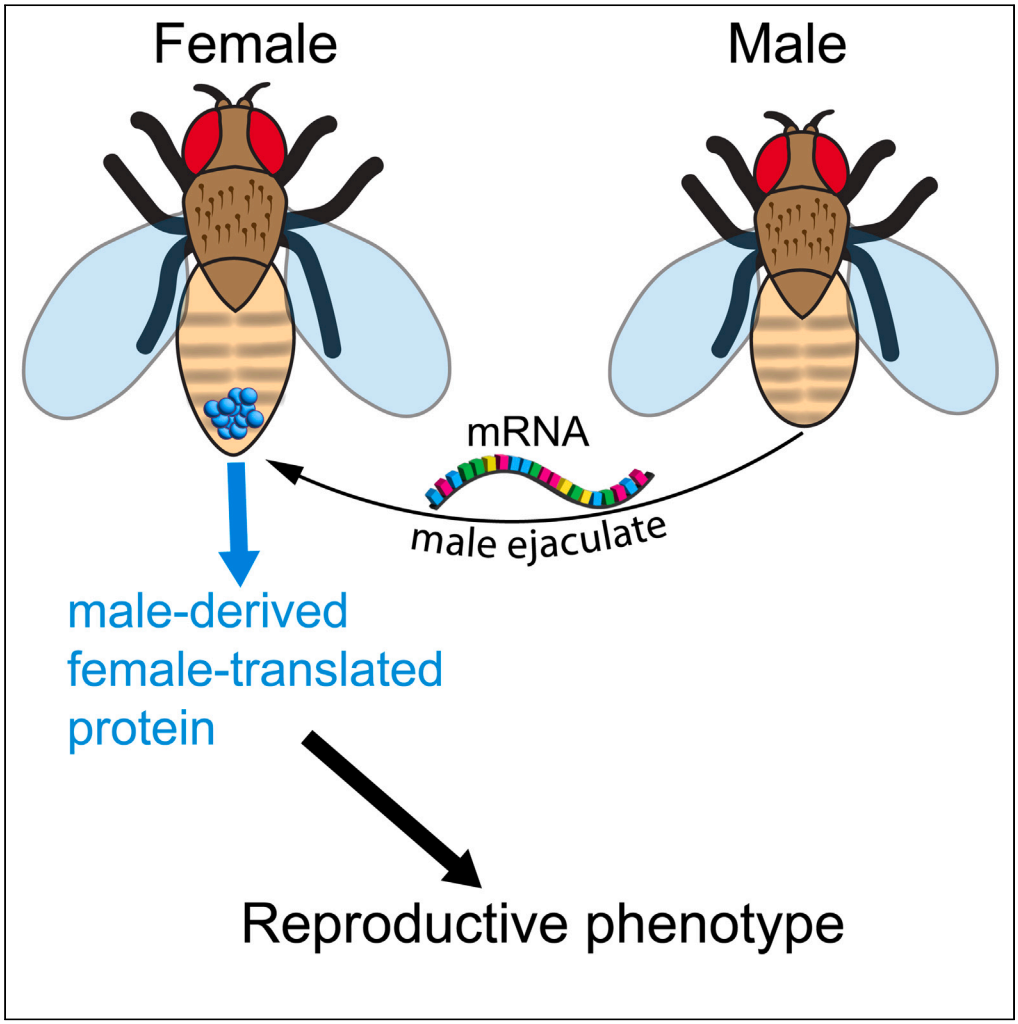
iScience

CellPress
OPEN ACCESS

Article

# Females translate male mRNA transferred during mating




Luciano M. Matzkin, Jeremy M. Bono, Helen K. Pigage, ..., Clinton C. Green, Jeffrey B. Callan, Stephen P. Delahunt II

lmatzkin@arizona.edu (L.M.M.)
jbono@uccs.edu (J.M.B.)

## Highlights

Coding and non-coding RNA are a common feature of male seminal fluid

We developed an experimental approach to investigate the role of transferred RNA

Male-derived female-translated proteins (mdFTPs) contribute to proteome of females

mdFTPs have significant functional effects on reproductive phenotypes

Matzkin et al., iScience 27, 110442
August 16, 2024 
https://doi.org/10.1016/j.isci.2024.110442

iScience

CellPress
OPEN ACCESS

Article

# Females translate male mRNA transferred during mating

Luciano M. Matzkin,[1,2,3,7,8,*] Jeremy M. Bono,[4,7,*] Helen K. Pigage,[4] Carson W. Allan,[1] Fernando Diaz,[1,5] John R. McCoy,[4] Clinton C. Green,[4] Jeffrey B. Callan,[4,6] and Stephen P. Delahunt II[4]

## SUMMARY

**Although RNA is found in the seminal fluid of diverse organisms, it is unknown whether it is functional within females. We developed a proteomic method (VESPA, Variant Enabled SILAC Proteomic Analysis) to test the hypothesis that *Drosophila* male seminal fluid RNA is translated by females. We found 67 male-derived, female-translated proteins (mdFTPs) in female lower reproductive tracts, many with predicted functions relevant to reproduction. Knockout experiments indicate that mdFTPs play diverse roles in postmating interactions, affecting fertilization success, and the formation/persistence of the insemination reaction mass, a trait hypothesized to be involved in sexual conflict. These findings advance our understanding of reproduction by revealing a mechanism of postmating molecular interactions between the sexes that strengthens and extends male influences on reproduction in previously unrecognized ways. Given the diverse species that carry RNA in seminal fluid, this discovery has broad significance for understanding molecular mechanisms of cooperation and conflict during reproduction.**

## INTRODUCTION

Reproductive success depends on complex molecular interactions between males and females that integrate morphological, physiological, and behavioral responses to mating.[1,2] Although fundamental to all sexually reproducing species, the mechanistic bases of many important postmating molecular interactions are still poorly understood in most organisms.[1,2] Males contribute more than just sperm to this process. Male seminal fluid is a complex mixture of diverse components, which not only aid in sperm survival and delivery to the oocyte but also interact directly with female reproductive tissues to facilitate fertilization and trigger lasting changes in female physiology and behavior.[1–4] While most attention has focused on the important roles played by seminal fluid proteins (SFPs) in driving the female postmating response, the functional significance of other seminal fluid components has received less attention.[1,2,4–6]

The presence of coding and noncoding RNA in sperm and seminal fluid is a common feature of male ejaculates, having been found in diverse organisms including humans, *Drosophila*, and mosquitoes.[5,7–13] While sperm RNA has received increasing attention and effects on developing offspring have been established, the functional significance of seminal fluid RNA remains unknown.[5,14] In some species, including humans, seminal fluid RNA is carried in extracellular vesicles (EVs).[5,15] The potential for EVs to serve as a mechanism for interorganismal signaling leads to the hypothesis that male RNA could be delivered to recipient cells in the female reproductive tract where it might perform critical functions.[5,15] The protein-coding potential of mRNA and diverse regulatory roles non-coding RNA plays in many fundamental biological processes[16] suggest a novel mechanism by which males may mediate reproductive outcomes. This would have broad implications for understanding cooperative interactions between males and females that facilitate fertilization and antagonistic interactions within and between the sexes resulting from sexual selection and sexual conflict.

Building on our previous research showing the transfer of *Drosophila arizonae* male RNA to females during copulation,[9] here we test the hypothesis that male seminal fluid mRNA is translated into protein by females. Identification of male-derived, female-translated proteins (mdFTPs) requires not only differentiating male and female proteins within the female reproductive tract but also the source of transcripts from which these proteins were produced. To overcome this challenge, we developed VESPA (Variant Enabled SILAC Proteomic Analysis), an experimental approach and bioinformatic pipeline that combines interspecies hybridization with metabolic proteome labeling of whole flies[17–21] to identify mdFTPs (Figure 1).

[1]Department of Entomology, University of Arizona, Tucson, AZ, USA
[2]BIO5 Institute, University of Arizona, Tucson, AZ, USA
[3]Department of Ecology and Evolutionary Biology, Tucson, AZ, USA
[4]Department of Biology, University of Colorado Colorado Springs, Colorado Springs, CO, USA
[5]Present address: Department of Life, Earth and Environmental Sciences, West Texas A&M University, Canyon, Texas, USA
[6]Present address: Graduate Professional Studies, Rabb School of Continuing Studies, Brandeis University, Waltham, Massachusetts, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: lmatzkin@arizona.edu (L.M.M.), jbono@uccs.edu (J.M.B.)
https://doi.org/10.1016/j.isci.2024.110442

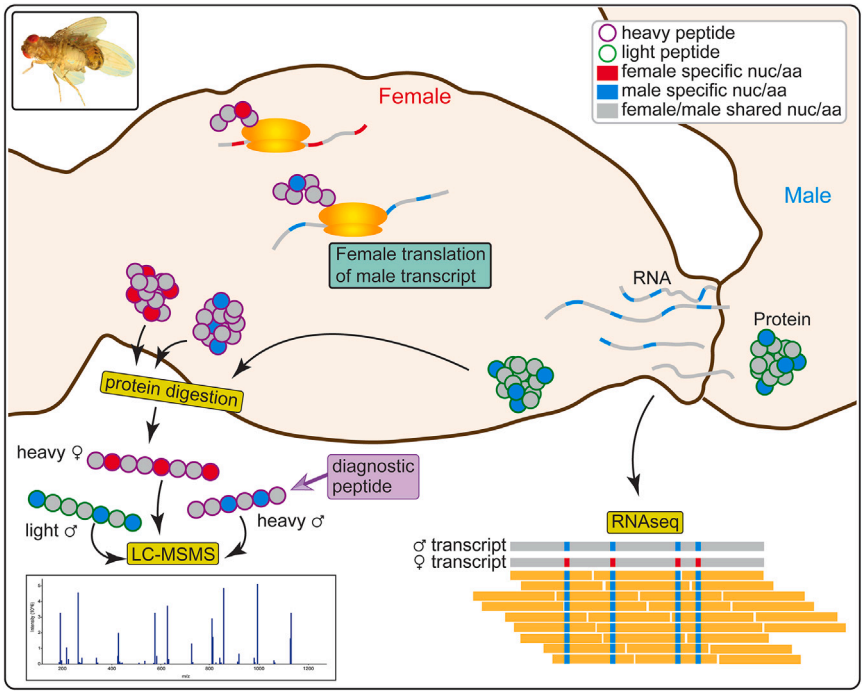



**Figure 1. Overview of VESPA**

VESPA differentiates proteins made by males and females by isotopic labeling and the source of RNA transcripts by fixed nucleotide (nuc) differences between species. After mating, female reproductive tracts are removed, proteins are extracted and digested, and samples are analyzed by liquid chromatography-tandem mass spectrometry. Peptides that match the amino acid (aa) sequence of the male but carry the heavy label of the female are diagnostic for mdFTPs.

