Reproductive functions and genetic architecture of the seminal fluid and sperm proteomes of the mosquito *Aedes aegypti*

Ethan C. Degner1,§, Yasir H. Ahmed-Braimah2,§, Kiril Borziak3, Mariana F. Wolfner2,*; Laura C. Harrington1,*; Steve Dorus3,*

1 Department of Entomology, Cornell University, Ithaca, NY
2 Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY
3 Center for Reproductive Evolution, Syracuse University, Syracuse, NY

§ Authors contributed equally to this work
* MFW: mfw5@cornell.edu; LCH: lch27@cornell.edu; SD: sdorus@syr.edu

Abstract

The yellow fever mosquito, *Aedes aegypti*, transmits several viruses, including dengue, Zika, and chikungunya. Some proposed efforts to control this vector involve manipulating reproduction to suppress wild populations or replacing them with disease-resistant mosquitoes. The design of such strategies requires an intimate knowledge of reproductive processes, yet our basic understanding of reproductive genetics in this vector remains largely incomplete. To accelerate future investigations, we have comprehensively catalogued sperm and seminal fluid proteins (SFPs) transferred to females in the ejaculate using tandem mass spectrometry. By excluding female-derived proteins using an isotopic labelling approach, we identified 870 sperm proteins and 280 seminal fluid proteins. Functional composition analysis revealed parallels with known aspects of sperm biology and SFP function in other insects. To corroborate our proteome characterization, we also generated transcriptomes for testes and the male accessory glands—the primary contributors to *Ae. aegypti* sperm and seminal fluid, respectively. Differential gene expression of accessory glands from virgin and mated males suggests that protein translation is upregulated post-mating. Several SFP transcripts were also modulated after mating, but >90% remained unchanged. Finally, a significant enrichment of SFPs was observed on chromosome 1, which harbors the male sex determining locus in this species. Our study provides a comprehensive proteomic and transcriptomic characterization of ejaculate production and composition and thus provides a foundation for future investigations of *Ae. aegypti* reproductive biology, from functional analysis of individual proteins to broader examination of reproductive processes.

Introduction

The mosquito, *Aedes aegypti*, is the most important vector of arboviruses globally, transmitting viruses that cause dengue [1], Zika [2], chikungunya [3], and yellow fever [4]. Consequently, *Ae. aegypti* places a severe strain on public health infrastructure around the world [5]. Despite decades of effort to control mosquito populations, *Ae. aegypti*
continues to drive human epidemics of these diseases. New and improved control strategies are needed to prevent future outbreaks and mitigate disease burden.

Some promising control strategies under development target reproduction to suppress mosquito populations. For example, sterilized males can be released to suppress populations by impairing reproduction by their wild mates [6, 7]. Manipulating reproductive phenotypes may also provide a means of driving disease-refractory traits into a population (reviewed in [8]). One such strategy employs the intracellular bacterium Wolbachia, which, when introduced into Ae. aegypti, induces cytoplasmic incompatibility that allows the bacterium to spread in a population, potentially to fixation [9]. Cytoplasmic incompatibility causes sperm of males with Wolbachia to be incompatible with uninfected females’ eggs, whereas Wolbachia-positive females can reproduce with any male, regardless of infection status (reviewed in [10]), giving Wolbachia-positive individuals a fitness advantage over their uninfected counterparts. This bacterium also blocks or reduces transmission of several viruses, including dengue [11] and Zika [12]. Consequently, introduction of novel Wolbachia infections into vector populations is being explored as a transmission reducing strategy.

Designing mosquito control strategies that target reproduction requires an intimate knowledge of the cellular and molecular mechanisms that govern it. Yet, only a few functions of proteins involved in mosquito reproduction have been described to date. For example, in Ae. aegypti seminal fluid proteins (SFPs) induce several physiological and behavioral changes in females, including refractoriness to future mating [13–18], stimulation of oogenesis [19], enhanced survival [18], and the ability to fertilize eggs [20]. However, the molecular identity of active SFP components for this species remains elusive. Seminal fluid initiates sperm motility via the action of proteases in many insects (silkworm [21]; water strider [22]; Culex mosquito [23]), but the precise sperm proteins on which seminal fluid acts in Ae. aegypti have not been identified. Similarly, sperm-associated odorant receptors may control motility in Ae. aegypti, although the exact function and ligands of these receptors are unknown [24]. Finally, the mechanism by which Wolbachia induces cytoplasmic incompatibility has not been described in Ae. aegypti, but Wolbachia proteins contained in sperm are hypothesized to be involved [25, 26].

Identification of sperm proteins and SFPs that are transferred to females during copulation is an important objective to enable future investigations into specific reproductive processes. Components of the transferred ejaculate include sperm and seminal fluid—both of which play vital roles in mosquito reproduction. An Ae. aegypti seminal fluid proteome was first reported by Sirot et al. [27] based on mass spectrometry analyses. That study described 93 putative SFPs transferred during mating. While not a primary focus of their work, they also identified 101 putative sperm proteins. Later work identified more than twice as many SFPs from Ae. albopictus using a similar method [28]. Proteome complexity of other insects’ seminal fluid (reviewed in [29–31]) and sperm [32–34] suggests that more proteins remain to be identified in the Ae. aegypti ejaculate.

Here, we build on the foundational work of Sirot et al. [27], using tandem mass spectrometry (MS/MS) with greater sensitivity to identify the constituent proteins of both Ae. aegypti sperm and seminal fluid. We also profile the transcriptomes of the male accessory glands (MAG; before and after mating) and testes, the major source tissues for SFPs and sperm, respectively. Our proteomic characterization represents a nearly four-fold expansion of putative SFPs and a more than eight-fold expansion in the Ae. aegypti sperm proteome. Our results yield insights into the molecular function, genome organization, regulation, and evolution of sperm proteins and SFPs in this important disease vector. Ultimately, these proteomes provide a basis for future studies of mosquito reproduction and, potentially, a catalog of molecular targets for the
development of novel mosquito control methods.

**Materials and Methods**

**Rearing**

Mosquitoes were derived from a laboratory colony of *Ae. aegypti* that was established from individuals collected in Bangkok, Thailand, in 2011 and supplemented with wild caught mosquitoes every 2 - 3 y. Mosquitoes were reared as described previously by Degner and Harrington [35]. Briefly, eggs were hatched under vacuum pressure, and a day later 200 first instar larvae were transferred to rearing trays of 1 L deionized water with four fish food pellets as diet (except in the case of $^{15}$N-labeled females; see below). Pupae were separated by sex based on size, allowed to eclose individually in separate test tubes, and adults were transferred into single-sex cages with 10% sucrose provided *ad libitum*.

**Sperm isolation**

Males used for sperm protein sample preparation were aged between 5 and 8 days post-eclosion (dpe). Mature sperm were isolated from the seminal vesicles of each male. Males were dissected in physiological saline (133 mM NaCl, 2.63 mM KCl, 9.75 mM Na$_2$HPO$_4$, 3 mM KH$_2$PO$_4$, 2 mM CaCl$_2$, adjusted to pH 6.9; hereafter “saline”). The seminal vesicle was isolated from other tissue and consecutively transferred to two clean droplets of saline to remove any adherent fat body or other debris. Clean dissecting tools were used to transfer the seminal vesicle to a final droplet of saline where it was ruptured to release sperm. Sperm suspended in saline were transferred to a microcentrifuge tube on ice, and pooled sperm samples were flash frozen in liquid nitrogen every 2 h.

Two biological replicates included sperm combined from 400 and 470 randomly selected males, respectively. Pooled samples were centrifuged at 25,000 x g and the supernatant was removed, leaving 18µl saline with the pellet. An equal volume of 2x Laemmli buffer + 5% β-mercaptoethanol was added to the pellet, and samples were solubilized by sonicating for 30 s, boiling for 15 min, and re-sonicating for 30 s. To remove any particulate debris, samples were spun down at 10,000 x g for 10 min, and the supernatant was placed in a fresh tube. Protein was quantitated using a 1:169 dilution of the sample using the EZQ assay (Thermo Fisher Scientific, Waltham, MA). Protein used for subsequent mass spectrometry was standardized across biological replicates (16µg).

**Transferred ejaculate esolation**

Males were reared as described above. As larvae, females were labelled with $^{15}$N using the rearing methodology of Sirot *et al.* [27]. Briefly, a prototrophic yeast strain (D273-10B) was grown in media whose only nitrogen source was $^{15}$N ammonium sulfate (Cambridge Isotope Laboratories; Cambridge, MA). Yeast were grown to saturation, pelleted, and resuspended in PBS to a final volume of one sixteenth of the growth media. A few drops of yeast slurry were provided to newly hatched first instar larvae after vacuum hatching. One day after hatching, 200 larvae were placed in a rearing tray with 200 mL water from a previous cohort of $^{15}$N-yeast-reared mosquitoes and 800 mL of deionized water. Larvae were fed 4 mL of labeled yeast slurry a day after hatching, and again at 4 d after hatching. Pupae were isolated in individual tubes to ensure virginity upon eclosion, and females were put in an 8 L bucket cage with sucrose...
solution ad libitum. Sucrose was replaced every 2 d to preclude the introduction of unlabeled nitrogen via microbial contamination.

At 4 - 5 dpe, matings between labeled females and unlabeled males were observed as in Degner and Harrington [35]. After mating, females were immediately placed on ice and dissected within 3 min. Because mosquito seminal fluid is known to contain proteases [27], mosquitoes were dissected in saline with protease inhibitors (cOmplete Mini Protease Inhibitor Cocktail; Sigma Aldrich, St. Louis, MO). In contrast to Sirot et al. [27], we dissected only bursae (and not spermathecae), cutting the bursae just proximal to the vaginal lips. Upon excision, bursae (n = 35 - 43 per replicate) were transferred to 33 L 1x Laemmli buffer diluted from a 2x stock solution with saline, -mercaptoethanol and 1x protease inhibitors. Samples were sonicated for 30 s, boiled for 10 min, sonicated for 30 s, and centrifuged at 10,000 x g for 10 min. The supernatant was removed; 30 L were frozen at -80 ◦ C, and 2 µL were diluted 1:3 with buffer for protein quantitation using the EZQ assay (Thermo 194 Fisher Scientific, Waltham, MA). In parallel, we also prepared samples of bursae from virgin, labeled females from each cohort to assess the efficiency of 15 N-labeling. Mated and virgin bursae samples contained 18 and 8 g of protein, respectively.

