

✂ Author's Choice

# Quantitative Proteomics Identification of Seminal Fluid Proteins in Male *Drosophila melanogaster*<sup>✂</sup>

✉ Irem Sepil<sup>‡</sup>, Ben R. Hopkins<sup>‡</sup>, Rebecca Dean<sup>§</sup>, Marie-Laëtitia Thézénas<sup>¶</sup>, Philip D. Charles<sup>¶</sup>, Rebecca Konietzny<sup>¶</sup>, Roman Fischer<sup>¶</sup>, Benedikt M. Kessler<sup>¶</sup>, and Stuart Wigby<sup>‡</sup>

Seminal fluid contains some of the fastest evolving proteins currently known. These seminal fluid proteins (Sfps) play crucial roles in reproduction, such as supporting sperm function, and particularly in insects, modifying female physiology and behavior. Identification of Sfps in small animals is challenging, and often relies on samples taken from the female reproductive tract after mating. A key pitfall of this method is that it might miss Sfps that are of low abundance because of dilution in the female-derived sample or rapid processing in females. Here we present a new and complementary method, which provides added sensitivity to Sfp identification. We applied label-free quantitative proteomics to *Drosophila melanogaster* male reproductive tissue - where Sfps are unprocessed, and highly abundant - and quantified Sfps before and immediately after mating, to infer those transferred during copulation. We also analyzed female reproductive tracts immediately before and after copulation to confirm the presence and abundance of known and candidate Sfps, where possible. Results were cross-referenced with transcriptomic and sequence databases to improve confidence in Sfp detection. Our data were consistent with 125 previously reported Sfps. We found nine high-confidence novel candidate Sfps, which were both depleted in mated *versus* unmated males and identified within the reproductive tract of mated but not virgin females. We also identified 42 more candidates that are likely Sfps based on their abundance, known expression and predicted characteristics, and revealed that four proteins previously identified as Sfps are at best minor contributors to the ejaculate. The estimated copy numbers for our candidate Sfps were lower than for previously identified Sfps, supporting the idea that our technique provides a deeper analysis of the Sfp proteome than previous studies. Our results demonstrate a novel, high-sensitivity approach to the analysis of seminal fluid proteomes, whose appli-

cation will further our understanding of reproductive biology. *Molecular & Cellular Proteomics* 18: S46–S58, 2019. DOI: 10.1074/mcp.RA118.000831.

Seminal fluid, the non-sperm component of the ejaculate, is a highly complex matrix of bio-molecules including peptides and proteins (1, 2). Seminal fluid proteins (Sfps)<sup>1</sup> are typically produced in specialized secretory glands in males (such as the accessory glands in insects, and the prostate, seminal vesicles, bulbourethral glands and ampullary glands in mammals), and are transferred to females during copulation. Sfps can play roles in sperm capacitation, storage, competition and fertilization, and modulate female post-mating behavior and physiology (2–8). In humans, evidence is accumulating that Sfps contribute to sperm fertilization success, and Sfps have been suggested as important biomarkers of male infertility (9). Given the decline in male fertility over the last few decades (10), and increasing age-related male infertility because of later parenthood in developed countries (11), there is an urgent need for an improved molecular understanding of male reproduction. Proteomics will play an important part in driving forward these advances in the field of human male fertilization biology (12).

In polyandrous species (in which females mate with multiple males) Sfps can influence sperm competition, whereby the ejaculates of different males overlap in the female reproductive tract, and the sperm of different males compete for fertilization (3). Sfps evolve rapidly and are thought to be under sexual selection as a result of both sperm competition and co-evolutionary conflicts between males and females (5, 13–15). Understanding which Sfps contribute to male sperm competition outcomes is especially important in polyandrous insect pests, because the success of key biocontrol methods,

From the <sup>‡</sup>Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK; <sup>§</sup>Department of Genetics, Evolution and Environment, University College London, London, UK; <sup>¶</sup>ITDI Mass Spectrometry Laboratory, Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

✂ Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

Received May 2, 2018, and in revised form, July 12, 2018

Published, MCP Papers in Press, October 4, 2018, DOI 10.1074/mcp.RA118.000831

such as the Sterile Insect Technique, rely on release of the males with competitively successful ejaculates (16). Moreover, studies in mammalian models show that seminal plasma can even influence the health of offspring (2, 17). Given their important effects for male and female reproductive success and offspring health, considerable recent research effort has focused on proteomic analyses of Sfps, for a diverse range of taxa. However, identification of the complete set of proteins that are transferred to the female at mating remains a challenge.

Many mammals are amenable to artificial ejaculation techniques, where the ejaculate is obtained by abdominal massage/squeezing (18, 19), the usage of artificial vaginae (20) or electroejaculation (21). Although these methods allow for direct analyses of Sfps, they can produce abnormal or inconsistent ejaculates, such as seen in mice (22). Moreover, these techniques are taxonomically restricted, and of limited use for most insect or bird species. An alternative method for the identification of Sfps is whole-organism isotopic labeling methods, whereby the females are metabolically labeled – by feeding a diet enriched in a “heavy” isotope – then mated to unlabeled males. As a result, the female reproductive tract proteome contains labeled female-derived proteins and unlabeled male Sfps that can be distinguished and quantified. <sup>15</sup>N-labeled females have been used to characterize the Sfps of the fruit fly *Drosophila melanogaster*, house mice *Mus domesticus*, and dengue vector mosquito *Aedes aegypti* (22–24). Although isotopic labeling methods have been instrumental for allowing direct characterization of the seminal fluid proteome, they may not be able to detect all male Sfps. Sfps might interact with each other during and after copulation—either in the post ejaculation stage independent from the female, or once inside the female—and this interaction might lead to protein degradation or cleavage, to release biologically active products. For instance, the *D. melanogaster* Sfp, Acp26Aa, is rapidly cleaved within the mated female’s reproductive tract, whereupon two of its cleavage products induce ovulation (25) and is detectable by ELISA for only one hour after mating (26). If some Sfps are even faster processed they may be hard to detect within the mated female by proteomic methods. Sfps involved in conflicts between the sexes could be rapidly degraded by females if harm is minimized by impairing the Sfp’s function (27). Another potential disadvantage of analyzing female reproductive tract samples after mating is that the Sfps are diluted once they are inside the female, decreasing their relative abundance. Previous work in *D. melanogaster* suggests that only about 15% of peptides are from males in dissected female reproductive tracts (based on comparing the number of peptides without *versus* with <sup>15</sup>N-label) (G. Findlay, personal communication). Hence, it is likely

that methods aimed at identifying Sfps in mated female reproductive tract tissue samples may miss Sfps that are low in abundance within the female.

Here we present a new quantitative proteomic method, based on the comparison of the accessory gland proteomes of male fruit flies, *Drosophila melanogaster*, before and after mating. This method negates the above issues inherent to the analysis of female derived samples, and allows for the indirect, but potentially powerful, inference of candidate Sfps. *Drosophila* is a model species for ejaculate research and the study of ejaculate-mediated sexual selection and sexual conflict (5, 28, 29). The functions of several *D. melanogaster* Sfps have been investigated in detail, particularly in relation to their roles in modulating behavioral and physiological processes in the female (7, 30). Using <sup>15</sup>N-labeling of the female, Findlay *et al.*, 2008 identified 157 Sfps transferred from the male during copulation (23). A small number of other male-derived proteins have been identified in the reproductive tract of mated females in *Drosophila melanogaster*, bringing the total number to 163 proteins (31, 32). We refer to these proteins as “known Sfps.” In comparison 2,064 proteins have been identified in the human seminal fluid. Although this number is an order of magnitude more than the known *Drosophila* Sfps, it is still considerably lower than proteins detected in other human bodily fluids, as for example the 10,000 proteins detected in blood plasma (12). Hence, it has been suggested that the large range of human Sfp abundance could be hindering the detection of low abundance proteins, a problem that might be shared among taxa including *Drosophila*.

*D. melanogaster* Sfps are stored in the male reproductive tract secretory tissues - accessory glands, seminal vesicles, ejaculatory duct, ejaculatory bulb and testes (7), but we currently have a limited understanding of the storage locations for each Sfp (but see (33)). We describe a label-free quantitative proteomics method based on the comparison of male accessory gland proteomes for candidate Sfp identification in *D. melanogaster*. This method is particularly aimed at capturing less abundant or rapidly degrading Sfps, that may have been missed in previous studies. We based the study on the prediction that the abundance of Sfps in male reproductive tract secretory tissues would significantly decrease at copulation. As expected, we found that the majority of detected known Sfps were significantly less abundant following mating. Many more proteins were also depleted following mating, indicating possible contribution to the pool of Sfps. These were analyzed for the presence of a signal peptide for secretion, or to understand if the protein is exclusively expressed in accessory gland tissues. The proteins that meet these assumptions are suggested as candidate Sfps. No candidates passing both these filters were found in the reproductive tract of virgin females, lending further support to the idea that these are male-originating. Finally, by quantifying the proteome of the accessory glands and ejaculatory duct separately, we demonstrate that several known Sfps are mainly or entirely

<sup>1</sup> The abbreviations used are: Sfps, seminal fluid proteins; GASP, gel-aided sample preparation; DAVID, database for visualization and integrated discovery; GO, gene ontology.

