Postcopulatory reproductive processes: The role of female and male reproductive genes and proteins in speciation and sexual selection

A dissertation submitted by Nooria M. Al-Wathiqui

In partial fulfillment of the requirement for the degree of Doctor of Philosophy in

Biology

Tufts University

May 2016

ADVISOR: Dr. Sara M. Lewis

Abstract

Sexual selection is a powerful force that drives the evolution of reproductive traits and continues after mating is completed. Postcopulatory sexual selection involves molecular interactions between the male ejaculate and the female reproductive tract; these interactions are mediated by male seminal fluid proteins (SFPs), transferred in the male ejaculate, and female reproductive proteins (FRPs), secreted by the female reproductive tract. SFPs have been characterized in numerous species and shown to affect female behavior and physiology. However, they have been poorly studied in species where males transfer a packaged ejaculate, called a spermatophore. Many insects transfer spermatophores, yet the implication of this type of ejaculate transfer for postcopulatory interactions is unclear. Furthermore, we understand little about FRPs and their roles in postcopulatory sexual selection. Not only are female proteins potential mediators of sexual selection, but may play a role in reproductive isolation due to their potential for rapid evolution.

Here, I used a combination of RNA sequencing, proteomics, and metabolomics to characterize SFPs and FRPs in the spermatophore-transferring taxa: *Ostrinia nubilalis* moths, *Tribolium castaneum* beetles, and *Photinus pyralis* fireflies to determine how reproductive genes, proteins, and metabolites differentially regulate postcopulatory interactions between the sexes in different ecological contexts. First, I used these methods to identify male and female reproductive genes that could be mediating a postmating, prezygotic barrier acting

between *O. nubilalis* strains. I found that ECB males differentially express peptidases and odorant binding proteins between strains. After mating within- and across-strain, females of *O. nubilalis* also differentially expressed several reproductive genes, many of which are novel. In *T. castaneum*, I used experimentally enforced monandry to examine how relaxed postcopulatory sexual selection could influence reproductive gene expression. Monandrous males showed a shift in gene expression that indicated they may be increasing sperm or production of eggs in female mates. Finally, in *P. pyralis* fireflies I characterized the composition of the male spermatophore, which I found contains a number of peptidases and proteins related to the immune response. Across all three taxa, I found that both sexes express peptidases that may be important mediators of postcopulatory sexual interactions in these spermatophore-producing species.

ACKNOWLEDGEMENTS

This work would not have been possible without much support: financial, intellectual and emotional. I would like to start by thanking the funding agencies that provided the Lewis lab and myself with resources to conduct scientific research including Tufts University, The Society for the Study of Evolution, and the National science foundation. I would also like to show extreme gratitude to my advisor, Dr. Sara Lewis; an inspiring scientist and mentor. Her encouragement, support, scientific prowess, and enthusiasm helped to shape me as the scientist I have become.

To my committee members Frances Chew, Colin Orians, and Stephen Fuchs, thank you for your thought-provoking comments and advice throughout my time at Tufts University. A special thank you must be given to Erik Dopman, for guiding me through learning about computer programming and sequence analysis, as well as providing funding for various projects. To Steve Fuchs, thank you for provided me with advice and the resources to conduct proteomics on the ECB spermatophore. I would also like to thank my outside committee member, Laura Sirot, whose work in the Wolfner lab and now in her own lab inspired me to study post-mating interactions in the first place.

Part of what makes the Tufts University Biology department so great is the interactions that graduate students have with all faculty members, even those outside of their immediate committee. To Mitch McVey, Kelly McLaughlin, Susan Ernst and Susan Koegel: I thank you for providing me with guidance on

becoming an effective educator and for giving me advice as I pursue career options.

To Eillen Magnant, Elizabeth Palmer and Mike Grossi, no matter what ridiculous request or question I had, you were always able to help me. I will miss visiting with you during my coffee breaks.

I would also like to thank my collaborators at Tufts and MIT for working with me on various research projects. Rachael Bonoan in the Starks lab at Tufts University has an amazingly positive attitude and really inspired me to be as enthusiastic about the work we did together. Timothy Fallon and Jing-Ke Weng at MIT have provided me with resources and guidance as I worked on Chapter 6. Tim has been an invaluable resource for troubleshooting this project. I wish Tim the best of luck in finishing his own thesis work.

My labmates Adam South, Natasha Tigreros, and Amanda Franklin deserve a special acknowledgement, as they have taught me skills in and outside of the lab that I have used throughout my career as a graduate student. I already know that it is going to be difficult to not bother Amanda in her office every morning.

Finally I would like to thank my friends and family. Jen Mortensen and Esther Miller have been fantastic office mates and provided much advice and comic relief during the years we've worked together. They have also fed me dinner an embarrassing number of times. To my sisters, Nadia, Tamila, and Deena and my brother, Mishal, your texts, calls, and visits have helped make me feel at home, even halfway across the country. I'm so proud of all of your successes. To my parents, Mahmood and Susan, its hard to even find the words to express the

gratitude I have for everything you've given me. Thank you for supporting me through 23+ years of school. Finally, to my co-conspirator and husband, Robert Burns, I definitely would not have finished this thesis without your encouragement, love, and support. I am in awe of the hurdles you have overcome and the progress you have made throughout the time I've known you.

Table of Contents

Index of Tables

Index of Figures

Index of supplementary information

Supplementary Tables:

Supplementary Figures:

Chapter 1. Thesis Overview

> ‘‘We are, however, here concerned only with that kind of selection, which I have called sexual selection. This depends on the advantage which certain individuals have over other individuals of the same sex and species, in exclusive relation to reproduction.’’
>
> - Charles Darwin

Charles Darwin defined sexual selection over 150 years ago as the advantage that some individuals have over other individuals in relation to reproduction (Darwin, 1874). Research on sexual selection has largely focused on mechanisms of precopulatory sexual selection, including intersexual choice, where one choosy sex selects mates based on the quality of their phenotypic traits and intrasexual competition, where members of the same sex compete for access to mates (Andersson, 1994). However, both intra- and intersexual selection are not limited to this premating time period, as they continue to act both during mating (pericopulatory sexual selection), and after a mating pair has separated (postcopulatory sexual selection) (Parker, 1970; Simmons, 2001). Postcopulatory sexual selection has proven difficult to study as it takes place within the female reproductive tract and often involves molecular interactions between the male ejaculate and the female reproductive tract. As a result, postcopulatory interactions between males and females can be influenced by secretions produced by the male and female reproductive tract (Chen, 1984; Chapman, 2003). In males, elaborate accessory glands along with increased complexity of male ejaculates have likely evolved to mediate these postcopulatory interactions (Lewis & South, 2012). Complex male ejaculates not only serve to transfer sperm, but

can also contain nutritive substances, defensive compounds, and seminal fluid proteins (SFPs) (Poiani, 2006; Perry et al., 2013; Boggs, 1990). SFPs have garnered particular interest as they likely mediate many aspects of postcopulatory sexual selection (Wolfner, 2009; Laflamme & Wolfner, 2013; Ravi Ram & Wolfner, 2007; Sirot et al., 2009; Gillott, 2003; Avila et al., 2011). For example, in *D. melanogaster*, SFPs are involved in many key postcopulatory processes, including reducing female receptivity to further mating, increasing oogenesis and oviposition, altering female sperm storage and use, and changing female feeding and sleep patterns (Ravi Ram et al., 2005; Laflamme & Wolfner, 2013; Sirot et al., 2009; Wolfner, 2009). Across the taxa studied thus far the protein classes mediating these processes appear to be largely conserved (Mueller et al., 2004); they include peptidases, peptidase regulators, immunity-related proteins, sperm-binding proteins, and peptide hormones.

While SFPs have been characterized in many insects (Gillott, 2003; Avila et al., 2011), female reproductive secretions have been poorly studied. Although female reproductive tracts also contain glandular and secretory tissues, the products of these tissues have been studied in only a few insect species, including *Pieris* butterflies (Meslin et al., 2015), *Drosophila* fruit flies (Prokupek et al., 2008; 2009; McGraw et al., 2004; Kelleher et al., 2007; Bono et al., 2011; Swanson, 2004), *Apis* honeybees (Baer et al., 2009), and *Ostrinia* moths (Al-Wathiqui et al., 2014). These studies demonstrate that female reproductive tissues secrete substances that may directly interact with male ejaculates, including

peptidases and peptidase regulators, immunity-related proteins, as well asproteins important for hormone signaling and sperm maintenance.

In some cases, direct interactions between male and female reproductive proteins have been demonstrated to play a role in postcopulatory sexual selection (Wolfner, 2009; Sirot et al., 2009; Ravi Ram & Wolfner, 2007; Laflamme & Wolfner, 2013). For example, *D. melanogaster* males transfer an SFP called ovulin, an ovulation-inducing prohormone (Herndon & Wolfner, 1995; Heifetz et al., 2000). Cleavage of this protein to its most active form involves both male and female peptidases (Park & Wolfner, 1995). Another SFP, called the sex peptide, which decrease female remating and increases egg production (Chen, 1984), requires a network of at least 13 male and female proteins to function properly (Yapici et al., 2008; Findlay et al., 2014). Thus, this work demonstrates the importance of both male and female reproductive proteins to our understanding of postcopulatory sexual selection.

In addition to being important for post-mating sexual selection, reproductive proteins may play an important role in species divergence. Both male and female reproductive genes have been shown to evolve rapidly (e.g. Swanson et al., 2001; Swanson, 2004; Prokupek et al., 2010; Kelleher et al., 2011); among *Drosophila* species, reproductive evolution is faster in lineages with more promiscuous mating systems, where sexual selection is intensified, when compared to monogamous lineages (Wagstaff, 2005; Swanson et al., 2001; Kelleher et al., 2007). Processes inducing evolutionary pressure on reproductive proteins include sexual conflict, where the sexes differ in their desired outcome of

reproduction (Arnqvist & Rowe, 2013). Sexual conflict can lead to a coevolutionary arms race between the sexes for control over reproductive outcomes (Arnqvist & Rowe, 2013; Andersson, 1994). As these reproductive proteins evolve rapidly, this can lead to divergence between populations (Swanson & Vacquier, 2002). Processes likely mediated by reproductive proteins, including male-male sperm competition and sexual conflict, have been linked to reproductive isolation between populations (Manier et al., 2013; Yeates et al., 2013).

Dysfunctional interactions between male and female reproductive proteins could occur due to divergence in either protein sequence or in the regulatory elements that control gene expression. Gene expression is of particular interest as it can change rapidly to respond to stimuli and can evolve over time (Mank et al., 2013). Recent advances in sequencing technology make it possible to determine how differences in gene expression may contribute to divergence in postcopulatory traits.

Thus far, the study of reproductive genes and proteins has been biased towards identifying SFPs in dipteran species where males transfer a liquid, sperm-containing ejaculate. In many other insects, males transfer sperm in a package called a spermatophore (Mann, 1984). However, male reproductive proteins have only been characterized in a few spermatophore-producing species (South et al., 2011; Walters & Harrison, 2010; Marshall et al., 2009; Civetta et al., 2006). Furthermore, female reproductive genes have only been characterized in two spermatophore producing species (Al-Wathiqui et al., 2014; Meslin et al., 2015).

Spermatophores can differ greatly in their structure and function (Mann, 1984). They also change male-female postcopulatory interactions, as spermatophores can require further processing before females are able to remate. For example, in most lepidopterans, males transfer sperm in a large, tough spermatophore (Wedell, 2005). Once it enters the female's bursa copulatrix, presence of the spermatophore activates female stretch receptors and causes the female to become unreceptive to future coutship (Sugawara, 1979). Lepidopteran females have evolved a chitinized structure, the signum, which along with bursal muscle contraction, aids in degrading the spermatophore, allowing females to remate (Cordero, 2005; Galicia et al., 2008).

In this thesis, I characterize male and female reproductive genes and proteins for three different spermatophore-producing species: *Ostrinia nubilalis* moths, *Tribolium castaneum* beetles, and *Photinus pyralis* fireflies. These taxa differ in their life histories and mating systems allowing for the study of these proteins in different ecological contexts. The research in this thesis is presented in six chapters that provide insight into how male and female reproductive proteins contribute to sexual selection and speciation. Chapter 2 , Chapter 3, and Chapter 4 focus on identifying male and female reproductive genes and proteins in the two divergent strains of the European corn borer moth (*Ostrinia nubilalis*). Chapter 2, published in 2014, presents the first identification of female reproductive genes in the bursa copulatrix and bursal gland in a lepidopteran. Chapter 3 identifies male reproductive genes and proteins for the two divergent strains of *O. nubilalis* moths, and how these genes may contribute to the known PMPZ barrier. Chapter

4 investigates postcopulatory changes in gene expression within the female reproductive tract of *O. nubilalis* moths. This chapter also examines differences in gene expression after females had mated with same and opposite strain males to determine how female reproductive genes may contribute to a known PMPZ barrier. Chapter 5 uses experimental evolution in the flour beetle *T. castaneum* to determine how male and female reproductive genes change in response to strength of sexual selection. *T. castaneum* beetles were used in this study because they are a highly promiscuous species and a model organism for sexual selection research. Chapter 6 focuses on understanding the synthesis, composition and fate of the spermatophore gift in *P. pyralis* fireflies. We characterize the *P. pyralis* male spermatophore because it is an important mediator of male differential paternity success. By examining reproductive proteins in these different insect taxa and from different evolutionary perspectives, this work contributes to our understanding of how both male and female reproductive genes and proteins function in sexual selection and reproductive isolation.

Chapter 2: Using RNA sequencing to characterize female reproductive genes between "Z" and "E" Strains of European Corn Borer moth (*Ostrinia nubilalis*)

Abstract

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*). 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 muscle 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. Between strains, 2% and 4% of genes were differentially regulated in

the bursa copulatrix and bursal gland, respectively. Divergently expressed genes in the bursa copulatrix were exclusively biased toward protease-like functions and 51 proteases or protease inhibitors were divergently expressed overall. Our study represents 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 post-mating, pre-zygotic 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.

Introduction:

Sexual selection is a powerful evolutionary force that can drive species divergence (Kirkpatrick M & Ravigné, 2002; West-Eberhard, 1983; Ritchie, 2007; Panhuis et al., 2001). 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 (Eberhard & Cordero, 1995; Smith, 1984; Simmons, 2001). 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 (Arnqvist & Rowe, 2005). 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 (Arnqvist & Rowe, 2005; Andersson, 2006) or divergence between populations (Birkhead & Pizzari, 2002; Andersson & Simmons, 2006; Parker & Patridge, 1998). Differences in reproductive traits between populations can quickly result in post-mating, pre-zygotic barriers (Coyne & Orr, 2004), 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 (Price, 1997; Birkhead et al., 2009; Howard et al., 1998). Conspecific sperm precedence is widespread and has been demonstrated in many insect species, including fruit flies, ground crickets and flour beetles (Price, 1997; Howard et al., 1998; Hewitt et al., 1989; Bella et al., 1992; Wade et al., 1994) . Of several possible mechanisms underlying conspecific sperm precedence in *Drosophila mauritiana* and *Drosophila similans*, biased sperm use by females was found to be a key determinant (Manier et al., 2013). Female *Drosophila* are able to favor conspecific males by preferentially

storing sperm in separate storage organs (Manier et al., 2013). 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 (Howard, 1999; Swanson & Vacquier, 2002).

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 (Avila et al., 2011). In fruit flies, *Drosophila melanogaster* males produce over 100 seminal proteins that are transferred to females during mating (Ram & Wolfner, 2007; Chapman, 2008; Sirot et al., 2009). These proteins have profound effects on female behavior and physiology, including changes in lifespan, ovulation, feeding habits and sperm storage patterns (Price, 1997; Ram & Wolfner, 2007; Sirot et al., 2009). 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 (Swanson & Vacquier, 2002). Seminal fluid proteins have been comprehensively characterized in several insect taxa, including fruit flies, mosquitoes, honeybees, crickets, flour beetles, butterflies, and bedbugs (Ram & Wolfner, 2007; Wolfner, 2002; Sirot et al., 2008; Dottorini et al., 2007; Reinhardt et al., 2008; Collins et al., 2006; Andres, 2006; South et al., 2011; Walters &

Harrison, 2008; Walters & Harrison, 2010). 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 (Chapman, 2003), to date female reproductive genes have been comprehensively studied in very few taxa including mosquitos, fruit flies, and honeybees (Rogers et al., 2008; Mack et al., 2006; Lawniczak & Begun, 2004; McGraw et al., 2004; McGraw et al., 2008; Prokupek et al., 2008; Prokupek et al., 2009; Baer et al., 2009). Unsurprisingly, female reproductive genes have been best characterized in *Drosophila* species including: *D. melanogaster*, *D. simulans*, *D. arizonae* and *D. mojavensis* (Mack et al., 2006; Lawniczak & Begun, 2004; McGraw et al., 2004; McGraw et al., 2008; Prokupek et al., 2008; Prokupek et al., 2009). 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 (Mack et al., 2006; Lawniczak & Begun, 2004; McGraw et al., 2004; McGraw et al., 2008; Prokupek et al., 2008; Prokupek et al., 2009). 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 (Suarez, 2005; Bloch-Qazi et al., 1998), while proteases and protease inhibitors have been shown to be required for activation of ovulation-inducing seminal fluid proteins in *D. melanogaster* (Wolfner, 2009). 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 (Mack et al., 2006;

McGraw et al., 2004; Zhong et al., 2013). 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 (Prokupek et al., 2008).

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 (Kristensen, 2003; Arnqvist & Rowe, 2005). 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 (Walters & Harrison, 2008; Walters & Harrison, 2010); however, researchers have yet to identify female genes from any structure in the lepidopteran female reproductive tract.

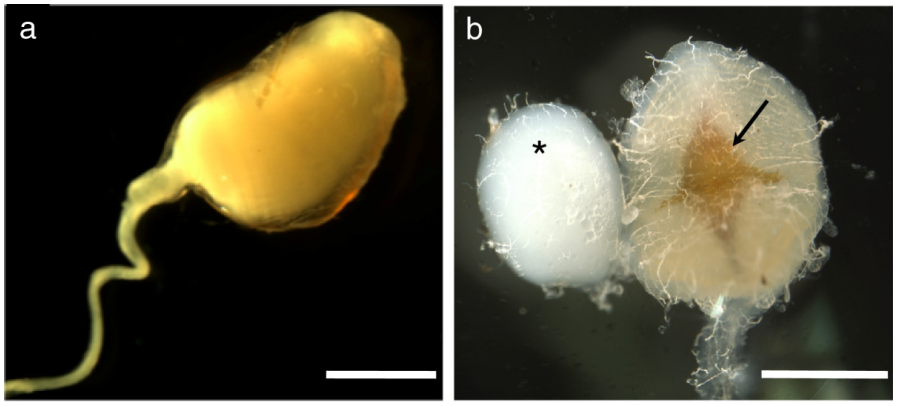


Figure 1. Male and female ECB reproductive structures. **a**. An ECB male spermatophore transferred to a female upon mating. **b**. The virgin female bursa copulatrix with signum (indicated by arrow) and the bursal gland (indicated by*). Notice that even in virgins the bursal gland is filled with

Most male lepidopterans transfer their ejaculate in a package called a spermatophore (Figure 1a) (Chapman, 2003; Kristensen, 2003). 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 (Chapman, 2003; Kristensen, 2003). 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 (McNamara et al., 2009; Gavrilets et al., 2001) and in *D. melanogaster,* identified male reproductive proteins act to reduce female receptivity to future mates (Wolfner, 2009; Kubli, 2003). On the other hand, females in many taxa gain material and genetic benefits from multiply mating (Jennions & Petrie, 2012; South & Lewis, 2011), 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 (Lincango et al., 2013). 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* (Drecktrah & Brindley, 1967). 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 (Drecktrah & Brindley, 1967). After mating, the bursal gland is similarly filled, but with an opaque fluid (Drecktrah & Brindley, 1967). 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 secreting 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 (Coyne & Orr, 2004), in which the two incipient species split approximately 75,000 to 150,000 years ago through the evolution of manifold reproductive barriers

(Malausa et al., 2007; Dopman et al., 2009). Females of both strains mate multiply (Drecktrah & Brindley, 1967) and suffer reduced longevity after mating (Fadamiro & Baker, 1999), 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 (Dopman et al., 2009). However, the mechanism underlying this gametic isolation is unknown (Dopman et al., 2009). 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 (Fadamiro & Baker, 1999; Royer & McNeil, 1991). 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 RNAlater (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 RNAlater at -80˚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 μg 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 for 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 X 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 (Lindgreen, 2012). 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 ≥ 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 (Langmead & Salzberg,

2012). Bowtie2 uses the Burrow-Wheeler transformation to index a reference, then searches the index until it finds an alignment for a specific read (Langmead & Salzberg, 2012). 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, EU532444.1, 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 (Pearson et al., 1997).

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 (Grabherr et al., 2011). 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 (Grabherr et al., 2011; Haas et al., 2013). 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 (Conesa & Gotz, 2008; Gotz et al., 2008). 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 dbxreftable and the gene product table of the GO database. Finally, Blast2go computed an annotation score for all possible GO terms for each sequence (Conesa & Gotz, 2008; Gotz et al., 2008).

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 EDAseq (Risso et al., 2011). EDAseq 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 (Risso et al., 2011). 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, EDAseq 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) (McCarthy et al., 2012; Robinson & Oshlack, 2010; Robinson et al., 2009; Robinson & Smyth, 2007). 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 (McCarthy et al., 2012). Genes were considered differentially expressed if they had a false-discovery rate (FDR) of <0.01.

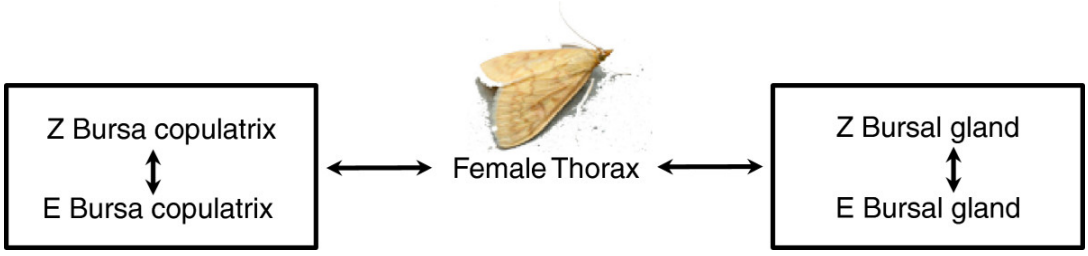


Figure 2. Differential expression analysis tissue comparisons. Our generalized linear model used to calculate differentially expressed genes between E- and Z-strain ECB included the following comparisons: female thorax versus the female bursa copulatrix, female thorax versus bursal gland, Z-strain versus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

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 ≤ 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 (Iseli et al., 1999), 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 motif using SignalP 4.0 (Petersen et al., 2011), which uses a neural network-based method to identify signal peptides. We then used TMHMM 2.0 to identify sequences with

transmembrane helices (Sonnhammer et al., 1998), 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. similans*, *D. melanogaster*, *D. arizonae*, *Apis melifera* and *Anopheles gambiae* (Rogers et al., 2008; McGraw et al., 2004; Prokupek et al., 2008; Prokupek et al., 2009; Baer et al., 2009; Kelleher et al., 2007; Bono et al., 2011; Swanson, 2004). 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 cuttoff 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 (Dopman et al., 2009), 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≤0.05) to identify relevant GO categories and genes that might account for reduced fecundity after between-strain matings (Conesa & Gotz, 2008).

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 mean assembled transcript lengths are greater than or equal to those reported in similar studies using the same sequencing technology (Feng et al., 2012; Wang et al., 2010). 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 files 2 and 3).

Table 1. **Assembly statistics**. Assembled library statistics for E and Z strain bursa copulatrix, bursal gland, and thorax assembly.

Total # base pairs (bp)	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

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 1). 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). The bursal gland was also enriched for 6 gene ontology categories directly related to sexual reproduction and gamete production (additional file 4). 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).

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

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

†Number of putative proteins with secretion signal peptides.
‡Number of genes with at least one transmembrane helix.

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. similans, D. melanogaster, D. arizonae,* and *A. melifera,* but no 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 1).

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 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.

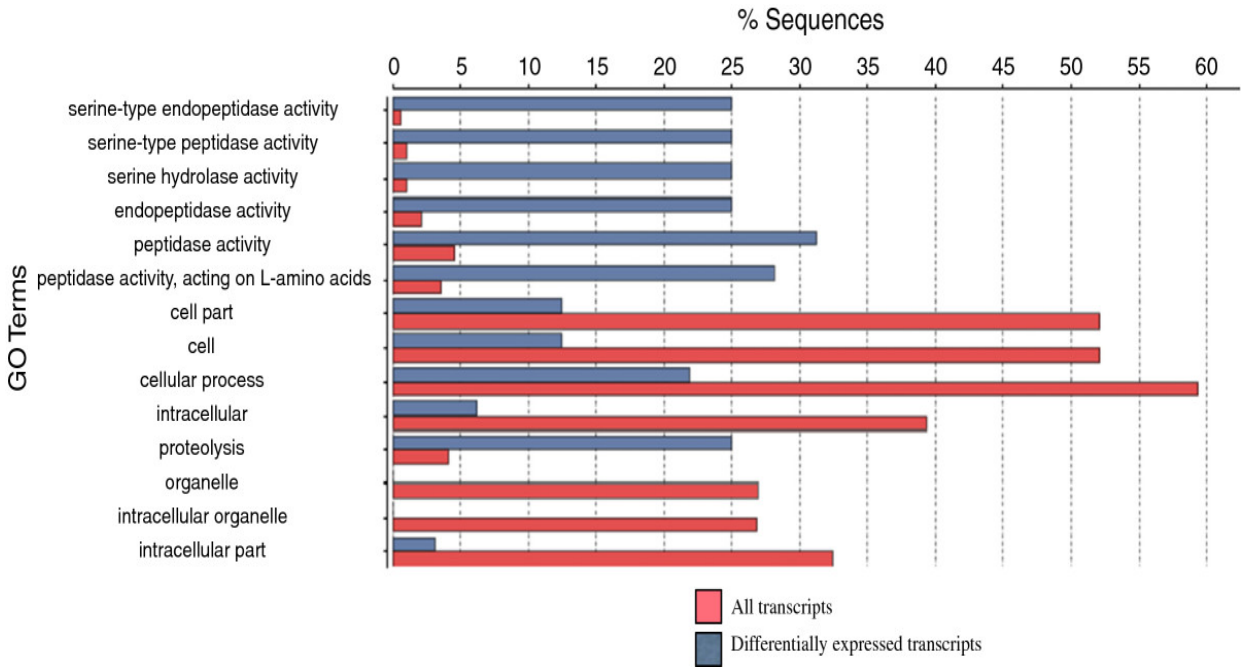


Figure 3. Enriched and depleted gene ontology categories in the ECB bursa copulatrix. Fisher's exact test results for enriched and depleted gene ontology categories found in the bursa copulatrix compared to the entire transcriptome. All enriched gene ontology categories were related to protease activity in the bursa copulatrix. Pink bars represent the gene ontology categories found in the entire transcriptome, while the gray bars represent the gene ontology categories found for the genes differentially expressed between E- and Z- strains of ECB.

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 (Laflamme & Wolfner, 2013; Lawniczak & Begun, 2007). 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 (Table 4a, b).

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 (Laflamme & Wolfner, 2013).

Bursa Copulatrix

As the site of initial interaction between male ejaculate and the female reproductive tract (Drecktrah & Brindley, 1967), 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 (Chapman, 2003), and within Papilionidae butterflies signum complexity correlates with the thickness of the outer covering on the spermatophore (Sanchez et al., 2011, Scriber et al., 1995). However, a recent study on the microstructure of the bursa copulatrix in Tortricidae moths identified pores that were suggested to perform a secretory function (Lincango et al., 2013), 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 copultrix compared to the female thorax were statistically significantly enriched for GO classes related to muscle structure or activity (supplementary table 1), 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 (Benoist, 1998). Similar patterns were found in the reproductive tract of female *D. melanogaster*, in which muscle contraction genes were up-regulated in response to mating (Mack et al., 2006). 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 (Suarez, 2005).

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 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 (Kristensen, 2003). 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), 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). 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 (Castrillon et al., 1993). 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 (Parker, 2006). Such antagonistic sexual coevolution has the potential to contribute to divergence between closely related populations (Arnqvist & Rowe, 2005; Chapman et al., 2003). Proteases and protease inhibitors are two classes of proteins that have been shown to be under positive selection in male reproductive tracts in *Drosophila spp.* (Laflamme & Wolfner, 2013; Wong et al., 2008). 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 (Prokupek et al., 2008; Prokupek et al., 2009; Lawniczak & Begun, 2007; Kelleher et al., 2011). 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 (Ram & Wolfner, 2007; Laflamme & Wolfner, 2013). For example, in *D. arizonae* 12 digestive proteases were specifically expressed in the female reproductive tract and demonstrated signs of positive selection (Kelleher et al., 2011). 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 (Kelleher et al., 2011).

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 4a), as were seven of the proteases found in the bursal gland (Table 4b). 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 tryptase 5 (Table 4a). Tryptase 5 has been shown to decrease male spermatozoa motility in humans and may be involved in fertility (Weidinger, 2003). Another interesting protein, pacifastin-related serine protease

inhibitor, was also found to be up-regulated in E-strain female bursa copulatrix tissues (Table 4a). Pacifastins have been shown to regulate the immune response, reproduction and phase transition in many insects (Breugelmans et al., 2009). 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 (Gaspari et al., 2004). 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 (Laflamme & Wolfner, 2013). Serine proteases have also been linked to sperm activation and immune response (Laflamme & Wolfner, 2013). This is relevant because mating and sperm storage often leads to changes in regulation of immune response in the female reproductive tract (Mack et al., 2006; Prokupek et al., 2009; Baer et al., 2006), 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.

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

Gene Name	Flybase ID	ECB gene	Molecular Function	Homologous Gene Found In
Actin 57B	FBgn0000044	comp26239_c0_seq1	Structural constituent of cytoskeleton	Drosophila similans, Drosophila arizonae, Drosophila melanogaster
Actin 5C	FBgn0000042	comp26239_c0_seq1	Structural constituent of cytoskeleton	Drosophila similans, Drosophila arizonae, Drosophila melanogaster
Actin 87E	FBgn0000046	comp26239_c0_seq1	Expressed in larval muscle	Drosophila similans, Drosophila arizonae, Drosophila melanogaster
NADH dehydrogenase subunit 4	FBgn0262952	comp26239_c0_seq1, comp25774_c0_seq1	NADH dehydrogenase activity	Drosophila similans, Drosophila arizonae, Drosophila melanogaster
Ribosomal protein LP0	FBgn0000100	comp6763_c0_seq1, comp26468_c0_seq1	Structural constituent of ribosome	Drosophila similans
Ribosomal protein 5a	FBgn0002590	comp26375_c0_seq1, comp6724_c0_seq1	Structural constituent of ribosome	Drosophila similans
Aldolase	FBgn0000064	comp13686_c0_seq1	Fructose-bisphosphate aldolase activity	Drosophila similans, Drosophila melanogaster
Alpha spectrin	FBgn0250789	comp13976_c0_seq1	Actin binding	Drosophila similans
Myosin heavy chain	FBgn0026059	comp26288_c0_seq1	ATP binding	Drosophila similans
Alpha tubulin	FBgn0003884	comp26314_c0_seq1	GTP binding	Drosophila similans
Calcium ATPase	FBgn0263006	comp6792_c0_seq1	calcium-transporting ATPase activity	Drosophila similans
Culex quinquefascitaus transport protein	FBgn0021953	comp15624_c0_seq1	Catalytic activity	Drosophila similans
Receptor of activated protein kinase C1	FBgn0000273	comp7107_c0_seq1	ATP binding	Drosophila similans
Elongation factor 2B	FBgn0000559	comp26371_c0_seq1	GTP binding	Drosophila similans, Drosophila arizonae
Heat shock protein 83	FBgn0001233	comp18191_c0_seq1	ATP binding	Drosophila similans, Drosophila melanogaster

Ribosomal protein L3	FBgn0020910	comp26384_c0_seq1	Structural constituent of ribosome	Drosophila similans
V-ATPase	FBgn0027779	comp18376_c0_seq1	Proton-transporting ATPase activity	Drosophila melanogaster
Ribosomal protein L4	FBgn0003279	comp6770_c0_seq1	Structural constituent of ribosome	Drosophila melanogaster
Topomyosin 1	FBgn0003721	comp6703_c0_seq1	Actin binding	Drosophila melanogaster
Protein C kinase 98E	FBgn0003093	comp309823_c0_seq1	ATP binding	Drosophila melanogaster
ATPase	FBgn0013672	comp26423_c0_seq1	Hydrogen-exporting ATPase activity	Drosophila arizonae
Beta Tubulin 56D	FBgn0003887	comp13374_c0_seq1	GTP binding	Drosophila arizonae, Apis melifera
Ubiquitin	FBgn0010288	comp26322_c0_seq1	Ubiquitin thiolesterase activity	Drosophila arizonae

Table 4. Differentially expressed proteases and protease inhibitors between strains. ECB sequences that showed differentially expressed proteases and protease inhibitors between strains (FDR<0.05) for the bursa copulatrix and the bursal gland. A. There were a total of 34 ECB sequences that showed homology to proteases or protease inhibitors that were also differentially expressed in the bursa copulatrix. Of these 7 were up-regulated in the Z-strain. B. There were a total of 17 ECB sequences that showed homology to proteases or protease inhibitors that were also differentially expressed between strains in the bursal gland. Of these 7 were up-regulated in the Z strain.
A.

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

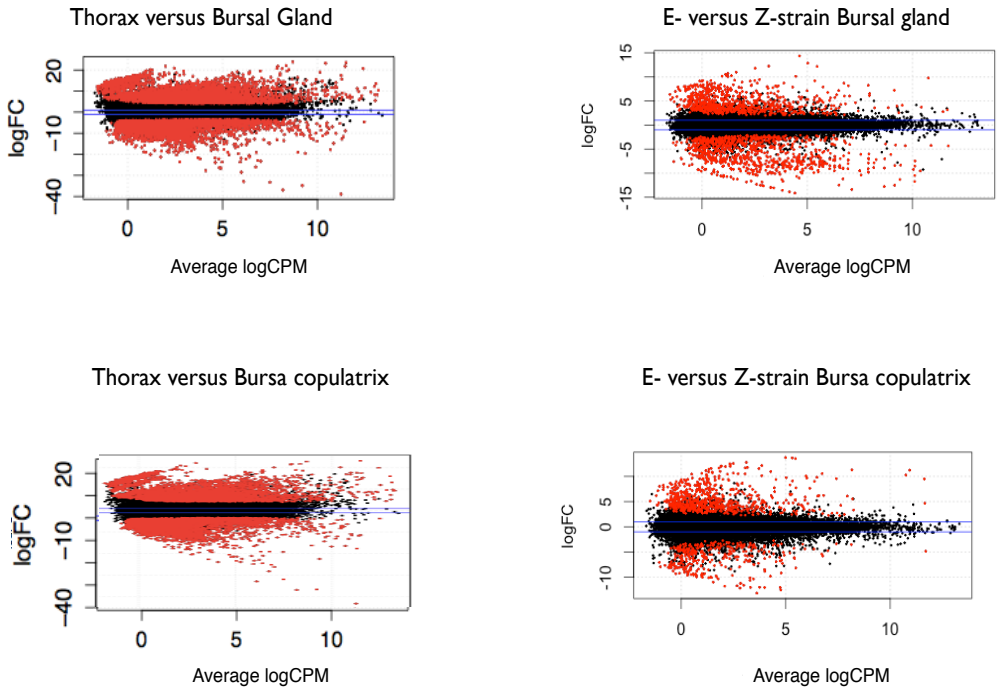
72215	serine protease	up in E	serine-type endopeptidase activity
20820	prism serine protease inhibitor 1	up in E	protease inhibitor
83118	secreted trypsin-like serine protease	up in E	serine-type endopeptidase activity
295616	angiotensin converting enzyme	up in E	metallopeptidase activity
14604	chymotrypsin-like 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
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
4699	astacin-like metalloendopeptidase	up in E	peptidase activity
9559	trypsin like protein	up in Z	peptidase activity

B.

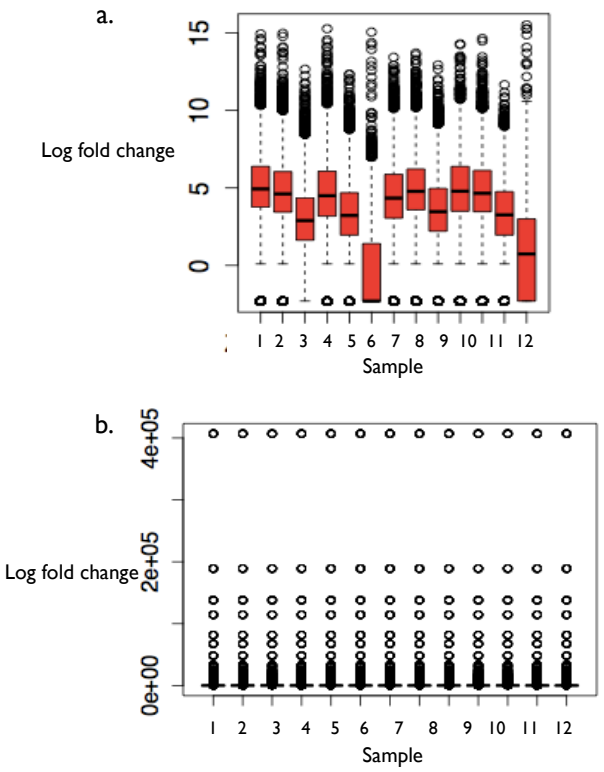
Bursal gland sequence name	Homologous protein	Direction of differential expression*	Predicted function
15139	retinoid-inducible serine carboxypeptidase-like	up in E	serine-type carboxypeptidase activity
144499	clip domain serine protease 11 precursor	up in E	serine-type endopeptidase activity
32437	wap 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

Supplementary Material

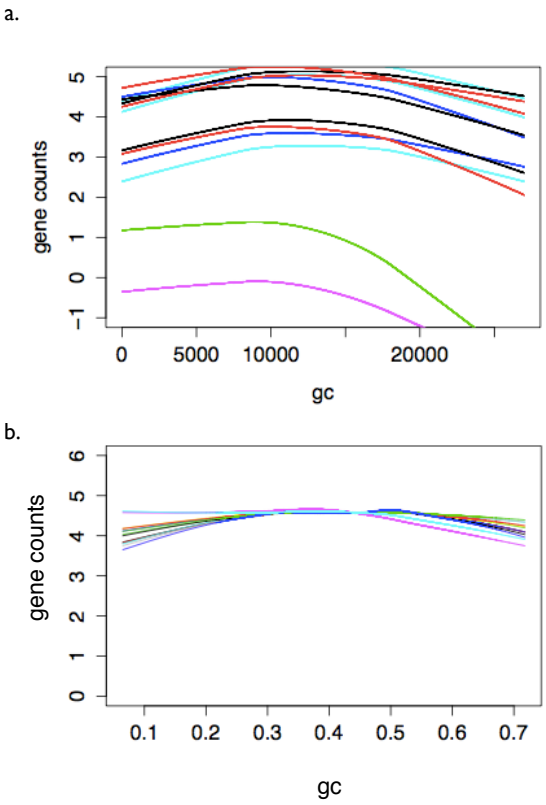
Supplementary Figure 1. MA plots for each comparison of interest. These plots shows the tagwise log-fold-change against the log-counts per million for each gene in a tissue library. Each dot on the graph represents an individual gene. The blue lines across each plot represent a 2 fold change in expression. All red points show differentially expressed genes with a FDR < 0.01 and all black dots are genes that were not significantly differentially expressed.



Supplementary Figure 2.**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.



Supplementary Figure 3. GC content normalization. Lowess regression of (a.) non-normalized and (b.) normalized GC content.



Supplementary Table 1.**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.

Bursal Gland GO Term	**Bursal Gland GO Term Continued**	**Bursa copulatrix GO Term**	**Both GO Terms**
actin binding	mitochondrial inner membrane	actomyosin structure organization	actin cytoskeleton organization
actin cytoskeleton	mitochondrial membrane part	cellular component organization or biogenesis at cellular level	actin filament-based process
adherens junction	mitochondrial respiratory chain	cellular respiration	ATP synthesis coupled electron transport
anatomical structure development	mitochondrial respiratory chain complex I	contractile fiber	cellular component organization or biogenesis
anatomical structure formation involved in morphogenesis	morphogenesis of an epithelium	contractile fiber part	cellular macromolecular complex subunit organization
anatomical structure morphogenesis	morphogenesis of embryonic epithelium	DNA metabolic process	cytoplasm
anchoring junction	multicellular organism reproduction	electron transport chain	cytoplasmic part
biological regulation	multicellular organismal development	energy derivation by oxidation of organic compounds	cytoskeleton organization
cell	multicellular organismal reproductive process	head segmentation	gene expression
cell development	NADH dehydrogenase complex	I band	generation of precursor metabolites and energy
cell differentiation	negative regulation of	intracellular non-membrane-bounded	

	biological process	organelle	intracellular
cell migration	negative regulation of cellular process	mitochondrial ATP synthesis coupled electron transport	intracellular organelle
cell motility	nucleus	mitochondrial matrix	intracellular organelle part
cell part	organ development	mitochondrial part	intracellular part
cellular component organization	organ morphogenesis	mitochondrion	macromolecular complex
cellular component organization at cellular level	organelle envelope	muscle cell differentiation	NADH dehydrogenase (quinone) activity
cellular component organization or biogenesis at cellular level	organelle inner membrane	myofibril	NADH dehydrogenase (ubiquinone) activity
cellular developmental process	oxidoreductase activity, acting on NADH or NADPH	myofibril assembly	NADH dehydrogenase activity
cellular process	post-embryonic morphogenesis	non-membrane-bounded organelle	organelle
cellular process involved in reproduction	post-embryonic organ morphogenesis	respiratory electron transport chain	organelle organization
cellular protein metabolic process	protein complex	ribosome	organelle part
cytoskeletal protein binding	protein localization	sarcomere	oxidative phosphorylation
developmental process	protein modification by small protein conjugation	sarcomere organization	oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as

			acceptor
developmental process involved in reproduction	protein modification by small protein conjugation or removal	striated muscle cell differentiation	
dorsal closure	protein modification by small protein removal	structural constituent of ribosome	
embryonic morphogenesis	protein transporter activity	structural molecule activity	
establishment or maintenance of cell polarity	regulation of biological process	translation	
gamete generation	regulation of cellular component organization	Z disc	
germ cell development	regulation of cellular process		
Golgi apparatus	respiratory chain		
imaginal disc development	respiratory chain complex I		
imaginal disc morphogenesis	rhabdomere development		
initiation of dorsal closure	sensory organ development		
instar larval or pupal development	sexual reproduction		
instar larval or pupal morphogenesis	small GTPase mediated signal transduction		
intracellular membrane-bounded	system development		

organelle			
localization	tissue development		
localization of cell	tissue morphogenesis		
macromolecule localization	tube development		
membrane-bounded organelle	tube morphogenesis		
metamorphosis			
mitochondrial envelope			

Chapter 3. *Molecular dissection of nuptial gifts in divergent strains of Ostrinia moths*

ABSTRACT

Seminal fluid proteins (SFPs) produced in the male accessory glands and ejaculatory duct are subject to strong sexual selection, often evolve rapidly, and therefore may play a key role in reproductive isolation and species formation. However, little is known about reproductive proteins for species in which males transfer ejaculate to females using a spermatophore package. By combining RNA sequencing and proteomics, we characterize putative SFPs, identify proteins transferred in the male spermatophore, and identify candidate genes contributing to a one-way gametic incompatibility between Z and E strains of the European corn borer moth (*Ostrinia nubilalis*). We found that the accessory glands and ejaculatory duct secrete over 200 highly expressed gene products, including peptidases, peptidase regulators, and odorant binding proteins. Comparison between *Ostrinia* strains revealed that accessory gland and ejaculatory duct sequences with hormone degradation and peptidase activity were among the most extremely differentially expressed. However, most spermatophore peptides lacked reproductive tissue bias or canonical secretory signal motifs, and roughly ¼ may have been produced elsewhere before being sequestered by the male accessory glands during spermatophore production. In addition, most potential gene candidates for postmating reproductive isolation did not meet standard criteria for predicted SFPs and nearly three quarters were novel, suggesting that both postmating sexual interactions and gametic isolation likely involve

molecular products beyond traditionally recognized SFPs.

Introduction:

Sexual selection can act as a powerful force shaping the evolutionary trajectory of traits and helping to drive speciation (Panhuis et al., 2001). Divergence in sexual traits may often occur by intersexual selection and by positive or negative correlated evolution of male and female traits (including sexual conflict: (Parker & Partridge, 1998; Arnqvist et al., 2000; Gavrilets, 2014; Gavrilets & Hayashi, 2005; Chapman et al., 2003). For example, after mating, male seminal fluid proteins from male ejaculate interact within the female reproductive tract with the ejaculates of other males and with female reproductive proteins (Simmons, 2001; Ram & Wolfner, 2007; Wolfner, 2009; Sirot et al., 2009). These interactions are thought to drive the rapid evolution of such reproductive proteins in mammals (Makalowski & Boguski, 1998), marine gastropods (Metz et al., 1998), and insects (Swanson, 2004; Swanson & Vacquier, 2002; Swanson et al., 2001). Rapid molecular evolution increases the likelihood of male-female protein incompatibilities between populations as a byproduct, increasing the potential for failed reproduction and reduced fertilization success (Price et al., 2001; Larson et al., 2011).

The evolution of postmating interactions between the sexes has been recognized as an important factor for reproductive isolation. In *Drosophila simulans* and *D. mauritiana* fruit flies, sperm competition among multiple males for fertilizations has been linked to postmating, prezygotic reproductive isolation (Manier et al., 2013). Cryptic female choice, in which female exert control over

sperm storage and usage, has also been implicated in postmating, prezygotic reproductive isolation in salmon and trout, where ovarian fluid inhibits conspecific sperm from fertilizing females eggs (Yeates et al., 2013). Although reproductive proteins are likely mediators of both sperm competition and cryptic female choice, their potential role in reproductive isolation is just beginning to be explored.

To understand postmating sexual selection and how this may contribute to species divergence, male reproductive proteins need to be characterized for diverse organisms. In many insects the male ejaculate is often complex and along with sperm, contains sugars, salts, defensive compounds, and seminal fluid proteins (SFPs) (Chen, 1984; Gillott, 2003; Simmons, 2001; Poiani, 2006). SFPs are produced in and secreted by the male accessory glands and male ejaculatory duct and transferred to the female in the male ejaculate (Chen, 1984). These proteins have been shown to exert profound effects on female behavior and physiology after mating. In *D. melanogaster*, SFPs have been shown to decrease female receptivity to future mates, increase egg laying, decrease longevity, increase feeding, and decrease sleep (Ram et al., 2005; Sirot et al., 2009; Wolfner, 2009).

In numerous insects, males transfer their sperm packaged into a complex spermatophore manufactured by male accessory glands. This sperm-containing packaging adds an additional level of complexity to male and female reproductive interactions and conflict (Chapman, 2003; Lewis & South, 2012; Lewis et al., 2014). In Lepidoptera, where spermatophore transfer is ubiquitous (Wedell,

2005), the male spermatophore can act as a type of mating plug that temporarily prevents females from re-mating (Sugawara, 1979). During mating, females receive the spermatophore inside a muscular organ called the bursa copulatrix (henceforth, the bursa). Distension of the bursa stimulates stretch receptors, which cause females to become unreceptive to additional mating (Sugawara 1979); therefore, a female will not remate until the spermatophore is partially degraded. Lepidopteran females have evolved several traits that function in spermatophore degradation, including female reproductive secretions (Al-Wathiqui et al., 2014; Meslin et al., 2015) and a muscular reproductive tract containing a chitinized structure called the signum (Cordero, 2009; Galicia et al., 2008; Sánchez et al., 2011). As these traits are expected to evolve rapidly in response to sexual conflict, they may cause variation in fertilization success and contribute to reproductive isolation between populations (Parker & Patridge, 1998; Chapman et al., 2003; Gavrilets & Hayashi, 2005).

A few recent studies have investigated gene expression in the female bursa and identified secretions that contribute to degrading the spermatophore (Al-Wathiqui et al., 2014; Meslin et al., 2015) . However, specific male reproductive genes and proteins may interact with the bursa to accomplish or inhibit spermatophore degradation but these products have yet to be fully characterized. One previous study examining seminal proteins in two *Heliconius* butterfly species, *H. erato* and *H. melpomone*, found several secreted proteases and cysteine-rich proteins (Walters & Harrison, 2010). Based on their similarity to *D. melanogaster* SFPs, these male proteins may be responsible for mediating

physiological and behavioral changes in females after mating (Wolfner, 2009; Laflamme & Wolfner, 2013).

Here, we characterize male reproductive proteins in two divergent strains of the European corn borer (ECB, *Ostrinia nubilalis*), a species for which female reproductive genes have previously been characterized (Al-Wathiqui et al., 2014). Z and E strains differ in sex-pheromone communication and are estimated to have diverged approximately 100,000 years ago (Malausa, 2005; Dopman et al., 2010). Of seven strong reproductive isolating barriers between strains, one acts during the postmating, prezygotic (PMPZ) time point and consists of reduced and asymmetric oviposition. Specifically, Z-strain females that mate with E-strain males experience a 30% reduction in egg laying; however, the percentage of fertilized eggs is not affected (Dopman et al., 2010). This PMPZ reproductive barrier may arise through dysfunctional interactions at any stages preceding egg deposition (i.e. sperm storage, egg maturation, fertilization, and oviposition) and could involve interactions between known female reproductive products (Al-Wathiqui et al., 2014) and SFPs or other products produced by the male accessory glands. In this study, we compare males from these two ECB strains to: 1) characterize seminal fluid proteins transferred to females and 2) identify candidate genes that may be involved in the PMPZ isolating barrier acting in ECB.

Methods:

Moth strains

Z and E strain European corn borer moths were collected as caterpillars, pupae, and adults from New York State, USA (n = 500 males and n = 500 females each from East Aurora, Geneva and Bouckville). Laboratory populations have been maintained by mass rearing ~200 adults per generation and have been previously described (Dopman et al., 2005).

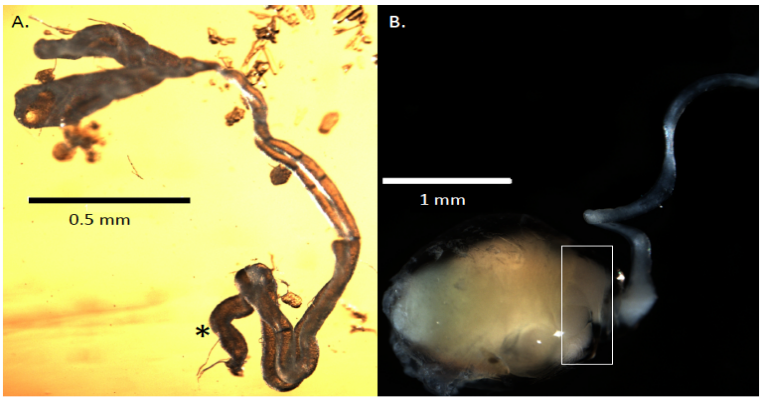


Figure 1. Male reproductive structures of European corn borer moths. A. The male accessory gland (upper left) shown with a small portion of the ejaculatory duct (denoted by *). These structures secrete proteins transferred to the female in a sperm-containing package known as the spermatophore. B. Male spermatophore provided by a Z-strain ECB male, dissected from the female reproductive tract immediately after mating. The sperm-containing ampulla, indicated by the white box, contains sperm, while the composition of the remaining spermatophore is unknown.

Sample Preparation and RNA Sequencing

A total of 18 accessory glands and 18 ejaculatory ducts were dissected from two-day-old virgin males for each strain (Figure 1a). Tissues were pooled from 6 individuals and used to create each of the six accessory gland and six ejaculatory duct libraries (three per strain). Following Al-wathiqui et al. 2014, tissues were stored in RNAlater at -80˚C prior to total RNA extraction (RNeasy Midi kit, Qiagen, California) and cDNA library construction (Illumina Truseq

RNA sample preparation kit v2, San Diego, CA) (Al-Wathiqui et al., 2014). The six accessory gland cDNA libraries and six ejaculatory duct cDNA libraries were then sequenced using an Illumina HiSeq2000 (50bp single-end reads). To compare patterns across reproductive and non-reproductive tissues, two male thorax libraries, one per strain, were also sequenced for each strain. Male thorax cDNA libraries were created using the SMART cDNA library 6.7 protocol (Takara Bio, Otsu, Shiga, Japan) and sequenced using 40 bp paired-end reads on an Illumina GA IIx, with a 200 bp insert length.

Sequence Processing and Assembly

After sequencing, Trimmomatic v0.17 was used to cleave Illumina sequencing adapters and remove the first and last five base pairs from each read as these usually have low sequencing quality scores (Lindgreen, 2012). After adapter and quality trimming, only reads ≥ 36 bp were retained. Following trimming, mtRNA and rRNA contamination was removed from each library using the short read aligner, Bowtie 2 and the "very-sensitive" setting (Langmead & Salzberg, 2012; Langmead et al., 2009). We then used FASTQ/A Collapser to reduce the number of identical sequences in our read libraries without affecting read counts (http://hannonlab.cshl.edu/fastx_toolkit/).

After all quality control steps were completed, we used Trinity version r2013-02-25 to assemble the six male accessory gland libraries, six male ejaculatory duct libraries, and two male thorax libraries into a single *de novo* transcriptome using recommended settings (Grabherr et al., 2011). After

assembly, redundancy was reduced at each locus by selecting the longest transcript. To identify putative homologs to known proteins in other organisms, a Blast search was conducted between each sequence and the entire NCBI non-redundant protein database. All sequences with significant Blast hits (e-value $\leq 10^{-3}$) were then mapped and annotation scores were computed for all possible gene ontology terms using Blast2GO (Conesa et al., 2005; Conesa & Götz, 2008; Götz et al., 2008).

RNA-Seq

To account for differences in RNA-seq library preparation between thorax tissue and male reproductive tissue, all libraries were normalized prior to read mapping and differential expression using EDAseq (Risso et al., 2011). EDAseq allows for within-lane normalization to account for differences in read lengths and nucleotide bias, which could result from using different preparation methods that differ in the number of PCR amplifications, and between-lane normalization to account for differences in sequencing depth across sequencing lanes.