**Tandem mass spectrometry analysis**

Solubilized proteins were separated on a 1-dimensional SDS-PAGE gel and split into 6 fractions, with two biological replicates run in parallel (Figure S1). Gel fractions were reduced in dithiothreitol, alkylated with iodoacetamide, and digested with trypsin. Lyophilized, digested proteins were reconstituted in 0.5% formic acid and subjected to nanoLC-ESI-MS/MS analysis using an Orbitrap Fusion Tribrid mass spectrometer (Thermo-Fisher Scientific, San Jose, CA) equipped with nanospray Flex Ion Source, and coupled with a Dionex UltiMate 3000RSLC nano system (Thermo, Sunnyvale, CA). Peptide samples were injected onto a PepMap C-18 RP nano trap column (5 µm, 100 µm i.d. x 20mm, Dionex) with nanoViper fittings at 20 µL/min flow rate for desalting. Samples were then separated on a PepMap C-18 RP nano column (2 µm, 75 µm x 15cm) at 35 ◦ C, followed by elution on a 90 min gradient of 5% to 35% acetonitrile in 0.1% formic acid at 300 nL/min. Finally, a 5 min ramping to 90% acetonitrile in 0.1% formic acid and a 5 min hold at this eluent completed each run cycle. Between cycles, the column was re-equilibrated for 25 min using 0.1% formic acid. The Orbitrap Fusion was run in positive spray ion mode with spray voltage set at 1.6 kV and a source temperature at 275 ◦ C. External calibration for FT, IT, and quadrupole mass analyzers was performed. In data-dependent acquisition analysis, the instrument was operated using FT mass analyzer in MS scan to select precursor ions followed by 3 s “Top Speed” data-dependent CID ion trap MS/MS scans at 1.6 m/z quadrupole isolation for precursor peptides with multiple charged ions above a threshold ion count of 10,000 and normalized collision energy of 30%. MS survey scans at a resolving power of 120,000 (fwhm at m/z 200), for the mass range of m/z 375-1575. Dynamic exclusion parameters were set at repeat count 1 with a 20 s repeat duration, an exclusion list size of 500, and 40 s of exclusion duration with ±10 ppm exclusion mass width. The activation time was 10 ms 219 for CID analysis. All data were acquired under Xcalibur 3.0 operation software (Thermo-Fisher Scientific). All post-quantitation sample preparation was conducted at the Cornell Biotechnology Resource Center.

**Peptide identification and protein annotation**

Raw data from each MS/MS run was analyzed by X!Tandem [36] and Comet [37] against the Ae. aegypti L5.0 protein database (GCF_002204515.2. [38]). Only the longest protein isoform of each gene was included in the search database, resulting in a
database of 14,626 proteins. For X!Tandem, a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 10.0 PPM were used. For Comet, a fragment bin tolerance of 1.0005 with a 0.4 offset and a parent ion tolerance of 10.0 PPM were used. Iodoacetamide derivative of cysteine was specified as a fixed modification, whereas oxidation of methionine and deamidation of glutamine and asparagine were specified as variable modifications. Peptides were allowed up to two missed trypsin cleavage sites. All downstream analyses were conducted using the Trans-Proteomic Pipeline (TPP v5.0 (Typhoon) rev 0; [39]). False Discovery Rates (FDRs) were estimated with a randomized decoy database using PeptideProphet [40], employing accurate mass binning model and the nonparametric negative distribution model. X!Tandem and Comet PeptideProphet results were merged using iProphet [41], to provide more robust peptide identification. Peptide identifications were accepted if they could be established at greater than 95.0% iProphet probability, and protein assignations were accepted if they could be established at greater than 99.0% probability. Proteins that contained identical peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy parsimony principles.

Verification of labelling efficiency

For ejaculate samples, protein from virgin female controls was run on the same gel as protein from their mated counterparts, and labeling efficiency was verified on a representative fraction in each cohort’s virgin sample (Figure S1). No peptides or proteins were identified using our statistical criteria in the labelled virgin female samples when searched using standard, unlabelled mass parameters. Thus, whole female labelling was complete in relation to MS/MS peptide and protein identification and precluded the identification of female proteins from mated female bursae.

Protein quantitation

Protein quantitation was conducted using the semi-quantitative spectral counting approach implemented by the APEX Quantitative Proteomics Tool [42]. The top 50 identified proteins with the highest protein identification probabilities were utilized as the training dataset. The 35 physicochemical properties available in the APEX tool were used for prediction of peptide detection/non-detection in the construction of a training dataset file. Protein probabilities ($O_i$) were computed using the Random Forest classifier algorithm trained with the data set generated in the previous step. APEX protein abundances per sample were calculated using the protXML file generated by ProteinProphet.

Experimental design and statistical rationale

To ensure the reproducibility of protein identifications, two biological replicates of each tissue were analyzed and robust identification criteria were applied (see above). Each replicate was prepared from independent cohorts of mosquitoes. To control for the possibility that unlabelled female-derived proteins were identified in our ejaculate samples, we assessed labelling efficiency by conducting mass spectrometry on virgin bursae alone, using a representative gel slice from each biological replicate (Figure S1). False Discovery Rates (FDRs) were estimated with a randomized decoy database using PeptideProphet [40], employing accurate mass binning model and the nonparametric negative distribution model. Corrections for multiple testing were applied where appropriate to ensure the conservative nature of statistical tests.
Transcriptome analysis of testes and male accessory glands

Testes were harvested from males at 1 dpe and transferred to TRIzol. Because mature sperm are actively produced at this age [43], and spermatogenesis is at its peak [44,45], testes at this age likely contain the majority of transcripts that contribute to the testicular sperm proteome. Male accessory glands (MAG) including the connecting ejaculatory duct were dissected from virgin males aged 6 and 8 dpe. We also analyzed MAG from mated males at the same age. Previous work has demonstrated that Ae. aegypti males become depleted after mating with three to five females in succession, and seminal fluid is slowly regenerated over 48 h [46]. In our study, we provided males with four virgin females for a period of 8 h to allow for male seminal fluid depletion. On average, each male mated with more than three females in this period (as determined by dissection of females’ spermathecae). Males were dissected in saline 16 h after their female mates had been removed. We generated four biological replicates from independent cohorts for each treatment, and each replicate contained combined tissue from 20-40 (testes) or 40-60 (MAG) males. Total RNA was extracted from each sample in Trizol following manufacturer’s instructions (Invitrogen, Carlsbad, CA). Poly-A mRNA was isolated and cDNA libraries were prepared using the QuantSeq 3’ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Vienna, Austria). Amplified cDNA products were run on an AATI Fragment Analyzer (Advanced Analytical Technologies, Inc.; Hialeah, FL) so that the cDNA was of sufficient concentration for sequencing. Library concentrations were balanced using digital PCR [47], and each of the 12 uniquely barcoded samples were sequenced in one lane using the Illumina HiSeq 2500 platform with 100 bp reads. All sequencing was conducted at the Cornell Biotechnology Resources Center.

For further analysis of the transcriptomes, we included additional, publicly available data to evaluate tissue-biased gene expression. These include gonadectomized male carcass (SRP075464; [48]) and a virgin female reproductive tract sample (SRP068996; [49]). Raw RNA-seq reads were processed by trimming the first 10 bases from the 3’ position, followed by quality trimming of both ends to a minimum quality Phred score of 20 (Sickle v.1.210; [50]). Processed reads were 293 bp long and mapped to the Ae. aegypti genome (VectorBase release L5.1; [38]) using Hisat2 (v.2.1.0; [51]) with default parameters, and transcript abundance was estimated for each sample with StringTie (v.1.3.4; [52]). Raw counts for each sample were extracted from the StringTie abundance estimates using the auxiliary “prepDE.py” script provided on the StringTie website (https://ccb.jhu.edu/software/stringtie/). Signal peptides in the translated transcriptome were predicted in silico using a local installation of SignalP (v.4.1; [53]).

We used raw counts from the RNA-seq samples to (1) classify genes based on tissue-biased expression in the MAG and testes, and (2) identify genes differentially expressed between virgin and mated MAG. Count matrices were filtered to remove low abundance transcripts (counts per million < 5). First, we compared expression levels between the focal tissue (testes or virgin MAG) to three other tissues: gonadectomized male carcass [48], a virgin female reproductive tract sample [49], and virgin MAG or testes (present study). We classified genes as testes- or MAG-biased if they had >2-fold higher transcript abundance compared to other samples at a minimum FDR cutoff of 0.05 (edgeR v.3.23.2; [54]). Finally, we identified differentially expressed genes between virgin and mated MAGs as having >2-fold abundance difference at an FDR cutoff of 0.01. Using the same differential expression criteria, we also re-analyzed data from female reproductive tracts of virgin and just-mated females [49] to identify transcripts putatively transferred to females in the ejaculate using current annotated gene models.
Chromosomal distribution of male reproductive genes

To evaluate the chromosomal distribution of SFPs, sperm proteins and MAG/testes-biased genes, we calculated the expected number of genes for each class on each chromosome, assuming a random distribution of genes across the genome. We then multiplied the total number of genes within each class by the expected proportion for each chromosome (based on the proportion of total genes on that chromosome) to establish an observed/expected ratio. We also calculated this ratio for a 123 Mb region on chromosome 1 that surrounds the sex determining locus and has low rates of 317 recombination \cite{55}. A 318 \( \chi^2 \) test (d.f. = 1) for each gene class per chromosome was used to test for biased representation.

Orthology relationships and gene ontology analysis

Species orthology was assessed using a local installation of OrthoDB with default BLAST and clustering parameters \cite{56}. Protein sequences for \textit{Ae. aegypti} and \textit{Ae. albopictus} were retrieved from NCBI (GCF\textunderscore002204515.2 and GCF\textunderscore001876365.2, respectively), and \textit{Drosophila melanogaster} protein sequences were retrieved from FlyBase (r6.18). Protein sequences for each species were filtered to retain only the longest isoform for each gene. Sperm proteins and SFPs for \textit{Ae. albopictus} and \textit{D. melanogaster} were based on previous mass spectrometry based proteomic studies \cite{28,32,57}. Because SFP identifications by Boes \textit{et al.} \cite{28} were made using a de novo transcriptome (the \textit{Ae. albopictus} genome had not yet been released), we converted the original SFP identifications to their current accession numbers by blasting assembled transcripts from Boes \textit{et al.} \cite{28} to the assembled \textit{Ae. albopictus} genome \cite{58}. All Gene Ontology (GO) analyses were conducted using GOseq \cite{59}, with gene lengths derived from the longest transcript of each assembled StringTie gene. GO terms were extracted from BLAST results against the SwissProt database (www.uniprot.org).

Data Availability

Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD010293 and 10.6019/PXD010293. The R analysis pipeline used for transcriptomic analysis is available as part of this project’s GitHub repository (https://github.com/YazBraimah/Aegypti.ejaculatomics). Raw RNA-seq reads from this study can be obtained from the Sequence Read Archive (SRA) accession SRP158536. RNA-seq data from Sutton \textit{et al.} \cite{48} and Alfonso-Parra \textit{et al.} \cite{49} can be accessed at SRA accessions SRP075464 and SRP068996, respectively.

