Supplementary material

The *Drosophila* seminal proteome and its role in postcopulatory sexual selection

Philosophical Transactions of The Royal Society B

Authors: Stuart Wigby§ Nora C. Brown§, Sarah E. Allen, Snigdha Misra, Jessica L. Sitnik, Irem Sepil, Andrew G. Clark & Mariana F. Wolfner

Supplemental Data 1
This dataset summarizes previously published sets of candidate and high confidence Sfps in *D. melanogaster*. Information for each gene includes: descriptive genomic information from FlyBase (such as chromosomal position, updated gene identifiers, predicted protein domains, and gene ontology terms); the type of evidence that supports a given gene's inclusion on the list (either proteomic, transcriptomic, or expressed sequence tag); if transfer to females upon mating has been experimentally shown [1]; if that protein has been experimentally identified in the mating plug [2]; if at least one transcript encodes a protein with a predicted signal peptide (SignalP5.0 [3]); whether that gene has published reproductive functions; which tissues of the male reproductive tract its corresponding protein has been experimentally found in [4]. Additional studies that have independently identified the same gene as a candidate Sfp are also noted [1,2,4–21]. The "Transferred Supplemental" tab represents those genes found transferred to females during mating, but which are likely sperm proteins [1]. The "Non-transferred Supplemental" tab are the candidate or low confidence genes which have not been experimentally shown to be transferred to females during mating, but have otherwise been shown to be expressed in male reproductive tissues or shown expression correlation with known high confidence Sfps (but which lack signal peptides or have predicted cellular housekeeping functions, as described in Table S1).

Supplemental Data 2
This dataset is based on the results of a DIOPT search (https://www.flyrnai.org/diopt [22]) of an input of all *Drosophila melanogaster* Sfps, and output of *Homo sapiens* proteins. It thus gives matches between the *D. melanogaster* Sfps and human proteins. We supplemented the table by indicating (a) whether Sfps are known to have functions in *D. melanogaster* PCSS, (b) whether the human proteins form part of the seminal plasma proteome (taken from reference [23]), (c) the expression patterns of human genes in the prostate and testes and how specific their expression is to particular tissues (taken from reference [24]), (d) the ratio of prostate or human gene expression relative to the sum of expression in all other organs combined, to indicate genes that are enriched in these reproductive tissues (e) functional details of the human proteins (for high and moderate hits only) taken from Uniprot [25]. Rows for which there is a high or moderate match between a *D. melanogaster* Sfp and a human gene of reproductive function (either an Sfp or a relevant UniProt function) are coloured light orange (rows 2-217). Where there are hits to different isoforms of the same human gene, text is coloured red. The data is ordered 1) by whether the *D. melanogaster* Sfp matches a human reproductive gene (yes>no), 2) whether the human gene is a human Sfp (yes>no), 3) DIOPT rank (High>Moderate>Low), and 4) DIOPT score (High>Low).

Supplemental Tables

Table S1: Characteristics used to differentiate high confidence Sfps from low confidence "candidate" Sfps. There are 292 high confidence Sfps, 8 transferred candidate (likely sperm)* proteins, and 313 candidate proteins. Note that it possible that genuine Sfps which are made in the seminal vesicle or testis and bind to sperm might be listed in the sperm proteome [26]. Sfps, by definition, are proteins secreted by the male reproductive tract that are transferred to females during mating as part of the ejaculate. Therefore, the best experimental evidence to support the claim that a protein is an Sfp are approaches such as [1] which used heavy nitrogen labelling and proteomics to directly identify male-derived transferred proteins in the female. However, owing to relative protein abundance and stability, additional lines of evidence (such as accessory gland expression and the presence of a signal peptide) may also be useful means of supporting whether or not a particular gene is an Sfp (and are typically used in studies of non-model organisms when heavy nitrogen labelling is not feasible or possible). For proteins which have not been experimentally verified to be transferred to females, those with the appropriate expression pattern (with a signal peptide) should be viewed as more likely candidates than proteins with predicted intracellular functions (regardless of whether or not they have a signal peptide).