To conduct differential expression analysis, normalized reads were mapped to our transcriptome using Bowtie 2 and the "very-sensitive" setting (Langmead & Salzberg, 2012; Langmead et al., 2009), and then the normalized read counts for each gene were tested for differential gene expression using edgeR (Robinson & Oshlack, 2010; Robinson et al., 2009; Robinson & Smyth, 2007; McCarthy et al., 2012). edgeR uses empirical Bayes methods to estimate gene-specific variation. We used a generalized linear model approach in which we

assessed differential expression with ECB strain and tissue type as factors. For all comparisons, genes were considered significantly differentially expressed if they had a false-discovery rate (FDR) of < 0.01 and an expression $\log_2$ fold change (LFC) of ≥ 2.

We applied two commonly used criteria to identify putative male SFPs (Sirot et al., 2008; Walters & Harrison, 2010). First, we looked for genes with up-regulated gene expression in reproductive tissue compared to thorax. Second, we looked for the presence of secretory peptide motifs, which indicate that a protein is secreted from the cell it is produced in and possibly transferred to females during copulation. To identify secretory peptides, we created a protein database using ESTscan to estimate protein sequences from our *de novo* transcriptome (Iseli et al., 1999). The resulting protein sequences were then analyzed using SignalP 4.0 to detect the presence of signal peptides (Petersen et al., 2011).

Proteomics

A proteomics approach was used to identify reproductive proteins in Z and E strain male spermatophores. Spermatophores were collected from females immediately after mating. Mating trials consisted of placing two males and one female of the same strain in a paper cup (10 cm diameter). Mating cups were kept in an incubator during the dark cycle and monitored every 15 minutes for signs of mating (16L:8D, 27°C, 70% R.H.). When mating occurred, the mating pair was monitored and the spermatophore was dissected out of the female immediately

after mating stopped. Six spermatophores were obtained from males of each strain.

After spermatophore collection, the sperm containing ampulla, and the spermatophore tail were removed (Figure 1b) and the remaining tissue was stored at -80°C. All six spermatophores were pooled to create a single sample per strain. Spermatophores were crushed in SDS using a mortar and pestle, and centrifuged briefly. The supernatant was collected, and boiled for 5 minutes. Each sample was then run on a 10% Tris-Glycine gel just until the entire sample migrated out of the well and the resulting two bands (one per strain) were fixed and stained with Coomassie brilliant blue, cut from the gel and sent to the Harvard Mass Spectrometry and Proteomics Resource Laboratory (Cambridge, MA). Samples were enzymatically digested and analyzed using a HPLC coupled with nanoelectrospray tandem mass spectrometry on a Thermo Fisher LTQ-Orbitrap mass spectrometer. The resulting spectra were annotated with Proteome Discoverer software (Thermo Scientific, NY) using the combined UniProt Swiss-Prot database. For the remaining un-annotated peptide sequences, we used a Blast search against our male reproductive tissue transcriptome to further characterize and annotate these sequences.

Candidate speciation genes

To identify male reproductive genes that could be involved in a reproductive isolating barrier, we looked for genes that were significantly differentially expressed between ECB strains. First, we determined which of our

putative SFP sequences were differentially expressed between E and Z strains (LFC ≥ 2, FDR ≤ 0.01). We also characterized differential expression between strains for other sequences expressed in the accessory glands and ejaculatory duct. We chose to look for differentially expressed sequences that we did not predict to be SFPs because differentially expressed genes that are not specific to or secreted from the reproductive tract could still play a role in reproductive isolation (Findlay et al., 2008). For this comparison only sequences that were significantly upregulated in male accessory gland or ejaculatory duct tissue by a LFC ≥ 10 and an FDR ≤ 0.01 were considered. This more stringent cutoff was used to focus our analysis to non-SFP genes that are highly differentially expressed between strains, as these may be more likely to be involved in reproductive isolation.

Results:

Male Transcriptome Assembly

From both Z and E strains, we obtained ~400,000,000 raw single end reads from the accessory glands and ejaculatory ducts, and ~60,000,000 paired-end reads from male thorax tissue. The resulting assembly contained 46,771 loci with a relatively large N50 score of 7,351 bp, a minimum sequence length of 201 bp, and a maximum sequence length of 18,129 bp. We reduced the complexity of this assembly by selecting the longest transcript at each locus, resulting in 44,285 sequences in the final transcriptome. Of these, 46.3% (23,761, e-value ≤0.001) showed similarity to sequences in other organisms and 21.4% (9,491) were successfully annotated with gene ontology information.

Putative SFPs

Compared to the male thorax, 4,877 and 1,424 transcripts were significantly up-regulated in the ejaculatory duct and accessory gland, respectively (LFC ≥ 2 and FDR ≤ 0.01). Of transcripts with reproductive tissue bias, 217 also contained a secretory signal and thus were considered putative SFPs (Figure 2, Supplementary Table 1). More than half (119/217) were expressed in both the accessory gland and ejaculatory duct. Of those showing tissue specificity, ≈23% (50/217) were primarily expressed in the accessory gland and ≈22% (47/217) were primarily expressed in the ejaculatory duct.

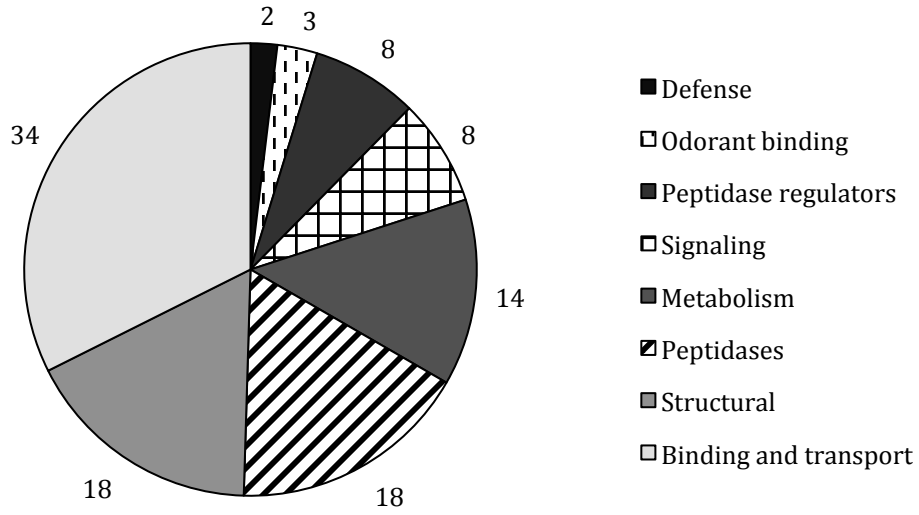


Figure 2. Gene Ontology category distribution for putative SFPs in *O. nubilalis*.

We were able to identify annotation terms for roughly two thirds (146/217) of putative SFPs. Gene products included peptidases and peptidase regulators (Figure 2, Supplementary Table 1). Peptidases differed in their hydrolysis mechanisms (Polgar, 1989) and were identified as metallopeptidases

(Figure 2, Supplementary Table 1), cysteine peptidases (Figure 2, Supplementary Table 1), and serine peptidases and regulators (Figure 2, Supplementary Table 1). We were also able to identify genes with activities associated with binding ligands such as pheromones and odor molecules (Leal, 2013). We identified two putative SFPs with odorant binding activity (comp7771_c0_seq1 and comp21759_c0_seq1). We also found one pheromone binding protein that was up-regulated in the accessory glands compared to the male thorax (comp18104_c0_seq1).

Proteomics

We identified 93 Z-strain spermatophore proteins and 24 E-strain spermatophore proteins (Table 1). Four putative SFPs with various activities had peptide counterparts: peptidyl-prolyl cis-trans isomerase 5 (comp17755_c0_seq1), protein disulfide isomerase (comp20773_c0_seq1), a carboxypeptidase (comp22488_c0_seq1), and sphingomyelin phosphodiesterase (comp8052_c0_seq1).

In the Z strain, we identified 8 peptidases and one peptidase regulator that are transferred to the female in the spermatophore. These consisted of two serine peptidases called vitellin-degrading protease (comp237626_c0_seq1 and comp20483_c0_seq1), which are responsible for degrading vitellin found in insect eggs (Chapman, 2003; Ikeda et al., 1990). The Z strain spermatophore also contained a cysteine protease, cathepsin L (comp311204_c0_seq1). Cathepsin L

is a lysosomal endopeptidase that can be secreted and interacts with structural proteins, such as collagen and fibronectin (Ishidoh & Kominami, 1995).

Three proteins involved in insect immunity were found in the Z strain spermatophore. These included one sequence encoding a lysozyme (comp22897_c0_seq1), defense protein hdd11 (comp25125_c0_seq1), and two phenoloxidase subunits (comp22426_c0_seq1, subunit 1 and comp21505_c0_seq1, subunit 2; Table 1). Both lysozymes and defense protein hdd11 are antibacterial proteins. Lysozymes are glycoside hydrolases that damage bacterial cell walls (Ellison, 1991), however the precise function of defense protein hdd11 is unknown (Bao, 2003). Prophenoloxidases are a key component of the insect melanogenesis defense response, in which pathogens are encapsulated by melanin (González-Santoyo & Cordoba-Aguilar, 2011).

In general, fewer proteins were identified in the E strain spermatophore (Table 1). However, like Z strain spermatophores, we identified proteases and defense proteins that are transferred to the female. These included cathepsin-L (comp311204_c0_seq1) and two proteins that are involved in the insect immune response. The two immune response proteins included phenoloxidase (comp22426_c0_seq1) , which was also found in Z-strain spermatophores, and dermcidin (P81605), an anionic antimicrobial protein in humans (Schittek et al., 2001; Narayana & Chen, 2015).

Differential expression

Our analysis of between-strain expression patterns first focused on the 217 genes that we identified as putative SFPs. In the accessory gland, we found 4

sequences that were differentially expressed (LFC ≥ 2) between males of the Z and E strains (Supplementary Table 2). None of these sequences were successfully annotated. In the ejaculatory duct, 3 sequences were differentially expressed between strains (LFC ≥ 2), and 1 was functionally annotated: comp1865_c0_seq1 shows ionotropic glutamate receptor activity and was up-regulated in the E-strain (LFC = –3.8; Supplementary table 2).

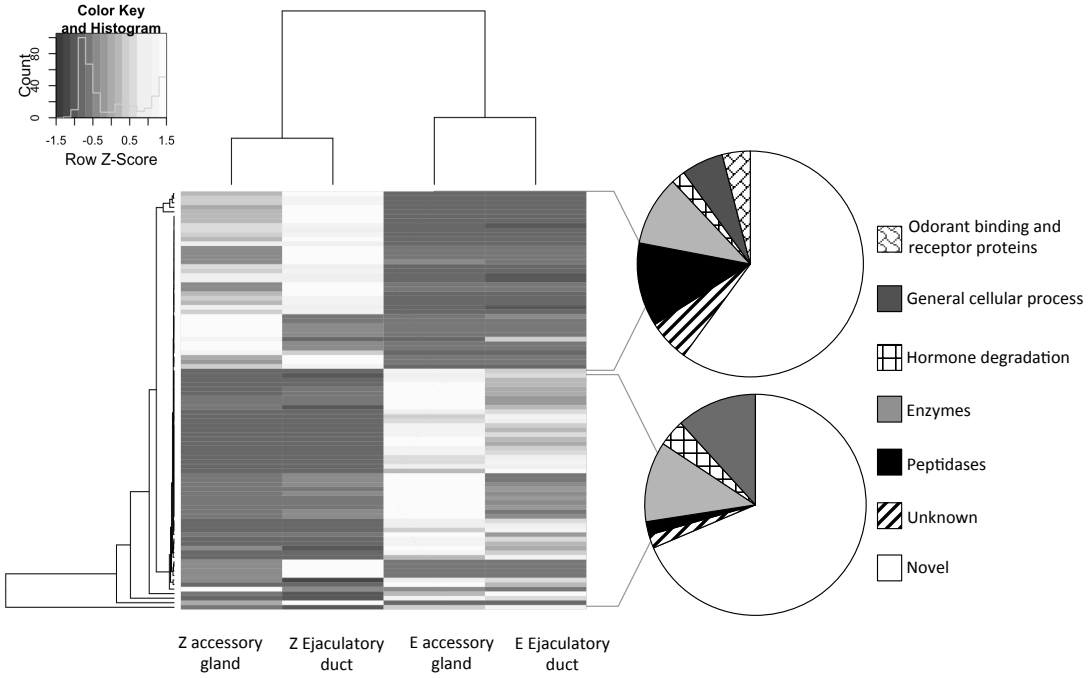


Figure 3. Reproductive genes that are differentially expressed between *Ostrinia nubilalis* strains and their gene ontology categories. The heat map represents normalized read counts converted to z-scores for each tissue type and strain. Lighter shading indicates higher expression level. Gene ontology pie charts summarize functional terms for genes up-regulated in at least one reproductive tissue in the Z-strain (top) or the E-strain (bottom).

We identified 1,361 ejaculatory duct transcripts and 1,550 accessory glands transcripts that were highly up-regulated in these tissues compared to male thorax (LFC ≥ 10 and FDR ≤ 0.01), but they lacked canonical secretory motifs so were not classified as SFPs (Figure 3 & Supplementary Table 2). Of these, 48 sequences were significantly differentially expressed between strains in the

ejaculatory duct, 41 were highly differentially expressed between strains in the accessory gland, and 3 were differentially expressed in both tissues (LFC ≥ 2, FDR ≤ 0.01). Genes that were up-regulated in the Z-strain showed significant homology to proteases, odorant binding proteins and odorant receptors (Figure 3). There were five serine peptidases that were significantly up-regulated in the Z-strain compared to the E-strain in the male reproductive tissues with LFC between 10.0 and 12.5. Genes that were up-regulated in the E-strain showed significant homology to proteins related to juvenile hormone degradation (Figure 3 & Supplementary Table 2). In both strains, the majority ($59/99 \approx 60\%$) of differentially expressed genes were novel and lacked hits to known genes.

Discussion:

The female reproductive tract is the primary venue for postmating sexual selection and gametic isolation, but our understanding of the specific mechanisms underlying postmating interactions is still limited. Much research has emphasized the transfer of male SFPs to females for these interactions, and because many SFPs rapidly evolve they make good candidates for reproductive isolation. However, male ejaculates are complex and known to contain diverse compounds in addition to SFPs (Poiani, 2006; Perry et al., 2013). To disentangle this complexity, we profiled the transcriptome of SFP-producing tissues as well as the proteome of ejaculate that males transfer to females. We found that a large number of putative SFPs are produced by ECB males (217) and they are similar in number and composition to other species (Avila et al., 2011; Bonilla et al., 2016).

However, most SFPs did not appear the male spermatophore (4/217 ≈ 2%). ECB spermatophores contained an additional 108 proteins that included peptidases, peptidase regulators, anti-microbial proteins, and odorant binding proteins. Some of these spermatophore proteins appear to be derived from non-reproductive tissue (28% lacked matches to reproductive tissue transcripts) or were possibly broadly expressed (16% were biased towards non-reproductive tissue), while another 20% show reproductive tissue bias yet lack classical SFP secretory signals. Furthermore, while most of the 99 candidate genes for postmating isolation between ECB strains were novel and thus may be rapidly evolving like SFPs (59/99 ≈ 60%), only seven gene candidates were actually recognized as SFPs. More work is needed to confirm these broad trends, but our work on moths join recent findings in crickets (Bonilla et al. 2016) and flies (Findlay et al. 2008) in suggesting a nuanced and possibly limited role of traditionally defined SFPs as mediators of postmating sexual selection and gametic reproductive isolation.

ECB male seminal fluid proteins

Peptidases and peptidase regulators are a conserved class of proteins that have been found in the seminal fluid of many taxa, from humans to arthropods (Laflamme & Wolfner, 2013; Avila et al., 2011; Gillott, 2003). In insects, they are known to play important roles in postmating processes. Serine peptidases and inhibitors in *D. melanogaster* are responsible for SFP transfer (aquarius) (Findlay et al., 2014), full induction of egg laying (seminase), and semen liquefaction (protein C inhibitor) (Pilch & Mann, 2006). Metallopeptidases have also been

shown in fruit flies to be involved in induction of egg-laying, sperm storage (Ram et al., 2006), and immune response following bacterial infection (Mueller et al., 2007).

In ECB, we found that males transfer serine peptidases, metallopeptidases and serine peptidase inhibitors to females in their ejaculate. Our list of putative SFPs included nine genes that were annotated as serine peptidases and 5 genes annotated as serine peptidase inhibitors (Figure 2, Supplementary Table 1). Although we were not able to identify serine peptidases and inhibitors in the male spermatophore, we did identify a zinc metalloprotease and vitellin-degrading proteases that are transferred to females in the Z strain spermatophore (Table 1). A zinc metallopeptidase in *D. melanogaster*, encoded by *CG11864*, processes two other male SFPs, sex peptide and Acp36DE, after all three proteins have entered the female reproductive tract (Ram et al., 2006). Vitellin-degrading proteases were characterized in the silk moth, *Bombyx mori*, as cleaving two different yolk proteins during embryogenesis, and in sea squirts and sea urchins the acrosome reaction is assisted by two vitellin envelope degrading proteases (Kim et al., 2008). In ECB, these proteins may be transferred in the spermatophore to mediate the acrosome reaction between male sperm and the female egg.

In humans and fruit flies, males also transfer cysteine proteases to females in their ejaculate (LaFlamme and Wolfner, 2013). We identified cathepsin L, a cysteine peptidase, in the spermatophores of both Z and E-strain males (Table 1). The function of cathepsins in reproduction is poorly understood; however, these proteins have been shown to play important roles in both digestion and egg

maturation. For example, in the cowpea weevil, *Callosobruchus maculatus,* cathepsin L is responsible for digestive proteolysis in the gut (Pedra et al., 2003). In mice, cathepsin L has been shown to be up-regulated in pre-ovulatory egg follicules, where it is predicted to play a role in follicular rupture (Robker et al., 2000). Cathepsins in the ECB spermatophore may potentially act to mediate male-female sexual conflict by cleaving and disabling the female reproductive proteins or by aiding in post-mating egg maturation.

We also identified two SFPs that were annotated as odorant binding proteins (comp7771_c0_seq1 and comp21759_c0_seq1). Odorant binding proteins are mainly associated with olfaction; however they have now been found in male ejaculate of many different insects (Findlay et al., 2008; South et al., 2011; Sirot et al., 2008; Kelleher et al., 2009; Chapman, 2008). These proteins are small, soluble proteins that are found in high concentrations around olfactory-receptor neurons and are thought to bind odors and shuttle them to neurons for sensing (Hekmat-Scafe et al., 2002). The role of these molecules in reproduction remains poorly understood, but in *Helicoverpa armigera* and *H.assulta*, odorant binding proteins transferred to females in the male ejaculate and found in fertilized eggs act as an oviposition deterrent for other females (Sun et al., 2012). These proteins may serve a similar function in ECB, where larval cannibalization is common and such deterrence could enhance survival (Breden & Chippendale, 1989).

Divergence of genes between European corn borer strains

In the presence of gene flow, regions of the genome involved in reproductive isolation or recurrent local adaptation will tend to remain differentiated while other gene regions become homogenized (Dopman et al. 2005, Ting et al. 2000; Harrison 2012). Hence, if genes encoding reproductive proteins have played an important role in speciation, greater sequence and/or regulatory differentiation is expected compared to genes encoding other proteins expressed in the same tissue.

We focused on identifying candidate genes that could be involved in postmating reproductive isolation between ECB strains based on patterns of differential regulation, and found five serine peptidases were up-regulated in the Z-strain ejaculatory duct compared to the E-strain ejaculatory duct (Supplementary Table 2). Peptidases and peptidase regulators are particularly attractive targets for further characterization as mediators of reproductive isolation because of their involvement in multiple postmating interactions between the sexes (Laflamme & Wolfner, 2013). In the *Allonemobius* complex of crickets, one trypsin-like serine protease called *Ejac-SP* has been shown to mediate a postmating, prezygotic barrier, in which heterospecific males fail to induce egg-laying after mating (Marshall et al., 2009). Proteases that were differentially expressed between ECB strains may play a similar role in induction of egg laying. Specifically, the 30% reduction in egg laying previously documented (Dopman et al., 2010) when Z-strain females mate with E-strain males could result from inability of E-strain males to activate relevant pathways controlling egg-laying in Z-strain females.

In Z-strain male accessory glands, we found up-regulation of a gene annotated as juvenile hormone (JH) diol kinase (comp16883_c0_seq1; Supplementary Table 2), a protein involved in the JH degradation pathway. JH is an essential insect hormone controlling a myriad of processes including inhibiting pheromone production after mating and stimulating production and uptake by oocytes of the egg yolk precursor, vitellogenin (Hartfelder, 2000; Cusson & McNeil, 1989). In *Aedes* mosquitoes a reduction in female JH titer is required for completion of oogenesis (Shapiro et al., 1986). Thus, up-regulation of JH diol kinase in the accessory glands of Z-strain ECB males may facilitate oogenesis by reducing female JH titers. Additional studies are needed to determine how these differentially expressed genes influence ECB reproduction.

Finally, we identified an odorant binding protein secreted from the male ejaculatory duct (comp11571_c0_seq1; Supplementary Table 2) that was up-regulated in Z-strain compared to E-strain males. However, additional work is needed to elucidate the function of odorant binding proteins in reproduction and species divergence.

Conclusion:

To understand how reproductive proteins contribute to sexual selection and reproductive isolation, we must characterize various organisms across the speciation continuum, from diverged populations to species. Here, we found evidence for proteases along with genes encoding proteins involved in egg development and egg production as possible molecular drivers of postmating isolation in diverged pheromone strains of moths. Peptidases, peptidase

regulators, and odorant binding proteins were found among the more than 200 gene products that had traditional signals of SFPs in male accessory glands and ejaculatory duct. Finally, proteomic analysis of male spermatophores revealed similar functions as reproductive tissue, despite limited overlap of constituent SFPs. Overall, our results suggest that gametic isolation and postmating sexual selection may involve molecular products beyond traditionally recognized SFPs.

Table 1. *O. nubilalis* spermatophore proteins and their functions identified using LC-MS/MS.

Strain	Sequence ID*	Sequence length (# of aa)	Score	Protein annotation	LFC relative to thorax**	Secretion signal
Z						
	Peptidases					
	comp27218_c0_seq1 - AAO21504.1	803	44.28	metalloendopeptidase	4.77	N
	comp29124_c0_seq2 - JF339040.1	929	86.42	aminopeptidase	<1.00	Y
	comp237626_c0_seq 1- XP_ 013186764.1	281	28.87	vitellin-degrading protease precursor	-8.70	Y
	comp22787_c0_seq1 - XP_ 013181289.1	437	6.86	zinc metalloproteinase nas-14-like	10.30	N
	comp22488_c0_seq1 - NP_ ADU33188.1	**468**	**7.64**	**carboxypeptidase**	**16.60, 11.32**	**N**
	comp20483_c0_seq1 - XP_ 013186764.1	288	5.32	vitellin-degrading protease precursor	<1.00	Y
	comp311204_c0_seq 1*- XP_ 013201208.1	265	49.15	Cathepsin L	-10.76, -8.21	N
	Q9HWK6*	462	37.46	Lysyl endopeptidase	-	-
	Peptidase regulators					
	comp13866_c0_seq1 - AFV46312.1	401	2.36	serpin 1	<1.00	Y
	Immune response					
	comp22897_c0_seq1 - ADU33188.1	162	6.77	lysozyme	<1.00	Y
	comp25125_c0_seq1 - NP_ AGV28583.1	189	3.02	defense protein hdd11	<1.00	Y
	comp21505_c0_seq1 - ABC59699.2	521	69.28	phenoloxidase subunit 2	-7.20	N
	comp22426_c0_seq1 *- BAM76811.1	308	25.43	phenoloxidase subunit 1-like	<1.00	N
	Unknown function					
	comp24798_c0_seq1 - KPJ14123.1	66	0.00	zinc finger partial	<1.00	N
	Other Enzymes					
	comp13624_c0_seq1 - EHJ69256.1	450	28.19	rab gdp dissociation inhibitor alpha	4.09, 4.73	N
	comp19425_c0_seq1 - KPJ11631.1	383	46.96	fructose-bisphosphate aldolase isoform x3	-8.04, -6.98	N
	comp30133_c0_seq1 *- AHV85216.1	340	39.64	glyceraldehyde-3-phosphate dehydrogenase	<1.00	N
	comp17755_c0_seq1 - XP_ 013197881.1	**217**	**21.62**	**peptidyl-prolyl cis-trans isomerase 5**	**3.58**	**Y**
	comp30119_c0_seq1 - XP_ 013133256.1	380	16.34	arginine kinase	-11.56, 10.70	N
	comp20773_c0_seq1 - NP_ 001298728.1	**496**	**13.38**	**protein disulfide isomerase**	**5.03**	**Y**

	comp13978_c0_seq1 - XP_ 014369064.1	110	12.65	long-chain-fatty-acid--ligase 4 isoform x1	13.79, 14.56	N
	comp14182_c0_seq1 - EHJ65007.1	553	8.88	atp synthase subunit mitochondrial	-5.60, -6.39	N
	comp24842_c0_seq1 - XP_ 013185295.1	349	8.37	cytosolic malate partial	<1.00	N
	comp29531_c0_seq1 - XP_ 013189570.1	705	7.71	glucose dehydrogenase	8.76, 21.33	N
	comp20964_c0_seq1 - KPI99141.1	210	7.06	peroxiredoxin 1	<1.00	N
	comp13811_c0_seq1 - NP_ 001298507.1	533	6.69	atp synthase subunit mitochondrial	<1.00	N
	comp29735_c0_seq1 - XP_ 013183407.1	194	6.42	nucleoside diphosphate kinase	8.06, 8.22	N
	comp8052_c0_seq1-XP_ 012549342.1	651	5.61	sphingomyelin phosphodiesterase	11.80, 10.76	Y
	comp27660_c0_seq1 - XP_ 013200244.1	539	4.99	pyruvate kinase-like isoform x1	-5.90, -5.91	N
	comp29695_c0_seq1 - XP_ 013191606.1	106	4.07	mitochondrial cytochrome c oxidase subunit 6c	-10.58, -9.53	N
	comp26299_c0_seq1 - NP_ 001298596.1	521	3.07	dihydrolipoamide dehydrogenase	4.45, 4.18	N
	comp25717_c0_seq1 - P31401.1	524	2.97	v-type proton atpase subunit b	<1.00	N
	comp19712_c0_seq1 - XM_013337708	276	2.91	nad mitochondrial-like	8.27, 9.14	N
	comp32255_c0_seq1 - ALD03682.1	246	2.85	nadh dehydrogenase	-8.92, -8.17	N
	comp28164_c0_seq3 - XP_ 013189812.1	667	2.72	succinate dehydrogenase	<1.00	N
	comp30728_c0_seq1 *- KPJ09319.1	585	2.71	catalase	4.25, 4.72	N
	comp28794_c0_seq1 - XP_ 013196670.1	503	2.67	fumarate mitochondrial-like	<1.00	N
	comp12038_c0_seq1 - EHJ76293.1	181	2.49	reverse transcriptase	-7.21, -7.83	N
	comp28630_c0_seq1 - XP_ 013196204.1	412	2.37	aspartate cytoplasmic	<1.00	N
	comp24438_c0_seq1 - XP_ 013174474.1	320	2.30	aldose reductase-like isoform x2	<1.00	N
	comp18221_c0_seq1 - KPI95301.1	430	2.26	delta-1-pyrroline-5-carboxylate mitochondrial	-4.08	N
	comp29759_c0_seq1 - XP_ 013174327.1	85	2.16	atp synthase subunit mitochondrial	-9.71, -8.79	N
	comp7575_c0_seq1-XP_ 013181141.1	982	1.75	alanine--trna cytoplasmic	4.91, 5.33	N
	Q9GV28	433	21.47	Chitinase-like protein EN03		-
	P54385	562	11.07	Glutamate dehydrogenase, mitochondrial		-
	P54385	562	11.07	Glutamate dehydrogenase,		-

				mitochondrial		
	P86699	223	9.39	Arginine kinase (Fragments)		-
	P29523	550	9.14	Membrane-bound alkaline phosphatase		-
	P86205	174	4.20	V-type proton ATPase catalytic subunit A (Fragments)		-
	Q24751	228	3.44	ATP synthase subunit beta, mitochondrial (Fragment)		-
	P31402	226	3.03	V-type proton ATPase subunit E		-
	Structural					
	comp30033_c0_seq1 - XP_ 017021020.1	148	52.96	truncated actin-4	<1.00	N
	P86221	290	16.23	Tubulin beta-4B chain (Fragments)		-
	P06604	449	13.05	Tubulin alpha-2 chain		-
	P13602	443	3.21	Tubulin beta-2 chain		-
	Q1HPU0	284	7.03	Tropomyosin-1		-
	P31816	283	9.49	Tropomyosin		-
	P07837	376	31.42	Actin, muscle-type A2		-
	P05661	1962	38.15	Myosin heavy chain, muscle		-
	comp20388_c0_seq1 - NP_ 001036887.1	456	3.21	tubulin beta-1 chain	<1.00	N
	comp20478_c0_seq1 - EGI67764.1	457	31.78	tubulin alpha-1 chain	-3.81	N
	comp19059_c0_seq1 - XP_ 013183021.1	288	12.88	tropomyosin-1	<1.00	N
	comp7754_c0_seq1- XP_ 013194450.1	1970	89.00	myosin heavy muscle isoform x1	<1.00	N
	comp569864_c0_seq 1- XP_ 009124146.1	126	1.67	histone h2b	<1.00	N
	comp233829_c0_seq 1- CDJ26240.1	130	8.46	histone h1	<1.00	N
	General cellular processes					
	Q24789	665	6.92	Heat shock cognate 70 kDa protein	-	-
	P02826	214	4.41	Heat shock 70 kDa protein cognate 1 (Fragments)	-	-
	comp20290_c0_seq1 *- XP_ 013166797.1	620	7.37	heat shock cognate 70 protein	5.07	N
	comp231404_c0_seq 1- XP_ 013188166.1	606	12.15	heat shock protein 70	<1.00	N
	Transport protein					
	comp14647_c0_seq1 - XP_ 013197881.1	462	5.62	endoplasmin	15.44, 13.04	N
	comp7707_c0_seq1* - NP_ 001040482.1	361	5.97	phosphate carrier mitochondrial	-7.83, -7.47	N

	comp29080_c0_seq1 - AEM75020.1	1701	253.67	vitellogenin	<1.00	N
	comp13646_c0_seq1 - XP_013188369.1	271	12.21	14-3-3 protein epsilon	<1.00	N
	comp25675_c0_seq1 - XP_013178606.1	185	3.42	zinc finger matrin-type protein 5	8.28	N
	comp22697_c0_seq1 - XP_AGG56522.1	129	2.70	histidine triad nucleide-binding protein 1	<1.00	N
	comp23908_c0_seq1 - NP_001040376.1	433	2.55	multifunctional protein ade2	<1.00	N
	comp29425_c0_seq1 - AHA86297.1	175	2.47	translationally controlled tumor protein	4.75, 5.07	N
	comp29159_c0_seq2 - XM_013320462.1	2868	2.21	protein binding	<1.00	N
	comp25173_c0_seq1 - XP_011554061.1	434	0.00	sorting nexin-27	<1.00	N
	P22297	681	3.07	Transferrin	-	-
	Q86YZ3	2850	8.30	Hornerin	-	-
	Q2F637	247	31.37	14-3-3 protein zeta	-	-
	comp29454_c0_seq1 - ABC79625.1	463	64.82	imaginal disc growth factor 4	8.76, 9.95	N
	comp22391_c0_seq2 - XP_004928610.1	590	3.17	protein rop isoform x1	<1.00	N
	comp7802_c0_seq1- XP_013196226.1	850	10.48	translation elongation factor 2	8.47	N
	comp27413_c0_seq1 - XP_004928900.1	249	2.32	mediator of rna polymerase ii transcription subunit 31	14.78	N
	comp29800_c0_seq1 - Q962U1.1	223	10.03	ribosomal protein L13	7.88, 7.11	N
	comp29927_c0_seq1 - XP_013189428.1	886	88.80	long form	<1.00	N
	comp7751_c0_seq1- XP_EDM13556.1	125	97.84	polyubiquitin-b- partial	<1.00	N
	P62976	658	89.76	Polyubiquitin	-	-
	comp829637_c0_seq1- AEG19552.1	107	4.48	storage protein	<1.00	Y
	comp807774_c0_seq1- AEG19552.1	88	25.28	storage protein	<1.00	N
	P09179	747	2.41	Sex-specific storage-protein 1	-	-
E						
	Peptidases					
	Q9HWK6	462	9.47	lysyl endopeptidase	-	-
	comp311204_c0_seq1*- XP_013201208.1	265	3.48	cathepsin L	-10.76, -8.21	N
	P54814	402	3.50	26S protease regulatory subunit 8	-	-
	comp25815_c0_seq1 - P54814.1	404	3.50	26s protease regulatory subunit 8	<1.00	N

	Structural					
	P10982	140	13.10	Actin-1 (Fragment)	-	-
	comp25213_c0_seq1 - XP_ 014369841.1	258	3.60	proteasome subunit alpha type-3	<1.00	N
	comp18938_c0_seq1 - XP_ 012547868.1	402	3.36	spectrin alpha chain isoform x2	12.47, 13.19	N
	comp25514_c0_seq2 - AFC34081.1	268	2.53	aquaporin isoform x2	-3.45	N
	comp7784_c0_seq1- XP_ 003977419.2	137	1.99	histone C	<1.00	N
	Immune response					
	P81605	110	2.86	dermcidin	-	-
	comp22426_c0_seq1 *- BAM76811.1	308	3.00	phenoloxidase subunit 1-like	<1.00	N
	Enzymes					
	P29401	623	2.97	Transketolase	-	-
	B0SC36	280	0.00	Diaminopimelate epimerase	-	-
	comp7707_c0_seq1* - NP_ 001040482.1	361	6.63	phosphate carrier mitochondrial	-7.83, -7.47	N
	comp28815_c1_seq2 - XP_ 013193819.1	1096	6.48	atp-citrate synthase	4.08	N
	comp30133_c0_seq1 *- AHV85216.1	340	3.86	glyceraldehyde-3-phosphate dehydrogenase	<1.00	N
	comp28189_c0_seq2 - AFO54708.2	844	3.37	glycogen phosphorylase	-8.91, -8.47	N
	comp30728_c0_seq1 *- KPJ09319.1	585	2.08	catalase	4.25, 4.72	N
	comp26186_c0_seq1 - XP_013177014.1	427	7.15	protein disulfide-isomerase a6	<1.00	Y
	comp28427_c0_seq1 - AID66693.1	725	2.98	peroxisomal multifunctional enzyme type 2	<1.00	N
	General cellular processes					
	comp20290_c0_seq1 *-XP_ 013166797.1	620	2.03	heat shock cognate 70 protein	5.07	N
	comp30844_c0_seq1 - XP_ 013186559.1	1255	5.64	vigilin	7.80, 5.84	N
	Q64542	1203	3.51	Plasma membrane calcium-transporting ATPase 4	-	-
	comp19529_c0_seq1 - XM_011550139.1	246	3.30	protein transport protein sec23a isoform x1	<1.00	N

Bold sequences were also identified as SFPs using a transcriptomic approach

- **These proteins were found in both E and Z spermatophores**

**** Negative numbers indicate a gene is up-regulated in male thorax**

Supplementary Methods 1.

Single LC-MS/MS experiment was performed on a LTQ Orbitrap Velos (Thermo Fischer) equipped with Waters (Milford, MA) NanoAcquity HPLC pump. Peptides were separated onto a 100 µm inner diameter microcapillary trapping column packed first with approximately 5 cm of C18 Aqua resin (5 µm, 100 Å, Waters, Milford, MA) followed by 15 cm of Aqua resin (1.8 µm, 200 Å, Milford, MA). Separation was achieved through applying a gradient from 5–27% ACN in 0.1% formic acid over 90 min at 200 nl min−1. Electrospray ionization was enabled through applying a voltage of 1.8 kV using an home-made electrode junction at the end of the microcapillary column and sprayed from fused silica pico tips (New Objective, MA). The LTQ Orbitrap Velos was operated in data-dependent mode for the mass spectrometry methods. The mass spectrometry survey scan was performed in the Orbitrap in the range of 395 –1,800 m/z at a resolution of 6 × 104, followed by the selection of the twenty most intense ions (TOP20) for CID-MS2 fragmentation in the Ion trap using a precursor isolation width window of 2 m/z, AGC setting of 1,000, and a maximum ion accumulation of 200 ms. Singly charged ion species were not subjected to CID fragmentation. Normalized collision energy was set to 35 V and an activation time of 10 ms. Ions in a 10 ppm m/z window around ions selected for MS2 were excluded from further selection for fragmentation for 60 s.

Supplementary table 1. Putative SFPs secreted from the *O. nubilalis* male accessory glands and ejaculatory duct.

Gene function	Transcript ID	Description	Homology	LFC	FDR	E-value
Peptidases						
	comp22488_c0_seq1-BAH23565.1	carboxypeptidase	*Bombyx mori*	16.6, 11.3	$2.5x10^{-8}$, $2.3x10^{-5}$	0
	comp28131_c0_seq1‡ - XP_011547818.1	carboxypeptidase B	*Plutella xylostella*	9.3	$7.0x10^{-4}$	0
	comp29067_c0_seq3-XP_004928966.1	carboxypeptide D	*Bombyx mori*	4.7, 5.5	$1.5x10^{-4}$, $1.1x10^{-3}$	0
	comp7756_c0_seq1*-AEW46900.2	cathepsin-like protease	*Chilo suppressalis*	10.2	$1.9x10^{-7}$	$6.5x10^{-23}$
	comp15721_c0_seq1-NP_001036904.1	furin-like convertase precursor	*Bombyx mori*	5.3, 9.6	$5.2x10^{-3}$, $4.9x10^{-6}$	$2.0x10^{-31}$
	comp19883_c0_seq2-XP_001600543.1	lysosomal aspartic protease	*Nasonia vitripennis*	8.5	$1.1x10^{-4}$, $6.9x10^{-5}$	$3.4x10^{-6}$
	comp22094_c0_seq1-BAH23566.1	male reproductive organ serine protease 1	*Bombyx mori*	10.7, 9.5	$3.1x10^{-10}$, $1.2x10^{-8}$	$3.2x10^{-155}$
	comp23849_c0_seq2*-AFK93534.1	pattern recognition serine proteinase	*Bombyx mori*	5.5	$8.0x10^{-3}$	0
	comp7717_c0_seq1-AEU11470.1	seminal fluid protein HACP001	*Heliconius erato*	31.6, 21.1	$1.9x10^{-8}$, $1.4x10^{-5}$	$2.2x10^{-154}$
	comp21136_c0_seq1-AEU11614.1	seminal fluid protein HACP026	*Heliconius erato*	9.5	$6.9x10^{-3}$, $5.8x10^{-4}$	$2.0x10^{-114}$
	comp19489_c0_seq1-ADJ58583.1	serine protease easter-like	*Heliconius melpomene*	15.3, 15.8	$3.8x10^{-10}$, $2.2x10^{-10}$	$2.0x10^{-76}$
	comp16437_c0_seq1-XP_004926199.1	traffiking protein particle complex subunit 12	*Bombyx mori*	6.9, 6.0	$2.3x10^{-4}$, $1.2x10^{-3}$	$6.8x10^{-102}$
	comp21256_c0_seq1‡-XP_004930560.1	trypsin-7	*Bombyx mori*	7.7	$3.2x10^{-3}$	$6.0x10^{-136}$
	comp27256_c0_seq1-XP_004929234.1	venom dipeptidyl peptidase 4-like isoform X2	*Bombyx mori*	7.2, 7.8	$1.0x10^{-4}$, $4.2x10^{-5}$	0
	comp20793_c0_seq1-XP_011555901.1	venom serine carboxypeptidase	*Plutella xylostella*	17.0, 11.5	$7.2x10^{-8}$, $4.0x10^{-5}$	$8.4x10^{-151}$
	comp9248_c0_seq1-EHJ74305.1	ubiquitin carboxyl-terminal hydrolase 20	*Bombyx mori*	6.5, 5.7	$2.7x10^{-3}$, $9.0x10^{-3}$	$1.6x10^{-53}$
	comp16071_c0_seq1-XP_004928968.1	rhomboid related protein 3	*Bombyx mori*	5.8, 6.7	$1.8x10^{-3}$, $5.2x10^{-4}$	$3.9x10^{-37}$
	comp14013_c0_seq1‡-XP_011553661.1	zinc metalloproteinase nas 14	*Plutella xylostella*	16.8	$2.5x10^{-3}$	$8.2x10^{-123}$
Peptidase regulators						
	comp25535_c0_seq2‡ - AEW46891.2	serine protease inhibitor 012	*Chilo suppressalis*	17.3	$1.9x10^{-3}$	$5.3x10^{-119}$
	comp13601_c0_seq1-EHJ75455.1	boophilin-like protein	*Danaus plexippus*	13.6, 12.3	$1.2x10^{-4}$, $3.8x10^{-4}$	$4.1x10^{-35}$
	comp21981_c0_seq2*-AAV91427.1	protease inhibitor 5	*Lonomia obliqua*	8.7	$7.3x10^{-4}$	$1.7x10^{-17}$

	comp14214_c0_seq1‡-ABG72726.1	protease inhibitor-like protein	*Antheraea mylitta*	5.5	$3.4x10^{-3}$	$7.9x10^{-5}$
	comp29971_c0_seq1‡-XP_011551777.1	pacifastin light chain	*Plutella xylostella*	20.1	$3.5x10^{-4}$	$3.0x10^{-7}$
	comp27798_c0_seq1* -NP_001139701.1	serine protease inhibitor 7	*Bombyx mori*	7.8	$2.1x10^{-3}$	$8.6x10^{-142}$
	comp26547_c0_seq1‡ -EHJ70410.1	serine protease inhibitor serpin 1	*Mamestra configurata*	15.5	$1.0x10^{-3}$	$6.0x10^{-55}$
	comp26839_c0_seq1 – XP_011562342.1	inter alpha-trypsin inhibitor	*Plutella xylostella*	4.2, 4.1	$1.8x10^{-4}$, $2.8x10^{-3}$	0
Odorant binding proteins						
	comp7771_c0_seq1*-AGH70105.1	odorant binding protein	*Spodoptera exigua*	10.5	$2.9x10^{-3}$	$3.4x10^{-66}$
	comp21759_c0_seq1‡-AGS36748.1	odorant binding protein	*Sesamia inferens*	10.1	$3.4x10^{-4}$	$6.5x10^{-36}$
	comp18104_c0_seq1‡ -BAG71416.1	pheromone binding protein	*Mythimna separata*	17.0	$5.3x10^{-3}$	$1.3x10^{-26}$
Other enzymes						
	comp17755_c0_seq1‡ -BAM19138.1	peptidyl-prolyl cis-trans isomerase	*Papilio polytes*	3.6	$5.3x10^{-3}$	$1.7x10^{-105}$
	comp28323_c0_seq2*-XP_004928565.1	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase-like isoform X1	*Bombyx mori*	4.8	$4.4x10^{-3}$	0
	comp28117_c0_seq2*-BAM20407.1	acid phosphatase	*Papilio polytes*	6.3	$5.9x10^{-4}$	$1.5x10^{-177}$
	comp12301_c0_seq1‡-EHJ67325.1	acid sphingomyelinase-like phosphodiesterase	*Danaus plexippus*	7.4	$1.1x10^{-3}$	0
	comp19855_c0_seq1-AJN91202.1	carboxylesterase	*Cnaphalocrocis medinalis*	5.1, 5.4	$3.4x10^{-4}$, $1.8x10^{-4}$	$2.1x10^{-165}$
	comp27906_c0_seq5*-AJN91202.1	carboxylesterase	*Cnaphalocrocis medinalis*	5.5	$1.6x10^{-3}$	0
	comp370_c0_seq1-XP_004925597.1	E3 ubiquitin-protein ligase TRIM9	*Bombyx mori*	15.8, 14.2	$1.7x10^{-5}$, $1.4x10^{-4}$	$2.4x10^{-130}$
	comp19181_c0_seq1*-	Glucose-methanol-choline oxidoreductase	*IPS*	7.1	$5.8x10^{-3}$	
	comp26795_c0_seq1-EHJ75790.1	multiple inositol polyphosphate phosphatase 1	*Danaus plexippus*	6.9, 8.4	$2.1x10^{-3}$, $3.3x10^{-4}$	$1.1x10^{-112}$
	comp26611_c0_seq1-AGG55020.1	nitrilase	*Heliothis subflexa*	4.8, 7.1	$7.7x10^{-3}$, $2.2x10^{-4}$	$1.8x10^{-109}$
	comp28068_c1_seq1‡-XP_004931741.1	polypeptide N-acetylgalatosaminyltransferase 1	*Danaus plexippus*	7.6	$6.2x10^{-4}$	$1.2x10^{-22}$
	comp24545_c0_seq1‡-XP_004926787.1	polypeptide N-acetylgalatosaminyltransferase 1	*Bombyx mori*	7.5	$1.8x10^{-4}$	$6.7x10^{-85}$
	comp20773_c0_seq1‡-BAM17793.1	protein disulfide isomerase	*Papilio xuthus*	5.0	$9.2x10^{-4}$	0
	comp17055_c0_seq1-XP_011547875.1	protein goliath	*Plutella xylostella*	5.5, 5.2	$5.4x10^{-4}$, $1.0x10^{-3}$	$2.4x10^{-28}$

	comp26777_c0_seq1-XP_011549197.1	serine/threonine-protein kinase	*Plutella xylostella*	8.0, 5.2	$1.6x10^{-5}$, $2.8x10^{-3}$	0
	comp17461_c0_seq1-EHJ66151.1	serine/threonine-protein kinase	*Danaus plexippus*	6.1, 7.1	$9.0x10^{-3}$, $3.3x10^{-3}$	$1.0x10^{-94}$
	comp19590_c0_seq1-XP_011549248.1	tyrosine-protein phosphatase 69D	*Plutella xylostella*	5.5, 5.7	$7.7x10^{-4}$, $5.5x10^{-4}$	$9.3x10^{-124}$
	comp26585_c0_seq1-XP_004926247.1	dolichyl-diphosphooligosaccharide	*Bombyx mori*	8.6, 6.7	$3.8x10^{-8}$, $3.9x10^{-5}$	0
	comp28456_c1_seq1‡-NP_001138795.1	endoplasmic reticulum oxidoreductase 1	*Bombyx mori*	6.0	$2.3x10^{-4}$	0
	comp22798_c0_seq2*-EHJ69101.1	glucuronyltransferase 1	*Danaus plexippus*	5.3	$2.1x10^{-3}$	$7.8x10^{-169}$
	comp27891_c0_seq1-XP_004928791.1	protein purity of essence	*Bombyx mori*	7.1, 6.6	$2.9x10^{-5}$, $8.1x10^{-5}$	0
	comp26858_c0_seq1-ADX87345.1	yellow-b	*Heliconius melpomene*	8.8, 8.3	$1.1x10^{-6}$, $3.8x10^{-6}$	0
	comp22723_c0_seq1‡-BAM18176.1	yellow-x	*Papilio xuthus*	4.2	$2.8x10^{-3}$	0
	comp13878_c0_seq1*-NP_001161422.1	protein D2-like isoform X2	*Acyrthosiphon pisum*	27.2	$2.5x10^{-3}$	$6.3x10^{-47}$-
	comp29089_c0_seq2-XP_011559501.1	protocadherin-15	*Plutella xylostella*	6.5, 6.8	$2.7x10^{-4}$, $2.0x10^{-4}$	0
Carbohydrate metabolism/lipid metabolism/other metabolism						
	comp22901_c0_seq1-XP_011552102.1	beta-glucuronidase	*Plutella xylostella*	17.9, 9.6	$3.0x10^{-6}$, $2.8x10^{-3}$	0
	comp8043_c0_seq1-ADB85578.1	chitinase	*Ostrinia nubilalis*	26.5, 17.2	$7.3x10^{-6}$, $7.0x10^{-4}$	0
	comp26449_c1_seq1*-XP_004928921.1	glucose dehadrogenase	*Bombyx mori*	7.4	$6.4x10^{-3}$	0
	comp18753_c0_seq1-EHJ65511.1	cuticular protein analogous to peritrophins 1-G	*Danaus plexippus*	13.5, 21.7	$1.4x10^{-4}$, $2.9x10^{-7}$	$6.6x10^{-94}$
	comp5922_c0_seq1-XP_004923982.1	frizzled-10	*Bombyx mori*	10.9, 15.0	$9.9x10^{-4}$, $9.9x10^{-6}$	$1.4x10^{-21}$
	comp23892_c0_seq1-EHJ71114.1	glucosidase	*Danaus plexippus*	5.3, 5.7	$5.8x10^{-3}$, $3.5x10^{-3}$	0
	comp27158_c0_seq2-ABO20846.1	trehalase-1	*Omphisa fuscidentalis*	4.6, 7.1	$2.7x10^{-3}$, $1.6x10^{-5}$	0
	comp19685_c0_seq1-XP_004925820.1	group XV phospholipase A2	*Bombyx mori*	5.0, 4.6	$1.0x10^{-3}$, $3.4x10^{-4}$	$1.8x10^{-133}$
	comp19149_c0_seq1‡-EHJ78604.1	intestinal mucin	*Danaus plexippus*	7.7	$3.8x10^{-3}$	$3.0x10^{-39}$
	comp20004_c0_seq1-XP_011551568.1	protein 5NUC-like isoform X2	*Plutella xylostella*	8.9, 19.8	$4.7x10^{-5}$, $2.0x10^{-12}$	0
	comp27809_c0_seq1-ABI81756.1	N-acetylglucosaminidase	*Ostrinia furnacalis*	11.3, 7.4	$9.4x10^{-6}$, $1.5x10^{-3}$	0
	comp13894_c0_seq1-EHJ79051.1	phospholipase A2D	*Danaus plexippus*	12.2, 21.1	$8.4x10^{-4}$, $2.2x10^{-6}$	$9.5x10^{-74}$
	comp8132_c0_seq1-XP_011562491.1	PI-PLC X domain-containing protein 1	*Bombyx mori*	17.7, 25.5	$2.6x10^{-6}$, $4.1x10^{-9}$	0

	comp8052_c0_seq1-XP_012549342.1	**sphingomyelin phosphodiesterase**	*Bombyx mori*	**11.80, 10.76**	$5.7x10^{-6}$, $2.8x10^{-5}$	0
Binding and transport proteins						
	comp25928_c0_seq1‡-XP_004925396.1	cationic amino acid transporter 2	*Bombyx mori*	8.5	$1.9x10^{-4}$	$7.8x10^{-137}$
	comp3621_c0_seq1-XP_011550992.1	glycine receptor subunit alpha 4	*Plutella xylostella*	17.4, 15.1	$3.3x10^{-5}$, $3.0x10^{-4}$	0
	comp21812_c0_seq1*-XP_011569320.1	microsomal triglyceride transfer protein	*Plutella xylostella*	7.9	$2.4x10^{-4}$	$2.2x10^{-55}$
	comp20163_c0_seq1-EHJ76486.1	apolipophorin	*Danaus plexippus*	10.5, 10.6	$1.1x10^{-3}$, $1.1x10^{-3}$	0
	comp22583_c0_seq2-XP_011552627.1	basement membrane-specific heparan sulfate proteoglycan core protein	*Plutella xylostella*	6.4, 6.6	$9.0x10^{-3}$, $7.9x10^{-3}$	$4.1x10^{-72}$
	comp14879_c0_seq1-XP_011556351.1	basigin	*Plutella xylostella*	10.6, 10.6	$3.1x10^{-8}$, $4.9x10^{-8}$	$2.8x10^{-12}$
	comp19780_c0_seq1‡-BAB79277.1	calreticulin	*Galleria mellonella*	6.5	$6.3x10^{-5}$	0
	comp26364_c0_seq1-XP_004923501.1	calsyntenin-1	*Bombyx mori*	5.9, 7.3	$1.2x10^{-5}$, $3.0x10^{-6}$	0
	comp18962_c0_seq1*-XP_011560391.1	Down syndrome cell adhesion molecule-like protein 1	*Plutella xylostella*	4.3	$5.6x10^{-3}$	$6.6x10^{-113}$
	comp19808_c0_seq1‡-EHJ74096.1	furrowed	*Danaus plexippus*	11.1	$6.9x10^{-3}$	0
	comp13388_c0_seq1-BAC06463.1	gag-like protein	*Papilio xuthus*	6.5, 6.9	$8.2x10^{-4}$, $4.6x10^{-4}$	$5.0x10^{-5}$
	comp6572_c0_seq1-BAC06463.1	interleukin 1 receptor accessory protein like 2	*Bombyx mori*	13.3, 15.1	$2.1x10^{-6}$, $9.4x10^{-8}$	$1.3x10^{-129}$
	comp22742_c0_seq1-AIR96002.1	lectin 1	*Ostrinia furnacalis*	11.6, 11.0	$2.9x10^{-6}$, $8.1x10^{-6}$	$6.6x10^{-120}$
	comp22344_c0_seq1‡-NP_001108408.1	FK506-binding protein	*Bombyx mori*	4.1	$6.6x10^{-3}$	$1.1x10^{-129}$
	comp24695_c0_seq1*-XP_011563126.1	hemicentin-1	*Danaus plexippus*	4.8	$3.3x10^{-3}$	0
	comp18744_c0_seq1-XP_011563126.1	multipe epidermal growth factor like domains protein 11	*Plutella xylostella*	8.2, 9.2	$2.7x10^{-6}$, $3.6x10^{-7}$	$6.2x10^{-71}$
	comp24160_c0_seq1-XP_004922199.1	multiple coagulation factor deficiency protein 2	*Bombyx mori*	11.0, 7.7	$4.4x10^{-6}$, $5.2x10^{-4}$	$2.3x10^{-42}$
	comp23862_c0_seq1-XP_004924889.1	nipped B like	*Bombyx mori*	8.5, 8.5	$6.8x10^{-7}$, $9.3x10^{-7}$	0
	comp14074_c0_seq1‡-XP_004921981.1	mesencephalic astocyte-derived neutrophic factor	*Bombyx mori*	6.7	$1.0x10^{-3}$	$3.9x10^{-81}$
	comp27378_c0_seq1-XP_004923396.1	protein sel-1 homolog 1	*Bombyx mori*	6.3, 5.2	$1.7x10^{-5}$, $3.0x10^{-4}$	$7.0x10^{-94}$
	comp11077_c0_seq1*-XP_011552980.1	protein singed wings 2	*Plutella xylostella*	7.9	$7.3x10^{-4}$	$2.9x10^{-20}$

	comp1329_c0_seq1-XP_004926232.1	protein prickle	*Bombyx mori*	13.8, 12.4	$1.3x10^{-3}$, $4.3x10^{-3}$	$8.7x10^{-57}$
	comp9451_c0_seq1*-XP_004927145.1	protein takeout	*Bombyx mori*	12.5	$1.7x10^{-3}$	$1.6x10^{-113}$
	comp19268_c0_seq1*-XP_011549055.1	protein takeout	*Plutella xylostella*	10.2	$8.5x10^{-4}$	$1.2x10^{-145}$
	comp29009_c0_seq1-XP_011563898.1	sodium hydrogen exchanger isoform 3	*Plutella xylostella*	4.9, 9.9	$4.8x10^{-3}$, $8.8x10^{-7}$	0
	comp27689_c0_seq1-XP_004923512.1	solute carrier family 12 member 9	*Plutella xylostella*	9.2, 6.8	$4.2x10^{-4}$, $4.9x10^{-3}$	$1.7x10^{-172}$
	comp21662_c0_seq1‡-EZA62148.1	thap domain-containing particle	*Cerapachys biroi*	4.7	$2.5x10^{-3}$	$5.1x10^{-29}$
	comp22560_c0_seq1*-XP_004928023.1	transmembrane and TPR repeat containing protein	*Bombyx mori*	6.4	$1.5x10^{-4}$	0
	comp2641_c0_seq1-XP_004932130.1	tripartite motif-containing protein 45	*Bombyx mori*	13.1, 12.2	$4.7x10^{-3}$, $9.1x10^{-3}$	$7.6x10^{-77}$
	comp26798_c0_seq1-EHJ67488.1	vacuolar ATP synthase subunit s1	*Danaus plexippus*	6.1, 8.0	$3.6x10^{-5}$, $3.8x10^{-7}$	$3.0x10^{-124}$
	comp19745_c0_seq1*-XP_004932608.1	voltage-dependent T-type calcium channel subunit alpha-1G	*Bombyx mori*	5.2	$3.1x10^{-3}$	0
	comp14130_c0_seq1-XP_004929592.1	WD repeat and FYVE domain containing protein-3	*Bombyx mori*	6.0, 6.1	$7.8x10^{-3}$, $7.6x10^{-3}$	0
	comp26527_c0_seq1-XP_011564046.1	zinc transporter foi	*Plutella xylostella*	4.6, 12.2	$6.7x10^{-3}$, $8.1x10^{-9}$	0
	comp28326_c0_seq1*-XP_004925979.1	zinc transporter ZIP9	*Bombyx mori*	6.7	$4.2x10^{-4}$	$3.7x10^{-145}$
Structural						
	comp25198_c0_seq1‡-EHJ67543.1	cuticle protein	*Danaus plexippus*	12.5	$5.4x10^{-6}$	$3.7x10^{-46}$
	comp25297_c0_seq1*-AAV73780.1	cuticle protein 13	*Antheraea yamamai*	7.4	$2.7x10^{-3}$	$1.6x10^{-20}$
	comp26610_c1_seq1*-EHJ76583.1	cuticular protein RR-3 motif 148	*Danaus plexippus*	9.1	$3.3x10^{-3}$	$6.2x10^{-93}$
	comp14556_c0_seq1-EHJ68589.1	hdd-1	*Danaus plexippus*	8.0, 13.1	$7.3x10^{-3}$, $1.3x10^{-4}$	$1.3x10^{-48}$
	comp2952_c0_seq1-XP_004932814.1	laminin subunit gamma 1	*Bombyx mori*	15.7, 15.1	$1.1x10^{-3}$, $1.9x10^{-3}$	$4.8x10^{-95}$
	comp14514_c0_seq1-EHJ66123.1	membrane protein TMS1	*Danaus plexippus*	11.9, 15.2	$1.7x10^{-11}$, $8.9x10^{-16}$	$1.1x10^{-120}$
	comp22171_c0_seq1-XP_011554490.1	Niemann-pick c1	*Plutella xylostella*	5.8, 5.6	$1.6x10^{-3}$, $2.2x10^{-3}$	$6.9x10^{-78}$
	comp29684_c0_seq1-ACY95306.1	ribosomal protein I26	*Manduca sexta*	8.7, 8.2	$1.0x10^{-7}$, $5.5x10^{-7}$	$3.8x10^{-77}$
	comp7734_c0_seq1-ACY95306.1	ribosomal protein p2	*Manduca sexta*	7.8, 7.5	$6.9x10^{-7}$, $1.7x10^{-6}$	
	comp15459_c0_seq1-XP_004931261.1	transmembrane protein 161b	*Bombyx mori*	5.7, 5.1	$1.1x10^{-3}$, $3.5x10^{-3}$	$4.7x10^{-139}$
	comp22938_c0_seq1-XP_011555878.1	transmembrane 9 superfamily member 2	*Bombyx mori*	11.0, 10.0	$3.0x10^{-8}$, $2.8x10^{-7}$	0
	comp29027_c0_seq5-	vacuole membrane	*Danaus plexippus*	4.9, 6.2	$3.1x10^{-3}$,	0