Results

Sperm and ejaculate proteome characterization

To characterize \textit{Ae. aegypti} proteins transferred to females during mating and distinguish between sperm and non-sperm seminal fluid components, we used MS/MS to analyze proteins from (1) purified sperm isolated from male seminal vesicles and (2) whole ejaculate from bursae of mated females. For our analysis of the whole ejaculate, females were labelled with heavy nitrogen to preclude detection of female-contributed proteins. Labelling efficiency was determined to be complete with regard to peptide and protein identification (see Materials and Methods). Two replicates were analyzed per sample. High levels of reproducibility were observed, including 85\% of sperm proteins and 78\% of ejaculate proteins identified in both biological replicates. Similar levels of
inter-replicate consistency have been reported for sperm proteomes of other organisms [34]. To maximize protein identification, we combined spectra across biological replicates for our final proteome determination, resulting in 54,894 peptide spectral matches (PSMs) for sperm, and 30,801 for the ejaculate (Table S1). The nearly two-fold disparity in PSMs is due to the contribution of labelled female proteins in the ejaculate sample. In total, 870 and 811 proteins with at least 2 unique PSMs were identified in sperm and ejaculate, respectively (Table S2). Sperm proteins were identified by an average of 11.4 unique PSMs and 62.8 total PSMs per protein. Ejaculate proteins were identified with an average of 8.0 unique PSMs and 37.8 total PSMs. As was expected given the substantial contribution of sperm cells to ejaculate composition, extensive overlap was observed between sperm and ejaculate proteomes; 516 proteins were detected in both samples, while 354 proteins were only identified in sperm and 295 proteins were uniquely detected in the ejaculate (Figure 1A).

The primary goal of this study was to use MS/MS with higher sensitivity and accuracy to expand upon the prior characterization of Ae. aegypti SFPs and sperm by Sirot et al. [27]. They identified 74 SFPs that mapped to the recently refined Ae. aegypti genome [38]; some of the 93 SFPs described by Sirot et al. [27] do not map to the new genome or are now part of larger, fused gene models. Of these, we identified 60 (81%) in our ejaculate sample, 32 of which were also identified in our purified sperm sample. It is noteworthy that we detected an additional 5 SFPs from Sirot et al. [27], but these were not included in our SFP list because they did not meet our two unique peptide inclusion criteria. As such our proteomic characterization expands the previous Ae. aegypti SFP characterization.

Refined seminal fluid protein classification

Because the ejaculate is a complex mixture of sperm and seminal fluid, we applied stringent inclusion criteria to our ejaculate samples to refine the delineation between SFPs and sperm proteins. We first removed 10 proteins involved in protein translation (i.e. ribosomal proteins, translation initiation factors, and elongation factors; Table S2) from the list of putative SFPs. These proteins exhibit ubiquitous patterns of expression, including both MAG and testes, and are unlikely to be bona fide secreted SFPs. Although we cannot rule out that they are secreted SFPs, their presence may be the result of cell rupture during seminal secretion [60] or transfer of MAG cells to the female, as has been described in D. melanogaster [61]. To reduce the possible inclusion of sperm proteins that were absent in our sperm proteome but present in the ejaculate (perhaps due to low abundance), we define “high confidence SFPs” as proteins with a minimum of 6 total PSMs in the ejaculate, 2 unique PSMs, and not present in our sperm proteome; this resulted in 177 high-confidence SFPs (Table S2).

Previous analyses of insect sperm proteomes have consistently identified proteins generally considered to be SFPs (i.e. highly expressed in the MAG and believed to be secreted molecules transferred to females as non-sperm components [32,62,63]. To identify proteins predominantly produced by the MAG, but identified in both our sperm and ejaculate sample, we used a 2.5-fold greater abundance threshold in the ejaculate relative to sperm. This resulted in the identification of 103 additional putative SFPs, which we label as “sperm/SFP overlap” (Table S2), including 53 with 5-fold greater protein abundance in the ejaculate relative to sperm. In total, this results in a combined SFP proteome of 280 proteins.

To better understand the relationship between candidate SFPs identified in the current study and by Sirot et al. [27], we examined protein abundance and transcript levels of both protein sets (50 of the 74 SFPs identified by those authors are present in our SFP proteome). This revealed that SFPs identified by Sirot et al. [27] were significantly more abundant than the remainder of our 280 SFPs (Figure 1B): Wilcoxon
Figure 1. Characteristics of the *Ae. aegypti* sperm and ejaculate proteomes. (A) Venn diagram representing the number of proteins identified by ≥2 unique peptides using LC-MS/MS in sperm (red) and ejaculate (blue) proteomes. (B) Scatter plot of the protein (log2PSM) and mRNA (log2TPM) abundances of the 177 SFPs and 103 overlap SFPs. Proteins classified as SFPs in the Sirot et al. study are highlighted in green and those not identified in Sirot et al. [27] are highlighted in blue. Dotted line is a linear model fit, and gray shading represents the 95% confidence interval. (C) Percentage of proteins containing a predicted signal peptide sequence among SFPs, sperm, and SFP/sperm overlap protein classes; all proportions are significantly different from each other ($\chi^2$, $p < 0.05$). (D) Heatmap representing the mRNA abundance profile of male reproductive classes identified in this study. Cladograms on the left and top represent Pearson cluster grouping of genes and samples, respectively. Annotation bars on the left indicate the protein classification for each gene. (E) Violin plots displaying the fold-change (log2) distribution of mRNA abundance between MAG and testes samples for each male reproductive protein class. Width of violins represents the number of proteins at any given TPM ratio; boxes represent inner quartiles and outliers drawn using the Tukey method. Asterisks indicate groups that are significantly different from each other (Wilcoxon Signed rank test; *** $= p < 0.001$).
rank sum test; $W = 3786, p < 0.001$), consistent with the higher sensitivity and coverage of new MS/MS methods utilized in this study. Proteins identified by Sirot et al. [27] also exhibited significantly higher levels of MAG expression (Figure 1B; Wilcoxon rank sum test; $W = 3005, p < 0.001$). We therefore conclude that the sensitivity of our SFP characterization resulted in the addition of a far greater number of low abundance SFPs.

To evaluate our SFP characterization, we explored three types of analyses. First, we determined the presence of predicted signal peptides in identified proteins — a hallmark of secreted proteins. Amongst our high-confidence SFPs, $\sim 33\%$ contained signal peptides, in comparison to only $\sim 9\%$ of sperm proteins ($\chi^2 = 69.2, p < 0.001$). Additionally, $\sim 17\%$ of the sperm/SFP overlap proteins had a predicted signal peptide — also significantly more than the sperm proteome ($\chi^2 = 6.47, p = 0.01$), and less than high-confidence SFPs ($\chi^2 = 7.7, p = 0.006$; Figure 1C). Second, we used RNA-seq to characterize the MAG and testis transcriptome, the predominant source tissues of proteins in our analysis. As expected, SFPs (Wilcoxon Signed rank test; $p < 0.001$) and sperm proteins (Wilcoxon Signed rank test; $p < 0.001$) exhibit MAG- and testis-biased expression, respectively, while the sperm/SFP overlap proteins exhibit less tissue-biased expression levels overall (Wilcoxon Signed rank test; $p = 0.36$; Figure 1D-E). Third, we assessed the amount of high-confidence SFP protein abundance variation explained by variation in MAG expression levels. This revealed a significant correlation between protein and transcript abundance ($R^2 = 0.49, F = 88.24, p < 0.001$; Figure 1B). In conjunction, these results support our proteomic characterization of SFPs secreted from the MAG, including proteins present in both sperm and ejaculate samples but highly enriched in the ejaculate proteome.

**Gene ontology of sperm and seminal fluid proteomes**

Gene ontology (GO) categories enriched in mosquito sperm are largely similar to those found in other insects’ sperm (Table 1, Table S3; [32,34,62,63]). Proteins associated with mitochondria and the axoneme are the most enriched cellular components in the sperm proteome. Proteasome components are also enriched. Over-represented biological processes in sperm include nucleotide biosynthesis, metabolic processes related to the tricarboxylic acid cycle, and proteins regulating ciliar function. Nucleotide binding, ion transport, and oxidoreductase activity are enriched molecular functions in sperm.

In our high-confidence SFP proteome, extracellular structure was significantly enriched amongst cellular component categories, further supporting the accuracy of our SFP identification. Over represented biological processes include proteolysis and both carbohydrate and amide metabolism. Significantly enriched molecular functions include hydrolase activity and peptidase activity. This observation is consistent with the widespread presence of proteolytic enzymes and regulators in SFPs of other insects (reviewed in [64]). The sperm/SFP overlap proteome shares several enriched categories with both the sperm and high-confidence SFP proteomes, including carboxylic acid metabolic processes and nucleoside binding. Several new GO terms emerge in this protein set as well, including ATP hydrolysis coupled proton transport, vesicle-mediated transport, and protein folding (Table 1, Table S3).

**Orthology with sperm proteins and SFPs in other species**

We next examined orthology of sperm proteins and SFPs in two different species: *D. melanogaster*, given its well-characterized sperm proteome and SFPs [32,32], and *Ae. albopictus*, which is the closest species to *Ae. aegypti* with characterized SFPs [28]. Orthology was determined between the complete genome of all three species, and then orthologs of *Ae. aegypti* SFPs and sperm proteins also classified as SFPs or sperm...
| Dataset                      | Ontology                | GO Term                                      | Proteins (total) | FDR       |
|-----------------------------|-------------------------|----------------------------------------------|------------------|-----------|
| Sperm Proteome              | Biological Process      | organonitrogen compound metabolic process    | 102 (978)        | 5.30E-11  |
|                             |                         | carboxylic acid metabolic process            | 100 (553)        | 1.90E-23  |
|                             |                         | electron transport chain                     | 13 (55)          | 6.50E-05  |
|                             | Cellular Component      | mitochondrion                                | 150 (834)        | 1.20E-45  |
|                             |                         | microtubule                                  | 36 (210)         | 1.30E-07  |
|                             |                         | cilium                                        | 32 (154)         | 1.70E-08  |
|                             |                         | proteasome complex                           | 31 (47)          | 2.10E-25  |
|                             |                         | dynein complex                               | 19 (40)          | 1.70E-11  |
|                             | Molecular Function      | nucleoside phosphate binding                 | 187 (1629)       | 5.70E-20  |
|                             |                         | hydrolase activity                           | 163 (2153)       | 4.70E-04  |
|                             |                         | oxidoreductase activity                      | 103 (839)        | 2.00E-13  |
|                             |                         | threonine-type endopeptidase activity        | 14 (15)          | 8.30E-16  |
| High-confidence SFP proteome| Biological Process      | organonitrogen compound metabolic process    | 30 (978)         | 0.009     |
|                             |                         | carboxylic acid metabolic process            | 21 (553)         | 0.033     |
|                             | Cellular Component      | extracellular region part                    | 33 (1116)        | 0.029     |
|                             | Molecular Function      | hydrolase activity                           | 62 (2153)        | 0         |
|                             |                         | nucleoside phosphate binding                 | 43 (1629)        | 0.04      |
|                             |                         | peptidase activity                           | 27 (739)         | 0.006     |
| Sperm/SFP overlap           | Biological Process      | carboxylic acid metabolic process            | 29 (553)         | 7.20E-11  |
|                             |                         | vesicle-mediated transport                   | 15 (536)         | 0.017     |
|                             |                         | protein folding                              | 10 (113)         | 2.60E-05  |
|                             | Molecular Function      | ATP hydrolysis coupled proton transport       | 8 (25)           | 4.30E-08  |
|                             |                         | oocyte microtubule cytoskeleton polarization  | 3 (10)           | 0.014     |

proteins in the other species were identified. In the comparison with *Ae. albopictus*, we focus solely on SFPs because a thorough sperm proteome is lacking. Overall, ~87% and ~98% of proteins in the *Ae. aegypti* genome have an ortholog (either as one to-one or one-to-many relationships) in *D. melanogaster* and *Ae. albopictus*, respectively. Among our identified proteins unique to the *Ae. aegypti* sperm proteome, 760 (99%) have an ortholog in *D. melanogaster*, and 451 (59%) of these are also found in the *D. melanogaster* sperm proteome (Table S2; [32,63]). Out of the 280 *Ae. aegypti* SFPs characterized in our study (including 177 high-confidence and 103 sperm/SFP overlap proteins), 275 (98%) have orthologs in *D. melanogaster*. Of these, only 11 have also been identified as *D. melanogaster* SFPs (4%; Table S2). Proteins that contribute to the seminal fluid proteome have therefore diverged extensively during Dipteran evolution. Orthologs were identified in the *Ae. albopictus* genome for 275 (98%) of our SFPs. Of the *Ae. albopictus* SFPs identified to date [28], 86 (43%) were classified as SFPs in our study (Table S2).