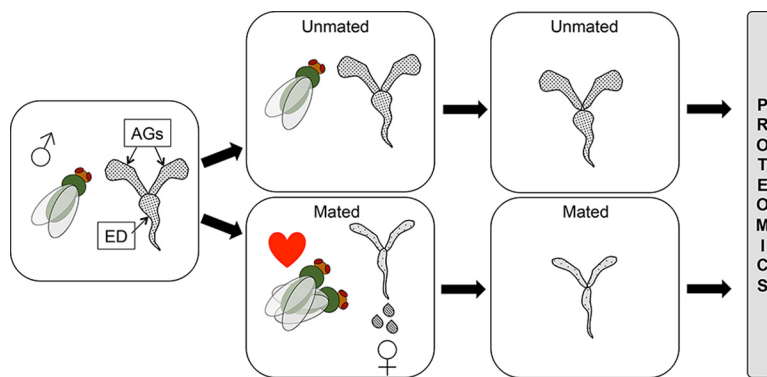


FIG. 1. **Experimental design.** Males are expected to lose seminal fluid proteins (Sfps) from the accessory glands (AGs) and ejaculatory duct (ED) at copulation as they are transferred to females. By analyzing protein abundance in the AGs and ED immediately after copulation *versus* in unmated males we can infer Sfps that are likely transferred. Sfps should be significantly more abundant in unmated males than in mated males.

stored in the ejaculatory duct rather than in the accessory glands.

#### EXPERIMENTAL PROCEDURES

##### *Stock and Fly Maintenance*

We used a lab-adapted, outbred Dahomey wild-type stock for all our experimental males, which has been maintained in large population cages with overlapping generations since 1970. All flies were maintained at 25 °C on a 12:12 L:D cycle and fed Lewis medium (34). Adult flies were maintained in 36 ml plastic vials containing Lewis medium supplemented with *ad libitum* live yeast grains. Flies were reared using a standard larval density method by placing ~200 eggs on 50 ml of food in 250 ml bottles (35). Virgins were collected on ice anesthesia within 8h of eclosion and were assigned to their experimental group.

##### *Experimental Design and Statistical Rationale*

The general approach to find candidate Sfps was to identify proteins in the male reproductive glands that significantly decreased in abundance after copulation (Fig. 1). We then used expression data from ModENCODE and sequence data from UniParc to determine if these proteins are exclusively expressed in the accessory glands, and if they are secreted (36–38), as is expected but not mandatory for Sfps. We also identified the proteins that significantly increase in abundance in the female reproductive tracts after mating to validate the candidate Sfps. Finally; we compared Sfp abundances in accessory glands and ejaculatory ducts to determine where they are stored. For Male Data Set 1 we generated four, and for Male Data Set 2 and Female Data Set we generated five biological replicates per condition following previously published power analyses indicating that a minimum of four biological replicates are required for reliable measurement of 1.5 and greater fold changes (setting power at 0.8 and confidence threshold at 0.05, and assuming a combined technical and biological variation of 25% (39)). For the Male Tissue Data Set we generated three biological replicates per condition, as power analyses (performed using Progenesis QIP internal tool, see below) of the previous data sets indicated that at least 60% of proteins (at 0.8 power) were quantifiable with only three replicates.

##### *Male Reproductive Gland Proteomes*

To obtain the quantitative proteome of male reproductive glands before and after mating, we used samples from two independent

experiments. These experiments, detailed below, provide a range of conditions for males, which might improve our power to identify proteins if Sfp expression is context-dependent. In the first experiment, male age and mating history were experimentally varied, and in the second experiment the adult social environment (male group size) was experimentally varied (supplemental Fig. S1). Any effects of age, mating history or social environment on Sfp abundance *per se*, are beyond the scope of the current study (Sepil *et al.*, *in prep*; Hopkins *et al.* *in prep*), but were controlled statistically to maximize power (see Statistical Analyses below).

*Male Data Set 1: Males of Varying Age and Mating History*—Samples were collected from male flies of experimentally varied age and mating history, as follows (supplemental Fig. S1a). Upon eclosion, Dahomey males were housed in groups of 12, either all males (single sex group) or consisting of three virgin males and nine virgin Dahomey females (mixed sex group). Males were allowed to age in their group vials for up to 5 weeks. Males from three age classes were used: 1 week, 3 weeks and 5 weeks old. The single sex group flies were transferred once per week, and the mixed sex group flies were transferred twice a week to fresh vials using light CO<sub>2</sub> anesthesia at each transfer. During the transfers, dead or escaped females were replaced with similarly aged mated females. To minimize female co-aging effects in the 5-week-old mixed sex groups, females were replaced at 3 weeks with virgin 3–5 days old females, reared using the same procedures as above. To minimize density effects on mating opportunity in the mixed sex group vials, two vials of the same treatment were merged when a single male was left in a vial owing to previous mortality or censoring. The males from the mixed sex groups were merged into single sex groups of 10–12 males 5 days before sample collection, in order to allow them to replenish their Sfps.

The day before the sample collection, 210 virgin females were placed individually in vials. On the day of sample collection, 35 males from each treatment (1 week old single sex group, 3 weeks old single sex group, 5 weeks old single sex group, 1 week old mixed sex group, 3 weeks old mixed sex group, 5 weeks old mixed sex group) were added to the individually housed female vials for mating. The mated males were flash frozen in liquid nitrogen 25 min after the start of the mating. These flies formed the “newly mated” male groups (supplemental Fig. S1a).

Another 35 males from each treatment (*i.e.* 1 week old single sex group, 3 week old single sex group, 5 week old single sex group, 1 week old mixed sex group, 3 week old mixed sex group, 5 week old mixed sex group) were flash frozen in liquid nitrogen without being exposed to females. These flies formed the “unmated” male groups.

Hence each of the six treatments had a “newly mated” and an “unmated” sample that were paired for further analysis (supplemental Fig. S1a). We repeated this experiment three times to produce four independent biological replicates. We thawed flash frozen males and dissected their accessory glands and ejaculatory duct on ice in PBS buffer. We did not include the ejaculatory bulb and the ejaculatory duct was incised approximately at the distal end. 19 reproductive glands from males of the same treatment and replicate (out of a potential of 35) were pooled in 25  $\mu$ l PBS buffer on ice. Hence, we had six paired newly mated male and unmated male samples from each replicate and 24 paired newly mated male and unmated male samples in total (i.e. 48 samples overall).

**Male Data Set 2: Varied Social Exposure**—Upon eclosion, males were randomly allocated to one of three single sex group size treatments: individually housed (treatment 1), housed in pairs (treatment 2), or housed in groups of eight (treatment 8). The males were aged in their treatment vials for 4 days (supplemental Fig. S1b). The day before sample collection, 105 virgin females were placed individually in vials. On the day of sample collection, 35 males from each treatment (1, 2, and 8) were added to the female vials for mating. The mated males were flash frozen in liquid nitrogen 25 min after the start of the mating. These flies formed the “newly mated” male groups (supplemental Fig. S1b).

Another 35 males from each treatment (1, 2, and 8) were flash frozen in liquid nitrogen as virgins without being exposed to females. These flies formed the “unmated” male groups. Hence each treatment (1, 2, 8) had a “newly mated” and “unmated” sample that were paired for further analysis. We repeated this experiment to produce five independent biological replicates. Flash frozen males were dissected as outlined above. Twenty reproductive glands from males of the same treatment and replicate were pooled in 25  $\mu$ l PBS buffer on ice (supplemental Fig. S1b). Hence, we had three paired newly mated male and virgin male samples from each replicate and 15 paired newly mated male and virgin male samples in total. Overall, we had 30 samples.

#### Female Reproductive Tract Proteome

**Female Data Set**—Upon eclosion Dhomey females and Dhomey males were aged in single sex groups of 12 for 3 days. The day before sample collection 35 females were placed individually in vials. On the day of sample collection, a single male was added to each female vial for mating. The mated females were flash frozen in liquid nitrogen 30 min after the start of the mating. These flies formed the “newly mated” female group. Another 35 females were flash frozen in liquid nitrogen as virgins without being exposed to males. These flies formed the “virgin” female group. The newly mated and virgin samples were paired for further analysis. We repeated this experiment to obtain five independent, biological replicates. Flash frozen females were thawed and their reproductive tracts (uterus, spermathecae, ovaries and the seminal receptacle, excluding the ovaries) were dissected on ice in PBS buffer. 20 reproductive tracts from females of the same treatment and replicate were pooled in 25  $\mu$ l PBS buffer on ice. Hence, we had five paired newly mated female and virgin female samples in total. Overall, we had 10 samples.

#### Male Accessory Glands and Ejaculatory Duct Proteome

**Male Tissue Data Set**—Upon eclosion, males were aged in single sex groups of 12 for 3 days. 70 males were flash frozen in liquid nitrogen as virgins. We repeated this procedure two more times to have three independent, biological replicates. Flash frozen males were thawed and randomly allocated to one of three dissection regimes: “Accessory Gland” (AG) regime flies only had their accessory glands dissected out, “Ejaculatory Duct (ED) regime flies only had

their ejaculatory duct (excluding ejaculatory bulb) dissected out and “Both” (BO) regime flies had both their accessory glands and ejaculatory duct dissected out. All three biological replicates were split into AG, ED, and BO regime dissection groups. 20 reproductive tissues from males of the same dissection group and replicate were pooled in 25  $\mu$ l PBS buffer on ice. Overall, we had nine samples.