## RESULTS

### Male seminal fluid RNA is translated by females

To identify mdFTPs using VESPA, we metabolically labeled the proteome of female *Drosophila mojavensis* using L-Lysine-$^{13}C_6$,$^{15}N_2$ (Lys 8). Heavy-labeled *D. mojavensis* females were mated with unlabeled *D. arizonae* males, which allows proteins produced by males and females to be differentiated within the female reproductive tract. While most peptide sequences are shared between the species, unique peptides resulting from fixed nonsynonymous substitutions distinguish the source of translated mRNA transcripts. Peptides that match the sequence of the male but the heavy isotope label of the female are diagnostic for mdFTPs (Table 1).

Lower reproductive tracts of heavy labeled *D. mojavensis* females mated to unlabeled *D. arizonae* males were removed at 6 h postmating for proteomic analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The 6-h time point was chosen based on previous results showing male RNA was still detectable in the female reproductive tract at this time[9] and because this allows the capture of proteins translated over the course of several hours. This experiment was independently repeated three times. Mass spectra were analyzed using MaxQuant[22] and MSFragger,[23] with a combined database containing all annotated proteins in *D. mojavensis* and *D. arizonae*. These analyses resulted in a total of 290,733 (MaxQuant) and 889,134 (MSFragger) peptide spectrum matches (PSMs) across all three replicates using a false-discovery rate (FDR) threshold of 0.01. To find evidence for mdFTPs, we identified heavy-labeled PSMs that matched the *D. arizonae* database sequence (heavy-*arizonae*, HA PSMs; Table 1). To further assess confidence in HA PSMs, we compared features of these identifications with light-*mojavensis* (LM; Table 1) PSMs. We do not expect LM PSMs since the proteome of female *D. mojavensis* was heavy labeled. Thus, any LM identifications are either false positives or due to incomplete labeling. Although we used a conservative FDR, this provides an additional control to assess whether the set of HA peptides is likely to be enriched for false positives. If HA peptides are enriched for false positives, we would expect to identify similar numbers compared to LM peptides since search algorithms should be equally likely to produce erroneous HA or LM hits. We identified an average of 7.2X more HA PSMs relative to LM PSMs across replicates (range: 4.1–9.1; Table S1), which further bolsters confidence in the overall validity of the set of HA PSMs. Output data from MaxQuant and MSFragger for diagnostic HA PSMs is included in Table S2.

After subsequent filtering to remove potential false positives resulting from polymorphism or leucine/isoleucine substitutions, a total of 234 and 187 unique HA peptides were identified by MaxQuant and MSFragger, respectively (Table S3). To be considered a mdFTP, we required proteins to be identified by at least two heavy peptides (i.e., we filtered out so-called "one-hit wonders"), one of which had to be a diagnostic HA peptide while the second could be diagnostic or non-diagnostic (heavy-*arizonae*-*mojavensis*, HAM peptides; Table 1). This resulted in a total of 166 unique mdFTPs. Twenty-six mdFTPs were identified by multiple distinct HA peptides (range: 2–7), and 58 were identified in two or three replicates (Table S4). Fifty-nine mdFTPs are potentially produced exclusively from male mRNA, as we did

**Table 1. Description of the six different types of peptides that could be identified by database searches**

| Peptide | Description |
|---|---|
| Heavy-*mojavensis* (HM) | Heavy labeled peptide with unique *mojavensis* sequence. Diagnostic for female proteins. |
| Heavy-*arizonae-mojavensis* (HAM) | Heavy labeled peptide with sequence shared by both species. Non-diagnostic for female proteins and mdFTPs. |
| Light-*arizonae* (LA) | Unlabeled peptide with unique *arizonae* sequence. Diagnostic for male SFPs. |
| Light-*arizonae-mojavensis* (LAM) | Unlabeled peptide with sequence shared by both species. Expected for male SFPs or incomplete label incorporation in female proteins. |
| Light-*mojavensis* (LM) | Unlabeled peptide with unique *mojavensis* sequence. Not expected because *mojavensis* female proteome was heavy labeled. Either false positives or from incomplete labeling. |
| Heavy-*arizonae* (HA) | Heavy labeled peptide with unique *arizonae* sequence. **Diagnostic for mdFTPs.** |

not identify HM peptides (heavy-*mojavensis*; Table 1) that would indicate the protein also being produced from female RNA. It is important to note that shotgun proteomics relies on a data-dependent acquisition strategy, where only a subset of peptides is selected for analysis. The stochastic nature of the selection process often results in peptides that are identified in only a subset of replicates.[24,25] Given that mdFTPs can only be identified by the relatively small number of diagnostic peptides that differ in sequence between *D. mojavensis* and *D. arizonae*, those with multiple diagnostic HA peptides, identified in multiple replicates, or by both search algorithms, have the strongest level of support (67 mdFTPs; Figure 2A). Confidence in these mdFTPs is further strengthened by the high quality of associated diagnostic peptide identifications. On average, mdFTP diagnostic peptides had lower posterior error probabilities (PEP; MaxQuant) and higher peptide prophet confidence scores (MSFragger) relative to LM peptides, and these metrics were similar to all other non-LM peptides (high-support mdFTPs; Figure S1). The remaining 99 mdFTPs were only identified in a single replicate, and, on average, diagnostic peptide confidence scores (PEP and peptide prophet probability) were similar to LM peptides and lower than highly supported mdFTP and other non-LM peptides (low support mdFTPs; Figure S1). Nevertheless, these mdFTPs are supported by a median of seven heavy peptides total when including non-diagnostic peptides (HAM peptides; Table S5). Moreover, female diagnostic HM peptides were not identified for 41 of these proteins, which further suggests they may be produced from male RNA. We conclude that while some mdFTPs in the low-support group are likely valid identifications, the group as a whole may be enriched for false positives.

## mdFTPs are supported by the presence of male RNA transcripts in female reproductive tracts

To investigate whether male RNA transcripts of mdFTPs were more likely to be detected in the reproductive tracts of heterospecifically mated females than transcripts of other proteins, we performed RNA-seq of *D. mojavensis* female lower reproductive tracts after mating with *D. arizonae* males. Although male transcripts can be identified by fixed nucleotide differences between the species, this is complicated by the fact that some male-transferred transcripts could be of low abundance and/or also be produced by females, resulting in observed variable sites in sequencing data. To evaluate evidence for male transcripts, we developed a metric called the RNA Transfer Index (RTI), which is the proportion of variable sites in a gene having at least five reads of male origin. Overall, transcripts of mdFTPs showed evidence of significant enrichment of male reads compared to non-mdFTPs. Specifically, mean RTI was ~4X higher in the set of 67 highly supported mdFTPs and ~2X higher in the remaining set compared to non-mdFTPs (Figure 2B). Transcripts of 11 mdFTPs appear to be supplied almost exclusively by males (>95% of reads are of male origin), whereas other mdFTP transcripts ranged from being mostly male in origin to mostly female. The lack of female diagnostic HM peptides for many mdFTPs with female RNA transcripts suggests that the presence of female transcripts is not necessarily indicative of translation into protein. Overall, transcriptomic data provides strong support for a positive association between mdFTPs identified through proteomics and the presence of male RNA transcripts identified by RNA-seq in female lower reproductive tracts.

## mdFTPs are largely distinct from proteins transferred in the male seminal fluid

Male seminal fluid proteins (SFPs) play critical functional roles in the postmating response of diverse organisms.[2,4] To determine whether mdFTPs overlap with SFPs, we analyzed the proteomes of unlabeled *D. arizonae* female reproductive tracts immediately after copulation with Lys4 (L-lysine-2HCL,4,4,5,5-D4) labeled *D. arizonae* males. This analysis identified a total of 207 SFPs that are transferred to females during mating (Table S6). Only ~14% overlapped with the overall set of mdFTPs (Figure 2C), suggesting that mdFTPs may target distinct aspects of the postmating process or function through alternative pathways relative to SFPs.

## Functional clusters implicate mdFTPs in the postmating response

Gene ontology (GO) and protein domain enrichment analyses of mdFTPs revealed several enriched functional clusters with links to the postmating response (Figure 3A). For example, proteins with a CAP domain were highly enriched. This domain is associated with cysteine-rich secretory proteins (CRISPs), which have diverse roles in fertility in mammals and insects.[26–28] GO-term analysis performed on mdFTP orthologs in *D. melanogaster* (126/166 mdFTPs had orthologous calls) revealed enrichment for processes related to metabolism and oxidative stress response. This is consistent with the fact that sperm must remain viable within the female reproductive tract potentially

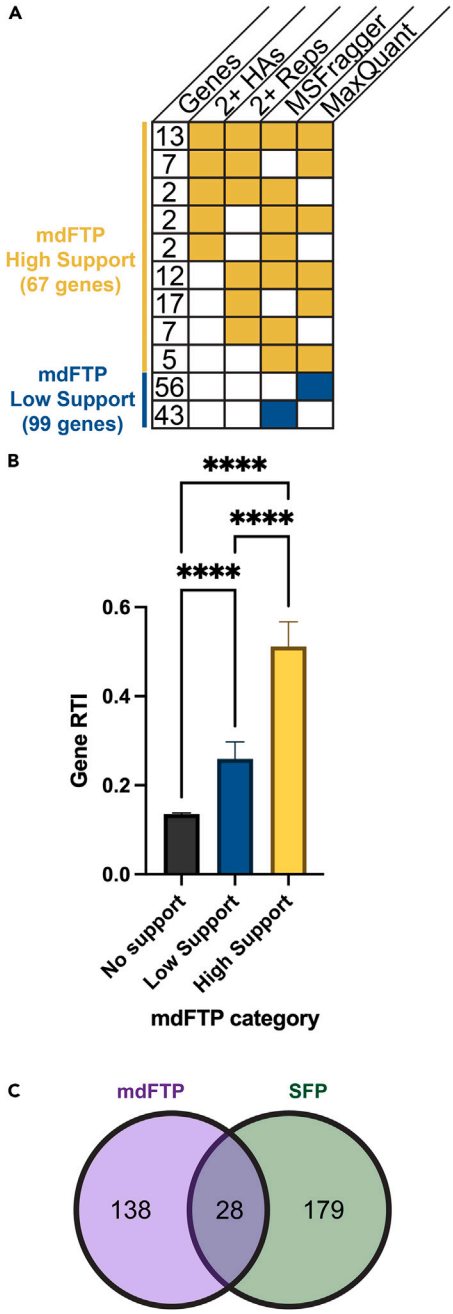



**Figure 2. mdFTPs are supported by multiple lines of evidence**

(A) All mdFTPs were identified by a minimum of two heavy peptides, with at least one being diagnostic. Figure indicates the number of genes per mdFTP type, and the filled boxes indicate the properties of each type. The most strongly supported mdFTPs include those identified by multiple diagnostic peptides, in two or more replicates, and/or by both MaxQuant and MSFragger.