**MAG and testis transcriptome characterization and differential expression**

We used short-read RNA sequencing to examine gene expression in testes and MAGs to (1) identify transcripts with tissue-biased expression, (2) compare transcript and protein abundance, (3) assess the transfer of male RNAs to females during mating and (4) characterize the effect of mating on MAG gene expression. In both MAG and testes, we find a significant correlation between transcript abundance and protein abundance for...
Figure 2. Protein and mRNA abundance relationships. (A) and (B) Scatterplots of normalized mRNA and protein abundance in MAG/ejaculate (A) and testes/sperm (B). The line and gray shading represent a linear model fit and its 95% confidence interval, respectively. Correlation coefficients are indicated. (C) and (D) Volcano plots of protein abundance differences for all proteins detected in both the ejaculate and sperm samples. Proteins that show >2-fold mRNA expression-bias in MAG or testes tissue are indicated in blue (MAG-biased) or red (testes-biased), and the size of each point corresponds to the number of unique peptides detected for each protein.

Because males regenerate seminal fluid over the course of 48 h after depleting their reserves by repeated insemination [46], we reasoned that MAG transcriptional regulation after mating might inform our understanding of pathways required to restore depleted SFPs. Differential expression analysis of virgin and mated MAGs' transcriptomes revealed a significant bias towards gene upregulation in mated males, with 320 transcripts that are upregulated and 126 that are downregulated after mating (binomial test; \( p < 0.001 \); Figure 3A). In contrast to downregulated transcripts—which
Table 2. Gene Ontology analysis of upregulated genes in MAGs after mating.

| Ontology              | GO Term                                      | Proteins (total) | FDR       |
|-----------------------|----------------------------------------------|------------------|-----------|
| Biological Process    | protein targeting to ER                      | 9 (17)           | 1.88E-07  |
|                       | ubiquitin-dependent protein catabolic process | 21 (197)         | 5.70E-06  |
|                       | Response to stress                           | 56 (1321)        | 8.11E-04  |
|                       | cellular amino acid metabolic process        | 19 (246)         | 0.001     |
|                       | signal peptide processing                    | 4 (7)            | 0.004     |
| Molecular Function    | aminoacyl-tRNA ligase activity               | 12 (52)          | 2.46E-06  |
|                       | threonine-type endopeptidase activity        | 7 (15)           | 1.36E-05  |
| Cellular Component    | proteasome complex                           | 17 (47)          | 1.18E-12  |
|                       | endoplasmic reticulum part                   | 33 (660)         | 0.005     |
|                       | small nucleolar ribonucleoprotein complex    | 5 (18)           | 0.015     |
|                       | organelle inner membrane                     | 18 (288)         | 0.031     |
|                       | nucleolus                                    | 20 (345)         | 0.035     |

are not enriched for any functional category—upregulated transcripts are enriched for several GO categories, many of which are consistent with this tissue’s primary function of producing secreted proteins (Table 2). For example, amino acid metabolism and aminoacyl-tRNA ligase activity are enriched. Enrichment of proteins involved in ubiquitin-dependent protein catabolism and those making up components of the proteasome suggests that there may be extensive protein recycling as seminal fluid is regenerated. Furthermore, the over representation of proteins associated with endoplasmic reticulum targeting and signal peptide processing is consistent with the vesicle-mediated apocrine secretion by which some of mosquito seminal fluid is reported to be produced [60]. We also note that a variety of immune-related genes are upregulated in MAGs after mating, including a family of Defensin antimicrobial genes (Figure 3A). Lastly, a similar pattern was observed among our characterized SFPs, of which 19 were upregulated after mating and only five that decreased in abundance (binomial test; p < 0.001; Figure 3B). Interestingly, upregulated SFPs included six different cytoplasmic tRNA ligases and a catalytic subunit of a signal peptidase responsible for post-translation removal of secretion signal peptides, consistent with the importance of protein production and secretion in the mated MAG (see Discussion).

Paternal mRNA transfer during mating

Previously, we have shown that males transfer mRNA to females in the ejaculate [49]. Using the newly annotated genome [38], we re-analyzed data from those experiments and identified 106 transcripts, including 41 protein coding genes and 17 long non-coding RNAs, that are putatively transferred to females. Identification was based on >2-fold increase in transcript abundance in females immediately after mating, followed by a subsequent decrease in transcript abundance. The remainder of the identified transcripts (48 of 106) are currently unannotated. Using the transcriptomic data in the current study, we determined that the vast majority of these transcripts have MAG-biased expression, with only two transcripts exhibiting testis-biased expression (Figure 3C). The MAG therefore appears to be a primary source of RNA transferred to females in the ejaculate. In total, 27 proteins encoded by transferred mRNA transcripts were identified in our proteomes, with 22 in the high-confidence SFP proteome, three in the sperm/SFP overlap proteome, and five that were in the sperm proteome. Interestingly, the putatively transferred transcripts whose products were present in our seminal fluid proteome encode highly abundant proteins that were on average six times more plentiful than the remainder of the seminal fluid proteins. Lastly, transferred transcripts exhibited a general trend towards down-regulation in the MAGs of mated
males compared to MAGs of virgins, including 13 transferred transcripts that were significantly down-regulated after mating between virgin and mated MAGs (Figure 3D);
this pattern may be due to their transfer without replenishment by the MAG at the time of dissection.

**SFPs are enriched on chromosome 1**

Sex chromosomes present exclusively in males, such as the mammalian and insect Y chromosome, are highly enriched for genes with male-biased function, including many critical to spermatogenesis and sperm function \([32, 65, 66]\). Although *Ae. aegypti* lack heteromorphic sex chromosomes, chromosome 1 harbors a region of robust linkage disequilibrium surrounding the recently characterized male sex determination locus, *Nix* \([55, 67]\). To assess the enrichment of male-biased genes on this chromosome we examined the physical distribution of SFPs, sperm proteins, and MAG/testes-biased genes by chromosome. This revealed a \(~1.5\)-fold enrichment of SFPs on chromosome 1 (62 observed, 39 expected; \(\chi^2 = 12.4, p < 0.001\)) and \(~1.2\)-fold enrichment of sperm proteins on chromosome 2 (377 observed, 316 expected; \(\chi^2 = 11.6, p < 0.001\); Figure 4). Despite the strong enrichment of SFPs on chromosome 1, we did not observe an over-representation of SFPs in the tightly linked region in the vicinity of the male determining locus, *Nix* (15 observed, 16 expected; \(\chi^2 = 0.08, p = 0.8\)). Together, these results suggest that chromosome 1 does exhibit aspects of male-biased functional composition, but that this signature is not readily detectable in the linked region that harbors the male sex-determining factor *Nix*. Non-uniform physical distribution of sperm protein across autosomes has also been observed in *Drosophila* \([63]\). Although the functional significance of this remains to be determined, it is possible that clustering facilitates co-expression during spermatogenesis, which is characterized by progressive genome silencing during the histone-to-protamine repackaging transition.

![Figure 4](image)

**Figure 4.** Chromosomal distribution of genes encoding SFPs, sperm proteins, or transcripts with MAG biased expression or testis biased expression. The y-axis represents the ratio of observed/expected number of genes for each of the chromosomes in *Ae. aegypti*, and the dashed line represents the expectation under no enrichment/depletion. Asterisks indicate significant enrichment (\(\chi^2; \ast \ast p < 0.01\)).

**Discussion**

A rapidly expanding body of evidence supports the critical roles of seminal fluid proteins (SFPs) in a wide array of reproductive phenotypes (reviewed in \([29]\)). Although this has been most extensively investigated in *Drosophila*, seminal peptides are also associated with post-mating behavioral and physiological responses in mosquitoes such
as *Ae. aegypti* [15,16,18]. The primary goals of this study were to comprehensively catalog male proteins transferred to *Ae. aegypti* females during insemination and establish a reliable methodology for delineating between sperm proteins and SFPs. To accomplish this, we (1) conducted an in-depth proteomic characterization of sperm, (2) utilized a whole-female labelling approach to identify unlabelled male proteins transferred by the male during insemination and (3) characterized the transcriptomes of the testis and male accessory gland (MAG). Importantly, we note that the whole-female labelling approach has been employed previously in *Ae. aegypti* but the assignment of proteins as SFPs was limited by the lack of information regarding proteins found in sperm. Thus, distinctions between sperm proteins and SFPs were previously difficult to achieve. It is also noteworthy that advances in MS/MS sensitivity and accuracy have resulted in far greater power of detection in our study, and our analysis has also benefited tremendously from the recent resequencing and reannotation of the *Ae. aegypti* genome [38]. Our proteomic characterization resulted in a nearly four-fold expansion of the current *Ae. aegypti* seminal fluid proteome, and an eight-fold expansion of identified proteins in sperm. Together, this *Ae. aegypti* "ejaculatome" provides a foundation for future molecular studies of mosquito reproduction and associated applications to control mosquito populations (see below).

**Proteome characteristics independently validate identification**

Our work differs from previous SFP characterization studies [27,28,57] in that our classification was supported by a detailed knowledge of sperm proteome composition. Nonetheless, several independent validation approaches were helpful in assessing the quality of our proteomic characterization. For example, we quantified the proportion of proteins with predicted secretion signals and analyzed transcriptome profiles in testes and MAGs. As would be predicted, SFPs identified in this study possessed a significantly higher proportion of predicted secretion signals than sperm proteins and were, on average, highly specific or biased towards expression in the MAG. Additionally, analysis of the functional composition of our proteomes revealed that they were closely aligned with the results of previous sperm [32–34, 63] and SFP studies in insects [30, 57].