#### Sample Preparation

All samples described above were stored at  $-80^{\circ}\text{C}$  until sample preparation for proteomic analysis. The samples were macerated with a clean pestle and washed with 25  $\mu$ l of Pierce RIPA Buffer. Then they were digested using the standard gel-aided sample preparation (GASP) protocol as described previously (40). In brief, samples were reduced with 50 mM DTT for 10 to 20 min. Protein lysate was mixed with an equal volume of 40% acrylamide/Bis solution (37.5:1. National Diagnostics, Atlanta, Georgia) and left at room temperature for 30 min to facilitate cysteine alkylation to propionamide. 5  $\mu$ l TEMED and 5  $\mu$ l 10% APS were added to trigger acrylamide polymerization. The resulting gel plug was shredded by centrifugation through a Spin-X filter insert without membrane (CLS9301, Sigma/Corning, Darmstadt, Germany). Gel pieces were fixed in 40% ethanol/5% acetic acid before 2 successive rounds of buffer exchange with 1.5 M Urea, 0.5 M Thiourea and 50 mM ammonium bicarbonate which were removed with acetonitrile. Immobilized proteins were digested with trypsin (Promega, Madison, Wisconsin) overnight and peptides extracted with two rounds of acetonitrile replacements. Peptides were first dried before desalting using Sola SPE columns (Thermo, Waltham, Massachusetts) and resuspended in 2% ACN, 0.1% FA buffer prior LC-MS/MS analysis.

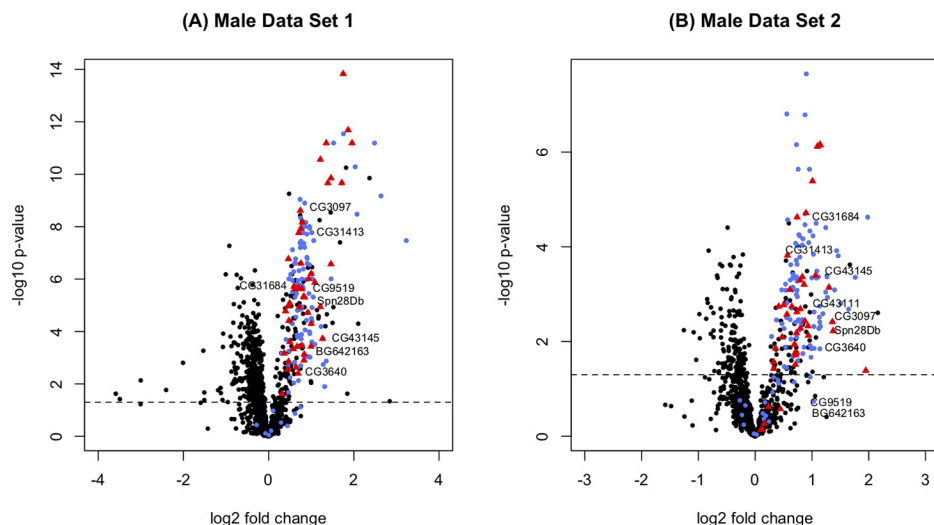
#### LC-MS/MS

Peptide samples were analyzed on a LC-MS/MS platform consisting of a Dionex Ultimate 3000 and a Q-Exactive (Male Data Set 2) or Q-Exactive HF (other data sets) mass spectrometer (Thermo). After peptide loading in 0.1% TFA in 2% ACN onto a trap column (PepMAP C18, 300  $\mu$ m x5 mm, 5  $\mu$ m particle, Thermo), peptides were separated on an easy spray column (PepMAP C18, 75  $\mu$ m x 500 mm, 2  $\mu$ m particle, Thermo) with a gradient 2% ACN to 35% ACN in 0.1% formic acid in 5% DMSO.

MS spectra were acquired in profile mode with a resolution of 70,000 on the Q-Exactive (QE-HF: 60,000) with an ion target of  $3 \times 10^6$ . The instrument was set to pick the 15 (QE-HF: 12) most intense features for subsequent MS/MS analysis at a resolution of 17,500 (QE-HF: 30,000), a maximum acquisition time of 128ms (QE-HF: 45ms), an AGC target of  $1 \times 10^5$ , an isolation width of 1.6Th (QE-HF: 1.2Th) and a dynamic exclusion of 27 s.

#### Processing of MS Data

RAW files were imported into Progenesis QIP (version 3.0.6039.34628) using default settings. MS/MS spectra were exported as MGF files using the 200 most intense peaks without deconvolution for searching. The *Drosophila melanogaster* UniProt reference proteome was used as a search target in all cases. Database retrieval dates were 30/03/2015 (21361 sequences) for Male Data Set 1, 15/02/2017 (23302 sequences) for Male Data Set 2, and 07/06/2017 (23285 sequences) for the remaining data sets. Male Data Set 1 was searched in PEAKS 7. The other data sets were searched in Mascot 2.5.1. In both search engines we used the following parameters: 10 ppm precursor mass accuracy, 0.05 Da fragment mass accuracy, Oxidation (M), Deamidation (N, Q) and Propionamide (K) as variable modifications, Propionamide (C) as a fixed modification, and two missed cleavage sites. We applied 1% FDR at peptide level (both



**FIG. 2. Volcano plot displaying all proteins detected in (A) Male Data Set 1 and (B) Male Data Set 2 that were identified by at least two unique peptides.** The  $\log_2$  fold change [ $\log_2$  (unmated) -  $\log_2$  (newly mated)] is shown on the x axis and significance is displayed on the y axis as the negative logarithm ( $\log_{10}$  scale) of the *fdr* corrected *p* value. Known Sfps are colored in blue. The candidate Sfps identified in this study are displayed as triangles and colored in red. The high-confidence candidate Sfps are named. The rest of the proteins are colored black. The significance cutoff ( $p < 0.05$ ) is highlighted with a dashed line.

search engines use a target-decoy method for FDR estimation) and an additional Mascot ion score cutoff of 20 before importing search results into Progenesis, where protein quantification was calculated using the Top3 method. Quantitative protein data were further normalized/processed as described below.

#### *In Silico Protein Annotation*

We used SignalP and UniProt to predict whether a protein was likely to be secreted, by checking for the presence of a signal peptide (36, 37). We used ModENCODE to check for exclusive expression in the accessory glands (38). UniProt was also used to deduce protein function. Lastly the Database for Visualization and Integrated Discovery (DAVID) was used for gene ontology (GO) enrichment analysis (41, 42). The resulting *p* values were corrected for multiple testing by the Benjamini-Hochberg procedure.

#### *Statistical Analysis*

All analyses were conducted using R v. 3.4.0 (43). Each data set (Male Data Set 1, Male Data Set 2, Female Data Set, and Male Tissue Data Set) was analyzed separately. Only proteins identified with at least two unique peptides were included in the final data set. Quantitative data generated by Progenesis was normalized by log transforming the intensities [ $\log_2(x + 1)$ ]. We followed the method of Keilhauer *et al.* (2015) to determine a “background proteome” for median centering purposes (44). Briefly, we calculated the standard deviation of the intensity profile for each identified protein, ranked the proteins according to the standard deviation of their profile, and selected the bottom 90% of the data. This background proteome was used to median center the distribution of each sample. For the female reproductive tract data set, quantified proteins were confirmed with spectral counts for each condition, as some proteins are expected to be present only in a subset of samples. We treated the proteins that had fewer than three spectral counts in total (among the five replicates in mated or virgin samples) as absent from those samples.

Paired *t*-tests were performed to compare protein intensities between paired male samples (unmated and newly mated male samples

of the same treatment and replicate) and paired female samples (virgin and newly mated female samples of the same replicate). The resulting *p* values were corrected for multiple testing using Benjamini-Hochberg procedure. The  $\log_2$  fold change between the means of the two groups and the negative  $\log_{10}$  of *fdr*-corrected *p* values were plotted against each other to create volcano plots. The quantification data was also used to calculate the abundance of each protein in Male Data Set 1 and Male Data Set 2 separately. Then the known Sfps, the functionally important Sfps and the candidate Sfps were ranked in abundance to compare the estimated copy numbers of candidate and functionally important Sfps against known Sfps in these samples. The significance of the abundance differences was calculated using Kruskal-Wallis rank sum tests.

For the Male Tissue Data Set we ran linear mixed effect models on the subset of known seminal proteins and the high-confidence candidate Sfps identified in this study to test whether the proteins are significantly more abundant in different tissues. We used the *nlme* package in R. For each protein, the initial model included the dissection regime (AG, DU or BO) as a fixed factor and the replicate number as a random factor. Again, the resulting *p* values were corrected for multiple testing using Benjamini-Hochberg procedure.

## RESULTS

### *Male Reproductive Gland Proteomes*

Two data sets of pooled male reproductive tracts were analyzed independently. Candidate Sfps were then identified by applying a set of criteria across the results of both data set analyses.

