

RESEARCH ARTICLE

Open Access

Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling

Matthew D Dean^{1,2*}, Geoffrey D Findlay³, Michael R Hoopmann³, Christine C Wu⁴, Michael J MacCoss³, Willie J Swanson³ and Michael W Nachman²

Abstract

Background: Seminal fluid plays an important role in successful fertilization, but knowledge of the full suite of proteins transferred from males to females during copulation is incomplete. The list of ejaculated proteins remains particularly scant in one of the best-studied mammalian systems, the house mouse (*Mus domesticus*), where artificial ejaculation techniques have proven inadequate. Here we investigate an alternative method for identifying ejaculated proteins, by isotopically labeling females with ¹⁵N and then mating them to unlabeled, vasectomized males. Proteins were then isolated from mated females and identified using mass spectrometry. In addition to gaining insights into possible functions and fates of ejaculated proteins, our study serves as proof of concept that isotopic labeling is a powerful means to study reproductive proteins.

Results: We identified 69 male-derived proteins from the female reproductive tract following copulation. More than a third of all spectra detected mapped to just seven genes known to be structurally important in the formation of the copulatory plug, a hard coagulum that forms shortly after mating. Seminal fluid is significantly enriched for proteins that function in protection from oxidative stress and endopeptidase inhibition. Females, on the other hand, produce endopeptidases in response to mating. The 69 ejaculated proteins evolve significantly more rapidly than other proteins that we previously identified directly from dissection of the male reproductive tract.

Conclusion: Our study attempts to comprehensively identify the proteins transferred from males to females during mating, expanding the application of isotopic labeling to mammalian reproductive genomics. This technique opens the way to the targeted monitoring of the fate of ejaculated proteins as they incubate in the female reproductive tract.

Keywords: seminal fluid, ejaculate, evolution

Background

Successful fertilization occurs through complex interactions among a diversity of proteins that mediate the final fusion of male and female pronuclei. In internally fertilizing species, sperm are accompanied by a non-sperm component of seminal fluid that functions in a variety of contexts. In mammals, this seminal fluid derives from several compartments of the male reproductive tract, the experimental removal of which leads to reductions in fertility success [1,2], smaller litter sizes

[3] and delays in oocyte penetration and embryonic development [4-6]. Seminal fluid also influences sperm motility and physiological status [7-11], suppresses the female immune system [12-14], protects sperm from neutrophil attack in the female reproductive tract [15,16], prepares the uterus for implantation [17], and alters female mating behavior [18,19]. In insects, seminal fluid induces egg laying and proper sperm storage [20-24] and mediates sperm competition outcomes [25-30].

Some properties of ejaculated proteins suggest they may be a source of sexual conflict. In many animal species, including worms [31,32], insects [33], reptiles [34-36], and mammals [37-40], ejaculated proteins

* Correspondence: matthew.dean@usc.edu

¹Molecular and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA, USA

Full list of author information is available at the end of the article

coagulate to form a copulatory plug (also referred to as a mating plug or vaginal plug). By blocking access to the uterus and oviducts, the plug is thought to be an adaptation by which males inhibit the passage of sperm from competitor males, thus protecting their reproductive investment. This hypothesis predicts that the copulatory plug is on average deleterious to females because it inhibits future mate choice. In mice, the copulatory plug is probably effective at inhibiting sperm from other males, because it remains intact for approximately 24 hours, females are truly fertile for about 4-12 hours during the estrus cycle, and sperm are not stored across estrus cycles [40]. Nevertheless, multiple paternity is still common [41,42]. Species which do not form copulatory plugs usually show alternative means of mate-guarding, or have mating ecologies that tend towards monogamy where mate-guarding would be unnecessary [38,40]. However, some apparently monogamous species of rodents like *Peromyscus polionotus*, in which sexual conflict is expected to be less severe, also form a copulatory plug [43].

Additional hypotheses for the function of the copulatory plug include male-female signaling necessary for proper implantation of embryos. For example, copulatory stimulation is necessary to prime the female uterus for implantation [44,45], and the plug may function in this context. The hypothesis that the plug prevents leakage of semen is inconsistent with experiments showing that removal of the plug does not inhibit fertilization, pregnancy, or parturition [46,47]. Similarly, the hypothesis that the plug acts as a reservoir regulating the release of sperm [48] is inconsistent with plug transfer experiments in guinea pigs [47].

A better understanding of the functions of seminal fluid requires a fuller picture of the proteins that are transferred from males to females in the ejaculate. Using house mice (*Mus domesticus*) as a model system, we mated vasectomized males to females that had been metabolically labeled with a heavy isotope of nitrogen, ¹⁵N. We then used mass spectrometry to identify unlabeled, ejaculated proteins directly from the female reproductive tract. We identified 69 ejaculated proteins from female reproductive tracts 6-14 hours post-coitus. Using current functional annotations, we showed that seminal fluid was significantly enriched for genes that participate in two main processes: protection from oxidative stress and endopeptidase inhibition. We also found that more than a third of all identified spectra mapped to just seven proteins known to form the copulatory plug, suggesting a large portion of the ejaculate is dedicated to the formation of this structure. By comparing mated to unmated females, we found that females produced endopeptidases in response to mating. Interestingly, the 69 ejaculated proteins were a non-random

subset of the ~500 proteins that we previously identified directly from dissected regions of the male reproductive tract [49]. The ejaculated proteins we detected here evolved significantly more rapidly than the other male reproductive proteins. These patterns are consistent with the hypothesis that sexual selection has driven the evolutionary dynamics of ejaculated proteins. Future testing of this hypothesis is made possible by the techniques implemented here.

Methods

Mice used

Breeding and genotypes followed Dean et al. [49]. We generated F1 progeny from crosses between two different wild-derived inbred strains of *Mus domesticus* (female LEWES/Eij x male WSB/Eij). F1 mice were then mated with each other to identify proteins transferred during mating. F1 mice were used rather than fully inbred strains to avoid the deleterious effects of inbreeding. We paired parental female LEWES/Eij mice with male WSB/Eij mice for one week, then separated them so the dam gave birth in isolation. At 21 days postpartum, F1 males were weaned individually, and F1 females were weaned in groups. Males were weaned individually because grouped males have comparatively reduced fertility [50], probably due to suppression by dominant males. F1 females labeled with ¹⁵N (see below) were then mated to unlabeled, vasectomized F1 males. All husbandry and experimental manipulations were approved by the University of Arizona Institutional Animal Care and Use Committee.

We measured the size of copulatory plugs in an additional set of mice derived from wild parents trapped more than 100m apart around Tucson, AZ, USA and then crossed in the laboratory. Wild derived F1 males were then mated to a common female genotype (F1 of female LEWES/Eij x male WSB/Eij crosses). In total, copulatory plugs were measured from 149 crosses from 47 different F1 males, derived from 9 wild caught sires and 15 wild caught dams.

Isotopic labeling of females

Artificial ejaculation techniques such as electroejaculation produce abnormal and inconsistent ejaculates in mice [51,52], so we instead employed isotopic labeling to differentiate male- and female-derived proteins [53]. ¹⁵N-enriched diets were prepared by combining ¹⁵N-labeled *Spirulina platensis* (>99 atom percent excess, Spectra Gases Inc., now part of Cambridge Isotope Laboratories, Inc., Andover, MA) with protein-free rodent diet (TD 93328, Harlan, Indianapolis, IN) in a 1:2 (mass:mass) ratio as previously described [54,55]. The two food types were ground into a homogenous powder with a mortar and pestle and worked into a

dough by slowly adding water (roughly 5-6 ml water/30 grams powder mixture). The dough was formed into 1.5 cm³ pellets and placed in a food dehydrator set at 54°C until completely dry.

