

RESEARCH ARTICLE

Open Access

# Using RNA sequencing to characterize female reproductive genes between Z and E Strains of European Corn Borer moth (*Ostrinia nubilalis*)

Nooria Al-Wathiqui\*, Sara M Lewis and Erik B Dopman\*

## Abstract

**Background:** Reproductive proteins often evolve rapidly and are thought to be subject to strong sexual selection, and thus may play a key role in reproductive isolation and species divergence. However, our knowledge of reproductive proteins has been largely limited to males and model organisms with sequenced genomes. With advances in sequencing technology, Lepidoptera are emerging models for studies of sexual selection and speciation. By profiling the transcriptomes of the bursa copulatrix and bursal gland from females of two incipient species of moth, we characterize reproductive genes expressed in the primary reproductive tissues of female Lepidoptera and identify candidate genes contributing to a one-way gametic incompatibility between Z and E strains of the European corn borer (*Ostrinia nubilalis*).

**Results:** Using RNA sequencing we identified transcripts from ~37,000 and ~36,000 loci that were expressed in the bursa copulatrix or the bursal gland respectively. Of bursa copulatrix genes, 8% were significantly differentially expressed compared to the female thorax, and those that were up-regulated or specific to the bursa copulatrix showed functional biases toward muscle activity and/or organization. In the bursal gland, 9% of genes were differentially expressed compared to the thorax, with many showing reproduction or gamete production functions. Of up-regulated bursal gland genes, 46% contained a transmembrane region and 16% possessed secretion signal peptides. Divergently expressed genes in the bursa copulatrix were exclusively biased toward protease-like functions and 51 proteases or protease inhibitors were divergently expressed overall.

**Conclusions:** This is the first comprehensive characterization of female reproductive genes in any lepidopteran system. The transcriptome of the bursa copulatrix supports its role as a muscular sac that is the primary site for disruption of the male ejaculate. We find that the bursal gland acts as a reproductive secretory body that might also interact with male ejaculate. In addition, differential expression of proteases between strains supports a potential role for these tissues in contributing to reproductive isolation. Our study provides new insight into how male ejaculate is processed by female Lepidoptera, and paves the way for future work on interactions between post-mating sexual selection and speciation.

**Keywords:** Female reproductive genes, Speciation, Next-generation sequencing, Lepidoptera

\* Correspondence: Nooria.Al\_wathiqui@tufts.edu; Erik.Dopman@tufts.edu  
Department of Biology, Tufts University, Medford, MA 02155, USA

## Background

Sexual selection is a powerful evolutionary force that can drive species divergence [1-4]. Although many studies focus on how organisms choose mates during pre-mating sexual selection, the process is not limited to courtship, but rather occurs across multiple time points before, during, and after copulation [5-7]. Post-mating sexual selection has proven challenging to study because it can involve interactions between the female reproductive tract and the male ejaculate on a molecular level. Such interactions include male-male sperm competition, sexual conflict involving male and female proteins, and cryptic female choice. Sexual conflict is of particular interest because it arises from the divergent reproductive interests of males and females and thus may represent an important component of post-mating interactions [8]. Many mechanisms of post-mating sexual selection involve co-evolutionary arms races between the sexes, a process that can lead to rapid trait evolution within [8,9] or divergence between populations [10-12]. Differences in reproductive traits between populations can quickly result in post-mating, pre-zygotic barriers [13], potentially playing a powerful role in species formation.

Most work on post-mating, pre-zygotic barriers has focused on interactions between the female reproductive tract and male sperm. Specifically, cryptic female choice has been identified as a possible mechanism for conspecific sperm precedence, in which multiply mated females produce more offspring sired by conspecific rather than heterospecific mates [14-16]. Conspecific sperm precedence is widespread and has been demonstrated in many insect species, including fruit flies, ground crickets and flour beetles [14,16-19]. Of several possible mechanisms underlying conspecific sperm precedence in *Drosophila mauritiana* and *Drosophila simulans*, biased sperm use by females was found to be a key determinant [20]. Female *Drosophila* are able to favor conspecific males by preferentially storing sperm in separate storage organs [20]. Although multiply mated females seem to be able to bias paternity of their offspring, the nature of the interactions between male and female reproductive proteins that might lead to such differential sperm use is unclear [21,22].

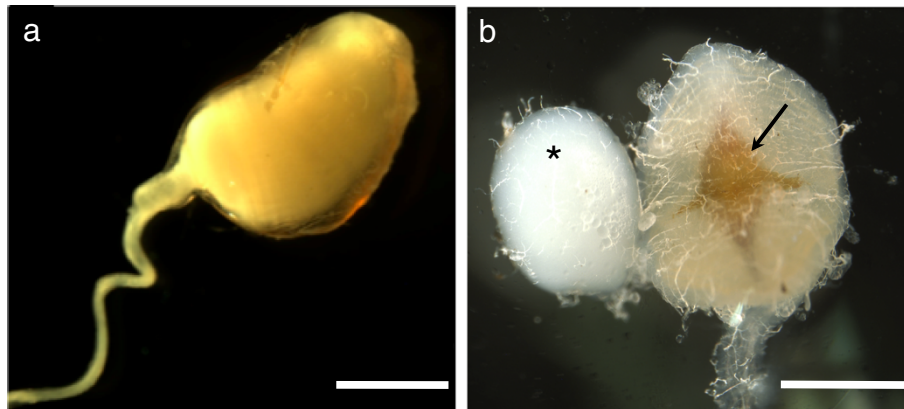
After mating, the female reproductive tract interacts not only with male sperm, but also with seminal fluid. In many taxa, male reproductive proteins are produced in accessory glands or the ejaculatory duct, which are then transferred to the female as components of the male ejaculate. Collectively, these non-sperm components of the ejaculate are called seminal fluid proteins and they are quite numerous [23]. In fruit flies, *Drosophila melanogaster* males produce over 100 seminal proteins that are transferred to females during mating [24-26]. These proteins have profound effects on female behavior and physiology,

including changes in lifespan, ovulation, feeding habits and sperm storage patterns [14,24,26]. Not only do male reproductive proteins have important effects on females, they are potentially powerful drivers of post-mating, pre-zygotic reproductive isolation because many of them evolve rapidly [22]. Seminal fluid proteins have been comprehensively characterized in several insect taxa, including fruit flies, mosquitoes, honeybees, crickets, flour beetles, butterflies, and bedbugs [24,27-35]. In contrast, we know very little about the many possible interacting female reproductive proteins for any one species.

Although the reproductive tracts of female insects also contain secretory tissue [36], to date female reproductive genes have been comprehensively studied in very few taxa including mosquitos, fruit flies, and honeybees [37-44]. Unsurprisingly, female reproductive genes have been best characterized in *Drosophila* species including: *D. melanogaster*, *D. simulans*, *D. arizonae* and *D. mojavensis* [38-43]. Many of these investigations have identified proteases and protease inhibitors, as well as genes related to muscle activity, immune response, and energy metabolism in female reproductive tracts [38-43]. Genes with these functions are predicted to mediate interactions with male ejaculate after mating. Indeed, muscle activity is a key component of female-mediated sperm storage and ejaculate processing [45,46], while proteases and protease inhibitors have been shown to be required for activation of ovulation-inducing seminal fluid proteins in *D. melanogaster* [47]. Furthermore, immune and energy metabolism genes appear to be important for the demands of egg production and oviposition or to protect females from male-introduced pathogens [38,40,48]. Many relevant female reproductive proteins are secreted from female tissue or are transmembrane, as these are likely to directly interact with male ejaculate or act as receptors for male seminal fluid proteins [42].

To reach a comprehensive understanding of the mechanisms by which male and female reproductive genes drive post-mating sexual selection and ultimately species divergence, studies of reproductive proteins must consider organisms with different mating systems and different levels of sexual conflict. As the second largest insect order with ~170,000 known species, moths and butterflies comprising the order Lepidoptera are ideal for study because the degree of multiple mating by females, and thus the opportunity for sexual conflict, positively correlates with speciation rate [49,50]. However, very little is known about male and female reproductive genes in Lepidoptera. Two comprehensive studies, both in Heliconid butterflies, have identified male reproductive proteins in lepidopterans [34,35]; however, researchers have yet to identify female genes from any structure in the lepidopteran female reproductive tract.