**Male Data Set 1**—From the 48 samples where 19 male reproductive tracts were pooled, we found a total of 1811 proteins, 1333 of which were identified by at least two unique peptides (supplemental Data S1; supplemental Data S2). We detected 109 (out of a total of 163) known Sfps (Fig. 2A - blue colored proteins), of which 100 were significantly more abundant in unmated samples ( $p \leq 0.02$ ;  $0.3 < \log_2$  fold change <

FIG. 3. Venn diagram displaying the protein overlap between Male Data Set 1 and Male Data Set 2 (A) for known Sfps significantly higher ( $p \leq 0.035$ ) in unmated samples and (B) for the rest of the proteins significantly higher ( $p \leq 0.049$ ) in unmated samples.

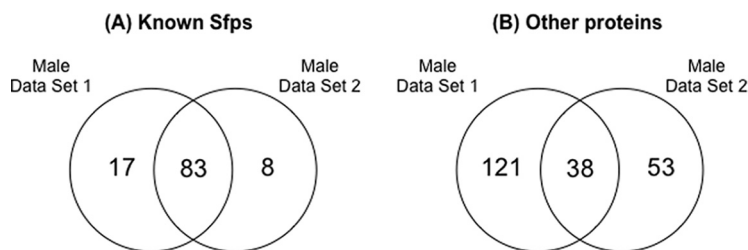


TABLE I

The functional categories of the novel candidate Sfps that are identified in this study

Functional category	Number of new candidates detected
Unknown function	20
Protease	8
Protease inhibitor	5
Lipid metabolism	3
Sperm storage/post-mating behaviour	2
Chitin binding	2
DNA interactions	2
Defense/immunity	1
Cell adhesion	1
Carbohydrate interactions	1
Cell redox homeostasis	1
Metal ion binding	1
Determination of adult lifespan	1
Hormone metabolism	1
Odorant binding	1
Pathogenesis	1
TOTAL	51

3.232; Fig. 2A; Fig. 3A). A further 159 proteins (Fig. 2A - the rest of the proteins on the upper right arm of the volcano plot) were found to be significantly more abundant in unmated samples ( $p \leq 0.048$ ;  $0.106 < \log_2$  fold change  $< 2.842$ ; Fig. 3B). Below we apply a set of criteria to these proteins to derive our new candidate Sfp proteins.

**Male Data Set 2**—From the 30 samples where 20 male reproductive tracts were pooled, we found a total of 2025 proteins, of which 1279 were identified by at least two unique peptides (supplemental Data S1; supplemental Data S2). We detected 109 known Sfps (Fig. 2B - blue colored proteins) and of these 91 were significantly more abundant in unmated samples ( $p \leq 0.036$ ;  $0.29 < \log_2$  fold change  $< 1.982$ ; Fig. 2B; Fig. 3A). Male Data Set 1 and Male Data Set 2 have 83 known Sfps in common that are significantly more abundant ( $p \leq 0.035$ ) in the unmated treatments (Fig. 3A). Another 91 proteins (Fig. 2B - the rest of the proteins on the upper right arm of the volcano plot) were found to be significantly more abundant in unmated samples ( $p \leq 0.049$ ;  $0.277 < \log_2$  fold change  $< 2.161$ ). 38 of these were shared with Male Data Set 1 (Fig. 3B).

#### Candidate Sfps from Male Data Sets

For the proteins that were found to be significantly more abundant in unmated samples in either male data set (excluding the known Sfps) we checked whether they met a set of criteria to determine candidate Sfps. These criteria were:

(1) Significantly higher abundance ( $p \leq 0.05$ ) in unmated male samples (Male Data Set 1) and, if present, higher abundance in unmated male samples (Male Data Set 2)

(2) Significantly higher abundance ( $p \leq 0.05$ ) in unmated male samples (Male Data Set 2) and, if present, higher abundance in unmated male samples (Male Data Set 1)

(3) Presence of a signal peptide

(4) Exclusive expression in accessory glands

We considered proteins that met at least three of the criteria as candidate Sfps. 51 proteins met at least three criteria and are suggested as novel Sfp candidates (Fig. 2). Functional classifications among these 51 proteins included proteases, protease inhibitors, function in sperm storage, chitin binding, lipid metabolism, DNA interactions and female post-mating behavior modification (Table I; supplemental Table S1). These classes are highly similar to the functional classes of known Sfps (23). DAVID analysis for enriched GO terms within the 51 candidate Sfps (using the complete list of known Sfps plus the

candidate Sfps as background) revealed enrichment for presence in extracellular region ( $p = 0.0033$ ). Moreover, candidate Sfps were significantly less abundant than known Sfps in both Male Data Set 1 ( $\chi^2_1 = 38.883$ ;  $p = 4.499e^{-10}$ ) and Male Data Set 2 ( $\chi^2_1 = 28.92$ ;  $p = 7.542e^{-8}$ ; Fig. 4).

We similarly checked for functional enrichment in the remaining up and down regulated proteins in both male data sets and largely detected no significant changes. The only exception was in an analysis of the proteins that were significantly more abundant ( $p \leq 0.049$ ) in newly mated males in Male Data Set 2 (Fig. 2B - proteins on the left arm of the volcano plot) against all the proteins detected in Male Data Set 2, which revealed enrichment for ribonucleoprotein activity ( $p = 3.4e^{-4}$ ), translation ( $p = 0.006$ ), ribosomal proteins ( $p = 0.008$ ), structural constituents of ribosomes ( $p = 0.016$ ) and ribosomes ( $p = 0.042$ ).

#### Female Reproductive Tract Proteome

**Female Data Set**—From the 10 samples where 20 female reproductive tracts were pooled, we found a total of 2150 proteins, of which 1482 were identified by at least two unique peptides (supplemental Data S1; supplemental Data S2). We detected 102 known Sfps, and of these 97 were significantly

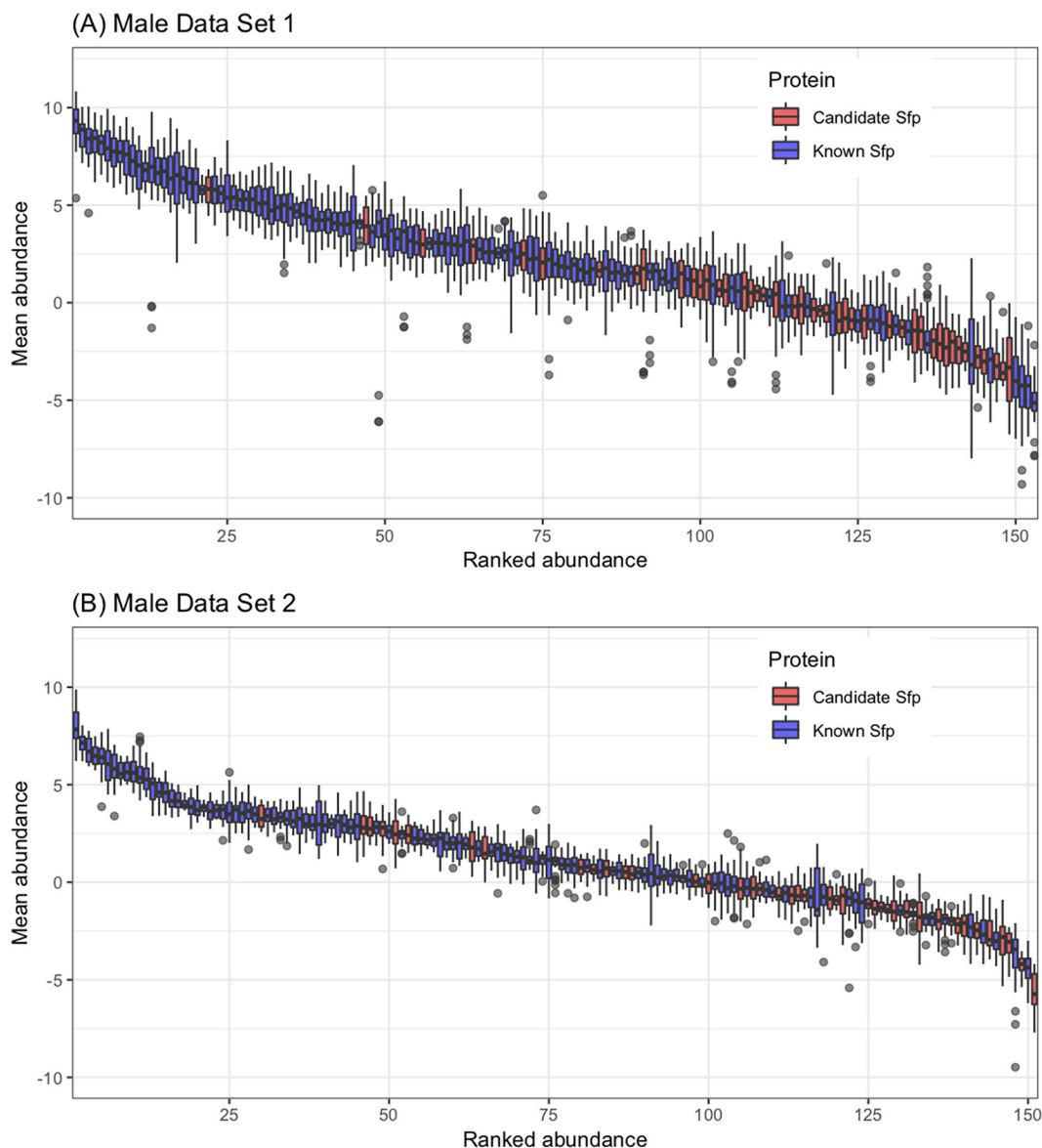
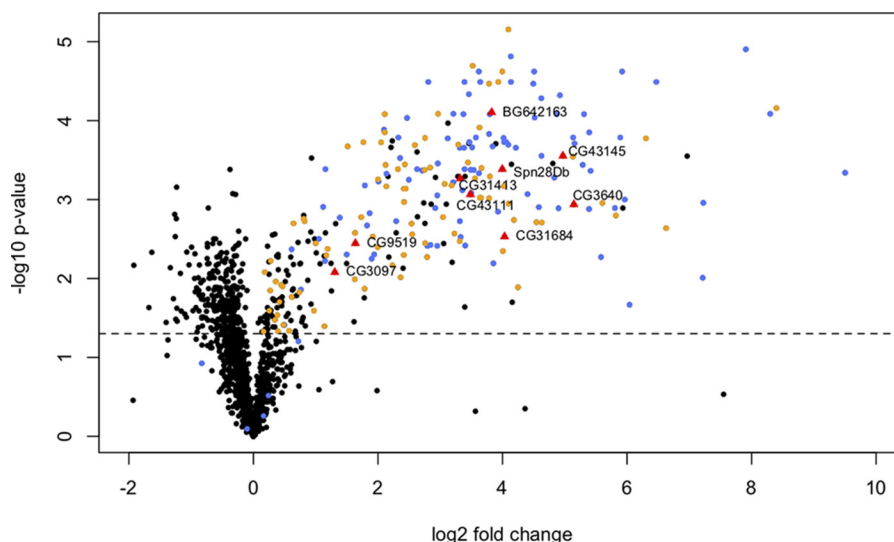