For example, our expanded sperm proteome was highly enriched for proteins related to flagellar structure, including microtubules, dynein complexes, and ciliar components, and proteins likely associated with the mitochondrial derivatives, which are a predominant structure in mosquito sperm [68,69] and that of other insects. Consistent with what has been described by Sirot *et al.* [27], as well as in other insects (reviewed in [29,30,57]) and humans [70], proteases were highly enriched amongst our high-confidence SFPs, supporting the likely accuracy of our expanded characterization (reviewed in [64]). The observed enrichment of vesicle-mediated transport proteins is also consistent with the fact that mosquito seminal fluid is in part produced by apocrine secretion [60]. Additionally, exosomes and other vesicles are believed to play a role in a variety of post-insemination cellular interactions. For example, vesicles transferred in *Drosophila* seminal fluid have been reported to fuse with sperm and interact with the female reproductive tract [71], exosomes of the mouse epididymis have recently been implicated in the control of sperm RNA stores [72], and the abundance of exosome markers in avian SFPs has led to speculation about vesicle-mediated mechanisms in post-testicular sperm maturation [73]. Therefore, the accuracy of our expanded proteomic characterization of sperm and SFP proteomes is corroborated by several independent lines of evidence.

It is important to note that, despite the application of stringent proteomic thresholds, some proteins could not be definitively assigned as either sperm protein or SFP. Previous studies in *Drosophila* and Lepidoptera have consistently identified known SFPs (such as Acp36DE) at appreciable abundance levels in sperm that have yet to be
combined with MAG secretions \cite{32,62,63}. Our identification of a relatively large protein set that is highly MAG-biased in expression but also present in sperm further suggests that the incorporation of “SFPs” during testicular sperm maturation occurs and is worthy of additional functional investigation. Although Drosophila expression profiles in the testis and accessory gland are quite distinct, many SFPs exhibit low levels of co-expression in the testis (Dorus, unpublished data). Our transcriptomic analyses here further support such patterns of co-expression. As such, dichotomous distinctions between sperm proteins and SFPs may be an oversimplification of a more nuanced relationship between these reproductive systems. We acknowledge this uncertainty in our classification of MAG-biased proteins that were also identified in our sperm proteome.

We also note that despite our expanded proteomic coverage, several proteins that we anticipated to be identified were absent. The most notable case was Head Peptide-1, a seminal fluid peptide which has been shown to be transferred in the ejaculate \cite{74,15} and has been reported to induce short term monogamy in the female after mating \cite{15}. Head Peptide-1, as is the case for many SFPs, undergoes extensive post-translational modification and may therefore be challenging to identify bioinformatically without a priori knowledge of the biochemical composition of the proteolytic products (such as in the case of the well-studied Drosophila Sex Peptide; \cite{75}). Another example was adipokinetic hormone (AAEL011996), which did not meet our two unique peptide inclusion threshold, although we did identify five copies of one peptide from its precursor protein that was also identified in Ae. albopictus seminal fluid \cite{28}. This protein has been postulated to contribute to sperm protection from oxidative stress \cite{76} and the regulation of feeding behavior \cite{77} in other insects. We suggest that the complexity of proteolytic pathways, governed both by male and female interacting proteins, is a major barrier in the use of shotgun proteomics to study SFP identity and function in the female reproductive tract. Future investigations would likely benefit from the inclusion of a targeted proteomic approach (reviewed in \cite{78}). Such approaches require an a priori list of candidate peptides; in Ae. aegypti, the neuropeptides and protein hormones catalogued by Predel et al. \cite{79} represent a useful pool of potentially important molecules.

**Evolution of male reproductive proteomes**

Male reproductive proteins, including SFPs, are consistently among the fastest evolving classes of protein (reviewed in \cite{80}). Although initially a goal of our study, conducting a robust analysis of the molecular evolution of protein identified in this study was limited by the availability of genomic resources appropriate for both inter- and intraspecific tests of positive selection. Obtaining high quality genomic data for different populations of Ae. aegypti has proven difficult, given the genome’s repetitive nature \cite{38,81,31}. Furthermore, this mosquito’s ability to move globally as diapausing eggs has allowed for frequent mixing and a complex population structure \cite{38,81,31}. The development of appropriate population level genetic data for the analysis of recent selective sweeps should be a priority in Ae. aegypti, as it has been in Anopheles gambiae \cite{34,85}. Furthermore, given the extent of molecular divergence between Ae. aegypti and Ae. albopictus \cite{28}, the development of genomic resources for a more closely related outgroup to Ae. aegypti will assist in understanding evolutionary patterns at the gene level. Despite these limitations, our analysis of orthology did reveal that the suite of proteins contributing to seminal fluid, but not sperm, has diverged substantially from other Dipterans. Although sperm proteins and SFPs possess levels of orthology to the Drosophila genome that are comparable to the genome as a whole, only 59% and 4% of orthology was observed when comparing the Ae. aegypti sperm and SFP proteomes (respectively) with those of Drosophila \cite{32,67}. While some of this disparity may be
attributed to differences in overall proteome size and coverage, such a stark contrast is nonetheless compelling evidence of tissue-specific evolutionary patterns. Orthology between *Ae. aegypti* SFPs and *Ae. albopictus* SFPs [28], while more extensive (43%), was still comparatively low compared to orthology between the sperm proteomes of *Ae. aegypti* and *D. melanogaster*—two distantly related Dipterans. These results suggest a process of “turn-over” in seminal fluid proteomes, whereby overall protein composition diverges rapidly even when there is evidence for conservation with regard to overarching molecular functions represented in seminal fluid. For example, a priori expectations about Gene Ontology enrichment were met for both *Ae. aegypti* sperm (*e.g.*, cilium and mitochondrial proteins) and SFPs (extracellular localization and hydrolase activity), despite overall SFP divergence. SFPs are a pronounced target of selection and have been discussed as a driver of sexual conflict (reviewed in [86]), and thus they are expected to rapidly diverge. By contrast, we note that strong conservation of sperm proteins exists across distant taxa, with different insect orders displaying 25% orthology between sperm proteomes [34], and even *D. melanogaster* and mammals with 20% sperm proteome orthology [32]. The overall lack of conservation in seminal fluid proteomes makes comparing the roles of specific SFPs across species difficult, but conserved molecular functions amongst SFPs will nevertheless allow the wealth of knowledge in *Drosophila* to be leveraged towards an understanding of SFP function in non-model insects.

Unlike other mosquitoes with heteromorphic sex chromosomes, Culicine mosquitoes (*e.g.*, *Aedes* and *Culex*) harbor male determining loci on undifferentiated, homomorphic sex chromosomes [87]. Theory predicts the evolution of heteromorphic sex chromosomes following the acquisition of a sex determining locus, suppression of recombination, and expansion of the non-recombining region. It remains unclear why homomorphic sex chromosomes appear to be retained in some taxa [88,89]. One proposed mechanism to mediate the selective effect of sexually antagonistic alleles on the promotion of recombination suppression is the establishment of efficient sex-biased expression [90]. Although previously lacking, the significant enrichment of SFPs on chromosome 1 is the first evidence in support of this hypothesis in *Ae. aegypti*. This trend was restricted to SFPs and was not observed for genes solely over-expressed in the MAG or testis. It is intriguing to speculate that this distinction between SFPs and other male reproductive genes might be due to the prevalence (and selective strength) of sexually antagonistic alleles specifically amongst SFPs, which may favor their localization on chromosome 1. This is consistent with their putative role as drivers of sexual conflict (reviewed in 86), including the mediation of female post-mating responses such as sexual receptivity and longevity [14,18].

**Functional relevance of abundant sperm proteins and SFPs**

Our sperm and SFP proteomes exhibited skewed abundance distribution with the top ten most abundant proteins comprising 25% and 17% of protein composition in sperm and SFPs, respectively. Interestingly, the most abundant sperm protein, cytosol aminopeptidase (AAEL006975), accounted for more than 7.4% of all protein and two other cytosol aminopeptidases were in the top ten most abundant proteins (AAEL000108, AAEL023987). These three proteins are orthologs of the eight sperm-leucyl aminopeptidases (S-LAPs) in *Drosophila* with similar expression patterns, including ∼1000-fold higher expression in testes than in MAG, ∼50 times more transcript in whole male carcasses than gonadectomized carcasses [91], and upregulation during later stages of spermatogenesis [48]. S-LAP orthologs constitute a significant proportion of the protein composition of *Drosophila* [92] and Lepidoptera [34,62]. Little is known about the specific function of S-LAPs, although it has been postulated that may serve a structural function given the inferred loss of enzymatic capacity of several S-LAPs during *Drosophila* evolution [92]. Additionally, a Y-linked S-LAP in *D.*
pseudoobscura has been implicated in a cryptic meiotic drive system, where suppression of this locus results in aberrant spermatogenesis and a higher proportion of X-bearing sperm [93]. It will be of great interest to establish the specific function of these proteins in mosquito sperm, given their high abundance and expression patterns during spermiogenesis. Furthermore, the proteins and transcripts involved in spermatogenesis described in this study may assist in the identification of other genes involved in meiotic drive systems (reviewed in [94]), which have been proposed as potential genetic means to reduce wild populations through the induction of sex ratio biases [95].

Although no SFP was as abundant as cytosol aminopeptidase in sperm, the top ten most abundant proteins ranged from 1.2 – 2.6% of the protein in our ejaculate sample. L-asparaginase (AAEL002796) was the most abundant SFP (61% more abundant than the next protein) and the tenth most abundant mRNA transcript in the MAG out of over 11,000 transcripts. While the relevance of the abundance of this enzyme is currently unclear, it may relate to several other notable observations. First, transcript AAEL020035, whose protein product is comprised of ~60% asparagine residues, is the single most abundant MAG transcript and was also, by far, the most abundant putatively transferred transcript. (We did not identify AAEL020035 in our SFP proteome but note that it results in few identifiable peptides because of its extreme amino acid composition). Second, asparagine tRNA-ligase was two times more abundant in seminal fluid than any other tRNA-ligase. Third, asparagine tRNA-ligase was upregulated in the MAG after mating and was the most abundant tRNA-ligase transcript. Together, these suggest that MAGs are well-equipped to produce ample protein with an asparagine amino acid bias. Finally, two other enzymes, aspartate transaminase (AAEL002399) and citrate synthase, are abundantly present in seminal fluid and could convert aspartate produced by asparaginase to oxaloacetate and citric acid, respectively. While it is premature to draw any firm conclusions based on these observations alone, it is intriguing to speculate that the SFP proteome has the capacity to conduct gluconeogenesis (of asparagine and potentially other amino acids) and that this may feed into the citric acid cycle. The citric acid cycle is believed to be functional in mammalian sperm (reviewed in [96]) and many citric acid cycle enzymes are present in our Ae. aegypti sperm proteome.