Three-week-old females were weaned from their mothers and immediately given ¹⁵N-enriched diet. In contrast, all males used in this experiment were fed regular diet. Female proteins will have a shifted mass as a result of incorporation of ¹⁵N. To gauge the effectiveness of our labeling strategy, we analyzed two non-reproductive tissues from a mated female: the liver, an organ with a relatively high rate of protein turnover, and the brain, which has a low rate of protein turnover. Under unlabeled search conditions, we identified five proteins from the liver and 103 proteins from the brain. These data confirmed that 15N labeling more effectively inhibited identification of female-derived proteins in tissues with faster protein turnover. As discussed below (*Analyzing an unmated female*), the low number of unlabeled proteins identified from the unmated female reproductive tract appears more similar to the high turnover liver tissue, suggesting that our labeling strategy was effective in masking female-derived proteins to enable detection of ejaculated proteins.

Vasectomy of males

Males approximately eight weeks of age were anesthetized with 2.5% avertin, then vasectomized using standard techniques [56]. We used vasectomized males because we were interested in the seminal fluid proteins and wanted to exclude the sperm proteome, which is complex [57-60]. Males of this genotype are sexually mature by eight weeks of age [61]. Cuts were closed using surgical clips and males were checked several times a day to monitor recovery. One week after vasectomy, clips were removed. One week following clip removal, males were mated to tester females that had been induced to ovulate using standard techniques [56,62]. These test matings confirmed libido and the absence of sperm in dissected female reproductive tracts. Males were mated to tester females in consecutive weeks; vasectomized males were mated to at least three tester females prior to mating with labeled females. In total, two vasectomized males were analyzed in the present study.

Mating and collection of samples

After three to four weeks of feeding on ¹⁵N chow, labeled females were induced to ovulate using standard techniques [56,62]. Immediately following administration of the hormone hCG, labeled females were paired with vasectomized males. Between 12 and 20 hours after initial pairing (likely to be 6-14 hours after mating), females were sacrificed and reproductive tracts were

removed. Internal fluids were stripped from both uteri and immediately frozen at -80°C, as were the copulatory plug, the remaining reproductive tract, the brain, and the liver. As a control, we collected a reproductive tract, brain, and liver from a labeled female that was exposed to a male but had not mated. In total, proteins from two mated females and one unmated female were analyzed with mass spectrometry.

Protein preparation and mass spectrometry

As a result of labeling, female-derived proteins were expected to have upward-shifted masses, making it possible to distinguish male- and female-derived proteins sampled from mated female reproductive tracts. Samples were generally prepared and analyzed by mass spectrometry as previously described [49,53] with some modifications. Tissue samples (dissected female reproductive tracts, liver, brain) were homogenized in 50 mM ammonium bicarbonate. The homogenate was centrifuged at 20,800 g for 5 min, and the soluble fraction was retained. Soluble proteins were quantified with a BCA assay (Thermo) and then mixed with PPS detergent (Protein Discoveries) to a final concentration of 0.1% PPS. Proteins were denatured, reduced and alkylated as described previously [63] and then digested with trypsin. PPS was hydrolyzed by the addition of HCl to a final concentration of 200 mM. Copulatory plugs were processed by placing slices of plug in 50 mM ammonium bicarbonate with 0.1% PPS and then sonicating 10 times with a probe sonicator, alternating 45 seconds of sonication with 45 seconds of ice incubation. Plug samples were then boiled for 2 min and homogenized with a pestle homogenizer. A few seconds of microcentrifugation removed remaining large pieces of solid plug, and the remaining, cloudy supernatant was then reduced, alkylated and trypsin digested as above.

Tryptic peptides of all samples were separated using 75-μm internal diameter fused silica HPLC columns packed with 35 cm of Jupiter C12 (4 μm, 90 Å; Phenomenex) reversed phase material. These columns were placed on-line with a LTQ-FT Ultra mass spectrometer (Thermo), and peptides were eluted over a 3-hour gradient. For each sample analyzed, we ran 5-7 technical replicates, each loading ~5 μg protein onto the column. Except as described below ("Accurate mass-directed tandem mass spectrometry"), mass spectra were obtained using data-dependent acquisition. We focused on four biological samples - two different copulatory plugs and two different uterine fluid samples isolated from two different matings - for analyses of reproductive proteins (Additional File 1).

In making protein identifications from the collected MS data, we purposely set our identification criteria to have a high false negative and low false positive rate to

lend confidence to protein identifications. MS2 files from each experiment were searched against two databases using the SEQUEST algorithm [64]: one database contained all proteins from the NCBI build 37 mouse genome, while the other contained randomly shuffled protein sequences representing decoy proteins. Results from these searches were analyzed with the PERCOLATOR program [65,66] to improve discrimination between correct and incorrect peptide-spectrum matches and to set a per-spectrum false discovery rate (FDR) of 0.01. However, previous research has shown that with a per-spectrum FDR of 1%, the peptide and the protein-level FDR can be much higher [8-11%, depending on the search algorithm used, 67]. Most of these false positive protein identifications were presumably those proteins identified with a single peptide. Thus, to consider a protein identified in this study, we required it to have been matched by at least two peptides, at least one of which was a unique match to a single region in the genome.

Normalized Spectral Abundance Factor (NSAF)

It is difficult to relate spectral counts to protein abundance because not all peptides within proteins are equally identifiable [68]. The acquisition of tandem mass spectrometry data is a semi-random process and is highly dependent on the presence of co-eluting molecular species. Signal suppression during electrospray ionization can potentially alter the mass spectrometry signal response within complex mixtures. Longer proteins may be more detectable simply because they are more likely to contain tryptic and ionizable peptides. Post-translational modifications such as glycosylation may further hinder identification of unmodified proteins.

Nevertheless, more abundant proteins should have a greater number of spectra mapping to their sequence compared to low abundance proteins [69,70]. As a rough proxy of relative protein abundance, we calculated the normalized spectral abundance factor (NSAF) [69,70], with some slight modifications. Here, we calculated a single experiment-wide NSAF for each gene by summing all spectral counts across the four main biological samples (two copulatory plugs, two uterine fluid samples), dividing this sum by the protein length, then dividing by the sum of this value across all genes. NSAF therefore ranges from 0 to 1 for each protein (actual observed range = 10^{-5} to 0.21, median = 0.002) and sums to 1 across all 69 identified proteins. Relatively high NSAF may indicate higher abundance in the sample, though the caveats discussed above suggest cautious interpretation.

for genes that encoded multiple alternative transcripts, we divided by the median transcript length; our results did not change if instead we divided by the shortest, the longest, or a randomly chosen transcript length. Our

results also did not change if we calculated NSAF separately for each of the four biological samples; we present the experiment-wide NSAF for simplicity. For spectra that mapped to more than one region of the genome, we divided the number of spectra by the number of regions it mapped to, adding the result to each gene's spectral count. However, as described above, a gene was only considered present if at least two different peptides mapped to it, at least one of which was a unique hit to that gene product.

For comparison, we re-analyzed the proteins identified from dissected regions of the male reproductive tract [49]. We calculated NSAF as described above, summing spectral count across the six distinct regions of the male reproductive tract sampled.

Evaluating Detection Sensitivity

Three targeted searches provided support that we identified most detectable ejaculated proteins. These three methods of evaluating detection sensitivity suggested that additional technical and/or biological replicates would not have yielded a substantially larger list of ejaculated proteins under the experimental conditions employed here.

Isolating insoluble proteins

In an attempt to detect male-derived proteins that could be bound to the female epithelium, we ran five technical replicates on the insoluble fraction of one of the mated female's reproductive tract. We isolated insoluble proteins by resuspending the pellet from centrifugation in 0.5% PPS and then sonicating twice with a probe sonicator.

Depletion of highly abundant proteins

In an attempt to unmask less abundant proteins, we re-analyzed one of the copulatory plug samples and one of the uterine fluid samples after depleting each of them of highly abundant immunoglobulin- and albumin-like proteins. We used the ProteoPrep ImmunoAffinity Albumin and IgG Depletion Kit (Sigma) to reduce levels of albumin and IgG proteins.