High Confidence Sfps	Candidate Sfps
Proteins shown experimentally to be transferred to females during mating [1,15]	Proteins transferred to females, but which are likely sperm proteins* based on their presence in the *Drosophila* sperm proteome
Proteins produced in the male reproductive tract that decrease in abundance after mating [16]	Proteins produced in the male accessory gland without a predicted signal peptide (not experimentally shown to be transferred to females)
Proteins produced in the male accessory gland with a predicted signal peptide [4,17–19]	Proteins produced in the male accessory gland with predicted functions relating to cellular housekeeping or other vital intracellular processes (not experimentally shown to be transferred to females)
Proteins identified in the mating plug with a predicted signal peptide [2]	
Genes that show expression correlation with known Sfps with a predicted signal peptide ("guilt by association") [20]	

Table S2: *D. melanogaster* Sfp genes matched to human seminal plasma and non-seminal plasma genes, using high and moderate confidence DIOPT hits. The full human proteome (20,050 proteins) was derived from reference [24]. Human seminal plasma genes are strongly overrepresented among the *D. melanogaster* hits to the human proteome, almost 3 times higher than expected by chance (proportion test; $\chi^2_1 = 241.2$, $p < 0.001$).

	Hits	No hits
Human seminal plasma genes	110	2036
Human other genes	169	17735

Supplemental Figures

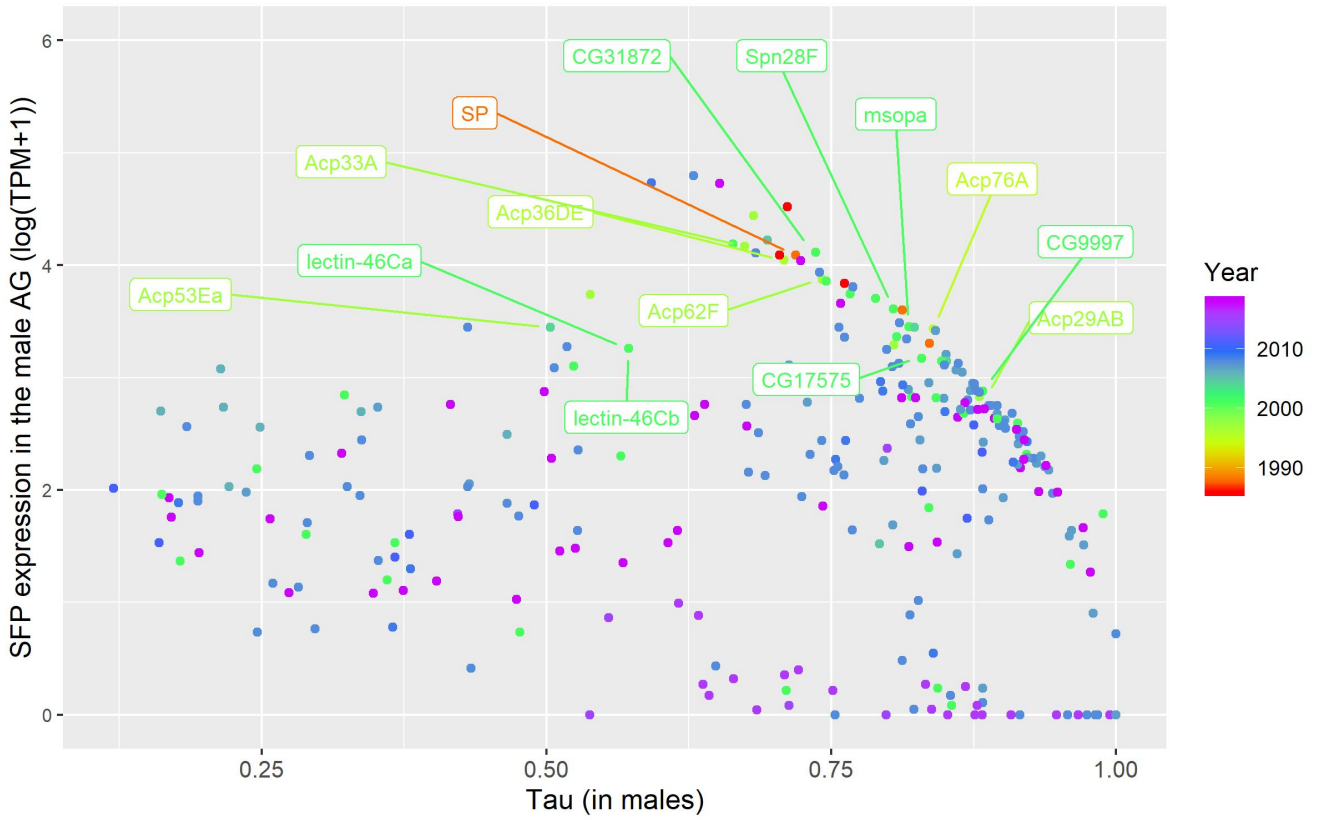

Figure S1: Expression dynamics of high confidence Sfps in *D. melanogaster:* expression specificity in male tissues as measured by Tau (which ranges from 0-1, corresponding to broad expression and complete tissue specificity, respectively [27]) plotted against log-transformed expression values from male accessory gland tissue (FlyAtlas2.0 [28]). The points are colored by the year in which a given Sfp was identified. Many of the earliest identified Sfps have extremely high expression in, and are highly specific to, the male accessory gland. The Sfps noted in Supplemental Data 2 as having a role in PCSS are labeled (note: lectin46Ca and lectin46Cb are the same genes as CG1652 and CG1656, respectively.)