Most male lepidopterans transfer their ejaculate in a package called a spermatophore (Figure 1a) [36,49].



**Figure 1** Male and female ECB reproductive structures. **a.** An ECB male spermatophore transferred to a female upon mating. **b.** The virgin female bursa copulatrix with signum (indicated by arrow) and the bursal gland (indicated by\*). Notice that even in virgins the bursal gland is filled with fluid. Scale bars represent 1 mm.

Although produced by males, the spermatophore is actually formed inside a large, sac-like structure inside the female called the bursa copulatrix (Figure 1b). Inside the bursa copulatrix, the spermatophore is broken open by the signum, a chitinized structure embedded in the muscular wall of the bursa copulatrix, and both sperm and male seminal fluid proteins are released into the female reproductive tract; however, the spermatophore remains in the bursa copulatrix as a visible structure for the entirety of the females life [36,49]. Thus, the bursa copulatrix represents a strong candidate arena for the resolution of sexual conflict and the origin of post-mating, pre-zygotic isolation. In general, males that prevent females from remating will achieve greater paternity success. In species where this is true, male traits will evolve that delay female remating. For example, males that transfer larger spermatophores are able to delay female remating for longer in some lepidopterans [51,52] and in *D. melanogaster*, identified male reproductive proteins act to reduce female receptivity to future mates [47,53]. On the other hand, females in many taxa gain material and genetic benefits from multiply mating [54,55], and therefore selection will favor morphological and biochemical traits that allow females to rapidly process male spermatophores. Although recent microstructural studies in Lepidoptera suggest the bursa copulatrix could have a secretory function, studies have yet to characterize any secretions from the structure [56]. Such secretions could be important for breaking down spermatophores or for interacting with male reproductive proteins.

A second structure found in some lepidopteran females that could mediate within- and between-species mating success is the bursal gland. Although patterns of evolutionary conservation remain unclear, the bursal gland is a prominent anatomical feature of the female reproductive system in the European corn borer moth, *Ostrinia nubilalis*

[57]. The bursal gland is dorsally connected directly to the bursa copulatrix by a short duct (Figure 1b) and is approximately 0.5-0.8 mm in diameter. In virgin females the bursal gland is filled with a white, translucent fluid, which then flows into the bursa copulatrix under pressure [57]. After mating, the bursal gland is similarly filled, but with an opaque fluid [57]. The function of this gland is currently unknown, however its proximity and direct connection to the bursa copulatrix and male spermatophore suggests that the bursal gland could function during spermatophore breakdown or to secrete female reproductive proteins regulating the activity of male reproductive proteins.

Here, we use next-generation RNA sequencing to characterize gene expression in the female bursa copulatrix and bursal gland as the first step towards determining how these tissues are involved in post-mating, pre-zygotic isolation in the European corn borer moth (hereafter, "ECB"). The Z and E strains of ECB are emerging textbook models for the study of speciation [13], in which the two incipient species split approximately 75,000 to 150,000 years ago through the evolution of manifold reproductive barriers [58,59]. Females of both strains mate multiply [57] and suffer reduced longevity after mating [60], conditions that are generally favorable for sexual conflict and the evolution of post-mating, pre-zygotic isolation. Consistent with this notion, one of seven barriers between strains, accounting for a ~30% reduction in gene flow, stems from reduced lifetime fecundity following between-strain mating. This post-mating, pre-zygotic incompatibility is asymmetric: Z-strain females that have mated with E-strain males lay significantly fewer eggs over their lifetime [59]. However, the mechanism underlying this gametic isolation is unknown [59]. By examining the transcriptome of bursa copulatrix and bursal gland reproductive tissues within and between ECB strains, we

characterize candidate genes that could be contributing to the egg-laying dysfunction. Specifically, we characterize the function of the bursa copulatrix and bursal gland and identify female reproductive genes that may be involved in isolation using the following criteria: (1) putative proteins that are secreted or membrane bound, (2) an up-regulation of transcripts that aid in muscle contraction and that may assist in spermatophore breakdown or sperm transfer to storage, (3) an up-regulation of proteases and protease inhibitors that could mediate male seminal fluid protein potential, and (4) differential expression between Z and E strains.

## Methods

### Sample preparation and sequencing

We collected bursa copulatrix and bursal gland tissues from 2-day old adult Z- and E-strain ECB females ( $n = 12$  per strain, Figure 1). At this stage females are reproductively mature [60,61]. As the goal of this study was to identify the reproductive function of these two structures and characterize differences between ECB strains and not to identify genes directly affected by mating, all female tissues were collected from virgins. The following dissections were done in RNeasy (Qiagen, California). First, females were sacrificed and the bursa copulatrix and attached bursal gland were removed from an incision in the female abdomen. Next, fat body was removed from both structures and, after separating the bursal gland from the bursa copulatrix, both tissues were stored in RNeasy at  $-80^{\circ}\text{C}$ . After tissue collection, total RNA was extracted from bursa copulatrix and bursal gland tissues using an RNeasy Midi kit (Qiagen, California). Bursal glands and bursa copulatrix tissues from four females were pooled by strain into three separate samples for each tissue type and each strain prior to an initial tissue homogenization step. This resulted in twelve samples, three bursal gland samples and three bursa copulatrix samples for each strain. RNA quantities were assessed using a Nanodrop and  $1\ \mu\text{g}$  of total RNA from each sample was used to create cDNA libraries (Illumina Truseq RNA sample preparation kit v2, San Diego, CA). To prepare samples for sequencing, mRNA was selected from each sample using poly-T-tail magnetic beads. Next, cDNA was synthesized using Superscript II (Invitrogen, Grand Island, NY) and Illumina adapters were attached to libraries for multiplexing prior to sequencing. cDNA strands were then amplified using 15 PCR cycles. Next, quality and quantity of cDNA was confirmed using a NanoDrop (Thermo Fisher Scientific Inc., Delaware) and an Agilent 2100 Bioanalyzer (Santa Clara, CA). Due to low sample quality for one Z-strain bursa copulatrix library, we did not sequence this sample, which left us with two Z-strain bursa copulatrix tissue libraries. Four Z- and E- bursal gland and bursa copulatrix cDNA libraries were multiplexed and sequenced in each of three lanes of two

Illumina flow cells. Libraries were sequenced using an Illumina HiSeq2000 with 50 bp single-end reads. To evaluate tissue-biased expression, we took advantage of previously sequenced Z- and E- female thorax libraries developed during a separate project. Z-strain and E-strain female thoraxes were collected from 2 females each. Briefly, female thorax cDNA libraries were created using the SMART cDNA library 6.7 protocol (Takara Bio, Otsu, Shiga, Japan). These libraries were sequenced using an Illumina GA IIx and  $40\times 40$  bp paired-end reads with a 200 bp insert length.

### Data preprocessing

After sequencing, all reads were subjected to quality control and trimming using Trimmomatic v0.17 to remove Illumina sequencing adapters and low quality reads [62]. Leading and trailing bases with a quality score  $< 5$  were trimmed from reads and then each read was trimmed by a sliding window with a width of 4 bp and minimum average quality of 15. After adapter and quality trimming, only reads  $\geq 36$  bp were retained.

Although we used magnetic beads to select for mRNA, our samples still contained small amounts of mtDNA and rRNA sequences. To remove these contaminants we used the short read aligner Bowtie 2 [63]. Bowtie2 uses the Burrow-Wheeler transformation to index a reference, then searches the index until it finds an alignment for a specific read [63]. We aligned our RNAseq reads to the complete ECB mitochondrial genome (NC\_003367.1), and all published ECB ribosomal sequences [AF336303.1, AF077013.1, DQ988989.1, AB568463.1, AY513653.1, JX683305.1, JX683313.1, AB568278.1, AB568276.1, AB568274.1, EU532443.1, EU532441.1, EU532439.1, EU53244.1, EU532442.1, EU532440.1, EU532438.1, AF349036.1] in the NCBI database using default parameters and removed these sequences. Identical reads were then collapsed using FastX toolkit to reduced library complexity and decrease the computational needs for transcriptome assembly [64].