	EHJ74814.1	protein 1 isoform 1X			$3.9x10^{-4}$	
	comp18839_c0_seq2-XP_011550988.1	vascular endothelial growth factor B-like isoform X2	*Plutella xylostella*	20.4, 15.5	$8.9x10^{-4}$, $5.9x10^{-3}$	$2.3x10^{-9}$
	comp28746_c0_seq10-XP_004933065.1	vesicular integral-membrane protein VIP36	*Bombyx mori*	9.0, 9.8	$2.1x10^{-8}$, $3.4x10^{-9}$	0
	comp2156_c0_seq1-ACX50393.1	silk protein p25	*Corcyra cephalonica*	12.5, 13.9	$4.0x10^{-3}$, $1.5x10^{-3}$	$8.8x10^{-95}$
	comp21813_c0_seq1-EHJ69021.1	16 kDa salivary protein	*Danaus plexippus*	9.1, 7.6	$6.7x10^{-9}$, $6.5x10^{-7}$	$1.5x10^{-57}$
	comp22694_c0_seq1*-XP_011558626.1	27 kDa glycoprotein	*Plutella xylostella*	7.6	$9.6x10^{-4}$	$2.6x10^{-96}$
	comp1512_c0_seq1‡-XP_011569370.1	protein skeletor	*Plutella xylostella*	12.8	$4.4x10^{-4}$	$1.4x10^{-70}$
Receptors and Proteins involved in signaling pathways						
	comp24559_c0_seq1-XP_004931417.1	semaphorin 5B	*Bombyx mori*	4.6, 4.6	$9.2x10^{-4}$, $1.1x10^{-3}$	0
	comp18375_c0_seq1-XP_011551273.1	semaphorin-1A	*Plutella xylostella*	6.1, 6.4	$7.6x10^{-3}$, $6.4x10^{-3}$	0
	comp27478_c0_seq1‡-NP_001116821.1	18 wheeler precursor	*Bombyx mori*	8.3	$3.4x10^{-4}$	0
	comp2184_c0_seq1-XP_011549620.1	ephrin type-B receptor 1_B	*Plutella xylostella*	12.3, 11.0	$1.0x10^{-3}$, $4.5x10^{-3}$	$2.8x10^{-42}$
	comp19292_c0_seq1-BAE94422.1	fibroblast growth factor receptor	*Spodoptera frugiperda*	4.9, 7.3	$1.6x10^{-3}$, $1.2x10^{-5}$	0
	comp20687_c0_seq1-ABM91320.1	integrin beta 1	*Ostrinia furnacalis*	4.1, 5.7	$5.3x10^{-3}$, $1.9x10^{-4}$	$8.0x10^{-166}$
	comp1865_c0_seq1*-AIG51930.1	ionotropic glutamate receptor	*Helicoverpa armigera*	13.3	$3.6x10^{-4}$	$1.7x10^{-36}$
	comp10741_c0_seq1*-ACJ06652.1	LGR1	*Spodoptera littoralis*	5.7	$2.2x10^{-4}$	0
Defense response						
	comp25838_c0_seq2*-NP_001243674.1	thrombospondid type 1	*Bombyx mori*	4.3	$9.6x10^{-3}$	$4.1x10^{-108}$
	comp18005_c0_seq1-EHJ66058.1	gloverin	*Danaus plexippus*	13.0, 11.6	$6.1x10^{-5}$, $2.6x10^{-4}$	$7.2x10^{-19}$
Proteins involved in RNA/DNA replication and cell homeostasis						
	comp28223_c0_seq1-XP_011564234.1	polypyrimidine tract-binding protein	*Plutella xylostella*	7.1, 6.2	$8.6x10^{-6}$, $8.6x10^{-5}$	$1.2x10^{-117}$
	comp21089_c0_seq2*-XP_004921565.1	post-GPI attachment to proteins factor 3-like isoform X1	*Bombyx mori*	7.3	$2.0x10^{-5}$	$1.1x10^{-96}$
	comp889_c0_seq1‡-XP_004921906.1	probable rho GTPase-activating protein	*Bombyx mori*	13.1	$5.4x10^{-3}$	$3.6x10^{-39}$
	comp20022_c0_seq1‡-ABO45233.1	reverse transcriptase	*Ostrinia nubilalis*	16.2	$1.9x10^{-3}$	$3.2x10^{-12}$
	comp13967_c0_seq1*-ABO45233.1	reverse transcriptase	*Ostrinia nubilalis*	15.7	$8.4x10^{-11}$	$3.6x10^{-32}$
	comp13171_c0_seq1-XP_004932297.1	active breakpoint cluster region protein	*Bombyx mori*	7.8, 7.4	$8.3x10^{-6}$, $3.0x10^{-5}$	0

	comp5691_c0_seq1-XP_011548557.1	atp-dependent rna helicase ddx54	*Plutella xylostella*	18.8, 17.7	$9.4x10^{-9}$, $8.9x10^{-8}$	$2.1x10^{-81}$
	comp27233_c0_seq4-XP_004924123.1	bmp and activin membrane-bound inhibitor	*Bombyx mori*	5.1, 4.1	$4.1x10^{-4}$, $3.9x10^{-3}$	$1.4x10^{-109}$
	comp27222_c0_seq1-XP_004927743.1	calnexin	*Bombyx mori*	5.0, 5.4	$7.3x10^{-4}$, $3.0x10^{-4}$	0
	comp6355_c0_seq1‡-XP_004921672.1	dnaJ homolog subfamily C member 16	*Bombyx mori*	7.2	$5.9x10^{-3}$	0
	comp20242_c0_seq1-XP_004924464.1	dnaJ homolog subfamily C member 3	*Bombyx mori*	8.3, 7.2	$6.5x10^{-6}$, $7.0x10^{-5}$	$5.7x10^{-91}$
	comp23329_c0_seq1‡-ADI61811.1	endonuclease-reverse transcriptase	*Bombyx mori*	5.7	$1.1x10^{-3}$	$6.0x10^{-86}$
	comp18845_c0_seq1-XP_004928941.1	eukaryotic translation initiation factor 3 subunit A	*Bombyx mori*	6.5, 8.6	$1.1x10^{-3}$, $5.2x10^{-5}$	$2.3x10^{-166}$

‡ Sequences were only found to be up-regulated in the accessory glands compared to the male thorax.

*Sequences were only found to be up-regulated in the ejaculatory duct compared to the male thorax.

† Information from the accessory gland is listed first followed by the information from the ejaculatory duct.

Supplementary Table 2. Sequences that are significantly differentially expressed between strains of ECB males in the accessory gland and the ejaculatory duct.

Strain and tissue	Gene functional category Sequence id-Description	Homology	Log_2FC	FDR	Secretion signal (Y/N)	E-value
E strain Ejaculatory duct						
	Hormone degradation					
	comp16883_c0_seq1-juvenile hormone esterase	**Manduca sexta**	-11.8	$1.86x10^{-15}$	N	$2.5x10^{-148}$
	Peptidases					
	comp10003_c0_seq1-tryptase	Bombyx mori	-11.6	$2.7x10^{-15}$	N	$5.0x10^{-26}$
	Other enzymes					
	comp53758_c0_seq1-farnesyl diphosphate synthase	Choristoneura fumiferana	-11.5	$4.4x10^{-8}$	N	$9.9x10^{-53}$
	comp96487_c0_seq1-E3 ubiquitin-ligase sinah partial	Zootermopsis nevadensis	-11.3	$4.8x10^{-8}$	N	$4.9x10^{-169}$
	comp182764_c0_seq1-aldo-keto reductase AKR2E4-like	Papilo xuthus	-10.8	$5.6x10^{-6}$	Y	$2.2x10^{-31}$
	comp11261_c0_seq1-multidrug resistance 1A-like	Fukomys damarensis	-11.1	$2.5x10^{-9}$	N	
	Metabolism					
	comp29676_c0_seq1-hypothetical protein involved in metabolic processes	Dendrolimus punctatus cypovirus 22	-11.2	$2.4x10^{-6}$	N	$9.8x10^{-45}$
	Transport					
	comp143474_c0_seq1-nose resistant to fluoxetine 6-like	Harpegnathos saltator	-10.5	$2.7x10^{-13}$	N	
	Proteins involved in RNA/DNA replication and cell homeostasis					
	comp151919_c0_seq1-enhancer of mRNA-decapping 4-like	Plutella xylostella	-10.1	$8.6x10^{-11}$	N	$5.9x10^{-20}$
	comp29678_c0_seq1-RNA-dependent RNA polymerase	Tobacco bushy top virus	-11.2	$9.8x10^{-7}$	N	$5.0x10^{-147}$
	Unknown conserved proteins					
	comp31166_c0_seq1-hypothetical protein	Chronic bee paralysis virus	-12.1	$1.0x10^{-6}$	N	$4.5x10^{-9}$
	Receptors					
	comp1865_c0_seq1-ionotropic glutamate receptor		-6.0	$1.4x10^{-3}$	Y	
	Novel					
	comp46998_c0_seq1-novel		-12.8	$4.1x10^{-12}$	N	
	comp39935_c0_seq1-novel		-12.7	$1.3x10^{-9}$	N	

	comp12436_c0_seq1-novel		-12.6	$5.1x10^{-8}$	N	
	comp41446_c0_seq1-novel		-12.2	$7.34x10^{-8}$	N	
	comp30321_c0_seq1-novel		-12.0	$1.1x10^{-5}$	N	
	comp8654_c0_seq1-novel		-11.8	$3.6x10^{-9}$	N	
	comp30512_c0_seq1-novel		-11.6	$2.0x10^{-5}$	Y	
	comp132623_c0_seq1-novel		-11.4	$2.3x10^{-13}$	N	
	comp78683_c0_seq1-novel		-11.1	$1.2x10^{-9}$	N	
	comp53811_c0_seq1-novel		-10.9	$4.1x10^{-7}$	N	
	comp171333_c0_seq1-novel		-10.9	$3.1x10^{-12}$	N	
	comp54037_c0_seq1-novel		-10.9	$8.9x10^{-15}$	N	
	comp91159_c0_seq1-novel		-10.8	$7.8x10^{-13}$	N	
	comp55452_c0_seq1-novel		-10.7	$1.4x10^{-7}$	N	
	comp129817_c0_seq1-novel		-10.5	$1.3x10^{-3}$	N	
	comp198931_c0_seq1-novel		-10.5	$3.3x10^{-6}$	N	
	comp5489_c0_seq1-novel		-10.5	$1.0x10^{-5}$	N	
	comp11005_c0_seq1-novel		-10.5	$7.45x10^{-7}$	N	
	comp29610_c0_seq2-novel		-10.5	$4.1x10^{-6}$	N	
	comp57950_c0_seq1-novel		-10.3	$2.0x10^{-11}$	N	
	comp35163_c0_seq1-novel		-10.2	$2.7x10^{-3}$	Y	
	comp86944_c0_seq1-novel		-10.1	$1.6x10^{-7}$	N	
	comp52490_c0_seq1-novel		-10.1	$5.3x10^{-9}$	N	
	comp16320_c0_seq1-Novel		-10.0	$3.4x10^{-12}$	N	
	comp8973_c0_seq1-novel		-3.8	$4.4x10^{-3}$	Y	
E-strain accessory glands						
	Hormone degradation					
	comp16883_c0_seq1-juvenile hormone esterase	**Manduca sexta**	-10.9	$1.6x10^{-13}$	N	$2.5x10^{-148}$
	Proteins involved in RNA/DNA replication and cell homeostasis					
	comp4149_c0_seq1-ATP synthase F0 subunit 6	Ixodes persulcatus	-10.8	$2.6x10^{-3}$	N	$9.2x10^{-31}$
	comp327_c0_seq1- NAD dehydrogenase subunit	Ixodes persulcatus	-11.3	$4.5x10^{-3}$	Y	$5.3x10^{-55}$
	comp105050_c0_seq1-novel		-12.2	$6.9x10^{-6}$	N	
	comp47551_c0_seq1-novel		-11.8	$4.0x10^{-16}$	N	
	comp15514_c0_seq1-novel		-11.7	$6.6x10^{-3}$	N	
	comp75857_c0_seq1-novel		-11.2	$1.5x10^{-10}$	N	
	comp46998_c0_seq1-novel		-11.1	$1.2x10^{-9}$	N	
	comp5864_c0_seq1-novel		-11.0	$7.2x10^{-6}$	N	
	comp8675_c0_seq1-novel		-11.0	$1.4x10^{-7}$	Y	
	comp87817_c0_seq1-novel		-10.3	$8.2x10^{-9}$	N	
	comp5603_c0_seq2-novel		-10.0	$1.5x10^{-8}$	N	
	comp8973_c0_seq1-novel*		-3.8	$4.4x10^{-3}$	Y	

Z-strain ejaculatory duct						
	Peptidases					
	comp30823_c0_seq1-serine-endopeptidase	Heliconius erato	10.0	4.5×10^{-5}		2.00×10^{-166}
	comp13511_c0_seq1-serine-endopeptidase	Heliconius erato	12.5	2.7×10^{-6}	Y	3.3×10^{-60}
	comp29421_c0_seq1-serine endopeptidase	Bombyx mori	10.2	1.0×10^{-4}		0
	comp29831_c0_seq1-serine-endopeptidase	Danaus plexippus	10.2	1.9×10^{-4}		2.7×10^{-145}
	comp7702_c0_seq1- serine-endopeptidase	Heliconius erato	10.4	1.7×10^{-4}	Y	6.2×10^{-162}
	Odorant binding and odorant receptors					
	comp11571_c0_seq1-odorant binding protein	Bombyx mori	10.3	2.8×10^{-10}	Y	2.8×10^{-36}
	Other enzymes					
	comp9075_c0_seq1-alpha-fucosyltransferase	Plutella xylostella	10.7	2.8×10^{-4}	N	5.1×10^{-98}
	Unknown conserved proteins					
	comp21645_c0_seq1-unknown conserved protein	Acyrthosiphon pisum	10.8	5.1×10^{-10}	N	1.91×10^{-54}
	Receptors					
	comp6450_c0_seq2-glycine receptor	Bombyx mori	10.8	6.5×10^{-14}	Y	$1.9X10^{-53}$
	Novel					
	comp31509_c0_seq1-novel		10.5	1.4×10^{-4}	N	
	comp7490_c0_seq1-novel		10.6	1.0×10^{-10}	N	
	comp9192_c0_seq1-novel		10.7	1.6×10^{-3}	N	
	comp14614_c0_seq1-novel		11.5	3.3×10^{-3}	Y	
	comp8741_c0_seq1-novel		11.8	1.6×10^{-4}	N	
	comp7274_c0_seq1-novel		12.1	3.3×10^{-4}	N	
	comp13096_c0_seq1-novel		10.1	1.4×10^{-13}	N	
	comp13690_c0_seq2-novel		4.5	2.0x10-7	Y	
Z-strain Accessory glands						
	Hormone degradation					
	comp81285_c0_seq1-juvenile hormone diol kinase	Ostrinia furnacalis	10.0	7.1×10^{-3}	Y	2.6×10^{-124}
						Peptidases
	comp83286_c0_seq1-mitochondrial-processing peptidase	Papilio xuthus	11.1	8.1×10^{-5}	N	0
	Odorant binding and odorant receptors					
	comp60717_c0_seq1-odorant receptor	Ostrinia nubilalis	11.7	2.1×10^{-4}	N	0
	Transmembrane transport					
	comp199729_c0_seq1-voltage-dependent calcium channel	Plutella xylostella	10.1	2.62E-06	N	2.0×10^{-119}

	comp21137_c0_seq2-proton-coupled amino acid transporter 2	Bombyx mori	10.4	1.2 x10^{-3}	Y	2.8x10^{-80}
						Other enzymes
	comp25798_c0_seq1-malate dehydrogenase	Bombyx mori	10.5	6.4 x10^{-3}	N	0
	comp196282_c0_seq1-sodium potassium-transporting ATPase	Bombyx mori	10.6	3.6 x10^{-4}	N	1.1x10^{-87}
	comp147762_c0_seq1-sodium potassium -transporting ATPase	Plutella xylostella	10.6	2.8x10^{-3}	N	3.7x10^{-101}
	Proteins involved in RNA/DNA replication and cell homeostasis					
	comp58932_c0_seq1-gag-pol poly	Plutella xylostella	10.7	3.0x10^{-18}	N	0
	comp105991_c0_seq1-DNA polymerase	Glyptapantel es flavicoxis	13.2	2.7x10^{-21}	N	0
	comp120371_c0_seq1-retrovirus-related Pol poly from transposon TNT 1-94	Nicotiana tabacum	11.4	1.0x10^{-12}	N	0
	Unknown conserved proteins					
	comp117763_c0_seq1-uncharacterized protein LOC106712049	Tribolium castaneum	10.6	1.9x10^{-9}	Y	1.7x10^{-30}
	comp269618_c0_seq1-uncharacterized protein OBRU01_05634	Bombyx mori	10.4	8.3 x10^{-7}	N	9.5x10^{-129}
	Novel					
	comp15519_c0_seq1-novel		10.0	1.9 x10^{-4}	N	
	comp89511_c0_seq1-novel		10.1	8.87E-09	N	
	comp110530_c0_seq1-novel		10.1	1.39E-09	N	
	comp134168_c0_seq1-novel		10.1	4.3 x10^{-4}	N	
	comp12283_c0_seq1-novel		10.1	9.4 x10^{-4}	N	
	comp250035_c0_seq1-novel		10.2	2.08E-05	N	
	comp10803_c0_seq1-novel		10.3	9.8 x10^{-4}	N	
	comp70238_c0_seq1-novel		10.4	0.001440629	N	
	comp13096_c0_seq1-novel		10.6	7.03E-15	N	
	comp45107_c0_seq1-novel		10.6	0.001083207	N	
	comp160588_c0_seq1-novel		10.6	1.21E-10	N	
	comp137196_c0_seq1-novel		10.8	5.86E-07	N	
	comp15115_c0_seq1-novel		10.9	9.46E-09	N	
	comp177628_c0_seq1-novel		10.9	7.23E-07	Y	
	comp71851_c0_seq1-novel		10.9	6.0 x10^{-4}	N	
	comp17798_c0_seq1-novel		12.7	3.1 x10^{-4}	N	
	comp51750_c0_seq1-novel		11.2	9.5 x10^{-4}	N	
	comp79091_c0_seq1-novel		11.3	2.66 x10^{-5}	N	

	comp52058_c0_seq1-novel		11.3	4.74×10^{-7}	N	
	comp12301_c0_seq1-novel*		6.0	7.4×10^{-8}	Y	
	comp26192_c0_seq8-novel*		5.9	2.5×10^{-4}	Y	
	comp30059_c0_seq1-novel*		9.4	3.8×10^{-3}	Y	

Bold sequences are present in both tissues

* Sequences also identified as putative sfp

Chapter 4. Postmating transcriptional changes in the female reproductive tract of the European corn borer moth

Abstract

Mating triggers a cascade of physiological and behavioral responses in females that persist after copulation. Seminal fluid proteins contained within male ejaculates are known to initiate some responses, but our understanding of how females mediate these reactions remains limited. Few studies have examined postmating transcriptional changes within ejaculate-receiving organs within females or how these changes might depend on the identity of the male. Furthermore, whereas males of many insects transfer packaged ejaculates, transcriptional dynamics have mainly been examined in dipterans, in which males transfer a free ejaculate. To identify genes that may be important in mediating female physiological responses in a spermatophore-producing species, we sequenced the transcriptomes of the ejaculate-receiving organs and examined postmating gene expression within and between pheromone strains of the European corn borer moth (ECB), *Ostrinia nubilalis*. After within-strain mating, significant differential expression of 978 transcripts occurred in the female bursa or its associated bursal gland, including peptidases, transmembrane transporters, and hormone processing genes; such genes may potentially play a role in postmating male-female interactions. We also identified 14 transcripts from the bursal gland that were differentially expressed after females mated with cross-strain males, representing candidates for previously observed postmating reproductive isolation between ECB strains.

Introduction:

Mating triggers a long-lasting cascade of female physiological and behavioral responses that persist well beyond the conclusion of copulation (Lawniczak & Begun, 2004). These responses include decreased receptivity to mating, altered patterns of resource allocation, and increased oogenesis and oviposition (Sirot et al., 2009; Wolfner, 2009a; Gillott, 2003; Avila et al., 2011), and thus may play a central role in postmating sexual selection. Understanding the mechanisms underlying these physiological changes requires examining the patterns of gene expression occurring within both male and female reproductive tissues.

To date, considerable research has focused on identifying male genes and their protein products transferred to females during mating. These proteins, called seminal fluid proteins (SFPs), are produced by male accessory glands and the male ejaculatory duct (Avila et al., 2011) and have previously been characterized for *Tribolium* beetles (South et al., 2011), *Aedes* mosquitoes (Sirot et al., 2011; Boes et al., 2014), *Gryllus* crickets (Andres, 2006), *Heliconius* butterflies (Walters & Harrison, 2010), *Lutzomyia* sand flies (Azevedo et al., 2012), *Apis* honeybees (Baer et al., 2009) and *Drosophila* fruit flies (Wolfner et al., 1997; Swanson et al., 2001; Findlay et al., 2008). In dipterans, where SFPs have been most extensively studied, they have profound effects on females, including changes in sperm storage, oogenesis, oviposition, feeding, remating behavior, and lifespan (Avila et al., 2011; Heifetz & Wolfner, 2004; Wolfner, 2009a; Ravi Ram et al., 2005). For example, in *Drosophila melanogaster*, male reproductive

proteins induce both ovulation and sperm storage (Ravi Ram & Wolfner, 2007; Laflamme & Wolfner, 2013; Wolfner, 2009).

Although male reproductive proteins clearly have important effects on females, the female reproductive tract does not merely function as passive storage for the male ejaculate. Rather, studies have demonstrated active production of enzymes and membrane-bound transporters by portions of the female reproductive tract. Such female reproductive proteins have been characterized in several insects including *Drosophila* fruit flies (Mack et al., 2006; McGraw et al., 2008; Prokupek et al., 2008), *Apis* honeybees (Baer et al., 2009; Collins et al., 2006), *Anopheles* mosquitoes (Rogers et al., 2008), and *Pieris rapae* butterflies (Meslin et al., 2015a). Across these insect species, proteins in the female reproductive tract are important for protein cleavage, nutrient processing and the immune response (McGraw et al., 2004b; Prokupek et al., 2008).

To date, female response to mating has been examined mainly for species where males transfer a free ejaculate (but see Meslin et al. 2015). However, in numerous insects, males transfer their sperm packaged into a complex spermatophore manufactured by male accessory glands (Chapman, 2003; Lewis & South, 2012; Lewis et al., 2014). Such male spermatophores are ubiquitous among the Lepidoptera (Wedell, 2005), and often contain protein that females use for egg production and somatic maintenance (Boggs & Gilbert, 1979; Oberhauser, 1989). During mating, females receive the spermatophore inside a muscular organ called the bursa copulatrix (henceforth, the bursa). The resultant distension of the bursa stimulates stretch receptors, which causes females to become unreceptive to

additional mating (Sugawara 1979). While the presence of the spermatophore in the bursa thus reduces female remating, in the Ditrysia (a clade comprising ~98% of all Lepidoptera; Regier et al. 2009) oviposition continues unimpeded because females have a distinct reproductive opening for egg-laying.

Based on the dynamic postmating interactions that occur between the sexes within Lepidoptera, we expect sexual selection and sexual conflict to be major forces driving the evolution of female reproductive anatomy, physiology, and molecular traits. However, surprisingly few studies have looked at the female side of postmating interactions. Evidence suggests that lepidopteran females have evolved several unique reproductive traits related to processing male spermatophores. Embedded within the inner bursal wall is a chitinized, toothed structure called the signum (Chapman 2003). Aided by muscular contractions of the bursa, the signum functions to physically tear open the tough outer envelope of the male spermatophore (Cordero 2005, Galicia et al. 2008). Loss of this structure in several monandrous lineages suggests that the lepidopteran signum represents a reproductive trait selected under polyandry to speed up the degradation of male spermatophores (Sánchez et al., 2011), potentially allowing females to acquire more nuptial gifts. Recently, a comprehensive study of post-mating changes in gene expression in females of the Cabbage White butterfly *Pieris rapae* (Meslin et al. 2015) revealed that the female bursa also functions as a digestive organ to chemically break down male spermatophores, as well as to absorb and transport products of spermatophore digestion.

Lepidopteran females also have a prominent, semi-spherical gland attached to the anterior bursa, yet we have just begun to elucidate the function of this bursal gland. In *Ostrinia nubilalis*, the European corn borer, a previous study of gene expression within the bursa and bursal gland of virgin females (Al-Wathiqui et al., 2014) found evidence that the bursal gland also functions in spermatophore digestion by producing peptidases that may interact with male reproductive proteins. Continued study of both the bursa and bursal gland is critical to our understanding of male-female interactions not only within the Lepidoptera, but for other spermatophore-producing species as well.

Characterizing female reproductive proteins will not only enhance our understanding of postcopulatory sexual selection, it may also provide insight into mechanisms underlying postmating, pre-zygotic reproductive isolation. Due to sexual selection and conflict, male and female reproductive proteins evolve rapidly (Prokupek et al., 2010; Swanson, 2004; Swanson et al., 2001). These forces can lead to the evolution of postmating, prezygotic reproductive barriers between closely related species, which have been demonstrated in several insects (Price et al., 2001; Larson et al., 2011). Divergence between populations can occur due to changes in the coding regions of genes or in how genes expression is regulated. In *Drosophila mojavensis*, when females mate with conspecific males, changes in gene expression were shown to play a role in postmating incompatibilities (Bono et al., 2011). This shift in gene regulation between conspecific and heterospecific matings likely represents failed interactions between male and female reproductive proteins. By examining changes in gene

expression between populations, we can elucidate how regulation of gene expression contributes to such divergence.

Here, we examine patterns of gene expressions that occur after mating within the female reproductive tract of the European corn borer (ECB). Both sexes mate with multiple partners during their lifetimes (Fadamiro & Baker, 1999), suggesting that post-copulatory sexual selection and conflict have been important in shaping female reproductive traits. Furthermore, ECB moths are an emerging model for speciation with two distinct strains, called Z and E, which currently exhibit multiple reproductive barriers (Dopman et al., 2010). One such barrier stems from reduced female fecundity following cross-strain matings, and accounts for a 30% reduction in gene flow (Dopman et al., 2010). This postmating, prezygotic incompatibility is asymmetric: Z-strain females lay significantly fewer eggs after mating with an E-strain male. The mechanisms underlying this gametic isolation are unknown. Here, we use RNAseq to characterize how gene expression changes in the bursa and the bursal gland as females begin to process the male spermatophore, store sperm, and lay eggs. We also sought to identify candidate genes important for postmating male-female interactions and that could be responsible for the egg-laying dysfunction by examining differences in female gene expression at two timepoints after mating with either same-strain or cross-strain males.

Results:

De Novo Assembly

To characterize changes in female gene expression after mating, we sequenced RNA isolated from the bursa and bursal gland dissected from: 1) virgin females, 2) females that had just mated and 3) females 24 h after mating (Figure 1). In total, ~15,000,000 reads were obtained from each of our 29 sequenced libraries. All 29 libraries, including virgin female tissues, were assembled into a female reproductive tissues assembly with 40,952 contiguous sequences, a mean sequence length of 828 bp, and an n50 length of 1,412 bp.

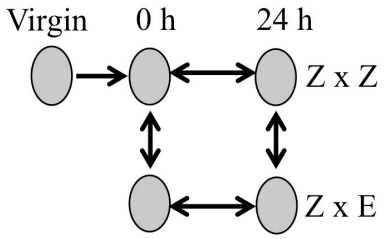


Figure 1. Experimental design comparing gene expression of Z- strain *O. nubilalis* females between 0 and 24 h timepoints and across mating types (same vs. cross-strain mating) for both the bursa and bursal gland.

Transcriptional Changes at Mating

Immediately after copulation (average duration 45 minutes), females have begun to store sperm and are beginning to break down the male spermatophore. Relative to that of virgin females (Figure 2A), within the first hour after mating (0 h timepoint) the female bursa was enlarged to contain the male spermatophore (Figure 2B), and the female bursal gland was filled with an opaque fluid (Figure 2C). By examining female reproductive tissues at this 0 h timepoint, we aimed to identify short-term changes in gene expression associated with mating and spermatophore presence.

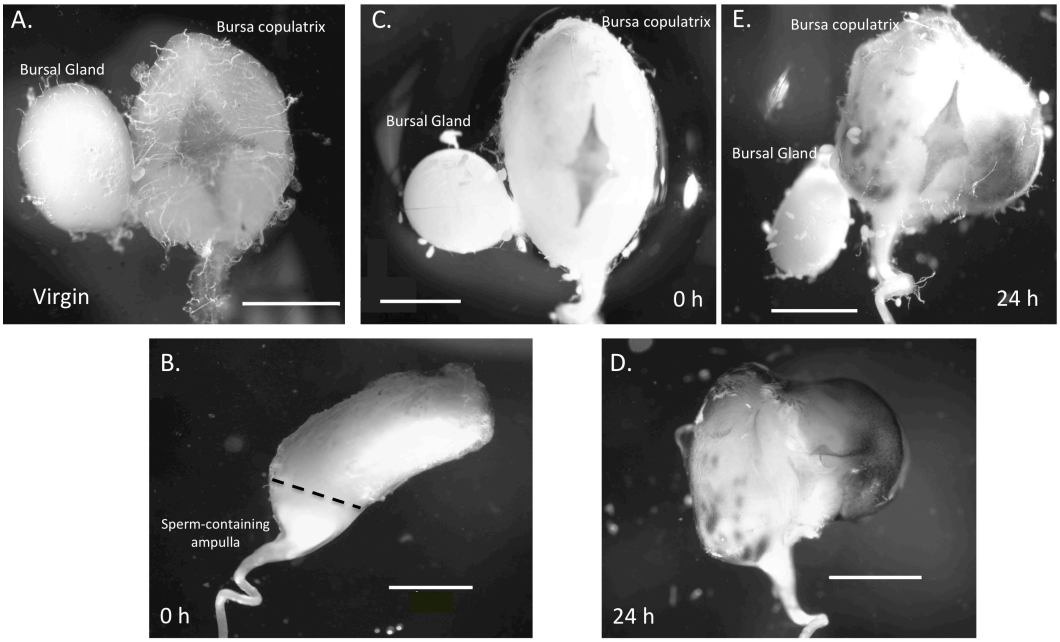


Figure 2. Changes in *O. nubilalis* female and male reproductive structures at different timepoints after mating. The female bursa copulatrix and bursal gland in virgin females (A; previously published in Al-Wathiqui et al., 2014), and at 0 h (n=18; C), and 24 h (n=18; E) timepoints. (B) and (D) show the male spermatophore removed from the female bursa copulatrix at 0 h and 24 h respectively. Scale bars represent 1 mm.

In the bursa, we found 345 genes that were significantly differentially expressed between virgin females and females immediately after mating; more than half of these genes (229) were significantly up-regulated (we used a log fold change ≥ 2 and a FDR ≤ 0.01 as criteria for significant up- or and down-regulation). Of the 229 up-regulated genes in the bursa immediately after mating, 60 were annotated and 169 genes had no annotation information associated with them (referred to as novel genes in Table 1).

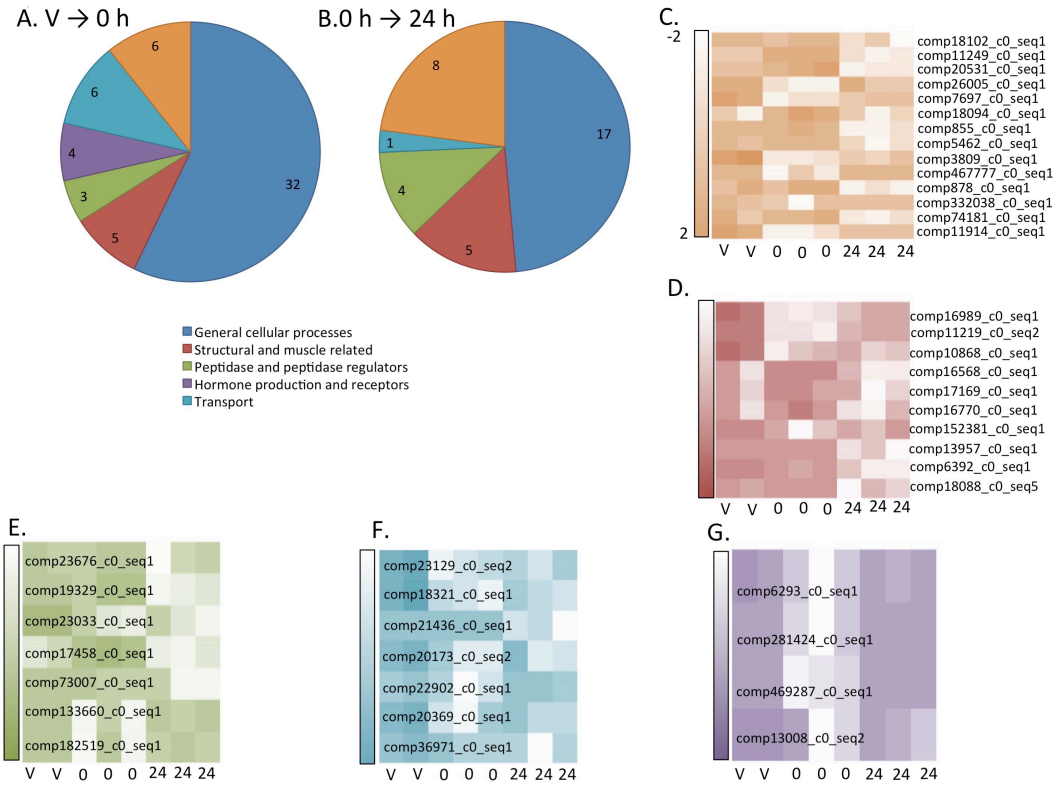


Figure 3. Female *O. nubilalis* gene expression in the bursa copulatrix before and after mating. (A) Gene ontology categories of up-regulated genes in virgin females compared to females 0 h after mating and (B) Gene ontology categories of up-regulated genes at 0 h compared to 24 h after mating. (C-G) Each heat map shows the expression patterns of genes (calculated using normalized read counts) up-regulated in the bursa at either the 0 h or 24 h timepoint for each of the annotation groups represented in the pie chart besides those genes annotated as proteins belonging to the "general cellular process" category. For each graph, genes that are up-regulated appears as a lighter color, corresponding to a z-score of -2 and genes that are down-regulated appear darker, corresponding to a z-score of 2 as shown in C. The genes ontology categories represented include "Other enzymes" (C), "Structural and muscle related" (D), "Peptidases and peptidase regulators" (E), "Transport" (F), and "Hormone production and receptors" (G).

Annotated genes included peptidases, genes involved in hormone production and binding, and sugar transporters (Figure 3A). The peptidases included homologs to two serine peptidases (comp182519_c0_seq1 and comp133660_c0_seq1) and a zinc metallopeptidase (comp23033_c0_seq1; Table 1); none of these genes contained secretion signal peptides. Several genes showed

significant homology to hormone receptors, including two nuclear hormone receptors, and comp13008_c0_seq2, an ecdysone-induced nuclear receptor. We also identified a gene that showed homology to juvenile hormone acid methyltransferase, a component of the juvenile hormone production pathway (comp6293_c0_seq1). We also found 116 down-regulated genes, that were annotated mainly as proteins with functional categories of DNA and protein binding, and cell maintenance and replication (Supplementary Table 1). We also identified two trehalose transport proteins (comp36971_c0_seq1 and comp23129_c0_seq2), which are important for transporting trehalose from its production site in the fat body to the hemolymph (Kanamori et al., 2010).

In the female bursal gland, we found 633 genes that were differentially expressed between virgin females and females immediately after mating. As in the bursa, more than half (351) of these were significantly up-regulated compared to virgin females, and 96 could be annotated (Figure 4a). Of the annotated proteins, 12 were annotated as peptidases (Table 2); six showed homology to serine-type peptidases with secretion signal peptides. Three of these secreted serine peptidases showed homology to the protein easter (comp30165_c0_seq1, comp15119_c0_seq1, and comp 73007) and one showed homology to the protein snake (comp15456_c0_seq1; Table 2). Both easter and snake are involved in establishing the dorsoventral axis during embryo development and potentially the immune response (Bevlin and Anderson, 1996).

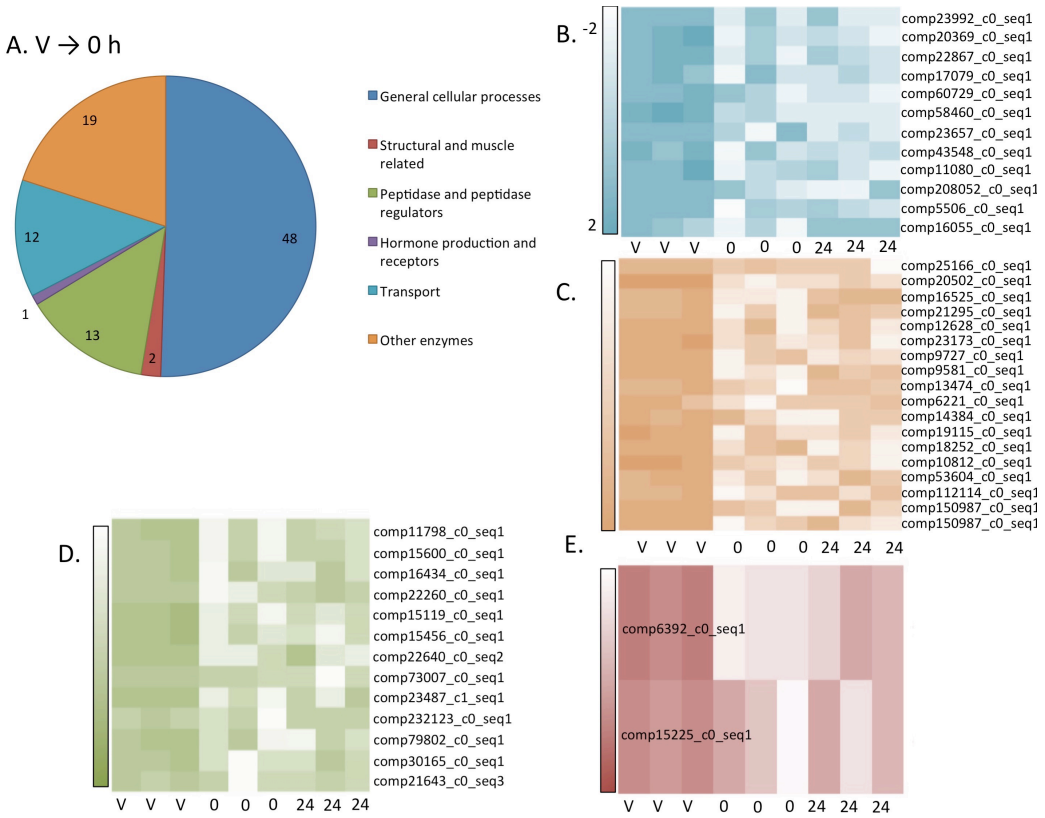


Figure 4. Female *O. nubilalis* gene expression in the bursal gland before and after mating. (A) Gene ontology categories of up-regulated genes in the virgin female bursal glands compared to females 0 h after mating. There was only one gene that was significantly up-regulated at the 24 h timepoint in the bursal gland, so there is no pie chart for this timepoint. (B-E) Each heat map shows the expression patterns of genes (calculated using normalized read counts) up-regulated in the bursal gland at either the 0 h timepoint for each of the annotation groups represented in the pie chart besides those genes annotated as proteins belonging to the "general cellular process" category. For each graph, genes that are up-regulated appears as a lighter color, corresponding to a z-score of -2 and genes that are down-regulated appear darker, corresponding to a z-score of 2 as shown in B. The genes ontology categories represented include "Transport" (B), "Other enzymes" (C), "Peptidases and peptidase inhibitors", and "Structural and muscle related" (D).

We found 282 down-regulated genes at 0 h postmating in the bursal gland (Supplementary Table 2), including three secreted serine peptidases (comp12052_c0_seq1, comp187511_c0_seq1, and comp12494_c0_seq1). We also identified five genes with significant homology to secreted peptidase

regulators, three of which act to inhibit chymotrypsin-like serine peptidases (comp167709_c0_seq1, comp19292_c0_seq1, and comp158200_c0_seq1), one inhibits serine peptidases (comp17954_c0_seq), and one inhibits cysteine peptidases (comp7457_c0_seq1; Supplementary Table 2).

Transcriptional changes 24 h after mating

By examining female reproductive tissues at this 24 h timepoint, we aimed to identify long-term changes in gene expression associated with female egg production and spermatophore degradation. At 24 h postmating, our dissections revealed that the spermatophore had become melanized (Figure 2D). In addition, the tough outer envelope of the male spermatophore had been torn open and the female signa was imbedded in the spermatophore (Figure 2 D & 2E). At this 24 h timepoint, the female bursal gland remained filled with an opaque fluid (Figure 2E).

We found 315 genes in the female bursa that were differentially expressed between the 0 h and 24 h timepoints. Eighty-three of these were significantly up-regulated, 35 of which were annotated These up-regulated genes belonged to GO functional classes (Figure 3B), that included general cellular processes, transport proteins (Figure 3C), structural proteins (Figure D), and peptidases and their inhibitors (Figure E). One peptidase that was up-regulated in the bursa (comp73007_c0_seq1); this sequence was previously up-regulated in the bursal gland at the 0 h timepoint. Between the 0 h and 24 h timepoints, 230 genes were down-regulated in the female bursa, 50 of which could be annotated. These

included transport proteins (Figure 3F; Supplementary Table 1) and hormone production and hormone receptor proteins (Figure 3G; Supplementary Table 1).

In contrast, gene expression in the bursal gland was relatively unchanged between 0 h and 24 h timepoints, with only 37 genes being significantly differentially expressed. Only two of these were significantly up-regulated: one could not be annotated, while the other (comp537_c0_seq1) showed homology to melanotransferrin, a cell-surface glycoprotein important for iron transport (Table 2). Out of all 37 bursal gland genes that were differentially expressed between 0 and 24 h postmating, 27 were significantly down-regulated, and these 27 genes were also differentially expressed between virgin females and 0 h postmating (Figure 4B, C, D, and E). Nine of these down-regulated genes were annotated, and their functional categories included DNA and protein binding, and cell maintenance (Supplementary Table 2).

Response to mating within versus across strains

To identify genes potentially important for reproductive isolation, we examined how gene expression changed in reproductive tissue of Z-strain females that had mated with either same-strain or cross-strain males. Immediately after same-strain mating, three bursal genes were up-regulated compared to cross-strain matings, but these could not be annotated (Table 3). In the bursal gland, no genes were significantly differentially expressed between the two mating types at 0 h postmating.

Next, we looked for differentially expressed genes in the female reproductive tract of female who had with mated with within and across-strain males 24 hours after mating. At this timepoint, we found no differentially expressed genes between the two mating types in the female bursa, yet 11 genes were differentially expressed in the bursal gland. Nine of these genes were annotated (Table 3), showing homology to proteins involved in metal ion binding (comp537_c0_seq1), protein binding (comp118661_c0_seq1), and protein phosphorylation (comp125320_c0_seq1).

Discussion:

Our study provides insights into how gene expression changes in the female reproductive tract of ECB, a spermatophore transferring species. Within the first 24 h postmating ECB females have stored male sperm, begun the process of breaking into the male spermatophore, and laying eggs. By examining changes in the bursa and bursal gland immediately after mating, we have identified genes that may be involved in these important postmating processes. We found that, as predicted by an earlier study (Al-Wathiqui et al, 2014), the female bursa does not act as a secretory organ after mating. Instead, genes that were up-regulated in the bursa immediately after mating were involved in, hormone reception, and sugar transport. At 24 h postmating, these categories of genes were down-regulated in the bursa, indicating that there is an initial short-term response to mating that takes place in the bursa. We were also able to confirm our previous hypothesis that the bursal gland functions as a secretory organ and functions to secrete proteases immediately after mating. Following mating, we also noted a down-

regulation of peptidase regulators in the bursal gland. After 24 h, gene expression in the bursal gland remained stable, as few differentially expressed genes were observed. We also identified 14 genes in the female reproductive tract of Z-strain ECB that were differentially expressed after females had mated within or across strains. These represent candidate genes responsible for the postmating, prezygotic reproductive isolating barrier acting between two ECB strains.

Short-term changes in gene expression

In many insects, females do not maintain a complement of fully mature eggs; more typically, eggs are matured after mating (Chapman, 2003). Furthermore, mating has been shown to increase egg maturation and oviposition in a number of insects, including *D. melanogaster* fruitflies (Chapman et al., 1995)*, Ceratitis capitata* fruit flies (Chapman et al., 1998), *Aedes agypti* mosquitoes (Klowden & Chambers, 1991), and ECB (Fadamiro & Baker, 1999). Here, we found immediate up-regulation of homologs to two hormone receptors important for vitellogenesis, a key process in egg production (Figure 3A and G). Two *Drosophila* proteins called E75 and HR38 act as nuclear receptor proteins involved in regulating the hormone ecdysone, which is responsible for activating yolk protein precursor genes in the female fat body and mediating the production of vitellogenin (Deitsch et al., 1995a; Deitsch et al., 1995b; Pierceall et al., 1999; Cruz et al., 2012; Xu et al., 2010). E75 is an ecdysone- induced nuclear receptor that, in mosquitoes, is activated by 20-hydroxyecdysone at the onset of vitellogenesis (Pierceall et al., 1999; Cruz et al., 2012). HR38 acts as a negative

regulator of the ecdysone receptor pathway that includes E75 (Xu et al., 2010). Up-regulation of these two genes immediately postmating may indicate that females are beginning to shift from inhibiting to promoting vitellogenesis.

Two genes with homology to trehalose transporters were up-regulated immediately after mating (Table 1). The trehalose transporter, Tret1, is responsible for transporting trehalose synthesized in insect fat body to hemolymph (Kanamori et al., 2010). Postmating up-regulation of trehalose transporters has also been shown in the female bursa of *Pieris rapae* butterflies (Meslin et al., 2015). In ECB, the bursa is surrounded by fat body and trehalose transporters may function in supplying energy to support active muscle contractions needed to break down male spermatophores.

Peptidases and their inhibitors are a conserved class of male reproductive proteins that have also been shown to be expressed in female reproductive tissue. Peptidases and peptidase regulators have garnered substantial attention because they may mediate male-female reproductive interactions (Laflamme & Wolfner, 2013). In characterizing male reproductive proteins in *D. melanogaster*, researchers have identified peptidases responsible for multiple post-mating processes, including sperm activation and storage and increasing egg maturation and oviposition. At least three SFPs in *D. melanogaster* (Ovulin, Acp36DE and CG11864) have been shown to undergo processing into their active form inside the female reproductive tract (Ravi Ram et al., 2005). For example, ovulin is cleaved into smaller peptides inside the female reproductive tract by male and possibly by female factors, which allows for release of mature oocytes from the

ovary (Heifetz et al., 2001; Ravi Ram & Wolfner, 2007; Wolfner, 2009). Acp36DE, required for sperm storage, is also processed by peptidases inside the female (Ram et al., 2006).

Previously, we characterized the virgin female bursal gland as secreting many peptidases, which we hypothesized could be important for postmating intersexual interactions (Al-Wathiqui et al, 2014); in the current study, we found more support for this hypothesis. We identified six peptidases with secretion signal peptides that were up-regulated immediately after mating compared virgin females (Figure 4 A and D). Four were serine peptidases, which are among the most common class of peptidases identified in male and female insects (Laflamme & Wolfner, 2013). These included genes with homology to the proteins snake and easter, which constitute part of the Toll pathway in *Drosophila melanogaster*, a development and immunity pathway (Belvin & Anderson, 1996). Activated by the presence of bacteria and fungi, in *D. melanogaster* females the Toll pathway is activated by the SFP sex peptide, (Peng et al., 2005). Mating can introduce microbes into the female reproductive tract, and up-regulation of immune response genes occurs in the *D. melanogaster* female reproductive tract after mating (McGraw et al., 2004). In ECB females, up-regulation of these genes may protect against infection and enhance female survival. Another serine peptidase secreted from the bursal gland that was up-regulated immediately after mating was a chemotrypsinogen-like protein (comp16434_c0_seq1). This is a digestive peptidase usually found in the insect midgut (Mazumdar-Leighton & Broadway, 2001). In *P. rapae*, digestive peptidases have also been found in the

female reproductive tract and are thought to aid in spermatophore processing (Meslin et al., 2015). Up-regulation of this gene in the ECB bursal gland may similarly serve to help break down the male spermatophore.

Peptidase regulators may also be important in mediating post-mating interactions between the sexes. We identified five secreted peptidase regulators, including four serine peptidases regulators, that were down-regulated in the bursal gland immediately postmating (Supplementary Table 2). In *D. melanogaster*, two peptidase inhibitors were similarly down-regulated in whole female flies after mating (McGraw et al., 2004) .

Long-term changes in gene expression between 0-24 h

In polyandrous species where males deposit spermatophores in the female reproductive tract, females benefit from rapidly processing male ejaculates in order to mate again. Two mechanisms appear to be involved in breaking down the male spermatophore. Mechanical disruption is accomplished through movements of the muscular bursa; as the bursa contracts, the chitinized, toothed signum imbedded in the bursal wall physically abrades the spermatophore (Cordero, 2005; Sugawara, 1979). Biochemical digestion of the spermatophore subsequently takes place, although this process is less well understood.

A recent study using the cabbage white butterfly, *P. rapae* has begun to elucidate the roles of mechanical and chemical digestion in spermatophore degradation (Meslin et al., 2015b). Meslin and colleagues examined gene expression in the female bursa of *P. rapae* at three timepoints (0 h, 24 h, and 72 h)

after mating. Muscle-related genes showed high expression levels in the bursa, along with a number of digestive genes encoding peptidases, nutrient processing genes, and transporters; 11 genes, including lipases, were expressed only in the bursa (Meslin et al., 2015b).