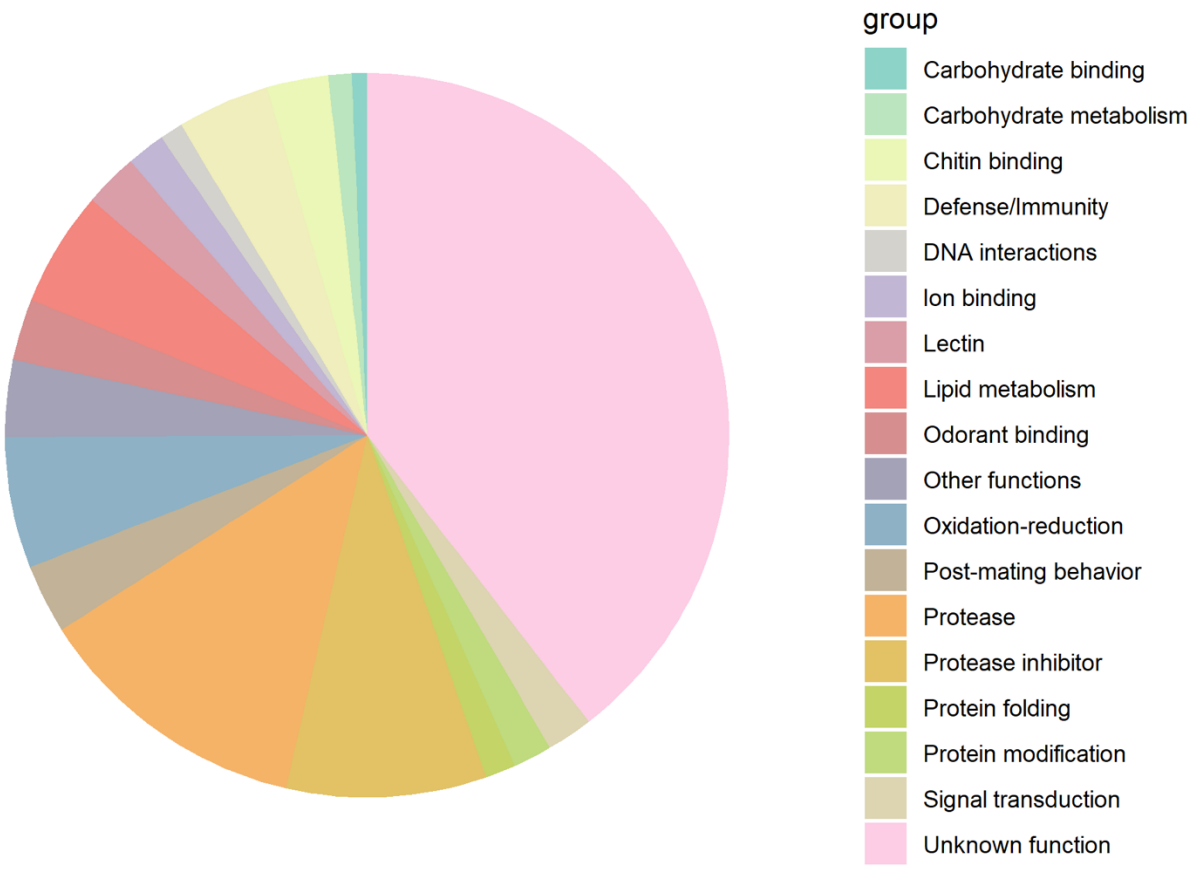

Figure S2: Functional protein classes (derived from FlyBase) associated with the 292 high confidence *Drosophila* Sfps.