### De novo sequence assembly

We used the Trinity program suite to assemble all 13 tissue libraries including female bursa copulatrix and bursal gland libraries, as well as the two thorax libraries into a single assembly [65]. Trinity uses the inchworm, chrysalis, and butterfly software modules to create a de novo assembly. First, inchworm assembles reads into unique sequences. Chrysalis then clusters sequences into contiguous sequences and a de Bruijn graph is created for each cluster of contiguous sequences. Lastly, butterfly uses the de Bruijn graphs to construct transcripts. For all steps, we used settings recommended by developers, including merging the assembly at a kmer of 25 [65,66]. Finally, following de novo assembly, we



selected the longest transcript at each locus to eliminate redundancy.

#### Annotation

To annotate our assembled transcriptome, we used the program Blast2Go [67,68]. First, putative homologs were identified by performing a blastx search of the entire NCBI non-redundant protein database (e-value cutoff  $10^{-3}$ ). For all sequences with significant blast hits, four different mappings were conducted. First, BLAST accession numbers are used to find gene names and symbols from NCBI gene\_info and gene2accession. Then, gene\_info identifiers are used to retrieve UniProt IDs using PSD, UniProt, SwissProt, TrEMBL, RefSeq, GenPept and PDB databases. In the final two steps, BLAST accessions were searched in the dbxref table and the gene product table of the GO database. Finally, Blast2go computed an annotation score for all possible GO terms for each sequence [67,68].

#### Differential expression analysis

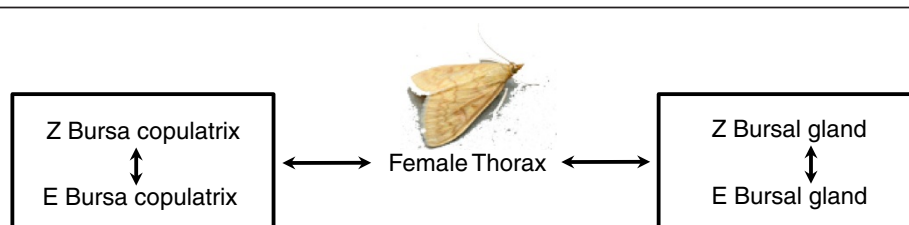
To identify differentially expressed sequences, we first mapped our reads back to our assembled transcriptome using Bowtie 2. For read mapping, we used the 'very sensitive' setting in Bowtie 2 because preliminary trials indicated that this setting resulted in the most uniquely aligned reads. Differences between the SMART and TRUseq library preparation protocols could potentially lead to biases related to GC content, read length, and sequencing depth in each library. To help control for these and other possible biases, we then normalized our libraries prior to differential expression analysis using the programs EDAsseq [69]. EDAsseq performs within-lane normalization to account for differences in gene length and GC content and between-lane normalization to account for differences in sequencing depth [69]. Within-lane normalization uses global scaling normalization, which separates genes into equally sized bins based on GC-content and then matches different parameters of the count distribution across bins. For between-lane normalization, EDAsseq uses a full-quantile normalization procedure that forces equal library sizes across lane.

After normalized read counts were obtained, we used the R package, edgeR to identify differentially expressed genes for all comparisons of interest using the normalized counts for each library (Figure 2) [70-73]. edgeR uses empirical Bayes methods to estimate gene-specific variation. As we were interested in four comparisons in particular, we used a generalized linear model approach in which we assessed differential expression with strain and tissue type as factors (Figure 2). Our model did not include an interaction term. Finally, a GLM likelihood ratio test was used to identify differentially expressed genes [70]. Genes were considered differentially expressed if they had a false-discovery rate (FDR) of  $<0.01$ .

#### Characterizing bursa copulatrix and bursal gland function

We used a three-pronged approach to characterize the reproductive function of the bursa copulatrix and the bursal gland. First, we adopted a common method to determine the specific functions of these tissues by ignoring housekeeping genes that have similar expression profiles across reproductive and non-reproductive tissues. For all of the remaining transcripts with significant expression differences between the bursa copulatrix and thorax, or between the bursal gland and thorax, gene annotations were pulled and enriched/depleted gene ontology categories were identified using a two-tailed fisher's exact test in Blast2go with a term filter cutoff of  $FDR \leq 0.05$ . Our entire non-redundant transcriptome containing transcripts from all three tissue types was used as the null distribution of GO categories.

Second, we identified signal peptides and transmembrane helices from the bursa copulatrix and bursal gland non-redundant transcriptome. For the purpose of identifying secreted and transmembrane proteins in the bursa copulatrix and bursal gland, we used a tblastx to remove all thorax sequences from our transcriptome. Next, to estimate predicted protein sequence from female bursa and bursal gland RNA-seq libraries, we used ESTscan [74], in which biases in hexanucleotide usage in coding versus non-coding regions and a Hidden Markov Model are used to predict protein-coding sequences. Subsequently, we identified sequences containing a secretion



**Figure 2 Differential expression analysis tissue comparisons.** Our generalized linear model used to calculate differentially expressed genes between E- and Z- strain ECB included the following comparisons: female thorax versus the female bursa copulatrix, female thorax versus bursal gland, Z-strain versus E-strain bursa copulatrix, and Z-strain versus E-strain bursal glands.

motif using SignalP 4.0 [75], which uses a neural network-based method to identify signal peptides. We then used TMHMM 2.0 to identify sequences with transmembrane helices [76], in which a Hidden Markov Model is used to predict integral membrane proteins.

Our last step to characterize the function of the bursa copulatrix and the bursal gland was to identify putative ECB homologs of female reproductive genes in other organisms. We obtained genes lists from studies on female reproductive genes for the following taxa: *D. simulans*, *D. melanogaster*, *D. arizonae*, *Apis mellifera* and *Anopheles gambiae* [37,40,42-44,77-79]. These studies either had the goal of identifying female reproductive genes or looked at expression changes in mated females compared to virgin females. Our search yielded a list of 2,952 contigs, which were then used as queries in a BLAST search against our transcriptome without thorax sequences using the tblastx algorithm and an e-value cutoff of  $10^{-5}$ .

#### Comparison of E- and Z- reproductive genes

As a final approach to examine the bursal gland and bursa copulatrix for possible roles in post-mating sexual selection or post-mating, pre-zygotic isolation, we explored patterns of gene expression between Z and E strain females. Although differential expression alone could be viewed as evidence in support of a functional relationship and mechanism underlying dysfunctional inter-strain oviposition [59], here we emphasize enriched or depleted functional terms. After identifying differentially expressed genes between E and Z strain bursas and bursal glands respectively, we used a two-tailed fisher's exact test (cutoff of  $FDR \leq 0.05$ ) to identify relevant GO categories and genes that might account for reduced fecundity after between-strain matings [67,68].

## Results

### De novo sequence assembly

Single-end Illumina sequencing of 11 ECB female reproductive tissue samples yielded more than ~ 700 million raw reads. Paired-end Illumina sequencing of 2 ECB thorax samples yielded ~ 6 million raw reads. The assembled transcriptome of all 13 libraries contained 92,335 transcripts belonging to ~ 51,000 loci with a mean sequence length of 991 bp and a minimum of 201 bp (Table 1). This is likely an overestimate of the number of loci represented in our transcriptome due to de novo assembly limitations. Our

**Table 1 Assembly statistics**

Total # base pairs	Number of assembled sequences	Min (bp)	Median (bp)	Mean (bp)	Max (bp)	n50 (bp)	n50 length (bp)
50868239	51307	201	535	991	15912	8297	1752

Assembled library statistics for E and Z strain bursa copulatrix, bursal gland, and thorax assembly.

mean assembled transcript lengths are greater than or equal to those reported in similar studies using the same sequencing technology [80,81]. Prior to library normalization, log-fold change of read counts between samples differed, which can bias differential expression results. GC-content of each sample also differed prior to normalization. After normalization, gene level counts and GC-content between samples were all equal across libraries (Additional file 1: Figure S2 and Additional file 2: Figure S3).