The most abundant seminal fluid proteins also exhibited a strong enrichment for function in protein cleavage. Proteins with protease, dipeptidase, and aminopeptidase activity represent 12 of the top 28 most abundant proteins present in the seminal fluid proteome. Proteolytic functions have been described previously in the seminal fluid of Ae. aegypti [27], Ae. albopictus [28], Cx. quinquefasciatus [97], and several non-mosquito taxa [21,22] (reviewed in [64,98]), and are a common function of many insects’ seminal fluid. Based on studies in other insects, functions of these enzymes may include the activation of sperm motility or the cleavage of propeptides into their active forms [99]. Our seminal fluid proteome also contains abundant enzymes that catabolize smaller substrates, such as amino acids and carbohydrates. Taken together, the enzymatic cocktail in seminal fluid may be well equipped to break down many of the molecules they contain. Seminal fluid proteomes were also enriched for proteins involved in maintaining proton and redox homeostasis. We identified several proteins contributing to V-type proton ATPases, which use ATP to regulate pH via proton transport. Maintaining an optimal pH in seminal fluid may allow for efficient sperm motility (reviewed in [100]). Regulating pH may also create an ideal environment in which enzymatic reactions occur. Seminal fluid also contained several proteins that function to neutralize free radicals, such as catalase (AAEL013407), peroxidase (AAEL013171), and several dehydrogenases. Regulating the physiochemical environment in seminal fluid is likely critical for the function and protection of sperm prior to their storage in the female’s long term storage organs (spermathecae).
Ejaculate RNAs Transferred to Females

There has been much conjecture about the importance of spermtazoal RNA to fertility [101] and recent work has confirmed that the regulation of sperm ncRNA stores in the mammalian epididymis is necessary for proper embryogenesis [102]. Little is known about the function of spermtazoal RNAs in insects, although they have been demonstrated to have substantial functional coherence, including an overwhelming enrichment of loci involved in translation [103]. New data in this study allowed us to probe for patterns in previously described transcripts that are putatively transferred to females during mating. A total of 106 transcripts were identified, including both coding and non-coding transcripts, and a majority of these exhibit high levels of expression in the MAG. Based on our SFP proteome, most of the protein coding transcripts are translated at high levels. Their high expression in the MAG suggests that they may simply hitchhike into seminal fluid with other secreted molecules. Alternatively, as has been demonstrated in *Drosophila*, they could be transferred in intact MAG cells [61], or via vesicles derived from the MAG [71]. Interestingly these vesicles, which may carry RNA cargo including miRNAs, fuse with sperm and have the capacity to interact with the female reproductive tract. Some male-derived transcripts are detectable in the female for up to 24 h post-mating [46], and it has been postulated that they could be used by females in some capacity [104]. In *Ae. aegypti*, both vesicles and RNAs are transferred in the ejaculate to the female, but their fate and function have not been investigated. Whether they impact the female or her future offspring is an intriguing, and potentially important, line of future investigation.

Mosquito control and future directions

Understanding the molecular architecture of *Ae. aegypti* reproduction holds great potential for vector control strategies. Mosquito reproduction is an ideal control target to reduce vector populations and the burden of disease transmission. The most direct application of this study will be the identification of modulators of female reproductive behavior. Mosquito SFPs induce behavioral responses that prevent female remating [14,15,17], including short term mating refractory behavior [15]. To date, the molecule(s) responsible for long term refractoriness has yet to be identified. Given the strength and duration of responses to low SFP “doses” [14], identification of the responsible proteins will provide powerful tools for manipulating female reproduction in a species-specific manner. In addition, such knowledge may provide a molecular metric by which the quality of males in modified mosquito release strategies (such as those employing sterile or Wolbachia-infected males; reviewed in [105]) may be monitored and optimized. Functional analysis of specific sperm proteins and SFPs may yield insights into processes such as sperm motility and activation [21,23], sperm storage [106], and sperm-egg recognition [107]. Very few studies have explored these processes in *Ae. aegypti* (reviewed in [108]). A mechanistic understanding of complex post-copulatory male-by-female interactions is critical to genetically modified mosquito release strategies that manipulate reproduction. Our detailed characterization of the male contributions to these interactions should serve as the foundation for the design and improvement of vector control strategies that limit the transmission of arboviruses that cause serious human illness and mortality.
Supporting Information

Figure S1

One-dimension SDS-PAGE gel separation of samples. Sperm (A) and ejaculate (B) protein fractions (1 - 6) analyzed by LC-MS/MS. Biological replicates were run in parallel on 10% bis-tris SDS- PAGE gel with MOPS running buffer and stained with colloidal Coomassie. V, virgin bursae; M, mated bursae. Labeling efficiency of females was verified through the analysis of fraction 5 from virgin females.

Table S1

Database of all proteins identified by tandem mass spectrometry in this study. Classification of each protein is based on criteria in text. Proteomic (probability of identification, percent coverage, total PSMs, and unique PSMs) and transcriptomic (TPM in virgin MAG, mated MAG, and testes) data are included, as well as whether each protein includes a signal peptide, and how each protein was previously classified by Sirot *et al.* [27].

Table S2

Full GO analysis for high confidence SFPs, sperm/SFP overlap, and sperm proteins. CC, cellular component; BP, biological process; MF, molecular function.

Acknowledgments

We thank Sylvie Pitcher, Sheng Zhang, Jen Grenier, Peter Schweitzer, and the staff at the Cornell Biotechnology Resource Center for technical support, and Laura Sirot for experimental guidance and feedback. This study was supported by *NIH/NIAID grant R01AI095491* to MFW and LCH, *NIH/NICHD grant R21HD088910* to SD and MFW, a Cornell Graduate School fellowship to ECD, and a Cornell Entomology Department Griswold grant to ECD and LCH. YHAB was supported by *NIH/NICHD grant R01HD059060* to MFW and Andrew G. Clark. RNA-seq data and mass spectrometry data were made possible by NIH grants 1S10OD010693-01 and 1S10OD017992-01, respectively.

References

1. S Bhatt, P W Gething, O J Brady, J P Messina, A W Farlow, C L Moyes, J M Drake, J S Brownstein, A G Hoen, O Sankoh, M F Myers, D B George, T Jaenisch, G R W Wint, C P Simmons, T W Scott, J J Farrar, and S I Hay. The global distribution and burden of dengue. *Nature*, 496(7446):504–507, April 2013.

2. T Chouin-Carneiro, A Vega-Rua, M Vazeille, A Yebakima, R Girod, D Goindin, M Dupont-Rouzeyrol, R Lourenco-de Oliveira, and A B Failloux. Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika virus. *PLOS Neglected Tropical Diseases*, 10(3):e0004543, March 2016.

3. G Pialoux, B A Gauzere, S Jaureguiberry, and M Strobel. Chikungunya, an epidemic arbovirosis. *Lancet Infect Dis*, 7(5):319–327, May 2007.
4. T P Monath and P F Vasconcelos. Yellow fever. *Journal of Clinical Virology*, 64:160–173, March 2015.

5. A C Fredericks and A Fernandez-Sesma. The burden of dengue and chikungunya worldwide: implications for the southern United States and California. *Ann Glob Health*, 80(6):466–475, November 2014.

6. W Klassen and C F Curtis. History of the Sterile Insect Technique. In V A Dyck, J Hendrichs, and A S Robinson, editors, *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management*, pages 3–36. Springer Netherlands, Dordrecht, 2005.

7. D O Carvalho, A R McKemey, L Garziera, R Lacroix, C A Donnelly, L Alphey, A Malavasi, and M L Capurro. Suppression of a field population of *Aedes aegypti* in Brazil by sustained release of transgenic male mosquitoes. *PLOS Neglected Tropical Diseases*, 9(7):e0003864, 2015.

8. V M Macias, J R Ohm, and J L Rasgon. Gene drive for mosquito control: where did it come from and where are we headed? *International Journal of Environmental Research and Public Health*, 14(9), September 2017.

9. A A Hoffmann, B L Montgomery, J Popovici, I Iturbe-Ormaetxe, P H Johnson, F Muzzi, M Greenfield, M Durkan, Y S Leong, Y Dong, H Cook, J Axford, A G Callahan, N Kenny, C Omodei, E A McGraw, P A Ryan, S A Ritchie, M Turelli, and S L O’Neill. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature*, 476(7361):454–457, 2011.

10. D LePage and S R Bordenstein. *Wolbachia*: Can we save lives with a great pandemic? *Trends in parasitology*, 29(8):385–393, August 2013.

11. T Walker, P H Johnson, L A Moreira, I Iturbe-Ormaetxe, F D Frentiu, C J McMeniman, Y S Leong, Y Dong, J Axford, P Kriesner, A L Lloyd, S A Ritchie, S L O’Neill, and A A Hoffmann. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature*, 476(7361):450–453, August 2011.

12. H L Dutra, M N Rocha, F B Dias, S B Mansur, E P Caragata, and L A Moreira. *Wolbachia* blocks currently circulating Zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. *Cell Host Microbe*, 19(6):771–774, June 2016.

13. R W Gwadz, G B Jr Craig, and W A Hickey. Female sexual behavior as the mechanism rendering *Aedes aegypti* refractory to insemination. *The Biological Bulletin*, 140(2):201–214, 1971.

14. M E Helinski, P Deewatthanawong, L K Sirot, Mariana F Wolfner, and L C Harrington. Duration and dose-dependency of female sexual receptivity responses to seminal fluid proteins in *Aedes albopictus* and *Ae. aegypti* mosquitoes. *J Insect Physiol*, 58(10):1307–1313, October 2012.

15. L B Duvall, N S Basrur, H Molina, C J McMeniman, and L B Vosshall. A peptide signaling system that rapidly enforces paternity in the *Aedes aegypti* mosquito. *Current Biology*, 27(23):3734–3742 e5, December 2017.

16. M S Fuchs, G B Craig, and D D Despommier. The protein nature of the substance inducing female monogamy in *Aedes aegypti*. *Journal of insect physiology*, 15(701):701–709, 1969.
17. M S Fuchs, G B Jr Craig, and E A Hiss. The biochemical basis of female monogamy in mosquitoes. I. Extraction of the active principle from *Aedes aegypti*. *Life sciences*, 7(16):835–839, 1968.

18. S M Villarreal, S Pitcher, M E H Helinski, L Johnson, Mariana F Wolfner, and L C Harrington. Male contributions during mating increase female survival in the disease vector mosquito *Aedes aegypti*. *J Insect Physiol*, 108:1–9, July 2018.

19. M J Klowden and G M Chambers. Male accessory gland substances activate egg development in nutritionally stressed *Aedes aegypti* mosquitoes. *Journal of insect physiology*, 37(10):721–726, 1991.

20. V Adlakha and M K Pillai. Involvement of male accessory gland substance in the fertility of mosquitoes. *Journal of insect physiology*, 21(8):1453–1455, August 1975.