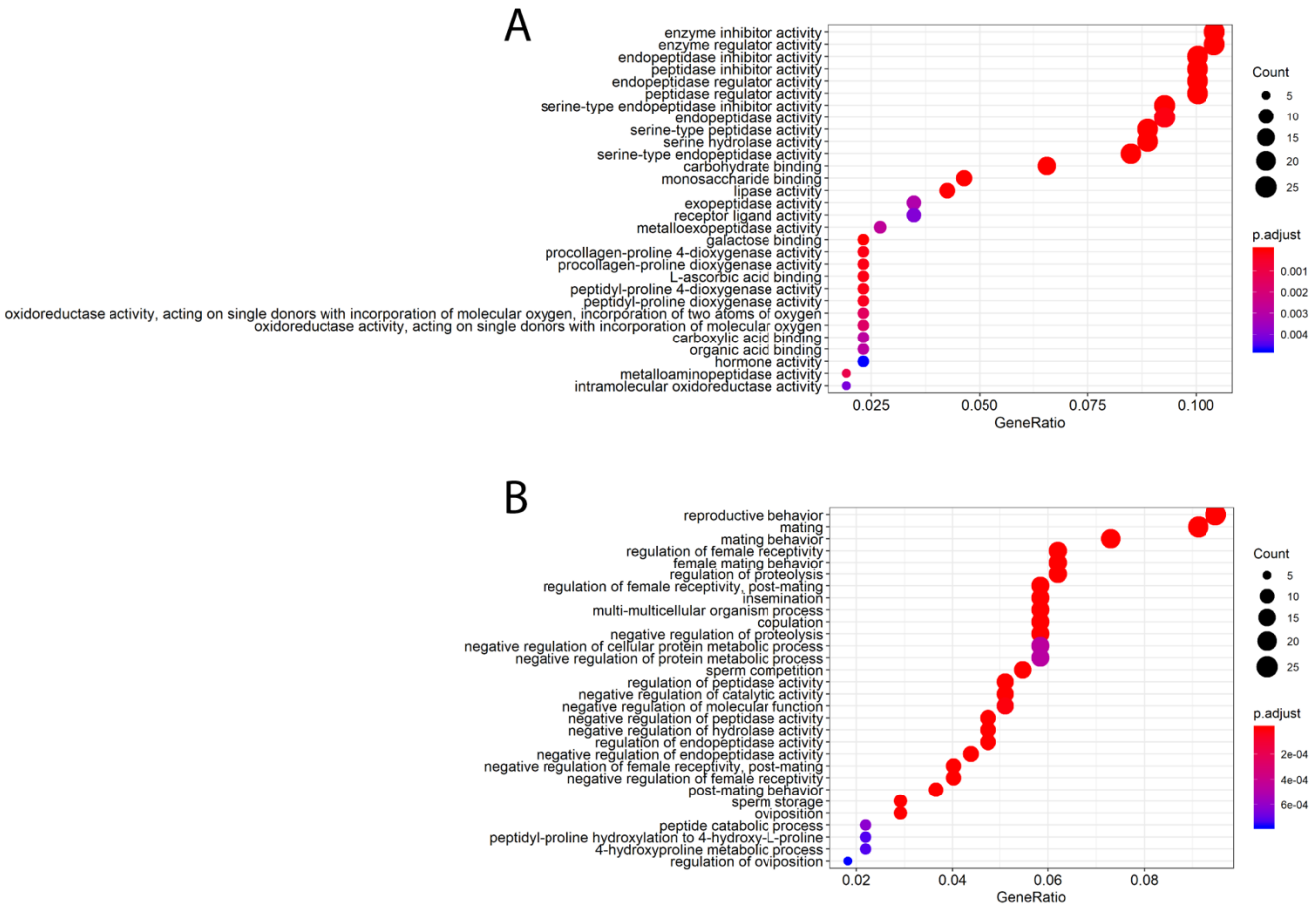

Figure S3: Dot plots of GO enrichment terms (A—Molecular Function; B—Biological Process) of high confidence Sfps in *Drosophila* using ClusterProfiler [29]. The input gene set was the list of Sfps, while the background set was all genes within the *Drosophila* genome. The size of each dot indicates the number of genes associated with a given term, and the color indicates the Benjamini-Hochberg-adjusted p-value. GeneRatio is the number of genes with a given term divided by the number of input genes.