### Characterizing reproductive function of the bursa copulatrix and bursal gland tissue

Our first approach to examining the reproductive function of the bursa copulatrix and bursal gland was to characterize differences in gene expression between reproductive and non-reproductive tissues. A total of 2,982 transcripts were differentially expressed between the bursa copulatrix and the thorax, representing 8% of all bursa copulatrix genes, whereas 3,316 genes were differentially expressed between the bursal gland and the thorax, representing 9% of all bursal gland genes (Additional file 3: Figure S1). For gene ontology terms enriched in the bursa copulatrix, 20% corresponded to categories related to muscle activity and organization ( $n = 596$ ), while the bursal gland had 3% of genes enriched for the same categories ( $n = 89$ ) (Additional file 4: Table S1). The bursal gland was also enriched for 6 gene ontology categories directly related to sexual reproduction and gamete production (Additional file 4: Table S1). Of all up-regulated differentially expressed transcripts in the bursa copulatrix compared to the thorax, 2% contained a signal peptide indicating they are secreted from the bursa copulatrix and 7% contained at least one transmembrane helix, while the bursal gland and thorax comparison yielded 16% of sequences with a signal peptide and 46% with at least 1 transmembrane helix (Table 2).

Next, we examined the results from our cross-species female reproductive gene comparison to identify conserved classes of female reproductive genes across insect taxa. We found 23 putative ECB homologs in *D. simulans*, *D. melanogaster*, *D. arizonae*, and *A. mellifera*, but no

**Table 2 Up-regulated bursa copulatrix and bursal gland genes with secretion signal peptides or transmembrane motifs**

Tissue	Secretion signal peptide <sup>†</sup>	TMHMM <sup>‡</sup>
Bursa copulatrix	2%	7%
Bursal gland	16%	46%

Percent of bursa copulatrix and bursal gland genes that were up-regulated compared to the female thorax and contain either a secretion signal peptide or at least one transmembrane motif. The bursal gland has a higher percent of predicted proteins with secretion signal peptides and transmembrane helices.

<sup>†</sup>Number of putative proteins with secretion signal peptides.

<sup>‡</sup>Number of genes with at least one transmembrane helix.

homologous sequences in *A.gambiae* (Table 3). Across all four species with significant blast hits to ECB bursa copulatrix or bursal gland genes, many possessed gene ontology functions related to muscle contraction (Table 3).

### Comparison of E- and Z- reproductive genes

To identify female reproductive genes that may be contributing to reproductive isolation between Z and E ECB strains, we searched for differentially expressed genes between strains. Between the Z and E strain bursa copulatrix

tissues, there were 864 genes with significant differential expression and for bursal gland tissues we found 1,390 significantly differentially expressed genes between strains (Additional file 3: Figure S1).

Subsequently, we examined enriched and depleted gene ontology classes with genes that were differentially expressed between ECB strains in either the bursa copulatrix or the bursal gland. Here, we found 7 gene ontology categories, including proteolysis and serine-type peptidase and endopeptidase activity, which were significantly

**Table 3 Between-species comparison of female reproductive genes**

Gene name	Flybase ID	ECB gene	Molecular function	Homologous gene found in
Actin 57B	FBgn0000044	comp26239_c0_seq1	Structural constituent of cytoskeleton	<i>Drosophila simulans</i> , <i>Drosophila arizonae</i> , <i>Drosophila melanogaster</i>
Actin 5C	FBgn0000042	comp26239_c0_seq1	Structural constituent of cytoskeleton	<i>Drosophila simulans</i> , <i>Drosophila arizonae</i> , <i>Drosophila melanogaster</i>
Actin 87E	FBgn0000046	comp26239_c0_seq1	Expressed in larval muscle	<i>Drosophila simulans</i> , <i>Drosophila arizonae</i> , <i>Drosophila melanogaster</i>
NADH dehydrogenase subunit 4	FBgn0262952	comp26239_c0_seq1, comp25774_c0_seq1	NADH dehydrogenase activity	<i>Drosophila simulans</i> , <i>Drosophila arizonae</i> , <i>Drosophila melanogaster</i>
Ribosomal protein LP0	FBgn0000100	comp6763_c0_seq1, comp26468_c0_seq1	Structural constituent of ribosome	<i>Drosophila simulans</i>
Ribosomal protein 5a	FBgn0002590	comp26375_c0_seq1, comp6724_c0_seq1	Structural constituent of ribosome	<i>Drosophila simulans</i>
Aldolase	FBgn0000064	comp13686_c0_seq1	Fructose-bisphosphate aldolase activity	<i>Drosophila simulans</i> , <i>Drosophila melanogaster</i>
Alpha spectrin	FBgn0250789	comp13976_c0_seq1	Actin binding	<i>Drosophila simulans</i>
Myosin heavy chain	FBgn0026059	comp26288_c0_seq1	Actin/ATP binding	<i>Drosophila simulans</i>
Alpha tubulin	FBgn0003884	comp26314_c0_seq1	GTP binding	<i>Drosophila simulans</i>
Calcium ATPase	FBgn0263006	comp6792_c0_seq1	Calcium-transporting ATPase activity	<i>Drosophila simulans</i>
Culex quinquefasciatus transport protein	FBgn0021953	comp15624_c0_seq1	Catalytic activity	<i>Drosophila simulans</i>
Receptor of activated protein kinase C1	FBgn0000273	comp7107_c0_seq1	ATP binding	<i>Drosophila simulans</i>
Elongation factor 2B	FBgn0000559	comp26371_c0_seq1	GTP binding	<i>Drosophila simulans</i> , <i>Drosophila arizonae</i>
Heat shock protein 83	FBgn0001233	comp18191_c0_seq1	ATP binding	<i>Drosophila simulans</i> , <i>Drosophila melanogaster</i>
Ribosomal protein L3	FBgn0020910	comp26384_c0_seq1	Structural constituent of ribosome	<i>Drosophila simulans</i>
V-ATPase	FBgn0027779	comp18376_c0_seq1	Proton-transporting ATPase activity	<i>Drosophila melanogaster</i>
Ribosomal protein L4	FBgn0003279	comp6770_c0_seq1	Structural constituent of ribosome	<i>Drosophila melanogaster</i>
Topomyosin 1	FBgn0003721	comp6703_c0_seq1	Actin binding	<i>Drosophila melanogaster</i>
Protein C kinase 98E	FBgn0003093	comp309823_c0_seq1	ATP binding	<i>Drosophila melanogaster</i>
ATPase	FBgn0013672	comp26423_c0_seq1	Hydrogen-exporting ATPase activity	<i>Drosophila arizonae</i>
Beta Tubulin 56D	FBgn0003887	comp13374_c0_seq1	GTP binding	<i>Drosophila arizonae</i> , <i>Apis mellifera</i>
Ubiquitin	FBgn0010288	comp26322_c0_seq1	Ubiquitin thiolesterase activity	<i>Drosophila arizonae</i>

Comparison between ECB female reproductive genes and female reproductive genes identified in other insect taxa. The gene function category represents the top blast hit for the gene that ECB show homology to.

enriched in the bursa copulatrix compared to our transcriptome and 7 that were depleted (Figure 3). For the bursal gland, we found one gene ontology class, structural constituent of cuticle, that was enriched compared to our transcriptome and one gene ontology category, intracellular, that was depleted.

To further explore female reproductive genes that were differentially expressed between strains, we examined our gene ontology lists for proteases and protease inhibitors. We focused on these classes because they mediate male–female post-mating interactions in *Drosophila* and are rapidly evolving, which makes these proteins likely to be involved in sexual conflict [82,83]. By manually searching annotations lists after a Fisher’s exact test was run in Blast2go, we found 44 proteases and 7 protease inhibitors with differential expression in both strains and in both tissues combined (Tables 4 and 5).