Accurate mass-directed tandem mass spectrometry

We also used an analytical method to direct the mass spectrometer to specifically fragment male-derived peptides that had not been previously sampled in a prior technical replicate [71]. We re-analyzed one of the plug samples and one of the uterine fluid samples, first running one technical replicate using data-dependent acquisition. We then used the HARDKLÖR algorithm [71,72] to identify peaks from MS1 signals that were predicted to come from a peptide with a natural abundance isotope distribution (i.e., an unlabeled male peptide). We constructed a list of these peptides' m/z (\pm 10 ppm) and elution times (\pm 1.5 min) off of the HPLC column and used this list to direct the mass spectrometer's peptide sampling for two subsequent technical replicates. If

no peptides on the list were detected at a given elution time, the instrument used standard data-dependent acquisition to sample peptides from that MS1 scan. Finally, we compared the number of proteins and peptides identified from three technical replicates that used this method to the number identified by three standard, data-dependent technical replicates performed on the same samples.

Testing for functional overrepresentation

We took two approaches to identify important functions in ejaculated proteins. First, we tested for statistical enrichment of genes with particular Gene Ontology functional annotations [73], using ONTOLOGIZER version 2.0 [74], with the "Term-for-Term" calculation method and Bonferroni-corrected $P < 0.05$. Among the 69 ejaculated proteins, 68 could be linked to Gene Ontology data. Second, we qualitatively examined genes to look for commonality of function among proteins with high NSAF.

Analyzing female-derived proteins

Analyzing an unmated female

As a negative control, we attempted to identify unlabeled proteins from a female that had undergone ^{15}N labeling for three weeks and was paired with a male for approximately 20 hours, but where copulation did not take place, as confirmed by the absence of a copulatory plug. In theory, we should not identify unlabeled proteins unless i) certain proteins fail to incorporate ^{15}N , for example proteins with low rates of turnover, or ii) the male mounted and transferred some proteins without true ejaculation. We identified two large hemoglobin families, an actin family, and *SVS4* when searching mass spectra from this virgin female's reproductive tract under the assumption of naturally occurring isotope distributions. The hemoglobin and actin families could plausibly be explained by their apparently high abundance - by chance we may have sampled a few relatively unlabeled peptides. Identification of *SVS4*, from five spectra derived from two uniquely mapping peptides, was surprising because this is a quintessential seminal vesicle secretion that is derived from the male reproductive tract. It is possible that mounting without ejaculation occurred and some male proteins were transferred at a low level. Notably, unlabeled *SVS4* was identified with roughly two orders of magnitude more spectra from mated females, suggesting the *SVS4* identified in the virgin female was an anomaly and that this is truly a male-transferred protein.

Labeled protein searches

Although this experiment was specifically designed to identify ejaculated proteins, we also identified female-derived proteins that could be induced from mating. We performed SEQUEST searches in which we adjusted the

search parameters to find proteins that were labeled with 95% ^{15}N incorporation. Specifically, we altered the SEQUEST search parameters such that the expected molecular mass of each amino acid was increased by (0.95 Daltons) \times (the number of nitrogen atoms in the amino acid), which corresponds to an expected 95% labeling. We analyzed the two copulatory plug samples in this manner. Because the SEQUEST algorithm allows some deviation between the theoretical mass of a peptide and the mass observed by the mass spectrometer, assuming an additional mass of 0.95 Daltons/nitrogen atom would not necessarily preclude identification of labeled proteins with similar levels of ^{15}N incorporation (e.g., 92% labeled peptides may still be identified).

Estimating evolutionary rate and adaptive evolution

We analyzed pairwise d_N/d_S estimates of all genes in the genome that have one-to-one orthologs between mouse and rat, taken from Dean et al. [49]. Briefly, all orthology assignments and sequences were downloaded from Ensembl version 48, NCBI mouse build 37 (<http://www.ensembl.org>). Protein sequences were aligned using CLUSTALW version 1.83 [75], associated with their coding DNA sequences using REVTRANS version 1.5 [76], and d_N/d_S estimated using the method of Goldman and Yang [77] as implemented in PAML version 3.15 [78]. We removed any genes with fewer than 100 aligned codons, an estimated $d_N > 1$, or an estimated $d_S \geq 0.381$ as quality control measures [details in 49]. We analyzed the full genome in this manner.

Tests for recurrent positive selection were also taken from Dean et al. [49], who analyzed evolutionary rates across five species with the phylogeny of ((mouse, rat), human, (dog, cow)). Briefly, a gene was considered to have experienced a history of recurrent adaptive evolution if five criteria were met: 1) the data fit the M8 model significantly better than M7 at $P < 0.01$ [79], 2) the data fit the M8 model significantly better than M8a at $P < 0.01$ [80], 3) the additional class of d_N/d_S estimated by M8 was greater than 1.1, 4) at least 1% of the codons belonged to this additional class of d_N/d_S , and 5) Fixed Effect Likelihood (FEL) analyses [81] revealed significant evidence of positive selection in at least one codon [$d_N/d_S > 1.1$ at $P < 0.10$, the p-value recommended by 82]. As a quality control measure, we excluded any genes whose pairwise d_S exceeded twice the genome median across any of the pairwise combinations of species [details in 49]. We analyzed the full genome in this manner.

Results

Identification of ejaculated proteins from the female reproductive tract

We directly identified ejaculated proteins from four biological samples: the two copulatory plugs and two

samples of the uterine fluids, from two different male-female matings. The costs associated with isotopic labeling inhibited additional sampling. We considered a gene to be positively identified if at least two different peptides mapped to it, at least one of which mapped uniquely to a single location in the genome. With these criteria, we identified 69 genes total (Additional File 1) from 27,565 spectra representing 827 different peptides, 795 of which mapped to a single location in the genome. Each gene was identified with a median of 80 spectra, seven different peptides (a median six of which mapped uniquely to that gene), at a median coverage of 21.4% of the protein. The median number of spectra per gene is ~ four times lower than the mean number of spectra per gene (= 399 spectra), indicating that a relatively few genes were identified with a high number of spectra. Genome duplications and high relatedness among certain gene families prevented some gene identifications because associated peptides did not map to a single genomic location. These ambiguous gene identifications are not considered further here but are presented in Additional File 2.

Evaluating detection sensitivity suggests most detectable proteins were identified

Technical replication verified that most detectable proteins were identified under our experimental conditions. The two uterine samples were each run through five technical replicates, and the two plugs were each run through seven technical replicates. Only four additional proteins were identified in the sixth and seventh plug replicates combined (Figure 1). Furthermore, proteins identified for the first time in later technical replicates showed lower median NSAF (Figure 2), suggesting most proteins that were reasonably abundant (and detectable) had been sampled.

Three targeted searches provided additional evidence that we identified most detectable ejaculated proteins. First, we isolated insoluble proteins from the female reproductive tract. In this insoluble fraction, we identified an additional six proteins that were not identified in any other samples (*POU domain class 4 transcription factor 1, elastin, DEAH box polypeptide 9, AT rich interactive domain 1B, histone cluster 1 H1e, and tubulin beta 2c*, identified with 2, 2, 3, 4, 8, and 26 spectra, respectively). Second, we re-analyzed one of the copulatory plug samples and one of the uterine fluid samples after depleting each of them of immunoglobulin- and albumin-like proteins, which were highly represented in early technical replicates. Only four additional proteins were newly detected (*major urinary protein 4, transferrin, aldolase 1 A isoform, and cathepsin L*, identified with 2, 2, 3, and 7 spectra, respectively) in depleted samples. Third, we re-ran several experiments after directing