FIG. 4. **Boxplot of the relative abundance of known Sfps and candidate Sfps found in (A) Male Data Set 1 and (B) Male Data Set 2.** Protein abundances were averaged across all the samples in the experiment and were sorted by decreasing order. Known Sfps are colored in blue and candidate Sfps are colored in red.

more abundant in mated samples ( $1.25e^{-5} < p < 0.039$ ;  $0.3 < \log_2$  fold change  $< 3.232$ ; Fig. 5). Although the known Sfps were consistently in higher abundance in mated flies, the data appeared to indicate the presence of some Sfps in virgin females at low abundance. The genes for some of these known Sfps are expressed in virgin females, which could explain their presence, but others are thought to be exclusively expressed in the male accessory glands (38). Of the 73 known Sfps previously identified as exclusively expressed in male accessory glands, virgin samples had more than two spectral counts for four of the proteins (range of 5 to 27 spectral counts), whereas mated samples had more than two spectral counts for 72 proteins (range of 3 to 1017 spectral counts). The other 28 known Sfps had expression profiles

in other tissues (38). Of these 28 Sfps, virgin samples had more than two spectral counts for 14 proteins (range of 4 to 94 spectral counts), whereas mated samples had for all proteins (range of 4 to 201 spectral counts). Another 203 proteins were found to be significantly more abundant ( $p \leq 0.049$ ) in mated female samples. 89 of these proteins are known sperm proteins and are found in the *Drosophila melanogaster* sperm proteome II (45). No enrichment was detected when the rest of the proteins significantly higher in mated females (supplemental Data S3; Fig. 5 - black colored proteins on the upper right arm of the volcano plot) were checked against all the female proteins. However, analysis of the proteins that are significantly more abundant ( $p \leq 0.049$ ) in virgin females (Fig. 5 - proteins on the upper left arm of the volcano plot) revealed



**FIG. 5. Volcano plot displaying all proteins detected in the Female Data Set that were identified by at least two unique peptides.** The  $\log_2$  fold change [ $\log_2$  (newly mated) -  $\log_2$  (virgin)] is shown on the x axis and significance is displayed on the y axis as the negative logarithm ( $\log_{10}$  scale) of the fdr corrected  $p$  value. Known Sfps are colored in blue; sperm proteins are colored in orange and high-confidence candidate Sfps are named, displayed as triangles and colored in red. The significance cutoff ( $p < 0.05$ ) is highlighted with a dashed line.

enrichment for immunoglobulin-like domains (supplemental Table S2).

The 51 candidate Sfps identified from the male data sets using four criteria were checked for two further criteria: **(5)** Significantly higher abundance in mated female samples; and **(6)** Presence in mated and absence in virgin female samples based on spectral counts. Nine of the 41 novel candidate Sfps also met these additional criteria and are therefore classified as high-confidence candidate Sfps (Fig. 5, Table II). Three of these high-confidence candidate Sfps have unknown functions (CG3640, CG43111, BG642163), two are protease inhibitors (CG43145, Spn28Db), one is a protease (CG3097), and one function in cell redox homeostasis (CG31413), lipid metabolism (CG31684) and hormone metabolism (CG9519).

#### Confirmed Known Sfps from Male and Female Data Sets

The known Sfps that were found to be significantly more abundant ( $p \leq 0.05$ ) in unmated samples in either male data set or that were found to be significantly more abundant ( $p \leq 0.05$ ) in mated samples in the female data set were classified as confirmed known Sfps. In total, 125 out of the 163 known Sfps were confirmed in our study (supplemental Table S3). Three known Sfps (CG5267, Sfp79B and Sfp84E) were similarly abundant and one known Sfp (CG15116) was less abundant in unmated samples in either male data set, hence these are at best minor contributors to the ejaculate. We also checked whether Sfps which are known to have functional importance - Acp26Aa, SP, Acp36DE and sex peptide network proteins (CG17575, Lectin-46Ca, Lectin-46Cb, Sems, CG9997, Aqrs, Antr, and Intr) are on average more abundant

TABLE II

The list and classes of the new proteins that are identified as high-confidence candidate Sfps in this study

Protein name	Functional category
CG31413	Cell Redox Homeostasis
CG9519	Hormone metabolism
CG31684	Lipid metabolism
CG3097	Protease
CG43145	Protease inhibitor
Spn28Db	Protease inhibitor
CG3640	Unknown function
BG642163	Unknown function
CG43111	Unknown function

than the rest of the known Sfps. Functionally important Sfps showed a high spread in abundance, but similar non-significant trends for higher abundance in both male data sets (Data set 1:  $\chi^2_1 = 2.9883$ ;  $p = 0.0838$ , Data set 2:  $\chi^2_1 = 2.6863$ ;  $p = 0.101$ ; Fig. 6). Though neither of these results is individually significant, because they test the same hypothesis, we combined the  $p$  values using Fisher's method to determine the overall probability (46). The combined  $p$  value is 0.0489, suggesting that overall the Sfps that have previously shown to have functional importance tend to appear more abundant in our data.

#### Male Accessory Glands and Ejaculatory Duct Proteome

**Male Tissue Data Set**—From the nine samples where 20 reproductive tissues were pooled, we found 1783 proteins, of which 1346 were identified by at least two unique peptides (supplemental Data S1; supplemental Data S2). Of the 117 known Sfps and high-confidence candidate Sfps detected,