### Discussion

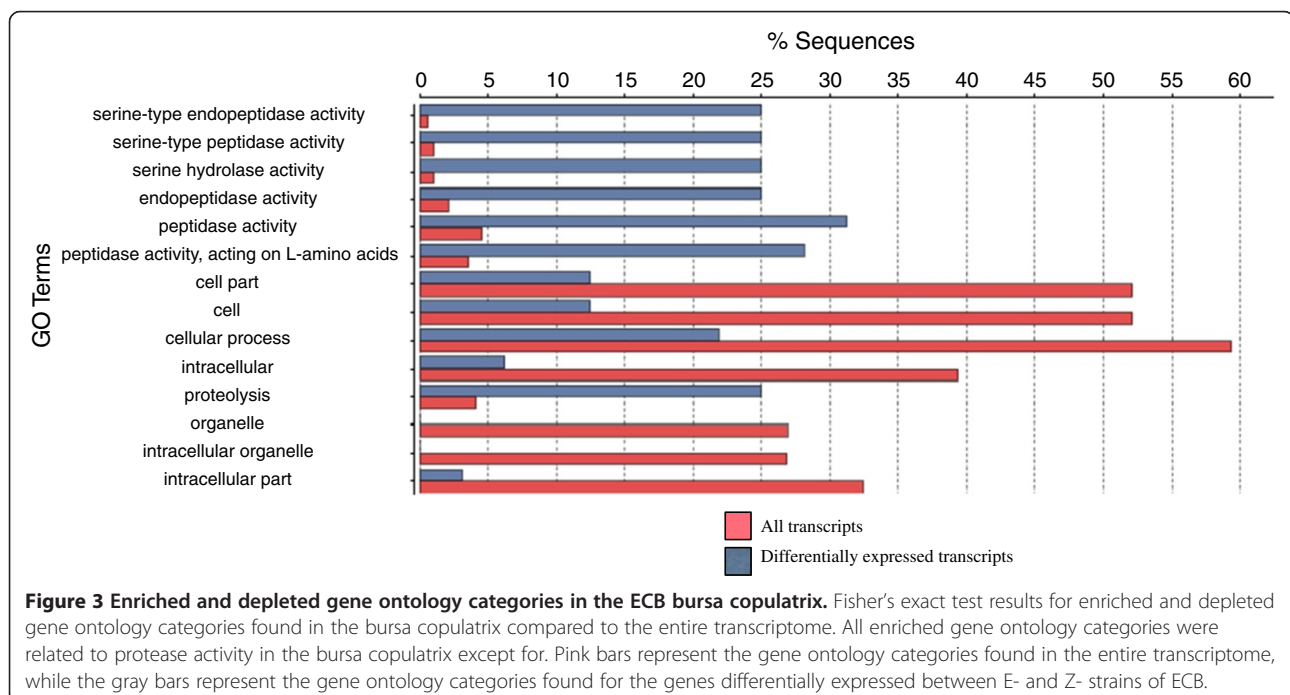
This study represents the first comprehensive characterization of female reproductive genes in any lepidopteran system. Using RNAseq, we identified female reproductive transcripts from the bursa copulatrix and bursal gland and from Z and E strain ECB and we characterized genes that may be involved in reproductive isolation between strains. The bursa copulatrix appears to act as a muscular sac, but it does not seem to secrete the variety of proteins found in the bursal gland, many of which are directly related to reproduction. We also found that most differentially expressed genes in the bursa copulatrix and many in the bursal gland were proteases, which could be important in

post-mating sexual selection and post-mating, pre-zygotic barriers. These are of particular interest because in other species, proteases are known to be involved in female interactions with male sperm and male seminal fluid proteins [82].

### Bursa copulatrix

As the site of initial interaction between male ejaculate and the female reproductive tract [57], the bursa copulatrix is likely to be an important arena for sexual conflict. Previous work has shown that the signum helps break open the male spermatophore [36], and within Papilionidae butterflies signum complexity correlates with the thickness of the outer covering on the spermatophore [84,85]. However, a recent study on the microstructure of the bursa copulatrix in Tortricidae moths identified pores that were suggested to perform a secretory function [56], such as processing the spermatophore.

The transcriptome of the bursa copulatrix would appear to support its role as a site for the mechanical disruption of the spermatophore. Up-regulated genes in the bursa copulatrix compared to the female thorax were statistically significantly enriched for GO classes related to muscle structure or activity (Additional file 4: Table S1), and many of the female reproductive genes in flies and honey bee showing homology with ECB sequences had muscle contraction functions (Table 3). For example, bursa copulatrix transcripts comp6763\_c0\_seq1, comp26288\_c0\_seq1, and comp6703\_c0\_seq1 had strong hits to actin 57B, myosin heavy chain, and topomyosin 1 respectively (Table 3), all key genes during muscle contraction [86].





**Table 4 Differentially expressed proteases and protease inhibitors in the bursa copulatrix**

Bursa copulatrix sequence name	Homologous protein	Direction of differential expression	Predicted function
4699	Astacin-like metalloendopeptidase	Up in E	Peptidase activity
6966	Trypsin	Up in E	Peptidase activity
333327	Prophenol oxidase activating enzyme 3	Up in E	Peptidase activity
12088	Seminal fluid protein hacp002	Up in E	Serine-type endopeptidase activity
215925	Serine protease 44-like	Up in E	Catalytic activity
100535	Trypsin zeta-like	Up in E	Catalytic activity
23804	Trypsin 7	Up in E	Hydrolase activity
10553	Cationic trypsin-like	Up in E	Hydrolase activity
5306	Seminal fluid protein hacp002	Up in E	Proteolysis
19982	Pacifastin-related serine protease inhibitor precursor	Up in E	Peptidase inhibitor activity
25107	Pacifastin-related serine protease inhibitor precursor	Up in E	Peptidase inhibitor activity
18635	Pacifastin-related serine protease inhibitor precursor	Up in E	Peptidase inhibitor activity
30890	Serine proteinase	Up in E	Peptidase activity
114622	Zinc carboxypeptidase	Up in E	Metallopeptidase activity
298440	Carboxypeptidase a5	Up in E	Peptidase activity
966	Carboxypeptidase b-like	Up in E	Metallopeptidase activity
72215	Serine protease	Up in E	Serine-type endopeptidase activity
20820	Prism serine protease inhibitor 1	Up in E	Protease inhibitor
83118	Secreted trypsin-like serine protease	Up in E	Serine-type endopeptidase activity
295616	Angiotensin converting enzyme	Up in E	Metallopeptidase activity
14604	Chymotrypsin-like protein	Up in E	Serine-type endopeptidase activity
121815	Serine protease 48	Up in E	Serine-type endopeptidase activity
18651	Tryptase 5	Up in E	Serine-type endopeptidase activity
25933	Nas-15 protein	Up in E	Metallopeptidase activity
46256	Zinc metalloproteinase nas-13 like	Up in E	Metallopeptidase activity
4699	Astacin-like metalloendopeptidase	Up in E	Peptidase activity
67637	Transmembrane protease serine 3	Up in E	Serine-type endopeptidase activity
4699	Astacin-like metalloendopeptidase	Up in E	Peptidase activity
9559	Trypsin like protein	Up in Z	Peptidase activity
31807	Neuronal pentraxin-2	Up in Z	Proteolysis
96526	Serine protease 1	Up in Z	Hydrolase activity
26516	Pancreatic trypsin inhibitor-like	Up in Z	Serine-type endopeptidase activity
12015	Colostrum trypsin	Up in Z	Serine-type endopeptidase inhibitor activity
254040	Angiotensin converting enzyme	Up in Z	Metallopeptidase activity
10123	Serine protease	Up in Z	Serine-type endopeptidase activity

ECB sequences that showed differentially expressed proteases and protease inhibitors between strains (FDR < 0.05) for the bursa copulatrix. Of 34 proteases and protease inhibitors in the bursa copulatrix, 7 were up-regulated in the Z-strain.

Similar patterns were found in the reproductive tract of female *D. melanogaster*, in which muscle contraction genes were up-regulated in response to mating [38]. Such consistent results across flies, honeybees, and moths suggest a conserved function of the female reproductive tract for muscle contraction across insect taxa. Indeed, even in taxa that lack male spermatophores, muscle contraction in the female reproductive tract has been shown to be important

in moving sperm into storage and for processing male ejaculates [45].

