On how to identify a seminal fluid protein: a response to Wigby et al.

Juan Hurtado[1,2], Francisca Cunha Almeida[1,2], Silvina Anahí Belliard[3], Santiago Revale[4], Esteban Hasson[1,2]

[1]Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), CABA, Argentina

[2]Instituto de Ecología, Genética y Evolución de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CABA, Argentina

[3]Laboratorio de Insectos de Importancia Agronómica, IGEAF (INTA), GV-IABIMO (CONICET), Buenos Aires, Argentina

[4]Alchemab Therapeutics Ltd., London, United Kingdom

Abstract

The choice of criteria to delimit a group or class is a subjective matter, even though the reasoning, the objectives, and the criteria themselves should always be clearly stated. This paper is part of a discussion about the criteria used to identify seminal fluid proteins (SFPs) in *Drosophila* species. SFPs are proteins that are transferred to females during copulation together with sperm. The only way to ascertain that a protein is an SFP is to prove that it is produced in a male reproductive organ and is found in the female reproductive tract after insemination. Nevertheless, the required methodology is labor-intensive and expensive, and therefore this kind of data is unlikely to be available for many species, precluding comparative and evolutionary studies on the subject. To conduct evolutionary analyses, in a previous study, we capitalized on the accumulated knowledge we have in the model species *D. melanogaster* to recommend a set of criteria for identifying candidate SFPs in other *Drosophila* species. Those criteria, based on transcriptomic evidence and *in silico* predictions from sequences, would allow a good balance between sensitivity (the inclusion of true SFPs) and specificity (the exclusion of false positives). In view of the criticism raised by another group, here we defend our criteria on one hand while accepting there is room for improvement on the other. The results are updated sets of criteria and SFPs that we believe can be useful in future evolutionary studies.

In 2020, Wigby et al. combined data from several studies on reproductive proteins to provide a list of 292 *Drosophila melanogaster* seminal fluid proteins (SFPs). They referred to these proteins as 'high-confidence SFPs'. By that moment, we were working on a review of the current knowledge on the inter-specific divergence and evolution of the SFP repertoire in the genus *Drosophila* (Hurtado et al., 2022). Using different criteria than those used by Wigby et al. (2020), we arrived at a different SFP list. In view of the differences between the two lists, they wrote a commentary in which they caution against the criteria we used. They focus on a specific portion of our list that is limited to genes showing high and enriched expression in the male accessory glands (MAGs), a pair of glands that constrict at ejaculation transferring their secretions into the anterior region of the ejaculatory duct. On one hand, the authors claim that our list is biased toward MAG-specific genes that may have evolved under lower selective constrain than other SFPs with more pleiotropic functions. On the other hand, they argue that our criteria are too stringent, leading to the exclusion of many good candidates.

Here, we respond to Wigby et al.'s criticisms by rationalizing our criteria and showing their pros and cons. Although we acknowledge that there is room for improvement, we explain why we think our criteria were proper in the context of our aims. We also compare Wigby et al.'s and our criteria in terms of the risks of including false positives and excluding false negatives. Finally, based on the criticisms raised by Wigby et al. and recent new evidence, we revised our criteria and present an updated list of high-confidence SFP candidates.

Most of the findings on *Drosophila* SFPs are restricted to *D. melanogaster* and, as we pointed out in our review, characterizing the seminal proteome of other species is imperative to fill important knowledge gaps on SFP evolution at a broader scale (Hurtado et al., 2022). In fact, one of the purposes of our previous study was to capitalize on the accumulated knowledge we have in *D. melanogaster* to propose criteria that may help identify candidate SFPs in other *Drosophila*. We agree with Wigby et al. that proteomics-based methods (e.g., proteomics combined with sex-specific isotopic labeling and quantitative proteomics) currently represent the "gold standard" in the field of SFP identification. However, if cheaper and less challenging techniques such as RNA-seq provide means by which identifying most SFPs with acceptable precision and sensitivity, criteria based on such techniques may help identify a large number of SFPs in many species.