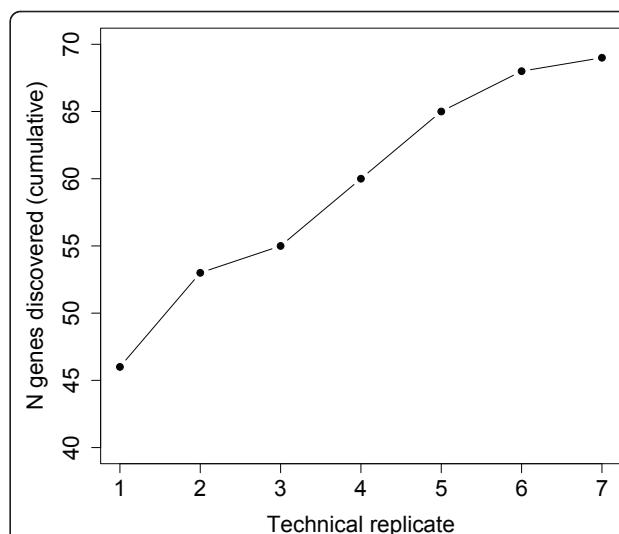


Figure 1 A cumulative distribution showing new genes discovered across technical replicates. The sixth and seventh technical replicate added a combined total of four new genes (out of 69 total), suggesting we have approached an asymptote of new gene discovery.

the mass spectrometer to only fragment peptides that had previously gone unanalyzed [72]. This directed sampling method had a minimal effect. A median of only 2 additional spectra were detected per gene for the copulatory plug sample, out of a total of 13,299 spectra used to identify 62 genes. For the uterine fluid sample, a

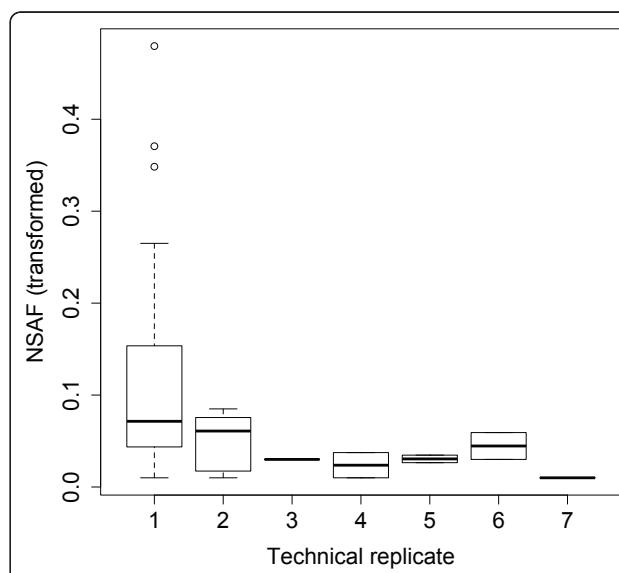


Figure 2 Genes discovered in the first replicate had higher NSAF (arcsin square root transformed) than genes discovered in later replicates. This pattern was seen in all four biological samples; we present one of the copulatory plug samples here. This result suggests that we have identified all reasonably abundant (and detectable) proteins under the experimental conditions employed.

median of 7 fewer spectra were detected per gene, out of a total of 9,725 spectra mapping to 50 genes. In sum, our evaluations of detection sensitivity provided support that we have identified the major ejaculated proteins present in the female, at least given the experimental conditions employed here.

Ejaculated proteins were statistically enriched for genes that protect from oxidative stress and inhibit endopeptidases

Two main branches in the Gene Ontology were significantly overrepresented among the 69 ejaculated proteins compared to the entire genome: *antioxidant activity* and *endopeptidase inhibitor activity*. Both functions were overrepresented among genes identified directly from male reproductive tissues [49]. Both functions were also overrepresented in human ejaculates, revealing commonalities among mammalian ejaculate function [49, their supplementary table 4, 83]. Five ejaculated proteins had *antioxidant activity* (compared to 57 of 14,720 annotated genes across the genome, Bonferroni-corrected $P < 0.01$). Six ejaculated proteins showed evidence of *endopeptidase inhibitor activity* (vs. 148/14,720 in the genome, Bonferroni-corrected $P < 0.02$).

Most spectra map to proteins associated with the copulatory plug

A large proportion of the proteins detected were associated with the copulatory plug. Of the 69 genes identified, 62 were found in the copulatory plug samples. It is thought that the copulatory plug forms via the action of the prostate-derived *transglutaminase 4*, which cross-links proteins of at least six seminal vesicle secretions - *SVS1*, *SVS2*, *SVS3a*, *SVS3b*, *SVS4*, and *SVS5* [84-87]. In total, these seven proteins were identified with 10,239 spectra, accounting for 37% of all identifiable spectra generated across the four biological samples (two copulatory plugs, two uterine fluid samples), in spite of the fact that their combined length accounted for only 8% of the combined length of all proteins identified.

To further explore the investment that males make in copulatory plugs, we made 149 crosses from 47 different F1 males derived from wild caught parents. These crosses using wild-caught mice were only used to assess natural variation in the weight of the copulatory plug; all other data in this manuscript were derived from F1 (male WSB/Eij x female LEWES/Eij) matings as described above. Approximately 12 hours after mating, the copulatory plug weighed a median 31 mg, which represented approximately 0.3% of the body weight of the females from which these plugs were collected. We corrected by female weight as a rough proxy for the size of the vaginal-cervix canal, which may constrain the size of the plug. By comparison,

a single testis from the male mice that formed these plugs accounts for a median 0.5% of its body mass, suggesting the plug represents a significant investment for males.

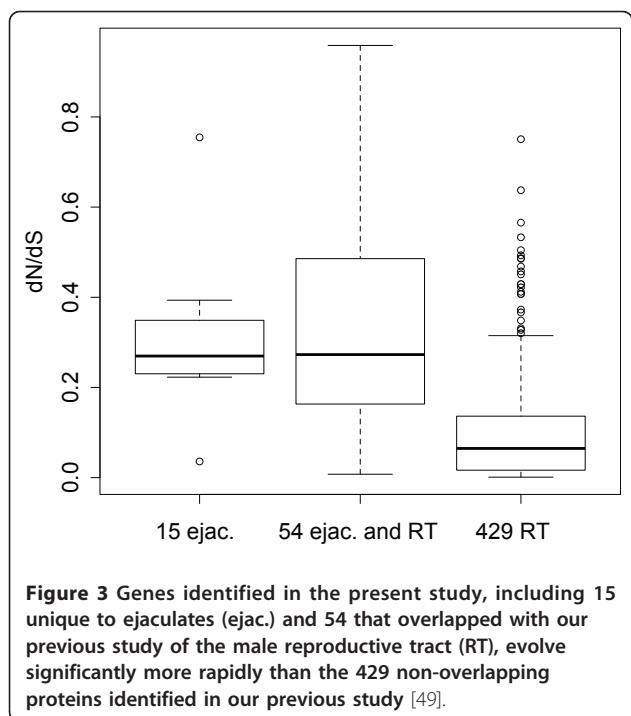
Female-derived proteins

To demonstrate another potential application of the differential labeling method, we identified ^{15}N -labeled (presumably female-derived) proteins by computationally adjusting the SEQUEST search algorithm to assume 95% ^{15}N incorporation into peptides. Three additional criteria facilitated identification of female-derived proteins that were indeed produced in response to mating. We required female-derived proteins to i) have a secretion signal at $P > 0.90$, as predicted by TargetP [88], ii) not be identified from an unmated ^{15}N -labeled female reproductive tract, and iii) not be identified as a male-derived seminal fluid gene. Using these criteria, we identified six female-derived proteins produced in response to mating - *lactotransferrin* (54 spectra, 14 peptides), *kallikrein-related peptidase 14* (14 spectra, 3 peptides), *lipocalin 2* (32 spectra, 2 peptides), *chloride channel calcium activated 3* (65 spectra, 15 peptides), *corneodesmosin* (, and *alpha-2-HS-glycoprotein* (6 spectra, 2 peptides). Two of these proteins (*lactotransferrin* and *kallikrein-related peptidase 14*) included domains indicative of endopeptidases [89-91], which are proteins that cleave other proteins.