(B) Mean RNA transfer index (RTI) per gene was higher for high-support mdFTPs [yellow in (A), $n = 67$] and low-support mdFTPs ($n = 99$) compared to all other genes with RNA-seq data ($n = 9727$). (GLMM: gene category $\chi^2 = 812.8$, $p = 2.2e^{-16}$). *Post hoc* tests were performed using Tukey's method (****$p \leq 0.0001$). Error bars represent standard error of the mean.

(C) Venn diagram showing overlap of mdFTPs with *D. arizonae* SFPs. Low overlap suggests mdFTPs may perform different functions from other proteins in the ejaculate or were not detected by our methods.

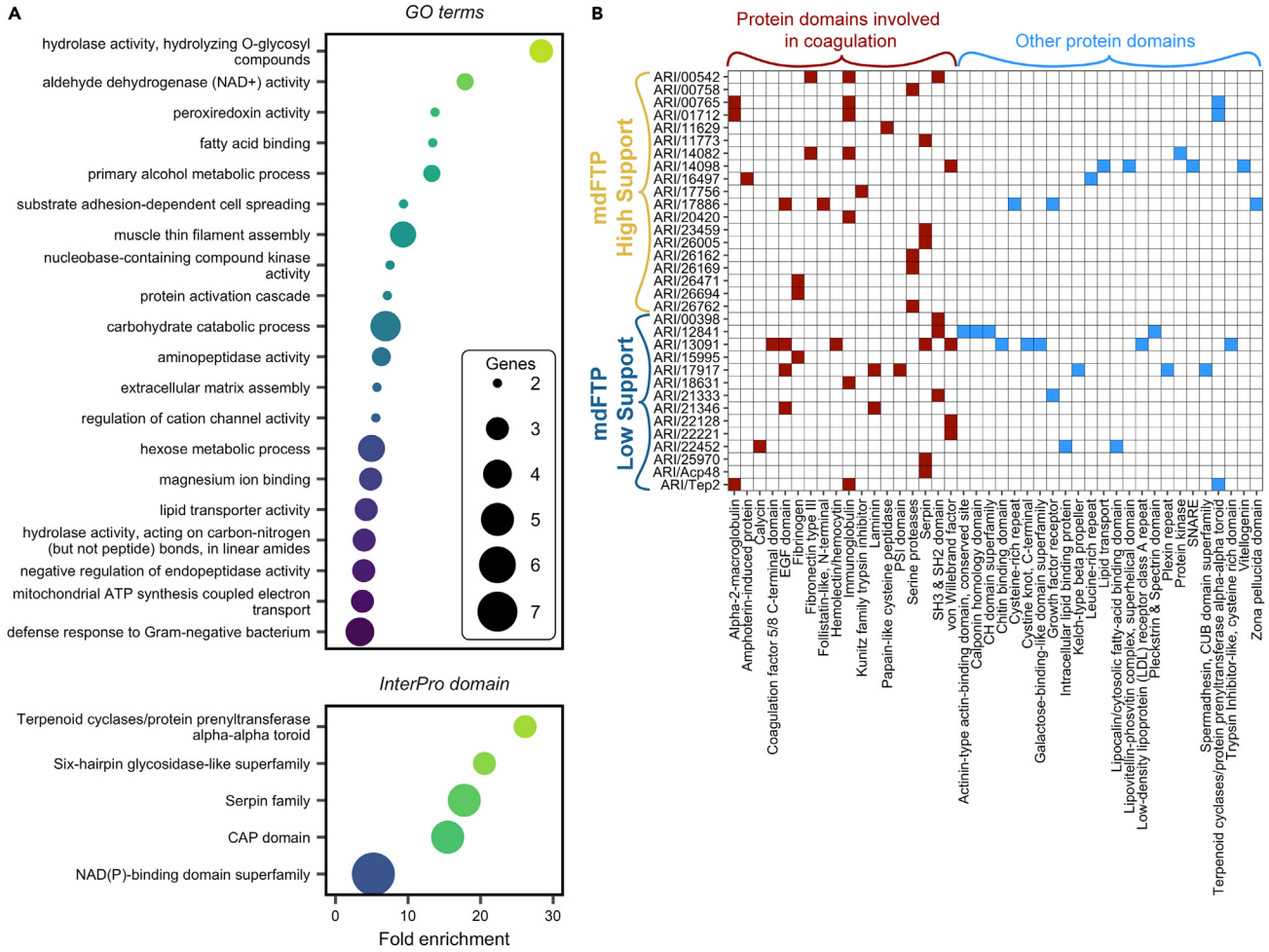



**Figure 3. mdFTPs have functional significance to reproduction**

(A) Enrichment of protein domains and GO terms for biological process and molecular function link mdFTPs to processes important for reproduction. This includes enrichment of CAP domains, which are linked to reproduction in diverse organisms, and two domains (serpin family and terpenoid cyclases/protein prenyltransferase alpha-alpha toroid) that have predicted involvement in coagulation. GO terms associated with energy production, oxidative stress response, and immunity were also enriched. Level of enrichment is indicated on the x axis and differentiated by color.

(B) List of mdFTPs with protein domains associated with coagulation. Additional domains not necessarily linked to coagulation are also noted. High-support mdFTPs refer to the 67 in yellow from Figure 2A, whereas low support mdFTPs represent the remaining 99.

for days to weeks prior to fertilization. Additionally, one enriched functional cluster was related to immunity. The regulation of immunity by mating is a key component of the postmating response in both vertebrates and invertebrates.[29,30] Our data suggest a potentially novel mechanism by which female immunity is regulated by the male ejaculate. More detailed data on significant functional clusters are found in Table S7.

## Many mdFTPs are associated with coagulation

In *D. mojavensis* and *D. arizonae*, a large opaque mass called the insemination reaction mass forms in the female reproductive tract after mating.[31,32] The composition of the reaction mass is unknown, but it is hypothesized to mediate sexual conflict over female remating rate.[31,33] In conspecific crosses, it persists for several hours, but in crosses between *D. mojavensis* and *D. arizonae* it lasts much longer, potentially even permanently sterilizing females.[31,33] Notably, at least two mdFTPs, ARI/13091 (CactusFlybase id: CFgn0008961; https://cactusflybase.arizona.edu/) and ARI/14098 (CFgn0013989) orthologous to *D. melanogaster hml* and *apolpp*, respectively, have demonstrated roles in hemolymph clotting in *Drosophila*.[34,35] This finding is especially interesting given that the reaction mass physically resembles a coagulatory response. In addition, two enriched protein domains, serpin family and terpenoid cyclases/protein prenyltransferase alpha-alpha toroid, have also been implicated in clotting and wound healing in vertebrates and some invertebrates.[36–38] Overall, at least 33 mdFTPs contain conserved protein domains associated with clotting and wound healing (Figure 3B), particularly in vertebrates.[39]

### Genes coding for mdFTPs are involved in formation and degradation of the reaction mass

Based on the large number of mdFTPs associated with coagulation, we hypothesized that the reaction mass is caused by a coagulatory response induced by SFPs and/or mdFTPs. To test this hypothesis, we used *D. arizonae* knockout lines that were established for two mdFTPs: *ARI/26694* (CFgn0007404) and *ARI/11629* (CFgn0010028). *ARI/26694* is a highly supported mdFTP that contains a fibrinogen domain and is transferred to females as both mRNA and protein. In vertebrates, fibrinogens are a critical component of the blood clotting cascade,[40] though invertebrate fibrinogens are not predicted to be involved in coagulation.[41] *ARI/11629* is a papain-like cysteine-peptidase transferred as RNA and protein, also from the list of highly supported mdFTPs. Previous research has demonstrated that this gene is rapidly evolving, as predicted for genes involved in sexual selection or sexual conflict.[42] Moreover, papain-like cysteine peptidases have been associated with coagulation in other organisms.[43] We compared the size of the reaction mass in *D. arizonae* females mated to *D. arizonae* KO or wild-type (WT) males immediately after mating and at 6 h postmating. For *ARI/26694*, females mated to KO males initially had a smaller reaction mass, and the mass also degraded more slowly compared to females mated to WT males (Figure 4A; Figure S2A). Conversely, for *ARI/11629*, females mated to KO males had a larger reaction mass regardless of the time postmating (Figure 4B; Figure S2B). Although we currently cannot differentiate the contribution of protein transferred by the male from protein produced from male RNA by the female, these results suggest that genes coding for mdFTPs are involved in a key aspect of the postmating response with predicted implications for sexual selection and sexual conflict.

### Gene coding for mdFTP affects fertilization efficiency

To evaluate whether mdFTPs have broader effects on postmating responses beyond the reaction mass, we also tested whether *ARI/26694* and *ARI/11629* influence female oviposition behavior or fertilization success. We compared the number of eggs laid over seven days postmating and fertilization success for WT *D. arizonae* females mated to *D. arizonae* WT or KO males. Knockout of *ARI/11629* did not affect oviposition (Figure S3). In contrast, analysis of the *ARI/26694* experiment revealed significant genotype × day interaction. However, *post-hoc* comparisons indicated that the only difference in oviposition occurred at four days postmating, with females mated to WT males laying fewer eggs than those mated to KO males. Since the magnitude of the difference is small and females do not lay many eggs at this time, we conclude that there is no overall biologically significant difference in oviposition between females mated to WT or KO males (Figure S3). However, given the trend for females mated to WT males to lay more eggs in the first few days after mating, additional data may reveal a biologically significant effect.

Although *ARI/26694* KO did not influence hatching rate (Figure S4), fertilization success for females mated to KO *ARI/11629* males was markedly reduced at days one and three postmating (Figure 4C). This reduction in successful fertilization would have strong impacts on male and female fitness. Since this protein is also transferred in the seminal fluid, additional studies are necessary to tease apart the effects of protein supplied by the male from protein translated by females from male RNA. Nevertheless, these results suggest genes coding for mdFTPs have broad-ranging effects on the postmating response that extend for days after mating.

## DISCUSSION

Our study provides clear evidence that seminal fluid RNA is translated into protein by females, including proteins with demonstrated functional roles in reproduction. These results advance our understanding of the molecular mechanisms of reproduction by illuminating a previously unrecognized mode of postmating interaction between males and females. This has important implications for elucidating the intricate nature and timing of cooperative reproductive molecular interactions between the sexes and uncovers a novel arena in which sexual selection and sexual conflict might occur. Seminal fluid RNA has been found in diverse species, which suggests that RNA transfer has substantial benefits.

