary material, data and R code used to perform analyses can be found at: https://github.com/MartinGarlovsky/Dmon_ejaculate_proteomics. Additional data has been

592 submitted to Dryad (https://doi.org/10.5061/dryad.pvmcvdnhw). The mass spectrometry
593 proteomics data have been deposited to the ProteomeXchange Consortium
594 (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno et
595 al. 2016) with the dataset identifier PXD019634.
596

597 **REFERENCES**

598 Ahmed-Braimah, Y. H., R. L. Unckless, and A. G. Clark. 2017. Evolutionary dynamics of male
599 reproductive genes in the *Drosophila virilis* subgroup. G3 Genes Genomes Genet.
600 7:3145–3155.
601 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment
602 search tool. J. Mol. Biol. 215:403–410.
603 Al-Wathiqui, N., S. M. Lewis, and E. B. Dopman. 2018. Molecular dissection of nuptial gifts in
604 divergent strains of Ostrinia moths. Physiol. Entomol. 43:10–19.
605 Avila, F. W., L. K. Sirot, B. A. LaFlamme, C. D. Rubinstein, and M. F. Wolfner. 2011. Insect
606 seminal fluid proteins: identification and function. Annu. Rev. Entomol. 56:21–40.
607 Avila, F. W., and M. F. Wolfner. 2009. Acp36DE is required for uterine conformational changes
608 in mated *Drosophila* females. Proc. Natl. Acad. Sci. 106:15796–15800.
609 Bayram, H., A. Sayadi, J. Goenaga, E. Immonen, and G. Arnqvist. 2017. Novel seminal fluid
610 proteins in the seed beetle Callosobruchus maculatus identified by a proteomic and
611 transcriptomic approach. Insect Mol. Biol. 26:58–73.
612 Bayram, H., A. Sayadi, E. Immonen, and G. Arnqvist. 2019. Identification of novel ejaculate
613 proteins in a seed beetle and division of labour across male accessory reproductive
614 glands. Insect Biochem. Mol. Biol. 104:50–57.
615 Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F.
616 Pagès, Z. Trajanoski, and J. Galon. 2009. ClueGO: a Cytoscape plug-in to decipher
617 functionally grouped gene ontology and pathway annotation networks. Bioinformatics
618 25:1091–1093.
619 Birkhead, T. R., and T. Pizzari. 2002. Postcopulatory sexual selection. Nat. Rev. Genet. 3:262–
620 273.

621 Bono, J. M., L. M. Matzkin, K. Hoang, and L. Brandsmeier. 2015. Molecular evolution of
622 candidate genes involved in post-mating-prezygotic reproductive isolation. J. Evol. Biol.
623 28:403–414.
624 Bono, J. M., L. M. Matzkin, E. S. Kelleher, and T. A. Markow. 2011. Postmating transcriptional
625 changes in reproductive tracts of con- and heterospecifically mated Drosophila
626 mojavensis females. Proc. Natl. Acad. Sci. 108:7878–7883.
627 Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann. 2014. Accurate Proteome-
628 wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio
629 Extraction, Termed MaxLFQ. Mol. Cell. Proteomics MCP 13:2513–2526.
630 Cramer, E. R. A., M. Ålund, S. E. McFarlane, A. Johnsen, and A. Qvarnström. 2016. Females
631 discriminate against heterospecific sperm in a natural hybrid zone. Evolution 70:1844–
632 1855.
633 Devigili, A., J. L. Fitzpatrick, C. Gasparini, I. W. Ramnarine, A. Pilastro, and J. P. Evans. 2018.
634 Possible glimpses into early speciation: the effect of ovarian fluid on sperm velocity
635 accords with post-copulatory isolation between two guppy populations. J. Evol. Biol.
636 31:66–74.
637 Fedorka, K. M., W. E. Winterhalter, and B. Ware. 2011. Percieved sperm competition intensity
638 influences seminal fluid protein production prior to courtship and mating. Evolution
639 65:584–590.
640 Findlay, G. D., M. J. MacCoss, and W. J. Swanson. 2009. Proteomic discovery of previously
641 unannotated, rapidly evolving seminal fluid genes in *Drosophila*. Genome Res. 19:886–
642 896.
643 Findlay, G. D., X. Yi, M. J. MacCoss, and W. J. Swanson. 2008. Proteomics Reveals Novel
644 *Drosophila* Seminal Fluid Proteins Transferred at Mating. PLoS Biol 6:e178.
645 Firman, R. C., C. Gasparini, M. K. Manier, and T. Pizzari. 2017. Postmating female control: 20
646 years of cryptic female choice. Trends Ecol. Evol. 32:368–382.
647 Garlovsky, M. D., and R. R. Snook. 2018. Persistent postmating, prezygotic reproductive
648 isolation between populations. Ecol. Evol. 8:9062–9073.
649 Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature
650 403:886–889.
651 Goenaga, J., T. Yamane, J. Rönn, and G. Arnqvist. 2015. Within-species divergence in the
652 seminal fluid proteome and its effect on male and female reproduction in a beetle. BMC
653 Evol. Biol. 15:266.