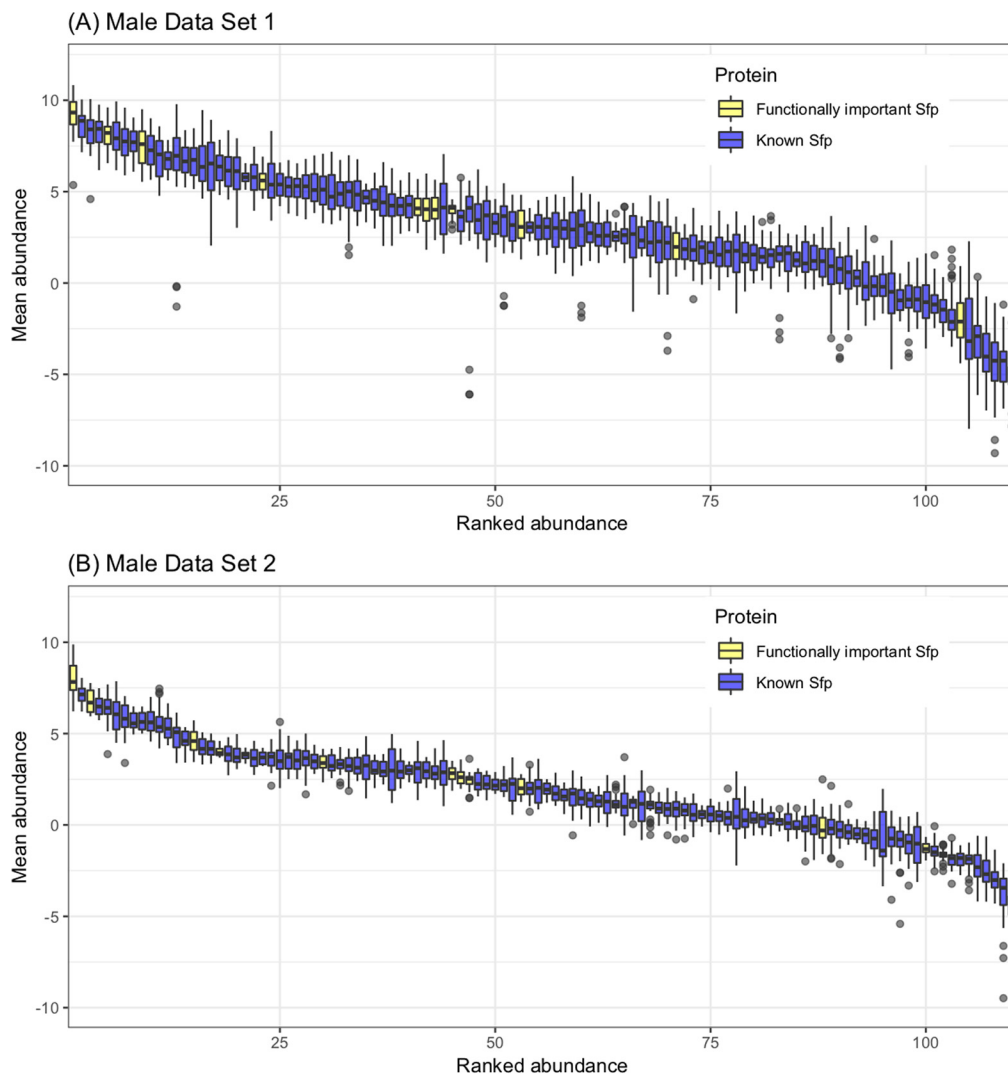


FIG. 6. Boxplot of the relative abundance of functionally important Sfps and the rest of the known Sfps found in (A) Male Data Set 1 and (B) Male Data Set 2. Protein abundances were averaged across all the samples in the experiment and were sorted by decreasing order. Functionally important Sfps are colored in yellow and the rest of the known Sfps are colored in blue.

109 varied in abundances between tissue samples. For 14 of these, protein abundances were significantly higher ( $p \leq 0.012$ ) in the ejaculatory duct than in the accessory gland (DU compared with AG). The abundances of these 14 proteins were similar between samples containing both the ejaculatory duct and accessory gland (BO) and DU samples, except for CG17242 where the protein was significantly more abundant ( $p \leq 0.034$ ) in the DU sample. Among these 14 proteins, 11 were also significantly more abundant ( $p \leq 0.047$ ) in the samples containing both the ejaculatory duct and accessory gland (BO) compared with the AG samples. Hence they are likely primarily or wholly ejaculatory duct specific (Fig. 7). The other three proteins, Est-6, NLaz and Obp56g, were considered candidate ejaculatory duct specific Sfps (supplemental Fig. S2). Although seven of the 11 proteins were known ejaculatory duct proteins (33, 47), the other four Sfps, CG17242, CG18067, CG31704, CG5402, had not previously been linked

to this tissue (supplemental Table S4). DAVID analysis of the 11 proteins against all the known Sfps did not identify any significant classes or functions for the putative ejaculatory duct specific Sfps.

#### DISCUSSION

We used label-free quantitative proteomics to identify candidate Sfps, by comparing the Sfp-producing tissues of males, and the reproductive tracts of females, before and after mating. Using this approach, our data showed consistency with 125 previously known Sfps, detected nine additional proteins that are highly likely to be Sfps, and identified a further 42 proteins as candidate Sfps. Lastly, we revealed that 11 Sfps are mainly stored in the ejaculatory duct, four of which were not previously linked with that tissue. Taken together, these results demonstrate how label-free quantitative proteomics methods, and our tissue-comparison approach,

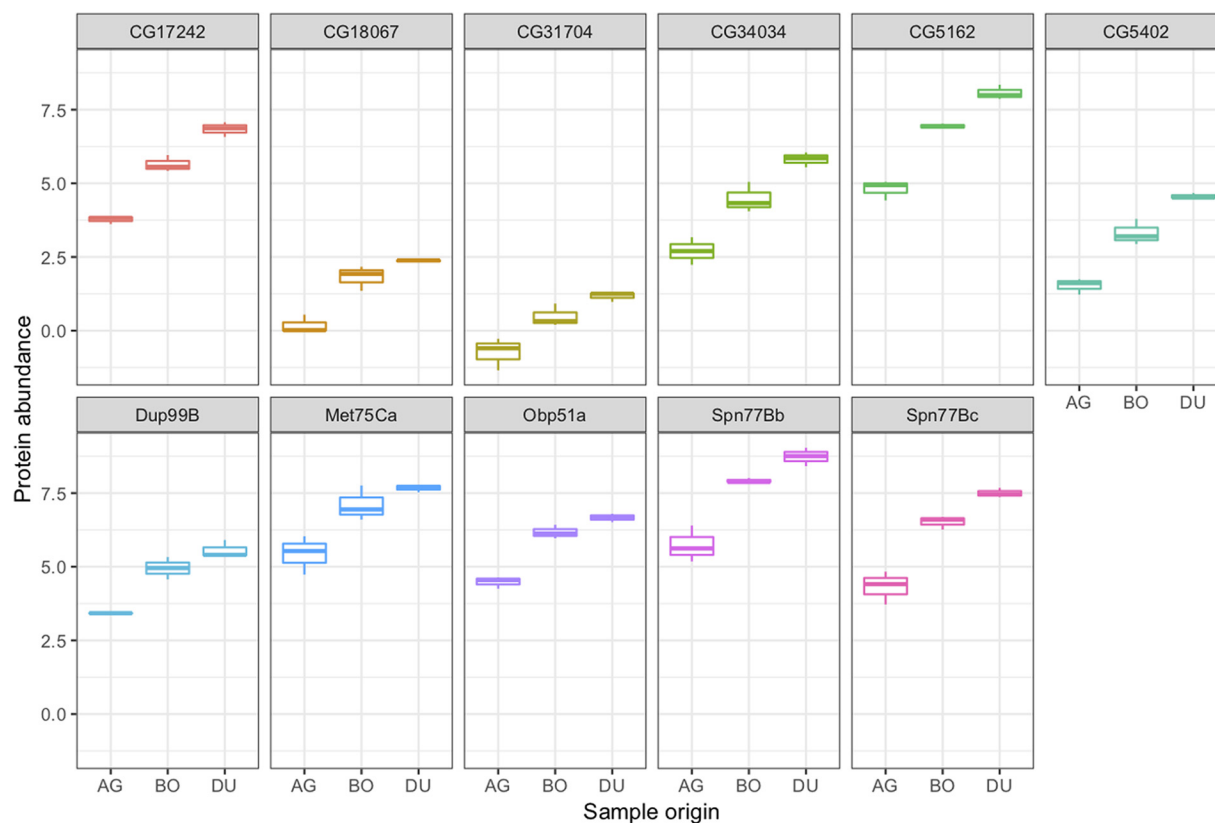


FIG. 7. Boxplot of the abundances of putative ejaculatory duct specific proteins in accessory gland only samples (AG), ejaculatory duct only samples (DU) and samples containing both the ejaculatory duct and accessory gland (BO). The 11 previously known Sfps were significantly more abundant ( $p \leq 0.05$ ) both in DU and in BO compared with AG.

could be used to complement labeling techniques to expand Sfp characterization and localization.

The approach we used here relies largely on just two principles. Sfps should decrease in quantity following mating in male secretory organs (Criteria 1 and 2), and Sfps should appear or increase in quantity following mating in female reproductive tracts (Criteria 5 and 6). Although previous studies have checked whether proteins appear in female reproductive tract following mating to identify Sfps without the usage of labeling techniques (48), a label-free quantitative proteomics approach using male accessory gland proteomes before and after mating has been lacking in the field. This is an important omission for two reasons. Approaches that are solely based on identifying proteins from the female reproductive tract might miss male-derived proteins that get rapidly cleaved/degraded during or soon after ejaculation in the male or female reproductive tract. Likewise, approaches using females are likely to overlook Sfps that are in low abundance, as they will be further diluted within the female reproductive tract. By comparing the reproductive tract proteome of males, our method has the potential to overcome these issues, and provides a complementary method to techniques that are using females.

In this study we used *Drosophila melanogaster*, a species that has its Sfps well characterized through  $^{15}\text{N}$ -labeling (23).

Here, we identify several proteins significantly decreasing in abundance following mating (from two independent male data sets). However, we used additional criteria to use the wealth of knowledge that exist for flies to expand the seminal fluid proteome. We checked whether these proteins had a signal peptide or were exclusively expressed in the accessory glands, as these are common qualities of Sfps. We considered proteins that met at least three out of the first four criteria to be considered a candidate Sfp. Based on all this information, 51 candidate Sfps were identified. We subsequently analyzed female reproductive tracts immediately after copulation to verify the presence of the candidate Sfps, where possible (Criteria 5 and 6). Nine of the 51 candidate Sfps were detected in the females following mating. As the data from the female reproductive tract confirmed the transfer of these nine proteins we suggest them as high-confidence candidate Sfps. The other 42 of the 51 candidate Sfps are of interest as they might represent the set of proteins that avoid detection in females for the reasons set above. Our criteria are conservative because not all Sfps are secreted or exclusively expressed in accessory glands but should ensure that most of our candidates are genuine Sfps. However, because these proteins were not found elevated in abundance in the female reproductive tract after mating we cannot exclude the possibility that these proteins deplete in the male during copulation

for reasons other than being transferred to the female and are therefore not Sfps. Targeted approaches, analyses of increased sensitivity, and measurements taken at earlier copulation time intervals should be used in the future to confirm which of these candidates are true Sfps.