By 2018, according to our bibliographic search, the number of well-established *D. melanogaster* SFPs for which there is clear evidence of being part of the seminal fluid transferred from males to females during mating was 165 [Table S1 in Hurtado et al., (2022)]. They were confirmed as SFPs by using antibodies, applying mass spectrometry on the mating plug, or combining proteomics with sex-specific isotopic labeling (see references in Hurtado et al., 2022 or Wigby et al., 2020). In 2019, Sepil et al. (2019) implemented a high-sensitivity approach based on quantitative proteomics to identify SFPs that might have been missed by previous approaches. They searched for proteins that after mating become significantly less abundant in the MAGs or the ejaculatory duct but more abundant in the female reproductive tract. While most of the (by then) well-established SFPs were confirmed in that study, 45 proteins were revealed as novel SFP candidates. Nine of these candidates met stringent multiple criteria and were classified by the authors as novel high-confidence SFPs, extending the list of well-established SFPs to 174. In our review, we listed the genes encoding these 174 SFPs and termed them 'known seminal genes' (KSGs) [Table S1 in Hurtado et al., (2022)]. Given the relatively low proportion of the novel KSGs discovered by the recent Sepil et al.'s approach, we argue that most *D. melanogaster* SFPs are likely already known. Thus, we considered that an extended list of SFP genes including, aside from the KSGs, low-confidence candidates may end up with many false positives. Instead, at least for the analyses we performed, we found it preferable to use a more restricted list that includes KSGs alone or with a few additional high-confidence candidates implicating low false-positive risk. In this way, the list would have few (if any) false positives and likely most of the true positives.

Although some KSGs are known to be expressed in the testes, the ejaculatory duct, or the ejaculatory bulb, most (~80%) of them are supplied by the MAGs through merocrine secretion (reviewed in Hurtado et al., 2022). Given that the MAGs are highly specialized in producing and secreting SFPs (Avila et al., 2016), it is very unlikely that a non-housekeeping gene that is highly expressed in the MAGs and encodes a secretory protein does not encode an SFP. The protein product of such a gene, if not extremely unstable, is expected to be found in the MAGs' lumen and, if it is not completely retained there, be transferred to females at mating as part of the seminal fluid. For *D. melanogaster*, conveniently, there are RNA-seq databases (FlayAtlas2 and ModeENCODE) including the MAGs among the tissues for which expression level is reported for every protein-coding gene (Leader et al., 2018; The modENCODE Consortium, 2010); and secretory proteins can be efficiently predicted *in silico* on the basis of sequence information (Almagro Armenteros et al., 2017; 2019). These tools offer the opportunity to search for protein-coding genes that are 1) highly expressed in the MAGs (so their protein products are unlikely scarce there), 2) specifically or mainly expressed in the MAGs relative to other tissues (so it is unlikely they are housekeeping genes), and 3) predicted by different approaches to encode secretory proteins (so their protein products are likely secreted to the glands' lumen). We hold that the vast majority of the genes meeting these three conditions (hereinafter the MAG-secretion conditions) do encode SFPs, and thus they can be included as high-confidence SFP candidates. We identified 46 non-KSGs meeting stringent MAG-secretion conditions (see details in Hurtado et al., 2022) and dubbed them 'unconfirmed high-confidence candidates' (UHCCs). The protein products of 34 of them had been found in the MAGs, ejaculatory duct, or ejaculatory bulb (Sepil et al., 2019; Takemori & Yamamoto, 2009; Walker et al., 2006). As the UHCCs resemble the KSGs regarding both chromosomal location (Fig. 2 in Hurtado et al., 2022) and Ka/Ks ratio (Fig. 3 in Hurtado et al., 2022), we opted to combine them to generate an updated list of 220 *D. melanogaster* SFPs that was used for the analyses presented in our review.

As Wigby et al. (2022) argue, there are SFPs that do not meet the MAG-secretion conditions we evaluated because, for instance, they are expressed in male tissues different from the MAGs, are not conventionally secreted, or are not transcribed at very high levels in the MAGs. Therefore, by leaving aside undiscovered SFPs not highly expressed in the MAGs, our list of UHCCs is surely biased toward MAG-specific genes. Undoubtedly, gene expression data for other relevant male tissues like the ejaculatory duct and bulb would help generate a more comprehensive and unbiased list of UHCCs, but they are not available to date. Nonetheless, we acknowledge that 36 unconfirmed SFP candidates found by Sepil et al. (2019), whose abundance in the male reproductive tract decreases after mating, may also have been considered high-confidence candidates, especially those predicted as secretory proteins. Therefore, here we extend our criteria and update our list accordingly (see below).