654 Holman, L. 2009. *Drosophila melanogaster* seminal fluid can protect the sperm of other males.
655 Funct. Ecol. 23:180–186.
656 Holt, C., and M. Yandell. 2011. MAKER2: an annotation pipeline and genome-database
657 management tool for second-generation genome projects. BMC Bioinformatics 12:491.
658 Howard, D. J., S. R. Palumbi, L. M. Birge, and M. K. Manier. 2009. Sperm and speciation. Pp.
659 367–403 *in* Sperm biology: an evolutionary perspective. Academic press, Burlington,
660 MA.
661 Innocenti, P., and E. H. Morrow. 2009. Immunogenic males: a genome-wide analysis of
662 reproduction and the cost of mating in Drosophila melanogaster females. J. Evol. Biol.
663 22:964–973.
664 Jennings, J. H., R. R. Snook, and A. Hoikkala. 2014. Reproductive isolation among allopatric
665 *Drosophila montana* populations. Evolution 68:3095–3108.
666 Käll, L., A. Krogh, and E. L. L. Sonnhammer. 2004. A Combined Transmembrane Topology
667 and Signal Peptide Prediction Method. J. Mol. Biol. 338:1027–1036.
668 Karr, T. L. 2019. Reproductive Proteomics Comes of Age. Mol. Cell. Proteomics 18:S1–S5.
669 Karr, T. L., H. Southern, M. A. Rosenow, T. I. Gossmann, and R. R. Snook. 2019. The Old and
670 the New: Discovery Proteomics Identifies Putative Novel Seminal Fluid Proteins in
671 Drosophila. Mol. Cell. Proteomics 18:S23–S33. American Society for Biochemistry and
672 Molecular Biology.
673 Kelleher, E. S., and J. E. Pennington. 2009. Protease Gene Duplication and Proteolytic Activity
674 in Drosophila Female Reproductive Tracts. Mol. Biol. Evol. 26:2125–2134.
675 Kelleher, E. S., W. J. Swanson, and T. A. Markow. 2007. Gene duplication and adaptive
676 evolution of digestive proteases in Drosophila arizonae female reproductive tracts.
677 PLOS Genet. 3:e148.
678 Köttgen, M., A. Hofherr, W. Li, K. Chu, S. Cook, C. Montell, and T. Watnick. 2011. Drosophila
679 Sperm Swim Backwards in the Female Reproductive Tract and Are Activated via TRPP2
680 Ion Channels. PLOS ONE 6:e20031. Public Library of Science.
681 LaFlamme, B. A., K. Ravi Ram, and M. F. Wolfner. 2012. The *Drosophila melanogaster*
682 seminal fluid protease “Seminase” regulates proteolytic and post-mating reproductive
683 processes. PLOS Genet. 8:e1002435.
684 LaFlamme, B. A., and M. F. Wolfner. 2013. Identification and function of proteolysis regulators
685 in seminal fluid. Mol. Reprod. Dev. 80:80–101.
686 Lakovaara, S. 1969. Malt as a culture medium for *Drosophila* species. Drosoph. Inf. Serv.