The gene ontology enrichment analysis revealed that the new candidate Sfps we identified are more likely to be present in the extracellular region. The high number of proteases and protease inhibitors identified point toward a very delicately regulated protein system to support sperm function and female postmating behavior. *Drosophila* seminal fluid proteases are known to regulate proteolytic and post-mating reproductive processes (49), hence these candidates warrant further investigation.

Yet, there are no predicted functions for a third of the candidate Sfps. Functional analysis of specific candidates through loss of function or overexpression experiments would be necessary to elucidate the role of these proteins. It is also currently unknown if any of these proteins are cleaved or processed in the ejaculate or once inside the female, and further investigations are necessary to test these possibilities. However, as expected, our analyses did reveal that the candidate Sfps are significantly less abundant than the known Sfps. This finding strengthens the possibility that most of the candidate Sfps were missed out in previous studies using mated females because of their low copy number in the samples, or rapid processing upon ejaculation. Of the known Sfps, those that have been shown to strongly influence post-mating phenotypes overall show a weak but significant tendency toward higher abundance in our data sets. This might be because males tend to express more highly those Sfps that are of highest importance to reproductive success, or alternatively because proteins of higher abundance are more likely to be detected by researchers and therefore analyzed for their function. It will be illuminating to see in the future, when our list of Sfps is more complete and when many more proteins have been analyzed for their function, whether lower abundance proteins tend to be functionally redundant, or whether abundance is independent of functional importance.

Previously, 12 proteins were identified as duct specific in *D. melanogaster* (23, 33, 47, 50). However, only eight of these are known Sfps so only these were considered in our study. We verified seven of these and found four known Sfps to be ejaculatory duct specific. The one Sfp that was not verified is Est-6, which was classified as a candidate ejaculatory duct specific protein in our study. This is because the abundance of Est-6 was similar between accessory gland samples (AG) and combined accessory gland and ejaculatory duct samples (BO). Two of the four known Sfps suggested as ejaculatory duct specific (CG17242, CG31704) were detected in isotopically labeled females mated to DTA-E males, whose main cells are disrupted with diphtheria toxin under the expression of the *Acp95EF* promoter (23, 51) DTA-E male ejaculates lack both sperm and main-cell accessory gland proteins, leading

Findlay *et al.* (23) to conclude that these proteins might be secondary cell derived (23). However, our results support the idea that they are instead stored in the ejaculatory duct, which should also be unaffected by the DTA-E manipulation.

Investigating compositional changes in duct specific and accessory gland-specific proteins, in relation to male and female condition will provide insights as to whether structural compartmentalization influences ejaculate composition. In *Drosophila melanogaster*, it has already been shown that males can adaptively tailor the composition of proteins in the ejaculate to exploit the effects of a previous male's ejaculate. However, the mechanism by which males could adjust the composition of their ejaculate is currently unclear (52). In *Pieris rapae* butterflies, the distinct protein mixtures found in the spermatophore envelope and the inner matrix are stored in separate regions of the male reproductive tract and are transferred to the female sequentially (53). This partitioning is likely to have important implications for how males strategically tailor their ejaculates, or conversely how pathology in specific Sfp-producing compartments impacts ejaculate composition and quality. For example, the Sfps in the human seminal plasma are stored in multiple compartments, each with specific functions (e.g. prostate, ampullary glands, seminal vesicles, bulbourethral glands, and epididymis), thus infections in specific glands will have distinct signatures in the seminal plasma (54). Improving our knowledge of the proteomic contribution of each gland is crucial if we are to understand the mechanisms that generate variation in ejaculate composition.

Gene ontology analysis of proteins that are significantly more abundant in newly mated males (the opposite to Sfps) identified enrichment for translation and ribosome related activity in one of the male data sets. This result is expected considering that males of *D. melanogaster* transfer about one third of their accessory gland contents to the female during each mating, and mating induces the rapid transcription and translation of Sfp genes (26, 55). However, enrichment for translation in newly mated males was only detected in Male Data Set 2, where males were uniformly young (4 days old) but not Male Data Set 1 where males were up to 5 weeks old. Koppik & Fricke (2017) have recently reported a decrease in male Sfp gene expression with advancing age (56), which could explain why no enrichment was observed in Male Data Set 1, which included older males. Similarly, semen volume is known to decrease with age in humans, whereas sperm concentration does not (11). This suggests that at least part of human male reproductive aging is non-sperm components of the ejaculate. Investigating the effects of aging on the male accessory gland proteome is the subject of ongoing work. Moreover, the proteins significantly more abundant in virgin females were enriched for immunoglobulin-like domains that are involved in cell-cell recognition, cell-surface receptors and muscle structure (57). The suppression of proteins related to these functions might be because of the conformational

changes in the female reproductive tract following mating and warrants further investigation.

#### CONCLUSIONS


To understand the role of Sfps in reproduction, it is essential to characterize the full suite of seminal fluid products. In this study, we have described a label-free quantitative proteomics method for Sfp identification that can potentially identify proteins that avoid detection in labeling techniques using females, such as those that are quickly degraded and/or low abundance. We propose both techniques to be used in conjunction for reliable Sfp identification. Our data show that the method is also useful for deciphering the contribution of different male reproductive tissues to the seminal fluid proteome.

**Acknowledgments**—We thank Tim Karr for the invitation to contribute to this special issue and two anonymous reviewers for their constructive comments. We thank Jennifer Perry, Eleanor Bath and Juliano Morimoto for helping with the experiments.

#### DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE (58) partner repository with the data set identifier PXD009451. Annotated spectra for Male Data Set 1 (searched with PEAKS) can be viewed using PRIDE Inspector. The other data sets (searched with Mascot) may be viewed using MS-Viewer (<http://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer>) using the following retrieval keys - Male Data Set 2: jbcy6huzg; Male Accessory Gland and Ejaculatory Duct proteomes: fv0oxzmu0d; Female Reproductive Tract proteome: bhhtu7oslq.

\* I.S. and S.W. were supported by a BBSRC fellowship to S.W. (BB/K014544/1). B.M.K., P.D.C. and R.F. are supported by the Kennedy Trust and John Fell Funds. R.D. is supported by Marie Curie Actions (grant agreement 655392). B.R.H. is funded by the EP Abraham Cephalosporin-Oxford Graduate Scholarship with additional support from the BBSRC DTP.

 This article contains supplemental material.

|| To whom correspondence should be addressed: Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK. Tel.: +441865-271234; E-mail: irem.sepil@zoo.ox.ac.uk.

Author contributions: I.S., B.R.H., and S.W. designed research; I.S., B.R.H., M.-L.T., R.K., and S.W. performed research; I.S., B.R.H., R.D., M.-L.T., P.D.C., R.K., and R.F. analyzed data; I.S., P.D.C., R.F., B.K., and S.W. wrote the paper; R.D. and M.-L.T. contributed new reagents/analytical tools.