One advantage of transferring RNA to females instead of, or in addition to, protein is that it provides a mechanism to temporally extend the production of key proteins. This is consistent with the fact that most male proteins are degraded within a few hours in the highly proteolytic environment of the female reproductive tract.[44–46] Transcriptome data from this study, and our previous study,[9] indicate that *D. arizonae* male RNA persists within the female reproductive tract for at least 6 h postmating (longer time points have not been tested). Temporal control over protein production could be beneficial in several contexts. For example, sperm must migrate through the female reproductive tract, enter sperm storage organs, and remain viable for days to weeks, while also competing with sperm from other males for access to female gametes. The continued production of proteins associated with these processes could be advantageous. Consistent with this, our results indicate that *ARI/11629* has effects on fertilization that extend for several days postmating. Males may thus benefit from transferring RNA transcripts, which results in continued production of this mdFTP beyond what the male could otherwise supply as protein. Similarly, recent evidence in *D. melanogaster* demonstrates that a "molecular handoff" between the sexes occurs within the female reproductive tract, whereby females gradually take over the production of key sperm-associated proteins initially supplied by males.[46] These proteins are enriched for processes such as energy metabolism, which is assumed to aid in sperm survival.[46] This type of cooperative maintenance of sperm viability between the sexes is predicted to be widespread across animals,[46] but the molecular mechanisms of how the molecular handoff occurs have not been elucidated. Orthologs of at least 12 mdFTPs overlap with this set of female-derived sperm-associated proteins in *D. melanogaster*.[46] Moreover, mdFTPs are highly enriched for processes involved in energy metabolism and oxidative stress response. This suggests that molecular continuity between the sexes may be facilitated, at least in part, by the intersexual RNA transfer described here.

*Drosophila arizonae* males may also benefit from temporal control over the production of proteins associated with the reaction mass in females. The reaction mass is hypothesized to serve as a mating plug that delays female remating,[31,33] though other functions are possible.[47,48] Males may benefit by using mdFTPs to prolong the persistence of the reaction mass, thereby extending the time until female remating.

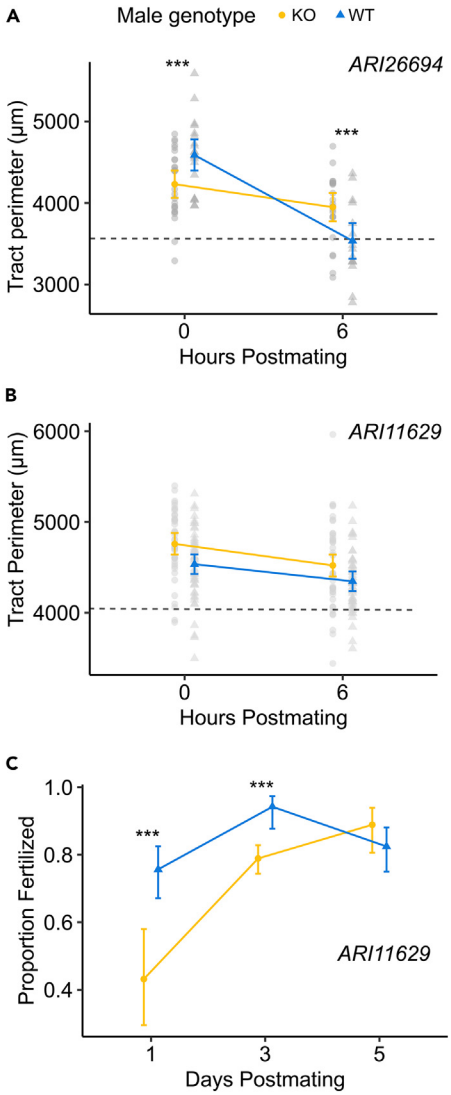



**Figure 4. Gene KO experiments demonstrate mdFTPs have functional effects on diverse reproductive processes**

(A) *D. arizonae* females mated to *D. arizonae* males with the *ARI26694* KO mutation had a smaller reaction mass initially, as measured by the average perimeter. However, it degraded more slowly compared to the reaction mass in females mated to WT males (two-way ANOVA: genotype × time interaction, F = 16.5, $p$ = 0.0001). *Post hoc* comparisons were performed using Tukey's method (**$p \leq 0.01$). KO-0 h: $n$ = 26; WT-0 h: $n$ = 35; KO-6 h: $n$ = 24; WT-6 h: $n$ = 15. Error bars represent 95% confidence intervals of estimated marginal means.

(B) *D. arizonae* females mated to *D. arizonae* males with the *ARI/11629* KO mutation had a larger reaction mass than females mated to WT males (two-way ANOVA: genotype, F = 12.0, $p$ = 0.0007). KO-0 h: $n$ = 42; WT-0 h: $n$ = 41; KO-6 h: $n$ = 41; WT-6 h: $n$ = 50. The dashed line indicates the mean perimeter of unmated female lower reproductive tract. Error bars represent 95% confidence intervals of estimated marginal means.

(C) *D. arizonae* females mated to *D. arizonae* males with a *ARI11629* KO mutation laid more unfertilized eggs on days one and three postmating compared to females mated to WT males (GLM: genotype × day interaction, $\chi^2$ = 17.2, $p$ = 0.0002). *Post hoc* comparisons were performed using Tukey's method (***$p > 0.001$). Error bars represent 95% confidence intervals of estimated marginal means. WT-D1: $n$ = 119; KO-D1: $n$ = 44; WT-D3: $n$ = 104; KO-D3 $n$ = 360; WT-D5 $n$ = 131; KO-D5 $n$ = 90.

Furthermore, while the reaction mass may prevent rapid remating, it also delays the onset of oviposition.[33,49] Thus, males may use mdFTPs to influence the timing of reaction mass degradation. A similar mechanism is observed in primates, where males transfer proteins to females that are involved in both the formation and degradation of a copulatory plug.[50] These hypotheses are consistent with our knockout experiments showing both pro- and anticoagulatory activity of genes coding for two mdFTPs associated with the reaction mass (*ARI/26694* and *ARI/11629*). Moreover, the overall set of mdFTPs with predicted roles in clotting include many proteins with both pro- and anticoagulatory activity (Figure 3B).

Formation of a reaction mass is relatively rare in *Drosophila*,[31] and the molecular bases of the formation and degradation of the mass have not been fully elucidated. We found relatively low overlap (~30%) between mdFTPs and the set of proteins associated with the *D. melanogaster*

mating plug,[51] which is structurally distinct from the reaction mass. Moreover, although mdFTPs include two proteins with known roles in *D. melanogaster* hemolymph coagulation (*hml* and *apolpp*),[34,35] other proteins known to play crucial roles in the coagulation cascade or mating plug formation were not observed among mdFTPs or SFPs.[34,35] These findings suggest that the reaction mass may be formed through distinct molecular pathways relative to other non-reproductive coagulatory responses in *Drosophila*. The two genes we experimentally demonstrated to be associated with the reaction mass include a fibrinogen protein (*ARI/26694*) and a cysteine peptidase (*ARI11629*), neither of which has an ortholog in *D. melanogaster*. While vertebrate fibrinogens play a critical role in blood clotting, previous studies have suggested that invertebrate fibrinogens do not function in coagulation.[41] Our results thus suggest a novel role of insect fibrinogens in coagulation. Much like vertebrate fibrinogens, which facilitate the assembly and degradation of fibrin clots,[52] *ARI/26694* influenced the initial size of the reaction mass, but also the rate of its degradation. Cysteine peptidases have known anticoagulant properties in other organisms,[43] which is consistent with our finding that females mated to *ARI/11629* KO males have a larger reaction mass than controls. Interestingly, mdFTPs include representatives of several protein classes associated with blood clotting in vertebrates[39] but that have not been linked to clotting in invertebrates[34,35] (Figure 3B). Some of these protein classes have broad functional roles, including other functions related to reproduction. Nevertheless, the confluence of all of them, along with the presence of the reaction mass, is striking. Overall, these findings indicate that the formation and degradation of the reaction mass is a highly orchestrated process that might be facilitated by the transfer of male RNA to females during copulation. If future studies confirm a direct role for mdFTPs in this process, this could have broad implications for understanding the formation and degradation of mating plugs, which have evolved independently in many taxa.

Aside from extending the timing of protein production, another benefit of transferring RNA to females is that transcripts could be targeted to specific cells in the female reproductive tract where they are translated. Localization of protein within female cells could provide functional capabilities that would not be possible for proteins transferred to the lumen of the female reproductive tract. Cell targeting could be facilitated if RNA is carried in EVs, which are capable of targeting cargo to specific cell types. We do not know how *D. arizonae* male seminal fluid RNA is packaged, but previous studies in *D. melanogaster* have shown that males transfer exosomes to females during copulation.[53] Although the contents of these exosomes have not been identified, they fuse with female epithelial cells and sperm and have demonstrated effects on female postmating physiology and behavior.[53–55] These findings, coupled with the fact that seminal fluid RNA in other organisms is packaged in EVs,[5,15] suggests this possibility warrants further study.

Sexual reproduction is a complex process that involves both cooperation and conflict between the sexes at the molecular level. Here, we illuminate a previously unrecognized mode of molecular interaction and interdependence between males and females. We show that males functionally alter the female postmating proteome by transferring RNA during copulation that is subsequently translated by females. Moreover, we demonstrate that genes coding for mdFTPs have diverse influences on postmating outcomes, including effects on fertilization success and formation/degradation of the reaction mass, a trait with predicted involvement in sexual conflict. Given the powerful coding and regulatory properties of RNA, coupled with the fact that RNA is likely a common feature of male ejaculates, this discovery has important implications for understanding the molecular mechanisms of reproduction across the Tree of Life.

### Limitations of the study

mdFTPs can only be identified by diagnostic peptides that differ in sequence between *D. mojavensis* and *D. arizonae*. Given that most peptides have the same sequence, our results likely underestimate the true number of mdFTPs.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Metabolic labeling of Drosophila
  - Protein isolation and LC-MS/MS
  - Identification of mdFTPs
  - HA peptide filtering and quality assessment
  - RNA sequencing of mated female reproductive tracts
  - Identification of D. arizonae male seminal fluid proteins
  - GO-term and protein domain enrichment analyses
  - Knockout experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RTI analysis
  - GO-term and protein domain enrichment analysis

- Statistical analysis of insemination reaction mass data
- Statistical analysis of fecundity data
- Statistical analysis of hatching/fertilization efficiency data

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110442.

## ACKNOWLEDGMENTS

We thank Chris Ebmeier from the Central Analytical Mass Spectrometry Facility and W.M. Keck Foundation Proteomics Resource at University of Colorado Boulder for assistance with proteomic analyses. We also thank Todd Schlenke for constructive comments on the manuscript. This work was supported by National Institutes of Health awards, 1R21HD097545-01 and 1R01HD112461-01, to L.M.M. and J.M.B.