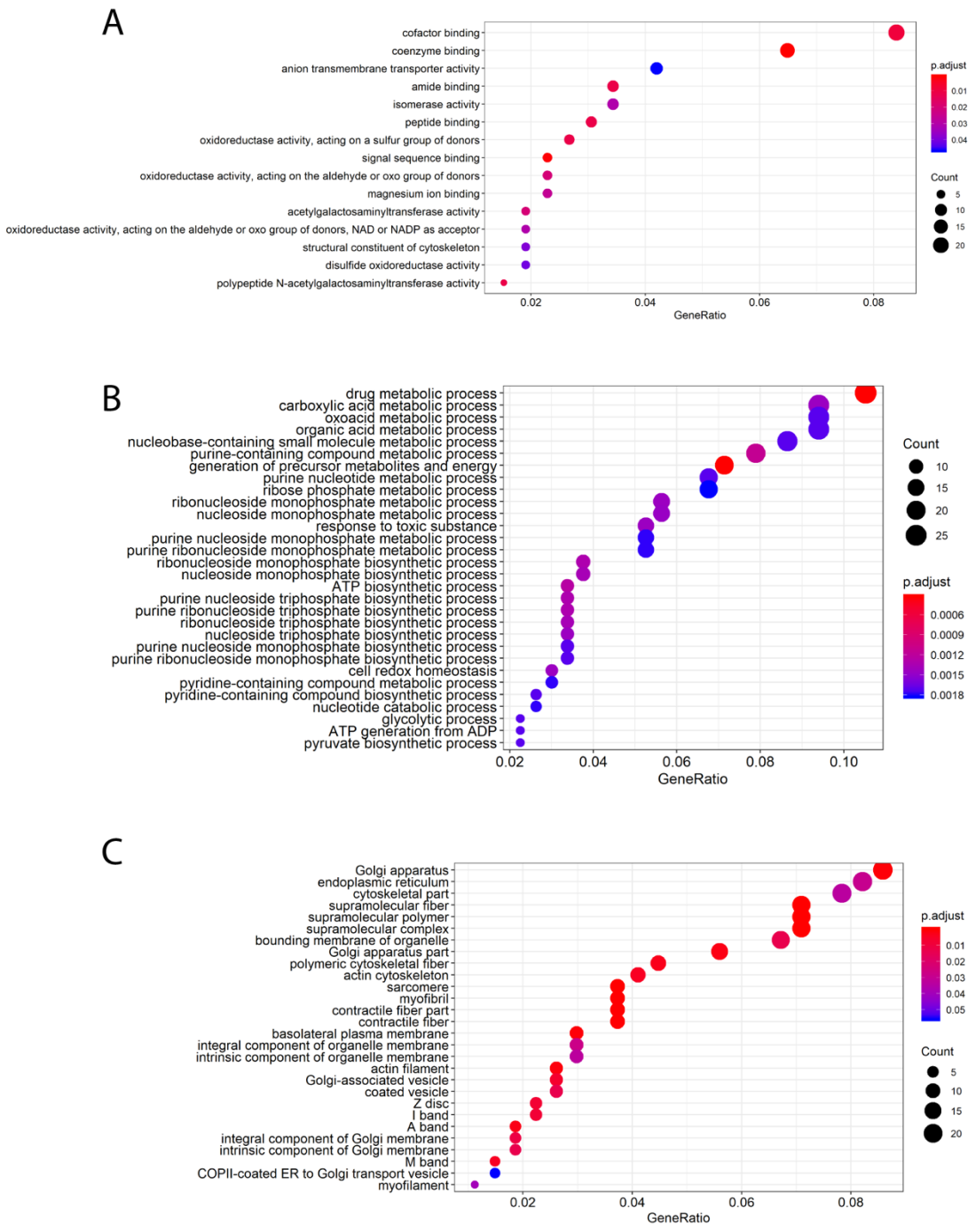

Figure S4: Dot plots of GO enrichment terms (A—Molecular Function; B—Biological Process; C—Cellular Compartment) of candidate Sfps in *Drosophila* using ClusterProfiler [29]. The input gene set was the list of candidate Sfps, while the background set was all genes within the *Drosophila* genome. The size of each dot indicates the number of genes associated with a given term, and the color indicates the Benjamini-Hochberg-adjusted p-value. GeneRatio is the number of genes with a given term divided by the number of input genes. Criteria for classification into the two lists was consistently applied, and enrichment of terms related to cellular function, metabolism, and intracellular compartments suggest many of these genes are not truly Sfps.