Wigby et al. (2022) pointed out that our criteria solely included genes highly and exclusively expressed in the MAGs that have a predicted signal peptide, but this criticism only applies to the 46 UHCCs. We clearly defined four alternative sets of conditions, mainly based on proteomic aspects, so a gene meeting any of them was also included in our list as a KSG (see details in Hurtado et al., 2022). In this sense, our criteria do not markedly differ from Wigby et al.'s. Although they neither explicitly define the alternative sets of evaluated conditions nor distinguish between confirmed and unconfirmed candidates, it is evident that not all of their "high-confidence" SFPs meet the same conditions. For instance, *polyph* (FBgn0033572) and *vsg* (FBgn0045823) are in their list on the bases of different types of evidence. While the evidence indicating that the protein product of *polyph* is transferred from the male into the female reproductive tract during mating is solely based on proteomics data, the inclusion of *vsg* is exclusively based on MAG transcription evidence (from an EST analysis) and the predicted ability to encode a signal peptide (Table S1 in Wigby et al., 2020). Aside from *vsg*, many other genes of Wigby et al.'s list (e.g., FBgn0033165, FBgn0037204, FBgn0052383, FBgn0032080, and FBgn0024234) were

considered SFP candidates solely because they meet MAG-secretion like conditions (i.e., based on expression patterns in the MAGs and the potential ability to encode a signal peptide). None of their protein products were detected by Findlay et al. (2008; 2009) among the male-derived proteins transferred at mating, by Sepil et al. (2019) among the proteins whose abundance decreases in the male reproductive tract or increases in the female reproductive tract right after mating, or by Avila et al. (2015) among the mating plug proteins. Therefore, both datasets, Wigby et al.'s and ours, probably bear similar biases.

A question that arises is to what extent the use of an SFP list including a group of candidates biased toward MAG expression and canonical secretion is problematic for evolutionary analyses, as Wigby et al. (2022) claim. We already showed that Ka/Ks ratio does not differ between the KSGs and our UHCCs (Fig. 3 in Hurtado et al., 2022). To further explore this potential bias, here we estimate gene birth and death rates across the *Drosophila* genus for the KSGs and UHCCs separately using Notung-2.9.1.5 (Chen et al., 2000; Darby et al., 2017), following the same approach we previously implemented for estimating such rates for both groups of genes together (Hurtado et al., 2022). Our new analysis revealed that the two groups have similar event rates (0.013 and 0.014 events/gene/million years for KSGs and UHCCs, respectively). Given the similarity between the KSGs and UHCCs in Ka/Ks ratio and turnover rates, if the addition of the UHCCs to our SFPs' list introduces an evolutionary constraint bias, such a bias is unlikely obvious or notable.

Another criticism raised by Wigby et al. (2022) is that our requisites for UHCCs were too stringent, and thus we left a number of good candidates out of the list. As mentioned above, we agree with Wigby et al. that our criteria can be improved to the extent of including the proteins proposed by Sepil et al. (2019) as SFP candidates. Nonetheless, the question of how stringent the requisites for SFP candidates should be is a debatable, subjective matter that should be addressed in light of the available evidence and particular purposes of the analysis. While the use of less stringent conditions would allow the inclusion of a larger number of good candidates, it would also increase the risk of including false positives. As we already stated, we favored the use of stringent conditions to search for UHCCs because we considered that most SFPs were already included in our list as KSGs. Wigby et al. (2020) decided to use less stringent conditions to search for candidates and, consequently, they ended up with a more extended list. Therefore, 99 of the genes Wigby et al. classified as high-confidence candidates were considered by us as neither KSGs nor UHCCs. Some of them, however, were allocated to our list as (non-high-confidence) candidates. This is the case, for instance, of some Sepil et al. (2019)'s candidates which, despite meeting the MAG-secretion conditions at some point, fail to reach the MAG expression level thresholds we used.

Eighty "high-confidence" candidates in Wigby et al.'s list are not even among our non-high-confidence candidates. For instance, FBgn0035933, FBgn0053462, and FBgn0259226 encode a predicted signal peptide but are barely expressed in the testes or MAGs and their protein products were not detected in any of the proteomic studies aimed to identify SFPs. While we cannot rule out the possibility that these genes encode SFPs, we consider that —until further evidence pointing so is found— most of them should not be listed as high-confidence SFP candidates. Similarly, neither FBgn0261989, FBgn0262484, FBgn0262536, nor *Drs* (FBgn0283461) should be classified as high-confidence candidates from our point of view. Although these genes were included in Wigby et al.'s list because they encode secretory proteins found in the mating plug (Avila et al., 2015; Wigby et al., 2020), they are barely or not expressed in any male reproductive tissue but are highly expressed specifically in the female spermatheca, from where they are likely secreted into the lumen of the female reproductive tract.