## AUTHOR CONTRIBUTIONS

Conceptualization: L.M.M. and J.M.B.; methodology: L.M.M., J.M.B., H.K.P., C.W.A., J.R.M., and C.C.G.; investigation: L.M.M., J.M.B., C.W.A., F.D., H.K.P., J.R.M., C.C.G., J.C., and S.D.; visualization: L.M.M. and J.M.B.; funding acquisition: L.M.M. and J.M.B.; project administration: L.M.M. and J.M.B.; supervision: L.M.M., J.M.B., H.K.P., and C.W.A.; writing—original draft: L.M.M. and J.M.B.; writing—review & editing: L.M.M., J.M.B., C.W.A., F.D., H.K.P., J.R.M., C.C.G., J.C., and S.D.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 8, 2024
Revised: May 31, 2024
Accepted: July 1, 2024
Published: July 2, 2024

## REFERENCES

1. Pitnick, S., Wolfner, M.F., and Dorus, S. (2020). Post-ejaculatory modifications to sperm (PEMS). Biol. Rev. *95*, 365–392. https://doi.org/10.1111/BRV.12569.
2. Pitnick, S., Wolfner, M.F., and Suarez, S.S. (2009). Ejaculate-female and sperm-female interactions. In Sperm Biology, T. Birkhead, D. Hosken, and S. Pitnick, eds. (Elsevier Ltd), pp. 247–304. https://doi.org/10.1016/B978-0-12-372568-4.00007-0.
3. Poiani, A. (2006). Complexity of seminal fluid: A review. Behav. Ecol. Sociobiol. *60*, 289–310. https://doi.org/10.1007/s00265-006-0178-0.
4. Schjenken, J.E., and Robertson, S.A. (2020). The female response to seminal fluid. Physiol. Rev. *100*, 1077–1117. https://doi.org/10.1152/PHYSREV.00013.2018.
5. Jodar, M. (2019). Sperm and seminal plasma RNAs: What roles do they play beyond fertilization? Reproduction *158*, R113–R123. https://doi.org/10.1530/REP-18-0639.
6. Scolari, F., Khamis, F.M., and Pérez-Staples, D. (2021). Beyond Sperm and Male Accessory Gland Proteins: Exploring Insect Reproductive Metabolomes. Front. Physiol. *12*, 729440. https://doi.org/10.3389/FPHYS.2021.729440/BIBTEX.
7. Fischer, B.E., Wasbrough, E., Meadows, L.A., Randlet, O., Dorus, S., Karr, T.L., and Russell, S. (2012). Conserved properties of *Drosophila* and human spermatozoal mRNA repertoires. Proc. Biol. Sci. *279*, 2636–2644. https://doi.org/10.1098/rspb.2012.0153.
8. Lalancette, C., Miller, D., Li, Y., and Krawetz, S.A. (2008). Paternal contributions: New functional insights for spermatozoal RNA. J. Cell. Biochem. *104*, 1570–1579. https://doi.org/10.1002/jcb.21756.
9. Bono, J.M., Matzkin, L.M., Kelleher, E.S., and Markow, T.A. (2011). Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated *Drosophila mojavensis* females. Proc. Natl. Acad. Sci. USA *108*, 7878–7883. https://doi.org/10.1073/pnas.1100388108.
10. Degner, E.C., Ahmed-Braimah, Y.H., Borziak, K., Wolfner, M.F., Harrington, L.C., and Dorus, S. (2019). Proteins, transcripts, and genetic architecture of seminal fluid and sperm in the mosquito *Aedes aegypti*. Mol. Cell. Proteomics *18*, S6–S22. https://doi.org/10.1074/mcp.RA118.001067E.
11. Ahmed-Braimah, Y.H., Wolfner, M.F., and Clark, A.G. (2021). Differences in Postmating Transcriptional Responses between Conspecific and Heterospecific Matings in *Drosophila*. Mol. Biol. Evol. *38*, 986–999. https://doi.org/10.1093/molbev/msaa264.
12. Alfonso-Parra, C., Ahmed-Braimah, Y.H., Degner, E.C., Avila, F.W., Villarreal, S.M., Pleiss, J.A., Wolfner, M.F., and Harrington, L.C. (2016). Mating-Induced Transcriptome Changes in the Reproductive Tract of Female *Aedes aegypti*. PLoS Neglected Trop. Dis. *10*, e0004451. https://doi.org/10.1371/journal.pntd.0004451.
13. Cridland, J.M., and Begun, D.J. (2023). Male-derived transcripts isolated from the mated female reproductive tract in *Drosophila melanogaster*. G3 (Bethesda). *13*, jkad202. https://doi.org/10.1093/G3JOURNAL/JKAD202.
14. Santiago, J., Silva, J.v., Howl, J., Santos, M.A.S., and Fardilha, M. (2021). All you need to know about sperm RNAs. Hum. Reprod. Update *28*, 67–91. https://doi.org/10.1093/HUMUPD/DMAB034.
15. Tamessar, C.T., Trigg, N.A., Nixon, B., Skerrett-Byrne, D.A., Sharkey, D.J., Robertson, S.A., Bromfield, E.G., and Schjenken, J.E. (2021). Roles of male reproductive tract extracellular vesicles in reproduction. Am. J. Reprod. Immunol. *85*, e13338. https://doi.org/10.1111/aji.13338.
16. Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77–94. https://doi.org/10.1016/J.CELL.2014.03.008.
17. Sury, M.D., Chen, J.X., and Selbach, M. (2010). The SILAC fly allows for accurate protein quantification in vivo. Mol. Cell. Proteomics *9*, 2173–2183. https://doi.org/10.1074/mcp.M110.000323.
18. Schober, F.A., Atanassov, I., Freyer, C., and Wredenberg, A. (2021). Quantitative Proteomics in *Drosophila* with Holidic Stable-Isotope Labeling of Amino Acids in Fruit Flies (SILAF). In Methods in Molecular Biology (Humana Press Inc), pp. 75–87. https://doi.org/10.1007/978-1-0716-0834-0_7.
19. Rosenberger, F.A., Atanassov, I., Moore, D., Calvo-Garrido, J., Moedas, M.F., Wedell, A., Freyer, C., and Wredenberg, A. (2021). Stable Isotope Labeling of Amino Acids in Flies (SILAF) Reveals Differential Phosphorylation of Mitochondrial Proteins Upon Loss of OXPHOS Subunits. Mol. Cell. Proteomics *20*, 100065. https://doi.org/10.1016/j.mcpro.2021.100065.
20. Chang, Y.C., Tang, H.W., Liang, S.Y., Pu, T.H., Meng, T.C., Khoo, K.H., and Chen, G.C. (2013). Evaluation of *Drosophila* metabolic labeling strategies for in vivo quantitative