In contrast, we found little evidence supporting a secretory role for the bursa copulatrix in ECB moths, at least for virgin females. Many of the genes that were up-regulated in the bursa copulatrix compared to the female thorax were not putatively secreted proteins (2%), consistent with the notion that the bursa copulatrix lacks an

**Table 5 Differentially expressed proteases and protease inhibitors in the bursal gland**

Bursal gland sequence name	Homologous protein	Direction of differential expression	Predicted function
15139	Retinoid-inducible serine carboxypeptidase-like	Up in E	Serine-type carboxypeptidase activity
144499	Clp domain serine protease 11 precursor	Up in E	Serine-type endopeptidase activity
32437	Wap four-disulfide core domain protein 2 precursor	Up in E	Peptidase inhibitor activity
24588	Cathepsin b	Up in E	Peptidase activity
26860	Serine protease	Up in E	Serine-type endopeptidase activity
6047	Serine protease easter-like	Up in E	Peptidase activity
26327	Vitellin-degrading protease precursor	Up in E	Proteolysis
78030	Serine protease 24	Up in E	Serine-type peptidase activity
26312	Vitellin-degrading protease precursor	Up in E	Proteolysis
26502	Vitellin-degrading protease precursor	Up in E	Proteolysis
7294	Seminal fluid protein hacp057	Up in Z	Cysteine-type peptidase activity
12342	bcp inhibitor	Up in Z	Cysteine-type peptidase activity
13247	Seminal fluid protein hacp057	Up in Z	Cysteine-type peptidase activity
21908	Seminal fluid protein hacp001	Up in Z	Serine-type endopeptidase activity
29657	Trypsin inhibitor precursor	Up in Z	Peptidase inhibitor activity
31807	Neuronal pentraxin-2	Up in Z	Proteolysis
10123	Serine protease	Up in Z	Serine-type endopeptidase activity

Footnote for Table 5: ECB sequences that showed differentially expressed proteases and protease inhibitors between strains (FDR < 0.05) for the bursal gland. Of 17 proteases and protease inhibitors in the bursal gland, 7 were up-regulated in the Z-strain.

important secretory function. Nevertheless, we did find that 7% of up-regulated transcripts possessed transmembrane motifs (Table 2), suggesting the presence of receptors or membrane channels that could interact with male-derived proteins.

#### Bursal gland

Prior to this study, the function of the conspicuous bursal gland present in many lepidopteran female reproductive tracts was completely unknown [49]. Given the direct connection between the bursa copulatrix and the bursal gland, we hypothesized that male-derived products could interact with the bursal gland in two possible ways: by female gland secretions moving into the bursa copulatrix, or by male ejaculate moving into the bursal gland from the bursa copulatrix. Both mechanisms are supported by our transcriptome results, with 16% of up-regulated bursal gland transcripts having secretion signal peptides and 46% having transmembrane motifs (Table 2). Furthermore, unlike the bursa copulatrix the bursal gland had fewer enriched functional categories related to muscle contraction when compared to female thorax (Additional file 4: Table S1), again suggesting mechanical spermatophore breakdown by the bursa copulatrix.

Gene expression in the bursal gland was statistically significantly enriched for many gene ontology categories related to reproduction, including sexual reproduction,

gamete generation, multicellular organism reproduction, cellular process involved in reproduction, and developmental processes involved in reproduction when compared to the female thorax (Additional file 4: Table S1). Of these, one ECB female putative protein stands out. This transcript of interest showed homology to purity essence, which has been shown to be involved in sperm individualization and male fertility [87]. Although its specific role in females is unknown, finding this product in a female reproductive tissue suggests that it also plays a role in female reproduction or fertility.

#### Sexual selection & reproductive isolation

Sexual conflict may extend beyond spermatophore breakdown, with this reproductive “arms race” continuing as males and females struggle for control over fertilization [88]. Such antagonistic sexual coevolution has the potential to contribute to divergence between closely related populations [8,89]. Proteases and protease inhibitors are two classes of proteins that have been shown to be under positive selection in male reproductive tracts in *Drosophila sp.* [82,90]. Although relatively few studies of female reproductive genes have been conducted thus far, these also suggest that proteases and protease inhibitors are important in male–female molecular interactions [42,43,83,91]. Proteases found in the male ejaculate and female reproductive tract have been predicted to co-regulate expression through activation or inhibition of proteolysis, or limit the

time for which a reproductive gene or protein is able to act [24,82]. For example, in *D. arizonae* 12 digestive proteases were specifically expressed in the female reproductive tract and demonstrated signs of positive selection [92]. The functional role of these proteases is unknown; however the adaptive evolution of digestive proteases in *D. arizonae* indicates that they likely play a role in male–female molecular interactions [92].

In ECB females, 2,254 transcripts were differentially expressed between the Z and E strains of ECB in either the bursa copulatrix and the bursal gland. Within the bursa copulatrix, 86% (6/7) of statistically significantly enriched categories dealt with protease function and 34 transcripts showed significant homology to proteases or protease inhibitors (Figure 3). Seven of the proteases found in the bursa copulatrix were over-expressed in Z-strain females compared to E-strain females (Table 4), as were seven of the proteases found in the bursal gland (Table 5). In E-strain female bursa copulatrix tissues, comp18651\_c0\_seq1 had increased expression with a log fold change of 10 and showed significant homology to trypsin 5 (Table 4). Trypsin 5 has been shown to decrease male spermatozoa motility in humans and may be involved in fertility [93]. Another interesting protein, pacifastin-related serine protease inhibitor, was also found to be up-regulated in E-strain female bursa copulatrix tissues (Table 4). Pacifastins have been shown to regulate the immune response, reproduction and phase transition in many insects [94]. Some proteins and inhibitors in the pacifastin family have been shown to have species-specificity in locusts, suggesting they could be important in reproductive isolation [95]. Differential expression of these proteases between strains has the potential to help explain the significant reduction in egg-laying that has been documented when Z-strain females mate with E-strain males during cross-strain matings, but further research using mated females is required to make any conclusions regarding this matter. Other proteases that were differentially expressed between ECB strains in both the bursa copulatrix and the bursal gland were serine and serine-like proteases. Serine proteases have been implicated in increased egg laying after mating seen in many organisms [82]. Serine proteases have also been linked to sperm activation and immune response [82]. This is relevant because mating and sperm storage often leads to changes in regulation of immune response in the female reproductive tract [38,43,96], thought to protect females against male-derived pathogens. Although we have identified these proteases in virgin females, these differentially expressed serine proteases present in the ECB female reproductive tract will provide a fruitful path for future study of post-copulatory interactions and post-mating, pre-zygotic barriers.

## Conclusions

To fully understand post-mating sexual selection and post-mating, pre-zygotic isolation we must examine reproductive transcripts and proteins in taxa with diverse reproductive structures, physiologies, and mating systems. Much has been learned from *Drosophila* concerning the molecular interplay between the sexes that takes place after mating, yet little is known about how these reproductive interactions contribute to divergence. Using the European corn borer, we examined female gene expression in the first portions of the female reproductive tract that come in contact with the male ejaculate. Thus, the sequences described here provide initial insight into male and female post-mating molecular interactions in a model for speciation. Our results indicate that sexual conflict over spermatophore breakdown and male–female molecular interactions are likely to be important in Lepidoptera. We found that the main role of the bursa copulatrix is like to be as a muscular sac that mechanically processes the male spermatophore, while the bursal gland appears to serve a secretory function, producing proteins that could interact with male reproductive proteins. We also found evidence that differential expression of proteases between recently diverged strains in both tissues may contribute to post-mating, pre-zygotic reproductive isolation. Our findings represent an important first step in understanding male–female interactions and the link between sexual selection and divergence in lepidopterans. In future studies, examining changes in gene expression profiles during spermatophore processing will provide additional insight into post-mating sexual interactions.

## Availability of supporting data

The data set supporting the results of this article is available in the NCBI Transcriptome shotgun assembly project under the accession GAVD00000000 and in the NCBI bio-projects under the accession number PRJNA232037.

## Additional files

**Additional file 1: Figure S2.** Stratified box-plots of the log-fold change of read count in each library. Stratified box-plots of the log-fold change of read count before (a) and after (b) normalization for each tissue library.

**Additional file 2: Figure S3.** GC content normalization. Lowess regression of (a.) non-normalized and (b) normalized GC content.

**Additional file 3: Figure S1.** MA plots for each comparison of interest. These plots show the tagwise log-fold-change against the log-counts per million for each gene in a tissue library. Each dot on the graph represents an individual gene. The blue lines across each plot represent a 2 fold change in expression. All red points show differentially expressed genes with a FDR < 0.01 and all black dots are genes that were not significantly differentially expressed.

**Additional file 4: Table S1.** Characterizing reproductive function of the bursa copulatrix and bursal gland tissue. Gene ontology classes in the bursa copulatrix, the bursal gland, and in both tissues compared to the female thorax. Each list of gene ontology categories is in alphabetical order.