687 Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad.
688 Sci. 78:3721–3725.
689 Löytynoja, A., and N. Goldman. 2010. webPRANK: a phylogeny-aware multiple sequence
690 aligner with interactive alignment browser. BMC Bioinformatics 11:579.
691 Manier, M. K., S. Lüpold, J. M. Belote, W. T. Starmer, K. S. Berben, O. Ala-Honkola, W. F.
692 Collins, and S. Pitnick. 2013. Postcopulatory sexual selection generates speciation
693 phenotypes in *Drosophila*. Curr. Biol. 23:1853–1862.
694 Markow, T. A. 2002. Perspective: Female remating, operational sex ratio, and the arena of
695 sexual selection in *Drosophila* species. Evolution 56:1725–1734.
696 Marshall, J. L., D. L. Huestis, C. Garcia, Y. Hiromasa, S. Wheeler, S. Noh, J. M. Tomich, and
697 D. J. Howard. 2011. Comparative proteomics uncovers the signature of natural selection
698 acting on the ejaculate proteomes of two cricket species isolated by postmating,
699 prezygotic phenotypes. Mol. Biol. Evol. 28:423–435.
700 Marshall, J. L., D. L. Huestis, Y. Hiromasa, S. Wheeler, C. Oppert, S. A. Marshall, J. M. Tomich,
701 and B. Oppert. 2009. Identification, RNAi knockdown, and functional analysis of an
702 ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. PLoS
703 ONE 4.
704 Mattei, A. L., M. L. Riccio, F. W. Avila, and M. F. Wolfner. 2015. Integrated 3D view of
705 postmating responses by the *Drosophila melanogaster* female reproductive tract,
706 obtained by micro-computed tomography scanning. Proc. Natl. Acad. Sci. 112:8475–
707 8480.
708 McDonough, C. E., E. Whittington, S. Pitnick, and S. Dorus. 2016. Proteomics of reproductive
709 systems: Towards a molecular understanding of postmating, prezygotic reproductive
710 barriers. J. Proteomics 135:26–37.
711 McGeary, M. K., and G. D. Findlay. 2020. Molecular evolution of the sex peptide network in
712 Drosophila. J. Evol. Biol. n/a.
713 Meslin, C., T. S. Cherwin, M. S. Plakke, B. S. Small, B. J. Goetz, N. I. Morehouse, and N. L.
714 Clark. 2017. Structural complexity and molecular heterogeneity of a butterfly ejaculate
715 reflect a complex history of selection. Proc. Natl. Acad. Sci. 114:E5406–E5413.
716 Mueller, J. L., K. R. Ram, L. A. McGraw, M. C. Bloch Qazi, E. D. Siggia, A. G. Clark, C. F.
717 Aquadro, and M. F. Wolfner. 2005. Cross-species comparison of *Drosophila* male
718 accessory gland protein genes. Genetics 171:131–143.