proteomic analyses with applications to early pupa formation and amino acid starvation. J. Proteome Res. *12*, 2138–2150. https://doi.org/10.1021/pr301168x.
21. Xu, P., Tan, H., Duong, D.M., Yang, Y., Kupsco, J., Moberg, K.H., Li, H., Jin, P., and Peng, J. (2012). Stable isotope labeling with amino acids in *Drosophila* for quantifying proteins and modifications. J. Proteome Res. *11*, 4403–4412. https://doi.org/10.1021/pr300613c.
22. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319. https://doi.org/10.1038/nprot.2016.136.
23. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.
24. Liu, H., Sadygov, R.G., and Yates, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. *76*, 4193–4201. https://doi.org/10.1021/AC0498563/SUPPL_FILE/AC0498563SI20040123_114906.PDF.
25. Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J.L., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., et al. (2010). Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography—Tandem Mass Spectrometry. J. Proteome Res. *9*, 761–776. https://doi.org/10.1021/PR9006365.
26. Gonzalez, S.N., Sulzyk, V., Weigel Muñoz, M., and Cuasnicu, P.S. (2021). Cysteine-Rich Secretory Proteins (CRISP) are Key Players in Mammalian Fertilization and Fertility. Front. Cell Dev. Biol. *9*, 3438. https://doi.org/10.3389/FCELL.2021.800351/BIBTEX.
27. Laflamme, B.A., and Wolfner, M.F. (2013). Identification and Function of Proteolysis Regulators in Seminal Fluid. Mol. Reprod. Dev. *80*, 80–101. https://doi.org/10.1002/MRD.22130.
28. 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. https://doi.org/10.1146/annurev-ento-120709-144823.
29. Lawniczak, M.K.N., Barnes, A.I., Linklater, J.R., Boone, J.M., Wigby, S., and Chapman, T. (2007). Mating and immunity in invertebrates. Trends Ecol. Evolut. *22*, 48–55. https://doi.org/10.1016/j.tree.2006.09.012.
30. Morrow, E.H., and Innocenti, P. (2012). Female postmating immune responses, immune system evolution and immunogenic males. Biol. Rev. *87*, 631–638. https://doi.org/10.1111/j.1469-185X.2011.00214.x.
31. Alonso-Pimentel, H., Tolbert, L.P., and Heed, W.B. (1994). Ultrastructural examination of the insemination reaction in *Drosophila*. Cell Tissue Res. *275*, 467–479. https://doi.org/10.1007/BF00318816.
32. Markow, T.A., and Ankney, P.F. (1988). Insemination Reaction in *Drosophila*: Found in Species Whose Males Contribute Material to Oocytes Before Fertilization. Evolution (N Y) *42*, 1097–1101. https://doi.org/10.2307/2408926.
33. Knowles, L.L., and Markow, T.A. (2001). Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. Proc. Natl. Acad. Sci. USA *98*, 8692–8696. https://doi.org/10.1073/pnas.151123998.
34. Karlsson, C., Korayem, A.M., Scherfer, C., Loseva, O., Dushay, M.S., and Theopold, U. (2004). Proteomic Analysis of the *Drosophila* Larval Hemolymph Clot. J. Biol. Chem. *279*, 52033–52041. https://doi.org/10.1074/JBC.M408220200.
35. Dziedziech, A., Shivankar, S., and Theopold, U. (2020). *Drosophila melanogaster* Responses against Entomopathogenic Nematodes: Focus on Hemolymph Clots. Insects *11*, 62. https://doi.org/10.3390/INSECTS11010062.
36. Lagrange, J., Lecompte, T., Knopp, T., Lacolley, P., and Regnault, V. (2022). Alpha-2-macroglobulin in hemostasis and thrombosis: An underestimated old double-edged sword. J. Thromb. Haemost. *20*, 806–815. https://doi.org/10.1111/JTH.15647.
37. Pike, R.N., Buckle, A.M., le Bonniec, B.F., and Church, F.C. (2005). Control of the coagulation system by serpins. Getting by with a little help from glycosaminoglycans. FEBS J. *272*, 4842–4851. https://doi.org/10.1111/J.1742-4658.2005.04880.X.
38. Theopold, U., Li, D., Fabbri, M., Scherfer, C., and Schmidt, O. (2002). The coagulation of insect hemolymph. Cell. Mol. Life Sci. *59*, 363–372. https://doi.org/10.1007/S00018-002-8428-4/METRICS.
39. Furie, B., and Furie, B.C. (1988). The molecular basis of blood coagulation. Cell *53*, 505–518.
40. Kattula, S., Byrnes, J.R., and Wolberg, A.S. (2017). Fibrinogen and Fibrin in Hemostasis and Thrombosis. Arterioscler. Thromb. Vasc. Biol. *37*, e13–e21. https://doi.org/10.1161/ATVBAHA.117.308564.
41. Hanington, P.C., and Zhang, S.-M. (2011). The Primary Role of Fibrinogen-Related Proteins in Invertebrates Is Defense, Not Coagulation. J. Innate Immun. *3*, 17–27. https://doi.org/10.1159/000321882.
42. Bono, J.M., Matzkin, L.M., Hoang, K., and Brandsmeier, L. (2015). Molecular evolution of candidate genes involved in post-mating-prezygotic reproductive isolation. J. Evol. Biol. *28*, 403–414. https://doi.org/10.1111/jeb.12574.
43. Ozhelvaci, F., and Steczkiewicz, K. (2023). Identification and classification of papain-like cysteine proteinases. J. Biol. Chem. *299*, 104801. https://doi.org/10.1016/j.jbc.2023.104801.
44. 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. https://doi.org/10.1016/J.IBMB.2005.05.001.
45. Wolfner, M.F. (2011). Precious Essences: Female Secretions Promote Sperm Storage in *Drosophila*. PLoS Biol. *9*, e1001191. https://doi.org/10.1371/JOURNAL.PBIO.1001191.
46. McCullough, E.L., Whittington, E., Singh, A., Pitnick, S., Wolfner, M.F., and Dorus, S. (2022). The life history of *Drosophila* sperm involves molecular continuity between male and female reproductive tracts. Proc. Natl. Acad. Sci. USA *119*, e2119899119. https://doi.org/10.1073/PNAS.2119899119.
47. Schneider, M.R., Mangels, R., and Dean, M.D. (2016). The molecular basis and reproductive function(s) of copulatory plugs. Mol. Reprod. Dev. *83*, 755–767. https://doi.org/10.1002/MRD.22689.
48. Avila, F.W., Wong, A., Sitnik, J.L., and Wolfner, M.F. (2015). Don't pull the plug! the *Drosophila* mating plug preserves fertility. Fly (Austin) *9*, 62–67. https://doi.org/10.1080/19336934.2015.1120931.
49. Kelleher, E.S., and Markow, T.A. (2007). Reproductive Tract Interactions Contribute to Isolation in *Drosophila*. Fly (Austin) *1*, 33–37. https://doi.org/10.4161/fly.3840.
50. Dorus, S., Evans, P.D., Wyckoff, G.J., Sun, S.C., and Lahn, B.T. (2004). Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat. Genet. *36*, 1326–1329. https://doi.org/10.1038/ng1471.
51. McDonough-Goldstein, C.E., Pitnick, S., and Dorus, S. (2022). *Drosophila* female reproductive glands contribute to mating plug composition and the timing of sperm ejection. Proc. Biol. Sci. *289*, 20212213. https://doi.org/10.1098/RSPB.2021.2213/.
52. Chapin, J.C., and Hajjar, K.A. (2015). Fibrinolysis and the control of blood coagulation. Blood Rev. *29*, 17–24. https://doi.org/10.1016/J.BLRE.2014.09.003.
53. Corrigan, L., Redhai, S., Leiblich, A., Fan, S.J., Perera, S.M.W., Patel, R., Gandy, C., Mark Wainwright, S., Morris, J.F., Hamdy, F., et al. (2014). BMP-regulated exosomes from *Drosophila* male reproductive glands reprogram female behavior. J. Cell Biol. *206*, 671–688. https://doi.org/10.1083/jcb.201401072.
54. Hopkins, B.R., Sepil, I., Bonham, S., Miller, T., Charles, P.D., Fischer, R., Kessler, B.M., Wilson, C., and Wigby, S. (2019). BMP signaling inhibition in *Drosophila* secondary cells remodels the seminal proteome and self and rival ejaculate functions. Proc. Natl. Acad. Sci. USA *116*, 24719–24728. https://doi.org/10.1073/pnas.1914491116.
55. Delbare, S.Y.N., Jain, A.M., Clark, A.G., and Wolfner, M.F. (2023). Transcriptional programs are activated and microRNAs are repressed within minutes after mating in the *Drosophila melanogaster* female reproductive tract. BMC Genomics *24*, 356. https://doi.org/10.1186/S12864-023-09397-Z.
56. Benowitz, K.M., Allan, C.W., Jaworski, C.C., Sanderson, M.J., Chen, X., and Matzkin, L.M. (2022). Chromosome-length genome assemblies of cactophilic *Drosophila* illuminate Running Title: Genomes of cactophilic *Drosophila*. Preprint at BioRxiv. https://doi.org/10.1101/2022.10.16.512445.
57. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093. https://doi.org/10.1093/BIOINFORMATICS/BTP101.
58. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.
59. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
60. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324.
61. Broad Institute. Picard Toolkit. 2019. https://broadinstitute.github.io/picard/.

62. Choi, S., Kim, H., and Paek, E. (2017). ACTG: novel peptide mapping onto gene models. Bioinformatics *33*, 1218–1220. https://doi.org/10.1093/BIOINFORMATICS/BTW787.
63. Dainat, J. Another Gff Analysis Toolkit to Handle Annotations in Any GTF/GFF Format. (Version v0.6.0). Zenodo.
64. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033.
65. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
66. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. https://doi.org/10.1101/GR.129684.111.
67. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92. https://doi.org/10.4161/FLY.19695.
68. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.
69. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/BIOINFORMATICS/BTU031.
70. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205. https://doi.org/10.1093/NAR/GKZ401.
71. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. J Stat Softw *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.
72. Fox, J., and Weisberg, S. (2019). An R Companion to Applied Regression Third (Sage Publications).
73. Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. Biom. J. *50*, 346–363. https://doi.org/10.1002/BIMJ.200810425.
74. Henrik, S., Bolker, B., Westfall, J., and Aust, F. (2016). Afex: Analysis of Factorial Experiments. https://cran.r-project.org/web/packages/afex/index.html.
75. Lenth, R. (2019). Emmeans: estimated marginal means. Aka Least-Squares Means. https://cran.r-project.org/package=emmeans.
76. Brooks, M.E., Kristensen, K., Koen, J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., and Bolker, B. (2017). glmmTMB Balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. *9*, 378–400.
77. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/NAR/GKAB1038.
78. Coleman, J.M., Benowitz, K.M., Jost, A.G., and Matzkin, L.M. (2018). Behavioral evolution accompanying host shifts in cactophilic *Drosophila* larvae. Ecol. Evol. *8*, 6921–6931. https://doi.org/10.1002/ECE3.4209.
79. Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. *10*, 757. https://doi.org/10.15252/MSB.20145625.
80. Diaz, F., Allan, C.W., Chen, X., Coleman, J.M., Bono, J.M., and Matzkin, L.M. (2022). Divergent evolutionary trajectories shape the postmating transcriptional profiles of conspecifically and heterospecifically mated cactophilic *Drosophila* females. Commun. Biol. *5*, 842. https://doi.org/10.1038/s42003-022-03758-2.
81. Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly Efficient Targeted Mutagenesis of *Drosophila* with the CRISPR/Cas9 System. Cell Rep. *4*, 220–228. https://doi.org/10.1016/j.celrep.2013.06.020.
82. Markow, T.A., Beall, S., and Matzkin, L.M. (2009). Egg size, embryonic development time and ovoviviparity in *Drosophila* species. J. Evol. Biol. *22*, 430–434. https://doi.org/10.1111/J.1420-9101.2008.01649.X.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Biological samples** | | |
| *Drosophila arizonae* | National *Drosophila* Species Stock Center, Ithaca, NY | *NDSSC: 15081-1271.41* |
| *Drosophila mojavensis* | National *Drosophila* Species Stock Center, Ithaca, NY | *NDSSC: 15081-1353.01* |
| **Chemicals, peptides, and recombinant proteins** | | |
| Lys8: L-Lysine·2HCl ($^{13}C_6$, 99%;$^{15}N_2$, 99%) | Cambridge Isotopes Laboratory, Inc, Tewksbury, MA | CNLM-291-H-PK |
| Lys4: L-Lysine·2HCl (4,4,5,5-$D_4$, 96–98%) | Cambridge Isotopes Laboratory, Inc, Tewksbury, MA | DLM-2640-PK |
| Sera-Mag™ Carboxylate-Modified Magnetic Beads & SpeedBeads | Cytiva, Wilmington, DE | 24152105050350 |
| Trypsin/Lys-C Mix, Mass Spec Grade | Promega, Madison, WI | V5071 |
| CleanCap® Cas9 mRNA | TriLink Biotechnologies, San Diego, CA | L-7606 |
| Qiagen DNeasy Blood & Tissue Kit | Qiagen, Hilden, Germany | 69504 |
| KAPA LTP Library Preparation Kit | Roche, Basel, Switzerland | KK8421 |
| **Deposited data** | | |
| Proteomic raw data from this study | ProteomeXchange Consortium | PRIDE: PXD041195, PXD041260 |
| Protein database for both *D. mojavensis* and *D. arizonae* | From Benowitz et al.,[56] located at OSF | https://osf.io/vhyzc/ |
| Raw genomic reads | NCBI's SRA repository | NCBI: PRJNA949702 |
| **Experimental models: Organisms/strains** | | |
| *Saccharomyces cerevisiae* (MATalpha leu2D0 lys2D0 ura3D0 in background BY4729) | Horizon Discovery LTD | https://horizondiscovery.com/en/non-mammalian-research-tools/products/yeast-knockout#description |
| **Software and algorithms** | | |
| MaxQuant (ver. 2.0.3.0) | Tyanova et al.[22] | https://www.maxquant.org/ |
| MSFragger (ver. 3.6 within FragPipe ver. 19.0 | Kong et al.[23] | https://msfragger.nesvilab.org/ |
| ClueGO | Bindea et al.[57] | https://apps.cytoscape.org/apps/cluego |
| ImageJ | Schneider et al.[58] | https://imagej.net/ij/ |
| Trimmomatic | Bolger et al.[59] | http://www.usadellab.org/cms/?page=trimmomatic |
| bwa-mem | Li and Durbin[60] | https://bio-bwa.sourceforge.net/ |
| Picard v2.27.5 | Broad Institute[61] | https://broadinstitute.github.io/picard/ |
| ACTG | Choi et al.[62] | https://prix.hanyang.ac.kr/download/ACTG.jsp |
| AGAT | Dainat[63] | https://agat.readthedocs.io/en/latest/ |
| bedtools2 | Quinlan and Hall[64] | https://bedtools.readthedocs.io/en/latest/index.html |
| samtools | Li et al.[65] | http://www.htslib.org/ |
| VarScan2 | Kobolt et al.[66] | https://varscan.sourceforge.net/ |
| SnpEff | Cingolani et al.[67] | https://pcingola.github.io/SnpEff/ |
| HISAT2 | Kim et al.[68] | http://daehwankimlab.github.io/hisat2/ |
| InterProScan | Jones et al.[69] | https://www.ebi.ac.uk/interpro/search/sequence/ |
| WebGestalR | Liao et al.[70] | https://cran.r-project.org/web/packages/WebGestaltR/index.html |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| lme4 package for R | Bates et al.[71] | https://cran.r-project.org/web/packages/lme4/index.html |
| car package for R | Fox and Weisberg[72] | https://cran.r-project.org/web/packages/car/index.html |
| multcomp package for R | Hothorn et al.[73] | https://cran.r-project.org/web/packages/multcomp/index.html |
| afex package for R | Henrik et al.[74] | https://cran.r-project.org/web/packages/afex/index.html |
| emmeans package for R | Lenth[75] | https://cran.r-project.org/web/packages/emmeans/index.html |
| glmmTMB package for R | Brooks et al.[76] | https://cran.r-project.org/web/packages/glmmTMB/index.html |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to lead contact, Luciano Matzkin (lmatzkin@arizona.edu).