## Abbreviations

ECB: European corn borer; GO: Gene ontology.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

NA, SML, and EBD conceived of the study, NA conducted the experiment, and NA, SML, and EBD wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Crista Burke for assisting in library preparation, Rick Harrison and Steve Bogdanowicz for female thorax library preparation, and Charlie Linn and Callie Musto for ECB specimens. This work was supported by a USDA-NIFA (2010-65106-20610) and NSF (DEB-1257521) to EBD. All sequences reads were deposited in the NCBI short-read archive and can be accessed using the following accession numbers.

Received: 23 December 2013 Accepted: 3 March 2014

Published: 12 March 2014

## References

- Kirkpatrick M, Ravigné V: **Speciation by natural and sexual selection: models and experiments.** *Am Nat* 2002, **159**:S22–S35.
- West-Eberhard MJ: **Sexual selection, social competition, and speciation.** *Q Rev Biol* 1983, **58**:155–183.
- Ritchie MG: **Sexual selection and speciation.** *Annu Rev Ecol Syst* 2007, **38**:79–102.
- Panhuis TM, Butlin R, Zuk M, Tregenza T: **Sexual selection and speciation.** *Trends Ecol Evol* 2001, **16**:364–371.
- Eberhard WG, Cordero C: **Sexual selection by cryptic female choice on male seminal products - a new bridge between sexual selection and reproductive physiology.** *Trends Ecol Evol* 1995, **10**:493–496.
- Smith RL: *Sperm Competition and the Evolution of Animal Mating Systems.* Massachusetts: Academic Pr; 1984.
- Simmons LW: *Sperm Competition and its Evolutionary Consequences in the Insects.* New Jersey: Princeton University Press; 2001.
- Arnqvist G, Rowe L: *Sexual Conflict: Monographs in Behavior and Ecology.* New Jersey: Princeton University Press; 2005.
- Andersson M: *Sexual Selection.* New Jersey: Princeton University Press; 1994.
- Birkhead TR, Pizzari T: **Evolution of sex: postcopulatory sexual selection.** *Nat Rev Genet* 2002, **3**:262–273.
- Andersson M, Simmons LW: **Sexual selection and mate choice.** *Trends Ecol Evol* 2006, **21**:296–302.
- Parker GA, Partridge L: **Sexual conflict and speciation.** *Phil Trans Roy Soc B* 1998, **353**:261.
- Coyne JA, Orr HA: *Speciation.* Massachusetts: Sinauer Associates, Inc.; 2004.
- Price C: **Conspecific sperm precedence in *Drosophila*.** *Nature* 1997, **388**:663–666.
- Birkhead TM, Hosken DJ, Pitnick SS: *Sperm Biology: an Evolutionary Perspective.* Pennsylvania: Elsevier Ltd; 2009.
- Howard DJ, Gregory PG, Chu J, Cain ML: **Conspecific sperm precedence is an effective barrier to hybridization between closely related species.** *Evolution* 1998, **52**:511–516.
- Hewitt GM, Mason P, Nichols RA: **Sperm precedence and homogamy across a hybrid zone in the alpine grasshopper *Podisma pedestris*.** *Heredity* 1989, **62**:343–353.
- Bella JL, Butlin RK, Ferris C, Hewitt GM: **Asymmetrical homogamy and unequal sex ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies.** *Heredity* 1992, **68**:345–352.
- Wade MJ, Patterson H, Chang NW, Johnson NA: **Postcopulatory, prezygotic isolation in flour beetles.** *Heredity* 1994, **72**(Pt 2):163–167.
- Manier MK, Lüpold S, Belote JM, Starmer WT, Berben KS, Ala-Honkola O, Collins WF, Pitnick S: **Postcopulatory sexual selection generates speciation phenotypes in *Drosophila*.** *CURBIO* 2013, **23**:1853–1862.
- Howard DJ: **Conspecific sperm and pollen precedence and speciation.** *Annu Rev Ecol Syst* 1999, **30**:109–132.
- Swanson WJ, Vacquier VD: **Reproductive protein evolution.** *Annu Rev Ecol Syst* 2002, **33**:161–179.
- Avila FW, Sirot LK, Laflamme BA, Rubinstein CD, Wolfner MF: **Insect seminal fluid proteins: identification and function.** *Annu Rev Entomol* 2011, **56**:21–40.
- 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.
- Chapman T: **The soup in My Fly: evolution. Form and function of seminal fluid proteins.** *Plos Biol* 2008, **6**:e179.
- Sirot L, Laflamme B, Sitnik J, Rubinstein C, Avila F, Chow C, Wolfner M: **Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study.** *Adv Genet* 2009, **68**:23–56.
- Wolfner MF: **The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*.** *Heredity* 2002, **88**:85–93.
- Sirot LK, Poulson RL, Caitlin McKenna M, Ginary H, Wolfner MF, Harrington LC: **Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction.** *Insect Biochem Mol Biol* 2008, **38**:176–189.
- Dottorini T, Nicolaides L, Ranson H, Rogers DW, Crisanti A, Catteruccia F: **A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior.** *P Natl Acad Sci* 2007, **104**:16215–16220.
- Reinhardt K, Wong CH, Georgiou AS: **Detection of seminal fluid proteins in the bed bug, *Cimex lectularius*, using two-dimensional gel electrophoresis and mass spectrometry.** *Parasitology* 2008, **136**:283.
- Collins AM, Caperna TJ, Williams V, Garrett WM, Evans JD: **Proteomic analyses of male contributions to honey bee sperm storage and mating.** *Insect Mol Biol* 2006, **15**:541–549.
- Andres JA: **Molecular evolution of seminal proteins in field crickets.** *Mol Biol Evol* 2006, **23**:1574–1584.
- South A, Sirot LK, Lewis SM: **Identification of predicted seminal fluid proteins in *Tribolium castaneum*.** *Insect Mol Biol* 2011, **20**:447–456.
- Walters JR, Harrison RG: **EST analysis of male accessory glands from *Heliconius* butterflies with divergent mating systems.** *BMC Genomics* 2008, **9**:592.
- Walters JR, Harrison RG: **Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies.** *Mol Biol Evol* 2010, **27**:2000–2013.
- Chapman RF: *The Insects: Structure and Function.* Cambridge: Cambridge University Press; 2003:1–16.
- Rogers DW, Whitten MMA, Thailayil J, Soichot J, Levashina EA, Catteruccia F: **Molecular and cellular components of the mating machinery in *Anopheles gambiae* females.** *P Natl Acad Sci* 2008, **105**:19390–19395.
- Mack PD, Kapelinikov A, Heifetz Y, Bender M: **Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*.** *P Natl Acad Sci USA* 2006, **103**:10358–10363.
- Lawnczak MK, Begun DJ: **A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females.** *Genome* 2004, **47**:900–910.
- McGraw LA, Gibson G, Clark AG, Wolfner MF: **Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*.** *Curr Biol* 2004, **14**:1509–1514.
- McGraw LA, Clark AG, Wolfner MF: **Post-mating gene expression profiles of female *Drosophila melanogaster* in response to time and to four male accessory gland proteins.** *Genetics* 2008, **179**:1395–1408.
- Prokupek A, Hoffmann F, Eyun S-I, Moriyama E, Zhou M, Harshman L: **An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes.** *Evolution* 2008, **62**:2936–2947.
- Prokupek AM, Kachman SD, Ladunga I, Harshman LG: **Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*.** *Insect Mol Biol* 2009, **18**:465–475.
- Baer B, Eubel H, Taylor NL, O'Toole N, Millar AH: **Insights into female sperm storage from the spermathecal fluid proteome of the honeybee *Apis mellifera*.** *Genome Biol* 2009, **10**:R67.
- Suarez SS: **Sperm transport in the female reproductive tract.** *Human Repro Update* 2005, **12**:23–37.
- Bloch Qazi MC, Aprille JR, Lewis SM: **Female role in sperm storage in the red flour beetle, *Tribolium castaneum*.** *Comp Biochem Phys A* 1998, **120**:641–647.
- Wolfner MF: **Battle and ballet: molecular interactions between the sexes in *Drosophila*.** *J Hered* 2009, **100**:399–410.
- Zhong W, McClure CD, Evans CR, Mlynski DT, Immonen E, Ritchie MG, Priest NK: **Immune anticipation of mating in *Drosophila*: Turandot M promotes immunity against sexually transmitted fungal infections.** *Proc B* 2013, **280**:20132018.