In the present analysis of bursal gene expression in mated ECB females, we found that several genes were up-regulated between 0 h and 24 h with homology to structural and muscle related-proteins (Figure 3B and D); these included talin-1 (comp17169_c0_seq1), and muscle-specific protein 20 (comp18088_c0_seq5). Muscle-specific protein 20 is found in synchronous muscle in *D. melanogaster* (Ayme-Southgate et al., 1989), while talin-1 helps maintain cell integrity during muscle contractions (Vigoreaux, 1994). Up-regulation of these genes may be associated with increased activity of bursal muscle as it initiates spermatophore processing.

In the bursal gland of ECB females, gene expression remained relatively stable between 0 and 24 h. We did find down-regulation of trehalose transport genes that were initially up-regulated after mating (Figure 3B; Supplementary Table 2). This expression pattern could indicate a transient increase in trehalose transport to support the muscular contractions involved in initially breaking open the spermatophore. Although we chose to limit our time course to 24 h (at which point females have started to lay fertilized eggs), it may be worthwhile to examine expression in the bursal gland at later timepoints.

Differences in female gene expression following same- versus cross-strain mating

Sexual conflict extends beyond mating, as the reproductive arms race continues while males and females struggle for control over fertilization (Arnqvist & Rowe, 2005; Chapman et al., 2003). Such antagonistic sexual coevolution has the potential to contribute to divergence between closely related populations, as may have occurred in a other insects (Price et al., 2001; Wade et al., 1994; Larson et al., 2011). Two ECB strains show an asymmetric postmating reproductive barrier that manifests as significantly reduced egg production by Z-strain females that have mated with E-strain males (Dopman et al., 2010). We examined differences in gene expression of Z-strain females following matings to same- *versus* cross-strain males to identify candidate genes that could be involved in reproductive isolation.

Fourteen genes were significantly differentially expressed between these mating types in the female bursa at 0 h and the bursal gland 24 h after mating, and nine were successfully annotated (Table 3). One up-regulated gene expressed in the female bursal gland following within-strain matings was 5-aminolevulinate synthase (Table 3), which is up-regulated in rat follicular tissue during ovulation (Espey & Richards, 2002). This gene controls heme supply to many tissues, and when heme biosynthesis is blocked in the bug *Rhodnius prolixus*, females lay significantly fewer eggs due to reduced heme binding protein in the hemolymph (Braz et al., 2001). Expression of 5-aminolevulinate synthase in the ECB bursal gland following within-strain mating might increase female egg production by

facilitating heme transport to different portions of the female reproductive tract. Thus, altered gene expression in the bursal gland after females have mated with E-strain males may be associated reduced egg production that has been previously observed.

Interestingly, some of the most highly differentially expressed genes between the two mating types were novel genes (Table 3) that did not show homology to known genes in any organism listed in the NCBI non-redundant database. Reproductive proteins are known to be rapidly evolving (Panhuis, 2006; Swanson & Vacquier, 2002). It is possible that the novel genes differentially expressed between the two ECB mating types may have diverged from known proteins. These gene sequences represent ideal candidates for further studies in ECB measuring the presence and extent of evolutionary pressures on female reproductive proteins.

While this study provides new insight into postmating changes in female gene expression, functional studies will be needed to determine the roles played by these genes in ECB reproduction. Our discovery that multiple female reproductive genes are differentially expressed between ECB mating types also opens exciting new avenues for research on mechanisms of postmating reproductive isolation.

Methods:

Insect rearing and mating

European corn borer moth populations of both strains have been maintained at Tufts University by mass rearing ~200 adults per generation; methods have been previously described (Dopman et al., 2005). All parings were conducted in an incubator (26°C) with a 8L:16D light cycle.

To conduct same- and cross-strain parings, two-day-old Z-strain females were paired with either two-day-old Z-strain males or two-day old E-strain males. All parings were conducted in 6 X 10 cm paper cups with mesh covers to allow for airflow. Pairs were monitored every 15 minutes until mating was observed. After a mating pair separated, the mated female was randomly assigned to be dissected either immediately (0 h; n=60 or 24 h; n=60) after the completion of mating. For each female, the bursa and bursal gland were dissected under 20X magnification in RNAlater. The male spermatophore was removed from the bursa and the bursa tissue was washed in RNAlater to avoid contamination of male-derived RNA. The fluid in the bursal gland was not removed because previous work demonstrated that the bursal gland acts as a secretory sac (Al-Wathiqui et al., 2014). All fat body was removed from the bursa and bursal gland and tissues were stored at -80°C in RNAlater.

RNA extractions and library prep

Prior to RNA extraction, tissues from 6 individuals were pooled to create 3 replicates for each mating type and timepoint. Total RNA was extracted (RNeasy Midi kit, Qiagen, California) and cDNA library was constructed

(Illumina Truseq RNA sample preparation kit v2, San Diego, CA). A total of 12 bursal gland cDNA libraries (n = 3 per strain and per timepoint) and 12 bursa cDNA libraries (n = 3 per strain and per timepoint) were sequenced using an Illumina HiSeq2000 (50bp single-end reads, 12 libraries per lane).

Virgin female libraries for bursal gland and bursa tissues created in a previous study (Al-Wathiqui et al., 2014) were used for comparisons with females 0 h post-mating. All virgin females were also from the Tufts University ECB colony and 2-days old at the time of dissection. For each replicate, 4 tissues were pooled prior to RNA extraction. This resulted in 2 bursa copulatrix samples and 3 bursal gland samples from virgin females. These samples were also prepared using the Illumina Truseq RNA sample preparation kit v2 and were sequenced on an Illumina HiSeq2000 (50 bp single-end reads).

Quality control and assembly

After sequencing, all libraries from virgin and mated female tissues were subjected to multiple quality control steps. Illumina adapters and trailing bases were removed from all libraries using Trimmomatic (Lindgreen, 2012), and all libraries were assembled into a single transcriptome using Trinity version 2013-02-25 and default program settings (Grabherr et al., 2011). After transcriptome assembly, the longest transcript was used as the best-assembled transcript for each locus. Assembly statistics were obtained using the python script, asemstats.py.

Differential expression

Specific tissues and timepoints that we compared are shown in figure 2. To identify how gene expression changed in the bursa and bursal gland

immediately after mating, we compared gene expression between Z-strain virgin female tissues and female tissues dissected from the female immediately after mating with a Z-strain male. We also compared genes expression from Z-strain females dissected immediately after mating with a Z-strain male, and female tissues dissected 24 h after mating with a Z-strain male to identify genes important in spermatophore breakdown. We also made these two comparisons for Z-strain females that mated with an E-strain male to identify genes that could be involved in reproductive isolation.

Due to the differences in when and how virgin and mated tissue libraries were prepared, we first normalized the read counts produced by each sample. All read counts were normalized using EDAseq, which uses within- and between-sequencing lane normalization to account for differences in read-length and GC content of reads (Risso et al., 2011). Using this program, we were able to minimize any differences between the virgin female libraries and the mated female libraries that could be due to variation in sample prep or date of sequencing (Supplementary Figures 1 and 2). Normalized reads were then mapped to our transcriptome using Bowtie 2 (Langmead et al., 2009; Langmead & Salzberg, 2012), and tested for differential gene expression. We then identified differentially expressed genes based on average expression of each gene across replicates using edgeR (McCarthy et al., 2012; Robinson & Oshlack, 2010; Robinson et al., 2009; Robinson & Smyth, 2007) to determine how genes change in response to mating and during spermatophore processing. Genes were considered differentially expressed if they had a false-discovery rate (FDR) of <

0.01 and an expression log fold change ≥ 2 (log_2FC). We examined both up- and down-regulated genes at each timepoint; however, we focus here on up-regulated genes as the majority of down-regulated genes in each tissue and each timepoint were related to the general cell maintenance (Supplemental Table 1).

To determine the function of significantly differentially expressed genes, we used Blast2go to identify homologous sequences in other species (Gotz et al., 2008; Conesa & Götz, 2008; Conesa et al., 2005). First, a blastx was conducted for all sequences against the NCBI non-redundant protein database using an e-value cutoff of $\leq 1.0X10^{-3}$. Next, conserved protein motifs were identified for each sequence using InterProScan, followed mapping and annotation analyses conducted using Blast2Go. The annotation step requires that sequences have a top blast hit with an e-value of $\leq 1X10^{-6}$. We then used the gplots package Heatmap.2 to create heat maps of annotated and up-regulated genes in the bursa and bursal gland (Warnes et al., 2008). The z-scores used to plot the data were calculated from the normalized read counts of each gene at each timepoint for each replicate. We used z-scores to scale rows, so that genes with similar expression patterns are closer to each other on the heat map.

We then used SignalP 4.0 (Petersen et al., 2011)and TMHMM 2.0 (Sonnhammer et al., 1998) to identify sequences that contain secretion signals and transmembrane helices. To do this, we first used TransDecoder to predict the open reading frames for our transcriptome, as SignalP and TMHMM require protein sequence as an input (Haas et al., 2013). This allowed us to characterize

sequences that are secreted from or act as receptors in the reproductive tissue and could potentially interact with male sfps.

Table 1. Sequences significantly up-regulated in the ECB bursa copulatrix at the 0 h and 24 h timepoints after mating.

Time point	Category	Sequence ID and description	Log_2 FC	FDR	Signal peptide	Trans-membrane region	e-value
0 h							
	Peptidases						
		comp182519_c0_seq1-serine protease nudel	-6.2	0.0001	N	Y	0
		comp133660_c0_seq1-serine protease nudel	-5.0	0.0004	N	N	$5.4x10^{-133}$
		comp23033_c0_seq1-carboxypeptidase n subunit 2-like	-2.4	0.004	N	N	$3.0x10^{-137}$
	Structural and muscle related						
		comp152381_c0_seq1-clip-associating protein	-6.7	0.004	N	N	$1.8x10^{-34}$
		comp11219_c0_seq2-alpha skeletal muscle	-3.1	$2x10^{-5}$	N	N	$3.3x10^{-64}$
		comp10868_c0_seq1-heat shock protein	-3.1	0.007	N	N	$2.2x10^{-67}$
		comp6392_c0_seq1-cuticular protein 65av	-3.1	0.002	N	N	$2.2x10^{-47}$
		comp16989_c0_seq1-truncated actin-4	-2.0	0.002	N	N	$1.0x10^{-109}$
	Hormone production and receptors						
		comp281424_c0_seq1-probable nuclear hormone receptor hr38	-5.3	$7x10^{-5}$	N	N	$4.1x10^{-158}$
		comp469287_c0_seq1-probable nuclear hormone receptor hr38	-4.9	$5x10^{-5}$	N	N	$2.2x10^{-148}$
		comp13008_c0_seq2-e75 nuclear receptor	-3.5	0.003	N	N	$1.7x10^{-80}$
		comp6293_c0_seq1-juvenile hormone acid methyltransferase	-3.4	0.007	N	N	$1.8x10^{-114}$
	Metabolism						
		comp282419_c0_seq1-1-phosphatidylinositol -bisphosphate phosphodiesterase gamma-1 isoform x2	-5.3	0.008	N	N	$3.8x10^{-112}$
		comp26052_c0_seq1-tyrosine hydroxylase	-4.5	0.005	N	N	0
		comp23209_c0_seq1-lipase 3-like	-3.9	0.0008	Y	N	0
		comp16329_c0_seq1-myrosinase 1-like	-3.5	0.0003	Y	N	0
		comp10645_c0_seq1-inositol hexakisphosphate kinase 2 isoform x2	-2.5	0.0005	N	N	$3.5x10^{-136}$
	Immune response and stress response						
		comp15243_c0_seq1-growth arrest and dna damage-inducible protein gadd45 alpha	-3.1	0.0002	N	N	$4.9x10^{-96}$
		comp187526_c0_seq1-interleukin enhancer-binding factor 2 homolog	-4.4	0.003	N	N	$2.3x10^{-27}$
	Transport						
		comp20173_c0_seq2-solute carrier family 25 member 35-like isoform x1	-3.9	$1x10^{-5}$	Y	N	$1.8x10^{-115}$

		comp36971_c0_seq1-facilitated trehalose transporter tret1-like	-4.3	0.003	N	Y	0
		comp22902_c0_seq1-open rectifier potassium channel protein 1 isoform x1	-3.1	0.006	N	Y	0
		comp20369_c0_seq1-organic cation transporter	-2.4	0.004	N	Y	0
		comp18321_c0_seq1-cationic amino acid transporter 3 isoform x1	-2.2	0.0002	N	Y	$9.6x10^{-150}$
		comp23129_c0_seq2-facilitated trehalose transporter tret1-like	-2.0	0.009	N	Y	0
	DNA , RNA, Protein Binding and folding						
		comp54887_c0_seq1-early growth response protein 2	-7.8	0.006	N	N	1.1x10-^{102}
		comp3014_c0_seq1-heat shock protein 70	-6.7	0.001	N	N	$7.8x10^{-114}$
		comp1957_c0_seq1-heat shock protein 70	-5.7	0.006	N	N	$2.9x10^{-70}$
		comp38119_c0_seq1-heat shock protein 70	-5.0	0.001	N	N	$5.2x10^{-50}$
		comp36096_c0_seq1-heat shock protein 70	-4.7	0.004	N	N	0
		comp9890_c0_seq1-zinc finger protein 704	-3.9	0.0005	N	N	$3.9x10^{-139}$
		comp7858_c0_seq1-PREDICTED: uncharacterized protein LOC105283242	-3.62	0.002	N	N	0
		comp23313_c0_seq1-arf-gap with dual ph domain-containing protein 1-like isoform x1	-3.6	$5x10^{-7}$	N	N	0
		comp21137_c0_seq1-dna-binding protein d-ets-4	-3.6	$3x10^{-5}$	N	N	0
		comp8433_c0_seq1-eukaryotic translation initiation factor 4e type 2	-6.0	0.0006	N	N	$9.3x10^{-47}$
		comp16641_c0_seq1-hypothetical protein KGM_08119	-3.3	0.0002	N	N	$6.6x10^{-8}$
		comp21528_c0_seq2-programmed cell death protein 6 isoform x2	-3.0	0.0001	N	N	8.0E-115
		comp1187_c0_seq1-rna-directed dna polymerase from mobile element jockey-like	-3.0	0.004	N	N	0
		comp12261_c0_seq1-transcription factor hivep3 isoform x2	-2.6	0.0001	N	N	0
		comp17033_c0_seq1-zinc finger protein 350-like isoform x1	-2.5	0.003	N	N	$5.1x10^{-55}$
		comp14847_c0_seq1-ets dna-binding protein pokkuri	-3.1	$5x10^{-6}$	N	N	0
		comp16752_c0_seq1-cryptochrome 2	-2.1	0.005	N	N	$2.1x10^{-105}$
		comp22263_c0_seq1-transcription factor hivep3	-2.2	0.0003	N	N	0
	Membrane components and receptors						
		comp152197_c0_seq1-5-hydroxytryptamine receptor 2a isoform x1	-5.5	0.008	N	Y	$4.2x10^{-132}$
		comp10279_c0_seq1-23 kda integral membrane	-3.5	$5X10^{-5}$	N	Y	$1.3x10^{-48}$

	comp13400_c0_seq2-g-protein coupled receptor mth2-like	-3.1	0.003	N	Y	$2.5x10^{-103}$
	comp140513_c0_seq1-progestin and adipoq receptor family member 3	-2.8	0.002	N	Y	0
	comp18852_c0_seq3-progestin and adipoq receptor family member 3	-2.7	0.0002	N	N	$2.2x10^{-9}$
	comp21994_c0_seq1-23 kda integral membrane	-2.5	0.0002	N	Y	$4.1x10^{-90}$
	comp6344_c0_seq1-cd63 antigen	-2.0	0.007	N	Y	$4.0x10^{-137}$
Other enzymes						
	comp332038_c0_seq1-e3 ubiquitin-protein ligase ubr3 isoform x1	-4.5	0.009	N	N	$5.5x10^{-42}$
	comp3809_c0_seq1-reverse transcriptase	-2.9	0.005	Y	N	$1.7x10^{-8}$
	comp467777_c0_seq1-cytochrome p450	-4.3	0.009	N	N	2.2x10-86
	comp26005_c0_seq1-inositol hexakisphosphate kinase 2 isoform x2	-2.7	$7x10^{-5}$	N	Y	$4.7x^{-10}$
	comp7697_c0_seq1-serine threonine-protein kinase sik2	-2.4	0.003	N	N	0
	comp11914_c0_seq1-serine threonine-protein kinase sik2	-2.3	0.0009	N	N	0
Signal transduction						
	comp14401_c0_seq1-socs2-12 protein	-3.7	$8x10^{-8}$	N	N	$6.1x10^{-11}$
	comp11972_c0_seq1-suppressor of cytokine signaling 2	-3.6	$8x10^{-8}$	N	N	$1.4x10^{-79}$
	comp12579_c0_seq1-ras-like gtp-binding protein	-2.2	0.009	N	N	$1.8x10^{-114}$
24 h						
Peptidases						
	comp23676_c0_seq1-uncharacterized protein LOC103515293	-2.1	0.007	N	N	0
	comp73007_c0_seq1- serine protease easter-like	-2.9	0.008	Y	N	$3.7x10^{-106}$
	comp17458_c0_seq1-tripeptidyl-peptidase 2	-2.7	0.002	N	N	$5.4x10^{-63}$
	comp19329_c0_seq1-caspase-1	-2.1	0.0008	N	Y	0
Structural and muscle related						
	comp13957_c0_seq1- isoform x1	-3.7	0.007	Y	N	$4.8x10^{-69}$
	comp17169_c0_seq1-talin-1 isoform x1	-3.0	0.003	N	N	$1.7x10^{-137}$
	comp16568_c0_seq1-microtubule-actin cross-linking factor 1 isoform x1	-2.5	0.0005	N	N	0
	comp18088_c0_seq5-muscle-specific protein 20-like	-2.2	0.002	N	N	$1.5x10^{-113}$
	comp16770_c0_seq1-apolipophorins	-2.0	0.004	N	N	$2.1x10^{-53}$
Immune response and Stress response						
	comp10978_c0_seq1-proteoglycan-4	-2.1	0.008	N	N	$2.6x10^{-34}$
	comp23927_c0_seq1-lethal essential for life l2efl	-2.1	0.0002	N	N	$1.5x10^{-103}$
Transport						
	comp21436_c0_seq1-amino acid	-2.7	$5x10^{-5}$	N	Y	0

	transporter					
DNA , RNA, Protein Binding and folding						
	comp537_c0_seq1-melanotransferrin	-4.4	0.001	N	N	$6.9x10^{-38}$
	comp19745_c0_seq1-uncharacterized f-box lrr-repeat	-3.2	0.0002	N	N	$1.4x10^{-138}$
	comp231_c0_seq1-small optic lobes protein	-3.1	0.005	N	N	$1.6x10^{-51}$
	comp23796_c0_seq1-irregular chiasm c-roughest protein isoform x1	-3.0	0.0002	Y	Y	0
	comp20507_c0_seq1-kelch-like protein 5	-2.5	0.0005	N	N	0
	comp22791_c0_seq1-ankyrin-1-like isoform x1	-2.4	0.008	N	N	0
	comp161254_c0_seq1-kinesin-like protein kif16b	-2.2	0.005	N	N	$7.5x10^{-178}$
	comp21974_c0_seq1-low-density lipoprotein receptor-related protein 2	-2.1	0.005	N	N	0
	comp22186_c0_seq1-low-density lipoprotein receptor-related protein 2	-2.1	0.007	N	N	0
	comp17896_c0_seq1-related protein rab-24-like	-2.0	0.002	N	Y	$2.1x10^{-119}$
Membrane components and receptors						
	comp3491_c0_seq1-otoferlin-like isoform x2	-4.8	0.0002	N	N	$1.6x10^{-151}$
	comp21040_c0_seq1-elongation of very long chain fatty acids protein aael008004-like	-2.7	$8x10^{-5}$	N	N	$3.5x10^{-19}$
	comp15329_c0_seq1-elongation of very long chain fatty acids protein aael008004-like	-2.5	$3x10^{-5}$	N	Y	$6.3x10^{-156}$
	comp82206_c0_seq1-inositol - trisphosphate receptor isoform x1	-2.8	0.002	N	Y	$2.2x10^{-84}$
	comp20585_c0_seq1-platelet glycoprotein v	-2.3	0.003	N	N	0
Other enzymes						
	comp5462_c0_seq1-serine threonine-protein kinase nlk	-3.8	0.0003	N	N	$1.7x10^{-174}$
	comp18102_c0_seq1-peptide methionine sulfoxide reductase	-3.6	0.0002	**N**	N	$6.1x10^{-142}$
	comp878_c0_seq1-e3 ubiquitin-protein ligase hecw2-like	-3.1	0.009	N	N	$3.8x10^{-47}$
	comp74181_c0_seq1-ubiquitin-conjugating enzyme e2q-like protein 1	-2.8	0.006	N	N	$4.7x10^{-12}$
	comp11249_c0_seq1-ubiquitin conjugation factor e4 b	-2.8	$2x10^{-6}$	N	N	$1.1x10^{-112}$
	comp18094_c0_seq1-low quality protein: baculoviral iap repeat-containing protein 6	-2.4	0.002	N	N	$3.3x10^{-149}$
	comp20531_c0_seq1-methyltransferase-like protein 13	-2.2	0.001	N	N	$9.3x10^{-81}$
	comp855_c0_seq1-adenylate cyclase type 2	-3.7	0.0004	N	Y	0

All bold sequences were found to be differentially expressed at both timepoints.

Table 2. Sequences significantly up-regulated in the ECB bursal gland at the 0 h and 24 h timepoints after mating.

Time point	Category	Sequence ID and description	Log_2FC	FDR	Signal peptide	Trans-membrane region	e-value
0 h							
	Peptidases						
		comp232123_c0_seq1-protein rhomboid	-4.0	0.001	N	Y	$1.3x10^{-68}$
		comp79802_c0_seq1-serine protease easter-like	-3.7	0.002	N	N	$4.9x10^{-16}$
		comp23487_c1_seq1-pattern recognition serine proteinase precursor	-3.6	$1x10^{-5}$	Y	N	$1.3x10^{-110}$
		comp16434_c0_seq1-chymotrypsinogen-like protein	-3.5	0.006	N	Y	0
		comp73007_c0_seq1-serine protease easter-like	-3.4	0.001	Y	N	$1.2x10^{-107}$
		comp15600_c0_seq1-aael012143- partial	-2.9	0.0007	N	N	$9.0x10^{-11}$
		comp22640_c0_seq2-gag-pol polyprotein	-2.9	0.003	N	N	0
		comp11798_c0_seq1-caspase-3	-2.9	0.002	N	N	$1.7x10^{-42}$
		comp15119_c0_seq1-serine protease easter-like	-2.3	0.010	Y	N	0
		comp21643_c0_seq3-venom protease-like isoform x2	-2.3	0.003	Y	Y	0
		comp30165_c0_seq1-serine protease easter-like isoform x1	-2.2	0.0008	Y	N	0
		comp15456_c0_seq1-serine protease snake-like	-2.0	0.006	Y	N	$2.2x10^{-71}$
	Peptidase regulators						
		comp22260_c0_seq1-ovalbumin-related protein x-like	-2.6	0.005	N	Y	$2.9x10^{-141}$
	Structural and muscle related						
		comp6392_c0_seq1-endocuticle structural glycoprotein abd-5-like	-3.4	$8x10^{-6}$	N	N	$2.2x10^{-47}$
		comp15225_c0_seq1-peritrophin a	-3.3	0.0002	N	N	$1.0x10^{-147}$
	Hormone production and receptor						
		comp6293_c0_seq1-juvenile hormone acid methyltransferase	-5.2	$1x10^{-7}$	N	N	$8.6x10^{-115}$
	Metabolism						
		comp19923_c0_seq1-lipase 3-like	-4.9	0.0004	Y	N	0
		comp23209_c0_seq1-lipase 3-like	-4.2	$2x10^{-06}$	Y	N	0
		comp23038_c0_seq1-lipase 3-like	-4.0	$3x10^{-5}$	Y	N	0
		comp61573_c0_seq1-glutathione s-transferase delta 1	-2.8	0.0005	N	N	$9.2x10^{-137}$
		comp48849_c0_seq1-pancreatic triacylglycerol lipase	-2.2	0.004	N	N	$1.2x10^{-127}$
	Immune response and stress response						
		comp54007_c0_seq1-protein lethal essential for life-like	-4.7	0.001	N	N	$2.7x10^{-102}$
		comp10868_c0_seq1-protein lethal essential	-2.6	0.004	N	N	$2.2x10^{-67}$

	for life-like					
Transport						
	comp23657_c0_seq1-multidrug resistance protein homolog 49-like isoform x1	-5.3	$3x10^{-7}$	N	Y	0
	comp208052_c0_seq1-synaptic vesicle glycoprotein 2c-like	-4.4	0.003	N	Y	$1.6x10^{-41}$
	comp5506_c0_seq1-sodium potassium-transporting atpase subunit alpha-like	-3.5	0.006	N	Y	$1.1x10^{-71}$
	comp23992_c0_seq1-zinc transporter zip1-like isoform x2	-3.0	0.004	N	Y	$1.2x10^{-170}$
	comp11080_c0_seq1-sodium potassium-transporting atpase subunit alpha-like	-2.9	0.002	N	Y	$8.6x10^{-124}$
	comp58460_c0_seq1-synaptic vesicle glycoprotein 2c-like	-2.6	0.0005	N	Y	$2.5x10^{-143}$
	comp43548_c0_seq1-alpha-tocopherol transfer	-2.6	0.004	N	N	0
	comp22867_c0_seq1facilitated trehalose transporter tret1-like	-2.4	$7x10^{-5}$	Y	Y	0
	comp17079_c0_seq1-sodium-dependent multivitamin transporter	-2.4	0.009	N	Y	0
	comp60729_c0_seq1-synaptic vesicle glycoprotein 2b-like	-2.3	0.005	N	Y	0
	comp20369_c0_seq1-organic cation transporter	-2.2	0.0005	N	Y	0
	comp16055_c0_seq1-g-protein coupled receptor mth2-like	-2.2	0.004	Y	Y	$2.4x10^{-119}$
DNA , RNA, Protein Binding and folding						
	comp7858_c0_seq1-PREDICTED: uncharacterized protein LOC105283242	-10.8	$2x10^{-14}$	N	N	0
	comp16802_c0_seq1-odorant binding protein	-2.0	0.008	Y	N	$1.4x10^{-74}$
	comp22476_c0_seq1-connectin-like	-9.4	$1x10^{-7}$	Y	Y	0
	comp118703_c0_seq1-gtpase-activating protein	-2.9	0.0006	N	N	$6.7x10^{-64}$
	comp14847_c0_seq1-ets dna-binding protein pokkuri	-2.3	$2x10^{-5}$	N	N	0
	comp85798_c0_seq1-tata-binding protein-associated factor 172	-2.5	0.007	N	Y	$1.9x10^{-53}$
	comp17580_c0_seq1-protein atonal homolog 8	-2.4	0.0003	N	N	$1.6x10^{-28}$
	comp89516_c0_seq1-isoform b	-8.2	$4x10^{-5}$	N	N	$1.2x10^{-53}$
	comp3014_c0_seq1-heat shock protein 70	-4.8	0.002	N	N	$7.8x10^{-114}$
	comp38119_c0_seq1-heat shock protein 70	-4.5	0.0001	N	N	$5.2x10^{-50}$
	comp36096_c0_seq1-heat shock protein 70	-4.2	0.0005	N	N	0
	comp17731_c0_seq1-enhancer of split mgamma	-3.9	0.0001	N	N	$9.8x10^{-172}$
	comp21137_c0_seq1-dna-binding protein d-ets-4	-3.9	$3x10^{-09}$	N	N	0
	comp12549_c0_seq1-jerky protein homolog-like	-3.8	0.006	N	N	$3.3x10^{-27}$
	comp120_c0_seq2-reverse ribonuclease integrase	-3.7	0.0006	N	Y	$8.2x10^{-56}$

		comp10325_c0_seq1-probable ribonuclease zc3h12d isoform x1	-3.4	0.0005	N	N	$1.2x10^{-32}$
		comp23313_c0_seq1-arf-gap with dual ph domain-containing protein 1-like isoform x1	-2.9	$3x10^{-7}$	N	N	0
		comp74794_c0_seq1-serine threonine-protein kinase tricorner isoform x2	-2.7	0.002	N	N	8.9E-42
		comp19567_c0_seq2-zinc finger protein 629-like isoform x1	-2.7	0.008	N	Y	0
		comp26954_c0_seq1-apolipoprotein d-like	-2.6	$5x10^{-6}$	Y	Y	$3.0x10^{-152}$
		comp16641_c0_seq1-hypothetical protein KGM_08119	-2.4	0.0004	N	N	$6.6x10^{-8}$
		comp29764_c0_seq1-pdz domain-containing protein gipc3	-2.4	$1x10^{-4}$	N	N	0
		comp23668_c0_seq2-annexin b9-like isoform x2	-2.3	$4x10^{-5}$	N	N	0
		comp15131_c0_seq1-atp-binding cassette sub-family g member 1	-2.2	0.0004	N	Y	0
		comp22107_c0_seq1-hemicentin-1-like isoform x2	-2.2	0.002	Y	Y	0
		comp121957_c0_seq1-rna-directed dna polymerase from mobile element jockey-like	-2.0	0.002	N	N	$3.6x10^{-45}$
		comp22667_c0_seq1-adhesion-like transmembrane protein	-2.0	0.006	Y	N	$1.7x10^{-169}$
		comp21169_c0_seq1-differentially expressed in fdcp 6 homolog	-2.0	0.009	N	N	0
	Membrane components and receptors						
		comp255565_c0_seq1-integrin beta-ps-like isoform x2	-2.4	0.008	N	Y	$1.1x10^{-146}$
		comp214888_c0_seq1-PREDICTED: uncharacterized protein LOC106104156	-5.6	0.002	N	Y	$3.9x10^{-36}$
		comp10279_c0_seq1-23 kda integral membrane	-4.2	$3x10^{-10}$	N	Y	$1.3x10^{-48}$
		comp16892_c0_seq1-ninjurin a	-4.5	0.001	N	Y	$3.9x10^{-98}$
		comp20173_c0_seq2-solute carrier family 25 member 35-like isoform x1	-3.4	$1 x10^{-6}$	Y	N	$1.8x10^{-115}$
		comp28393_c0_seq1-elongation of very long chain fatty acids protein aael008004 isoform x1	-2.5	0.008	N	Y	0
		comp13400_c0_seq2-g-protein coupled receptor mth2-like	-2.5	0.002	N	Y	$2.1x10^{-118}$
		comp21994_c0_seq1-23 kda integral membrane	-2.4	$1 x10^{-5}$	N	Y	$4.1x10^{-90}$
		comp19624_c0_seq1-androgen-dependent tfpi-regulating	-2.1	0.005	N	Y	$6.4x10^{-90}$
		comp33781_c0_seq1-gonadotropin-releasing hormone ii receptor isoform x1	-2.0	0.004	N	Y	$4.7x10^{-101}$
		comp18852_c0_seq3-progestin and adipoq receptor family member 3	-2.0	0.0006	N	N	$2.2x10^{-9}$
	Other enzymes						
		comp23173_c0_seq1-4-coumarate-- ligase 1-like	-2.2	0.0007	N	N	0
		comp9581_c0_seq1inositol-trisphosphate 3-kinase a isoform x3	-4.4	$8x10^{-9}$	N	N	$3.2x10^{-28}$

		comp922_c0_seq1-aldehyde partial	-4.1	0.0003	N	N	0
		comp150987_c0_seq1-apyrase-like	-3.5	0.003	N	N	$1.7x10^{-97}$
		comp112114_c0_seq1-focal adhesion kinase 1 isoform x1	-3.3	0.005	N	N	$5.3x10^{-51}$
		comp9727_c0_seq1-uridine diphosphate glucosyltransferase	-2.7	0.0001	N	N	0
		comp13474_c0_seq1-pyrazinamidase nicotinamidase	-3.2	0.002	N	N	0
		comp25166_c0_seq1-allantoinase-like	-2.6	0.006	N	N	0
		comp16525_c0_seq1-inositol-trisphosphate 3-kinase a isoform x3	-2.6	$4.76x10^{-6}$	N	N	$4.8x10^{-167}$
		comp14384_c0_seq1-aldo-keto reductase akr2e4-like	-2.5	0.008328534	Y	Y	$1.7x10^{-147}$
		comp18252_c0_seq1-inositol oxygenase-like	-2.5	0.009	N	N	$3.2x10^{-83}$
		comp53604_c0_seq1-carbonyl reductase	-2.5	0.005	N	Y	$3.1x10^{-118}$
		comp19115_c0_seq1-probable methylcrotonoyl- carboxylase beta mitochondrial	-2.4	$7.8x10^{-5}$	N	N	0
		comp21295_c0_seq1-rho-associated protein	-2.4	$7x10^{-5}$	Y	N	$8.9x10^{-27}$
		comp18356_c0_seq1-probable cytochrome p450 9f2	-2.3	0.010	N	Y	0
		comp12628_c0_seq1-uncharacterized oxidoreductase -like	-2.1	0.006	N	N	0
		comp6221_c0_seq1-tyrosine-protein phosphatase non-receptor type 9-like	-2.3	0.0005	N	N	$1.1x10^{-21}$
		comp10812_c0_seq1-helicase polq-like	-2.5	$6x10^{-5}$	N	N	$1.3x10^{-118}$
		comp20502_c0_seq1-superoxide dismutase	-2.1	$6x10^{-5}$	N	N	$1.9x10^{-68}$
	Biosynthesis						
		comp14397_c0_seq1-2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase-like	-2.7	0.0003	N	N	$7.7x10^{-80}$
		comp6624_c0_seq1-probable pterin-4-alpha-carbinolamine dehydratase	-2.3	0.003	N	N	$7.5x10^{-73}$
		comp17850_c0_seq1-adenosine deaminase cecr1-like	-2.2	0.004	Y	N	0
24 h							
	DNA , RNA, Protein Binding and folding						
		comp537_c0_seq1-metallotransferrin	-2.4	0.007	N	N	$6.9x10^{-38}$

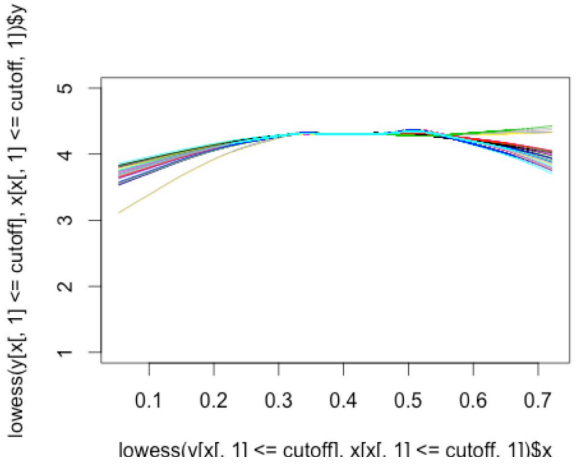
All bold sequences were found to be differentially expressed at both timepoints.

Table 3. ECB female bursa and bursal gland genes significantly differentially expressed after same- and cross-strain matings.

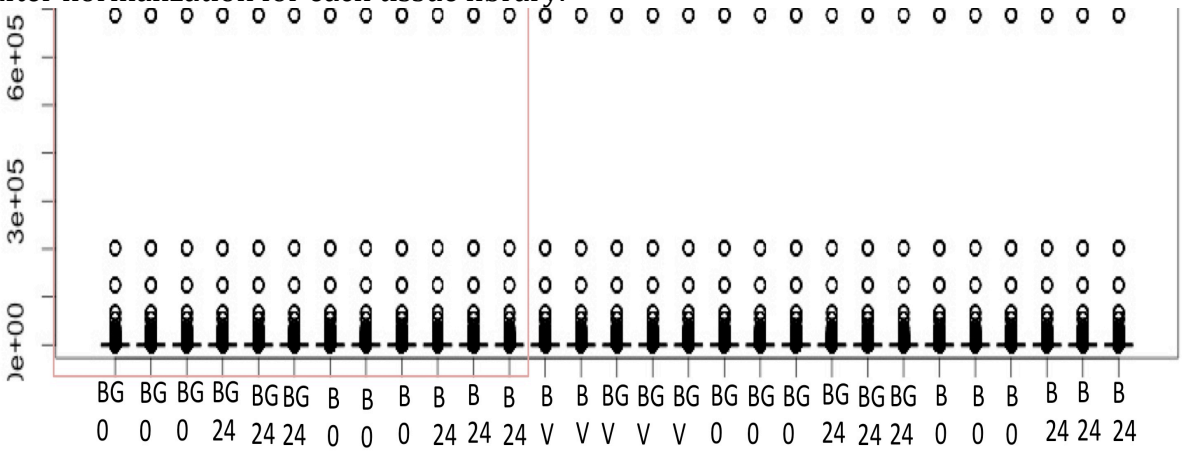
Comparison	Category	Sequence ID and description	Log_2FC	FDR	e-value
BG at 0 h	No DE genes				
BG at 24 h	DNA , RNA, Protein Binding and folding				
		comp537_c0_seq1-melanotransferrin	2.6	0.008	$3.5x10^{-39}$
		comp15304_c0_seq1-zinc finger protein basonuclin-2-like	3.3	0.009	$1.7x10^{-49}$
		comp34863_c0_seq1-tyrosine-protein phosphatase non-receptor type 13-like isoform x3	2.6	0.009	$6.3x10^{-142}$
		comp165526_c0_seq1- uncharacterized protein LOC106120954, partial	3.3	0.004	$1.8x10^{-74}$
		comp180772_c0_seq1-pre-mrna-processing factor 39	-2.6	0.009	$6.4x10^{-25}$
	Other Enzymes				
		comp3323_c0_seq1-5-aminolevulinate mitochondrial	-8.9	0.009	$4.0x10^{-45}$
		comp118661_c0_seq1-mitogen-activated protein kinase kinase kinase 15 isoform x1	4.5	0.001	$2.0x10^{-29}$
		comp125320_c0_seq1-tyrosine-protein kinase fps85d isoform x2	3.5	0.009	$1.0x10^{-71}$
	Metabolism				
		comp87585_c0_seq1-inorganic pyrophosphatase	-4.5	0.009	$7.1x10-^{33}$
	Novel				
		comp7507_c0_seq1	-11.6	0.001	
		comp110801_c0_seq1	-3.5	0.009	
B at 0 h	Novel				
		comp10103_c0_seq1	3.7	0.001	
		comp22744_c0_seq5	8.6	0.006	
		comp111094_c0_seq1	3.5	0.003	
B at 24 h	No DE genes				

Supplementary Material:

Supplementary figure 1. Lowess regression of normalized GC content in each library.



Supplementary Figure 2. Stratified boxplots of the log fold change of reads after normalization for each tissue library.



Supplementary Table 1: Significantly down-regulated genes in the ECB female bursa copulatrix at two post-mating time points

Time point	GO description	Sequence ID	Sequence description	Log2 Fold Change	FDR	e-value	SignalP	TMHMM
0 h	Binding							
		comp5447_c0_seq1	transcription factor ken 2 isoform x4	2.0	0.008678939	3.50E-97	N	N
		comp12027_c0_seq1	ankyrin-3-like isoform x1	2.1	0.003420911	0.00E+00	N	N
		comp9705_c0_seq1	tigger transposable element-derived protein 6-like	2.3	0.003599382	6.70E-20	N	N
		comp6536_c0_seq1	prefoldin subunit 6	2.3	0.008128154	1.80E-44	N	N
		comp10211_c0_seq1	tigger transposable element-derived protein 6-like	2.6	0.000432736	1.30E-19	N	N
		comp20132_c0_seq1	PREDICTED: uncharacterized protein K02A2.6-like	2.6	0.00030967	0.00E+00	N	N
		comp19190_c0_seq1	aael002168- partial	2.7	0.000295229	5.50E-27	N	N
		comp13674_c0_seq2	serine threonine-protein kinase genghis khan	2.8	0.006611813	2.60E-43	N	N
		comp3106_c0_seq1	rna-binding protein 26 isoform x1	3.0	0.003689713	1.20E-28	N	N
		comp2428_c0_seq1	protein son of sevenless	3.7	0.008745951	3.00E-113	N	N
		comp14205_c0_seq1	PREDICTED: uncharacterized protein LOC101745853	3.9	0.000188406	8.10E-116	N	N
		comp17263_c0_seq2	zinc finger bed domain-containing protein 4-like	4.6	0.001541139	2.00E-05	N	N
		comp24287_c0_seq1	rna-dependent rna polymerase	7.3	0.001080587	1.30E-25	N	N
	Structural and muscle							
		comp51704_c0_seq1	talin-1 isoform x3	2.1	0.006112858	1.70E-97	N	N
		comp17485_c0_seq1	myosin light chain smooth muscle	2.8	0.00396624	0.00E+00	N	N
		comp109406_c0_seq1	proline-rich extensin-like protein epr1	3.7	0.002071053	2.90E-38	N	N
	Cell functions and RNA/DNA replication and repair							
		comp23955_c0_seq1	ribosomal protein l39	2.6	0.001179454	3.80E-27	N	N
		comp25672	dna-directed rna	2.1	0.00787967	8.80E-78	N	N

		_c0_seq1	polymerase ii subunit rpb11		7			
		comp6941_c0_seq1	ribosome maturation protein sbds	2.8	0.000513219	9.90E-169	N	N
		comp17224_c0_seq1	dna primase small subunit	3.0	0.000108474	0.00E+00	N	N
		comp16630_c0_seq1	condensin complex subunit 2	3.3	0.007951458	6.00E-134	N	N
	Enzymes							
		comp6566_c0_seq1	diamine acetyltransferase 2	2.3	0.008433766	1.30E-27	N	N
		comp18094_c0_seq1	baculoviral iap repeat-containing protein 6	2.5	0.00232568	6.10E-158	N	N
		comp17317_c0_seq1	mitochondrial ribonuclease p protein 1 homolog	2.8	0.006463878	0.00E+00	N	N
		comp9059_c0_seq1	estrogen sulfotransferase	3.0	0.009176136	0.00E+00	N	N
		comp1751_c0_seq1	tyrosine-protein phosphatase 69d isoform x1	3.6	0.003172787	6.60E-63	N	N
	Development							
		comp13586_c0_seq1	cullin-2	2.7	0.006481985	1.30E-50	N	N
	Signaling							
		comp107130_c0_seq1	adenomatous polyposis coli isoform x1	3	0.007230301	2.50E-19	N	N
	transport							
		comp82206_c0_seq1	inositol - trisphosphate receptor isoform x1	3.3	0.000241257	4.70E-94	N	Y
		comp537_c0_seq1	melanotransferrin	6.2	8.17E-08	1.60E-38	N	N
	Peptidase							
		comp18522_c0_seq1	26s proteasome complex subunit dss1	4.7	1.38E-06	3.60E-23	N	N
24 h								
	Membrane							
		comp5225_c0_seq1	probable g-protein coupled receptor 139	2.1	0.002947275	0.00E+00	N	Y
		comp275198_c0_seq1	tyramine receptor	3.5	0.001442608	0.00E+00	N	N
		comp161932_c0_seq1	cadherin isoform g	4.0	0.002793534	9.90E-133	N	N
		comp261860_c0_seq1	tyramine receptor	4.1	0.000291198	6.10E-53	N	Y
	Binding							
		comp23313_c0_seq1	arf-gap with dual ph domain-containing protein 1-like	2.1	0.000400994	0	N	N
		comp232_c0_seq1	protein isoform b-like	2.8	0.00493618	0	N	N

		comp23528_c0_seq1	fibronectin type iii domain-containing protein 5-like	2.8	0.007229497	5.70E-179	Y	Y
		comp15567_c0_seq1	iglon family member 5 isoform x3	3.5	0.000743479	4.40E-129	N	N
		comp293533_c0_seq1	similar to CG7991	3.5	0.00036998	3.70E-45	N	N
		comp7316_c0_seq1	kv channel-interacting protein 1	4.1	5.94E-07	3.30E-116	N	N
		comp458244_c0_seq1	transposon ty3-g gag-pol polyprotein isoform x2	4.6	0.00591815	3.50E-108	N	N
		comp263851_c0_seq1	kinesin-like protein kif21a	4.7	0.001202091	8.10E-48	N	N
		comp294862_c0_seq1	retroelement polyprotein	6.2	0.002060244	1.60E-105	N	N
	Enzymes							
		comp332038_c0_seq1	e3 ubiquitin-protein ligase ubr3 isoform x1	3.5	0.008467276	5.50E-42	N	
		comp15760_c0_seq1	protein kinase dc2	3.8	6.51E-05	0.00E+00	N	
		comp4533_c0_seq1	ceramide glucosyltransferase	2.2	0.004354428	1.40E-44	N	Y
		comp22445_c1_seq1	cystathionine beta-synthase	2.3	4.73E-05	0.00E+00	N	N
		comp7869_c0_seq1	ceramide glucosyltransferase	3.9	0.000158535	6.60E-150	N	Y
		comp467777_c0_seq1	cytochrome p450	4.0	0.001900369	2.20E-86	N	N
		comp162789_c0_seq1	rieske-domain protein neverland	6.3	0.003281384	3.00E-165	N	N
		comp12539_c0_seq1	sulfotransferase 1c4	12.3	0.000414817	1.30E-93	N	N
	Metabolism							
		comp22879_c0_seq1	probable 4-coumarate-- ligase 2	2.3	0.001600182	0	N	N
		comp168035_c0_seq1	udp-glycosyltransferase ugt46a1	2.6	0.000408862	8.00E-148	Y	Y
		comp21155_c0_seq1	sterol carrier protein 2 3-oxoacyl- thiolase	2.8	0.002451445	1.10E-172	N	N
		comp212864_c0_seq1	4-coumarate-- ligase 1-like	2.9	0.004557422	8.10E-174	N	N
		comp165543_c0_seq1	phospholipase a2-like	3.9	0.002781379	5.70E-74	Y	Y
		comp174247_c0_seq1	15-hydroxyprostaglandin dehydrogenase	8.8	0.000109	1.60E-66	N	N
	Peptidase							
		comp13270_c0_seq1	seminal fluid protein hacp057	3.1	0.003032672	7.70E-17	Y	Y
		comp22975_c0_seq1	cytosolic carboxypeptidase 1-like isoform x4	2.6	0.000418585	0	N	N

		comp182519_c0_seq1	serine protease nudel	3.3	0.009899643	0.00E+00	N	Y
	Cell functions and RNA/DNA replication and repair							
		comp21137_c0_seq1	dna-binding protein d-ets-4	2.7	8.46E-05	0	N	N
		comp281424_c0_seq1	probable nuclear hormone receptor hr38	5.6	3.30E-07	4.10E-158	N	N
		comp469287_c0_seq1	nuclear receptor subfamily 4 group a member 2 isoform x2	6.0	1.46E-08	2.50E-130	N	N
	Transport							
		comp647173_c0_seq1	glutamate-gated chloride channel isoform x11	2.8	0.003787758	6.40E-170	N	N
		comp16972_c0_seq1	facilitated trehalose transporter tret1-like	3.0	5.63E-05	2.00E-133	N	Y
		comp22902_c0_seq1	open rectifier potassium channel protein 1 isoform x1	3.3	0.000282507	0.00E+00	N	Y
		comp6354_c0_seq1	facilitated trehalose transporter tret1-like	3.6	0.000229346	1.70E-179	N	Y
		comp116077_c0_seq1	synaptic vesicle glycoprotein 2b-like	4.5	0.000783354	2.40E-149	N	Y
		comp208052_c0_seq1	synaptic vesicle glycoprotein 2b-like	5.8	1.94E-06	1.60E-41	N	Y
		comp492743_c0_seq1	synaptic vesicle glycoprotein 2b-like	8.8	3.09E-07	5.70E-45	N	N
	Signaling and receptors							
		comp14401_c0_seq1	socs2-12 protein	2.8	6.38E-07	6.10E-11	N	N
		comp21518_c0_seq1	semaphorin-2a-like isoform x1	3.2	8.06E-06	7.60E-96	N	N
		comp18067_c0_seq1	semaphorin-2a isoform x1	3.5	2.81E-06	0.00E+00	N	N
		comp220380_c0_seq1	semaphorin-2a	6.3	5.03E-07	1.70E-52	N	N
		comp317193_c0_seq1	bone morphogenetic protein receptor type-1b	3.1	0.000278163	7.00E-32	N	N
	Structural and muscle							
		comp271144_c0_seq1	kelch-like protein diablo	2.8	0.008446331	1.40E-50	N	N
		comp142076_c0_seq1	myosin heavy muscle isoform x17	3.0	0.004750705	3.80E-14	N	N
	Defense							
		comp182822_c0_seq1	toll receptor 18 wheeler	3.4	0.003378262	3.80E-32	N	Y

Supplementary Table 2: Significantly down-regulated genes in the bursal gland at two post-mating time points

Time point	GO description	Sequence ID	Sequence description	Log2 Fold Change	FDR	e-value	SignalP	TMHMM
0 h	RNA/DNA replication and repair							
		comp199416_c0_seq1	dna helicase	2.0	0.009804205	0	N	N
		comp15013_c0_seq2	histone-lysine n-methyltransferase trithorax	2.1	0.006006689	0	N	N
		comp213635_c0_seq1	dna-directed rna polymerase ii subunit rpb1-like	3.8	0.000151936	7.10E-35	N	N
		comp4341_c0_seq1	reverse partial	3.9	0.002143476	0.00E+00	N	Y
		comp208218_c0_seq1	forkhead box protein n3-like isoform x2	4.3	0.00084458	9.30E-35	N	Y
		comp207914_c0_seq1	protein gooseberry isoform x2	13.2	4.70E-05	0.00E+00	N	N
	Binding							
		comp222258_c0_seq2	lim homeobox protein lhx1 isoform x3	2.0	0.003189173	1.40E-12	N	N
		comp186778_c0_seq1	rhophilin-2 isoform x1	2.1	5.66E-05	8.10E-86	N	N
		comp194225_c0_seq1	leucine-rich repeat-containing protein 40-like	2.2	0.000516118	0	N	N
		comp128449_c0_seq1	transposable element p transposase isoform x3	2.3	0.003337599	7.10E-46	Y	N
		comp237221_c0_seq1	PREDICTED: uncharacterized protein LOC106137125	2.3	0.000928762	0.00E+00	N	N
		comp172993_c0_seq1	protein disabled	2.5	0.008908321	1.50E-122	N	N
		comp337081_c0_seq1	thap domain-containing protein	2.7	0.009476307	3.00E-04	N	N
		comp136249_c0_seq1	transcription factor ken	2.7	0.001636202	2.10E-19	N	N
		comp219911_c0_seq1	zinc finger cchc domain-containing protein 8 homolog	2.7	0.00367063	6.20E-36	N	N
		comp169224_c0_seq1	protein numb isoform x2	2.8	0.006515253	1.10E-75	N	N

		comp459_c0_seq1	PREDICTED: uncharacterized protein LOC105842437	3.0	0.008037794	1.20E-74	N	N
		comp18307_c0_seq1	PREDICTED: uncharacterized protein LOC106136491	3.0	0.005731557	0.00E+00	N	N
		comp16844_c0_seq1	gastrula zinc finger	3.2	0.002516787	8.60E-50	Y	N
		comp2151_c0_seq1	homeotic protein empty spiracles-like	3.4	0.004399474	1.80E-71	N	N
		comp179570_c0_seq1	rna-directed dna polymerase from transposon x- partial	4.1	0.000742147	2.10E-145	N	N
		comp12091_c0_seq1	kruppel-homolog 1	4.8	0.000203501	4.50E-18	N	N
		comp234232_c0_seq1	protein 5nuc-like	5.1	0.006771171	0.00E+00	N	N
		comp1324_c0_seq1	PREDICTED: uncharacterized protein LOC106129409	6.1	0.007261962	0.00E+00	N	Y
		comp92033_c0_seq1	calcium and integrin-binding family member 2	9.5	0.009630271	4.80E-112	N	N
		comp180373_c0_seq1	PREDICTED: uncharacterized protein LOC103573902	9.7	0.006666175	8.40E-144	N	N
	Metabolism							
		comp22208_c0_seq1	2-hydroxyacylsphingosine 1-beta-galactosyltransferase	2.1	0.004486316	0	Y	Y
		comp14463_c0_seq1	retinol dehydrogenase 11-like	2.2	0.002874685	1.50E-95	N	Y
		comp14821_c0_seq1	probable chitinase 2	11.5	0.001053279	0.00E+00	Y	N
	Membrane channels and components							
		comp1911_c0_seq1	innexin inx1	2.3	0.004899436	0	N	Y
		comp23077_c0_seq1	23 kda integral membrane	2.8	0.008671473	2.90E-73	N	Y
		comp4041_c0_seq1	neural cell adhesion molecule 2-like isoform x2	5.3	0.006429716	7.90E-36	N	N
		comp325233_c0_seq1	serine-rich adhesin for platelets-like	8.7	0.003081658	5.90E-29	N	N

		comp13933_c0_seq1	elongation of very long chain fatty acids protein 1-like	9.2	0.00014993	7.90E-102	N	Y
		comp17180_c0_seq1	cd63 antigen-like	11.4	0.001794798	5.10E-84	Y	Y
		comp21596_c0_seq4	vascular endothelial growth factor b-like isoform x2	13.9	0.000151936	2.70E-14	Y	N
		comp21722_c0_seq3	vascular endothelial growth factor a-like isoform x1	15.5	3.27E-05	6.60E-15	N	Y
	Enzymes							
		comp12891_c0_seq1	phospholipase c beta 1	2.5	0.003337599	0	N	N
		comp3728_c0_seq1	phospholipase c beta 1	2.8	0.005594116	2.20E-55	N	N
		comp20870_c0_seq1	arylalkylamine n-acetyltransferase	3.2	0.001070447	8.20E-126	N	N
		comp23410_c0_seq1	peptidylglycine alpha-hydroxylating monooxygenase	2.5	0.007062375	0	Y	N
		comp22835_c0_seq1	gmp reductase 1-like	2.6	0.007970183	0	Y	Y
		comp624_c0_seq1	slit homolog 3	3.2	7.34E-05	0	Y	N
		comp23301_c0_seq1	lower subunit	3.4	0.000158348	8.1W-116	Y	N
		comp16388_c0_seq1	protein jagged-1	3.9	0.00890702	2.80E-77	N	N
		comp6540_c0_seq1	ferritin subunit	4.0	0.002463782	5.00E-35	N	N
		comp13450_c0_seq1	ferritin subunit	4.1	0.000288481	1.50E-32	N	N
		comp231739_c0_seq1	pancreatic triacylglycerol lipase-like	5.4	0.000372312	3.20E-69	Y	N
		comp370704_c0_seq1	indole-3-acetaldehyde oxidase-like	8.4	0.000255025	0.00E+00	N	N
		comp10945_c0_seq1	cytochrome p450 4v2-like	11.2	0.001003372	1.70E-145	N	N
	Transmembrane transport							
		comp82206_c0_seq1	inositol - trisphosphate receptor isoform x1	2.7	0.000571755	4.70E-94	N	Y
		comp19696_c0_seq1	atp-binding cassette sub-family a member 3-like	3.6	0.001636757	7.00E-106	N	Y
		comp17409_c0_seq1	atp-binding cassette sub-family a member 3-like	3.6	0.002043022	6.10E-162	N	Y
		comp135933_c0_seq1	voltage-dependent l-type calcium channel subunit beta-2 isoform x4	3.7	0.000155241	2.20E-128	N	N

		comp174455_c0_seq1	atp-binding cassette sub-family a member 3-like	4.3	0.000392316	3.90E-84	N	N
	Structural and muscle related							
		comp184560_c0_seq1	disheveled-associated activator of morphogenesis 1 isoform x2	2.9	0.002893856	8.40E-168	N	N
	Immunity related proteins							
		comp13038_c0_seq1	18 wheeler partial	2.9	0.001488593	0	N	Y
		comp389758_c0_seq1	protein toll	4.2	0.000142893	6.70E-143	N	N
		comp34113_c0_seq1	PREDICTED: uncharacterized protein LOC105395759	17.5	1.12E-05	3.90E-08	Y	N
	Peptidases							
		comp20161_c0_seq1	reverse ribonuclease integrase	3.1	0.00444919	0	N	N
		comp12952_c0_seq1	serine protease easter-like	5.1	0.00056473	4.10E-145	N	N
		comp13178_c0_seq1	seminal fluid protein hacp010	6.8	0.001388874	0.00E+00	Y	N
		comp11118_c0_seq1	chymotrypsin-like elastase family member 2a	7.6	0.000570796	0.00E+00	N	N
		comp187511_c0_seq1	seminal fluid protein hacp038	7.6	9.28E-07	0.00E+00	Y	N
		comp12494_c0_seq1	seminal fluid protein hacp001	12.1	4.76E-06	1.80E-161	Y	N
	Peptidase regulators							
		comp13766_c0_seq2	bcp inhibitor	4.6	0.001197109	3.60E-12	N	Y
		comp7457_c0_seq1	bcp inhibitor	14.7	3.57E-07	4.00E-16	Y	Y
		comp17954_c0_seq1	PREDICTED: uncharacterized protein LOC106124191	7.3	0.000192238	5.80E-13	Y	N
		comp37318_c0_seq1	amyloid beta a4 protein	8.9	7.53E-06	1.10E-11	N	N
		comp158200_c0_seq1	kazal-type serine protease inhibitor domain-containing protein	9.2	9.43E-07	3.80E-07	Y	Y
		comp19292_c0_seq1	kazal-type proteinase inhibitor	11.0	9.44E-06	2.70E-10	Y	N
		comp167709_c0_seq	antichymotrypsin-2-like isoform x1	10.3	0.000204817	3.50E-141	Y	N

		1						
		Odorant binding protein and extracellular						
		comp8780_c0_seq1	general odorant-binding protein 56a-like	6.1	2.43E-05	1.20E-67	Y	N
		comp151397_c0_seq1	venom allergen 5-like	7.7	0.000366614	6.00E-78	N	Y
		comp14135_c0_seq1	antitrypsin isoform 2	10.0	0.001199438	2.60E-132	Y	N
24 h								
	Hormone							
		comp6293_c0_seq1	juvenile hormone acid methyltransferase	4.3	7.49E-05	1.80E-114	N	N
	RNA/DNA replication and repair							
		comp9415_c0_seq1	clip-associating protein	2.7	0.009845233	3.00E-54	N	N
		comp17731_c0_seq1	enhancer of split mgamma	4.4	0.000143308	9.70E-172	N	N
		comp19708_c0_seq1	protein lozenge	2.3	0.00684954	0.00E+00	N	N
	Membrane channels and components							
		comp16055_c0_seq1	g-protein coupled receptor mth2-like	2.5	0.007675502	1.50E-112	Y	Y
	Binding							
		comp23313_c0_seq1	arf-gap with dual ph domain-containing protein 1-like	2.1	0.003223413	0.00E+00	N	N
		comp29764_c0_seq1	pdz domain-containing protein gipc3	2.1	0.007362388	0	N	N

Chapter 5: Experimental manipulation of mating system alters reproductive gene expression in *Tribolium casteneum* beetles

Abstract

Sexual selection, widely recognized as a powerful force driving the evolution of both male and female reproductive traits, continues after mating through interactions between male ejaculates and the female reproductive tract. Males transfer complex ejaculates containing seminal fluid proteins that interact with female-secreted reproductive proteins to mediate postcopulatory sexual interactions. Using an experimental evolution approach, we manipulated mating systems of *Tribolium castaneum* flour beetles to gain insight into how variation in sexual selection intensity alters the expression and sequence divergence of reproductive genes in both sexes. Following 12 generations of enforced monogamy *versus* polygamy we used RNAseq to: 1) elucidate the reproductive transcriptome of both sexes, and 2) compare male and female gene expression between mating system lines. Comparisons between selection lines with contrasting mating systems revealed significant alterations in reproductive gene expression in both sexes. This experimental evolution study provides new insight into how variation in postcopulatory sexual selection intensity can alter the expression and sequence divergence of reproductive genes. In addition, this study further elucidates the reproductive transcriptome of *Tribolium* flour beetles, an important stored product pest.