The more stringent conditions we applied for searching for unconfirmed candidates do not account for all the differences between Wigby et al.'s and our list. Strikingly, the more relaxed criteria applied by

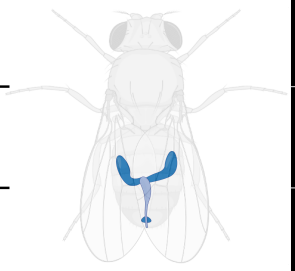
Wigby et al. left 24 of our UHCCs out of their list (e.g., FBgn0260462, FBgn0085328, and FBgn0263249). Besides being very highly expressed in the MAGs and encoding secretory proteins, the proteins of most of these UHCCs were detected by Sepil et al. (2019) in the male reproductive tract or the reproductive tract of mated females. Therefore, they seem better SFP candidates than many of Wigby et al.'s candidates with low MAG expression levels and for which proteomics-based evidence is lacking (e.g., FBgn0010406, FBgn0053462, and FBgn0262005).

Three other genes in our list [FBgn0038395, *S-Lap7* (FBgn0033868), and FBgn0054034] were excluded from Wigby et al.'s list because they are known to encode sperm proteins. Although SFPs are defined as non-sperm components of the ejaculate (i.e., ejaculate proteins produced by non-sperm cells of the male reproductive tract), they may be also produced by sperm cells (McCullough et al., 2022). The protein products of these three genes were confirmed to be transferred —as part of the soluble fraction of the ejaculate— into the female reproductive tract at mating (Findlay et al., 2008). Furthermore, they were detected in the MAGs and ejaculatory duct proteomes (Sepil et al., 2019) and are predicted to locate in the extracellular space. Therefore, we believe they should be listed as high-confidence SFPs.

A few months after our review was published, McCullough et al. (2022) combined quantitative proteomics with sex-specific isotopic labeling to monitor the post-testicular life history of the sperm proteome. Interestingly, their results reveal that many SFPs associate with sperm cells during or soon after ejaculation, which may have important implications for the functions of the seminal fluid proteome. One hundred thirty-two KSGs, 18 UHCCs, and 18 of our non-high-confidence candidates not classified as KSGs were among the SFPs that the authors found associated with sperm in the postmating uterus. These results confirm that at least 39% of our UHCCs are true SFP genes.

Overall, we hold that our criteria implicate acceptable precision and sensitivity. We recognize, however, that they can be improved; and we believe that the relaxation of the MAG expression requirements for those genes with protein products that were detected in the MAGs, ejaculatory duct, or ejaculatory bulb, will lead to a more comprehensive and less MAG-biased list of SFPs. In view of this and the recent findings by McCullough et al. (2022) on the sperm proteome, we decided to revise our criteria to provide an updated list of *D. melanogaster* SFPs (Fig. 1). In brief, our expanded criteria now incorporate Sepil et al. (2019)'s candidates and non-canonical sperm proteins that were found in the female uterus associated with sperm (McCullough et al., 2022) but exclude our previous unconfirmed candidates whose protein products have not been found in the male reproductive tract.

condition	evidence	criteria									
		KSGs					UHCCs				
		1	2	3	4	5	6	7	8	9	10
A: gene with enriched expression in the MAGs	transcriptomics			✓				✓		✓	
B: gene highly and mainly expressed in the MAGs	transcriptomics				✓	✓					✓
C: protein found in the MAGs, ED, or EB	proteomics			✓		✓		✓	✓	✓	✓
D: Sepil et al. (2019)'s candidate	proteomics, transcriptomics, and *in silico* prediction						✓				
E: canonical sperm protein	proteomics				X				X		
F: mating plug protein	proteomics			✓				✓			
G: gene expressed in the female reproductive tract	transcriptomics			X				X			
H: confirmed SFP before Sepil et al. (2019)	proteomics or immuno-detection	✓									
I: transferred male-derived sperm-associated protein	proteomics				✓	✓			✓	✓	
J: Sepil et al. (2019)'s high-confidence candidate	proteomic, transcriptomic, and *in silico* prediction		✓								
K: protein with signal peptide	*in silico* prediction from sequence								✓	✓	✓
L: extracellular-space protein	*in silico* prediction from sequence			✓		✓	✓		✓	✓	✓
well-established SFP (KSG)							X	X	X	X	X