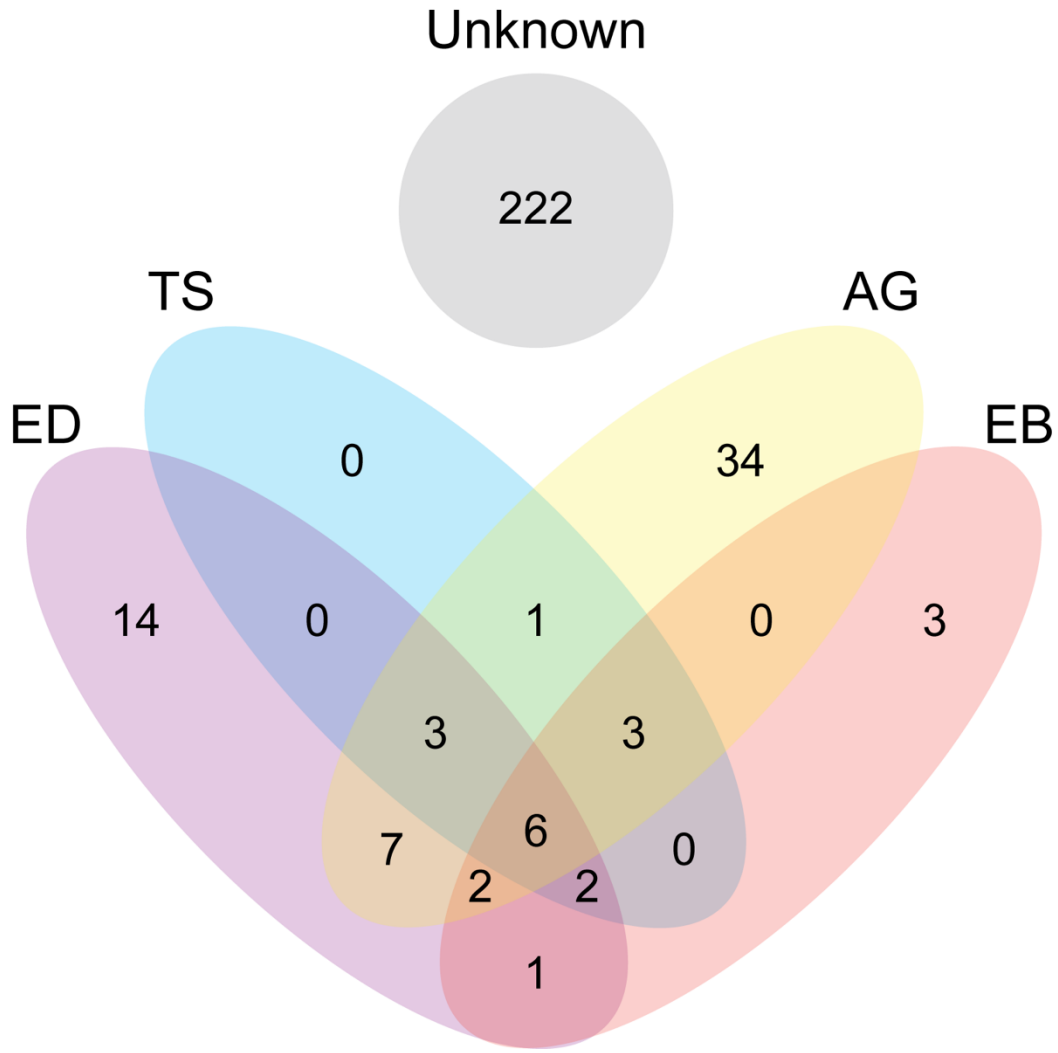

Figure S5: Inferred site of synthesis or storage of 72 high confidence Sfps based on proteomics of dissected male reproductive tract tissues (ED—Ejaculatory Duct; TS—Testes; AG—Accessory Gland; EB—Ejaculatory Bulb [4,16]). Of these 72 proteins, the majority are derived from the male accessory gland, though there are Sfps specific to other male tissues, such as the ejaculatory duct and bulb.