The 69 ejaculated proteins identified were a non-random subset of proteins produced in the male reproductive tract

Previously [49], we identified 506 proteins from six distinct regions of the reproductive tract - seminal vesicles, anterior prostate (a.k.a. the coagulating gland), ventral prostate, dorsolateral prostate, bulbourethral diverticulum, and the bulbourethral gland (a.k.a. Cowper's gland) - from the same genotype analyzed here (an F1 male derived from a cross between a male WSB/Eij and a female LEWES/Eij). We re-analyzed those data with the same criteria presented above, producing a list of 483 total single-region proteins (Additional File 3). We found that 54 genes overlapped between the two studies, while 429 genes that were detected in our previous study of the male reproductive tract were not identified here. For simplicity, we refer to these as the 429 "non-overlapping" proteins. If we required only a single uniquely mapping peptide (rather than requiring at least two peptides mapped, at least one of which was unique), we still only observed 72 of the 483 previously identified proteins.

The 54 overlapping genes evolved significantly more rapidly than the 429 non-overlapping genes (Figure 3). Of the 54 overlapping genes, 29 had a one-to-one ortholog in rat and produced estimates of evolutionary rate



that satisfied various measures of quality control (see Materials and Methods). The median d_N/d_S for these 29 genes ($d_N/d_S = 0.27$, Q1-Q3 = 0.16-0.49) was significantly higher than the median estimated d_N/d_S for the 429 non-overlapping genes ($N = 303$ of 429 non-overlapping genes with quality one-to-one orthologs, median $d_N/d_S = 0.06$, Q1-Q3 = 0.02-0.14) (Wilcoxon Rank Sum Test [WRST] $W = 7,336$, $P < 10^{-8}$) (Figure 2). In addition to these sequence-based metrics, the 54 overlapping genes had fewer one-to-one orthologs between mouse and rat compared to the non-overlapping genes (29/54 vs. 303/429, respectively[<http://www.ensembl.org>, version 48], Fisher's Exact Test $P < 0.02$). This result suggests these genes are evolving so rapidly that orthology is difficult to detect, that they undergo more gene conversion which obscures orthology, and/or that they experience higher rates of gene birth and death.

These patterns of rapid evolution derived from mouse-rat comparisons were robust to the precise set of non-overlapping genes investigated. All patterns remained statistically significant even if we compared the 54 overlapping genes to the 88 (of 429) non-overlapping genes that i) have a one-to-one ortholog found in human ejaculates [83], and ii) have a one-to-one ortholog in rat. These additional comparisons represented an attempt to control for possible protein contamination, and to focus on those proteins that show the most evidence of being ejaculated [following, 49].

Unfortunately, we cannot perform deeper evolutionary analyses for most of these genes because orthology

across the five mammalian genomes analyzed here (mouse, rat, dog, human, cow) is lacking. It is possible that rapid evolution has obscured orthology assignment. Similar patterns have been observed in insects [92]. Of the 54 overlapping proteins, only 15 have orthologs across the five species, which is a significantly smaller proportion than the 216 (of 429) non-overlapping proteins that have orthologs across the five species (FET, $P = 0.001$). Of the 15 overlapping proteins with orthologs, two showed statistically significant evidence of adaptive evolution according to the five criteria above (*tissue inhibitor of metalloproteinase 1* and *plasminogen activator urokinase*), which was not significantly different than the 17 adaptively evolving genes identified from the 216 non-overlapping proteins with orthologs (FET, $P = 0.36$). Attempts to gain power by analyzing more closely related genomes of rabbit, guinea pig, kangaroo rat, and squirrel (<http://www.ensembl.org>) were inconclusive due to the low coverages of these additional genomes (data not presented).

Of the 69 ejaculated proteins detected in the present study, 15 were not observed in our previous analysis of the male reproductive tract (Figure 3). These proteins may derive from regions of the male reproductive tract not sampled in our previous study, for example the ampullary gland, a small swelling in the vas deferens. It is also possible some of these 15 proteins were more easily detected after ejaculation into the female reproductive tract. These 15 proteins evolved at a rate similar to the 54 overlapping proteins (Figure 3).

Rapid evolution of female-derived endopeptidases, male-derived endopeptidase inhibitors, and copulatory plug genes

Female-derived endopeptidases and male-derived endopeptidase inhibitors evolve relatively rapidly, although our study is underpowered given the low number of genes in both categories. In pairwise mouse-rat estimates, the female-derived endopeptidases *lactotransferrin* and *kallikrein related peptidase 14* showed a d_N/d_S of 0.78 and 0.32, respectively, values that are substantially higher than the genome of median 0.13. Furthermore, *lactotransferrin* showed statistically significant evidence of recurrent positive selection across a phylogeny of five mammalian species (according to five criteria discussed previously, 1: the data fit the M8 model significantly better than M7 [$2\Delta L = 29.8$, $P < 10^{-6}$], 2: the data fit the M8 model significantly better than M8a [$2\Delta L = 22.9$, $P < 10^{-5}$], 3: the additional class of d_N/d_S estimated by M8 = 3.8, 4: an estimated 4.9% of codons belonged to this additional class, and 5: FEL analyses estimated that 2% of codons experienced $d_N/d_S > 1.1$ at $P < 0.10$). Only three male-derived endopeptidase inhibitors - *cystatin C*, *spink5*, and *timp1* - had high quality

orthologs between mouse and rat, but all three showed high d_N/d_S of 0.41, 0.49, and 0.52, respectively. *Timp1* showed statistically significant evidence of recurrent adaptive evolution across the five mammalian species (1: $2\Delta L = 9.81$, $P < 0.01$, 2: $2\Delta L = 4.82$, $P < 0.03$, 3: additional class of $d_N/d_S = 2.9$, 4: estimated 4.9% of codons belonged to this class, and 5: FEL estimated 1.4% of codons with $d_N/d_S > 1.1$), *spink5* did not, and *cystatin C* could not be analyzed due to a lack of orthology. Rapid evolution of female-derived endopeptidases and male-derived endopeptidase inhibitors is consistent with a model of sexual conflict between these two molecular classes [93,94], though additional functional experiments are required to evaluate this hypothesis further.

Proteins involved in the formation of the copulatory plug showed especially rapid evolution. Four genes known to form a large proportion of the copulatory plug - *SVS1*, *SVS2*, *SVS5*, and *Tgm4* (the other *SVS* genes drop out of pairwise mouse-rat comparisons due to either lack of orthology or failed quality control) - have d_N/d_S estimates of 0.36, 0.40, 0.67, and 0.33, respectively, which are approximately three or more times the genome median (0.13).

Discussion

A major finding over the past ~15 years is that male reproductive proteins diverge rapidly in sequence [reviewed by 95], gene birth/death processes [96-99], expression [100-103], and protein size or composition [104-107]. Adaptive evolution of copulatory plug proteins is especially strong in species with relatively high levels of polyandry [106,108-110]. In primates, copulatory plug proteins also show signs of rapid evolution [111,112], and the solidification intensity of the plug is positively correlated with the level of sperm competition [39]. In *Drosophila*, both male- and female-derived proteases have undergone rampant duplication, gene conversion, and/or adaptive evolution [93,113-115]. There are several hypotheses to account for this elevated rate of divergence, including adaptive evolution related to natural selection and/or intra- or inter-sexual selection. Disentangling these alternative hypotheses requires a better understanding of the function of ejaculated proteins. Here we used isotopic labeling to separate female- from male-derived proteins taken from the female reproductive tract, identifying 69 proteins that are transferred during mating.

Two functions - *antioxidant activity* and *endopeptidase inhibitor activity* - were significantly enriched among the 69 identified proteins. Sperm are particularly susceptible to oxidative stress as a result of their high metabolic rate, their high level of polyunsaturated fatty acids in their membranes, and their lack of most cytoplasmic components of the antioxidant system.

Oxidative stress can damage the paternal genome, leading to aberrant embryonic development [116]. Male hamsters that had their accessory glands surgically removed ejaculated sperm with elevated DNA damage compared to sham-operated controls [117]. In humans, sub-fertile men had a higher level of reactive oxygen species and lower antioxidant ability in their seminal fluid, compared to normally fertile men [118]. In some birds, more colorful males harbor sperm that are more resistant to oxidative stress, raising the possibility that males advertise their ability to protect sperm [119].