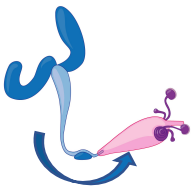
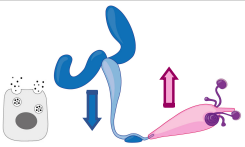
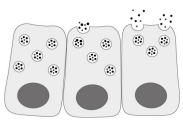




Fig. 1. Revised criteria for high-confidence *D. melanogaster* SFPs. The first column stands for the evaluated conditions. Numbered columns represent each of the alternative sets of conditions for SFP. Genes meeting set of conditions 1, 2, 3, 4, or 5 are included in the list as KSGs (i.e., genes encoding well-established SFPs or genes for which evidence showing so is —from our perspective— almost unequivocal). Genes meeting set 6, 7, 8, 9, or 10 are included as (still) UHCCs because the evidence supporting them as candidates is convincing but not

conclusive. **A:** Genes that are at least moderately expressed in the MAGs (>25 RPKM and >25 FPKM in modENCODE and FlyAtlas2, respectively), exhibit higher expression levels in this tissue relative to the testes and most of the other adult male tissues, and show MAG-expression enrichment relative to the male whole body (enrichment > 1 in FlyAtlas2). **B:** Genes that are at least 'very highly' expressed in the MAGs and exhibit expression enrichment in this tissue, or genes that are at least 'highly' expressed in the MAGs and exhibit at least 'very high' MAG-expression enrichment. In more detail, this condition includes genes that 1) show >100 RPKM and >100 FPKM in the MAGs (in modENCODE and FlyAtlas2, respectively) and higher expression levels in this tissue relative to the testes and most of the other adult male tissues (in FlyAtlas2), or 2) show >50 RPKM and >50 FPKM in the MAGs, higher expression levels there relative to the testes and most of the other adult male tissues, and a four times MAG-expression enrichment relative to the male whole body (enrichment > 4 in FlyAtlas2). **C:** Genes encoding proteins found in the MAGs, ejaculatory duct (ED), or ejaculatory bulb (EB) (Walker et al., 2006; Takemori & Yamamoto, 2009; Sepil et al., 2019). **D:** Genes classified as SFP candidates by Sepil et al. (2019). This includes genes that, besides showing high expression enrichment in the MAGs or encoding a signal peptide, encode proteins whose abundance in the MAGs or ejaculatory duct decreases right after mating. **E:** Genes encoding proteins of sperm collected from the male seminal vesicles, where sperm cells become motile and are stored until ejaculation (McCullough et al., 2022; Wasbrough et al., 2010). **F:** Genes encoding mating plug proteins (Avila et al., 2015; Wigby et al., 2020). **G:** Genes expressed in the female reproductive tract, with >10 FPKM in the ovaries or spermathecae (in FlyAtlas2). **H:** Genes encoding well-established SFPs before Sepil et al. (2019). These genes were confirmed as SFPs by using antibodies, applying mass spectrometry on the mating plug, or combining proteomics with sex-specific isotopic labeling (see references in Hurtado et al., 2022). **I:** Genes encoding male-derived proteins that are associated with transferred sperm in the uterus of recently mated females (McCullough et al., 2022). **J:** Genes classified as high-confidence SFPs by Sepil et al. (2019). This includes Sepil et al. (2019)'s candidates whose protein products become more abundant in the female uterus right after mating. **K:** Genes encoding a signal peptide according to SignalP v5.0 (Almagro Armenteros et al., 2019). **L:** Genes encoding secretory proteins according to DeepLoc v1.0 (Almagro Armenteros et al., 2017).

The new list contains 294 high-confidence SFP genes, 30 of which had not been previously proposed as SFPs (Table S1). Using the same denominations as before, we classified the genes into the —somewhat subjective— categories of KSGs and UHCCs. Two hundred thirty-six genes were considered KSGs and 58, UHCCs. We also provide a list of 105 genes that were classified as non-high-confidence SFP candidates, which includes novel candidates as well as previously predicted seminal genes (Table S2). More than three-quarters of the KSGs (77%) are mainly expressed in the MAGs and likely encode a secretory protein, and 150 KSGs (64%) meet our updated MAG-secretion conditions (conditions B, K, and L, Fig. 1). We found 189 *D. melanogaster* protein-coding genes meeting our updated MAG-secretion conditions, and only 29 of them are so far not classified as well-established SFP genes or high-confidence candidates based on proteomic evidence. These figures suggest that stringent criteria based on expression level in the MAGs (as well as the ejaculatory duct and bulb) and predicted ability to encode secretory proteins can provide reliable information to identify SFPs in other *Drosophila* species until proteomics-based evidence is generated.