Supplemental References

1. Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2008 Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol.* **6**, e178. (doi:10.1371/journal.pbio.0060178)
2. Avila FW, Cohen AB, Ameerudeen FS, Duneau D, Suresh S, Mattei AL, Wolfner MF. 2015 Retention of ejaculate by *Drosophila melanogaster* females requires the male-derived mating plug protein PEBme. *Genetics* **200**, 1171–1179. (doi:10.1534/genetics.115.176669)
3. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019 SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423. (doi:10.1038/s41587-019-0036-z)
4. Takemori N, Yamamoto MT. 2009 Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* **9**, 2484–2493. (doi:10.1002/pmic.200800795)
5. Chen PSS, Stumm-Zollinger E, Aigaki T, Balmer J, Bienz M, Böhlen P. 1988 A male accessory gland peptide that regulates reproductive behavior of female *D. melanogaster*. *Cell* **54**, 291–298. (doi:10.1016/0092-8674(88)90192-4)
6. Simmerl E, Schäfer M, Schäfer U. 1995 Structure and regulation of a gene cluster for male accessory gland transcripts in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **25**, 127–137. (doi:10.1016/0965-1748(94)00034-F)
7. Wolfner MF *et al.* 1997 New genes for male accessory gland proteins in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **27**, 825–834.
8. Coleman S, DrÄhn B, Petersen G, Stolorov J, Kraus K. 1995 A *Drosophila* male accessory gland protein that is a member of the serpin superfamily of proteinase inhibitors is transferred to females during mating. *Insect Biochem. Mol. Biol.* **25**, 203–207. (doi:10.1016/0965-1748(94)00055-M)
9. Schäfer U. 1986 Genes for male-specific transcripts in *Drosophila melanogaster*. *Mol. Gen. Genet. MGG* **202**, 219–225. (doi:10.1007/BF00331640)
10. Bertram MJ, Wolfner MF, Neubaum DM. 1996 Localization of the *Drosophila* male accessory gland protein Acp36DE in the mated female suggests a role in sperm storage. *Insect Biochem Mol Biol* **26**, 971–980.
11. Begun DJ, Lindfors HA. 2005 Rapid evolution of genomic Acp complement in the *melanogaster* subgroup of *Drosophila*. *Mol. Biol. Evol.* **22**, 2010–2021. (doi:10.1093/molbev/msi201)
12. Holloway AK, Begun DJ. 2004 Molecular evolution and population genetics of duplicated accessory gland protein genes in *Drosophila*. *Mol. Biol. Evol.* **21**, 1625–1628. (doi:10.1093/molbev/msh195)
13. Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, Shirras AD, Isaac RE. 2006 Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase. *Proteome Sci.* **4**, 1–10. (doi:10.1186/1477-5956-4-9)
14. Mueller JL, Ravi Ram K, McGraw LA, Qazi MCB, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005 Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**, 131–143.
15. Findlay GD, MacCoss MJ, Swanson WJ. 2009 Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res.* **19**, 886–895. (doi:10.1101/gr.089391.108)
16. Sepil I, Hopkins BR, Dean R, Thézénas M-L, Charles PD, Konietzny R, Fischer R, Kessler BM, Wigby S. 2019 Quantitative proteomics identification of seminal fluid proteins in male *Drosophila melanogaster*. *Mol. Cell. Proteomics* **18**, S46–S58. (doi:10.1074/mcp.RA118.000831)
17. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001 Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 7375–7379.
18. Ravi Ram K, Wolfner MF. 2007 Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr. Comp. Biol.* **47**, 427–445. (doi:10.1093/icb/icm046)
19. Sitnik JL, Gligorov D, Maeda RK, Karch F, Wolfner MF. 2016 The female post-mating response requires genes expressed in the secondary cells of the male accessory gland in *Drosophila melanogaster*. *Genetics* **202**, 1029–1041. (doi:10.1534/genetics.115.181644)
20. Ayroles JF, Laflamme BA, Stone EA, Wolfner MF, Mackay TFC. 2011 Functional genome annotation of *Drosophila* seminal fluid proteins using transcriptional genetic networks. *Genet. Res. (Camb).* **93**, 387–395. (doi:10.1017/S0016672311000346)
21. Monsma SA, Wolfner MF. 1988 Structure and expression of a *Drosophila* male accessory gland gene whose product resembles a peptide pheromone precursor. *Genes Dev* **2**, 1063–1073.
22. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE. 2011 An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**. (doi:10.1186/1471-2105-12-357)

23. Gilany K, Minai-Tehrani A, Savadi-Shiraz E, Rezadoost H, 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.
24. Fagerberg L *et al.* 2014 Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406. (doi:10.1074/mcp.M113.035600)
25. The Uniprot Consortium. 2019 UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515. (doi:10.1093/nar/gky1049)
26. Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, Polpitiya A, Petritis K, Karr TL. 2010 The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J. Proteomics* **73**, 2171–2185. (doi:10.1016/j.jprot.2010.09.002)
27. Yanai I *et al.* 2005 Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659. (doi:10.1093/bioinformatics/bti042)
28. Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. 2018 FlyAtlas 2: A new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res.* **46**, D809–D815. (doi:10.1093/nar/gkx976)
29. Yu G, Wang L-G, Han Y, He Q-Y. 2012 clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287. (doi:10.1089/omi.2011.0118)