Male seminal fluid was also significantly enriched for proteins with *endopeptidase inhibitor* activity. Such proteins are involved in a diversity of physiological functions including modulation of immune response and sperm capacitation. Dean et al. [49] hypothesized that endopeptidase inhibitors may protect the copulatory plug from degradation.

On the female side of the equation, two of the six identified female-derived genes, *lactotransferrin* and *kallikrein related-peptidase 14*, included domains indicative of endopeptidases. One possible function for female-derived endopeptidases is the degradation of the copulatory plug [49]. While there is some reference in the literature to the plug "falling out" or being easily dislodged by females or other males [56], in our extensive experience with wild-derived mice (like those of the present study), the plug is strongly attached to the tissues of the vagina and cervix, rarely visible externally, and requires considerable effort to dissect. Female-derived endopeptidases might degrade the plug and/or detach the plug from its close association to female tissue as an initial step in dislodgement.

Female-derived endopeptidases might be targeted by male-derived endopeptidase inhibitors. Of the six male-derived endopeptidase inhibitors identified above, three were characterized as I4 subfamily members and two as I1 subfamily members [the sixth is not characterized, merops.sanger.ac.uk 120]. Members of subfamily I1 are known to inhibit endopeptidases of the S1 family [121], like the female-derived *kallikrein related peptidase 14* that we identified here. The other female-derived endopeptidase that we identified, *lactotransferrin*, is part of the S60 family of endopeptidases, which is not known to be inhibited by any of the male-derived endopeptidase inhibitors identified here [120]. More direct experiments are needed to test whether female-derived endopeptidases and male-derived endopeptidase inhibitors interact directly.

Curiously, an additional 429 proteins previously identified in the male reproductive tract by Dean et al. [49] were not observed here. We consider three hypotheses to explain why we did not identify these 429 non-overlapping proteins in this study. One hypothesis is these 429 non-overlapping proteins were not ejaculated. Our

earlier work was based on tissue dissection and may therefore have included some contamination by non-ejaculated proteins. This hypothesis seems unlikely to be the main explanation because 327 of the 429 non-overlapping proteins had a one-to-one ortholog in humans, and of those, 114 were detected in human ejaculates [83]. We note that the general findings in either study were not altered if we confined analyses to those genes that had a one-to-one ortholog to a human-ejaculated gene.

A second hypothesis is that even though female proteins were labeled with heavy nitrogen, their presence still reduced the signal-to-noise ratio at various stages throughout the mass spectrometry pipeline employed here. This hypothesis also seems unlikely because technical replication (Figures 1,2) as well as three independent targeted searches (see *Evaluating Detection Sensitivity* in Results) all suggested we have identified most detectable proteins. Because we used the same mass spectrometry techniques in both studies and the same mouse genotype, the 429 non-overlapping proteins should have been detected if present, unless they were post-translationally modified in ways that make them undetectable only after ejaculation. Other technical artifacts associated with mass spectrometry, such as random loss of signal due to precise composition of co-eluting molecular species, predict a random subset of genes would be identified in our heavy isotope framework, which was not observed here.

A third hypothesis is that many of the 429 non-overlapping proteins were degraded in the female reproductive tract after ejaculation but prior to our sampling of female reproductive tracts. Wild-derived mice demonstrate complicated mating behaviors, so sampling female reproductive tracts immediately after ejaculation is difficult. Thus, for these initial experiments, female reproductive tracts were sampled 6-14 hours after copulation. During this interval, changes in the number and relative abundance of male proteins may have occurred. Consistent with this hypothesis, females produced endopeptidases in response to mating, which may actively degrade ejaculated proteins. Under this scenario, male proteins might be under selection to evolve rapidly, thus evading female degradation machinery. The 69 ejaculated proteins indeed evolved significantly more rapidly than other male reproductive proteins.

Conclusion

We applied isotopic labeling to directly identify 69 proteins transferred from males to females during mating. The techniques applied here make it possible to study the fate of ejaculated proteins over time. Future experiments can use targeted proteomic methods to follow *in vivo* the localization and degradation of specific male proteins in the female reproductive tract, to more fully appreciate their roles in reproduction and evolutionary fitness.

Additional material

Additional file 1: Male-derived genes detected in the female reproductive tract. The 69 genes that code proteins transferred from males to females.

Additional file 2: Ambiguous male-derived genes. 30 genes that were only identified with ambiguously mapping spectra from the female reproductive tract.

Additional file 3: Genes detected from the male reproductive tract. 483 genes identified from dissected regions of the male reproductive tract [a re-analysis of 49].

Acknowledgements

Debbie Stead and Stephanie Munger (U. Arizona) taught vasectomy techniques. Daniela Tomazela (U. Washington) provided guidance in using the ImmunoAffinity depletion kit, and Jan Aagaard (U. Washington) provided advice on protein preparation procedures. This research was supported by NIH fellowship F32GM070246-02 (MDD), and NSF and NIH grants (MWN).

Author details

¹Molecular and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA, USA. ²Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. ³Department of Genome Sciences, University of Washington, Seattle, WA, USA. ⁴Department of Cell Biology, University of Pittsburgh, Pittsburgh, PA, USA.

Authors' contributions

MDD conceived of the study, designed and performed the crossing experiments, analyzed all data, and wrote the manuscript. GDF helped conceive the study, designed and performed protein isolation and mass spectrometry experiments, and helped write the manuscript. MRH modified algorithms to evaluate peptide labeling and perform directed peptide sampling, and helped perform targeted mass spectrometry experiments. CCW provided heavy nitrogen chow and helped design experiments. MJM helped design experiments and oversaw all mass spectrometry experiments. WJS contributed to experimental design and data interpretation. MWN designed experiments, interpreted data, and contributed to the overall study. All authors have read and approved the final manuscript.

Received: 2 April 2011 Accepted: 10 June 2011 Published: 10 June 2011

References

1. Peitz B, Olds-Clarke P: Effects of seminal vesicle removal on fertility and uterine sperm motility in the house mouse. *Biol Reprod* 1986, 35:608-617.
2. Queen K, Dhabuwala CB, Pierrepont CG: The effect of the removal of the various accessory sex glands on the fertility of male rats. *J Reprod Fertil* 1981, 62:423-426.
3. Pang SF, Chow PH, Wong TM: The role of the seminal vesicles, coagulating glands and prostate glands on the fertility and fecundity of mice. *J Reprod Fertil* 1979, 56:129-132.
4. Henault MA, Killian GJ, Kavanaugh JF, Griel LC Jr: Effect of accessory sex gland fluid from bulls of differing fertilities on the ability of cauda epididymal sperm to penetrate zona-free bovine oocytes. *Biol Reprod* 1995, 52:390-397.
5. O WS, Chen HQ, Chow PH: Effects of male accessory sex gland secretions on early embryonic development in the golden hamster. *J Reprod Fertil* 1988, 84:341-344.
6. Carballada R, Esponda P: Effect of antibodies against seminal vesicle secretion on fertility in the rat. *Zygote* 1999, 7:223-231.
7. Kawano N, Yoshida M: Semen-coagulating protein, SVS2, in mouse seminal plasma controls sperm fertility. *Biol Reprod* 2007, 76:353-361.
8. Huang YH, Chu ST, Chen YH: A seminal vesicle autoantigen of mouse is able to suppress sperm capacitation-related events stimulated by serum albumin. *Biol Reprod* 2000, 63:1562-1566.
9. Peitz B: Effects of seminal vesicle fluid components on sperm motility in the house mouse. *J Reprod Fertil* 1988, 83:169-176.