21. S Nagaoka, K Kato, Y Takata, and K Kamei. Identification of the sperm-activating factor initiatorin, a prostatic endopeptidase of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 42(8):571–582, August 2012.

22. H Miyata, C D Thaler, L T Haimo, and R A Cardullo. Protease activation and the signal transduction pathway regulating motility in sperm from the water strider *Aquarius remigis*. *Cytoskeleton*, 69(4):207–220, April 2012.

23. C D Thaler, H Miyata, L T Haimo, and R A Cardullo. Waveform generation is controlled by phosphorylation and swimming direction is controlled by Ca2+ in sperm from the mosquito *Culex quinquefasciatus*. *Biology of reproduction*, 89(6), December 2013.

24. R Jason Pitts, Chao Liu, Xiaofan Zhou, Juan C Malpartida, and Laurence J Zwiebel. Odorant receptor-mediated sperm activation in disease vector mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7):2566–2571, 2014.

25. J F Beckmann and A M Fallon. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochemistry and Molecular Biology*, 43(9):867–878, September 2013.

26. J F Beckmann, J A Ronau, and M Hochstrasser. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature Microbiology*, 2:17007, March 2017.

27. Laura K Sirot, Melissa C Hardstone, Michelle E H Helinski, José M C Ribeiro, Mari Kimura, Prasit Deewatthanawong, Mariana F Wolfner, and Laura C Harrington. Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLOS Neglected Tropical Diseases*, 5(3):e989, 2011.

28. Kathryn E Boes, José M C Ribeiro, Alex Wong, Laura C Harrington, Mariana F Wolfner, and Laura K Sirot. Identification and characterization of seminal fluid proteins in the Asian tiger mosquito, *Aedes albopictus*. *PLOS Neglected Tropical Diseases*, 8(6):e2946, 2014.

29. Frank W Avila, Laura K Sirot, Brooke A LaFlamme, C Dustin Rubinstein, and Mariana F Wolfner. Insect seminal fluid proteins: identification and function. *Annual Review of Entomology*, Vol 56, 56:21–40, 2011.
30. B Baer, R Zareie, E Paynter, V Poland, and A H Millar. Seminal fluid proteins differ in abundance between genetic lineages of honeybees. Journal of Proteomics, 75(18):5646–5653, October 2012.

31. I Sepil, B R Hopkins, R Dean, M L Thezenas, P D Charles, R Konietzny, R Fischer, B M Kessler, and Stuart Wigby. Quantitative proteomics identification of seminal fluid proteins in male Drosophila melanogaster. bioRxiv, 2018.

32. E R Wasbrough, S Dorus, S Hester, J Howard-Murkin, K Lilley, E Wilkin, A Polpitiya, K Petritis, and T L Karr. The Drosophila melanogaster sperm proteome-II (DmSP-II). Journal of Proteomics, 73(11):2171–2185, October 2010.

33. R Zareie, H Eubel, A H Millar, and B Baer. Long-term survival of high quality sperm: insights into the sperm proteome of the honeybee Apis mellifera. Journal of Proteome Research, 12(11):5180–5188, November 2013.

34. E Whittington, Q Zhao, K Borziak, J R Walters, and S Dorus. Characterisation of the Manduca sexta sperm proteome: Genetic novelty underlying sperm composition in Lepidoptera. Insect Biochemistry and Molecular Biology, 62:183–193, July 2015.

35. E C Degner and L C Harrington. Polyandry depends on postmating time interval in the dengue vector Aedes aegypti. The American Journal of Tropical Medicine and Hygiene, 94(4):780–785, April 2016.

36. R Craig and R C Beavis. TANDEM: matching proteins with tandem mass spectra. Bioinformatics, 20(9):1466–1467, June 2004.

37. J K Eng, T A Jahan, and M R Hoopmann. Comet: an open-source MS/MS sequence database search tool. Proteomics, 13(1):22–24, January 2013.

38. B J Matthews, O Dudchenko, S Kingan, S Koren, I Antoshechkin, J E Crawford, W J Glassford, M Herre, S N Redmond, N H Rose, G D Weedall, Y Wu, S S Batra, C A Brito-Sierra, S D Buckingham, C L Campbell, S Chan, E Cox, B R Evans, T Fansiri, I Filipovic, A Fontaine, A Gloria-Soria, R Hall, V S Joardar, A K Jones, R G G Kay, V Kodali, J Lee, G J Lyckett, S N Mitchell, J Muehling, A Omer, F A Partridge, P Peluso, A P Aiden, V Ramasamy, G Rasic, S Roy, K Saavedra-Rodriguez, S Sharan, A Sharma, M Smith, J Turner, A M Weakley, Z Zhao, O S Akbari, W C Black, H Cao, A C Darby, C Hill, J S Johnston, T D Murphy, A S Raikhel, D B Sattelle, I V Sharakhov, B J White, L Zhao, E L Aiden, R S Mann, L Lambrechts, J R Powell, M V Sharakhova, Z Tu, H M Robertson, C S McBride, A R Haste, J Korlach, D E Neafsey, M Phillipps, and L B Vosshall. Improved Aedes aegypti mosquito reference genome assembly enables biological discovery and vector control. bioRxiv, 2017.

39. E W Deutsch, L Mendoza, D Shteynberg, J Slagel, Z Sun, and R L Moritz. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics: Clinical Applications, 9(7-8):745–754, August 2015.

40. A Keller, A I Nesvizhskii, E Kolker, and R Aebersold. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical Chemistry, 74(20):5383–5392, October 2002.
41. D Shteynberg, E W Deutsch, H Lam, J K Eng, Z Sun, N Tasman, L Mendoza, R L Moritz, R Aebersold, and A I Nesvizhskii. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Molecular & Cellular Proteomics*, 10(12):M111 007690, December 2011.

42. J C Braisted, S Kuntumalla, C Vogel, E M Marcotte, A R Rodrigues, R Wang, S T Huang, E S Ferlanti, A I Saeed, R D Fleischmann, S N Peterson, and R Pieper. The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics*, 9:529, December 2008.

43. A N Clements. *The Biology of Mosquitoes: Development, Nutrition, and Reproduction*, volume 1. New York, CABI Publishing, 1992.

44. A J Hatala, L C Harrington, and E C Degner. Age and body size influence sperm quantity in male *Aedes albopictus* (Diptera: Culicidae) mosquitoes. *J Med Entomol*, 55(4):1051–1054, June 2018.

45. A Ponlawat and L C Harrington. Age and body size influence male sperm capacity of the dengue vector *Aedes aegypti* (Diptera : Culicidae). *Journal of medical entomology*, 44(3):422–426, May 2007.

46. Catalina Alfonso-Parra, Frank W Avila, Prasit Deewatthanawong, Laura K Sirot, Mariana F Wolfner, and Laura C Harrington. Synthesis, depletion and cell-type expression of a protein from the male accessory glands of the dengue vector mosquito *Aedes aegypti*. *Journal of insect physiology*, 70:117–124, November 2014.

47. A A Morley. Digital PCR: A brief history. *Biomolecular Detection and Quantification*, 1(1):1–2, September 2014.

48. E R Sutton, Y Yu, S M Shimeld, H White-Cooper, and A L Alphey. Identification of genes for engineering the male germline of *Aedes aegypti* and *Ceratitis capitata*. *BMC genomics*, 17(1):948, November 2016.

49. C Alfonso-Parra, Y H Ahmed-Braimah, E C Degner, F W Avila, S M Villarreal, J A Pleiss, Mariana F Wolfner, and L C Harrington. Mating-induced transcriptome changes in the reproductive tract of female *Aedes aegypti*. *PLOS Neglected Tropical Diseases*, 10(2):e0004451, February 2016.

50. N A Joshi and J N Fass. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. 2011.

51. D Kim, B Langmead, and S L Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12(4):357–360, April 2015.

52. M Pertea, D Kim, G M Pertea, J T Leek, and S L Salzberg. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9):1650–1667, September 2016.

53. H Nielsen. Predicting secretory proteins with SignalP. *Methods in molecular biology (Clifton, N.J.)*, 1611:59–73, 2017.

54. M D Robinson, D J McCarthy, and G K Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, January 2010.
55. A Fontaine, I Filipovic, T Fansiri, A A Hoffmann, C Cheng, M Kirkpatrick, G Rasic, and L Lambrechts. Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, *Aedes aegypti*. *Genome Biology and Evolution*, 9(9):2322–2335, September 2017.

56. E M Zdobnov, F Tegenfeldt, D Kuznetsov, R M Waterhouse, F A Simao, P Ioannidis, M Seppey, A Loetscher, and E V Kriventseva. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, 45(D1):D744–D749, January 2017.

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

58. X G Chen, X Jiang, J Gu, M Xu, Y Wu, Y Deng, C Zhang, M Bonizzoni, W Dermauw, J Vontas, P Armbruster, X Huang, Y Yang, H Zhang, W He, H Peng, Y Liu, K Wu, J Chen, M Lirakis, P Topalis, T Van Leeuwen, A B Hall, C Thorpe, R L Mueller, C Sun, R M Waterhouse, G Yan, Z J Tu, X Fang, and A A James. Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44):E5907–15, November 2015.

59. M D Young, M J Wakefield, G K Smyth, and A Oshlack. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11(2):R14, 2010.

60. S Ramalingam. Secretion in the male accessory glands of *Aedes aegypti* (L.) (Diptera: Culicidae). *International Journal of Insect Morphology and Embryology*, 12(2/3):87–96, 1983.

61. A Leiblich, L Marsden, C Gandy, L Corrigan, R Jenkins, F Hamdy, and C Wilson. Bone morphogenetic protein- and mating-dependent secretory cell growth and migration in the *Drosophila* accessory gland. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47):19292–19297, November 2012.

62. E Whittington, D Forsythe, K Borziak, T L Karr, J R Walters, and S Dorus. Contrastting patterns of evolutionary constraint and novelty revealed by comparative sperm proteomic analysis in Lepidoptera. *BMC genomics*, 18(1):931, December 2017.

63. S Dorus, S A Busby, U Gerike, J Shabanowitz, D F Hunt, and T L Karr. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nature Genetics*, 38(12):1440–1445, December 2006.

64. B A LaFlamme and Mariana F Wolfner. Identification and function of proteolysis regulators in seminal fluid. *Molecular Reproduction and Development*, 80(2):80–101, February 2013.

65. A B Carvalho, B A Dobo, M D Vibranovski, and Andrew G Clark. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 98(23):13225–13230, November 2001.

66. B T Lahn and D C Page. Functional coherence of the human Y chromosome. *Science*, 278(5338):675–680, October 1997.
67. A B Hall, S Basu, X Jiang, Y Qi, V A Timoshevskiy, J K Biedler, M V Sharakhova, R Elahi, M A Anderson, X G Chen, I V Sharakhov, Z N Adelman, and Z Tu. Sex determination. A male-determining factor in the mosquito *Aedes aegypti*. *Science*, 348(6240):1268–1270, June 2015.