719 Panhuis, T. M., R. K. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation.
720 Trends Ecol. Evol. 16:364–371.
721 Parker, D. J., R. A. W. Wiberg, U. Trivedi, V. I. Tyukmaeva, K. Gharbi, R. K. Butlin, A. Hoikkala,
722 M. Kankare, M. G. Ritchie, and J. Gonzalez. 2018. Inter and intraspecific genomic
723 divergence in *Drosophila montana* shows evidence for cold adaptation. Genome Biol.
724 Evol. 10:2086–2101.
725 Perry, J. C., L. Sirot, and S. Wigby. 2013. The seminal symphony: how to compose an
726 ejaculate. Trends Ecol. Evol. 28:414–422.
727 Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0: discriminating
728 signal peptides from transmembrane regions. Nat. Methods 8:785–786.
729 Pillai, H., A. M. Shende, M. S. Parmar, A. A, S. L, A. Kumaresan, T. S. G, and S. K. Bhure.
730 2017. Regucalcin is widely distributed in the male reproductive tract and exerts a
731 suppressive effect on in vitro sperm capacitation in the water buffalo (Bubalus bubalis).
732 Mol. Reprod. Dev. 84:212–221.
733 Pitnick, S., M. F. Wolfner, and S. Dorus. 2020. Post-ejaculatory modifications to sperm
734 (PEMS). Biol. Rev. 95:365–392.
735 Pitnick, S., M. F. Wolfner, and S. S. Suarez. 2009. Ejaculate-female and sperm-female
736 interactions. Pp. 247–304 *in* Sperm biology: an evolutionary perspective. Academic
737 press, Burlington, MA.
738 Plakke, M. S., A. B. Deutsch, C. Meslin, N. L. Clark, and N. I. Morehouse. 2015. Dynamic
739 digestive physiology of a female reproductive organ in a polyandrous butterfly. J. Exp.
740 Biol. 218:1548–1555.
741 Plakke, M. S., J. L. Walker, J. B. Lombardo, B. J. Goetz, G. N. Pacella, J. D. Durrant, N. L.
742 Clark, and N. I. Morehouse. 2019. Characterization of Female Reproductive Proteases
743 in a Butterfly from Functional and Evolutionary Perspectives. Physiol. Biochem. Zool.
744 92:579–590.
745 Price, C. S. C., C. H. Kim, C. J. Gronlund, and J. A. Coyne. 2001. Cryptic reproductive isolation
746 in the *Drosophila simulans* species complex. Evolution 55:81–92.
747 R Core Team. 2018. R: A language and Environment for Statistical Computing. R Foundation
748 for Statistical Computing, Vienna, Austria.
749 Ramm, S. A., L. McDonald, J. L. Hurst, R. J. Beynon, and P. Stockley. 2009. Comparative
750 proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and
751 its functional significance in sperm competition. Mol. Biol. Evol. 26:189–198.

752 Ravi Ram, K., and M. F. Wolfner. 2007. Seminal influences: *Drosophila* Acps and the molecular
753 interplay between males and females during reproduction. Integr. Comp. Biol. 47:427–
754 445.
755 Reinhardt, K., and A.-C. Ribou. 2013. Females become infertile as the stored sperm's oxygen
756 radicals increase. Sci. Rep. 3:2888. Nature Publishing Group.
757 Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma
758 powers differential expression analyses for RNA-sequencing and microarray studies.
759 Nucleic Acids Res. 43:e47–e47.
760 Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for
761 differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl.
762 26:139–140.
763 Rowe, M., T. Albrecht, E. R. A. Cramer, A. Johnsen, T. Laskemoen, J. T. Weir, and J. T. Lifjeld.
764 2015. Postcopulatory sexual selection is associated with accelerated evolution of sperm
765 morphology. Evolution 69:1044–1052.
766 Rowe, M., E. Whittington, K. Borziak, M. Ravinet, F. Eroukhmanoff, G.-P. Sætre, and S. Dorus.
767 2019. Molecular Diversification of the Seminal Fluid Proteome in a Recently Diverged
768 Passerine Species Pair. Mol. Biol. Evol., doi: 10.1093/molbev/msz235.
769 Sagga, N., and A. Civetta. 2011. Male-female interactions and the evolution of postmating
770 prezygotic reproductive isolation among species of the Virilis subgroup. Int. J. Evol. Biol.
771 2011.
772 Sepil, I., B. R. Hopkins, R. Dean, M.-L. Thézénas, P. D. Charles, R. Konietzny, R. Fischer, B.
773 M. Kessler, and S. Wigby. 2019. Quantitative proteomics identification of seminal fluid
774 proteins in male *Drosophila melanogaster*. Mol. Cell. Proteomics 18:S46–S58.
775 Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B.
776 Schwikowski, and T. Ideker. 2003. Cytoscape: A Software Environment for Integrated
777 Models of Biomolecular Interaction Networks. Genome Res. 13:2498–2504.
778 Simmons, L. W., and J. L. Fitzpatrick. 2019. Female genitalia can evolve more rapidly and
779 divergently than male genitalia. Nat. Commun. 10:1312.
780 Singh, A., N. A. Buehner, H. Lin, K. J. Baranowski, G. D. Findlay, and M. F. Wolfner. 2018.
781 Long-term interaction between *Drosophila* sperm and sex peptide is mediated by other
782 seminal proteins that bind only transiently to sperm. Insect Biochem. Mol. Biol. 102:43–
783 51.