### Materials availability

*Drosophila arizonae* knockout lines are available from the lead contact upon request.

### Data and code availability

- **Data:** All mass spectrometry proteomics data have been deposited on the ProteomeXchange Consortium via the PRIDE[77] partner repository and are publicly available as of the date of publication with the dataset identifier PXD041195 and PXD041260. All MaxQuant and MSFragger analysis parameter and output files are available at OSF, https://osf.io/vhyzc, and are publicly available as of the date of publication. Raw RNA sequencing reads have been deposited on NCBI's SRA repository and are publicly available as of the date of publication under BioProject PRJNA949702.
- **Code:** All original code is available in this paper's Methods S1. DOIs are listed in the key resources table.
- **Additional information:** Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All experiments described used *D. mojavensis* females (National *Drosophila* Species Stock Center (NDSSC): 15081–1353.01) and *D. arizonae* males (NDSSC: 15081–1271.41). Genomes from these stocks are accessible at the cactusflybase site (https://cactusflybase.arizona.edu/). Knockout *D. arizonae* lines were also generated in the same *D. arizonae* stock. Fly stocks were maintained on standard banana food on a 12:12 light dark cycle at room temperature. Flies used in experiments were between 8 and 14 days old and had not mated prior to the experiment. Flies were randomly chosen for experiments.

## METHOD DETAILS

### Metabolic labeling of Drosophila

To distinguish the origin of proteins in the female reproductive tract, we metabolically labeled female flies, by modifying protocols developed previously for *D. melanogaster*.[17–21,46] Males were unlabeled. For labeling, we reared a lysine auxotrophic strain of *Saccharomyces cerevisiae* (MATalpha leu2Δ0 lys2Δ0 ura3Δ0 in background BY4729; Dharmacon, Inc. Lafayette, CO) in synthetic media containing yeast nitrogen base without amino acids, yeast synthetic drop-out medium with all amino acids except lysine, and isotopically labeled lysine (Lys8: L-Lysine-$^{13}C_6$,$^{15}N_2$; Cambridge Isotopes Laboratory, Inc, Tewksbury, MA). Cultures were grown in a shaking incubator for ~24 h at 30°C. Cells were pelleted by centrifugation, washed with sterilized water, lyophilized, and frozen at −20°C. We made food media for rearing flies by combining Lys8 labeled yeast, yeast nitrogen base without amino acids, pure molasses, agar, sterile water and methylparaben dissolved in ethanol as a preservative (Table S4). This mixture was autoclaved and dispensed into sterile glass vials in a biosafety cabinet. We placed *D. mojavensis* adults in population cages overnight on media containing baker's yeast and molasses to encourage oviposition. The following day, eggs were collected and sterilized/dechorionated by soaking in 2.5% sodium hypochlorite for 3 min. Eggs were rinsed with sterile water and transferred under aseptic conditions to vials containing Lys8 food (~100 eggs/vial) where they were reared until adult emergence. We collected unmated female flies within a few hours of eclosion and transferred them to vials containing sterilized food consisting of pure molasses, agar, water, and a flake of lyophilized Lys8 labeled yeast. Flies were transferred daily to new vials until they reached reproductive maturity. This protocol ensured that flies were only exposed to Lys8 labeled protein prior to mating experiments. Labeling efficiency was assessed by analyzing Lys8 labeled *D. mojavensis* male reproductive tract samples. Spectra were analyzed by MaxQuant[22] (ver. 2.0.3.0) and MSFragger[23] (ver. 3.6 within FragPipe ver. 19.0) with Lys8 as a variable modification. Over 99% of identified peptides carried the Lys8 label, indicating high labeling efficiency. Heavy-labeled unmated female *D. mojavensis* females were paired in vials containing sterilized molasses-agar food without yeast with unmated *D. arizonae* males that had been reared in standard banana-molasses food.[78] Copulations were

observed and mated females were isolated until we removed lower reproductive tracts at 6 h postmating. Female reproductive tracts were placed in 20 μL ice-cold 50 mM ammonium bicarbonate. Individual tubes were moved to −80°C once five tracts were collected. This experiment was repeated three times, with each biological replicate consisting of ∼45 lower reproductive tracts.

## Protein isolation and LC-MS/MS

Protein isolation and LC-MS/MS was performed at the **Central Analytical Mass Spectrometry Facility and W.M. Keck Foundation Proteomics Resource at University of Colorado Boulder.** Heterospecifically-mated *D. mojavensis* lower female reproductive tracts were denatured, reduced and alkylated with 5% (w/v) sodium dodecyl sulfate (SDS), 10 mM tris (2-carboxyethyl) phosphine hydrochloride (TCEP-HCl), 40 mM 2-chloroacetamide, 50 mM Tris pH 8.5 and boiled at 95°C for 10 min. Samples were prepared for mass spectrometry analyses using the SP3 method.[79] Carboxylate-functionalized speedbeads (Cytiva) were added to protein lysates. Acetonitrile was added to 80% (v/v) to precipitate protein and bind it to the beads. The protein-bound beads were washed twice with 80% (v/v) ethanol and twice with 100% acetonitrile. LysC/Trypsin mix (Promega) was added for approximately 1:50 protease to protein ratio in 50 mM Tris pH 8.5 and incubated rotating at 37°C overnight. Tryptic digests were cleaned using an Oasis HLB 1cc (10 mg) cartridge (Waters) according to the manufacturer. Samples were dried using a vacuum rotatory evaporator. To reduce sample complexity, samples were fractionated using a Waters M-class UPLC equipped with a photodiode-array detector. Samples were suspended in 50uL 0.16% (v/v) aqueous ammonia in water, then injected onto a 500um id x 150 mm long custom fabricated rpC18 column (UChrom C18 1.8 um 120A) and separated with a gradient 2%–40% acetonitrile with 0.16% (v/v) aqueous ammonia. Concatenated fractions (8–12 depending on the replicate) were collected and dried using a vacuum rotatory evaporator. Fractions were suspended in 0.1% TFA, 3% acetonitrile in water for LC/MS/MS analyses. Fractionated peptides were directly injected onto a Waters M-class column (1.7um, 120A, rpC18, 75um x 250 mm) and gradient eluted from 2% to 20% acetonitrile with 0.1%(v/v) formic acid over 100 min then 20%–32% acetonitrile over 20 min at 0.3uL/minute using a Thermo Ultimate 3000 UPLC (Thermo Scientific). Peptides were detected with a Thermo Q-Exactive HF-X mass spectrometer (Thermo Scientific) scanning MS1 at 120,000 resolution from 380 to 1580 m/z with a 45 ms fill time and 3E6 AGC target. The top 12 most intense peaks were isolated with 1.4 m/z window with a 100 ms fill time and 1E5 AGC target and 27% HCD collision energy for MS2 spectra scanned at 15,000 resolution. Dynamic exclusion was enabled for 25 s.

## Identification of mdFTPs

Mass spectra were analyzed using two different software programs: MaxQuant[22] (ver. 2.0.3.0) and MSFragger[23] (ver. 3.6 within FragPipe ver. 19.0). Database searches assumed proteins could be unlabeled or heavy labeled (Lys8) and included common variable modifications (Oxidation (M); Acetyl (Protein N-term)). PSM and protein FDR were set to 0.01. The 'requantify' and 'matching between runs' features were not utilized. The protein database for both *D. mojavensis* and *D. arizonae* was based on our recent assembly and annotation including all genome transcripts (19,778 for *D. mojavensis* and 19,747 for *D. arizonae*)[56] available at OSF (https://osf.io/vhyzc). All gene IDs used in this study follow annotation nomenclature set forth in the cactusflybase site.[56] PSMs were sorted to identify heavy *D. arizonae* (HA) PSMs, which are diagnostic for mdFTPs. We required mdFTPs to be supported by a minimum of one diagnostic peptide and at least one additional heavy peptide that could be diagnostic or non-diagnostic (HAM peptides; Table 1). Thus, so called 'one-hit-wonders' were filtered out of our mdFTP list. All mass spectrometry proteomics data have been deposited on the ProteomeXchange Consortium via the PRIDE[77] partner repository with the dataset identifier PXD041195 and PXD041260. All MaxQuant and MSFragger analysis parameter and output files are available at OSF.

## HA peptide filtering and quality assessment

We performed rigorous filtering to remove potential false positive HA peptide identifications due to polymorphism or leucine/isoleucine substitutions. Moreover, to further investigate the validity of HA peptide identifications we compared features of HA peptides with LM peptides, which could only be identified by error or because of incomplete label incorporation (Table S1).