Introduction:

Sexual selection is widely recognized as a powerful evolutionary force driving the evolution of both male and female reproductive traits (Darwin, 1859; Andersson, 1994). Contrary to Darwin's original conception, considerable evidence now indicates that sexual interactions often continue long past mating (reviewed by Simmons, 2001; Peretti & Aisenberg, 2015). Relative to precopulatory interactions, postcopulatory sexual selection is considerably more cryptic, as it encompasses complex molecular interactions between females and multiple male ejaculates that transpire within the hidden confines of the female reproductive tract (Simmons, 2001; Peretti & Aisenberg, 2015; Sirot et al., 2009; Findlay et al., 2014).

Such postcopulatory interactions have been postulated to be a major selective agent driving the elaboration of male ejaculates (Arnqvist & Nilsson 1999; Lewis & South, 2012). In diverse organisms ranging from humans to insects, males transfer ejaculates that contain not only sperm, but also seminal fluid proteins (SFPs) (Poiani, 2006; Perry et al., 2013). Produced by male reproductive tissues, SFPs in *Drosophila melanogaster* fruit flies have been shown to influence female behavior and physiology, including reducing the likelihood of re-mating, inducing ovulation and oviposition, promoting sperm storage, and inhibiting feeding (Wolfner, 2009; Heifetz & Wolfner, 2004; Ravi Ram & Wolfner, 2007; Wolfner, 2002). Importantly, although female reproductive proteins (FRPs) may also play important roles in postcopulatory sexual selection, such female proteins have received very little attention (Ah-King et al., 2014). However, FRPs have been characterized in a few insects, including *Apis* honeybees (Baer et al., 2009a), *Drosophila* fruit flies (McGraw et al., 2004b; Prokupek et al., 2008), *Pieris* butterflies (Meslin et al., 2015), and *Ostrinia* moths (Al-Wathiqui et al., 2014). Furthermore, both male and female reproductive proteins have been shown to evolve

rapidly in response to sexual selection and sexual conflict (Ellegren & Parsch, 2007; Swanson et al., 2001; Wilburn & Swanson, 2015; Swanson & Vacquier, 2002; Prokupek et al., 2010; Swanson, 2004).

Animal mating systems are a key factor determining the strength of postcopulatory sexual selection (Thornhill & Alcock, 2013; Choe & Crespi, 1997). Under monogamous mating systems, where females mate with a single male, postcopulatory sexual selection on reproductive traits is relaxed. Polygamous mating systems, where both sexes have multiple partners, intensifies postcopulatory sexual selection as it not only heightens male-male competition for access to fertilizations (Reuter et al., 2008; Fedorka et al., 2011; Simmons, 2001), but also allows females to exert cryptic (i.e. postcopulatory) female choice (Peretti & Aisenberg, 2015). However, very little research has addressed the crucial question concerning how these different mating systems might alter reproductive gene expression in both sexes (but see Gerrard et al., 2013; Immonen et al., 2014).

In this study, we applied experimental evolution to examine how enforced mongamy *versus* polygamy alters reproductive gene expression in both sexes of the flour beetle *Tribolium castaneum*. Experimental evolution studies, which monitor trait evolution after manipulating specific experimental conditions (Kawecki et al., 2012; Garland & Rose, 2009), provide a powerful tool to examine the consequences of altering sexual selection intensity (Arnqvist & Rowe, 2005; Ritchie, 2007; Garland & Rose, 2009). *T. castaneum* beetles have been used as an important model for sexual selection research over several decades (reviewed by (Fedina & Lewis, 2008)). In addition to detailed descriptions of *T. castaneum* mating behavior, reproductive anatomy and physiology (Al-Khalifa, 1981; Gadzama & Happ, 1974; Surtees, 1961; Bloch Qazi et al., 1996; Fedina & Lewis, 2004; Fedina, 2007), the internal processes of spermatophore

transfer, sperm release, and sperm storage within the female reproductive tract have also been well-studied (Fedina & Lewis, 2006; Bloch Qazi et al., 1996; Fedina, 2007). Additional work in *T. castaneum* has documented male-male sperm competition and cryptic female choice (Bloch Qazi et al., 1996; Fedina & Lewis, 2004; Fedina, 2007; Fedina & Lewis, 2006). Recent work has also identified SFPs manufactured by *T. castaneum* male accessory glands (South et al., 2011a; Xu et al., 2013; Parthasarathy et al., 2009). Finally, two previous studies in *T. castaneum* have successfully used experimental evolution to manipulate mating systems (Bernasconi & Keller, 2001, Michalczyk et al., 2010). Michalczyk et al. (2010) found that after 20 generations of male-biased sex ratios (1:6 ratio of F:M), *T. castaneum* males showed differences in their copulatory behaviors compared to males from female-biased lines (9:1 ratio of F:M). (Michalczyk et al., 2010b). Bernasconi & Keller (2001) demonstrated that even after their mother had experienced a single generations of polyandry, *T. castaneum* males showed higher second-male paternity success (P_2) compared to monandrous males. However, no studies *of Tribolium* beetles have used experimental evolution to determine how different mating systems might alter expression in male and female reproductive genes.

Because animal mating systems determine the strength of postcopulatory sexual selection, contrasting mating systems will generate distinct selection pressures that are predicted to alter gene expression and drive adaptive population divergence in reproductive traits. *Tribolium castaneum* is a highly polygamous stored products pest, with both sexes mating repeatedly over a long adult lifespan (Park, 1933; Sokoloff, 1974; Fedina & Lewis, 2008). Thus, we used RNAseq of *T. castaneum* males and females from polygamous lines to comprehensively characterize the reproductive transcriptome. We then compared reproductive gene expression in both sexes under experimentally enforced monogamy vs. polygamy. We predicted that relaxed postcopulatory sexual selection under monogamy would result in down-regulated expression of specific male genes

involved in sperm competition or inhibition of female remating. We also predicted that monogamous females would show down-regulation of genes involved in postcopulatory (cryptic) choice.

Methods:

Mating system selection lines

Beetles used in the selection lines were obtained from a Tufts University colony of wild-type beetles (originally derived from the Berkeley synthetic strain), and maintained following previously described methods (Lewis et al., 2004) in a dark incubator at 29°C and 70% RH.

Experimental evolution was conducted over 12 generations in three replicate lines each for two mating conditions: 1) *Monogamy* - a single adult female was paired with a single adult male in 2 g flour (15 mating pairs), or 2) *Polygamy* -15 adult females and 15 adult males were combined in 30 g flour to mimic a natural *T. castaneum* mating system. All adults were at least 2 wks post-eclosion to ensure sexual maturity, and identical beetle density was maintained across all lines. After a 10 day mating period, adults were removed and the flour containing eggs laid by all 15 females for each replicate line was placed in a new container with 270 g fresh flour. After 25 days, pupae were separated (approximately 1000 pupae per replicate line), and ~50 pupae were haphazardly selected and sexed from each replicate line. These were housed individually in 0.5 mL centrifuge tubes and monitored periodically until eclosion.

Tissue collection, RNA extraction and sample preparation

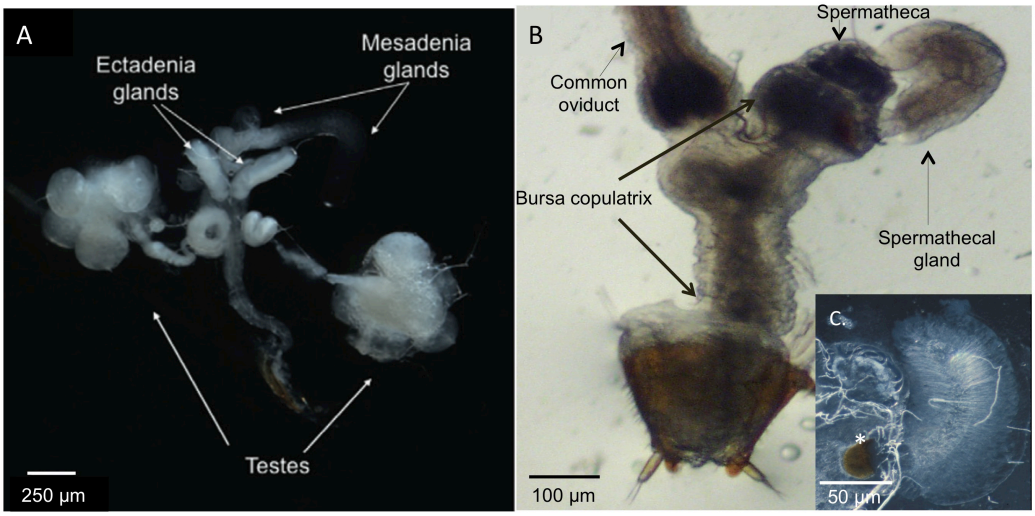


Figure 1. The male and female reproductive tracts of *T. castaneum* flour beetles. (A)The *T. castaneum* male reproductive system consists of two testes connected to the ejaculatory duct via the vas deferens and seminal vesicles. Also connected to the ejaculatory duct are two pairs of accessory glands: the ectadenia glands and the mesadenia glands (previously published in South et al., 2010). (B) In the female, the male spermatophore is formed within the bursa copulatrix. Shortly after mating, sperm are released from the spermatophore and stored initially in the anterior bursa (top right portion of bursa copulatrix), then in the spermatheca. (C)The female spermathecal gland is a conspicuous, feather-like structure attached to the anterior bursa; it contains a chitinized o-ring (*) near its base of the structure (pictured below *).

To examine the effect of mating system experimental manipulation on reproductive gene expression, we dissected reproductive accessory glands from both male and female reproductive tract. In *T.castaneum,* males possess two pairs of accessory glands: the ectadenia (EC) and the mesadenia glands (MES) (Figure 1A). Together with the testes, these glands contribute to the production of a spermatophore, a sperm-containing package (Bloch Qazi et al., 1996; Novaczewski & Grimnes, 1996; Sevener et al., 1992). We examined gene expression in the male EC and MES glands, in both cases compared to male thorax (Supplementary Figure 1).

The male spermatophore is transferred to the female bursa copulatrix during mating and then everts to release sperm into the female reproductive tract (Bloch Qazi et al., 1996; Fedina, 2007). Sperm are initially stored in the anterior portion of the bursa, but are soon transferred to the spermatheca for long-term storage (Figure 1B) (Bloch Qazi et al., 1996). Attached to the anterior bursa, near the spermatheca duct is a conspicuous, feather-like structure called the spermathecal gland (Figure 1C). Because the function of this gland is unknown, we also examined gene expression in the spermathecal gland relative to female thorax (Supplementary Figure 1).

After 12 generations of selection, reproductive tissue was dissected from 10 adults randomly chosen from across all replicate lines; tissue from these 10 individuals was pooled to form a single biological replicate, and three biological replicates were obtained for each male and female tissue type. To obtain the tissues of interest, beetles were cooled for 5 minutes in a -20°C freezer, then dissected under 100X magnification in RNAlater. The tissues dissected from males included the MES, EC, thorax, and the testes including the vas deferens (Figure 1A). From females, we dissected the thorax and the spermathecal gland, including the chitinized ring (Figure 1B & 1C). Tissue samples were stored at -80°C in RNAlater (n=36).

RNA was extracted by removing the RNAlater, freezing pooled tissue samples in liquid nitrogen, then homogenizing tissues by hand inside a 1.5mL centrifuge tube. Qiagen RNAeasy Tissue and Lipid kit was used to extract the homogenized tissue, and RNA quantity and quality was assessed using a Nanodrop 2000. An Illumina Truseq RNA sample preparation kit (v2) was used to prepare the samples for RNA sequencing following methods in Al-Wathiqui et al. (2014). Briefly, cDNA libraries were created from each sample, adapters were attached for multiplexing, and libraries were amplified. The resulting 36 samples were pooled at equal concentrations at the Tufts University

Core Facility and sequenced on three lanes of the sample flow cell on an Illumina HiSeq 2500, High Output v4 (single-end, 50 bp reads). In the polygamous lines, one of the three replicates of male testes tissue failed to sequence, so only two biological replicates were included in subsequent analyses for this tissue. Library quality was assessed using FastQC (Babraham Bioinformatics), which showed very little adapter contamination (>5%), so no further quality control was implemented.

Characterizing the reproductive transcriptome

To characterize the reproductive transcriptome of *T. castaneum* males and females, we first looked for genes that were up-regulated in reproductive tissues compared to non-reproductive thorax tissue. To identify genes that were up-regulated in reproductive tissue compared to thorax, we first mapped our RNA sequences to the *Tribolium* genome obtained from BeetleBase (Wang et al., 2007; Kim et al., 2009) using Bowtie 2 with very sensitive settings (Langmead & Salzberg, 2012; Langmead et al., 2009). Using HTSEq-count we then counted the number of reads mapped to each gene in the *Tribolium* genome (Anders et al., 2015). These read counts were normalized and then visualized in the Bioconductor package edgeR using a biological coefficient of variation (BCV) to described the variation in true abundance of each gene between biological replicates (Robinson et al., 2009). edgeR was then used to identify differentially expressed sequences between tissue types based on normalized read counts using a generalized linear model approach with no interaction term was used (McCarthy et al., 2012; Robinson & Oshlack, 2010; Robinson et al., 2009; Robinson & Smyth, 2007). Genes were considered significantly up-regulated if they had a Log_2 fold change (Log_2FC) ≥ 2 and a false discovery rate (FDR) ≤ 0.01. After differentially expressed genes were identified, we used both annotations and protein structural information

obtained from EnsemblMetazoa Biomart to characterize differentially expressed genes (Kinsella et al., 2011).

To identify putative SFPs and FRPs, we used two criteria: 1) up-regulation in reproductive tissue compared to thorax and 2) presence of a secretion signal. We then used the program SignalP to identify genes containing secretion signals, indicating secretion of a protein from the cells where it is produced.

Changes in gene expression between selection lines

We also identified male and female genes that were significantly differentially expressed in specific reproductive tissues between monogamous and polygamous lines (Supplementary Figure 1). Again, edgeR with a general linear model approach was used, with no interaction term specified. Genes were considered significantly differentially expressed between mating system lines if they showed a $Log_2FC \geq 2$ and a FDR of ≤ 0.01.

Results:
Sequencing statistics

We successfully obtained 35 sequenced tissue libraries. Of these libraries, library size ranged from 474 Mbp to 1,709 Mbp. The mean PHRED quality score for all libraries exceeded 35 (less than 1/3200 chance a base was called incorrectly). On average 74% of reads from each library aligned to the *T. castaneum* genome. BCV analysis indicated that the majority of biological replicates were clustered together by tissue type (Supplementary Figure 2).

Gene expression in male and female reproductive tissues

We identified 577 genes that were up-regulated in the MES glands of *T. castaneum* males compared to the male thorax (LogFC≥2); of these, 66% could be annotated with functional categories (Figure 2A). In the EC glands, 615 genes were up-regulated compared to male thorax and 65% were annotated (Figure 2B). Both tissues showed similar distributions of functional categories, including peptidases, peptidase regulators, transmembrane transport proteins, and odorant binding proteins (Figure 2).

In the spermathecal gland of *T. castaneum* females, there were 502 genes that were significantly up-regulated compared to female thorax. Of these, 75% were annotated with the majority belonging to functional categories of general cellular processes, transmembrane transport, and peptidase activity (Figure 4C).

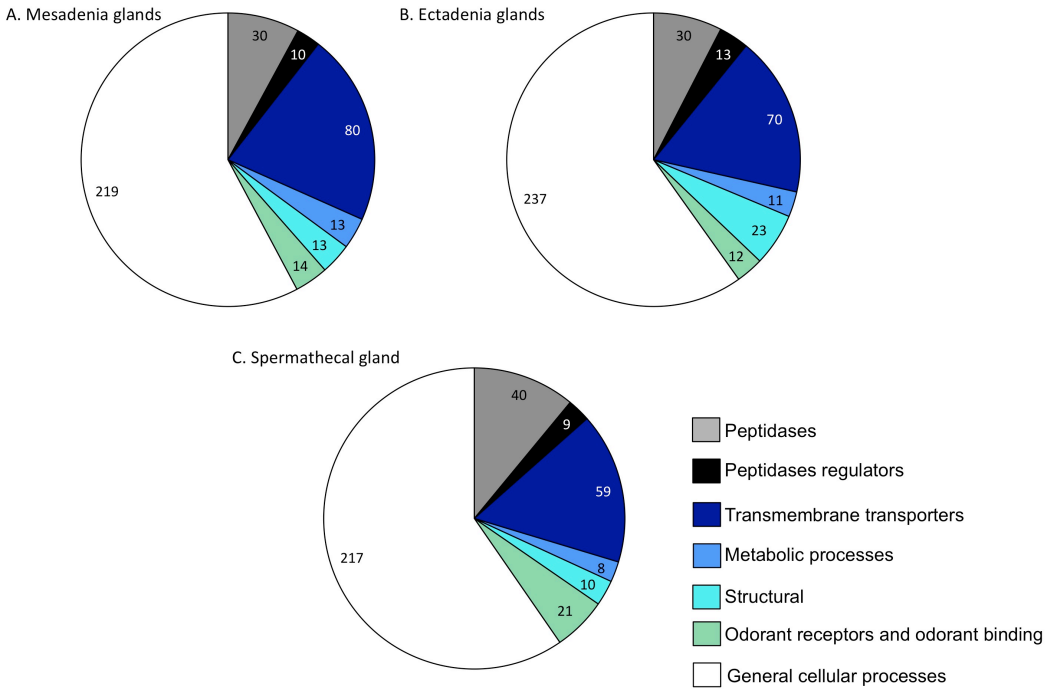


Figure 2. Distribution of gene ontology categories for genes up-regulated in the mesadenia (A), ectadenia (B), and spermathecal glands (C) compared to thorax in *T. castaneum beetles*.

Identification of putative male SFPs

In the MES glands of *T. castaneum* males, we identified 71 genes that were up-regulated compared to male thorax and also contained predicted secretion signals (Table 1), indicating that they are likely secreted and transferred to females in the male ejaculate. Eighteen were novel genes (no similarity to known proteins), while the remaining 53 genes were annotated as peptidases, ion transport proteins, immune response proteins, and other categories (Table 1). Two serine-type peptidases were identified (TC002061 and TC008930; Table 1). Two MES genes (TC016368 and TC000520) were annotated as the protein spaezle (Table 1), an activating ligand in the Toll pathway, which is important for development and immunity in *D. melanogaster* (Valanne et al., 2011). We also identified a gene (TC07737) which was annotated as attacin c, an anti-bacterial protein (Carlsson et al., 1998).

In the EC glands of *T. castaneum* males, we identified 88 genes up-regulated compared to male thorax and that also contained predicted secretion signals (Table 1). Of these, 76 were annotated as peptidases, peptidase regulators, binding proteins, and other functional categories (Table 1). Genes expressed in male EC glands included seven serine peptidases, four peptidases of unknown class, one serine peptidase regulator (serpin) and a peptidase regulator of unknown class (Table 1). Five of the serine proteases identified (TC001023, TC007019, TC008505, TC000545, and TC000550) were further characterized as trypsins, serine peptidases in the chymotrypsin family. We also identified four genes up-regulated in the EC glands (TC003109, TC007724, TC008768, and TC015304) that encoded insect cuticle proteins (Table 1), which interact with chitin to form the insect cuticle (Andersen et al., 1995). Another gene up-regulated in the male EC glands (TC005093) encoded the protein coleoptericin (Table 1), an antibacterial protein that inhibits bacterial cell division (Sagisaka et al., 2001).

Identification of putative FRPs

Of the genes up-regulated in the spermathecal gland of *T.castaneum* females, 119 contained predicted secretion signals, including 22 novel genes (Table 2). The remaining 97 genes encoded peptidases, odorant binding proteins, structural proteins, as well as other categories of proteins (Table 2). Of the 19 peptidases and eight peptidase regulators identified, 11 encoded serine peptidases; eight of which were further characterized as trypsins (Table 2). Another gene up-regulated in the female spermathecal gland (TC009364) encoded the cysteine protease cathepsin L (Table 2). Cathepsin L is a lysosomal endopeptidase that interacts with structural proteins, such as collagen and fibronectin, and which has been shown to be expressed in mammalian testes (Turk et al., 2012), as well as the *T. castaneum* larval gut (Denell et al., 2008). We also identified three genes (TC002092, TC005905, TC00887) encoding metallopeptidases, which use metal ions in their hydrolysis mechanisms (Polgar, 1989) and five genes encoding serpins (Table 2), which inhibit the action of serine peptidases (Silverman et al., 2001). The female spermathecal gland also showed up-regulation of six odorant binding proteins predicted to be secreted (Table 2). Odorant binding proteins bind pheromones and odor molecules and deliver them to odor-sensing nerves (Leal, 2013). Finally, several genes encoding structural proteins, including seven insect cuticle proteins, and Knickkopf, a protein responsible for chitin organization (Chaudhari et al., 2011) were also up-regulated in the female spermathecal gland (Table 2).

Differential gene expression between mating system lines

In the MES accessory glands of *T. castaneum* males, we found 34 genes that were significantly differentially expressed between experimental monagmous and polygamous lines. These included three trypsin-like serine-peptidases, one of which was up-regulated in monogamous compared to polygamous males (TC010906-RA) while the remaing two were down-regulated (TC004937 and TC000495) (Figure 3; Supplementary

Table 1). Two sequences encoding cytochrome P450s (TC002542 and TC015030) also showed up-regulation in the MES glands of monandrous males (Figure 3). Cytochrome P450s belong to a protein superfamily and function as terminal oxidase enzymes in electron transport chains (Scott & Wen, 2001; Scott et al., 1998). One gene (TC015030) encodes cytochrome P450 307B1, which is involved in production of ecdysteroid hormones important in insect reproduction (Hentze et al., 2013).

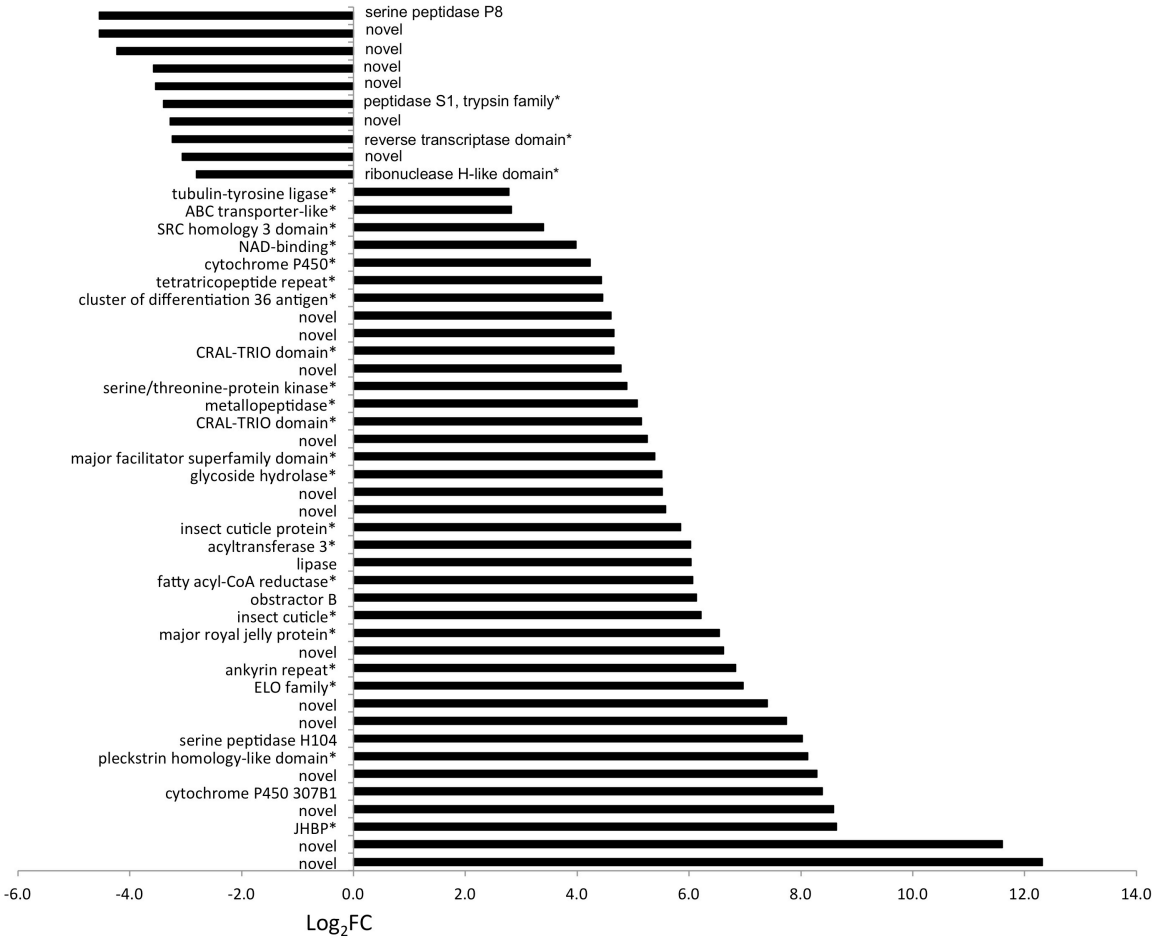


Figure 3. Expression level and annotation terms of genes differentially expressed between monandrous and polyandrous *T.casteneum* lines in male mesadenia glands. * denotes genes that are annotated with protein structural information only.

Another gene (TC001183) up-regulated in the MES glands of monandrous males encodes a juvenile hormone binding protein (Figure 3). Juvenile hormone binding proteins (JHBP) function to transport juvenile hormone, an insect hormone important for

development and reproduction, through the hemolymph to targeted areas, protecting it from degradation (Goodman et al., 1990).

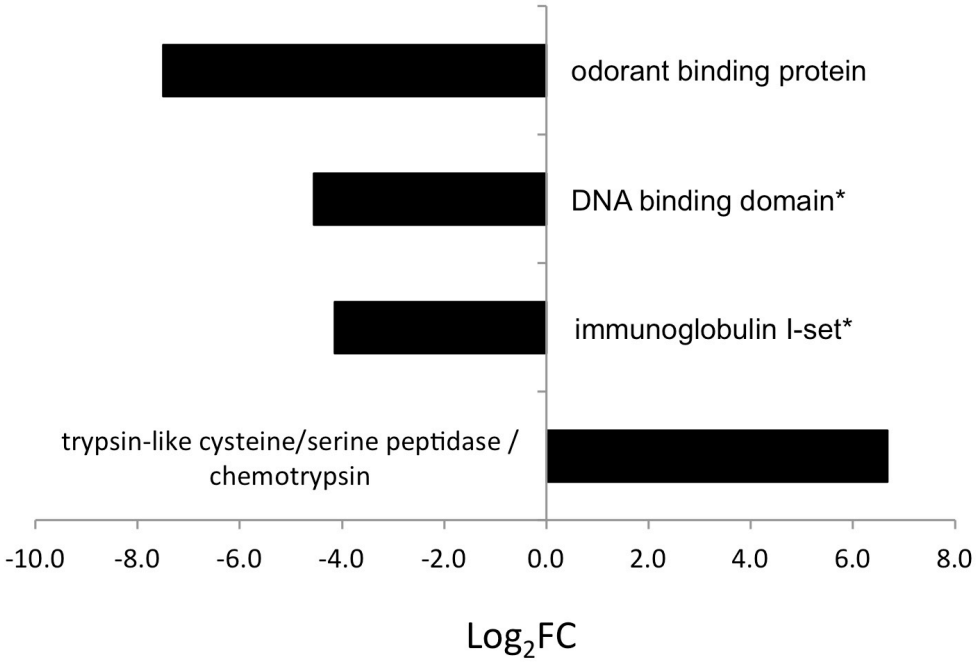


Figure 4. Expression level and annotation terms of genes differentially expressed between monagamous and polyandrous *T. casteneum* lines in male testes. * denotes genes that are annotated with protein structural information only.

We identified no genes in the EC glands that were significantly differentially expressed between selection lines. In male testes, however, we found four sequences significantly differentially expressed between mating system lines. Three of these were down-regulated in monogamous compared to polygamous males, including one (TC030341) that encodes an odorant receptor (odorant receptor 136; Figure 4). Odorant receptors are membrane-bound G-coupled protein receptors that detect odor molecules (Wicher et al., 2008). The remaining genes down-regulated in testes of monogamous males encode an unknown protein containing an immunoglobulin I-set domain (TC000863) and an unknown protein containing a DNA binding domain (TC000424).

Finally, we identified one sequence (TC030072) that was up-regulated in testes of males from monogamous lines, and which encoded a trypsin serine peptidase (Figure 4).

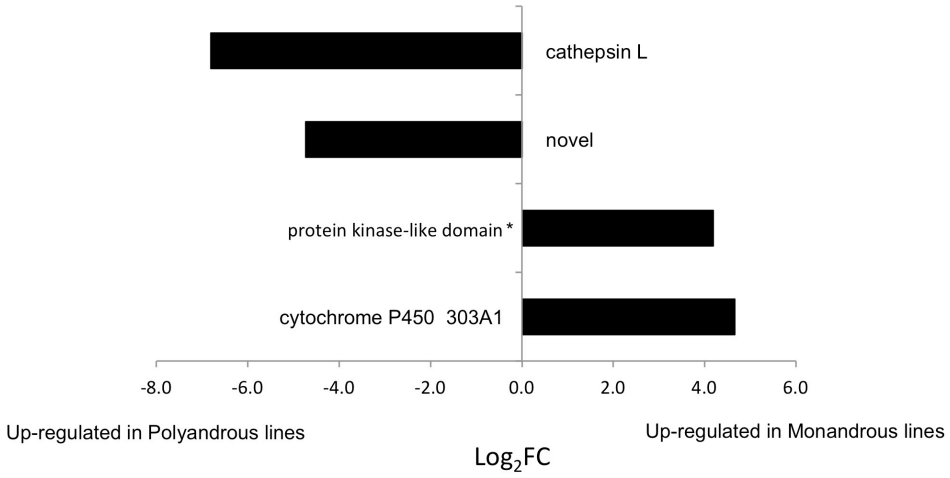


Figure 5. Expression level and annotation terms of genes differentially expressed between monagamous and polyandrous *T. casteneum* lines in female spermathecal glands. * denotes genes that are annotated with protein structural information only

In *T. castaneum* females, we identified one spermathecal gland sequence (TC009364) that was down-regulated in monogamous lines and annotated as cathepsin L (Figure 5). As described above, cathepsin L has been identified as a digestive protease in *T. castaneum* larval gut (Denell et al., 2008) and is also expressed in mammalian testes (Turk et al., 2012). Two sequences were up-regulated in the spermathecal glands of monogamous females: TC002542 encodes a cytochrome P450 and TC008101 encodes an uncharacterized protein with a protein kinase domain.

Discussion:

Using an experimental evolution approach through manipulating mating systems, this study provides new insight into how variation in sexual selection intensity can alter the expression of reproductive genes in both sexes. In addition, this study further elucidates the reproductive transcriptomes of *Tribolium* flour beetles, an important stored

product pest. Following 12 generations of enforced monogamy *verus* polygamy, we documented changes in reproductive gene expression and sequence divergence. Compared to males from polygamous lines, those from monogamous lines showed altered expression of reproductive genes encoding several trypsin-like serine peptidases, cytochrome P450s, a juvenile hormone binding protein, and an odorant receptor. Females from monogamous lines also showed altered reproductive gene expression, including a digestive protease, cathepsin L, and a cytochrome P450. While additional experiments will be necessary to pinpoint the functional roles played by these gene products in *T. castaneum* reproduction, these results represent a substantive contribution to our understanding of postcopulatory sexual interactions.

Altered reproductive gene expression in response to mating system manipulation

In *T. castaneum* males, three trypsin-like serine peptidases were differentially expressed between monandrous and polygamous lines in male mesadenia glands (Figure 3; Supplementary Table 1); one sequence was up-regulated in monandrous males (TC010906) and two were down-regulated (TC004937 and TC000495). In other insects, reproductive peptidases belonging to the trypsin family of serine peptidases have been shown to play key roles in numerous postcopulatory processes (reviewed in (Laflamme & Wolfner, 2013). For example, in both *D. melanogaster* and crickets in the *Allonemobius* complex, trypsin-like serine peptidases transferred by males are necessary to fully induce egg-laying by females (Laflamme et al., 2012; Marshall et al., 2009). In *D. melanogaster*, male-produced trypsins are also required to release sperm from female storage organs (Bonilla et al., 2015). Furthermore, trypsins have been shown to be important for sperm activation in both *Manduca sexta* (Friedlander et al., 2012) and *Bombxy mori* (Nagaoka et al., 2012). Because these *T. castaneum* sequences did not show orthology to specific

trypsin peptidases with known functions, further experiments will be needed to elucidate their functions in *T. castaneum* reproduction.

Our study also revealed significant down-regulation of odorant receptor 136 (TC030341) in the testes of monandrous males (Figure 4; Supplementary Table 1). In human sperm cells, the odorant receptor hOR17-4 binds chemical cues found in the female reproductive tract and mediates an increase in sperm swimming speed (Sun et al., 2012; Spehr et al., 2006). In insects, odorant receptors are also important for sperm motility (Leopold, 1976). For example in *A. gambiae* mosquitoes one odorant recptor, AgOrco, is localized to sperm flagella and activates flagellar movement when its ligands are present (Pitts et al., 2014). Based on the expression pattern we observed in *T. castaneum* selection lines, odorant receptor 136 may play a role in augmenting male sperm competitive ability when rival male sperm is present.

In *T. castaneum* males, the largest change in gene expression observed between selection lines occurred in juvenile hormone binding protein (JHBP; TC001183), which was strongly up-regulated ($Log_2FC=8.6$) in the mesadenia glands of monandrous males (Figure 3; Supplementary Table 1). Juvenile hormone (JH) controls a myriad of processes in insects, including female uptake of vitellogenin (egg yolk protein) into oocytes (Hartfelder, 2000). JHBP transports JH through the hemolymph to target sites (Goodman et al., 1990). In male *D. melanogaster* fruit flies, knockdown experiments have shown that JHBP plays a role in SFP production and in maintaining male mating behavior (Wilson et al., 2003; Shemshedini et al., 1990). *T. castaneum* males showed reduced accessory gland size, and lower sperm and SFP transfer to females after knock-down of juvenile hormone acid O-methyltransferase, a component of the JH biosynthesis pathway (Parthasarathy et al., 2009). However, when JHBP was knocked-down in *T. castaneum* males subsequently mated to females, female egg production remained unchanged (Xu et

al., 2013). Thus, although JHBP is strongly up-regulated in monogamous, its role in *T. castaneum* reproduction remains unclear.

Three encoding cytochrome P450s genes (TC009412, TC002542, and TC015030) were also up-regulated in the mesadenia glands of monandrous males (Figure 3; Supplementary Table 1). Cytochrome P450s are involved in the biosynthesis pathways for both JH and ecdysteroid hormones, both of which play a role in female egg production (Feyereisen, 1999; Scott & Wen, 2001). In *A. aegypti* mosquitoes, male ejaculate contains the ecdysteroid, 20-hydroxyecdysone, which interacts with a receptor in the female reproductive tract (MISO) to control egg maturation (Baldini et al., 2013). Ecdysteroids also mediate sperm production and release from male testes in *Spodoptera litura* moths (Seth et al., 2004), as well as male fertility in *Argyrotaenia* and *Choristoneura* leafroller moths (Sun et al., 2000). In monandrous *T. castaneum* males, we found up-regulation of cytochrome 307B1 (TC15030), which is predicted to mediate ecdysteroid biosynthesis in the mesadenia glands (Hentze et al., 2013). Hence, it is plausible that up-regulated cytochrome P450s in monandrous lines might increase male ecdysteroid production, thus increasing male sperm production and/or female egg maturation.

In *T. castaneum* male ectadenia glands, we found no differentially expressed genes between monandrous and polygamous lines. In the related species *Tenebrio molitor*, similar male glands secrete structural components of the spermatophore wall (Happ, 1992). Thus, in *T. castaneum* male ectadenia glands may serve a similar conserved function essential for male reproduction.

In *T. castaneum* females, we identified three spermathecal gland genes (TC009364, TC009412, and TC009252) that were differentially expressed between mating system selection lines (Figure 4). One gene (TC009364), encodes a cathepsin L, which acts as a digestive protease in the *T. castaneum* larval gut (Denell et al., 2008), and

was up-regulated (Log_2FC= -6.8; Supplementary Table 1) in females from polygamous lines. Up-regulation of cathepsin L suggests that it may play a role in cryptic female choice in polygamous *T. castaneum* females.

Conclusion:

Because animal mating systems determine the strength of postcopulatory sexual selection, contrasting mating systems should generate distinct selection pressures that are predicted to alter reproductive gene expression and drive adaptive sequence divergence. In an experimental evolution approach, we manipulated the mating system of *Tribolium castaneum*, a highly polygamous stored products pest. After 12 generations of enforced monogamy vs. polygamy, we used RNAseq to compare expression levels and sequence divergence in reproductive genes of both sexes. Compared to polygamous lines, males from monogamous lines showed altered expression of reproductive genes encoding several trypsin-like serine peptidases, cytochrome P450s, a juvenile hormone binding protein, and an odorant receptor. Females from monogamous lines also showed altered reproductive gene expression, including a digestive protease, cathepsin L, and a cytochrome P450. This study further characterized *T. castaneum* reproductive proteins, and found several peptidases and peptidase regulators secreted by both sexes and are likely to mediate postcopulatory processes. Finally, this study suggests that the conspicuous spermathecal gland present in *T. castaneum* females could play an important role in postcopulatory sexual selection, as it secretes odorant binding proteins, proteases, and protease regulators.

Table 1. Functional classes of up-regulated genes unique to and predicted to be secreted from two distinct male accessory glands in *T.castaneum*.

Tissue	Gene function category and Transcript ID	Gene description	Log_2 FC	FDR
Mesadenia Glands				
	Peptidases			
	TC002061	trypsin-like serine peptidase P32	3.3	$6.9x10^{-11}$
	TC008930	trypsin-like serine peptidase H88	5.7	$1.3x10^{-10}$
	TC014727	peptidase S9*	2.4	$5.4x10^{-6}$
	TC007101	protease associated domain*	2.1	$1.7x10^{-5}$
	TC005403	peptidase M28*	2.0	$1.2x10^{-5}$
	Peptidase regulators			
	TC005754	serpin 22	3.8	$9.1x10^{-4}$
	TC010868	peptidase inhibitor*	2.4	$4.8x10^{-5}$
	TC009375	endopeptidase inhibitor*	2.3	$3.8x10^{-8}$
	Hydrolases			
	TC000981	acetylcholinesterase	4.3	$9.4x10^{-11}$
	TC008420	acid phosphatase*	6.4	$1.1x10^{-5}$
	TC001186	NOTUM*	3.3	$1.3x10^{-5}$
	Transport			
	TC012624	GOLD domain*	3.6	$6.9x10^{-14}$
	TC006393	ligand-gated ion channel subunit*	3.5	$1.2x10^{-6}$
	TC010147	ion transport N-terminal domain*	5.0	$3.0x10^{-8}$
	TC005324	calcineurin-like phosphoesterase domain, apaH type*	5.0	$1.4x10^{-15}$
	TC015563	calycin*	3.4	$2.6x10^{-5}$
	TC000596	CAP domain*	3.5	$1.5x10^{-3}$
	TC000779	DnaJ domain*	2.5	$1.6x10^{-6}$
	TC004117	signal sequence receptor, delta*	2.0	$7.6x10^{-5}$
	Structural proteins			
	TC013808	insect cuticle protein*	3.3	$4.7x10^{-3}$
	TC015304	insect cuticle protein*	2.5	$8.3x10^{-4}$
	TC003574	EF-hand domain*	2.9	$8.0x10^{-9}$
	TC014175	EF-hand domain*	2.3	$1.0x10^{-7}$
	TC010099	immunoglobulin-like fold*	2.4	$5.3x10^{-4}$
	TC014162	immunoglobulin I-set*	2.3	$7.2x10^{-4}$
	Other enzymes			
	TC011151	acetyltransferase*	9.3	$6.6x10^{-7}$
	TC002612	choline dehydrogenase*	9.6	$6.3x10^{-12}$

	TC003715	Alkaline phosphatase-like *	2.9	$4.3x10^{-7}$
	TC000223	Protein Kinase C Substrate 80K-H	2.4	$6.2x10^{-9}$
	TC001643	thioredoxin-like fold*	2.3	$1.2x10^{-12}$
	TC004723	thioredoxin-like fold*	2.1	$1.5x10^{-6}$
	Cell signaling			
	TC014638	cystine-knot cytokine*	7.4	$5.0x10^{-14}$
	Oxidation- reduction and metal binding			
	TC007331	gamma interferon inducible lysosomal thiol reductase GILT*	3.0	$1.0x10^{-3}$
	TC013504	G-coupled protein receptor 3C*	2.5	$1.1x10^{-6}$
	TC001556	peroxidasin	2.4	$7.7x10^{-6}$
	TC011919	transferrin Fe BS*	2.0	$2.5x10^{-5}$
	Metabolic proteins			
	TC001770	glycoside hydrolase superfamily*	3.5	$2.1x10^{-3}$
	TC012734	glycoside hydrolase*	2.5	$5.6x10^{-3}$
	TC009725	lipase*	2.0	$6.6x10^{-4}$
	Protein and hormone binding , hormone production			
	TC001553	galactose-binding domain-like*	3.4	$9.3x10^{-5}$
	TC003182	CUB domain*	2.0	$2.8x10^{-4}$
	Immune response			
	TC000520	spaetzle	3.0	$4.3x10^{-9}$
	TC016368	spaetzle	6.2	$4.4x10^{-15}$
	TC007737	attacin c*	3.2	$1.4x10^{-3}$
	DNA binding, cell signaling, cell adhesion and cell division			
	TC006116	windbeutel	2.9	$1.2x10^{-11}$
	TC003986	torpedo*	2.7	$1.7x10^{-10}$
	TC010947	c-type lectin fold*	2.6	$3.1x10^{-7}$
	TC002063	concanavalin A-like lectin*	2.1	$6.9x10^{-3}$
	TC030769	homeodomain-like*	2.3	$9.9x10^{-9}$
	TC000278	fibroblast growth factor 8	2.1	$7.0x10^{-4}$
	TC008118	sterile alpha motif/pointed domain*	2.0	$7.9x10^{-10}$
	Unknown			
	TC011691	LD repeat*	4.3	$9.1x10^{-8}$
	TC003571	domain of unknown function protein 4779*	2.8	$6.8x10^{-6}$
Ectadenia Glands				
	Peptidases			
	TC000545	trypsin-like serine peptidase P11	5.2	$5.7x10^{-4}$
	TC000550	trypsin-like serine peptidase P15	6.0	$6.3x10^{-4}$
	TC001023	trypsin-like serine peptidase P22	5.4	$2.5x10^{-4}$
	TC011078	trypsin-like serine peptidase 126	3.2	$7.7x10^{-3}$

	TC009320	peptidase*	15.0	$1.6x10^{-10}$
	TC007019	serine peptidase p77	8.5	$3.8x10^{-11}$
	TC008505	serine peptidase h81	13.8	$1.8x10^{-6}$
	TC013173	trypsin-like serine peptidase *	14.4	$1.2x10^{-15}$
	TC013613	serine peptidase 144	3.5	$5.2x10^{-3}$
	TC002061	serine peptidase 32	2.2	$2.6x10^{-5}$
	TC011052	metallopeptidase m2*	2.0	$2.1x10^{-4}$
	Peptidase regulators			
	TC011162	peptidase inhibitor*	5.9	$2.0x10^{-8}$
	TC013310	serpin 28	2.4	$6.7x10^{-4}$
	Hydrolases			
	TC014100	chitin deacetylase 1	2.8	$2.5x10^{-4}$
	TC006732	carbesterase B*	10.8	$1.6x10^{-9}$
	Transport			
	TC002289	sodium/hydrogen exchanger	3.0	$2.4x10^{-12}$
	TC004373	protein disulfide oxidoreductase *	3.6	$1.1x10^{-14}$
	TC007269	amino acid transporter domain*	10.9	$6.4x10^{-6}$
	TC010500	transporter*	7.6	$1.4x10^{-12}$
	TC012232	sugar transporter*	3.5	$1.2x10^{-5}$
	TC013423	sugar transporter*	7.2	$1.5x10^{-10}$
	TC008334	neurotransmitter-gated ion-channel*	3.0	$3.3x10^{-3}$
	TC008266	GABA-gated ion channel	2.6	$2.1x10^{-3}$
	TC000779	DnaJ domain*	2.3	$1.0x10^{-5}$
	TC011911	triose-phosphate transporter domain*	2.1	$2.4x10^{-8}$
	Structural			
	TC013810	transmembrane protein 98*	6.6	$8.6x10^{-7}$
	TC003109	insect cuticle protein*	5.8	$1.2x10^{-5}$
	TC007724	insect cuticle protein*	5.0	$2.2x10^{-3}$
	TC008768	insect cuticle protein*	3.4	$7.3x10^{-3}$
	TC015304	insect cuticle protein*	2.1	$4.6x10^{-3}$
	TC009893	cuticular protein analogous to peritrophins 1-F	2.2	$8.7x10^{-4}$
	TC011661	tetraspanin*	6.5	$6.2x10^{-8}$
	TC002660	leucine rich protein*	15.7	$8.3x10^{-8}$
	TC000210	leucine-rich repeat-containing N-terminal*	2.3	$1.8x10^{-3}$
	TC006065	seven cysteines, N-terminal*	2.0	$1.0x10^{-7}$
	Oxidation- reduction and metal binding			
	TC005874	oxidoreductase*	2.4	$1.7x10^{-4}$
	TC011321	G-coupled protein receptor, rhodopsin-like, 7TM*	2.8	$4.0x10^{-7}$

	TC007259	G-coupled protein coupled receptor*	2.2	$4.9x10^{-3}$
	TC013504	G-coupled coupled receptor*	2.0	$7.0x10^{-5}$
	TC004797	zinc finger*	2.6	$2.2x10^{-5}$
	TC000175	haem peroxidase*	2.2	$7.2x10^{-5}$
	Metabolic proteins			
	TC008954	gylcosyltransferase*	3.2	$8.7x10^{-8}$
	TC006371	glycoside hydrolase 13*	4.2	$1.1x10^{-6}$
	TC009553	glycoside hydrolase 13*	6.4	$5.7x10^{-18}$
	TC005411	glycoside hydrolase, family 2*	3.4	$6.1x10^{-4}$
	TC013552	lipopolysaccharide-modifying protein*	2.2	$2.3x10^{-5}$
	TC014957	low-density lipoprotein (LDL) receptor class A*	2.1	$2.1x10^{-7}$
	TC015632	GPI mannosyltransferase*	2.1	$5.4x10^{-6}$
	Protein and hormone binding, hormone production			
	TC015295	Protein binding domain*	2.8	$3.2x10^{-4}$
	TC008678	chemosensory protein 14	5.3	$5.2x10^{-9}$
	TC011601	eicosanoid/glutathione metabolism protein*	2.5	$1.3x10^{-4}$
	TC015525	juvenile hormone binding protein*	6.0	$1.6x10^{-4}$
	TC005670	EMI domain*	3.7	$7.1x10^{-3}$
	TC001553	galactose-binding domain-like	2.7	$2.3x10^{-3}$
	TC007906	phosphoinositide-binding clathrin adaptor*	2.2	$7.7x10^{-6}$
	TC009862	CUB domain*	2.2	$8.2x10^{-3}$
	TC015030	cytochrome P450 307B1	10.8	$5.2x10^{-8}$
	Immune response			
	TC005093	coleoptericin*	5.2	$5.2x10^{-7}$
	TC005096	NAD binding protein*	5.3	$1.7x10^{-7}$
	TC000802	protein yellow	2.7	$3.7x10^{-3}$
	TC016368	spaetzle	2.6	$2.5x10^{-4}$
	DNA binding, cell, signaling, cell division, cell signaling			
	TC007774	abnormal spindle associated protein*	5.8	$1.5x10^{-4}$
	TC007129	homeobox domain*	4.4	$2.4x10^{-5}$
	TC006733	homeobox domain*	8.8	$3.8x10^{-16}$
	TC013059	homeobox domain*	5.3	$1.7x10^{-5}$
	TC007448	synaptotagmin*	2.4	$2.1x10^{-6}$
	TC009200	protein kinase domain*	5.6	$1.9x10^{-4}$
	TC004117	translocon-associated protein*	2.0	$9.3x10^{-5}$
	TC000163	zona pellucida domain*	4.6	$6.0x10^{-9}$
	TC007326	cystine-knot_cytokine *	4.7	$2.5x10^{-5}$
	TC006640	cystine-knot_cytokine *	7.5	$3.0x10^{-5}$

	Unknown			
	TC007307	domain of unknown function 4789	3.9	$5.2x10^{-5}$
	TC007252	Protein of unknown function DUF4728	6.6	$1.1x10^{-7}$
	TC006381	Protein of unknown function DUF4786	3.1	$1.0x10^{-3}$
	TC011827	Protein of unknown function DUF1676	2.8	$3.7x10^{-3}$
	TC003571	Protein of unknown function DUF4779	2.5	$5.9x10^{-5}$

* proteins annotated by protein functional domains only

Table 2. Functional classes of up-regulated genes unique to and predicted to be secreted from spermatheca in *T.castaneum*.