784 Sirot, L. K., A. Wong, T. Chapman, and M. F. Wolfner. 2015. Sexual conflict and seminal fluid
785 proteins: a dynamic landscape of sexual interactions. Cold Spring Harb. Perspect. Biol.
786 7:a017533.
787 South, A., and S. M. Lewis. 2011. The influence of male ejaculate quantity on female fitness:
788 a meta-analysis. Biol. Rev. 86:299–309.
789 Tsuda, M., J.-B. Peyre, T. Asano, and T. Aigaki. 2015. Visualizing Molecular Functions and
790 Cross-Species Activity of Sex-Peptide in *Drosophila*. Genetics 200:1161–1169.
791 Turissini, D. A., J. A. McGirr, S. S. Patel, J. R. David, and D. R. Matute. 2018. The rate of
792 evolution of postmating-prezygotic reproductive isolation in *Drosophila*. Mol. Biol. Evol.
793 35:312–334.
794 Tyanova, S., T. Temu, and J. Cox. 2016. The MaxQuant computational platform for mass
795 spectrometry-based shotgun proteomics. Nat. Protoc. 11:2301–2319.
796 VanKuren, N. W., and M. Long. 2018. Gene duplicates resolving sexual conflict rapidly evolved
797 essential gametogenesis functions. Nat. Ecol. Evol. 1.
798 Vesala, L., T. S. Salminen, M. Kankare, and A. Hoikkala. 2012. Photoperiodic regulation of
799 cold tolerance and expression levels of regucalcin gene in Drosophila montana. J.
800 Insect Physiol. 58:704–709.
801 Vizcaíno, J. A., A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-
802 Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, and H. Hermjakob. 2016. 2016
803 update of the PRIDE database and its related tools. Nucleic Acids Res. 44:D447–D456.
804 Oxford Academic.
805 Wagstaff, B. J., and D. J. Begun. 2005. Molecular Population Genetics of Accessory Gland
806 Protein Genes and Testis-Expressed Genes in Drosophila mojavensis and D. arizonae.
807 Genetics 171:1083–1101.
808 Walker, M. J., C. M. Rylett, J. N. Keen, N. Audsley, M. Sajid, A. D. Shirras, and R. E. Isaac.
809 2006. Proteomic identification of *Drosophila melanogaster* male accessory gland
810 proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase. Proteome
811 Sci. 4:9.
812 Wang, D., B. Eraslan, T. Wieland, B. Hallström, T. Hopf, D. P. Zolg, J. Zecha, A. Asplund, L.
813 Li, C. Meng, M. Frejno, T. Schmidt, K. Schnatbaum, M. Wilhelm, F. Ponten, M. Uhlen,
814 J. Gagneur, H. Hahne, and B. Kuster. 2019. A deep proteome and transcriptome
815 abundance atlas of 29 healthy human tissues. Mol. Syst. Biol. 15.

816 Whittington, E., T. L. Karr, A. J. Mongue, S. Dorus, and J. R. Walters. 2019. Evolutionary
817 Proteomics Reveals Distinct Patterns of Complexity and Divergence between
818 Lepidopteran Sperm Morphs. Genome Biol. Evol. 11:1838–1846.
819 Wigby, S., L. K. Sirot, J. R. Linklater, N. Buehner, F. C. F. Calboli, A. Bretman, M. F. Wolfner,
820 and T. Chapman. 2009. Seminal fluid protein allocation and male reproductive success.
821 Curr. Biol. 19:751–757.
822 Wolfner, M. F. 2009. Battle and ballet: molecular interactions between the sexes in *Drosophila*.
823 J. Hered. 100:399–410.
824 Wong, A., S. N. Albright, J. D. Giebel, K. R. Ram, S. Ji, A. C. Fiumera, and M. F. Wolfner.
825 2008. A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in
826 *Drosophila melanogaster*. Genetics 180:921–931.
827 Wright, A. E., P. W. Harrison, F. Zimmer, S. H. Montgomery, M. A. Pointer, and J. E. Mank.
828 2015. Variation in promiscuity and sexual selection drives avian rate of Faster-Z
829 evolution. Mol. Ecol. 24:1218–1235.
830 Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol.
831 24:1586–1591.
832
833