#### REFERENCES

- Poiani, A. (2006) Complexity of seminal fluid: A review. *Behav. Ecol. Sociobiol.* **60**, 289–310
- Mcgraw, L. A., Suarez, S. S., and Wolfner, M. F. (2015) On a matter of seminal importance. *BioEssays* **37**, 142–147
- Perry, J. C., Sirot, L., and Wigby, S. (2013) The seminal symphony: How to compose an ejaculate. *Trends Ecol. Evol.* **28**, 414–422
- Hopkins, B. R., Sepil, I., and Wigby, S. (2017) Seminal fluid. *Curr. Biol.* **27**, R404–R405
- Sirot, L. K., Wong, A., Chapman, T., and Wolfner, M. F. (2015) Sexual conflict and seminal fluid proteins: A dynamic landscape of sexual interactions. *Cold Spring Harb. Perspect. Biol.* **7**,
- Rodríguez-Martínez, H., Kvist, U., Emerudh, J., Sanz, L., and Calvete, J. J. (2011) Seminal plasma proteins: What role do they play? *Am. J. Reprod. Immunol.* **66**, 11–22
- Avila, F. W., Sirot, L. K., Laflamme, B. A., Rubinstein, C. D., and Wolfner, M. F. (2011) Insect seminal fluid proteins: identification and function. *Annu. Rev. Entomol.* **56**, 21–40
- Sirot, L. K., and Wolfner, M. F. (2015) in *Cryptic Female Choice in Arthropods: Patterns, Mechanisms and Prospects*, eds Peretti A. V., Aisenberg A. (Springer International Publishing, Cham), pp 351–384
- Gilany, K., Minai-Tehrani, A., Savadi-Shiraz, E., Rezadoost, H., and Lakpour, N. (2015) Exploring the human seminal plasma proteome: An unexplored gold mine of biomarker for male infertility and male reproduction disorder. *J. Reprod. Infertil.* **16**, 61–71
- Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Mindlis, I., Pinotti, R., and Swan, S. H. (2017) Temporal trends in sperm count: A systematic review and meta-regression analysis. *Hum. Reprod. Update* **23**, 646–659
- Johnson, S. L., Dunleavy, J., Gemmill, N. J., and Nakagawa, S. (2015) Consistent age-dependent declines in human semen quality: A systematic review and meta-analysis. *Ageing Res. Rev.* **19**, 22–33
- Jodar, M., Soler-Ventura, A., and Oliva, R. (2017) Semen proteomics and male infertility. *J. Proteomics* **162**, 125–134
- Chapman, T. (2008) The soup in my fly: Evolution, form and function of seminal fluid proteins. *PLoS Biol.* **6**, 1379–1382
- Ramm, S. A., McDonald, L., Hurst, J. L., Beynon, R. J., and Stockley, P. (2009) Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. *Mol. Biol. Evol.* **26**, 189–198
- Dean, M. D., Clark, N. L., Findlay, G. D., Karn, R. C., Yi, X., Swanson, W. J., MacCoss, M. J., and Nachman, M. W. (2009) Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol. Biol. Evol.* **26**, 1733–1743
- Alphey, L., Benedict, M., Bellini, R., Clark, G. G., Dame, D. A., Service, M. W., and Dobson, S. L. (2010) Sterile-insect methods for control of mosquito-borne diseases: an analysis. *Vector Borne Zoonotic Dis.* **10**, 295–311
- Bromfield, J. J., Schjenken, J. E., Chin, P. Y., Care, A. S., Jasper, M. J., and Robertson, S. A. (2014) Maternal tract factors contribute to paternal seminal fluid impact on metabolic phenotype in offspring. *Proc. Natl. Acad. Sci.* **111**, 2200–2205
- Baer, B., Heazlewood, J. L., Taylor, N. L., Eubel, H., and Millar, A. H. (2009) The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics* **9**, 2085–2097
- Borziak, K., Álvarez Fernández, -A., Karr, L. T., Pizzari, T., and Dorus, S. (2016) The Seminal fluid proteome of the polyandrous Red junglefowl offers insights into the molecular basis of fertility, reproductive ageing and domestication. *Sci. Rep.* **6**, 35864
- Druart, X., Rickard, J. P., Mactier, S., Kohnke, P. L., Kershaw-Young, C. M., Bathgate, R., Gibb, Z., Crossett, B., Tsikis, G., Labas, V., Harichaux, G., Grupen, C. G., and de Graaf, S. P. (2013) Proteomic characterization and cross species comparison of mammalian seminal plasma. *J. Proteomics* **91**, 13–22
- Tecirlioglu, R., Hayes, E., and Trounson, A. (2002) Semen collection from mice: electroejaculation. *Reprod. Fertil. Dev.* **14**, 363–371
- Dean, M. D., Findlay, G. D., Hoopmann, M. R., Wu, C. C., Maccoss, M. J., Swanson, W. J., and Nachman, M. W. (2011) Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* **12**, 306
- Findlay, G., Yi, X., MacCoss, M., and Swanson, W. (2008) Proteomics Reveals Novel *Drosophila* Seminal Fluid Proteins Transferred at Mating. *PLoS Biol.* **6**, e178
- Sirot, L. K., Hardstone, M. C., Helinski, M. E., Ribeiro, J. M., Kimura, M., Deewatthanawong, P., Wolfner, M. F., and Harrington, L. C. (2011) Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis* **5**, e989
- Heifetz, Y., Vandenberg, L. N., Cohn, H. I., and Wolfner, M. F. (2005) Two cleavage products of the *Drosophila* accessory gland protein ovulin can independently induce ovulation. *Proc. Natl. Acad. Sci.* **102**, 743–748

26. Sirot, L. K., Buehner, N. A., Fiumera, A. C., and Wolfner, M. F. (2009) Seminal fluid protein depletion and replenishment in the fruit fly, *Drosophila melanogaster*: An ELISA-based method for tracking individual ejaculates. *Behav. Ecol. Sociobiol.* **63**, 1505–1513
27. Sirot, L. K., Findlay, G. D., Sitnik, J. L., Frasher, D., Avila, F. W., and Wolfner, M. F. (2014) Molecular Characterization and Evolution of a Gene Family Encoding Both Female- and Male-Specific Reproductive Proteins in *Drosophila*. *Mol. Biol. Evol.* **31**, 1554–1567
28. Bateman, A. (1948) Intra-sexual sexual selection in *Drosophila*. *Heredity* **2**, 349–368
29. Chapman, T., Arnqvist, G., Bangham, J., and Rowe, L. (2003) Sexual conflict. *Trends Ecol. Evol.* **18**, 41–47
30. Ram, K. R., and Wolfner, M. F. (2007) Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr. Comp. Biol.* **47**, 427–445
31. Ravi Ram, K., Ji, S., and Wolfner, M. F. (2005) Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **35**, 1059–1071
32. Mueller, J. L., Page, J. L., and Wolfner, M. F. (2007) An ectopic expression screen reveals the protective and toxic effects of *Drosophila* seminal fluid proteins. *Genetics* **175**, 777–783
33. Takemori, N., and Yamamoto, M. T. (2009) Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* **9**, 2484–2493
34. Lewis, E. (1960) A new standard food medium. *Drosoph. Inf. Serv.* **34**, 117–118
35. Clancy, D. J., and Kennington, W. J. (2001) A simple method to achieve consistent larval density in bottle cultures. *Drosoph. Inf. Serv.* **84**, 168–169
36. Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cucho, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemerrier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Noupikel, N., Paesano, S., Pedruzzi, I., Pilboud, S., Pozzato, M., Pruess, M., Rivoire, C., Roehert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbruggue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and Zhang, J. (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169
37. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786
38. Celniker, S. E., Dillon, L. A. L., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K. P., and Waterston, R. H. (2009) Unlocking the secrets of the genome. *Nature* **459**, 927–930
39. Levin, Y. (2011) The role of statistical power analysis in quantitative proteomics. *Proteomics* **11**, 2565–2567
40. Fischer, R., and Kessler, B. M. (2015) Gel-aided sample preparation (GASP)-A simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15**, 1224–1229
41. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57
42. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13
43. Team, R. D. C. (2012) R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput. Vienna, Austria*, {ISBN} 3-900051-07-0, <http://www.R-project.org>
44. Keilhauer, E. C., Hein, M. Y., and Mann, M. (2015) Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol. Cell. Proteomics* **14**, 120–135
45. Wasbrough, E. R., Dorus, S., Hester, S., Howard-Murkin, J., Lilley, K., Wilkin, E., Polpitiya, A., Petritis, K., and Karr, T. L. (2010) The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J. Proteomics* **73**, 2171–2185
46. Sokal, R. R., and Rohlf, F. J. (1995) in *Biometry: the principles and practice of statistics in biological research* (Freeman, New York), pp 794–797. 3rd edn.
47. Saudan, P., Hauck, K., Soller, M., Choffat, Y., Ottiger, M., Spörri, M., Ding, Z., Hess, D., Gehrig, P. M., Klausner, S., Hunziker, P., and Kubli, E. (2002) Ductus ejaculatorius peptide 99B (DUP99B), a novel *Drosophila melanogaster* sex-peptide pheromone. *Eur. J. Biochem.* **269**, 989–997
48. Sirot, L. K., Poulson, R. L., Mckenna, M. C., Girnary, H. C., Wolfner, M. F., and Harrington, L. C. (2009) Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. *Insect Biochem. Mol. Biol.* **38**, 176–189
49. LaFlamme, B. A., Ravi Ram, K., and Wolfner, M. F. (2012) The *Drosophila melanogaster* seminal fluid protease “Seminase” regulates proteolytic and post-mating reproductive processes. *PLoS Genet.* **8**, 30–32
50. Sheehan, K., Richmond, R. C., and Cochrane, B. J. (1979) Studies of esterase 6 in *Drosophila melanogaster*. III. The developmental pattern and tissue distribution. *Insect Biochem.* **9**, 443–450
51. Kalb, J. M., DiBenedetto, A. J., and Wolfner, M. F. (1993) Probing the function of *Drosophila melanogaster* accessory glands by directed cell ablation. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8093–8097
52. Sirot, L. K., Wolfner, M. F., and Wigby, S. (2011) Protein-specific manipulation of ejaculate composition in response to female mating status in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **108**, 9922–9926
53. Meslin, C., Cherwin, T. S., Plakke, M. S., Small, B. S., Goetz, B. J., Morehouse, N. I., and Clark, N. L. (2017) Structural complexity and molecular heterogeneity of a butterfly ejaculate reflect a complex history of selection. *Proc. Natl. Acad. Sci.* **114**, E5406–E5413
54. La Vignera, S., Vicari, E., Condorelli, R. A., D'Agata, R., and Calogero, A. E. (2011) Male accessory gland infection and sperm parameters (review). *Int. J. Androl.* **34**, e330–e347
55. Herndon, L. A., Chapman, T., Kalb, J. M., Lewin, S., Partridge, L., and Wolfner, M. F. (1997) Mating and hormonal triggers regulate accessory gland gene expression in male *Drosophila*. *J. Insect Physiol.* **43**, 1117–1123
56. Koppik, M., and Fricke, C. (2017) Gene expression changes in male accessory glands during ageing are accompanied by reproductive decline in *Drosophila melanogaster*. *Mol. Ecol.* **26**, 6704–6716
57. Teichmann, S. A., and Chothia, C. (2000) Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. *J. Mol. Biol.* **296**, 1367–1383
58. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dienes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456