Transcript ID	Gene description	Log_2FC	FDR
Peptidases			
TC008930	trypsin-like serine peptidase h88	14.5	$7.0x10^{-44}$
TC015344	trypsin-like serine peptidase h158	14.0	$4.9x10^{-35}$
TC000887	peptidase S10, serine carboxypeptidase, active site*	12.5	$1.1x10^{-42}$
TC030065	trypsin-like serine peptidase p26	12.2	$2.8x10^{-21}$
TC008931	trypsin-like serine peptidase h89	11.8	$2.1x10^{-10}$
TC006424	trypsin-like serine peptidase p74	7.3	$2.3x10^{-9}$
TC015465	peptidase M2, peptidyl-dipeptidase A*	7.0	$6.6x10^{-8}$
TC013276	trypsin-like serine peptidase p135	6.7	$7.1x10^{-7}$
TC030073	trypsin-like serine peptidase p98	6.5	$5.0x10^{-4}$
TC004557	aminopeptidase N-like protein	5.2	$6.0x10^{-21}$
TC002092	ance-3	3.3	$5.7x10^{-8}$
TC012390	trypsin-like serine ppeptidase H129	2.8	$9.4x10^{-9}$
TC014727	peptidase S9B, dipeptidylpeptidase IV N-terminal*	2.5	$2.1x10^{-6}$
TC009364	cathepsin L	4.4	$7.5x10^{-7}$
TC005905	carboxypeptidase A	3.1	$8.7x10^{-6}$
TC006303	peptidase S28*	2.1	$8.5x10^{-3}$
TC030072	serine protease p97	2.7	$6.7x10^{-5}$
TC013415	serine protease P141	2.8	$7.1x10^{-4}$
TC001259	aminopeptidase N-like protein	7.0	$5.4x10^{-3}$
Peptidase regulators			
TC005741	serpin 43Ab*	10.8	$6.4x10^{-9}$
TC013310	serpin 28	3.5	$2.7x10^{-6}$
TC030077	serpin 5	3.3	$2.0x10^{-11}$
TC014237	serpin 30	2.7	$9.6x10^{-8}$
TC030076	serpin 4	2.7	$5.7x10^{-5}$
TC004553	proteinase inhibitor, propeptide*	2.7	$3.0x10^{-7}$
TC014664	Alpha-2-macroglobulin*	2.1	$4.0x10^{-3}$
TC007946	trypsin inhibitor-like*	2.2	$6.8x10^{-3}$
Other enzymes			
TC000602	cysteine-rich secretory protein, allergen V5*	14.5	$2.0x10^{-10}$
TC003941	thioredoxin-like fold*	14.3	$4.3x10^{-30}$
TC000596	ves allergen*	10.4	$2.7x10^{-13}$
TC015370	glycosyl hydrolase family 9 protein	2.1	$2.2x10^{-5}$
TC013591	short-chain dehydrogenase*	2.6	$2.4x10^{-9}$

Metabolic proteins			
TC000449	saposin-like*	3.6	$2.5x10^{-18}$
TC000336	GDSL lipase*	4.1	$1.0x10^{-9}$
TC007263	partial AB-hydrolase lipase domain*	3.2	$3.5x10^{-4}$
TC003715	sulfatase*	3.3	$1.4x10^{-8}$
TC001169	chitin binding domain*	5.1	$7.2x10^{-10}$
TC001350	chitin binding domain*	3.5	$1.9x10^{-4}$
TC002612	glucose-methanol-choline oxidoreductase*	6.3	$7.3x10^{-7}$
Protein binding			
TC008118	stromal interaction molecule*	3.5	$1.4x10^{-24}$
TC000863	immunoglobulin-like fold*	5.7	$1.4x10^{-14}$
TC004567	calcium-binding domain*	2.4	$1.4x10^{-8}$
DNA binding, cell signaling, cell adhesion and cell division			
TC015735	zinc finger*	9.8	$1.8x10^{-11}$
TC007448	cystine-knot cytokine*	3.4	$2.1x10^{-10}$
TC014638	cystine-knot cytokine*	3.7	$3.4x10^{-5}$
TC003986	torpedo*	2.6	$1.9x10^{-9}$
TC001363	hedgehog, N-terminal signalling domain*	5.8	$3.9x10^{-8}$
TC001549	coagulation factor 5/8 C-terminal type domain*	4.8	$2.7x10^{-16}$
TC012560	single domain Von Willebrand factor type C domain*	9.0	$7.0x10^{-11}$
TC012558	single domain Von Willebrand factor type C domain*	8.7	$8.3x10^{-5}$
TC010419	C-type lectin-like*	4.5	$2.0x10^{-11}$
TC016314	C-type lectin fold*	2.2	$1.5x10^{-6}$
TC014781	concanavalin A-like lectin*	2.7	$8.6x10^{-4}$
TC013016	cysteine-rich Golgi apparatus protein 1 repeat*	2.0	$2.2x10^{-6}$
TC006116	windbeutel	2.1	$2.6x10^{-6}$
TC005324	5'-Nucleotidase	2.1	$6.7x10^{-3}$
TC000798	5'-Nucleotidase/apyrase*	2.8	$5.1x10^{-5}$
TC011068	zona pellucida domain	2.6	$8.0x10^{-5}$
Structural proteins			
TC003109	insect cuticle protein*	10.5	$2.6x10^{-9}$
TC001115	insect cuticle protein*	6.8	$6.9x10^{-13}$
TC013818	insect cuticle protein*	6.0	$9.7x10^{-4}$
TC014685	insect cuticle protein*	5.1	$5.0x10^{-6}$
TC003832	insect cuticle protein*	3.7	$5.6x10^{-5}$
TC010893	insect cuticle protein*	3.1	$5.2x10^{-12}$
TC011142	insect cuticle protein*	2.5	$3.7x10^{-3}$

TC010653	knickkopf	2.9	$1.9x10^{-6}$
TC008645	SIFa preprohormone	3.7	$4.0x10^{-4}$
Transport proteins			
TC011400	ionotropic glutamate receptor*	4.1	$1.0x10^{-5}$
TC012232	sugar/inositol transporter*	3.0	$2.7x10^{-4}$
TC011911	triose-phosphate transporter domain*	2.2	$1.6x10^{-8}$
TC013741	lipid transport protein*	3.3	$5.0x10^{-3}$
Signal transduction			
TC009749	G protein-coupled receptor, rhodopsin-like*	3.5	$4.0x10^{-6}$
Developmental protein			
TC014086	wnt*	4.2	$3.9x10^{-6}$
TC004299	maverick*	4.0	$2.0x10^{-10}$
Immune response			
TC006229	major royal jelly protein*	4.8	$1.0x10^{-14}$
TC005096	coleoptericin*	10.4	$6.3x10^{-19}$
TC005093	coleoptericin*	10.0	$8.1x10^{-17}$
TC007738	attacin, C-terminal*	7.9	$3.1x10^{-15}$
TC007737	attacin, C-terminal*	4.9	$2.6x10^{-7}$
TC000520	spatzle	3.3	$3.4x10^{-10}$
Pheromone receptors, odorant binding, hormone binding proteins			
TC030445	odorant binding protein 19	12.2	$8.4x10^{-21}$
TC007742	odorant binding protein 13	8.0	$2.0x10^{-5}$
TC015902	chemosensory protein 18	8.1	$8.5x10^{-9}$
TC015950	chemosensory protein 19	7.6	$6.5x10^{-4}$
TC010064	odorant binding protein c05	4.1	$1.7x10^{-6}$
TC010063	odorant binding protein 10	3.1	$1.8x10^{-4}$
TC030450	odorant binding protein 23	2.5	$5.3x10^{-3}$
TC010069	odorant binding protein C07	3.7	$5.2x10^{-3}$
TC001183	haemolymph juvenile hormone binding*	6.1	$1.1x10^{-7}$
TC014397	hemolymph juvenile hormone binding*	3.0	$2.5x10^{-3}$
Unknown conserved proteins			
TC004632	protein of unknown function DM4	3.1	$4.0x10^{-6}$
TC003571	protein of unknown function 4779	3.7	$2.4x10^{-8}$
TC009456	protein of unknown function	2.9	$1.4x10^{-5}$
TC008051	protein of unknown function DUF3827	2.6	$1.2x10^{-6}$

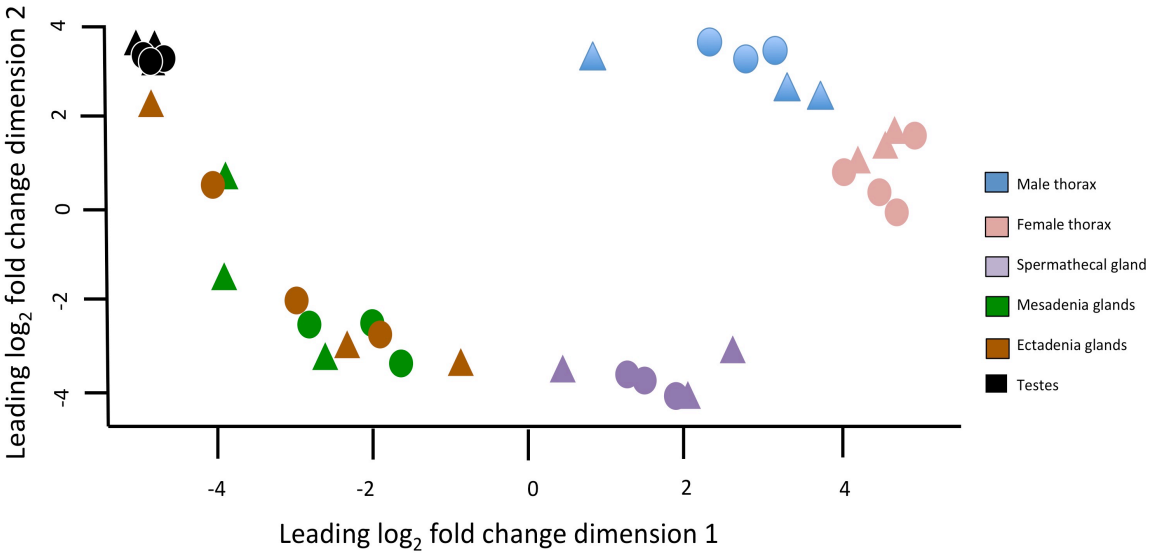
* proteins annotated by protein functional domain information from InterProScan only

Supplementary Materials

Supplementary Figure 1. Experimental design comparing gene expression between monandrous and polygamous *T. castaneum* beetle lines for both male and female reproductive tissues (MES = mesadenia glands; EC = ectadenia glands; TES = testes)

Polygamous lines	MES	EC	TES	Male Thorax	Spermathecal gland	Female Thorax
Monandrous lines	MES	EC	TES	Male Thorax	Spermathecal gland	Female Thorax

Supplementary figure 2. Biological coefficient of variation analysis of gene abundance (normalized read counts) across all male and female tissues libraries and replicate samples. Samples from polygamous individuals are represented by circles and monandrous individuals are represented by triangles.



Supplementary Table 1.Genes differentialy expressed between monandrous and polyandrous lines. Negative number indicate up-regulation in polyandrous lines.

Tissue	Gene name	Log_2FC	FDR	Gene description
Testes				
	TC030072	6.7	8.9x10-7	Trypsin-like serine peptidase
	TC000863	-4.1	2.3x10-5	Immunoglobulin I-set
	TC000424	-4.6	5.8x10-3	DNA binding domain
	TC030341	-7.5	6.0x10-7	odorant binding protein
Mesadenia glands				
	TC005634	12.3	1.6x10-3	novel
	TC002483	11.6	9.6x10-5	novel
	TC001183	8.6	9.6x10-3	JHBP
	TC007267	8.6	6.0x10-3	novel
	TC015030	8.4	6.0x10-3	cytochrome P450 307B1
	TC003112	8.3	4.2x10-3	novel
	TC013742	8.1	5.8x10-3	pleckstrin homology-like domain
	TC010906	8.0	2.7x10-3	serine peptidase H104
	TC001181	7.7	8.9x10-3	novel
	TC003508	7.4	6.0x10-3	novel
	TC016278	7.0	3.4x10-3	ELO family*
	TC001585	6.8	6.0x10-3	ankyrin repeat
	TC000803	6.6	2.8x10-3	novel
	TC003539	6.5	1.3x10-3	major royal jelly protein
	TC004547	6.2	5.2x10-3	insect cuticle
	TC011139	6.1	2.6x10-3	obstractor B
	TC015203	6.1	7.2x10-3	Fatty acyl-CoA reductase
	TC007185	6.0	4.2x10-3	lipase
	TC007129	6.0	2.1x10-3	acyltransferase 3
	TC013132	5.9	3.3x10-3	insect cuticle protein
	TC012466	5.6	4.7x10-3	novel
	TC012409	5.5	4.1x10-4	novel
	TC005412	5.5	1.9x10-3	glycoside hydrolase
	TC008459	5.4	3.4x10-3	major facilitator superfamily domain
	TC000955	5.2	4.3x10-3	novel
	TC014916	5.1	1.3x10-3	CRAL-TRIO domain
	TC011972	5.1	5.5x10-3	Metallopeptidase
	TC009252	4.9	3.4x10-3	serine/threonine-protein kinase
	TC012889	4.8	8.9x10-3	novel
	TC014917	4.7	1.3x10-3	CRAL-TRIO domain
	TC008369	4.7	4.3x10-3	novel

	TC003905	4.6	9.7x10-3	novel
	TC012756	4.5	4.0x10-5	cluster of differentiation 36 antigen
	TC003554	4.4	1.3x10-3	tetratricopeptide repeat
	TC009412	4.2	2.7x10-3	cytochrome P450
	TC015938	4.0	2.0x10-4	NAD-binding
	TC006854	3.4	3.4x10-3	SRC homology 3 domain
	TC012169	2.8	2.7x10-3	ABC transporter-like
	TC014642	2.8	6.0x10-3	tubulin-tyrosine ligase
	TC008582	-2.8	7.8x10-3	Ribonuclease H-like domain
	TC006068	-3.1	2.6x10-3	novel
	TC016174	-3.2	6.6x10-3	Reverse transcriptase domain
	TC016159	-3.3	6.2x10-3	novel
	TC004937	-3.4	5.9x10-3	Peptidase S1, trypsin family
	TC008583	-3.5	4.6x10-3	novel
	TC016160	-3.6	1.3x10-3	novel
	TC016388	-4.2	4.0x10-5	novel
	TC009693	-4.5	3.4x10-3	novel
	TC000495	-4.5	4.0x10-5	Serine peptidase P8
Spermathecal gland				
	TC002542	4.7	2.2x10-3	CYP303A1
	TC008101	4.2	2.2x10-3	Protein kinase-like domain
	TC016160	-4.7	3.0x10-3	novel
	TC009364	-6.8	2.47E-06	cathepsin L

Chapter 6. Molecular composition of male nuptial gifts in fireflies: shedding light on postcopulatory sexual selection

Abstract

Postcopulatory sexual selection is recognized as a key driver of reproductive trait evolution, including the machinery required to produce endogenous nuptial gifts. Despite the importance of such gifts, the molecular composition of the non-gametic components of male ejaculates and their interactions with female reproductive tracts remain poorly understood. During mating, males of *Photinus* fireflies transfer to females a coiled spermatophore gift manufactured by multiple reproductive glands. Here we combined transcriptomics of both male and female reproductive glands with proteomics and metabolomics to better understand the synthesis, composition and fate of the spermatophore in the common Eastern firefly, *Photinus pyralis*. Our transcriptome of male glands revealed up-regulation of several proteases that may enhance male fertilization success and activate female immune response. Using bottom-up proteomic analysis we identified 208 functionally annotated proteins that were transferred to the female in the male spermatophore, 68 of which we linked to up-regulated gene expression in specific male glands. Targeted metabolomic analysis also provided the first evidence, to our knowledge, that *Photinus* nuptial gifts contain lucibufagin, a previously described firefly defensive toxin. The reproductive tracts of female fireflies showed increased gene expression for several proteases that may be involved in egg production. This study offers new insights into the molecular composition of male spermatophores in a non-model organism, and extends our understanding of how nuptial gifts may mediate postcopulatory interactions between the sexes.

Introduction:

A powerful driver of evolutionary change, sexual selection is responsible for shaping myriad traits that impact reproductive success for individuals of each sex. It is now widely recognized that both intrasexual competition and intersexual choice continue to operate beyond the end of copulation (Simmons 2001; Peretti & Aisenberg 2015). These prolonged sexual interactions play out inside female reproductive tracts, and are accomplished through the action of diverse molecules manufactured by both sexes (Ravi Ram & Wolfner 2007). Polyandry results in temporal and spatial overlap of male ejaculates, and thus heightens selective pressure on males to maximize their own paternity success relative to that of their rivals (Kvarnemo & Simmons 2013; Simmons 2001; Eberhard 1996). One potential result is the elaboration of male reproductive glands that manufacture substances to increase relative reproductive success (Simmons 2001; Kvarnemo & Simmons 2013). During copulation, males in diverse species deliver to females not only sperm, but a complex array of seminal products including proteins, hormones, nucleic acids, lipids, amino acids, carbohydrates and defensive compounds (Poiani 2006; Perry et al. 2013). Known as endogenous nuptial gifts, these non-gametic components of male ejaculates can be transferred either via seminal fluid or within a sperm-containing package known as a spermatophore (Lewis & South 2012; Lewis et al. 2014). Among gift components, seminal fluid proteins (SFPs) have been characterized most extensively, particularly in humans and the fruit fly *Drosophila melanogaster*. (Avila et al. 2011). Male *D. melanogaster* transfer more than 200 distinct SFPs to their mates during copulation, and human seminal fluid contains approximately 100 different SFPs (Chapman 2008; Avila et al. 2011; Fung et al. 2004; Findlay et al. 2009; Findlay et al. 2014). Despite the remarkable diversity and rapid evolution of SFPs (Swanson & Vacquier 2002; Clark 2006), the major protein classes are highly conserved, suggesting

functional similarities across taxa even as distant as insects and mammals (Mueller et al. 2004; Laflamme & Wolfner 2013).

Interestingly, SFPs have been shown to dramatically alter female physiology and behavior once deposited inside the female reproductive tract. For instance, in *D. melanogaster* SFPs reduce female receptivity to further mating, increase oogenesis and oviposition, alter female sperm storage and use, and change female feeding and sleep patterns (Gillott 2003; Avila et al. 2011; Ravi Ram & Wolfner 2007). However, despite decades of research on nuptial gifts in select taxa, the detailed molecular mechanisms underlying how such gifts influence postcopulatory sexual selection remain largely unresolved. Transcriptomic studies of the male accessory glands (MAGs) that are responsible for manufacturing SFPs have been restricted primarily to *Drosophila* and other dipterans (Davies & Chapman 2006; Dottorini et al. 2007; Sirot et al. 2008). Although detailed anatomical descriptions of MAGs do exist for other taxa (Leopold 1976), their glandular products remain poorly characterized. Additionally, sexual selection research shows a recognized bias toward male reproductive traits (Ah-King et al. 2014). Thus, despite the central role that females play in postcopulatory sexual interactions, remarkably little is known about the products of female reproductive glands (Al-Wathiqui et al. 2014; Meslin et al. 2015; Prokupek et al. 2008; Prokupek et al. 2009; McGraw et al. 2004; Baer et al. 2009). Understanding the role of nuptial gifts in the context of sexual selection will require comprehensive analyses interrogating the molecular composition of male nuptial gifts as well as secretions from female reproductive tissues that receive and process male gifts.

Fireflies are bioluminescent beetles belonging to the family Lampyridae, which comprises ~2000 extant species (Capinera 2006). Their diverse life histories, sexual signals, and mating systems (Lewis 2009; Viviani 2001) have made fireflies an important group for understanding the evolution of nuptial gifts (Lewis et al. 2004; South et al.

2010). The firefly *Photinus pyralis* is a common North American species widely distributed across the eastern United States (Lloyd 1966). Historically important, *P. pyralis* was used for early studies focused on the biochemistry and physiology of bioluminescence (McElroy 1947; Buck 1948; Lloyd 1966) as well as precopulatory sexual selection (Vencl & Carlson 1998). Within the genus *Photinus*, males deliver nuptial gifts in the form of elaborate spermatophores (Lewis et al. 2004) that are manufactured by multiple reproductive accessory glands (Reijden & Monchamp 1997; Lewis et al. 2004). Because most *Photinus* fireflies do not eat in their adult stage, all reproductive activities must be fueled by stored resources acquired during larval feeding (Lewis & Christopher K Cratsley 2008). This is reflected in the decline of spermatophore size over successive matings (Cristopher K Cratsley et al. 2003). Although the production of nuptial gifts is costly for males, larger gifts are correlated with increased reproductive success (South & Lewis 2012). Male gifts also provide multiple benefits to females. Compared to females that mated only once, triply-mated *Photinus* females showed 73% greater lifetime fecundity (Rooney & Lewis 2002). Furthermore, females that receive larger nuptial gifts showed a 12-16% increase in their longevity (Rooney & Lewis 2002; South & Lewis 2012). Radiolabeling studies have shown that some spermatophore-derived proteins become incorporated into the developing oocytes of *Photinus* females (Rooney & Lewis 1999). Thus, nuptial gifts have major fitness consequences for both male and female fireflies.

Nearly 25% of all firefly species exhibit extreme sexual dimorphism: adult females completely lack wing development or have greatly reduced wings and are thus incapable of flight (Branham 2003). Physiological tradeoffs between flight and reproduction are well documented in other insects (Harshman & Zera 2007; Zera & Denno 1997), with flightlessness shifting resource allocation toward increased reproductive output (Roff 1990; Roff 1986; Roff & Fairbairn 1991; Harshman & Zera

2007; Nespolo et al. 2008; Zera & Denno 1997). Recent phylogenetic analysis revealed that female flightlessness has evolved repeatedly in the Lampyridae, typically followed by loss of male nuptial gifts (South et al. 2010). Such correlated evolution between male and female traits suggests that firefly nuptial gifts not only mediate postcopulatory sexual selection, but may also be intimately linked with patterns of female reproductive investment. Thus, better understanding of the molecular composition of firefly nuptial gifts may provide new insights into their role in postcopulatory sexual interactions as well as their influence on other key life history traits.

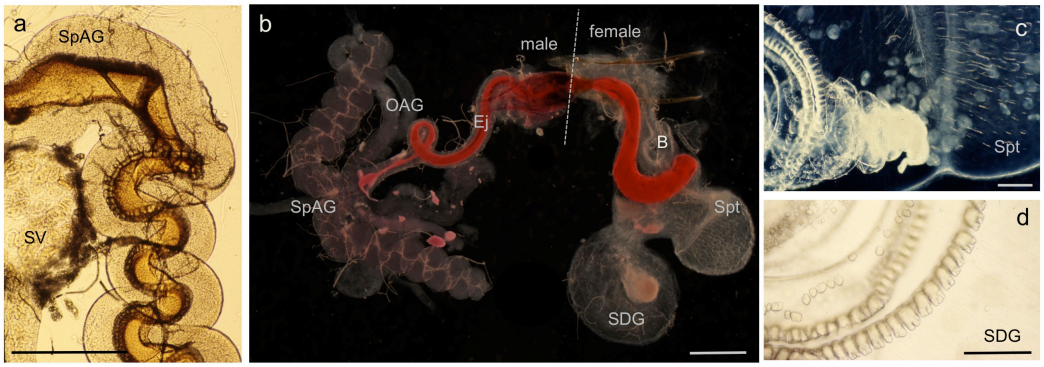


Figure 1. Nuptial gift formation, transfer and fate in *Photinus* fireflies. a) Spiral accessory glands (SpAG) manufacture the main structure of the spermatophore, visible as a dark secretion edged with serrated scales; the seminal vesicles (SV) store sperm rings that are packaged into the spermatophore upon transfer. b) During mating, the spermatophore (stained here with rhodamine b) moves through the male ejaculatory duct (Ej) into the female's reproductive tract. In addition to paired spiral glands, three other male accessory glands (OAG; long accessory gland not shown) contribute to the spermatophore. c) After transfer, sperm rings are released from the tip of the spermatophore and enter the female sperm storage organ, the spermatheca (Spt); the clear spermatophore sheath is visible (originally published in (Lewis et al., 2004)). d) The rest of the spermatophore moves into the spermatophore-digesting gland (SDG) where it disintegrates over the next 2-3 d. Scale bars are 500 um in a and b, 50 um in c and d.

In *P. pyralis*, as in many other *Photinus* fireflies, the male spermatophore is produced by four distinctive paired reproductive accessory glands (Fig. 1) (Reijden & Monchamp 1997). Most prominent are the tightly coiled spiral accessory glands

(SpAGs). Prior to mating, the lumen of each spiral gland contains a secretion edged with two longitudinal rows of serrated scales (Fig 1a). During mating, the spiral glands empty (Fig. 1b), and their secretions fuse to form the major structural component of the spermatophore. As it passes through the male ejaculatory duct (Ej, Fig. 1b), this spermatophore acquires additional material, including sperm rings that have been stored in the seminal vesicles (SV, Fig. 1a) and the contents of three additional pairs of tubular accessory glands: the short, medium, and long accessory glands, also known as other accessory glands (OAG) (Fig. 1b). Spermatophore transfer from the male to the female bursa copulatrix (B, Fig. 1b) takes about 30-60 min. Sperm rings are released into the female sperm storage organ, the spermatheca (Spt, Fig. 1c), then disperse as sperm become capacitated and begin swimming slowly in dense aggregations. The rest of the male spermatophore enters a specialized female structure known as the spermatophore-digesting gland (SDG) (Fig. 1b and d), where it disintegrates within 2-3 days after copulation.

Here, we adopted a multi-omics approach to interrogate the synthesis, content and fate of the spermatophore nuptial gift in *P. pyralis*. We sequenced and analyzed the transcriptomes of both male and female reproductive tissues, which revealed unique patterns of gene expression in these tissues. We further carried out bottom-up MS/MS proteomics and liquid chromatography-high-resolution accurate mass spectrometry (LC-HRAM-MS)-based metabolomics to explore the molecular composition of *P. pyralis* spermatophores at the protein and metabolite levels, respectively. This work expands our knowledge of nuptial gifts beyond existing model systems such as *Drosophila* and humans, and further sheds light on specific accessory gland functions and the molecular mechanisms of postcopulatory sexual selection in polyandrous animals.

Results:

Probing *P. pyralis* gene expression profiles by RNAseq

To elucidate gene expression in specific male accessory glands that manufacture nuptial gifts as well as in the female tissues that receive and process such gifts, we used RNAseq to assemble a transcriptome of *P. pyralis* reproductive tissues. We successfully demultiplexed a total of 320,271,148 reads into 18 separate libraries, each containing an average of 17,792,841 sequences. All libraries were assembled into 47,131 contigs with an average contig length of 1159 bp. Gene expression patterns indicated strong tissue-specificity, and biological coefficient of variation analysis based on normalized read counts demonstrated the expected clustering of biological replicates within each *P. pyralis* male tissue (Fig. 2).

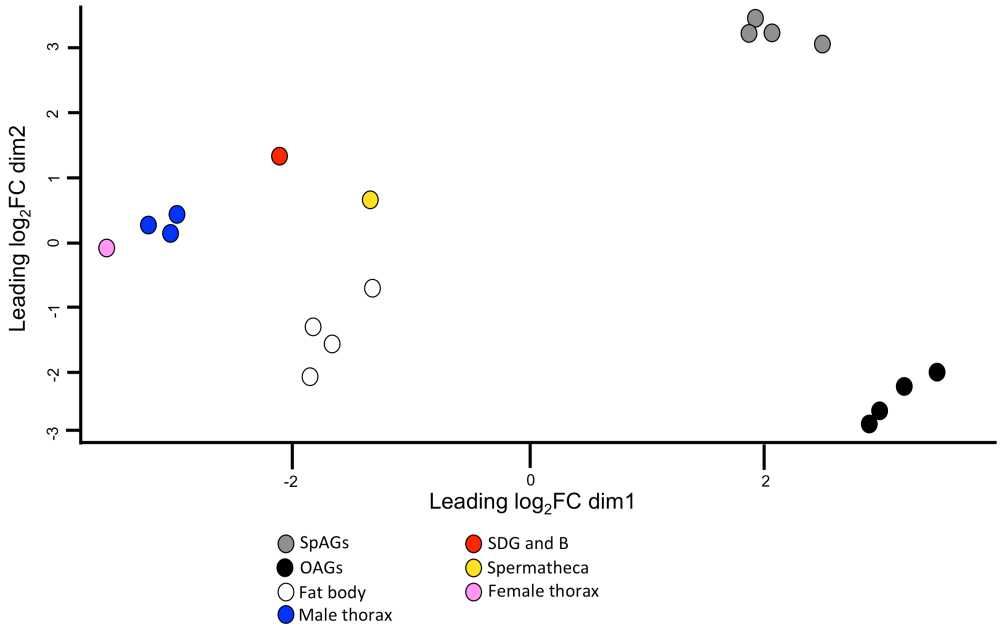


Figure 2. Biological coefficient of variation analysis of transcript abundance (normalized read counts) in *P. pyralis* male and female tissues. The x- and y-axes approximate the expression differences between samples using LogFC.

Differential gene expression in male reproductive glands

To examine *P. pyralis* differential gene expression in the SpAGs and the OAGs, we identified transcripts that showed a $\log_2$ fold change (logFC) ≥ 2 in these reproductive glands compared to male thorax, and also showed a false discovery rate (FDR) ≤ 0.01. Our transcriptome analysis identified 3294 putative transcripts that were significantly up-regulated in MAGs compared to male thorax (Supplementary Table 1). Male SpAGs and OAGs showed similar gene ontology (GO) functional categories (Molecular Function, Level III), including peptidases and peptidase regulators, metabolic processes, structural proteins, transmembrane transport and signal transduction (Fig. 3a). In male OAGs, 11.5% of the transcripts had functions related to peptidase and peptidase regulators, compared to only 4.8% of genes in male SpAGs (Fig 3a & b).

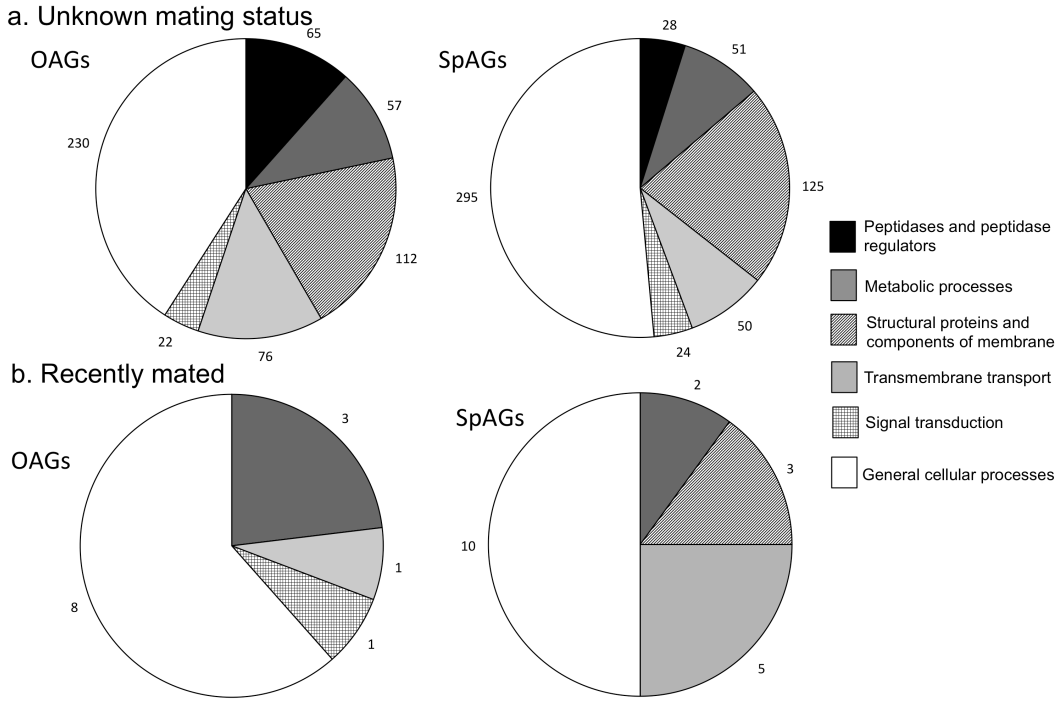


Figure 3. Distributions of gene ontology categories for *P. pyralis* genes up-regulated in males' other accessory glands (OAGs) and spiral accessory glands (SpAGs), both compared to thorax for: a) males whose mating status was unknown, and b) males that had mated within the previous 2 h.

T

o

gain insight into differentiated function between these male glands, we first identified sequences that were significantly up-regulated with LogFC ≥ 10 in either male SpAGs or OAGs compared to thorax, then identified sequences that were significantly differentially expressed between the two male gland types (LogFC ≥ 2; FDR ≤ 0.01). Comparison of GO functional categories for this subset of uniquely expressed genes confirmed that OAGs were mainly enriched in peptidase and peptidase regulator activity (Table 1).

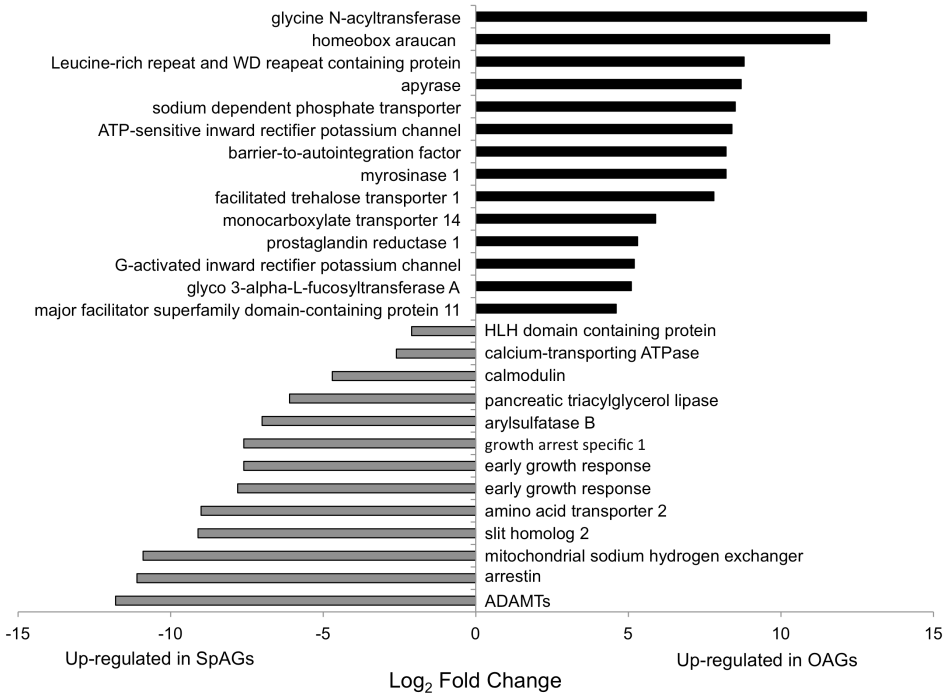


Figure 4. Comparison of differences in gene expression (log2 fold change) for annotated sequences co-expressed in OAGs and SpAGs of *P. pyralis* males.

We further characterized differences between male OAGs and SpAGs by comparing expression levels of sequences co-expressed in both tissues (Fig. 4). The 14 annotated genes that were up-regulated in male OAGs compared to SpAGs were predicted to be involved in general cellular processes. While the 13 genes up-regulated in male SpAGs compared to the OAGs included one identified as a homolog to a

metalloprotease, a disintegrin and metalloproteinase with a thrombospondin motif (ADAMTS) (Fig. 4).

Effects of mating on male gene expression

We also examined reproductive gene expression in *P. pyralis* males 2 h after mating, a time when they are actively manufacturing new spermatophores. The MAGs of recently mated males showed up-regulation of 402 genes compared to other males (Supplementary Table 1), with predicted secretion signals detected in 9% and 6% of genes that were significantly up-regulated in the SpAGs and OAGs, respectively. In comparison to other males (Fig. 3a), the OAGs of recently mated males showed greater enrichment in transmembrane transport function (Fig. 3b), primarily amino acid transporters. In contrast, the SpAGs of recently mated males showed greater enrichment in metabolic processes (Fig. 3b), particularly purine and cysteine metabolism.

Protein composition of the firefly nuptial gift

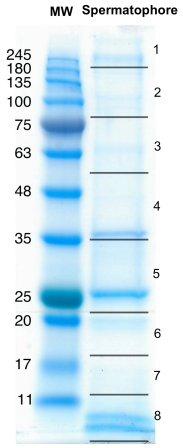


Figure 5. SDS-PAGE gel of soluble protein extract from a single *P. pyralis* male spermatophore with BLUEstain™ Protein MW ladder (numbers indicate gel sections excised for proteomic analysis).

To examine the composition of *P. pyralis* nuptial gifts, we dissected a spermatophore from a mated female immediately after copulation, separated solubilized proteins on a SDS-PAGE gel (Fig. 5), and examined protein composition by digestion of proteins into peptides followed by nano LC-HRAM-MS/MS proteomic analysis.

Combined with transcriptome data from *P. pyralis* male accessory glands and fat body, this approach allowed us to identify 425 proteins that were transferred to females in the male spermatophore. Of these, 208 were annotated (Supplementary Table 2). Based on our male transcriptome results, we were also able to determine the putative anatomical production site for 68 of these spermatophore proteins, based on our male transcriptome results (Table 2).

P. pyralis male SpAGs were identified as the production site for two serine peptidases. One of these, a transcript with homology to the peptidase snake (151_Ppyr_v3_TRINITY_DN10938_c0_g1_i1), showed a LogFC of 9.8 compared to male thorax (Table 2), which is within the top 8% of differentially expressed genes in this male gland. Snake is a member of the protease cascade that leads to the activation of the Toll pathway, which is important for *Drosophila* embryonic development and immune response activation (Valanne et al. 2011). Another peptidase (151_Ppyr_v3_TRINITY_DN8730_c0_g1_i1) showed homology to trypsin 1; with a LogFC of 8.5, this transcript is in the top 15% of most differentially expressed genes compared to male thorax (Table 2).

Proteomic analysis also confirmed presence in the *P. pyralis* spermatophore of several male reproductive proteins apparently manufactured by male OAGs (Table 2). Among the peptidases, one transcript (151_Ppyr_v3_TRINITY_DN14826_c0_g1_i1; LogFC= 3.5 compared to male thorax) showed significant similarity to Neprylisin 11 from *Tribolium castaneum* (Table 2). Another transcript showed homology to Neprilysin 2 (151_Ppyr_v3_TRINITY_DN12753_c1_g1_i4; LogFC= 10.8 compared to male thorax). Neprylisins are membrane-bound zinc metalloproteases that are responsible for the activation/inactivation of peptide hormones and neuropeptides (Thomas et al. 2005).

We also investigated the transcriptome of male fat body, an insect tissue possessing high metabolic and protein biosynthesis activity. The proteomics dataset of

the *P. pyralis* male spermatophore contained several proteins that appear to be synthesized in male fat body (Table 2). One was a cysteine protease, cathepsin L11 (151_Ppyr_v3_TRINITY_DN10232_c0_g1_i1; LogFC = 2.2 compared to male thorax), a lysosomal endopeptidase that can be secreted and interact with structural proteins, such as collagen and fibronectin (Ishidoh & Kominami 1995).

Metabolomics analysis of the firefly nuptial gift

To examine small molecule composition of the *P. pyralis* spermatophore, we conducted a LC-HRAM-MS metabolomics analysis aimed at elucidating compounds specifically enriched in the spermatophore compared to extracts from the male body with the posterior abdomen excised. In an untargeted metabolomics analysis, we noted several mass features exclusively present or present at significantly higher abundance in the spermatophore extract. However, these mass features did not match any compounds in the KEGG Database, suggesting they may represent specialized metabolites yet to be identified (Supplementary Table 3).

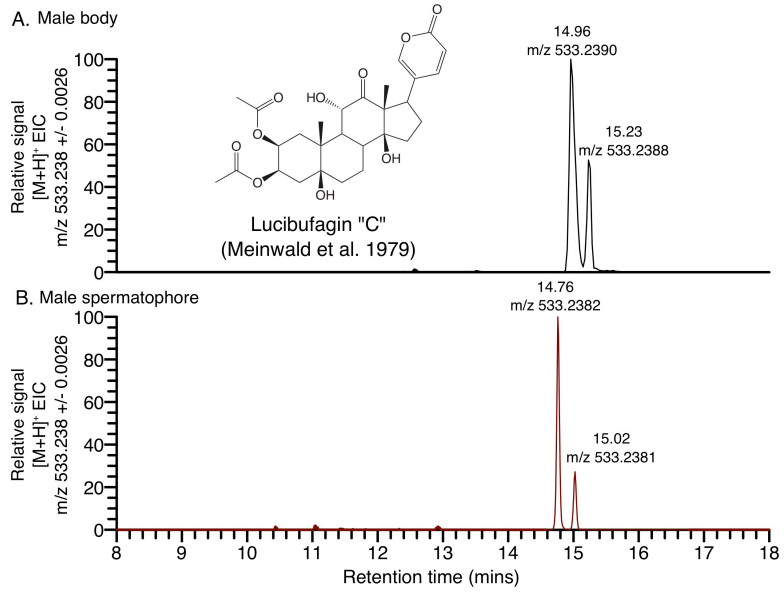


Figure 6. Positive ion mode extracted-ion-chromatograms (EIC) of the diacetylated lucibufagin $[M+H]^+$ exact mass from a LC-HRAM-MS analysis of a) *P. pyralis* male body (with posterior abdominal segments removed) and b) *P. pyralis* male spermatophore.

Using a targeted metabolomics analysis, we determined that a known firefly defense compound lucibufagin C, was present in both the spermatophore and male body. Lucibufagins have previously been shown to be a major class of anti-predator defense compounds in *Photinus* fireflies (Eisner et al. 1978; Meinwald et al. 1979). In the positive ion mode extracted-ion-chromatograms (Fig. 6), both tissues showed a large peak characteristic of lucibufagin C, as well as a smaller second peak likely representing a different isomer of diacetylated lucibufagin. This targeted analysis also identified *P. pyralis* pterin, a high abundance compound of unknown function previously co-purified with lucibufagin (Goetz et al. 1981). The identity of lucibufagin C and pterin in the spermatophore was confirmed by comparison of retention time, exact mass, and MS/MS fragmentation spectra between male body and spermatophore. These compounds were among the most abundant mass features detected in the male body extract (Supplementary Fig 1), and were identified without authentic chemical standards, as the feature retention time, exact mass, isotopic pattern, and fragmentation spectra were consistent with their respective structural identities.

Gene expression in the female reproductive tract

To determine how specific tissues might process the spermatophore and interact with male reproductive proteins, we examined differential gene expression in the female reproductive tract of *P. pyralis* relative to female thorax, although the single replicate available for female tissues meant that we could not test for statistical significance. However, using the more stringent criteria of a LogFC ≥ 3 and FDR <0.01, we identified numerous highly expressed genes in different portions of the female reproductive tract (Table 3).

The female bursa copulatrix (B) initially receives the male spermatophore, which is then moved into spermatophore-digesting gland (SDG) where the spermatophore is

degraded over several days. In the combined spermatophore-digesting gland and bursa (SDG/B) tissues, we found 80 transcripts that were up-regulated compared to female thorax, of which 33 were annotated (Table 3). We identified four sequences that showed homology to peptidases, including one sequence (151_Ppyr_v3_TRINITY_DN8737_c0_g1_i1; LogFC=4.9 compared to female thorax) with homology to angiotensin-converting enzyme, a zinc-metallopeptidase.

We also examined gene expression in the female spermatheca, where male sperm are stored prior to fertilization, and identified 80 up-regulated genes (39 annotated) in this female reproductive tissue (Table 3). As in the SDG/B, a sequence (151_Ppyr_v3_TRINITY_DN8737_c0_g1_i1; LogFC=6.2 compared to female thorax) with homology to angiotensin-converting enzyme was up-regulated in the female spermatheca, along with five other peptidases (Table 3). Another peptidase showed homology to neprilysin 2.

Discussion:

Despite recent advances, we remain in the early stages of deciphering the molecular interactions that transpire between male ejaculates and the female reproductive tract during and after mating. Clearly, a necessary first step is to identify the players on both sides. To gain insight into postcopulatory sexual interactions, we sequenced the transcriptomes of both male and female reproductive glands in the firefly *P. pyralis* and performed proteomics and metabolomics analyses of the male spermatophore gift. Firefly spermatophores are produced by several distinct male reproductive glands, and are delivered to and processed within the female reproductive tract. Our *de novo* transcriptome of male reproductive glands demonstrated up-regulation of several proteases and transport proteins, which may play important roles in postmating interactions. Combined with spermatophore proteomic analysis, we found 208 annotated

proteins packaged into the *P. pyralis* male spermatophore and transferred to females, and identified the putative anatomical production sites for 68 of these male proteins. Targeted metabolomics analysis also yielded the first evidence that *P. pyralis* males may incorporate lucibufagins, the primary antipredator defensive compounds in *Photinus* fireflies, into their nuptial gifts. We also examined gene expression in the female reproductive tract, and found up-regulation of several proteases. These results are discussed in greater detail below.

Molecular Composition of Firefly Nuptial Gifts

Recent work reveals proteases to be a conserved protein class in the male seminal gifts of diverse taxa (Laflamme & Wolfner 2013), suggesting their proteolytic roles help regulate postcopulatory interactions. This study demonstrates that the reproductive accessory glands of *P. pyralis* males synthesize serine proteases, metalloproteases, and cysteine proteases, many of which are packaged into the nuptial gift and delivered to females (Table 2).

We identified several metalloproteases produced by male accessory glands that are also transferred to females within the male spermatophore (Table 2). Metalloproteases transferred in seminal fluid of *D. melanogaster* have been linked to the induction of egg laying and are also important for spermatogenesis and fertilization (Thomas et al. 2005). Neprylisin 2, produced in *P. pyralis* male OAGs, has also been identified in male ejaculates of *Dermacentor variabilis* ticks (Sonenshine et al. 2011), and *Melanoplus sanguinipes* grasshoppers (Bonilla et al. 2015). When neprylisin-like 1 was knocked down in male mice, their mates produced smaller litters (Carpentier et al. 2004), and a study in *D. melanogaster* similarly found reduced fertility in females mated to neprylisin 2-knockdown males (Sitnik et al. 2014). Angiotensin-converting enzyme was shown here to be transferred to females in the *P. pyralis* male spermatophore (Supplementary Table 2) and also occurs in the seminal fluid of several insects, including *C. capitata* fruit flies

(Davies & Chapman 2006), *T. oceanicus* (Simmons et al. 2013) and *M. sanguinipes* grasshoppers (Bonilla et al. 2015), and the flour beetle *T. castaneum* (Xu et al. 2013). In *T. casteneum*, knockdown of angiotensin-converting enzyme in males led to the production of abnormal sperm and decreased egg production by their mates (Xu et al. 2013). We also identified ADAMTS, another metalloprotease, which was up-regulated in the spiral glands of *P. pyralis* males (Fig 4) and may be important for sperm fertilization ability (Aydos et al. 2016); however, this was not detected in our spermatophore proteomic analysis. Thus, several metalloproteases synthesized in *P. pyralis* male accessory glands may be important for increasing male fertilization success, perhaps through enhancing sperm storage and/or release.

Serine proteases represent a common component of insect ejaculates, and have been shown to mediate postmating physiological changes in females of several taxa. For example, trypsin-like serine proteases transferred in male nuptial gifts increase female oviposition in *A. socius* crickets (Marshall et al. 2009) and *D. melanogaster* fruit flies (Laflamme & Wolfner 2013). In *P. pyralis* males, the spiral glands produce two serine proteases, snake and trypsin 1, and we confirmed these are also transferred to females in spermatophores (Table 2). Snake acts as an important mediator of the Toll pathway immune response in mosquitoes and fruit flies and is part of a serine protease cascade controlling synthesis of drosomycin, an antifungal agent (Gorman & Paskewitz 2001; Hoffmann et al. 1996; Levashina et al. 1999). Thus, inclusion of these enzymes in *P. pyralis* nuptial gifts may potentially increase female immune response and reduce the likelihood of infection by microbial pathogens introduced during mating.

We also identified a cysteine protease, cathepsin L 11, a papain-like enzyme that appears to be produced by male fat body and transferred in the spermatophore to the female (Table 2). Cathepsin L is a lysosomal endopeptidase that can be secreted, and which degrades structural proteins such as collagen and fibronectin (Turk et al. 2012).

Cathepsins are responsible for digestive proteolysis in the gut of cowpea weevils *Callosobruchus maculatus* (Pedra et al. 2003), and may play a similar role in degrading the male spermatophore inside the spermatophore-digesting gland. Another possible function is suggested by high concentrations of cathepsin L found in pre-ovulatory follicles of mice where it may initiate follicular rupture and ovulation (Robker et al. 2000). Studies in other *Photinus* fireflies demonstrated that some spermatophore-derived proteins are incorporated into female oocytes (Rooney & Lewis 1999), so inclusion of cathepsin L in the *P. pyralis* nuptial gift may act to stimulate follicle degradation and egg maturation.

Certain *Photinus* fireflies are known to derive protection against their predators through biosynthesis of specialized toxic steroidal pyrones known as lucibufagins (Eisner et al. 1978; Meinwald et al. 1979). Notably, our metabolomic analysis provides preliminary evidence that *P. pyralis* males transfer detectable quantities of lucibufagins to females in their nuptial gift (Fig 6; Supplementary Table S3), as we identified lucibufagin derivatives in the spermatophore extract based on compared exact mass, fragmentation pattern, and retention time. We hypothesize that males may package lucibufagins into their nuptial gift, where these defense compounds could augment the female's own defenses to help protect the female or her eggs against predators or microbial attack. Previous studies have found that other insect males also transfer defensive chemicals to females within their spermatophores or seminal fluid (Lewis & South 2012). In many cases, such defensive compounds are derived from host plants, including pyrrolizidine alkaloids in *Utetheisa ornatrix* moths (Eisner & Meinwald 1995), cyanogenic glycosides in several *Heliconius* butterflies (Cardoso & Gilbert 2006), and vicilin-derived peptides in *Callosobruchus maculatus* cowpea beetles (Alexandre et al. 2011). In blister beetles (family Meloidae), however, males actively synthesize a toxic terpene, cantharidin, which they store in the MAG and transfer in their nuptial gifts (Nikbakhtzadeh et al. 2007). In

fireflies, further experiments are needed to definitively elucidate the source of male lucibufagins, to quantify amounts contained within male nuptial gifts, and to determine the extent to which male-derived lucibufagins may be used to defend the female or her eggs.

Gene Expression in Female Reproductive Tracts

Postcopulatory sexual interactions are evolutionarily important, yet have proven challenging to study because they typically take place within the female reproductive tract. Moreover, to date gene expression in those female reproductive tissues that receive and process male ejaculates has been examined for only a few taxa, including *Drosophila* spp. (Prokupek et al. 2008), the honeybee *Apis mellifera* (Baer et al. 2009), the corn borer moth *Ostrinia nubilalis* (Al-Wathiqui et al. 2014), and *Pieris rapae* butterflies (Meslin et al. 2015). In *Photinus* fireflies, after the male spermatophore is deposited in the female's bursa copulatrix, it enters a specialized gland (SDG) where it subsequently disintegrates over the next several days (Reijden & Monchamp 1997). Male sperm are stored and remain viable within the female's spermatheca for up to two weeks before fertilization (Demary 2005). In this study of *P. pyralis* fireflies, we demonstrated that the sperm- or spermatophore-receiving portions of the female reproductive tract express genes for proteases, protease inhibitors, and proteins involved in immune response and in maintaining sperm viability.

Female peptidases and peptidase regulators are likely to be important mediators of postcopulatory sexual interactions, and these have also been identified from female reproductive tracts of other insects (Prokupek et al. 2008; Al-Wathiqui et al. 2014; Meslin et al. 2015). In *D. melanogaster,* female peptidases and peptidase inhibitors interact with male SFPs and are required to process into their active forms at least three male proteins that induce egg-laying and reduce female receptivity to remating (Ravi Ram et al. 2005). Within the *Drosophila repleta* group, females peptidases and peptidase

inhibitors expressed in more promiscuous species show higher dN/dS ratios compared to monogamous species, indicating strong positive selection on these female reproductive proteins (Kelleher et al. 2007).

In this study, we identified several proteases that are expressed in the reproductive tract of *P. pyralis* females. In the spermatheca, we identified six peptidases, including neprilysin 2 (151_Ppyr_v3_TRINITY_DN42525_c0_g1_i1) that were up-regulated compared to the female thorax and in both female tissues, we found that a sequence encoding angiotensin-converting enzyme (151_Ppyr_v3_TRINITY_DN8737_c0_g1_i1) was up-regulated compared to the female thorax. As in male insects, female neprilysins have been shown to be critical in maintaining *D. melanogaster* fertility (Sitnik et al. 2014). When neprilysin 2 is knocked-down in *D. melanogaster* females, fewer eggs are laid and they show decreased viability (Sitnik et al. 2014). In *P. pyralis* females, neprilysin 2 may also play a role in regulating ovulation and maintaining egg viability. Angiotensin-converting enzyme was also found in both *P. pyralis* female tissues, and has previously been shown to be expressed in the bursa copulatrix and spermatheca of female *Lacanobia oleracea* moths (Ekbote et al. 2003). The function of this peptidase in the female reproductive tract has yet to be determined.

In summary, this study offers new insights into the molecular composition of the firefly spermatophore, and deepens our understanding of how such nuptial gifts can mediate postcopulatory interactions between males and females. One future challenge will be to perform functional studies in fireflies and other non-model organisms to determine how these reproductive proteins influence reproductive fitness of both sexes. Future studies examining intraspecific differences in nuptial gift composition will also shed light on the evolutionary forces that drive the origin and maintenance of nuptial gifts across taxa.

Methods:

Specimen and tissue collection

Photinus pyralis fireflies used in this study were collected at Mercer Meadows Pole Farm, Lawrenceville, NJ (40°18'23.4" N, 74°44'53.9"W) on 27 June and 11-12 July 2015, and identified based on male genitalia (Green 1956) and flash patterns. Both sexes were kept individually in plastic containers with sliced apple and damp paper towel. Mating status of field-collected individuals was unknown. Fireflies were kept in the lab for less than one week prior to experimentation.