68. S N Bao and W de Souza. Ultrastructural and cytochemical studies of the spermatid and spermatozoon of *Culex quinquefasciatus* (Culicidae). *Journal of Submicroscopic Cytology and Pathology*, 25(2):213–222, April 1993.

69. M Ndiaye, X Mattei, and O T Thiaw. Maturation of mosquito spermatozoa during their transit throughout the male and female reproductive systems. *Tissue and Cell*, 29(6):675–678, December 1997.

70. B Pilch and M Mann. Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biology*, 7(5):R40, 2006.

71. L Corrigan, S Redhai, A Leiblich, S J Fan, S M Perera, R Patel, C Gandy, S M Wainwright, J F Morris, F Hamdy, D C Goberdhan, and C Wilson. BMP-regulated exosomes from *Drosophila* male reproductive glands reprogram female behavior. *The Journal of Cell Biology*, 206(5):671–688, September 2014.

72. U Sharma, F Sun, C C Conine, B Reichhoff, S Kukreja, V A Herzog, S L Ameres, and O J Rando. Small RNAs are trafficked from the epididymis to developing mammalian sperm. *Developmental Cell*, 46(4):481–494 e6, August 2018.

73. K Borziak, A Alvarez-Fernandez, L Karr T, T Pizzari, and S Dorus. The seminal fluid proteome of the polyandrous red junglefowl offers insights into the molecular basis of fertility, reproductive ageing and domestication. *Scientific Reports*, 6:35864, November 2016.

74. C Naccarati, N Audsley, J N Keen, J H Kim, G J Howell, Y J Kim, and R E Isaac. The host-seeking inhibitory peptide, Aea-HP-1, is made in the male accessory gland and transferred to the female during copulation. *Peptides*, 34(1):150–157, March 2012.

75. P S Chen, E Stumm-Zollinger, T Aigaki, J Balmer, M Bienz, and P Bohlen. A male accessory gland peptide that regulates reproductive behavior of female *D. melanogaster*. *Cell*, 54(3):291–298, 1988.

76. A Bednarova, D Kodrik, and N Krishnan. Adipokinetic hormone exerts its anti-oxidative effects using a conserved signal-transduction mechanism involving both PKC and cAMP by mobilizing extra- and intracellular Ca2+ stores. *Comparative Biochemistry and Physiology, Part C: Toxicology & Pharmacology*, 158(3):142–149, September 2013.

77. T Konuma, N Morooka, H Nagasawa, and S Nagata. Knockdown of the adipokinetic hormone receptor increases feeding frequency in the two-spotted cricket *Gryllus bimaculatus*. *Endocrinology*, 153(7):3111–3122, July 2012.

78. E Borras and E Sabido. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics*, 17(17-18), September 2017.

79. R Predel, S Neupert, S F Garczynski, J W Crim, M R Brown, W K Russell, J Kahnt, D H Russell, and R J Nachman. Neuropeptidomics of the mosquito *Aedes aegypti*. *Journal of Proteome Research*, 9(4):2006–2015, April 2010.
80. W J Swanson and V D Vacquier. The rapid evolution of reproductive proteins. *Nat Rev Genet*, 3(2):137–144, February 2002.

81. V Nene, J R Wortman, D Lawson, B Haas, C Kodira, Z J Tu, B Loftus, Z Y Xi, K Megy, M Grabihr, Q H Ren, E M Zdobnov, N F Lobo, K S Campbell, S E Brown, M F Bonaldo, J S Zhu, S P Sinkins, D G Hogenkamp, P Amedeo, P Arensburger, P W Atkinson, S Bidwell, J Biedler, E Birney, R V Bruggner, J Costas, M R Coy, J Crabtree, M Crawford, B deBruyn, D DeCaprio, K Eighmeier, E Eisenstadt, H El-Dorry, W M Gelbart, S L Gomes, M Hammond, L I Hannick, J R Hogan, M H Holmes, D Jaffe, J S Johnston, R C Kennedy, H Koo, S Kravitz, E V Kriventseva, D Kulp, K LaButti, E Lee, S Li, D D Lovin, C H Mao, E Maucei, C F M Menck, J R Miller, P Montgomery, A Mori, A L Nascimento, H F Naveira, C Nusbaum, S O’Leary, J Orvis, M Perta, H Queeneville, K R Reidenbach, Y H Rogers, C W Roth, J R Schneider, M Schatz, M Shumway, M Stanke, E O Stinson, J M C Tubio, J P VanZee, S Verjovski-Almeida, D Werner, O White, S Wyder, Q D Zeng, Q Zhao, Y M Zhao, C A Hill, A S Raikhel, M B Soares, D L Knudson, N H Lee, J Galagan, S L Salzberg, I T Paulsen, G Dimopoulos, F H Collins, B Birren, C M Fraser-Liggett, and D W Severson. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 316(5832):1718–1723, June 2007.

82. M Moore, M Sylla, L Goss, M W Burugu, R Sang, L W Kamau, E U Kenya, C Bosio, L Munoz Mde, M Sharakova, and W C Black. Dual African origins of global *Aedes aegypti* s.l. populations revealed by mitochondrial DNA. *PLOS Neglected Tropical Diseases*, 7(4):e2175, 2013.

83. Q M Shi, H D Zhang, G Wang, X X Guo, D Xing, Y D Dong, L Xiao, J Gao, Q M Liu, A J Sun, C X Li, and T Y Zhao. The genetic diversity and population structure of domestic *Aedes aegypti* (Diptera: Culicidae) in Yunnan Province, southwestern China. *Parasites and Vectors*, 10(1):292, June 2017.

84. M K Lawniczak, S J Emrich, A K Holloway, A P Regier, M Olson, B White, S Redmond, L Fulton, E Appelbaum, J Godfrey, C Farmer, A Chinwalla, S P Yang, P Minx, J Nelson, K Kyung, B P Walecz, E Garcia-Hernandez, M Aguiar, L D Viswanathan, Y H Rogers, R L Strausberg, C A Sasaki, D Lawson, F H Collins, F C Kafatos, G K Christophides, S W Clifton, E F Kirkness, and N J Besansky. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, 330(6003):512–514, October 2010.

85. K R Reidenbach, D E Neafsey, C Costantini, N Sagnon, F Simard, G J Ragland, S P Egan, J L Feder, M A Muskavitch, and N J Besansky. Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in West and Central Africa. *Genome Biology and Evolution*, 4(12):1202–1212, 2012.

86. L K Sirot, A Wong, T Chapman, and Mariana F Wolfner. Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. *Cold Spring Harbor Perspectives in Biology*, 7(2):a017533, December 2014.

87. M A Toups and M W Hahn. Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics*, 186(2):763–766, October 2010.

88. B Charlesworth. The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, 6(2):149–162, February 1996.
89. B Charlesworth and D Charlesworth. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355(1403):1563–1572, November 2000.

90. B Vicoso, V B Kaiser, and D Bachtrog. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16):6453–6458, April 2013.

91. O S Akbari, I Antoshechkin, H Amrhein, B Williams, R Diloreto, J Sandler, and B A Hay. The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3: Genes—Genomes—Genetics*, 3(9):1493–1509, September 2013.

92. S Dorus, E C Wilkin, and T L Karr. Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of *Drosophila* sperm. *BMC genomics*, 12:177, April 2011.

93. C Ellison, C Leonard, E Landeen, L Gibilisco, N Phadnis, and D Bachtrog. Rampant cryptic sex chromosome drive in *Drosophila*. *bioRxiv*, 2018.

94. A K Lindholm, K A Dyer, R C Firman, L Fishman, W Forstmeier, L Holman, H Johannesson, U Knief, H Kokko, A M Larracuente, A Manser, C Montchamp-Moreau, V G Petrosyan, A Pomiankowski, D C Presgraves, L D Safronova, A Sutter, R L Uncleess, R L Verspoor, N Wedell, G S Wilkinson, and T A R Price. The ecology and evolutionary dynamics of meiotic drive. *Trends in ecology & evolution*, 31(4):315–326, April 2016.

95. A M Hammond and R Galizi. Gene drives to fight malaria: current state and future directions. *Pathogens and Global Health*, 111(8):412–423, December 2017.

96. P E Visconti. Sperm bioenergetics in a nutshell. *Biology of reproduction*, 87(3):72, September 2012.

97. K Stephens, R A Cardullo, and C D Thaler. *Culex pipiens* sperm motility is initiated by a trypsin-like protease from male accessory glands. *Molecular Reproduction and Development*, 85(5):440–448, May 2018.

98. B Baer, J L Heazlewood, N L Taylor, H Eubel, and A H Millar. The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics*, 9(8):2085–2097, April 2009.

99. J M Rhea, C Wegener, and M Bender. The proprotein convertase encoded by *amontillado* (*amon*) is required in *Drosophila* corpora cardiaca endocrine cells producing the glucose regulatory hormone AKH. *PLoS genetics*, 6(5):e1000967, May 2010.

100. Michael Werner and Leigh W Simmons. Insect sperm motility. *Biological reviews of the Cambridge Philosophical Society*, 83(2):191–208, 2008.

101. D Miller and G C Ostermeier. Spermatozoal RNA: why is it there and what does it do? *Gyncologie, Obstétrique & Fertilité*, 34(9):840–846, September 2006.

102. C C Conine, F Sun, L Song, J A Rivera-Perez, and O J Rando. Small RNAs gained during epididymal transit of sperm are essential for embryonic development in mice. *Developmental Cell*, 46(4):470–480, 2018.
103. B E Fischer, E Wasbrough, L A Meadows, O Randlet, S Dorus, T L Karr, and S Russell. Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proceedings of the Royal Society of London. Series B: Biological Sciences, 279(1738):2636–2644, July 2012.

104. J M Bono, L M Matzkin, E S Kelleher, and T A Markow. Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated Drosophila mojavensis females. Proceedings of the National Academy of Sciences of the United States of America, 108(19):7878–7883, May 2011.

105. R S Lees, J R Gilles, J Hendrichs, M J Vreysen, and K Bourtzis. Back to the future: the sterile insect technique against mosquito disease vectors. Current Opinion in Insect Science, 10:156–162, August 2015.

106. Frank W Avila, K Ravi Ram, Margaret C Bloch Qazi, and Mariana F Wolfner. Sex peptide is required for the efficient release of stored sperm in mated Drosophila females. Genetics, 186(2):595–600, 2010.

107. M E Perotti, F Cattaneo, M E Pasini, F Verni, and J H P Hackstein. Male sterile mutant casanova gives clues to mechanisms of sperm-egg interactions in Drosophila melanogaster. Molecular Reproduction and Development, 60(2):248–259, October 2001.

108. Ethan C Degner and Laura C Harrington. A mosquito sperm’s journey from male ejaculate to egg: mechanisms, molecules, and methods for exploration. Molecular Reproduction and Development, 83(10):897–911, 2016.