The consensus genome sequences of *D. mojavensis* and *D. arizonae* do not include potential polymorphisms that could confound identification of diagnostic peptides (e.g., one *D. mojavensis* segregating allele matches *D. arizonae*). Therefore, we collected the head and thorax of mated females from one of our experimental replicates and bulk sequenced the sample so that any HA peptides identified based on polymorphic sites could be filtered out. DNA from the head and thorax of 45 mated females was extracted using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The library was prepared using the KAPA LTP Library Preparation Kit (Roche, Basel, Switzerland) kit and sequenced on an Illumina HiSeq 4000 at Novogene (Beijing, China) to 114X coverage. Raw reads have been deposited on NCBI's SRA repository under BioProject PRJNA949702. Reads from *D. mojavensis* were trimmed using Trimmomatic[59] and mapped to the Sonora *D. arizonae* genome r0.93[56] using bwa-mem.[60] PCR duplicates were then removed using Picard v2.27.5.[61]

To determine if polymorphism existed within diagnostic peptides, we first mapped HA peptides to the *D. arizonae* genome r0.93[56] using ACTG.[62] To utilize ACTG, we generated a GFF file from the GTF using AGAT.[63] From the output, a BED file was generated of all HA peptides, and we used bedtools2[64] to extract the peptide regions. A mpileup file was then generated using samtools,[65] and variant calling was performed with VarScan2.[66] SnpEff[67] was used to identify synonymous and nonsynonymous changes in the candidate peptide regions. We retained peptides if they met the following two conditions: (1) at least one nonsynonymous change that was not a Leu-Ile or Ile-Leu (not distinguishable by mass spectrometry), (2) bulk genome sequencing of *D. mojavensis* revealed that the sequence was not polymorphic in *D. mojavensis* (≥95% of reads match the *D. mojavensis* reference genome).

To further investigate the validity of HA peptide identifications we compared the number of HA PSMs to LM (light-*mojavensis*) PSMs identified in each replicate. Since female proteomes were heavy labeled, any LM PSMs are assumed to be erroneous identifications or from

incomplete labeling (we excluded individual PSMs including both heavy and light quantifications since they are especially likely to include light peptides resulting from incomplete labeling). HA PSMs could be errors or could represent mdFTPs. Since LM and HA PSMs should be equally likely to be erroneously identified, we expect the number of each type to be approximately equal if most HA PSMs are errors.

### RNA sequencing of mated female reproductive tracts

We used a combination of DNA and RNA sequencing to establish a set of fixed nucleotide differences between *D. mojavensis* and *D. arizonae* since consensus genome sequences do not provide information on polymorphism within lines. For *D. arizonae*, we identified polymorphisms using genome sequencing reads and pooled RNA-seq reads from multiple life stages originally collected for the *D. arizonae* genome assembly.[56] For *D. mojavensis*, we used reads from the genome assembly and from head and thorax samples of mated flies described above. Reads from both species were mapped to the *D. arizonae* r0.93 assembly[56] using HISAT2[68] for RNA and bwa[60] for DNA. Duplicates were then removed using Picard v2.27.5,[61] bam files were merged, exons regions exported, mpileup generated using samtools[65] and variants called using VarScan2.[66] Variant called files for *D. arizonae* DNA and RNA sequencing were generated containing positions which had a *D. mojavensis* allele frequency >0.95 and a significant Fisher's Exact test ($p < 0.05$). We used these files to generate a list of 549,579 positions that represent fixed differences between the species.

To identify male derived RNA in the lower reproductive tract of females, we used RNA-seq data from,[80] in which heterospecifically-mated female reproductive tracts were collected under the same conditions as reported here. Reads from heterospecifically-mated *D. mojavensis* females 45 min and 6 h postmating (NCBI BioProject PRJNA777940) were trimmed with Trimmomatic,[59] duplicates removed with Picard v2.27.5,[61] and pooled and mapped to the *D. arizonae* r0.93 genome using HISAT2.[68] We used the list of fixed sites between the species to query the RNA-seq data to detect male reads (*D. arizonae*). Using a custom perl script, we determined the frequency of the male allele at each site. This was then used to calculate the RNA transfer index (RTI), which is the proportion of sites per gene that have 5 or more *D. arizonae* reads.

### Identification of D. arizonae male seminal fluid proteins

*Drosophila arizonae* SFPs were identified as part of a larger unpublished analysis comparing *D. mojavensis/D.arizonae* seminal fluid proteomes. *Drosophila arizonae* were reared in standard banana-molasses food[78] or food made with L-lysine-2HCL,4,4,5,5-D4 (Lys 4; Cambridge Isotopes Laboratory, Inc, Tewksbury, MA) as described above. Unmated Lys 4 males and unlabeled females were paired in vials and copulations were observed. Mated females were immediately frozen in liquid nitrogen and stored at −80°C until reproductive tracts were removed and placed in 50 mM ammonium bicarbonate. These tracts were combined in groups of five with mated *D. mojavensis* tracts that were collected in the same manner, except males were labeled with Lys8 (six biological replicates). Protein extraction, digestion, LC-MS/MS, and analysis with MaxQuant[22] (ver. 2.0.3.0) were performed as described above. *Drosophila arizonae* male SFPs were identified by filtering the protein list to include only proteins carrying the Lys4 label. We considered identified proteins to be SFPs if they had at least two Lys4 labeled peptides and were found in any of the six replicates.

### GO-term and protein domain enrichment analyses

To obtain functional predictions for mdFTPs we analyzed GO-terms and protein domains. GO-term enrichment was analyzed using ClueGO.[57] Since GO-terms are better annotated for *D. melanogaster*, we used *D. melanogaster* orthologs of mdFTPs (126/166 had orthologous calls). The background gene list for gene enrichment analysis included orthologs of all *D. arizonae* genes identified in *D. melanogaster*. To assess protein domain enrichment, we identified domains in the *D. arizonae* genome using InterProScan[69] to generate a database.

### Knockout experiments

We used CRISPR knockouts to test the effects of *ARI/26694* and *ARI11629* on aspects of the female postmating response. Genes were targeted using two sgRNAs designed against regions at the 5′ end of each gene following the method outlined by Bassett et al.[81] sgRNAs and Cas9 mRNA (TriLink Biotechnologies, San Diego, CA) were injected into *D. arizonae* embryos by Rainbow Transgenic Flies, Inc. (Camarillo, CA). For each gene, we generated homozygous lines for two different frameshift mutations (Figure S5), which were crossed to make transheterozygous males. Transheterozygous males were used in all experiments except the *ARI/11629* reaction mass experiment because one mutant line was lost prior to this experiment. On the morning of experiments, unmated WT females from a different *D. arizonae* line (ARTU2) were paired with either unmated WT or KO males and copulations were observed. For the reaction mass assay, females were separated from males and placed in liquid nitrogen either immediately after mating or 6 h postmating. Lower reproductive tracts were removed and photographed using a camera mounted to a Leica S9i dissecting microscope (37.5X magnification). Images were analyzed using ImageJ[58] software by using the freeform drawing tool to trace the outline of the lower female reproductive tract. Perimeter and area were calculated, and pixels were converted to μm using photo resolution.

For the oviposition assay, females were separated from males after copulation and placed alone in vials containing banana-molasses food. Flies were moved to new vials every 24 h for seven days and eggs were counted each day. The fertilization success assay was conducted in a different manner for the two genes. For *ARI/26694*, mated females were placed in groups of three in vials containing banana-molasses food. Flies were moved to a new vial every 24 h. Eggs were counted on days one, three, and five, and the number of hatching larvae was assessed for eggs laid on each of these days. This analysis does not differentiate between unfertilized eggs and embryonic mortality, but since we did not

detect differences in hatching, we did not perform additional experiments to tease these effects apart. For *ARI/11629*, we used an assay that better differentiated between embryonic mortality and fertilization deficiency. Mated females were separated from males and placed in population cages overnight on banana-molasses food with yeast paste. The food plate was removed in the morning and stored for 6 h so that developing embryos would be 6 to 22 h. Embryonic development time for these species is approximately 28 h.[82] We used a solution of 2.5% sodium hypochlorite to dechorionate embryos before they were fixed in a of 1:4 solution of 4% paraformaldehyde/heptane and then devitellinized in methanol. We stained embryos with 4′,6-diamidino-2-phenylindole (DAPI; 2.8 μg/mL) and analyzed embryonic development by fluorescence microscopy using a Leica DM5000B microscope (20X objective). We considered eggs to be unfertilized if no more than four nuclei were observed (representing the four products of female meiosis). We acknowledge that although the minimum age of embryos was 6 h, embryogenesis could have been arrested at an earlier stage of development if the embryo was inviable. If this occurred during the earliest syncytial divisions, these embryos would be difficult to distinguish from unfertilized eggs using our methodology. However, this possibility is unlikely given that we did not observe embryos at a range of development times between the earliest syncytial divisions and 6 h, as would be expected if inviability was common. Moreover, we would expect inviability to remain constant over time, but the phenotype we observed changed over time.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### RTI analysis

We compared the RTI among three categories of genes: highly supported mdFTPs, supported mdFTPs, and genes with no support for being mdFTPs. Comparisons were made using a generalized linear mixed model (GLMM) with binomial error distribution implemented in the 'lme4'[71] package for R. Gene category was treated as a fixed effect and gene id was treated as a random effect. An anova table was generated using the 'car'[72] package, and *post-hoc* testing with Tukey's adjustment was performed using the 'multcomp'[73] package.

### GO-term and protein domain enrichment analysis

We tested for enrichment in terms for biological process and molecular function using a right-sided hypergeometric test applying a Benjamini-Hochberg false-discovery rate threshold of 0.05. Enriched terms were grouped by ClueGO based on functional relationships. We report the average fold enrichment per group. We used the R package WebGestalR[70] to determine overrepresentation of domains in our mdFTPs using the *D. arizonae* as its background.

### Statistical analysis of insemination reaction mass data

Data were analyzed as a factorial type II Anova using the 'afex'[74] package for R. The model included male genotype, time, and their interaction as factors. Model fit was evaluated visually using residual plots. *Post-hoc* comparisons were analyzed using the 'emmeans'[75] package for R with Tukey's adjustment for multiple comparisons.

### Statistical analysis of fecundity data

Data were analyzed with the R package 'glmmTMB'[76] using a generalized linear mixed model (GLMM) with a negative binomial error distribution. The model included male genotype, day, and their interaction. Individual was treated as a random effect. Model fit was evaluated using residual plots.

### Statistical analysis of hatching/fertilization efficiency data

For *ARI/26694*, hatching data was analyzed with the 'afex'[74] package using a binomial generalized linear mixed model including genotype, day, and their interaction as factors. Vial was treated as a random effect. *Post-hoc* comparisons were analyzed with the 'emmeans'[75] using Tukey's adjustment for multiple comparisons. For *ARI/11629*, fertilization efficiency was analyzed using a binomial generalized linear model with male genotype, day, and their interaction as factors. Model fit was evaluated using residual plots. *Post-hoc* comparisons were performed with the emmeans[75] package using Tukey's adjustment for multiple testing.