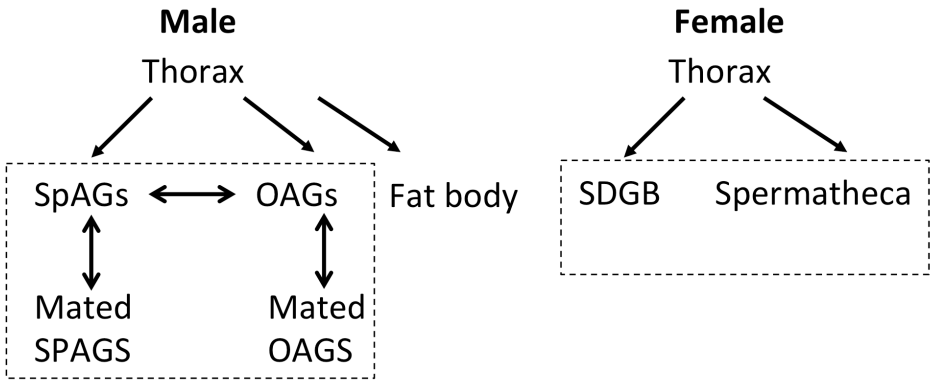


Figure 7. Planned comparisons between *P. pyralis* male and female tissues for differential expression analysis (reproductive tissues are enclosed within dashed lines). Two-way arrows indicate that both up- and down-regulated genes were examined, while one-way arrows indicate that only genes up-regulated in the destination tissue were examined (e.g. genes up-regulated in male accessory glands compared to male thorax).

We compared gene expression in male and female reproductive tissues as shown in Fig. 7. Tissues were collected from fireflies anesthetized at -20°C for 20 min then dissected under 40-70x magnification in RNAlater. From 12 *P. pyralis* males, the following tissues were dissected and pooled into 3 biological replicates (4 males each): spiral accessory glands (SpAG), other accessory glands (OAG), thorax, and fat body. Insect fat body is a metabolically active tissue responsible for protein synthesis; widely

distributed throughout the abdomen, fat body is abundant surrounding the male accessory glands. As fewer females were available, three females were pooled to produce a single biological replicate of the following tissues: spermatheca (Spt), spermatophore digesting gland and bursa (SDG/B), and thorax. All tissues were stored in RNAlater at -80°C until RNA extraction.

After mating, *Photinus* males immediately begin assembling a new spermatophore (Reijden & Monchamp 1997), thus we predicted that accessory glands of recently mated males would show higher transcription levels of functionally related genes. Males were mated with females in the lab, and then SpAGs and OAGs were dissected 2 h after mating pairs had separated. Tissues harvested from four recently mated males were pooled into one biological replicate and stored in RNAlater at -80°C until RNA extraction.

RNA extraction and sample preparation

Prior to RNA extraction, each pooled biological replicate was frozen in liquid nitrogen and homogenized in QIAzol lysis reagent (Qiagen, Valenica CA USA) using a mortar and pestle. RNA was extracted using RNeasy Lipid Tissue Kit (Qiagen), and Illumina sequencing libraries were prepared from total RNA enriched to mRNA with a polyA pulldown using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA). A total of 18 libraries were sequenced at the Whitehead Institute Genome Technology Core (Cambridge, MA) on two lanes of an Illumina HiSeq 2500 using rapid mode (PE 100 bp). Raw sequencing data has been uploaded to NCBI SRA (SRP078386).

Transcriptome assembly and differential expression analysis

Resulting RNA-Seq reads in FASTQ format were checked with the FastQC software package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and Illumina TruSeq3 adaptor contamination and low quality reads were removed by the Trimmomatic software package (http://www.usadellab.org/cms/?page=trimmomatic

(Bolger et al. 2014)), with the following parameters "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25". 185402330 paired reads pooled from all libraries remained post quality filtering. A de novo transcriptome was assembled from the pooled quality-filtered paired reads with Trinity 2.2.0 (Grabherr et al. 2011) using default parameters with the exception of "--min_glue 5 --min_kmer_cov 3", on a single high-memory server (Whitehead Institute). Candidate ORFs were translated in silico from the de novo transcriptome using Transdecoder 2.0.1 (Haas et al. 2013), with the minimum protein length set to 20 amino acids.

Expression analysis was performed using Trinity by the included "align_and_estimate_abundance.pl" script, with default parameters. This script utilizes Bowtie (Langmead & Salzberg 2012; Langmead et al. 2009) and RSEM (Li & Dewey 2011) to map reads to assembled transcripts and perform transcript quantification with expectation maximization respectively. We identified male and female genes that were significantly differentially expressed between specific tissues using the Bioconductor package edgeR (comparisons of interest shown in Figure 2; (McCarthy et al. 2012; Robinson & Oshlack 2010; Robinson et al. 2009; Robinson & Smyth 2007a). Male genes were considered significantly differentially expressed if they had a $\log_2$ fold change ≥ 2 (logFC) and a false discovery rate (FDR) ≤ 0.01. . We focused our subsequent analysis on male genes that showed significant up-regulation in either SpAGs or OAGs relative to male thorax. Female comparisons lacked replicates, so significant differential expression could not be assessed. Because of this lack of replication for female tissues, we are cautious in our conclusions drawn from this data. We also compared genes that were up-regulated compared to thorax in male SpAGs (LogFC≥10) to those up-regulated in male OAGs compared to male thorax (LogFC≥10) directly. This list of genes was then

analyzed for differential expression between the SpAGs and the OAGs to determine how the function of each tissue differs.

After differential expression analysis, all significantly differentially expressed genes were annotated using Blast2GO and InterProScan (Gotz et al. 2008; Conesa & Götz 2008; Conesa et al. 2005). To identify putative homologs, a Blast search was conducted between each sequence and the entire NCBI non-redundant protein database (Pruitt 2004). All sequences with significant Blast hits (e-value $\leq 10^{-10}$) were then mapped and annotation scores were computed for all possible gene ontology terms. We used InterProScan to obtain further protein domain/motif information, enabling us to identify protein domains that indicated secretion (Zdobnov & Apweiler 2001). Here, we only discuss sequences that that were successfully annotated using Blast2GO. All differentially expressed genes that were successfully annotated had e-values $\leq 6x10^{10}$.

Principal component analysis

We summarized multivariate variation in gene expression among various male and female tissues using principal component analysis. To normalize read counts, the trimmed mean of M-values normalization method was conducted for each transcript using edgeR (McCarthy et al. 2012; Robinson & Oshlack 2010; Robinson et al. 2009; Robinson & Smyth 2007b). Next a biological coefficient of variation analysis was conducted in edgeR.

Spermatophore proteomics

One hour after the initiation of stage II copulation (Lewis & Wang 1991), a mating pair of *Photinus pyralis* fireflies was separated, and the spermatophore was carefully dissected out from the female's reproductive tract. Upon removal from storage, the spermatophore was transferred into 50 µL of 2x Laemmli Sample Buffer (Bio-Rad) with 2% β-mercaptoethanol, and heated to 95˚C for 5 min. Sample (25 µL) was loaded onto a 12% percent discontinuous Laemelli SDS-PAGE gel. BLUEstain™ Protein ladder

(Gold Biotechnology) was loaded in a neighboring well for inferring protein size. Eight sections containing proteins ranging from >180 kDa to ~6 kDa were cut from the gel, and provided to the Whitehead Institute Proteomics Core Facility (Cambridge, MA). Thereafter the samples were digested with trypsin, and run individually on a Dionex Ultimate 3000 RSLCnano nanoflow LC coupled to a ThermoFisher Scientific Orbitrap Elite mass spectrometer. *In silico* translated ORFs from the Trinity *de novo* transcriptome were used as a search database to identify tryptic peptides from the samples. Mascot (Matrix Science, London, UK; version 2.5.1) was used as the proteomic search engine. Verification of peptide and protein identification and general analysis was performed in Scaffold (Supplementary methods 1; Scaffold version 4.4.8, Proteome Software Inc., Portland, OR). Raw proteomic data and peptide identifications have been uploaded to the EBI PRIDE database (https://www.ebi.ac.uk/pride/archive/) with the following accession number (PXD004005).

Spermatophore metabolomics

To examine the small molecule composition of the male spermatophore, we conducted an untargeted liquid-chromatography high-resolution accurate-mass mass-spectrometry (LC-HRAM-MS) metabolomic analysis aimed at elucidating compounds specifically enriched in the spermatophore. Again, a pair *P. pyralis* fireflies was separated shortly after mating, and the spermatophore carefully dissected out of the female's reproductive tract. Briefly, we compared mass features detected in 1:1 water:methanol extracts of the spermatophore and the body of an adult *P. pyralis* male whose posterior abdominal segments (including the lantern) had been removed. We conducted targeted analyses to look for lucibufagin, pterin, and several insect hormones, as well as an untargeted metabolomic analysis to identify any compounds enriched in male spermatophores. Data processing and analysis was performed with MZmine2 (Pluskal et al. 2010) (see supplementary methods for

details). Raw and feature called metabolomic data from the *P.pyralis* spermatophore and body have been uploaded to the EBI MetaboLights database (http://www.ebi.ac.uk/metabolights/) with the following accession number (MTBLS362).

Table 1. GO categories describing molecular function of genes expressed in reproductive accessory glands of *P. pyralis* males

Proposed functional class	Sequences unique to OAGs* (%, n)	Sequences unique to SpAGs† (%, n)
Transmembrane transport	3%, 6	3%, 2
Peptidases and Peptidase Regulation	8%, 14	0
Signal transduction	0.5%, 1	0
Developmental proteins	1%, 2	4%, 3
Unknown conserved proteins	0.5%, 1	0
General cellular processes	3%, 6	8%, 6
Novel	83%, 151	85%, 64

* Other accessory glands
† Spiral accessory glands

Table 2. Transcripts encoding spermatophore proteins and their proposed tissue of production in *P.pyralis* males.

Tissue and protein functional class	Sequence ID and description	e-value	% similarity	MW (kDa)	Gel Section
SpAGs					
Peptidases and peptidase regulators					
	151_Ppyr_v3_TRINITY_DN8730_c0_g1_i1-trypsin 1 like	2.4x10^{-58}	61.0%	28	5
	151_Ppyr_v3_TRINITY_DN10938_c0_g1_i1-serine protease snake	1.1x10^{-117}	70.8%	33	5
Unknown conserved proteins					
	151_Ppyr_v3_TRINITY_DN13942_c0_g1_i1-uncharacterized protein LOC656585	0	92.4%	76	3
General cellular processes					
	151_Ppyr_v3_TRINITY_DN11025_c0_g1_i1-Histone 2B	4.0x10^{-58}	99.2%	14	7
	151_Ppyr_v3_TRINITY_DN13259_c0_g1_i1-60 kDa heat shock protein	0	95.3%	61	3
	151_Ppyr_v3_TRINITY_DN8371_c0_g1_i1-elongation factor 1-alpha	0	95.3%	50	4
	151_Ppyr_v3_TRINITY_DN12873_c0_g1_i2-40S ribosomal S9	1.3x10^{-123}	99.5%	23	5
	151_Ppyr_v3_TRINITY_DN13943_c0_g1_i1-eukaryotic translation initiation factor 5A	5.0x10^{-99}	98.1%	18	7
	151_Ppyr_v3_TRINITY_DN1419_c0_g1_i1-40S ribosomal S13	1.2x10^{-95}	97.2%	17	7
Novel					
	151_Ppyr_v3_TRINITY_DN1435_c3_g5_i1			185	1
	151_Ppyr_v3_TRINITY_DN1435_c3_g6_i1			14	7
	151_Ppyr_v3_TRINITY_DN1435_c3_g8_i1			239	1
	151_Ppyr_v3_TRINITY_DN5809_c1_g1_i1			35	4
	151_Ppyr_v3_TRINITY_DN5981_c0_g1_i1			13	7
	151_Ppyr_v3_TRINITY_DN16695_c0_g1_i1			56	3
	151_Ppyr_v3_TRINITY_DN2808_c0_g1_i1			13	7
	151_Ppyr_v3_TRINITY_DN5177_c0_g1_i1			14	7
	151_Ppyr_v3_TRINITY_DN6688_c0_g1_i1			13	7
	151_Ppyr_v3_TRINITY_DN12693_c3_g3_i1			71	3
	151_Ppyr_v3_TRINITY_DN15103_c0_g2_i1			49	4
	151_Ppyr_v3_TRINITY_DN14054_c0_g1_i1			82	2
	151_Ppyr_v3_TRINITY_DN15103_c0_g1_i1			15	7
	151_Ppyr_v3_TRINITY_DN4568_c0_g1_i1			20	6
	151_Ppyr_v3_TRINITY_DN17631_c0_g1_i2			250	1
	151_Ppyr_v3_TRINITY_DN10347_c0_g1_i1			28	5
OAGs					
Peptidases and peptidase regulators					
	151_Ppyr_v3_TRINITY_DN14826_c0_g1_i1-	0	83.8%	89	2

	neprilysin-11				
	151_Ppyr_v3_TRINITY_DN17149_c1_g1_i1-aminopeptidase N	0	65.3%	105	2
	151_Ppyr_v3_TRINITY_DN12753_c1_g1_i4-neprilysin 2	$2.0x10^{-104}$	72.7%	31	5
Structural component of cell					
	151_Ppyr_v3_TRINITY_DN19550_c0_g1_i1-glycine rich cell wall structural	$7.3x10^{-19}$	65.7%	31	5
	151_Ppyr_v3_TRINITY_DN7522_c0_g1_i1-muscle specific protein 20	$3.3x10^{-111}$	92.3%	20	6
General cellular processes					
	151_Ppyr_v3_TRINITY_DN11298_c0_g1_i1-annexin B10	$7.7x10^{-144}$	80.8%	36	4
	151_Ppyr_v3_TRINITY_DN1632_c0_g1_i1-V-type proton ATPase subunit G	$4.2x10^{-33}$	94.8%	14	7
	151_Ppyr_v3_TRINITY_DN15081_c0_g1_i1-heat shock 70 kDa cognate 3	0	97.6%	72	3
	151_Ppyr_v3_TRINITY_DN16493_c0_g1_i1-Glutamate dehydrogenase	0	90.1%	61	3
	151_Ppyr_v3_TRINITY_DN7056_c1_g1_i1-arylsulfatase B	0	78.9%	58	4
	151_Ppyr_v3_TRINITY_DN17778_c0_g2_i6-insulin-like growth factor-binding complex	$4.6x10^{-25}$	52.6%	29	5
	151_Ppyr_v3_TRINITY_DN13283_c0_g1_i1-V-type proton ATPase subunit A	0	97.2%	68	3
	151_Ppyr_v3_TRINITY_DN1410_c0_g1_i1-V-type proton ATPase subunit B	0	99.0%	55	4
	151_Ppyr_v3_TRINITY_DN54_c0_g1_i1-disulfide-isomerase A3	0	85.8%	55	4
Novel					
	151_Ppyr_v3_TRINITY_DN30894_c2_g1_i1			30	5
	151_Ppyr_v3_TRINITY_DN8691_c0_g1_i1			61	3
	151_Ppyr_v3_TRINITY_DN17778_c0_g2_i2			63	3
	151_Ppyr_v3_TRINITY_DN17778_c0_g2_i1			44	4
	151_Ppyr_v3_TRINITY_DN14800_c0_g1_i1			69	3
	151_Ppyr_v3_TRINITY_DN13010_c0_g1_i4			44	4
	151_Ppyr_v3_TRINITY_DN6834_c0_g1_i2			35	5
	151_Ppyr_v3_TRINITY_DN17799_c6_g3_i11			50	4
	151_Ppyr_v3_TRINITY_DN15764_c0_g1_i1			48	4
	151_Ppyr_v3_TRINITY_DN10439_c0_g1_i1			15	7
	151_Ppyr_v3_TRINITY_DN2338_c0_g1_i1			50	4
	151_Ppyr_v3_TRINITY_DN13548_c0_g2_i1			45	4
	151_Ppyr_v3_TRINITY_DN12812_c0_g1_i1			23	6
	151_Ppyr_v3_TRINITY_DN16243_c0_g1_i1			12	7
	151_Ppyr_v3_TRINITY_DN17761_c0_g1_i1			13	7
	151_Ppyr_v3_TRINITY_DN2953_c0_g2_i1			20	6
	151_Ppyr_v3_TRINITY_DN16752_c1_g1_i1			82	2
	151_Ppyr_v3_TRINITY_DN6760_c0_g1_i1			53	4
	151_Ppyr_v3_TRINITY_DN2319_c1_g1_i1			11	8

	151_Ppyr_v3_TRINITY_DN14594_c0_g1_i1			83	2
	151_Ppyr_v3_TRINITY_DN17799_c6_g3_i7			22	6
Fat body					
Peptidases and peptidase regulators					
	151_Ppyr_v3_TRINITY_DN10232_c0_g1_i1-cathepsin L11	0	81.7%	63	3
General cellular processes					
	151_Ppyr_v3_TRINITY_DN15562_c0_g1_i1-beta-galactosidase	0	79.3%	75	3
Novel					
	151_Ppyr_v3_TRINITY_DN13548_c0_g2_i1			45	4
	151_Ppyr_v3_TRINITY_DN5991_c0_g1_i1			43	4
	151_Ppyr_v3_TRINITY_DN12416_c3_g1_i2			370	1
	151_Ppyr_v3_TRINITY_DN17128_c1_g1_i1			100	2
	151_Ppyr_v3_TRINITY_DN17195_c0_g2_i1			24	5
	151_Ppyr_v3_TRINITY_DN6374_c0_g1_i1			74	3

Table 3. Transcripts encoding female reproductive genes and their annotation description in *P.pyralis* fireflies.

Tissue and category	Sequence ID and description	e-value	% similarity
SDGB			
Peptidases and peptidase regulators			
	151_Ppyr_v3_TRINITY_DN17891_c2_g1_2-protease RNA-dependent RNA partial	$7.2x10^{-36}$	73.9%
	151_Ppyr_v3_TRINITY_DN10767_c0_g1_i1-protease/RNA-dependent RNA polymerase	$7.6x10^{-41}$	82.6%
	151_Ppyr_v3_TRINITY_DN15516_c0_g1_i5-dipeptidyl peptidase 9	0	93.1%
	151_Ppyr_v3_TRINITY_DN8737_c0_g1_i1-angiotensin-converting enzyme	0	89.9%
Transport			
	151_Ppyr_v3_TRINITY_DN12916_c0_g1_i1-solute carrier family 12 member 6 isoform X7	0	85.0%
	151_Ppyr_v3_TRINITY_DN15627_c0_g1_i5-plasma membrane calcium-transporting ATPase 2	0	94.1%
	151_Ppyr_v3_TRINITY_DN5842_c0_g1_i1-innexin inx1	0	88.9%
Development			
	151_Ppyr_v3_TRINITY_DN17388_c0_g1_i4-suppressor of hairless	0	96.1%
	151_Ppyr_v3_TRINITY_DN15473_c0_g1_i4-homeobox abdominal-B-like isoform X3	$4.6x10^{-77}$	70.4%
Unknown conserved proteins			
	151_Ppyr_v3_TRINITY_DN9101_c0_g1_i1-uncharacterized protein LOC664274	$6.1x10^{-58}$	83.0%
General cellular processes			
	151_Ppyr_v3_TRINITY_DN13365_c0_g1_i3-zinc finger 271-like	0	62.1%
	151_Ppyr_v3_TRINITY_DN16484_c0_g1_i2-DNA helicase MCM9	0	82.8%
	151_Ppyr_v3_TRINITY_DN16354_c1_g1_i2-splicing factor 1	$1.1x10^{-133}$	98.1%
	151_Ppyr_v3_TRINITY_DN12336_c0_g1_i1-E3 ubiquitin-ligase TRIM37-like	0	72.6%
	151_Ppyr_v3_TRINITY_DN13021_c0_g1_i1-Mitochondrial dicarboxylate carrier	$2.2x10^{-145}$	85.3%
	151_Ppyr_v3_TRINITY_DN17027_c2_g1_i5-eukaryotic translation initiation factor 2-alpha kinase 4	$1.0x10^{-76}$	49.8%
	151_Ppyr_v3_TRINITY_DN12815_c0_g2_i3-dihydropyrimidinase isoform X1	$8.9x10^{-174}$	88.5%
	151_Ppyr_v3_TRINITY_DN10803_c0_g1_i2-cyclin-L1 isoform X2	$2.6x10^{-109}$	80.7%
	151_Ppyr_v3_TRINITY_DN12251_c0_g1_i1-carnitine O-palmitoyltransferase mitochondrial	0	71.5%
	151_Ppyr_v3_TRINITY_DN13981_c0_g1_i2-ESF1 homolog	0	75.0%
	151_Ppyr_v3_TRINITY_DN13659_c0_g1_i5-RNA binding protein	$2.9x10^{-164}$	98.4%
	151_Ppyr_v3_TRINITY_DN17535_c1_g4_i3-Sphingomyelin phosphodiesterase	$1.5x10^{-75}$	73.0%
	151_Ppyr_v3_TRINITY_DN17637_c0_g2_i5-acyl- synthetase family member 4 isoform X3	$2.2x10^{-132}$	59.6%
	151_Ppyr_v3_TRINITY_DN15346_c0_g1_i1-alkaline	$8.4x10^{-178}$	67.6%

	phosphatase 4-like		
	151_Ppyr_v3_TRINITY_DN10122_c0_g2_i1-alpha-L-fucosidase	0	79.3%
	151_Ppyr_v3_TRINITY_DN11997_c0_g1_i1-polyprotein	$7.7x10^{-68}$	49.8%
	151_Ppyr_v3_TRINITY_DN17891_c2_g1_i1-polyprotein	$2.1x10^{-112}$	66.2%
	151_Ppyr_v3_TRINITY_DN12512_c0_g1_l-polyprotein	$4.0x10^{-57}$	66.4%
	151_Ppyr_v3_TRINITY_DN15814_c0_g1_i1-polyprotein	$1.5x10^{-66}$	76.7%
	151_Ppyr_v3_TRINITY_DN13377_c0_g1_i1-polyprotein	$3.8x10^{-31}$	31.7%
	151_Ppyr_v3_TRINITY_DN8502_c0_g1_i1-polyprotein	$3.3x10^{-24}$	56.6%
	151_Ppyr_v3_TRINITY_DN454_c0_g1_i1-polyprotein	$2.7x10^{-25}$	56.8%
	151_Ppyr_v3_TRINITY_DN17327_c1_g1_i3-blastopia polyprotein	0	63.7%
Spermatheca			
Peptidases and peptidase regulators			
	151_Ppyr_v3_TRINITY_DN16681_c1_g1_i1-probable aminopeptidase NPEPL1	0	87.7%
	151_Ppyr_v3_TRINITY_DN17891_c2_g1_i2-protease RNA-dependent RNA partial	$7.2x10^{-36}$	73.9%
	151_Ppyr_v3_TRINITY_DN8737_c0_g1_i1-angiotensin-converting enzyme	0	89.9%
	151_Ppyr_v3_TRINITY_DN42525_c0_g1_i1-neprilysin 2	0	89.5%
	151_Ppyr_v3_TRINITY_DN10767_c0_g1_i1-protease RNA-dependent RNA partial	$7.6x10^{-41}$	82.6%
	151_Ppyr_v3_TRINITY_DN15516_c0_g1_i5-dipeptidyl peptidase 9	0	93.1%
Chitin metabolism			
	151_Ppyr_v3_TRINITY_DN5290_c0_g1_i1-chitin deacetylase 1 precursor	0	96.5%
Structural/ components of membrane			
	151_Ppyr_v3_TRINITY_DN16210_c0_g1_i11-inward rectifier potassium channel 2-like isoform X1	0	82.1%
	151_Ppyr_v3_TRINITY_DN15532_c0_g1_i2-Aquaporin	$2.5x10^{-98}$	80.6%
	151_Ppyr_v3_TRINITY_DN2886_c0_g1_i1-lethal(3)malignant blood neoplasm 1	$9.4x10^{-52}$	82.9%
Transport			
	151_Ppyr_v3_TRINITY_DN12916_c0_g1_i1-solute carrier family 12 member 6 isoform X7	0	85.0%
	151_Ppyr_v3_TRINITY_DN15627_c0_g1_i3-plasma membrane calcium-transporting ATPase 2 isoform X1	0	95.8%
	151_Ppyr_v3_TRINITY_DN15627_c0_g1_i5-plasma membrane calcium-transporting ATPase 2 isoform X1	0	94.1%
Development			
	151_Ppyr_v3_TRINITY_DN17388_c0_g1_i4-suppressor of hairless	0	96.1%
	151_Ppyr_v3_TRINITY_DN15473_c0_g1_i8-homeobox abdominal-B isoform X1	$3.8x10^{-80}$	76.4%
	151_Ppyr_v3_TRINITY_DN15473_c0_g1_i4-homeobox abdominal-B-like isoform X3	$4.6x10^{-77}$	70.4%
	151_Ppyr_v3_TRINITY_DN17215_c0_g5_i5-homeobox extradenticle isoform X3	$6.2x10^{-178}$	84.0%
Unknown conserved proteins			
	151_Ppyr_v3_TRINITY_DN9101_c0_g1_i1-uncharacterized protein LOC664274	$6.1x10^{-58}$	83.0%

General cellular processes			
	151_Ppyr_v3_TRINITY_DN17737_c0_g1_i2-zinc finger 106	$2.2x10^{-74}$	61.2%
	151_Ppyr_v3_TRINITY_DN15941_c0_g1_i5-tetratricopeptide repeat 14 homolog isoform X2	$7.4x10^{-110}$	86.2%
	151_Ppyr_v3_TRINITY_DN16354_c1_g1_i2-splicing factor 1	$1.0x10^{-133}$	98.1%
	151_Ppyr_v3_TRINITY_DN13283_c0_g1_i1-V-type proton ATPase catalytic subunit A	0	97.2%
	151_Ppyr_v3_TRINITY_DN11767_c0_g2_i2-V-type proton ATPase subunit H isoform X1	0	92.3%
	151_Ppyr_v3_TRINITY_DN17027_c2_g1_i5-eukaryotic translation initiation factor 2-alpha kinase 4	$8.6x10^{-79}$	41.7%
	151_Ppyr_v3_TRINITY_DN15391_c2_g1_i2-four and a half LIM domains 2 isoform X7	$8.2x10^{-134}$	92.8%
	151_Ppyr_v3_TRINITY_DN10803_c0_g1_i2-cyclin-L1 isoform X2	$2.6x10^{-109}$	80.7%
	151_Ppyr_v3_TRINITY_DN7056_c1_g1_i1-arylsulfatase B	0	78.9%
	151_Ppyr_v3_TRINITY_DN17637_c0_g2_i5-acyl- synthetase family member 4 isoform X3	$2.2x10^{-132}$	59.6%
	151_Ppyr_v3_TRINITY_DN15346_c0_g1_i1-alkaline phosphatase 4-like	$8.5x10^{-178}$	67.6%
	151_Ppyr_v3_TRINITY_DN17839_c12_g9_i4-adenosylhomocysteinase 2 isoform X2	$2.3x10^{-146}$	98.2%
	151_Ppyr_v3_TRINITY_DN12336_c0_g1_i1-E3 ubiquitin-ligase TRIM37-like	0	72.6%
	151_Ppyr_v3_TRINITY_DN13021_c0_g1_i1-mitochondrial dicarboxylate carrier	$2.2x10^{-145}$	85.3%
	151_Ppyr_v3_TRINITY_DN17719_c0_g1_i1-ATP-binding cassette sub-family G member 8	0	84.5%
	151_Ppyr_v3_TRINITY_DN12765_c0_g1_i4-PAB-dependent poly(A)-specific ribonuclease subunit PAN3	0	88.4%
	151_Ppyr_v3_TRINITY_DN11997_c0_g1_i1-polyprotein	$7.7x10^{-68}$	76.7%
	151_Ppyr_v3_TRINITY_DN17891_c2_g1_i1-polyprotein	$2.1x10^{-112}$	73.6%
	151_Ppyr_v3_TRINITY_DN12512_c0_g1_i1-polyprotein	$3.9x10^{-57}$	66.2%
	151_Ppyr_v3_TRINITY_DN15814_c0_g1_i1-polyprotein	$1.5x10^{-66}$	49.8%
	151_Ppyr_v3_TRINITY_DN17327_c1_g1_i3-blastopia polyprotein	0	63.1%

Supplementary Materials

Supplemental Methods: Proteomics

DATABASE SEARCHING-- Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by the Mascot ExtractMSn utility. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.5.1). Mascot was set up to search translated ORFs from the de-novo transcriptome concatenated with common contaminants (45515 entries) assuming the digestion enzyme stricttrypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 20 PPM. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification and oxidation of methionine as a variable modification.

CRITERIA FOR PROTEIN IDENTIFICATION-- Scaffold (version Scaffold_4.4.8, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5383-92) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least 3 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Al et al Anal. Chem. 2003;75(17):4646-58). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Supplementary methods: Metabolomics

Collection

Photinus pyralis fireflies used in this study were collected at Mercer Meadows Pole Farm, Lawrenceville, NJ (40°18’23.4” N, 74°44’53.9”W) on 27 June and 11-12 July 2015, and identified based on male genitalia (Green, 1956) and flash patterns. Both sexes were kept individually in plastic containers with sliced apple and damp paper towel. Mating status of field-collected individuals was unknown. Fireflies were kept in the lab for less than one week prior to experimentation.

Tissue processing plus extraction

A single spermatophore dissected from a male Photinus pyralis was placed in 100 µL of 50% methanol. A single adult male Photinus pyralis was flash frozen in liquid nitrogen, and the posterior 2 abdominal segments (containing lantern & genetalia) were removed with a razor blade at 4 ˚C. The remaining anterior portion of the firefly, hereafter called the “body”, was placed in 150 µL 50% MeOH. Both tissues were macerated in the solvent, and intermittently sonicated in a water bath sonicator for 30 minutes, not letting the temperature rise above 40 ˚C. Post sonication, the extract was centrifuged in a benchtop centrifuge at 14,000 g @ 4˚C for 10 min to pellet tissue debris and other particulates. The clarified extract was filtered through a 0.2 µm PFTE filter (Filter Vial, P/No. 15530-100, Thomson Instrument Company).

Liquid chromatography
20 µL of the filtered extracts were separated on a UltiMate 3000 (Dionex) HPLC by reversed-phase chromatography on a 150 mm C18 Column (Kinetex 2.6 µm silica core shell C18 100Å pore, P/No. 00F-4462-Y0, Phenomenex), utilizing a gradient of Solvent A (0.1% formic acid in H2O) and Solvent B (0.1% formic acid in acetonitrile); 5% B for 2 min, 5-80% B over 40 min, 95% B for 4 min, and 5% B for 5 min; flow rate 0.8 mL/min. The flow from this chromatography was coupled to a Q-Exactive (Thermo-Scientific) mass spectrometer.

Mass spectrometer settings
The Q-Exactive mass spectrometer was configured to perform 1 MS1 scan from m/z 120-1250 followed by 1-3 data-dependent MS2 scans using HCD fragmentation with a stepped collision energy of 10, 15, 25 normalized collision energy (NCE). Data was collected as profile data. The instrument was always used within 7 days of the last mass accuracy calibration. The ion source parameters were as follows: spray voltage (+) at 3000 V, spray voltage (-) at 2000 V, capillary temperature at 275 ˚C, sheath gas at 40 arb units, aux gas at 15 arb units, spare gas at 1 arb unit, max spray current at 100 (µA), probe heater temp at 350 ˚C, ion source: HESI-II.

Data processing
The raw profile data in Thermo format was converted to mzML format using ProteoWizard MSConvert (Chambers et al. 2012) with 64-bit binary encoding precision, index writing, gzip compression of the whole file, zlib compression of peaklist data , and numpress linear compression of peaklist data. Data analysis was performed with MZmine2 2.19 (Pluskal et al. 2010) and Xcalibur 2.2 SP1.48 (Thermo Scientific).

Metabolite identification
1. Thermo .raw data was converted to .mzML by ProteoWizard MSConvert with the parameters specified above.

2. Raw data was imported from .mzML format files.

3. MS^1 and MS^2 profile mass spectra were detected/centroided with the “Mass Detector” module. The following parameters were used:
- Mass Detector: Exact mass
- Noise level: 1E4

4. Continuous MS^1 ions were assembled into mass traces with the “Chromatogram builder” module. The following parameters were used:
- Minimum time span (min): 0.1 the
- Minimum height: 1.0E4, and the
- m/z tolerance: 0.002 or 5ppm (whichever is greater)

5. Chromatograms were then deconvolved into individual peaks, using the “Chromatogram deconvolution” module, with the algorithm set to “Local minimum search” and following parameters:
- Chromatographic threshold: 1.0%

- Search minimum in RT range (min): 0.1
- Minimum relative height: 0.5%
- Minimum absolute height: 1.0E3
- Min ratio peak top/edge: 4
- Peak duration range (min): 0.0 – 5.0

6. The MS^1 mass spectra were then deisotoped with the "Isotopic peak grouper" module. The following parameters were used:
- m/z tolerance: 0.01025 Da or 20ppm (whichever is greater)
- Retention time tolerance: 0.2 (absolute (min))
- Monotonic shape: unchecked
- Maximum charge: 2
- Representative isotope: Most intense

7. Mass features were aligned with a correction for retention time deviation between the Photinus pyralis body and spermatophore extract, using the "RANSAC aligner" module with the following parameters:
- m/z tolerance: 8.0E-4 or 4ppm (whichever is greater)
- RT tolerance: 2.0%
- RT tolerance after correction: 0.25 (absolute (min))
- RANSAC iterations: 1000000
- Minimum number of points: 30.0%
- Threshold value: 0.08 (absolute (min))
- Linear model: unchecked
- Require same charge state: unchecked

8. Putative in-source fragments were identified by the "Fragment search" module with the following parameters:
- Retention time tolerance: 0.3 (absolute (min))
- m/z tolerance of MS2 data: 0.001 or 5ppm (whichever is greater)
- Max fragment peak height: 50%
- Min MS2 peak height: 1.0E5

9. Putative adduct ions were identified by the "Adduct search" module with the following parameters:
- Retention time tolerance: 0.3 (absolute (min))
- m/z tolerance of MS2 data: 0.0005 or 2ppm
- Max relative adduct peak height: 20%
- The following default MZmine2 adducts were selected:
    - [M+Na-H]
    - [M+K-H]
    - [M+NH3]

10. Putative complexed ions were detected by the "Complex search" module with the following parameters:
- Retention time tolerance: 0.3 (absolute (min))
- m/z tolerance of MS2 data: 0.001 or 5ppm (whichever is greater)
- Max complex peak height: 80.0%

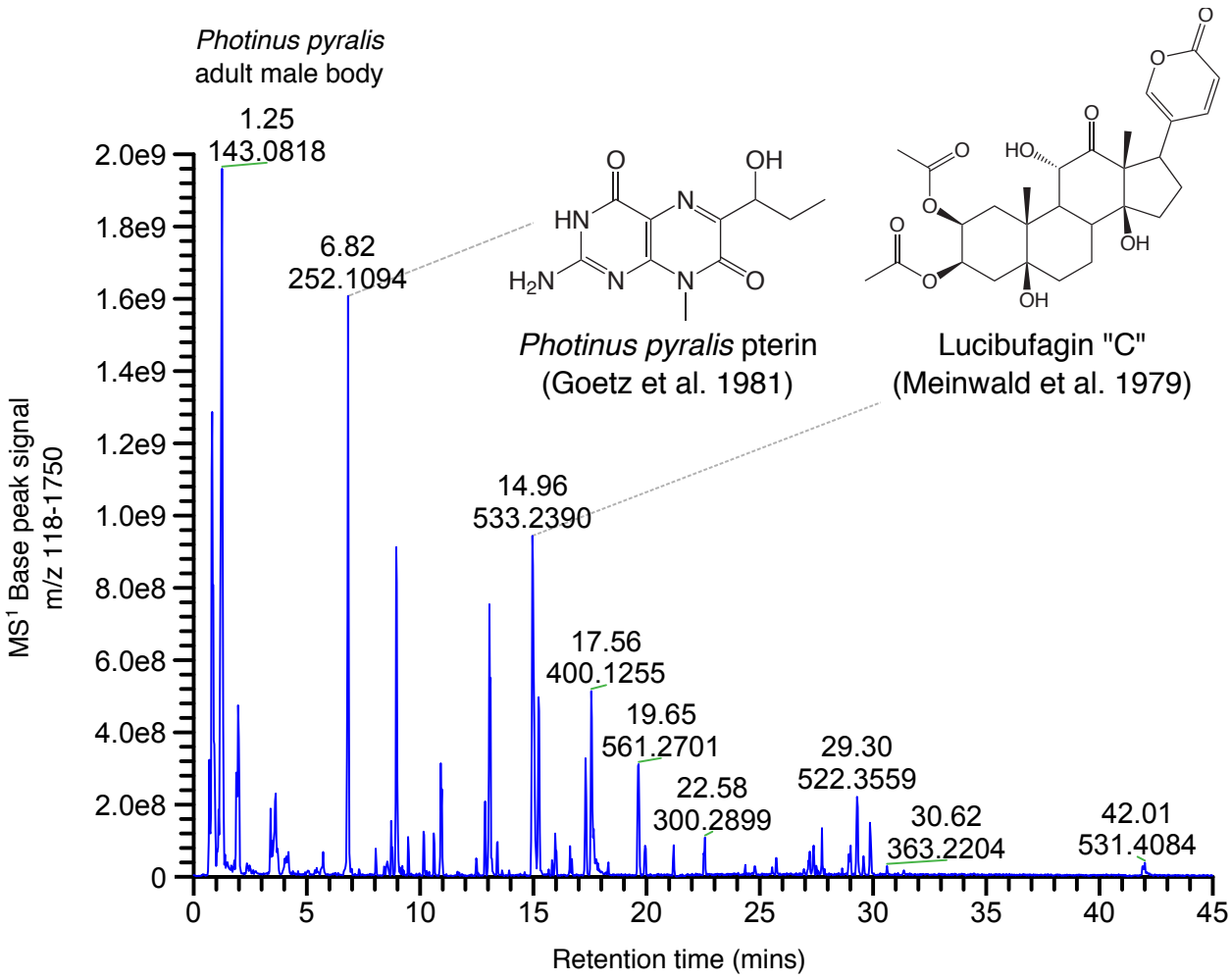
11. The resulting peaklist was exported as "spermatophore_vs_body_final_unfiltered.mzTab" (supplementary file)

12. The peaklist was filtered to remove putative complexes, adducts, and fragments, using the "peaklist rows filter" module, with the custom parameters:
- Text in identity: "Adduct" or "Complex" or "Fragment"
- Keep or remove rows: "Remove rows that match all criteria"
This step used a custom fork of MZmine2 2.1.9 (https://github.com/photocyte/mzmine2/commit/609bb3b1811b3df4ff632100591a2e449564eaf2). The change has been contributed to the main MZmine2 sourcecode repository.

13. The resulting peaklist without annotated adducts, complexes, and in-source fragment ions was exported as "spermatophore_vs_body_final_filtered.mzTab"

14. Resulting compounds were manually annotated using a combination of the "Online database search" module & "Formula prediction" module. When possible, MS^2 spectra were compared with the Metlin metabolite database (metlin.scripps.edu)

Supplementary Figure 1: C18 LC-HRAM-MS base peak chromatogram of a methanolic extract of a single whole male firefly with the last two abdominal segments removed (body).



Supplementary Table 1. Summary statistics for annotated sequences that were differentially expressed in *P. pyralis* male reproductive tissues.

Tissue comparison	Number of DE Sequences	Sequences Containing Transmembrane helices (n)	Sequences Containing Secretion signal (n)	Annotated sequences (%,n)
Male SpAGs and Thorax*	1891	362	206	30%, 573
Male OAGs *vs* Thorax*	1403	509	253	40%, 562
Male OAGs *vs* Mated Male OAG†	26	0	0	50%, 13
Male SpAGs *vs* Mated Male SpAGs†	47	0	0	43%, 20

* Statistics presented are for genes up-regulated in reproductive tissue compared to thorax
† Statistics presented are for genes up-regulated in tissues dissected from recently mated males compared to males with unknown mating status

Supplementary table 2. Annotations of sequences corresponding to spermatophore proteins in *P.pyralis*

Accession Number	Descritption
Antimicrobial and immune response proteins	
gi|1675192	dermcidin preproprotein [Homo sapiens]
151_Ppyr_v3_TRINITY_DN12083_c0_g1_i1	phenoloxidase
151_Ppyr_v3_TRINITY_DN16010_c0_g1_i1	phenoloxidase
151_Ppyr_v3_TRINITY_DN61042_c0_g1_i1	phenoloxidase
151_Ppyr_v3_TRINITY_DN36635_c0_g1_i1	phenoloxidase
Carbohydrate metabolic process	
151_Ppyr_v3_TRINITY_DN15451_c0_g1_i1	chitooligosaccharidolytic beta-N-acetylglucosaminidase
151_Ppyr_v3_TRINITY_DN14087_c0_g2_i1	chitinase EN03
151_Ppyr_v3_TRINITY_DN6995_c0_g1_i1	Neutral alpha-glucosidase AB
Signal transduction	
151_Ppyr_v3_TRINITY_DN12946_c0_g1_i1	14-3-3 epsilon
Peptidases and peptidase regulators	
151_Ppyr_v3_TRINITY_DN9669_c0_g1_i1	serine protease
151_Ppyr_v3_TRINITY_DN10938_c0_g1_i1	serine protease snake
151_Ppyr_v3_TRINITY_DN8730_c0_g1_i1	transmembrane protease serine 9
151_Ppyr_v3_TRINITY_DN5988_c0_g1_i1	angiotensin-converting enzyme
151_Ppyr_v3_TRINITY_DN15769_c1_g1_i1	aminopeptidase N
151_Ppyr_v3_TRINITY_DN17149_c1_g1_i1	aminopeptidase N
151_Ppyr_v3_TRINITY_DN10619_c0_g1_i1	carboxypeptidase Q
151_Ppyr_v3_TRINITY_DN16325_c0_g2_i1	Cathepsin L
151_Ppyr_v3_TRINITY_DN4855_c0_g1_i1	Cathepsin L
151_Ppyr_v3_TRINITY_DN16477_c0_g1_i1	cysteine ase CG12163
151_Ppyr_v3_TRINITY_DN10232_c0_g1_i1	digestive cysteine ase 1
151_Ppyr_v3_TRINITY_DN10909_c0_g1_i1	Serine protease easter
151_Ppyr_v3_TRINITY_DN13786_c0_g1_i1	Serine ase stubble
151_Ppyr_v3_TRINITY_DN16168_c0_g1_i1	gamma-glutamyltranspeptidase 1
151_Ppyr_v3_TRINITY_DN10673_c0_g1_i1	lysosomal aspartic protease
151_Ppyr_v3_TRINITY_DN14826_c0_g1_i1	membrane metallo-endopeptidase 1
151_Ppyr_v3_TRINITY_DN12753_c1_g1_i4	membrane metallo-endopeptidase 1
Development	
151_Ppyr_v3_TRINITY_DN9304_c0_g1_i1	transcription factor ken
Structural	
151_Ppyr_v3_TRINITY_DN4899_c0_g1_i1	actin
151_Ppyr_v3_TRINITY_DN6767_c0_g1_i2	annexin B9
151_Ppyr_v3_TRINITY_DN11298_c0_g1_i1	annexin B10
151_Ppyr_v3_TRINITY_DN40252_c0_g1_i1	Tubulin beta chain
151_Ppyr_v3_TRINITY_DN4678_c0_g1_i1	Tubulin beta chain

151_Ppyr_v3_TRINITY_DN11630_c0_g1_i4	troponin T
151_Ppyr_v3_TRINITY_DN11348_c0_g1_i10	troponin I
151_Ppyr_v3_TRINITY_DN11348_c0_g1_i11	troponin I
151_Ppyr_v3_TRINITY_DN8862_c0_g1_i1	tubulin alpha-1
151_Ppyr_v3_TRINITY_DN41067_c0_g1_i1	tubulin alpha-1C
151_Ppyr_v3_TRINITY_DN15129_c0_g1_i2	spectrin alpha chain
151_Ppyr_v3_TRINITY_DN7496_c0_g1_i1	troponin C
151_Ppyr_v3_TRINITY_DN12255_c0_g1_i1	cuticular RR-2
151_Ppyr_v3_TRINITY_DN2226_c0_g1_i1	cofilin actin-depolymerizing factor homolog
151_Ppyr_v3_TRINITY_DN1828_c0_g1_i1	endocuticle structural glyco bd-3
151_Ppyr_v3_TRINITY_DN11462_c0_g1_i1	Tubulin beta-1
151_Ppyr_v3_TRINITY_DN14107_c0_g1_i1	Tubulin alpha-1
151_Ppyr_v3_TRINITY_DN10944_c2_g1_i1	Spectrin beta chain
151_Ppyr_v3_TRINITY_DN5992_c0_g1_i1	myosin regulatory light chain 2
151_Ppyr_v3_TRINITY_DN6933_c0_g1_i1	myosin light chain alkali
151_Ppyr_v3_TRINITY_DN8786_c0_g1_i1	myophilin
151_Ppyr_v3_TRINITY_DN7522_c0_g1_i1	myophilin
151_Ppyr_v3_TRINITY_DN11793_c0_g1_i1	muscle-specific 20
151_Ppyr_v3_TRINITY_DN8157_c0_g1_i2	muscle LIM MIp84B
151_Ppyr_v3_TRINITY_DN48792_c0_g1_i1	muscle
151_Ppyr_v3_TRINITY_DN14440_c0_g1_i1	glycogen phosphorylase
151_Ppyr_v3_TRINITY_DN19550_c0_g1_i1	glycine-rich cell wall structure
151_Ppyr_v3_TRINITY_DN17477_c0_g1_i1	Laminin subunit gamma-1
151_Ppyr_v3_TRINITY_DN17711_c0_g1_i3	PDZ and LIM domain Zasp
151_Ppyr_v3_TRINITY_DN12682_c2_g1_i1	paxillin
151_Ppyr_v3_TRINITY_DN5860_c0_g1_i1	profilin
151_Ppyr_v3_TRINITY_DN55009_c0_g1_i1	pupal cuticle
General cellular processes	
151_Ppyr_v3_TRINITY_DN8346_c0_g1_i1	von Willebrand factor D and EGF domain-containing
151_Ppyr_v3_TRINITY_DN5244_c0_g1_i1	vacuolar sorting-associated 52
151_Ppyr_v3_TRINITY_DN6721_c0_g1_i1	ribosomal L14
151_Ppyr_v3_TRINITY_DN22312_c0_g1_i1	ubiquitin-conjugating enzyme E2
151_Ppyr_v3_TRINITY_DN9650_c0_g1_i1	transitional endoplasmic reticulum ATPase
151_Ppyr_v3_TRINITY_DN2260_c0_g1_i1	ubiquitin carboxyl-terminal hydrolase isozyme L3
151_Ppyr_v3_TRINITY_DN16145_c0_g1_i1	trifunctional enzyme
151_Ppyr_v3_TRINITY_DN7599_c0_g1_i1	succinate dehydrogenase
151_Ppyr_v3_TRINITY_DN22141_c0_g1_i1	superoxide dismutase
151_Ppyr_v3_TRINITY_DN14481_c1_g1_i1	superoxide dismutase
151_Ppyr_v3_TRINITY_DN12323_c0_g1_i1	triosephosphate isomerase
151_Ppyr_v3_TRINITY_DN9674_c0_g1_i1	60S ribosomal L9
151_Ppyr_v3_TRINITY_DN48836_c0_g1_i1	60S ribosomal L31

151_Ppyr_v3_TRINITY_DN22126_c0_g1_i1	60S ribosomal L23
151_Ppyr_v3_TRINITY_DN13259_c0_g1_i1	60 kDa heat shock mitochondrial
151_Ppyr_v3_TRINITY_DN12961_c1_g1_i2	6-phosphofructo-2-kinase fructose-2,6-bisphosphatase 1
151_Ppyr_v3_TRINITY_DN12873_c0_g1_i2	40S ribosomal S9
151_Ppyr_v3_TRINITY_DN10001_c0_g1_i1	40S ribosomal S5
151_Ppyr_v3_TRINITY_DN1419_c0_g1_i1	40S ribosomal S13
151_Ppyr_v3_TRINITY_DN58249_c0_g1_i1	40S ribosomal S10-like
151_Ppyr_v3_TRINITY_DN22108_c0_g1_i1	4-aminobutyrate mitochondrial
151_Ppyr_v3_TRINITY_DN8639_c0_g1_i1	39S ribosomal mitochondrial
151_Ppyr_v3_TRINITY_DN7521_c0_g1_i1	14-3-3 zeta isoform X1
151_Ppyr_v3_TRINITY_DN7521_c0_g1_i3	14-3-3 zeta isoform X1
151_Ppyr_v3_TRINITY_DN14336_c0_g1_i1	alpha-sarcomeric
151_Ppyr_v3_TRINITY_DN12220_c0_g1_i1	alpha-N-acetylgalactosaminidase
151_Ppyr_v3_TRINITY_DN10122_c0_g2_i1	alpha-L-fucosidase
151_Ppyr_v3_TRINITY_DN15260_c0_g1_i1	adenylate kinase isoenzyme 1 isoform
151_Ppyr_v3_TRINITY_DN53061_c0_g1_i1	adenylate kinase isoenzyme 1 isoform
151_Ppyr_v3_TRINITY_DN7600_c0_g1_i1	adenine phosphoribosyltranferase
151_Ppyr_v3_TRINITY_DN7470_c1_g1_i1	Beta-1,3-galactotransferase brn
151_Ppyr_v3_TRINITY_DN40076_c0_g1_i1	ATP synthase subunit mitochondrial
151_Ppyr_v3_TRINITY_DN13536_c0_g1_i1	ATP synthase subunit mitochondrial
151_Ppyr_v3_TRINITY_DN10157_c0_g1_i1	ATP synthase subunit mitochondrial
151_Ppyr_v3_TRINITY_DN1753_c0_g1_i1	ATP synthase subunit mitochondrial
151_Ppyr_v3_TRINITY_DN1656_c0_g1_i1	ATP synthase subunit
151_Ppyr_v3_TRINITY_DN10376_c0_g1_i1	ATP carrier
151_Ppyr_v3_TRINITY_DN34519_c0_g1_i1	ATP carrier
151_Ppyr_v3_TRINITY_DN11783_c0_g1_i1	aspartate mitochondrial
151_Ppyr_v3_TRINITY_DN7056_c1_g1_i1	arylsulfatase B
151_Ppyr_v3_TRINITY_DN8849_c0_g1_i1	Calmodulin
151_Ppyr_v3_TRINITY_DN7080_c0_g1_i1	C-type lectin isoform X1
151_Ppyr_v3_TRINITY_DN15349_c0_g1_i1	beta-glucuronidase
151_Ppyr_v3_TRINITY_DN15562_c0_g1_i1	beta-galactosidase-1 2
151_Ppyr_v3_TRINITY_DN4074_c0_g2_i1	Catalase
151_Ppyr_v3_TRINITY_DN50812_c0_g1_i1	Catalase
gi|2788141	casein kappa [Bos taurus]
gi|2780696	casein alpha-S2 [Bos taurus]
gi|3079434	casein alpha s1 [Bos taurus]
gi|99962	Chain B, Porcine E-Trypsin (E.C.3.4.21.4)
151_Ppyr_v3_TRINITY_DN6187_c0_g1_i1	CG31997
151_Ppyr_v3_TRINITY_DN42963_c0_g1_i1	CDGSH iron-sulfer domain-containing 2
151_Ppyr_v3_TRINITY_DN5028_c0_g1_i1	cdc42
151_Ppyr_v3_TRINITY_DN6022_c0_g1_i1	cytosolic none-specific dipeptidase
151_Ppyr_v3_TRINITY_DN10917_c0_g1_i1	cytoplasmic NADP-dependent isocitrate

151_Ppyr_v3_TRINITY_DN23117_c0_g1_i1	cytoplasmic 1
151_Ppyr_v3_TRINITY_DN4792_c0_g1_i1	cytochrome heme
151_Ppyr_v3_TRINITY_DN4994_c0_g1_i1	cytochrome c-2
151_Ppyr_v3_TRINITY_DN58646_c0_g1_i1	cytochrome c oxidase
151_Ppyr_v3_TRINITY_DN12285_c0_g1_i1	Cytochrome b-c1 complex subunit mitochondrial
151_Ppyr_v3_TRINITY_DN32056_c0_g1_i1	cytochrome b-c1 complex
151_Ppyr_v3_TRINITY_DN13818_c0_g1_i1	delta-1-pyrroline-5-carboxylate
151_Ppyr_v3_TRINITY_DN13560_c0_g1_i1	D-arabinitol dehydrogenase 1
gi|5853084	desmoplakin isoform I [Homo sapiens]
gi|11970374	desmoglein 1 preproprotein [Homo sapiens]
gi|13435361 (+1	desmocollin 1 isoform Dsc1a preproprotein [Homo sapiens]
151_Ppyr_v3_TRINITY_DN13270_c0_g1_i1	Volatage-dependent anion-selective channel
151_Ppyr_v3_TRINITY_DN6818_c0_g1_i1	V-type proton ATPase catalytic subunit S1
151_Ppyr_v3_TRINITY_DN1632_c0_g1_i1	V-type proton ATPase catalytic subunit G
151_Ppyr_v3_TRINITY_DN58678_c0_g1_i1	V-type proton ATPase catalytic subunit F
151_Ppyr_v3_TRINITY_DN2233_c1_g1_i1	V-type proton ATPase catalytic subunit E
151_Ppyr_v3_TRINITY_DN10946_c0_g1_i1	V-type proton ATPase catalytic subunit D1
151_Ppyr_v3_TRINITY_DN14413_c0_g1_i1	V-type proton ATPase catalytic subunit C
151_Ppyr_v3_TRINITY_DN1410_c0_g1_i1	V-type proton ATPase catalytic subunit B
151_Ppyr_v3_TRINITY_DN13283_c0_g1_i1	V-type proton ATPase catalytic subunit A
151_Ppyr_v3_TRINITY_DN2178_c0_g1_i1	UMP-CMP kinase
151_Ppyr_v3_TRINITY_DN8371_c0_g1_i1	elongation factor 1
151_Ppyr_v3_TRINITY_DN27341_c0_g1_i1	elongation factor 1
151_Ppyr_v3_TRINITY_DN54_c0_g1_i1	disulfide-isomerase A3
151_Ppyr_v3_TRINITY_DN15467_c0_g2_i1	dihydrolipoyllysine-reside acetyltransferase
151_Ppyr_v3_TRINITY_DN8547_c0_g1_i1	dihydrolipoamide dehydrogenase E3
151_Ppyr_v3_TRINITY_DN13943_c0_g1_i1	eukaryotic translation initiation factor 5A
151_Ppyr_v3_TRINITY_DN49586_c0_g1_i1	eukaryotic translation initiation factor 2 Y-linked
151_Ppyr_v3_TRINITY_DN5489_c0_g1_i1	enolase
151_Ppyr_v3_TRINITY_DN12567_c0_g1_i1	fructose-bisphosphatase aldolase
151_Ppyr_v3_TRINITY_DN14233_c0_g1_i1	fructose-bisphosphatase aldolase
151_Ppyr_v3_TRINITY_DN11409_c0_g1_i1	farnesoic acid O-methyltransferase
151_Ppyr_v3_TRINITY_DN11409_c0_g2_i1	farnesoic acid O-methyltransferase
151_Ppyr_v3_TRINITY_DN1067_c0_g1_i1	farnesoic acid O-methyltransferase
151_Ppyr_v3_TRINITY_DN10755_c0_g1_i1	NADP-dependent malic enzyme
151_Ppyr_v3_TRINITY_DN5182_c0_g1_i1	NADH-ubiquinone oxidoreductase subunit 8
151_Ppyr_v3_TRINITY_DN14301_c0_g1_i1	NADH dehydrogenase iron-sulfur
151_Ppyr_v3_TRINITY_DN11379_c0_g1_i1	NADH dehydrogenase 1
151_Ppyr_v3_TRINITY_DN22415_c0_g1_i1	NADH dehydrogenase 1
151_Ppyr_v3_TRINITY_DN1514_c0_g1_i1	NADH dehydrogenase 1

151_Ppyr_v3_TRINITY_DN5919_c0_g1_i1	NADH dehydrogenase 1
151_Ppyr_v3_TRINITY_DN10193_c0_g1_i1	NADH dehydrogenase 1
151_Ppyr_v3_TRINITY_DN9022_c0_g1_i1	glyceraldehyde-3-phosphate dehydrogenase
151_Ppyr_v3_TRINITY_DN7572_c0_g1_i1	Glutathione S-transferase
151_Ppyr_v3_TRINITY_DN16493_c0_g1_i1	Glutamate mitochondrial
151_Ppyr_v3_TRINITY_DN11197_c1_g1_i1	histone H4
151_Ppyr_v3_TRINITY_DN11025_c0_g1_i1	histone H2B
151_Ppyr_v3_TRINITY_DN1631_c0_g1_i1	high mobility group DSP 1
151_Ppyr_v3_TRINITY_DN17199_c2_g1_i1	Hermansky-Pudlak syndrome 4
151_Ppyr_v3_TRINITY_DN15081_c0_g1_i1	heat shock 70 kDa cognate 3
151_Ppyr_v3_TRINITY_DN12693_c3_g3_i1	heat shock 70
151_Ppyr_v3_TRINITY_DN22898_c0_g1_i1	heat shock 70
151_Ppyr_v3_TRINITY_DN22911_c0_g1_i1	heat shock 1
151_Ppyr_v3_TRINITY_DN5987_c0_g1_i1	guanine deaminase
151_Ppyr_v3_TRINITY_DN7428_c0_g1_i1	GTP-binding SAR1b
151_Ppyr_v3_TRINITY_DN8247_c0_g1_i1	proliferating cell nuclear antigen
151_Ppyr_v3_TRINITY_DN19462_c0_g1_i1	Programmed cell death 6
151_Ppyr_v3_TRINITY_DN9439_c0_g1_i1	Programmed cell death 6
151_Ppyr_v3_TRINITY_DN11875_c0_g1_i1	peroxiredoxin
151_Ppyr_v3_TRINITY_DN58267_c0_g1_i1	peptidyl-prolyl cis-trans isomerase 5
151_Ppyr_v3_TRINITY_DN7024_c0_g1_i1	peptidyl-prolyl cis-trans isomerase
151_Ppyr_v3_TRINITY_DN12493_c0_g1_i2	peptidoglycan-recognition LB
151_Ppyr_v3_TRINITY_DN14214_c0_g1_i1	lamin Dm0
151_Ppyr_v3_TRINITY_DN53411_c0_g1_i1	lamin Dm0
gi|12056468 (+1	junction plakoglobin [Homo sapiens]
151_Ppyr_v3_TRINITY_DN48962_c0_g1_i1	isocitrate dehydrogenase
151_Ppyr_v3_TRINITY_DN1856_c0_g1_i1	integument esterase
151_Ppyr_v3_TRINITY_DN17778_c0_g2_i6	insulin-like growth factor-binding complex
gi|5786458	hornerin [Homo sapiens]
151_Ppyr_v3_TRINITY_DN9981_c0_g1_i1	RNA-binding squid
151_Ppyr_v3_TRINITY_DN10620_c0_g1_i1	ras-related Rac1
151_Ppyr_v3_TRINITY_DN5313_c0_g1_i1	ras-like GTP-binding Rho 1
151_Ppyr_v3_TRINITY_DN8233_c1_g1_i1	Pyruvate kinase
151_Ppyr_v3_TRINITY_DN22331_c0_g1_i1	probable enoyl-mitochondrial
151_Ppyr_v3_TRINITY_DN13079_c0_g1_i1	probable citrate synthase
151_Ppyr_v3_TRINITY_DN13421_c0_g1_i1	polyubiquitin-B
151_Ppyr_v3_TRINITY_DN14322_c0_g1_i6	phospholipid hydroperoxide glutathione peroxidase
151_Ppyr_v3_TRINITY_DN14322_c0_g1_i5	phospholipid hydroperoxide glutathione peroxidase
151_Ppyr_v3_TRINITY_DN40713_c0_g1_i1	phosphoglycerate mutase 1
151_Ppyr_v3_TRINITY_DN14350_c0_g1_i1	phosphoglycerate kinase
151_Ppyr_v3_TRINITY_DN6698_c0_g1_i1	Nucleoside diphosphatase kinase

151_Ppyr_v3_TRINITY_DN8274_c0_g1_i1	Nucleoside diphosphatase kinase
151_Ppyr_v3_TRINITY_DN9600_c0_g1_i1	Mitochondrial-processing peptidase subunit beta
151_Ppyr_v3_TRINITY_DN10304_c0_g1_i1	mitochondrial hydriogen-transporting ATP synthase
151_Ppyr_v3_TRINITY_DN9630_c3_g1_i1	mitochondrial enolase superfamily member 1
151_Ppyr_v3_TRINITY_DN1525_c0_g1_i1	microsomal glutathione S-transferase 1
151_Ppyr_v3_TRINITY_DN1238_c0_g1_i1	MICOS complex subunit MIC13 homolog
151_Ppyr_v3_TRINITY_DN1766_c0_g1_i1	methylmalonate-semialdehyde dehydrogenase
151_Ppyr_v3_TRINITY_DN10178_c0_g1_i1	malate mitochondrial
151_Ppyr_v3_TRINITY_DN58181_c0_g1_i1	malate cytoplasmic
151_Ppyr_v3_TRINITY_DN15632_c0_g1_i2	long form
151_Ppyr_v3_TRINITY_DN10972_c0_g1_i1	leucine-rich repeat-containing G-coupled receptor 4
Conserved unknown proteins	
151_Ppyr_v3_TRINITY_DN13942_c0_g1_i1	uncharacterized protein LOC656585

Chapter 7. Summary and conclusion

The primary goal of this dissertation was to characterize male and female reproductive genes and proteins in three spermatophore-producing species: *Ostrinia nubilalis* moths, *Tribolium castaneum* flour beetles, and *Photinus pyralis* fireflies. The secondary goal was to characterize how male and female reproductive gene expression changes in response to two different ecological contexts: 1) reproductive isolation and 2) variation in sexual selection. Here, I review what I have learned about male and female reproductive proteins and how they could mediate postcopulatory processes important for both sexual selection and reproductive isolation.