10. Ignotz GG, Lo MC, Perez CL, Gwathmey TM, Suarez SS: Characterization of a fucose-binding protein from bull sperm and seminal plasma that may be responsible for formation of the oviductal sperm reservoir. *Biol Reprod* 2001, **64**:1806-1811.
11. Agrawal Y, Vanha-Perttula T: Effect of secretory particles in bovine seminal vesicle secretion on sperm motility and acrosome reaction. *J Reprod Fertil* 1987, **79**:409-419.
12. Anderson DJ, Tarter TH: Immunosuppressive effects of mouse seminal plasma components in vivo and in vitro. *J Immunol* 1982, **128**:535-539.
13. Peitz B, Bennett D: Inhibition of complement-mediated cytotoxicity of antisera by fluid secreted by the seminal vesicle of the house mouse. *J Reprod Immunol* 1981, **3**:109-116.
14. Thaler CJ: Immunological role for seminal plasma in insemination and pregnancy. *Am J Reprod Immunol* 1989, **21**:147-150.
15. Wartha F, Beiter K, Normark S, Henriques-Normark B: Neutrophil extracellular traps: casting the NET over pathogenesis. *Curr Opin Microbiol* 2007, **10**:52-56.
16. Alghamdi AS, Foster DN: Seminal DNase frees spermatozoa entangled in neutrophil extracellular traps. *Biol Reprod* 2005, **73**:1174-1181.
17. Robertson SA: Seminal fluid signaling in the female reproductive tract: lessons from rodents and pigs. *J Anim Sci* 2007, **85**:E36-44.
18. Carter CS, Schein MW: Sexual receptivity and exhaustion in the female golden hamster. *Horm Behav* 1971, **4**:191-200.
19. Goldfoot DA, Goy RW: Abbreviation of behavioral estrus in guinea pigs by coital and vagino-cervical stimulation. *J Comp Physiol Psychol* 1970, **72**:426-434.
20. Heifetz Y, Vandenberg LN, Cohn HI, Wolfner MF: Two cleavage products of the Drosophila accessory gland protein ovulin can independently induce ovulation. *Proc Natl Acad Sci USA* 2005, **102**:743-748.
21. Heifetz Y, Tram U, Wolfner MF: Male contributions to egg production: the role of accessory gland products and sperm in *Drosophila melanogaster*. *Proc R Soc Lond B Biol Sci* 2001, **268**:175-180.
22. Herndon LA, Wolfner MF: A Drosophila seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc Natl Acad Sci USA* 1995, **92**:10114-10118.
23. Ravi Ram K, Wolfner MF: Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr Comp Biol* 2007, **47**:427-445.
24. Wong A, Albright SN, Giebel JD, Ram KR, Ji S, Fiumera AC, Wolfner MF: A Role for Acp29AB, a Predicted Seminal Fluid Lectin, in Female Sperm Storage in *Drosophila melanogaster*. *Genetics* 2008, **180**:921-931.
25. Chapman T, Neubaum DM, Wolfner MF, Partridge L: The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. *Proc R Soc Lond B Biol Sci* 2000, **267**:1097-1105.
26. Harshman LG, Prout T: Sperm displacement without sperm transfer in *Drosophila melanogaster*. *Evolution* 1994, **48**:758-766.
27. Fiumera AC, Dumont BL, Clark AG: Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. *Genetics* 2007, **176**:1245-1260.
28. Fiumera AC, Dumont BL, Clark AG: Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. *Genetics* 2005, **169**:243-257.
29. Clark AG, Aguadé M, Prout T, Harshman LG, Langley CH: Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* 1995, **139**:189-201.
30. Price CSC, Dyer KA, Coyne JA: Sperm competition between *Drosophila* males involves both displacement and incapacitation. *Nature* 1999, **400**:449-452.
31. Palopoli MF, Rockman MV, TinMaung A, Ramsay C, Curwen S, Aduna A, Laurita J, Kruglyak L: Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. *Nature* 2008, **454**:1019-1022.
32. O'Brien SJ, Berman EJ, Estes JD, Gardner MB: Murine retroviral restriction genes Fv-4 and Akvr-1 are alleles of a single locus. *J Virol* 1983, **47**:649-651.
33. Rogers DW, Baldini F, Battaglia F, Panico M, Dell A, Morris HR, Catteruccia F: Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. *PLoS Biol* 2009, **7**:e1000272.
34. Devine MC: Copulatory plugs, restricted mating opportunities and reproductive competition among male garter snakes. *Nature* 1977, **267**:345-346.
35. Devine MC: Copulatory plugs in snakes: enforced chastity. *Science* 1975, **187**:844-845.
36. Moreira PL, Birkhead TR: Copulatory plug displacement and prolonged copulation in the Iberian rock lizard (*Lacerta monticola*). *Behav Ecol Sociobiol* 2004, **56**:290-297.
37. Dewsbury DA: Sperm competition in muroid rodents. In *Sperm competition and the evolution of animal mating systems*. Edited by: Smith RL. New York: Academic Press; 1984:547-571.
38. Hartung TG, Dewsbury DA: A comparative analysis of copulatory plugs in muroid rodents and their relationship to copulatory behavior. *J Mammal* 1978, **59**:717-723.
39. Dixson AF, Anderson MJ: Sexual selection, seminal coagulation and copulatory plug formation in primates. *Folia Primatologica* 2002, **73**:63-69.
40. Voss R: Male accessory glands and the evolution of copulatory plugs in rodents. *Occasional Papers of the Museum of Zoology, University of Michigan* 1979, **689**:1-27.
41. Firman RC, Simmons LW: The frequency of multiple paternity predicts variation in testes size among island populations of house mice. *J Evol Biol* 2008.
42. Dean MD, Ardlie KG, Nachman MW: The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). *Mol Ecol* 2006, **15**:4141-4151.
43. Baumgardner DJ, Hartung TG, Sawyer DK, Webster DG, Dewsbury DA: Muroid copulatory plugs and female reproductive tracts: a comparative investigation. *J Mammal* 1982, **63**:110-117.
44. Land RB, McGill TE: The effects of the mating pattern of the mouse on the formation of corpora lutea. *J Reprod Fertil* 1967, **13**:121-125.
45. Diamond M: Intromission pattern and species vaginal code in relation to induction of pseudopregnancy. *Science* 1970, **169**:995-997.
46. Firman RC, Simmons LW: Polyandry, sperm competition, and reproductive success in mice. *Behavioral Ecology* 2008, **19**:695-702.
47. Martan J, Shepherd BA: The role of the copulatory plug in reproduction of the guinea pig. *J Exp Zool* 1976, **196**:79-83.
48. Asdell SA: *Patterns of mammalian reproduction* Ithaca, New York: Comstock Publishing Company; 1946.
49. Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW: Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol Biol Evol* 2009, **26**:1733-1743.
50. Snyder RL: Fertility and reproductive performance of grouped male mice. In *Comparative aspects of reproductive failure*. Edited by: Benirschke K. New York: Springer-Verlag; 1967:458-472.
51. Snyder RL: Collection of mouse semen by electroejaculation. *Anat Rec* 1966, **155**:11-14.
52. Tecirlioglu RT, Hayes ES, Trounson AO: Semen collection from mice: electroejaculation. *Reprod Fertil Dev* 2002, **14**:363-371.
53. Findlay GD, Yi X, MacCoss MJ, Swanson WJ: Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* 2008, **6**:e178.
54. McClatchy DB, Dong MQ, Wu CC, Venable JD, Yates JR: 15N metabolic labeling of mammalian tissue with slow protein turnover. *J Proteome Res* 2007, **6**:2005-2010.
55. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR: Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* 2004, **76**:4951-4959.
56. Nagy A, Gertsenstein M, Vintersten K, Behringer R: *Manipulating the mouse embryo*. 3 edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2003.
57. Baker MA, Hetherington L, Reeves GM, Aitken RJ: The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. *Proteomics* 2008, **8**:1720-1730.
58. Stein KK, Go JC, Lane WS, Primakoff P, Myles DG: Proteomic analysis of sperm regions that mediate sperm-egg interactions. *Proteomics* 2006, **6**:3533-3543.
59. Cao W, Gerton GL, Moss SB: Proteomic profiling of accessory structures from the mouse sperm flagellum. *Mol Cell Proteomics* 2006, **5**:801-810.
60. Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL: Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. *Mol Biol Evol* 2010, **27**:1235-1246.