The list of *D. melanogaster* high-confidence SFPs is planned to be periodically updated and is available at https://github.com/hurtadojuan/SFPs.

Acknowledgments

We thank Julián Mensch for his insights on the first draft of the manuscript. This work was supported by The National Scientific and Technical Research Council (Argentina).

Supporting information

Table S1. List of the *D. melanogaster* genes classified as high-confidence SFP genes. The evaluated conditions for the applied criteria are shown for each gene.

Table S2. List of additional, non-high-confidence SFP candidate genes. The evaluated conditions for the applied criteria are shown for each gene.

Find both tables at https://github.com/hurtadojuan/SFPs.

References

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, 33(21), 3387–3395.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420–423.

Avila, F. W., Cohen, A. B., Ameerudeen, F. S., Duneau, D., Suresh, S., Mattei, A. L., & Wolfner, M. F. (2015). Retention of ejaculate by *Drosophila melanogaster* females requires the male-derived mating plug protein PEBme. *Genetics*, 200(4), 1171–1179.

Avila, F. W., Sánchez-López, J. A., McGlaughon, J. L., Raman, S., Wolfnert, M. F., & Heifetz, Y. (2016). Nature and functions of glands and ducts in the *Drosophila* reproductive tract. In E. Cohen & B. Moussian (Eds.), *Extracellular Composite Matrices in Arthropods* (pp. 411–444). Springer, Cham.

Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology*, 7(3–4), 429–447.

Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., & Durand, D. (2017). Xenolog classification. *Bioinformatics*, 33(5), 640–649.

Findlay, G. D., MacCoss, M. J., & Swanson, W. J. (2009). Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Research*, 19(5), 886–895.

Findlay, G. D., Yi, X., Maccoss, M. J., & Swanson, W. J. (2008). Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biology*, 6(7), e178.

Hurtado, J., Almeida, F. C., Belliard, S. A., Revale, S., & Hasson, E. (2022). Research gaps and new insights in the evolution of *Drosophila* seminal fluid proteins. *Insect Molecular Biology*, 31(2), 139–158.

Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A., & Dow, J. A. T. (2018). FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Research*, 46(D1), D809–D815.

McCullough, E. L., Whittington, E., Singh, A., Pitnick, S., Wolfner, M. F., & Dorus, S. (2022). The life history of *Drosophila* sperm involves molecular continuity between male and female reproductive tracts. *Proceedings of the National Academy of Sciences of the United States of America*, 119(11).

Sepil, I., Hopkins, B. R., Dean, R., Thézénas, M. L., Charles, P. D., Konietzny, R., Fischer, R., Kessler, B. M., & Wigby, S. (2019). Quantitative proteomics identification of seminal fluid proteins in male Drosophila melanogaster. *Molecular and Cellular Proteomics*, 18, S46–S58.

Takemori, N., & Yamamoto, M. T. (2009). Proteome mapping of the Drosophila melanogaster male reproductive system. *Proteomics*, 9(9), 2484–2493.

The modENCODE Consortium. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*, 330(6012), 1787–1797.

Walker, M. J., Rylett, C. M., Keen, J. N., Audsley, N., Sajid, M., Shirras, A. D., & Isaac, R. E. (2006).

Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase. *Proteome Science*, 4, 1–10.
Wasbrough, E. R., Dorus, S., Hester, S., Howard-Murkin, J., Lilley, K., Wilkin, E., Polpitiya, A., Petritis, K., & Karr, T. L. (2010). The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *Journal of Proteomics*, 73(11), 2171–2185.
Wigby, S., Brown, N. C., Allen, S. E., Misra, S., Sitnik, J. L., Sepil, I., Clark, A. G., Wolfner, M. F., Wigby, S., Clark, A. G., & Wolfner, M. F. (2020). The *Drosophila* seminal proteome and its role in postcopulatory sexual selection. *Philosophical Transactions of the Royal Society, B: Biological Sciences,* 375(1813), 20200072.
Wigby, S., Brown, N. C., Sepil, I., & Wolfner, M. F. (2022). On how to identify a seminal fluid protein: a commentary on Hurtado et al. *Insect Molecular Biology,* In press.