49. Kristensen NP: **Lepidoptera, moths and butterflies.** In *Handbook of Zoology*. 4th edition. Edited by Maximilian Fischer. New York: Walter de Gruyter, Inc; 2003.
50. Arnqvist G, Edvardsson M, Friberg U, Nilsson T: **Sexual conflict promotes speciation in insects.** *PNAS* 2000, **97**:10460–10464.
51. McNamara KB, Elgar MA, Jones TM: **Large spermatophores reduce female receptivity and increase male paternity success in the almond moth, *Cadra cautella*.** *Anim Behav* 2009, **77**:931–936.
52. Gavrillets S, Arnqvist G, Friberg U: **The evolution of female mate choice by sexual conflict.** *Proc B* 2001, **268**:531–539.
53. Kubli E: **Sex-peptides: seminal peptides of the *Drosophila* male.** *Cell Mol Life Sci* 2003, **60**:1689–1704.
54. Jennions MD, Petrie M: **Why do females mate multiply? A review of the genetic benefits.** *Biol Rev* 2012, **75**:21–64.
55. South A, Lewis SM: **The influence of male ejaculate quantity on female fitness: a meta-analysis.** *Biol Rev* 2011, **86**:299–309.
56. Lincango P, Fernández G, Baixeras J: **Microstructure and diversity of the bursa copulatrix wall in Tortricidae (Lepidoptera).** *Arth Struct Dev* 2013, **42**:247–256.
57. Drecktrah HG, Brindley TA: **Morphology of the internal reproductive systems of the European corn borer.** *Iowa State J Sci* 1967, **41**:467–480.
58. Malausa T, Dalecky A, Ponsard S, Audiot P, Streiff R, Chaval Y, Bourguet D: **Genetic structure and gene flow in French populations of two *Ostriniataxa*: host races or sibling species?** *Mol Ecol* 2007, **16**:4210–4222.
59. Dopman EB, Robbins PS, Seaman A: **Components of reproductive isolation between North American pheromone strains of the European corn borer.** *Evolution* 2009, **64**:881–902.
60. Fadamiro HY, Baker TC: **Reproductive performance and longevity of female European corn borer, *Ostrinia nubilalis*: effects of multiple mating, delay in mating, and adult feeding.** *J Insect Phys* 1999, **45**:385–392.
61. Royer L, McNeil JN: **Changes in calling behaviour and mating success in the European corn borer (*Ostrinia nubilalis*), caused by relative humidity.** *Entomol Exp Appl* 1991, **61**:131–138.
62. Lindgreen S: **AdapterRemoval: easy cleaning of next generation sequencing reads.** *BMC Res Notes* 2012, **5**:337.
63. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Meth* 2012, **9**:357–359.
64. Pearson WR, Wood T, Zhang Z, Miller W: **Comparison of DNA sequences with protein sequences.** *Genomics* 1997, **46**:24–36.
65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**:644–652.
66. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, Williams T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc* 2013, **8**:1494–1512.
67. Conesa A, Götz S: **Blast2GO: a comprehensive suite for functional analysis in plant genomics.** *Int J Plant Sci* 2008, **2008**:1–12.
68. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**:3420–3435.
69. Risso D, Schwartz K, Sherlock G, Dudoit S: **1471-2105-12-480-1.** *BMC Bioinformatics* 2011, **12**:480.
70. McCarthy DJ, Chen Y, Smyth GK: **Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.** *Nucleic Acids Res* 2012, **40**:4288–4297.
71. Robinson MD, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data.** *Genome Biol* 2010, **11**:R25.
72. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2009, **26**:139–140.
73. Robinson MD, Smyth GK: **Moderated statistical tests for assessing differences in tag abundance.** *Bioinformatics* 2007, **23**:2881–2887.
74. Iseli C, Jongeneel CV, Bucher P: **ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** *Proc Int Conf Intell Syst Mol Biol* 1999, **7**:138–148.
75. Petersen TN, Brunak S, Heijne Von G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Meth* 2011, **8**:785–786.
76. Sonnhammer E, Heijne Von G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175–182.
77. Kelleher ES, Swanson WJ, Markow TA: **Gene duplication and adaptive evolution of digestive proteases in *Drosophila* Arizonae female reproductive tracts.** *PLoS Genet* 2007, **3**:e148.
78. Bono JM, Matzkin LM, Kelleher ES, Markow TA: **Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated *Drosophila* mojavensis females.** *Proc Natl Acad Sci* 2011, **108**:7878–7883.
79. Swanson WJ: **Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection.** *Genetics* 2004, **168**:1457–1465.
80. Feng C, Chen M, Xu C-J, Bai L, Yin X-R, Li X, Allan AC, Ferguson IB, Chen K-S: **Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq.** *BMC Genomics* 2012, **13**:19.
81. Wang X-W, Luan J-B, Li J-M, Bao Y-Y, Zhang C-X, Liu S-S: **De novo characterization of a whitefly transcriptome and analysis of its gene expression during development.** *BMC Genomics* 2010, **11**:400.
82. Laflamme BA, Wolfner MF: **Identification and function of proteolysis regulators in seminal fluid.** *Mol Repro Dev* 2012, **80**:80–101.
83. Lawnczak MKN, Begun DJ: **Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila* melanogaster.** *Mol Biol Evol* 2007, **24**:1944–1951.
84. Sánchez V, Hernández-Baños BE, Cordero C: **The evolution of a female genital trait widely distributed in the Lepidoptera: comparative evidence for an effect of sexual coevolution.** *PLoS ONE* 2011, **6**:e22642.
85. Scriber JM, Tsubaki Y, Lederhouse RC: *Swallowtail Butterflies: Their Ecology and Evolutionary Biology.* Florida: Scientific Publishers Inc; 1995.
86. Benoist P: **Differential muscle-type expression of the *Drosophila* Troponin T gene: a 3-base pair microexon is involved in visceral and adult hypodermic muscle specification.** *J Biol Chem* 1998, **273**:7538–7546.
87. Castrillon DH, Gonczy P, Alexander S, Rawson R, Eberhart CG, Viswanathan S, DiNardo S, Wasserman SA: **Toward a molecular genetic analysis of spermatogenesis in *Drosophila* melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis.** *Genetics* 1993, **135**:489.
88. Parker GA: **Sexual conflict over mating and fertilization: an overview.** *Phil Trans Roy B* 2006, **361**:235–259.
89. Chapman T, Arnqvist G, Bangham J, Rowe L: **Sexual conflict.** *Trends Ecol Evol* 2003, **18**:41–47.
90. Wong A, Turchin MC, Wolfner MF, Aquadro CF: **Evidence for positive selection on *Drosophila* melanogaster seminal fluid protease homologs.** *Mol Biol and Evol* 2008, **25**:497–506.
91. 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.
92. Kelleher ES, Markow TA: **Duplication, selection and gene conversion in a *Drosophila* mojavensis female reproductive protein family.** *Genetics* 2009, **181**:1451–1465.
93. Weidinger S: **Mast cell-sperm interaction: evidence for tryptase and proteinase-activated receptors in the regulation of sperm motility.** *Human Repro* 2003, **18**:2519–2524.
94. Breugelmanns B, Simonet G, Van Hoef V, Van Soest S, Vanden Broeck J: **Pacifastin-related peptides: structural and functional characteristics of a family of serine peptidase inhibitors.** *Peptides* 2009, **30**:622–632.
95. Gaspari Z, Ortutay C, Perczel A: **A simple fold with variations: the pacifastin inhibitor family.** *Bioinformatics* 2004, **20**:448–451.
96. Baer B, Armitage SAO, Boomsma JJ: **Sperm storage induces an immunity cost in ants.** *Nature* 2006, **441**:872–875.

doi:10.1186/1471-2164-15-189

**Cite this article as:** Al-Wathiqui et al.: Using RNA sequencing to characterize female reproductive genes between Z and E Strains of European Corn Borer moth (*Ostrinia nubilalis*). *BMC Genomics* 2014 **15**:189.