61. Good JM, Handel MA, Nachman MW: Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* 2008, **62**:50-65.
62. Dean MD, Nachman MW: Faster fertilization rate in conspecific versus heterospecific matings in house mice. *Evolution* 2009, **63**:20-28.
63. Aagaard JE, Yi X, MacCoss MJ, Swanson WJ: Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. *Proc Natl Acad Sci USA* 2006, **103**:17302-17307.
64. Eng JK, McCormack AL, Yates JR: An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994, **5**:976-989.
65. Spivak M, Weston J, Bottou L, Kall L, Noble WS: Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. *J Proteome Res* 2009, **8**:3737-3745.
66. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ: Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007, **4**:923-925.
67. Balgley BM, Laudean T, Yang L, Song T, Lee CS: Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics* 2007, **6**:1599-1608.
68. Good DM, Coon JJ: Advancing proteomics with ion/ion chemistry. *BioTechniques* 2006, **40**:783-789.
69. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, Washburn MP: Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods (San Diego, Calif)* 2006, **40**:303-311.
70. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP: Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* 2006, **103**:18928-18933.
71. Hoopmann MR, Merrihew GE, von Haller PD, MacCoss MJ: Post analysis data acquisition for the iterative MS/MS sampling of proteomics mixtures. *J Proteome Res* 2009, **8**:1870-1875.
72. Hoopmann MR, Finney GL, MacCoss MJ: High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Anal Chem* 2007, **79**:5620-5632.
73. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000, **25**:25-29.
74. Robinson PN, Wollstein A, Bohme U, Beattie B: Ontologizing gene-expression microarray data: characterizing clusters with Gene Ontology. *Bioinformatics (Oxford, England)* 2004, **20**:979-981.
75. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994, **22**:4673-4680.
76. Wernersson R, Pedersen AG: RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* 2003, **31**:3537-3539.
77. Goldman N, Yang Z: A codon-based model of nucleotide substitution for protein-coding DNA sequences: a maximum likelihood approach. *J Mol Evol* 1994, **40**:725-736.
78. Yang Z: PAML, a program package for phylogenetic analysis by maximum likelihood. *CABIOS* 1997, **13**:555-556.
79. Yang Z, Nielsen R, Goldman N, Pedersen A-MK: Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, **155**:431-449.
80. Swanson WJ, Nielsen R, Yang Q: Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 2003, **20**:18-20.
81. Kosakovsky Pond SL, Frost SDW, Muse SV: HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England)* 2005, **21**:676-679.
82. Kosakovsky Pond SL, Frost SDW: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005, **22**:1208-1222.
83. Pilch B, Mann M: Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome biology* 2006, **7**:R40.
84. Lundwall Å, Peter A, Lovgren J, Lilja H, Malm J: Chemical characterization of the predominant proteins secreted by mouse seminal vesicles. *Eur J Biochem* 1997, **249**:39-44.
85. Lin H-J, Luo C-W, Chen Y-H: Localization of the transglutaminase cross-linking site in SVS III, a novel glycoprotein secreted from mouse seminal vesicle. *J Biol Chem* 2002, **277**:3632-3639.
86. Porta R, Esposito C, Gentile V, Marinello L, Peluso G, Metafora S: Transglutaminase-catalyzed modifications of SV-IV, a major protein secreted from the rat seminal vesicle epithelium. *Int J Pept Protein Res* 1990, **35**:117-122.
87. Fawell SE, Higgins SJ: Formation of rat copulatory plug: purified seminal vesicle secretory proteins serve as transglutaminase substrates. *Mol Cell Endocrinol* 1987, **53**:149-152.
88. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 2000, **300**:1005-1016.
89. Olsson AY, Lundwall A: Organization and evolution of the glandular kallikrein locus in *Mus musculus*. *Biochem Biophys Res Commun* 2002, **299**:305-311.
90. Olsson AY, Lilja H, Lundwall A: Taxon-specific evolution of glandular kallikrein genes and identification of a progenitor of prostate-specific antigen. *Genomics* 2004, **84**:147-156.
91. Cunningham GA, Headon DR, Conneely OM: Structural organization of the mouse lactoferrin gene. *Biochem Biophys Res Commun* 1992, **189**:1725-1731.
92. Walters J, Harrison R: EST analysis of male accessory glands from Heliconius butterflies with divergent mating systems. *BMC Genomics* 2008, **9**:592.
93. Lawniczak MKN, Begun DJ: Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. *Mol Biol Evol* 2007, **24**:1944-1951.
94. Kelleher ES, Pennington JE: Protease gene duplication and proteolytic activity in *Drosophila* female reproductive tracts. *Mol Biol Evol* 2009, **26**:2125-2134.
95. Clark NL, Aagaard JE, Swanson WJ: Evolution of reproductive proteins from animals and plants. *Reprod Fertil Dev* 2006, **131**:11-22.
96. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, et al: Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004, **428**:493-521.
97. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al: Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002, **420**:520-562.
98. Torgerson DG, Singh RS: Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha 4 proteasome subunits in *Drosophila*. *Genetics* 2004, **168**:1421-1432.
99. Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ: Duplication and selection on abalone sperm lysin in an allopatric population. *Mol Biol Evol* 2007, **24**:2081-2090.
100. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM: Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol* 2004, **21**:1308-1317.
101. Meiklejohn CD, Parsch J, Ranz JM, Hartl DL: Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 2003, **100**:9894-9899.
102. Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B: Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Res* 2006.
103. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S: Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 2005, **1850**-1854.
104. Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, Tomich JM, Howard DJ: Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. *Mol Biol Evol* 2011, **28**:423-435.
105. Coulthart MB, Singh RS: High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. *Mol Biol Evol* 1988, **5**:182-191.
106. Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P: Comparative proteomics reveals evidence for evolutionary diversification of rodent

- seminal fluid and its functional significance in sperm competition. *Mol Biol Evol* 2009, 26:189-198.
107. Civetta A, Singh RS: High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *J Mol Evol* 1995, 41:1085-1095.
108. Ramm SA, Parker GA, Stockley P: Sperm competition and the evolution of male reproductive anatomy in rodents. *Proc R Soc B* 2005, 272:949-955.
109. Karr RC, Clark NL, Nguyen ED, Swanson WJ: Adaptive evolution in rodent seminal vesicle secretion proteins. *Mol Biol Evol* 2008, 25:2301-2310.
110. Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD: Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol Biol Evol* 2008, 25:207-219.
111. Clark NL, Swanson WJ: Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* 2005, 1:e35.
112. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT: Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. *Nat Genet* 2004, 36:1326-1329.
113. Wong A, Turchin MC, Wolfner MF, Aquadro CF: Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. *Mol Biol Evol* 2008, 25:497-506.
114. Kelleher ES, Markow TA: Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* 2009, 181:1451-1465.
115. Kelleher ES, Clark NL, Markow TA: Diversity-enhancing selection acts on a female reproductive protease family in four subspecies of *Drosophila mojavensis*. *Genetics* 2011, 187:865-876.
116. Aitken RJ, Baker MA: Oxidative stress, sperm survival and fertility control. *Mol Cell Endocrinol* 2006, 250:66-69.
117. O W-s, Chen H, Chow PH: Male genital tract antioxidant enzymes - their ability to preserve sperm DNA integrity. *Mol Cell Endocrinol* 2006, 250:80-83.
118. Sharma RK, Pasqualotto FF, Nelson DR, Thomas AJ Jr, Agarwal A: The reactive oxygen species total antioxidant capacity score is a new measure of oxidative stress to predict male infertility. *Hum Reprod* 1999, 14:2801-2807.
119. Helfenstein F, Losdat S, Møller AP, Blount JD, Richner H: Sperm of colourful males are better protected against oxidative stress. *Ecol Lett* 2010, 13:213-222.
120. Rawlings ND, Barrett AJ, Bateman A: MEROPS: the peptidase database. *Nucleic Acids Res* 2010, 38:D227-233.
121. Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PGW, Irving JA, Lomas DA, Luke CJ, Moyer RW, et al: The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. *J Biol Chem* 2001, 276:33293-33296.

doi:10.1186/1471-2164-12-306

Cite this article as: Dean et al.: Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* 2011 12:306.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