Using *O. nubilalis* moths, we were able to identify reproductive proteins that could contribute to a previously demonstrated reproductive isolating barrier acting between divergent pheromone strains. This postmating, prezygotic (PMPZ) barrier causes an asymmetric reduction in egg-production when a Z strain female mates with an E strain male (Dopman et al., 2010). Many factors could mediate this reduction in egg laying, including failed interactions between FRPs and SFPs. In Chapters 2, 3, and 4, we aimed to identify FRPs and SFPs in *O. nubilalis* moths and identify candidates from these that could be involved in reproductive isolation. Chapter 2 reports investigations on the function of the *O. nubilalis* female bursa copulatrix and the bursal gland. Using RNA sequencing and differential expression analysis, we identified genes that were up-regulated and predicted to be secreted from these tissues. We found that the female bursa

copulatrix primarily acts as a muscular sac, as predicted by previous studies of the ultrastructure of this tissue in *Tortricidae* moths (Lincango et al., 2013), while the bursal gland had increased expression of genes encoding secreted and transmembrane proteins that might act as receptors for male SFPs. We also found that peptidases and peptidase regulators were differentially expressed between *O. nubilalis* strains in both the bursa copulatrix and bursal glands.

Along with FRPs, *O. nubilalis* males from separate strains may differ in the SFPs they transfer to females, which could lead to reproductive failures between divergent populations (Price et al., 2001; Larson et al., 2011). In Chapter 3, we characterized male SFPs using a combined RNA sequencing and proteomics approach that allowed us to identify proteins that male reproductive glands may be secreting, and to determine which of these are actually transferred to the female in the male spermatophore. We also used differential expression analyses to identify divergence in male reproductive gene expression between strains. We found that the ECB male spermatophore contained a number of peptidases, peptidase regulators, and odorant binding proteins that could be important for mediating male-female postcopulatory interactions (Leal, 2013; Laflamme & Wolfner, 2013). When we identified genes that were differentially expressed between strains, we found that numerous peptidases were more highly expressed in Z strain male accessory glands. Many genes with the highest level of differential expression between strains showed no homology to known proteins. These genes may represent rapidly evolving male reproductive proteins; these

could be candidates to be involved in postmating, prezygotic isolation in *O. nubilalis* moths.

Mating has been shown to cause changes in gene expression in the female reproductive tract for peptidase regulators, cytochrome P450s, immune-related genes, genes related to metabolism, along with other categories of genes (McGraw et al., 2004; Prokupek et al., 2008). One possible mechanism leading to reproductive isolation between strains of *O. nubilalis* moths may involve a failure by males to induce post-mating changes in female gene expression associated with the induction of female egg laying. In Chapter 4, we mated Z strain females to *O. nubilalis* males of either strain and used RNA sequencing and differential expression analysis to examine changes in gene expression in the bursa copulatrix and bursal gland. We found that Z strain females showed differences in gene expression in the bursal gland after mating within- and across-strain, including a number of genes of unknown function. Thus, the PMPZ reproductive barrier acting between *O. nubilalis* strains may be a result of failed interactions between novel male and female reproductive proteins.

As females mate more frequently, the intensity of postcopulatory sexual selection increases (Thornhill & Alcock, 2013; Choe & Crespi, 1997). In chapter 5, we aimed to determine how altering the intensity of postcopulatory sexual selection could influence gene expression in male and female reproductive tissues. To do this, we manipulated the mating system of *Tribolium castaneum* to enforce monogamy for 12 generations, thus reducing the intensity of sexual selection. We then used RNA sequencing to characterize gene expression in the

male accessory glands responsible for producing the male spermatophore and in female spermathecal glands, followed by differential expression analysis. In monandrous lines, *T. castaneum* males may shift gene expression in one pair of accessory glands to enable increased egg and/or sperm production. In females, the spermathecal gland of polyandrous females showed up-regulation of a peptidase that could potentially mediating cryptic female choice by activating or deactivating male SFPs.

In chapter 6, we used RNA sequencing, proteomics, and metabolomics approaches to identify the molecular composition of the spermatophore of *P. pyralis* fireflies. In *P. pyralis* fireflies male spermatophore size is correlated with reproductive success (South & Lewis, 2012) and proteins from the male spermatophore have been shown to be incorporated into female eggs (Rooney & Lewis, 1999). In spite of the apparent importance of the spermatophore for postcopulatory sexual selection in *Photinus* fireflies, its composition was completely unknown before we began our studies. Along with identifying the composition of the male spermatophore, we also aimed to provide insight into the function of particular female reproductive structures using RNA sequencing. Combined RNA sequencing and proteomics analyses revealed that *P. pyralis* males were transferring peptidases that may be important for sperm activation and for mediating the female immune response in their spermatophore. We also report that *P. pyralis* females secrete digestive peptidases from tissues in their reproductive tract, which may serve to digest the male spermatophore or activate/deactivate male SFPs. Finally, using metabolomics analyses we found

that male *P. pyralis* males may also transfer the defensive compound, lucibufagin, to females within their spermatophores.

By examining reproductive genes and proteins in three different spermatophore-producing taxa I have increased our knowledge regarding how this reproductive trait influences postcopulatory sexual interactions. Across all three taxa, peptidases and peptidase inhibitors are likely to mediate interactions between SFPs and FRPs. Specifically, across all three taxa we identified cathepsins, digestive proteases, whose function in reproduction is just beginning to be elucidated. I also found that control of the immune response after mating is an important aspect of male and female postcopulatory interactions. The unique characteristics of each study species represented in this thesis allowed me to characterize how reproductive proteins may be important for sexual selection and reproductive isolation.

There are many avenues for future research concerning male-female postcopulatory interactions and the role they play in sexual selection. Future research should focus on the function of peptidases and peptidase regulators in spermatophore producing species. Peptidases and regulators are a class of proteins found throughout insect tissues, but their function in male-female postcopulatory interactions for spermatophore producing species is completely unknown. Most importantly, comparative studies of male and female reproductive genes across taxa with varying life history and reproductive characteristics need to be conducted to determine how these traits affect male-female postcopulatory interactions. This will allow for a deeper understanding of how male and female

reproductive genes and proteins evolve in response to different types of selective pressures.

Literature cited

Adler D, Nenadic O, Zucchini W. (2003). Rgl: A r-library for 3d visualization with opengl. Interface.

Ah-King, M., Barron, A.B. & Herberstein, M.E. (2014). Genital evolution: Why are females still understudied? *PLoS biology*. **12** (5). pp. e1001851–8.

Al-Wathiqui, N., Lewis, S.M. & Dopman, E.B. (2014). Using RNA sequencing to characterize female reproductive genes between Z and E strains of European corn borer moths (*Ostrinia nubilalis*). *BMC Genomics*. **15** (1). pp. 1–13.

Alexandre, D., Linhares, R.T., Queiroz, B., Fontoura, L., Uchôa, A.F., Samuels, R.I., et al. (2011). Vicilin-derived peptides are transferred from males to females as seminal nuptial gift in the seed-feeding beetle *Callosobruchus maculatus*. *J Insect Physiol*. **57** (6). pp. 801–808.

Alumot, E., Lensky, Y. & Holstein, P. (1969). Sugars and trehalase in the reproductive organs and hemolymph of the queen and drone honey bees (*Apis mellifica* L. var. ligustica Spin.). *Com Biochem Physiol*. 28 (3). pp. 1419–1425.

Andersen, S.O., Hojrup, P. & Roepstorff, P. (1995). Insect cuticular proteins. *Insect Biochem Mol Biol.* **25** (2). pp. 153–176.

Andersson M. (1994). Sexual selection. New Jersey: Princeton University Press.

Andersson M, Simmons LW. (2006). Sexual selection and mate choice. *Trends Ecol Evol* **21**. pp.296–302.

Andres, J.A. (2006). Molecular evolution of seminal proteins in field crickets. *Mol Biol and Evol*. **23** (8). pp. 1574–1584.

Arnqvist, G. & Nilsson, T. (1999). The evolution of polyandry: Multiple mating and female fitness in insects. *Anim Behav*. **60** (2). pp. 145-164.

Arnqvist, G., Edvardsson, M, Friberg, U., & Nilsson, T.. (2000). Sexual conflict promotes speciation in insects. *Proc Nat Acad Sci*. **97** (19). pp. 10460–10464.

Arnqvist, G. & Rowe, L. (2005). Sexual conflict: Mongraphs in behavior and ecology. New Jersey: Princeton University Press.

Avila, F.W., Sirot, L.K., Laflamme, B.A., Rubinstein, C.D. & Wolfner, M.F. (2011). Insect seminal fluid proteins: Identification and function. *Annu Rev Entomol*. **56** (1). pp. 21–40.

Aydos, S.E., Yukselten, Y., Sunguroglu, A., Demircan, K. & Aydos, K. (2016).

Role of ADAMTS1 and ADAMTS5 in male infertility. *Andrologia*. Early view.

Ayme-Southgate, A., Lasko, P., French, C. & Pardue, M.L. (1989). Characterization of the gene for mp20: a *Drosophila* muscle protein that is not found in asynchronous oscillatory flight muscle. *J Cell Biol*. **108** (2). pp. 521–531.

Azevedo, R.V.D., Dias, D.B.S., Bretas, J.A.C., Mazzoni, C.J., Souza, N.A., Albano, R.M. *et al*. (2012) The transcriptome of *Lutzomyia longipalpis* (Diptera: Psychodidae) male reproductive organs. *PLoS ONE* **7**. pp. e34495–e34495.

Baer B, Eubel H, Taylor NL, O'Toole N, Millar AH. (2009). Insights into female sperm storage from the spermathecal fluid proteome of the honeybee *Apis mellifera*. *Genome Biol.* **10**. pp. R67.

Baer B, Armitage SAO, Boomsma JJ. (2006). Sperm storage induces an immunity cost in ants. *Nature.* **441**. pp. 872–875.

Baldini, F., Gabrieli, P., South, A., Valim, C., Mancini, F. & Catteruccia, F. (2013). The Interaction between a Sexually Transferred Steroid Hormone and a Female Protein Regulates Oogenesis in the Malaria Mosquito *Anopheles gambiae*. *PLoS biology*. **11** (10). pp. e1001695–11.

Bella JL, Butlin RK, Ferris C, Hewitt GM. (1992). Asymmetrical homogamy and unequal sex ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. *Heredity*. **68**. pp. 345–352.

Belvin, M.P. & Anderson, K.V. (1996). A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu Rev Cell Dev Biol*. **12** (1). pp. 393–416.

Benoist P. (1998). 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*. **273**. pp. 7538–7546.

Bernasconi, G. & Keller, L. (2001). Female polyandry affects their sons' reproductive success in the red flour beetle *Tribolium castaneum*. *J Evol Biol*. **14** (1). pp. 186–193.

Birkhead TM, Hosken DJ, Pitnick SS. (2009). Sperm biology: an evolutionary perspective. Pennsylvania: Elsevier Ltd.

Birkhead TR, Pizzari T. (2002). Evolution of sex: postcopulatory sexual selection. *Nat Rev Genet.* **3**. pp. 262–273.

Bloch Qazi, M.C.B., Herbeck, J.T. & Lewis, S.M. (1996). Mechanisms of Sperm Transfer and Storage in the Red Flour Beetle (Coleoptera: Tenebrionidae).

Ann Entomol Soc Am. **89** (6). pp. 892–897.

Bloch Qazi MC, Aprille JR, Lewis SM. (1998). Female role in sperm storage in the red flour beetle, **Tribolium castaneum**. *Comp Biochem Phys A*. **120**. pp. 641–647.

Bao, Y. (2003). cDNA cloning and expression of bacteria-induced Hdd11 gene from eri-silkworm, *Samia cynthia ricini*. *Comp Biochem Physiol C.* **136** (4). pp. 337–342.

Boes, K.E., Ribeiro, J.M.C., Wong, A., Harrington, L.C., Wolfner, M.F. & Sirot, L.K. (2014). Identification and characterization of seminal fluid proteins in the Asian Tiger Mosquito, *Aedes albopictus*. *PLoS Negl Trop Dis*. **8** (6). pp. e2946–14.

Boggs, C.L. & Gilbert, L.E. (1979). Male contribution to egg production in butterflies: evidence for transfer of nutrients at mating. *Science*. **206** (4414). pp. 83–84.

Boggs, C.L. (1990). A general model of the role of male-donated nutrients in female insects' reproduction. *American Naturalist*. pp. 598–617.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 doi:10.1093/bioinformatics/btu170.

Bonilla, M.L., Todd, C., Erlandson, M. & Andres, J. (2015). Combining RNA-seq and proteomic profiling to identify seminal fluid proteins in the migratory grasshopper *Melanoplus sanguinipes* (F). *BMC Genomics*. pp. 1–15.

Bono, J. M., Matzkin, L. M., Kelleher, E. S. and Markow, (2011). Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated *Drosophila mojavensis* females. *Proc Natl Acad Sci USA.* **108**. pp. 7878-7883.

Branham, M. (2003). The origin of photic behavior and the evolution of sexual communication in fireflies (Coleoptera: Lampyridae). *Cladistics*. **19** (1). pp. 1–22.

Braz, G.R., Abreu, L., Masuda, H. & Oliveira, P.L. (2001). Heme biosynthesis and oogenesis in the blood-sucking bug, *Rhodnius prolixus*. *Insect Biochem Mol Biol*. **31** (4-5). pp. 359–364.

Breden, F. & Chippendale, G.M. (1989). Effect of larval density and cannibalism on growth and development of the southwestern corn borer, *Diatraea grandiosella*, and the European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). *J Kans Entomol Soc*. **62** (3). pp. 307-315.

Breugelmans B, Simonet G, van Hoef V, Van Soest S, Vanden Broeck J. (2009). Pacifastin-related peptides: Structural and functional characteristics of a family of serine peptidase inhibitors. *Peptides*. **30**. pp. 622–632.

Buck, J.B. (1948). The anatomy and physiology of the light organ in fireflies. *Ann N Y Acad Sci*. **49** (3). pp. 397–485.

Cardoso, M.Z. & Gilbert, L.E. (2006). A male gift to its partner? Cyanogenic glycosides in the spermatophore of longwing butterflies (Heliconius). *Naturwissenschaften*. **94** (1). pp. 39–42.

Carlsson, A., Nyström, T., de Cock, H. & Bennich, H. (1998). Attacin--an insect immune protein--binds LPS and triggers the specific inhibition of bacterial outer-membrane protein synthesis. *Microbiology*. **144** (8). pp. 2179–2188.

Carpentier, M., Guillemette, C., Bailey, J.L., Boileau, G., Jeannotte, L., DesGroseillers, L. & Charron, J. (2004). Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. *Mol Cell Biol*. **24** (10). pp. 4428–4437.

Castrillon DH, Gonczy P, Alexander S, Rawson R, Eberhart CG, Viswanathan S, DiNardo S, Wasserman SA. (1993). Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster:* characterization of male-sterile mutants generated by single P element mutagenesis. *Genetics*. **135**. pp. 489.

Chaudhari, S.S., Arakane, Y. & Specht, C.A. (2011). Knickkopf protein protects and organizes chitin in the newly synthesized insect exoskeleton. *Proc Nat Acad Sci USA*. **108** (41). pp. 17028-17033.

Chapman, R.F. (2003). *The insects: structure and function*. Cambridge: Cambridge University Press.

Chapman, T., Liddle, L.F., Kalb, J.M., Wolfner, M.F. & Partridge, L. (1995). Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature*. **373** (6511). pp. 241–244.

Chapman, T., Takahisa, M., Smith, H.K. & Partridge, L. (1998). Interactions of mating, egg production and death rates in females of the Mediterranean fruitfly, *Ceratitis capitata*. *Proc R Soc B*. **265** (1408). pp. 1879–1894.

Chapman T. (2008). The Soup in My Fly: Evolution, Form and Function of Seminal Fluid Proteins. *Plos Biol.* **6**. e179.

Chapman, T., Arnqvist, G., Bangham, J. & Rowe, L. (2003). Sexual conflict. *TREE*. **18** (1). pp. 41–47.

Chen, P.S. (1984). The Functional-Morphology and Biochemistry of Insect Male Accessory-Glands and Their Secretions. *Annual Rev Entomol*. **29**. pp. 233–

255.

Choe, J.C. & Crespi, B.J. (1997). *The Evolution of Mating Systems in Insects and Arachnids*. Cambridge University Press.

Civetta, A., Braswell, W.E., Andrés, J.A., Maroja, L.S., Harrison, R.G., Howard, D.J. & Swanson, W.J. (2006). Identification and comparative analysis of accessory gland proteins in Orthoptera. *Genome*. **49** (9). pp. 1069–1080.

Clark, N.L. (2006). Evolution of reproductive proteins from animals and plants. *Reproduction*. **131** (1). pp. 11–22.

Collins, A.M., Caperna, T.J., Williams, V., Garrett, W.M. & Evans, J.D. (2006). Proteomic analyses of male contributions to honey bee sperm storage and mating. *Insect Mol Biol*. **15** (5). pp. 541–549.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21** (18). pp. 3674–3676.

Conesa, A. & Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics*. pp. 1–12.

Cordero, C. (2005). The evolutionary origin of signa in female Lepidoptera: natural and sexual selection hypotheses. *J Theor Biol*. **232** (3). pp. 443–449.

Cordero, C. (2009). On the function of cornuti, sclerotized structures of the endophallus of Lepidoptera. *Genetica*. **138** (1). pp. 27–35.

Coyne JA, Orr HA. (2004). Speciation. Massachusetts: Sinauer Associates, Inc.

Cratsley CK, Rooney JA, Lewis SM. (2003). Limits to Nuptial Gift Production by Male Fireflies, *Photinus ignitus*. *J Insect Behav*. **16**. pp. 361–370.

Cruz, J., Mane-Padros, D., Zou, Z. & Raikhel, A.S. (2012). Distinct roles of isoforms of the heme-liganded nuclear receptor E75, an insect ortholog of the vertebrate Rev-erb, in mosquito reproduction. *Mol Cell Endocrinol*. **349** (2). pp. 262–271.

Cusson, M. & McNeil, J.N. (1989). Involvement of juvenile hormone in the regulation of pheromone release activities in a moth. *Science.* **243** (4888). pp. 210–212.

Darwin, C. (1959). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. London: J. Murray

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. Philadelphia: J.Wanamaker.

Davies, S.J. & Chapman, T. (2006). Identification of genes expressed in the accessory glands of male Mediterranean fruit flies (Ceratitis capitata). *Insect Biochem Mol Biol.* **36** (11). pp. 846–856.

Deitsch, K. W., Chen, J.-S., and Raikhel, A. S. (1995a). Indirect control of yolk protein genes by 20-hydrocyecdysone in the fat body of the mosquito, *Aedes aegypti. Insect Biochem Mol Biol.* **25,** 449–454.

Deitsch, K. W., Dittmer, N., Kapitskaya, M. Z., Chen, J.-S., Cho, W.-L., and Raikhel, A. S. (1995b). Regulation of gene expression by 20-hydroxyecdysone in the fat body of *Aedes aegypti* (Diptera: Culicidae). *Eur. J. Entomol.* **92,** 237–244.

Demary, K.C. (2005). Sperm storage and viability in *Photinus* fireflies. *J Insect Physiol.* **51** (7). pp. 837–841.

Denell, R., Gibbs, R., Beeman, R.W., et al. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*. **452** (7190). pp. 949–955.

Dopman, E.B., Pérez, L., Bogdanowicz, S.M. & Harrison, R.G. (2005). Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proc Acad Nat Sci.* **102** (41). pp. 14706–14711.

Dopman, E.B., Robbins, P.S. & Seaman, A. (2010). Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution.* **64** (4). pp. 881–902.

Dottorini T, Nicolaides L, Ranson H, Rogers DW, Crisanti A, Catteruccia F. (2007). A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. *Proc Natl Acad Sci.* **104**. pp. 16215–16220.

Drecktrah HG, Brindley TA. (1967). Morphology of the internal reproductive systems of the European corn borer. *Iowa State J Sci.* **41**. pp. 467–480.

Eberhard WG, Cordero C. (1995). Sexual Selection by Cryptic Female Choice on Male Seminal Products - a New Bridge Between Sexual Selection and Reproductive Physiology. *TREE*. **10**. pp. 493–496.

Eberhard, W.G. (1996). *Female control: sexual selection by female cryptic choice*. Monographs in Behavior and Ecology. New Jersey: Princeton University Press.

Eisner, T., Wiemer, D.F., Haynes, L.W. & Meinwald, J. (1978). Lucibufagins: Defensive steroids from the fireflies *Photinus ignitu*s and *P. marginellus* (Coleoptera: Lampyridae). *Proc Nat Acad Sci.* **75** (2). pp. 905–908.

Eisner, T. & Meinwald, J. (1995). The chemistry of sexual selection. *Proc Nat*

Acad Sci. **92** (1). pp. 50–55.

Ekbote UV, Weaver RJ, Isaac RE. Angiotensin I-converting enzyme (ACE) activity of the tomato moth, *Lacanobia oleracea*: changes in levels of activity during development and after copulation suggest roles during metamorphosis and reproduction. *Insect Biochem Mol Biol* **33**. pp. 989–998 doi:10.1016/S0965-1748(03)00105-X (2003).

Ellegren, H. & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*. **8** (9). pp. 689–698.

Ellison, R. T. (1991). Killing of gram-negative bacteria by lactoferrin and lysozyme. *J Clinic Invest.* **88** (4). pp. 1080–1091.

Espey, L.L. & Richards, J.S. (2002). Temporal and spatial patterns of ovarian gene transcription following an ovulatory dose of gonadotropin in the rat. *Biol Repro.* **67** (6). pp. 1662–1670.

Fadamiro, H.Y. & Baker, T.C. (1999). Reproductive performance and longevity of female European corn borer, *Ostrinia nubilalis*: effects of multiple mating, delay in mating, and adult feeding. *J Insect Physiol*. **45** (4). pp. 385–392.

Fedina, T.Y. & Lewis, S.M. (2004). Female influence over offspring paternity in the red flour beetle Tribolium castaneum. *Proc B*. **271** (1546). pp. 1393–1399.

Fedina, T.Y. & Lewis, S.M. (2006). Proximal traits and mechanisms for biasing paternity in the red flour beetle *Tribolium castaneum* (Coleoptera : Tenebrionidae). *Behav Ecol Socio*. **60** (6). pp. 844–853.

Fedina, T.Y. (2007). Cryptic female choice during spermatophore transfer in *Tribolium castaneum* (Coleoptera: Tenebrionidae). *J Insect Physiol*. **53** (1). pp. 93–98.

Fedina, T.Y. & Lewis, S.M. (2008). An integrative view of sexual selection in *Tribolium* flour beetles. *Biol Rev*. **83** (2). pp. 151–171.

Fedorka, K.M., Winterhalter, W.E. & Ware, B. (2011). Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. *Evolution*. 65 (2). pp. 584–590.

Feng C, Chen M, Xu C-J, Bai L, Yin X-R, Li X, Allan AC, Ferguson IB, Chen K-S. (2012). Transcriptomic analysis of Chinese bayberry (Myrica rubra) fruit development and ripening using RNA-Seq. *BMC Genomics.* **13**. pp. 19.

Feyereisen, R. (1999). Insect P450 enzymes. *Ann Rev Entomol*. **44** (1). pp. 507–533.

Findlay, G.D., Yi, X., Maccoss, M.J. & Swanson, W.J. (2008). Proteomics reveals

novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol*,. **6.** pp. 178.

Findlay GD, MacCoss MJ, Swanson WJ. (2009). Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res.* **19**, 886–896 doi:10.1101/gr.089391.108.

Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. (2014). Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. *PLoS Genet.* **10**. pp. e1004108–16 doi:10.1371/journal.pgen.1004108 (2014).

Friedlander, M., Jeshtadi, A. & Reynolds, S.E. (2012). The structural mechanism of trypsin-induced intrinsic motility in *Manduca sexta* spermatozoa in vitro. *J Insect Physiol*. **47** (3). pp. 245–255.

Fung, K.Y.C., Glode, L.M., Green, S. & Duncan, M.W. (2004). A comprehensive characterization of the peptide and protein constituents of human seminal fluid. *The Prostate*. **61** (2). pp. 171–181.

Galicia, I, Sanchez V. & Cordero, C. (2008) On the function of signa, a genital trait of female Lepidotera. *Ann Entomol Soc Am.* **101**. pp. 786–793.

Garland, T. & Rose, M.R. (2009). *Experimental Evolution*. California: University of California Press.

Gaspari Z, Ortutay C, Perczel A. (2004). A simple fold with variations: the pacifastin inhibitor family. *Bioinformatics.* **20**. pp. 448–451.

Gavrilets S, Arnqvist G, Friberg U. (2001). The evolution of female mate choice by sexual conflict. *Proc B.* **268**. pp. 531–539.

Gavrilets, S. & Hayashi, T.I. (2005). Speciation and sexual conflict. *Evol Ecol.* **19** (2). pp. 167–198.

Gavrilets, S. (2014). Is sexual conflict an "Engine of Speciation"? *Cold Spring Harbor Perspect Biol.* **6** (12). pp. a017723–a017723.

Gerrard, D.T., Fricke, C., Edward, D.A., Edwards, D.R. & Chapman, T. (2013). Genome-Wide Responses of Female Fruit Flies Subjected to Divergent Mating Regimes W. J. Etges (ed.). *PLoS ONE*. **8** (6). p. e68136.

Gillott, C. (2003). Male accessory gland secretions: Modulators of female reproductive physiology and behavior. *Annu Rev Entomol.* **48** (1). pp. 163–184.

Goetz, M.A., Meinwald, J. & Eisner, T. (1981). Lucibufagins, IV. New defensive

steroids and a pterin from the firefly,*Photinus pyralis* (coleoptera: Lampyridae). *Experientia.* **37** (7). pp. 679–680.

González-Santoyo, I. & Córdoba- Aguilar, A. (2011). Phenoloxidase: a key component of the insect immune system. *Entomol Exp Appl.* **142** (1). pp. 1–16.

Goodman, W.G., Park, Y.C. & Johnson, J.A. (1990). Development and partial characterization of monoclonal antibodies to the hemolymph juvenile hormone binding protein of *Manduca sexta*. *Insect Biochem*. **20** (6). pp. 611–618.

Gorman, M.J. & Paskewitz, S.M. (2001). Serine proteases as mediators of mosquito immune responses. *Insect Biochem Mol Biol*. **31.** pp. 257-262.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J. & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. **36** (10). pp. 3420–3435.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol*. **29** (7). pp. 644–652.

Green, J.W. (1956). *Revision of the Nearctic Species of Photinus:(Lmpyridae: Coleoptera)*. *Proc Calif Acad Sci*. **28**. pp. 561-613.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. **8** (8). pp. 1494–1512.

Happ, G.M. (1992). Maturation of the male reproductive system and its endocrine regulation. *Ann Rev Entomol*. **37** (1). pp. 303–320.

Harrison, R.G. (2012) The language of speciation. *Evolution*. 66. pp. 3643–3657.

Harshman, L.G. & Zera, A.J. (2007). The cost of reproduction: the devil in the details. *TREE*. **22** (2). pp. 80–86.

Hartfelder, K. (2000). Insect juvenile hormone: from 'status quo' to high society. *Braz J Med Biol Res.* **33** (2). pp. 157–177.

Heifetz, Y., Lung, O., Frongillo, E.A. & Wolfner, M.F. (2000). The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary.

CURBIO. **10** (2). pp. 99–102.

Heifetz, Y., Tram, U. & Wolfner, M.F. (2001). Male contributions to egg production: the role of accessory gland products and sperm in *Drosophila melanogaster*. *Proc R Soc Lond [Biol]*. **268** (1463). pp. 175–180.

Heifetz, Y. & Wolfner, M.F. (2004). Mating, seminal fluid components, and sperm cause changes in vesicle release in the *Drosophila* female reproductive tract. *PNAS*. **101** (16). p. 6261.

Hekmat-Scafe, et al. (2002). Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster. Genome Research*. 12 (9). pp. 1357–1369.

Hentze, J.L., Moeller, M.E., Jørgensen, A.F., Bengtsson, M.S., Bordoy, A.M., Warren, J.T., et al. (2013). Accessory Gland as a Site for Prothoracicotropic Hormone Controlled Ecdysone Synthesis in Adult Male Insects. *PLoS ONE*. **8** (2). pp. e55131–10.

Herndon, L.A. & Wolfner, M.F. (1995). A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc Nat Acad Sci*. **92** (22). pp. 10114–10118.

Hewitt GM, Mason P, Nichols RA. (1989). Sperm precedence and homogamy across a hybrid zone in the alpine grasshopper *Podisma pedestris*. *Heredity.* **62**. pp. 343–353.

Hoffmann, J.A., Reichhart, J.M. & Hetru, C. (1996). Innate immunity in higher insects. *Curr Opin Immunol.* **8** (1). pp. 8–13.

Howard DJ, Gregory PG, Chu J, Cain ML. (1998). Conspecific sperm precedence is an effective barrier to hybridization between closely related species. *Evolution*. pp. 511–516.

Howard DJ. (1999). Conspecific sperm and pollen precedence and speciation. Annu Rev Ecol Syst. pp. 109–132.

Ikeda, M., et al. (1990). Purification and characterization of proteases responsible for vitellin degradation of the silkworm, *Bombyx mori*. *Insect Biochem.* **20** (7). pp. 725–734.

Immonen, E., Snook, R.R. & Ritchie, M.G. (2014). Mating system variation drives rapid evolution of the female transcriptome in Drosophila pseudoobscura. *Ecol Evol*. **4** (11). pp. 2186-2201.

Iseli C, Jongeneel CV, Bucher P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc 8th International Conf Intel Sys Mol Biol.* **7**. pp. 138–148.

Ishidoh, K. & Kominami, E. (1995). Procathepsin L degrades extracellular matrix proteins in the presence of glycosaminoglycans in vitro. *Biochem Biophys Res Commun.* **217** (2). pp. 624–631.

Jennions MD, Petrie M. (2012). Why do females mate multiply? A review of the genetic benefits. Biol Rev. **75**. pp. 21–64.

Kanamori, Y., Saito, A., Hagiwara-Komoda, Y., Tanaka, D., Mitsumasu, K., Kikuta, S., et al. (2010) The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involvedin trehalose import into peripheral tissues. *Insect Biochem Mol Biol*. **40** (1), pp. 30–7.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I. & Whitlock, M.C. (2012). Experimental evolution. *Trends in Ecology & Evolution*. **27** (10). pp. 547–560.

Kelleher ES, Swanson WJ, Markow TA. (2007). Gene Duplication and Adaptive Evolution of Digestive Proteases in *Drosophila arizonae* Female Reproductive Tracts. *PLoS Genet*. **3**. pp. e148.

Kelleher ES, Markow TA. (2009). Duplication, Selection and Gene Conversion in a *Drosophila mojavensis* Female Reproductive Protein Family. *Genetics.* **181**. pp. 1451–1465.

Kelleher ES, Clark NL, Markow TA. (2011). Diversity-Enhancing Selection Acts on a Female Reproductive Protease Family in Four Subspecies of *Drosophila mojavensis*. *Genetics.* **187**. pp. 865–876.

Kim, E., et al. (2008). Sperm penetration through cumulus mass and zona pellucida. *Internat J Dev Biol*. **52** (5-6). pp. 677–682.

Kim, H.S., Murphy, T., Xia, J., Caragea, D., Park, Y., Beeman, R.W., Lorenzen, M.D., Butcher, S., Manak, J.R. & Brown, S.J. (2009). BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. *Nucleic Acids Research*. **38**. pp. D437–D442.

Kinsella, R.J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*. pp. bar030–bar030.

Kirkpatrick M, Ravigné V. (2002). Speciation by Natural and Sexual Selection: Models and Experiments. *Am Nat.* **159**. pp. S22–S35.

Klowden, M.J. & Chambers, G.M. (1991). Male accessory gland substances activate egg development in nutritionally stressed *Aedes aegypti* mosquitoes. *J Insect Physiol*. **37** (10). pp. 721–726.

Kristensen NP. (2003). *Lepidoptera, moths and butterflies*. In *Handbook of Zoology. Volume 4*. Edited by Maximilian Fischer: New York: Walter de Gruyter, Inc.

Kubli E. (2003). Sex-peptides: seminal peptides of the *Drosophila* male. *Cell Mol Life Sci.* **60**. pp. 1689–1704.

Kvarnemo, C. & Simmons, L.W. (2013). Polyandry as a mediator of sexual selection before and after mating. *Proc B*. **368** (1613). pp. 20120042–20120042.

Laflamme, B.A. & Wolfner, M.F. (2013). Identification and function of proteolysis regulators in seminal fluid. *Mol Reprod Dev.* **80** (2). pp. 80–101.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10** (3). p. R25.

Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*. **9** (4). pp. 357–359.

Larson, E.L., Hume, G.L., Andrés, J.A. & Harrison, R.G. (2011). Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. *J Evol Biol*. **25** (1). pp. 174–186.

Lawniczak MK, Begun DJ. (2004). A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. *Genome.* **47**. pp. 900–910.

Lawniczak MKN, Begun DJ. (2007). Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster. Mol Biol Evol*. **24**. pp. 1944–1951.

Leal, W.S. (2013). Odorant reception in insects: Roles of receptors, binding proteins, and degrading enzymes. *Ann Rev Entomol*. **58** (1). pp. 373–391.

Leopold, R.A. (1976). The role of male accessory glands in insect reproduction. *Ann Rev Entomol*. **21.** pp. 199-221.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J.A. & Reichhart, J.-M. (1999). Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science*. **285** (5435). pp. 1917–1919.

Lewis, S.M. & Wang, O.T. (1991). Reproductive ecology of two species of *Photinus* fireflies (Coleoptera: Lampyridae). *Psyche*. **98** (4). pp. 293-307.

Lewis, S.M., Cratsley, C.K. & Rooney, J.A. (2004). Nuptial gifts and sexual

selection in *Photinus* Fireflies. *American Zoologist*. **44** (3). pp. 234–237.

Lewis, S.M. & Cratsley, C.K. (2008). Flash signal evolution, mate choice, and predation in Fireflies. *Annual Review of Entomology*. **53** (1). pp. 293–321.

Lewis, S.M. (2009). *Bioluminescence and sexual signaling in fireflies. Pp. 147–159 in V. B. Meyer-Rochow, ed. Bioluminescence in action: a collection of illuminating essays. Research Signpost, Kerala, India.*

Lewis, S. & South, A. (2012). The Evolution of Animal Nuptial Gifts. *Advances Stud Behav*. **44**. pp. 53-97.

Lewis, S.M., Vahed, K., Koene, J.M., Engqvist, L., Bussiere, L.F., Perry, J.C., Gwynne, D. & Lehmann, G.U.C. (2014). Emerging issues in the evolution of animal nuptial gifts. *Biol Lett*. **10** (7). pp. 20140336–20140336.

Li B, Dewey CN. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* **12**, 323 doi:10.1186/1471-2105-12-323.

Lincango P, Fernández G, Baixeras J. (2013). Microstructure and diversity of the bursa copulatrix wall in Tortricidae (Lepidoptera). *Arth Struct Dev*. **42**. pp. 247–256.

Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC Research Notes*. **5** (1). p. 337.

Lloyd, J.E. (1966). *Studies on the flash communication system in Photinus fireflies*.University of Michigan: Miscellaneous Publications. **130**. pp. 1-90.

Lloyd, J.E. (2008). Fireflies (Coleopltera: Lampyridae). In: Capinera JL. Encyclopedia of Entomology. pp. 1429-52 .

Love, M.I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15** (12). p. 550.

Mack, P.D., Kapelnikov, A., Heifetz, Y. and Bender, M. (2006) Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **103**. pp. 10358–10363.

Makalowski, W. & Boguski, M.S. (1998). Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Nat Acad Sci USA*. **95** (16). pp. 9407–9412.

Malausa T, Dalecky A, Ponsard S, Audiot P, Streiff R, Chaval Y, Bourguet D:(2007). Genetic structure and gene flow in French populations of two *Ostrinia* taxa: host races or sibling species? *Mol Ecol.* **16**. pp. 4210–4222.

Manier MK, Lüpold S, Belote JM, Starmer WT, Berben KS, Ala-Honkola O, Collins WF, Pitnick S. (2013). Postcopulatory Sexual Selection Generates Speciation Phenotypes in *Drosophila*. *CURBIO.* **23**. pp. 1853–1862.

Mank, J.E., Wedell, N. & Hosken, D.J. (2013). Polyandry and sex-specific gene expression. *Proc B*. **368** (1613). pp. 20120047–20120047.

Mann, T. (1984). *Spermatophores*. Springer Science & Business Media.

Marshall, J.L., Huestis, D.L., Hiromasa, Y., Wheeler, S., Oppert, C., Marshall, S.A., Tomich, J.M. & Oppert, B. (2009). Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a Cricket .*PLoS ONE*. **4** (10). pp. e7537.

Mazumdar-Leighton, S. & Broadway, R.M. (2001). Identification of six chymotrypsin cDNAs from larval midguts of *Helicoverpa zea* and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. *Insect Biochem Mol Biol*. **31** (6-7). pp. 633–644.

McCarthy, D.J., Chen, Y. & Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. **40** (10). pp. 4288–4297.

McElroy, W.D. (1947). The energy source for bioluminescence in an isolated system. *Proc Nat Acad Sci USA*. **33** (11). pp. 342–345.

McGraw, L.A., Gibson, G., Clark, A.G. & Wolfner, M.F. (2004). Genes Regulated by Mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Curr Biol*. **14** (16). pp. 1509–1514.

McGraw LA, Clark AG, Wolfner MF. (2008). Post-mating Gene Expression Profiles of Female *Drosophila melanogaster* in Response to Time and to Four Male Accessory Gland Proteins. *Genetics.* **179**. pp. 1395–1408.

McNamara KB, Elgar MA, Jones TM. (2009). Large spermatophores reduce female receptivity and increase male paternity success in the almond moth, Cadra cautella. *Anim Behav.* **77**. pp. 931–936.

Meinwald, J., Wiemer, D.F. & Eisner, T. (1979). Lucibufagins. 2. Esters of 12-oxo-2. beta., 5. beta., 11. alpha.-trihydroxybufalin, the major defensive steroids of the firefly Photinus pyralis (Coleoptera: Lampyridae). *J Am Chem Soc.* **101** (11). pp. 3055–3060.

Meslin, C., Plakke, M.S., Deutsch, A.B., Small, B.S., Morehouse, N.I. & Clark, N.L. (2015). Digestive organ in the female reproductive tract borrows genes from multiple Organ Systems to Adopt Critical Functions. *Mol Biol Evol*. **32** (6). pp. 1567–1580.

Metz, E.C., et al. (1998). Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proc Nat Acad Sci USA*. **95** (18). pp. 10676–10681.

Michalczyk, Ł., Millard, A.L., Martin, O.Y., Lumley, A.J., Emerson, B.C. & Gage, M.J.G. (2010). Experimental evolution exposes female and male responses to sexual selection and conflict in *Tribolium castaneum*. *Evolution*. **65** (3). pp. 713–724.

Mueller, J.L., Ripoll, D.R., Aquadro, C.F. & Wolfner, M.F. (2004). Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. *Proc Nat Acad Sci USA*. **101** (37). pp. 13542–13547.

Nagaoka, S., Kato, K., Takata, Y. & Kamei, K. (2012). Identification of the sperm-activating factor initiatorin, a prostatic endopeptidase of the silkworm, *Bombyx mori. Insect Biochem Mol Biol*. pp. 1–12.

Narayana, J.L. & Chen, J.-Y. (2015). Antimicrobial peptides: Possible anti-infective agents. *Peptides*. **72** (C). pp. 88–94.

Nespolo, R.F., Roff, D.A. & Fairbairn, D.J. (2008). Energetic trade-off between maintenance costs and flight capacity in the sand cricket (*Gryllus firmus*). *Functional Ecology*. **22** (4). pp. 624–631.

Nikbakhtzadeh, M.R., Dettner, K., Boland, W., Gäde, G. & Dötterl, S. (2007). Intraspecific transfer of cantharidin within selected members of the family Meloidae (Insecta: Coleoptera). *J Insect Physiol*. **53** (9). pp. 890–899.

Novaczewski, M. & Grimnes, K.A. (1996). *Histological characterization of the reproductive accessory gland complex of Tribolium anaphe (Coleoptera: Tenebrionidae)*. Tribolium Information Bulletin.

Oberhauser, K. S. (1989). Effects of spermatophores on male and female monarch butterfly reproductive success. *Behav Ecol Sociobiol.* **25** 237–246.

Panhuis TM, Butlin R, Zuk M, Tregenza T. (2001). Sexual selection and speciation. *TREE*. **16**. pp. 364–371.

Panhuis, T.M. (2006). Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. *Genetics*. **173** (4). pp. 2039–2047.

Park, M. & Wolfner, M.F. (1995). Male and female cooperate in the prohormone-like processing of a *Drosophila melanogaster* seminal fluid protein. *Dev Biol*. **171** (2). pp. 694–702.

Park, T. (1933). Studies In Population Physiology. II. Factors regulating initial growth of *Tribolium confusum* populations. *Journal of Experimental Zoology*.

65 (1). pp. 17–42.

Parker GA. (2006). Sexual conflict over mating and fertilization: an overview. *Proc B*. **361**. pp. 235–259.

Parker GA, Partridge L. (1998). Sexual conflict and speciation. *Proc B.* **353**. pp. 261.

Parker, G.A. (1970). Sperm competition and its evolutionary consequences in insects. *Biol Rev Camb Philoso Soc*. **45** (4). pp. 525.

Parthasarathy, R., Tan, A., Sun, Z., Chen, Z., Rankin, M. & Palli, S.R. (2009). Juvenile hormone regulation of male accessory gland activity in the red flour beetle, *Tribolium castaneum. Mech Dev*. **126** (7). pp. 563–579.

Pearson WR, Wood T, Zhang Z, Miller W. (1997). Comparison of DNA sequences with protein sequences. *Genomics*. **46**. pp. 24–36.

Pedra, J.H.F., Brandt, A., Westerman, R., Lobo, N., Li, H.-M., Romero-Severson, J., Murdock, L.L. & Pittendrigh, B.R. (2003). Transcriptome analysis of the cowpea weevil bruchid: identification of putative proteinases and alpha-amylases associated with food breakdown. *Insect Mol Biol*. **12** (4). pp. 405–412.

Peng, J., Zipperlen, P. & Kubli, E. (2005) *Drosophila* sex-peptide stimulates female innate immune system after mating via the Toll and Imd pathways. *Curr Biol.* **15**. pp. 1690–1694.

Peretti AV, Aisenberg A. (2015). Cryptic Female Choice in Arthropods. Switzerland: Springer International Publishing.

Perry, J.C., Sirot, L. & Wigby, S. (2013). The seminal symphony: how to compose an ejaculate. *TREE*. pp. 1–9.

Petersen, T.N., Brunak, S., Heijne, von, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. **8** (10). pp. 785–786.

Pierceall, W.E., Li, C., Biran, A., Miura, K., Raikhel, A.S. & Segraves, W.A. (1999). E75 expression in mosquito ovary and fat body suggests reiterative use of ecdysone-regulated hierarchies in development and reproduction. *Mol Cell Endocrinol*. **150** (1-2). pp. 73–89.

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

Pitts, R.J., Liu, C., Zhou, X., Malpartida, J.C. & Zwiebel, L.J. (2014). Odorant receptor-mediated sperm activation in disease vector mosquitoes. *Proc Nat*

Acad Sci USA. **111** (7). pp. 2566–2571.

Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*. **11** (1). pp. 395.

Poiani, A. (2006). Complexity of seminal fluid: a review. *Behav Ecol Sociobiol*. **60** (3). pp. 289–310.

Polgar, L. (1989). *Mechanisms of Protease Action*. CRC Press.

Price C. (1997). Conspecific sperm precedence in *Drosophila*. *Nature.* **388**. pp. 663–666.

Price, C.S., Kim, C.H., Gronlund, C.J. & Coyne, J.A. (2001). Cryptic reproductive isolation in the *Drosophila simulans* species complex. *Evolution*. **55** (1). pp. 81–92.

Prokupek, A., Hoffmann, F., Eyun, S.-I., Moriyama, E., Zhou, M. & Harshman, L. (2008). An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution*. **62** (11). pp. 2936–2947.

Prokupek AM, Kachman SD, Ladunga I, Harshman LG. (2009). Transcriptional profiling of the sperm storage organs of *Drosophila melanogaster*. *Insect Mol Biol*. **18**. pp. 465–475.

Prokupek, A.M., Eyun, S.-I., Ko, L., Moriyama, E.N. & Harshman, L.G. (2010). Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of *Drosophila melanogaster*. *J Evol Biol*. **23** (7). pp. 1386–1398.

Pruitt, K.D. (2004). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*. **33**. pp. D501–D504.

Ram, K. & Wolfner, M.F. (2007). Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integ Comp Biol*. **47** (3). pp. 427–445.

Ram, K.R., Sirot, L.K. & Wolfner, M.F. (2006). Predicted seminal astacin-like protease is required for processing of reproductive proteins in *Drosophila melanogaster*. *PNAS*. **103** (49). pp. 18674–18679.

Ram, K., Ji, S. & Wolfner, M.F. (2005). Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. *Insect Biochem Mol Biol*. **35** (9). pp. 1059–1071.

van der Reijden, E.D. & Monchamp, J.D. (1997). The formation, transfer, and

fate of spermatophores in *Photinus* fireflies (Coleoptera: Lampyridae). *Can j Zool.* **75** (8). pp. 1202–1207.

Regier, J.C., Zwick, A., Cummings, M.P., Kawahara, A.Y., Cho, S., Weller, S., et al. (2009). Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evol Biol*. **9** pp. 280

Reinhardt K, Wong CH, Georgiou AS. (2008). Detection of seminal fluid proteins in the bed bug, *Cimex lectularius,* using two-dimensional gel electrophoresis and mass spectrometry. *Parasitology*. **136**. pp. 283.

Reuter, M., Linklater, J.R., Lehmann, L., Fowler, K., Chapman, T. & Hurst, G.D.D. (2008). Adaptation to experimental alterations of the operational sex ratio in populations of *Drosophila melanogaster*. *Evolution*. 62 (2). pp. 401–412.

Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics*. **12** (1). pp. 480.

Ritchie MG. (2007). Sexual Selection and Speciation. *Annu Rev Ecol Evol Syst*. **38**. pp. 79–102.

Robinson, M.D. & Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*. **23** (21). pp. 2881–2887.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26** (1). pp. 139–140.

Robinson, M.D. & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. **11** (3). p. R25.

Robker, R.L., Russell, D.L., Espey, L.L., Lydon, J.P., O'Malley, B.W. & Richards, J.S. (2000). Progesterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases. *PNAS*. **97** (9). pp. 4689–4694.

Roff, D.A. (1986). The evolution of wing dimorphism in insects. *Evolution*. **40** (5). pp. 1009.

Roff, D.A. (1990). The evolution of flightlessness in insects. *Ecol Monogr*. **60**. pp. 1009-1020.

Roff, D.A. & Fairbairn, D.J. (1991). Wing dimorphisms and the evolution of migratory polymorphisms among the Insecta. *Am Zool*. **31**. pp. 243-251.

Rogers, D.W., Whitten, M.M., Thailayil, J., Soichot, J., Levashina, E.A. &

Catteruccia, F. (2008) Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *Proc Natl Acad Sci USA* **105**: 19390–19395.

Rooney, J. & Lewis, S.M. (1999). Differential allocation of male-derived nutrients in two lampyrid beetles with contrasting life-history characteristics. *Behav Ecol*. **10** (1). pp. 97–104.

Rooney, J. & Lewis, S.M. (2002). Fitness advantage from nuptial gifts in female fireflies. *Trans R Entomol Soc Lond*. **27** (3). pp. 373–377.

Royer L, McNeil JN. (1991). Changes in calling behaviour and mating success in the European corn borer (*Ostrinia nubilalis*), caused by relative humidity. *Entomol Exp Appl*. **61**. pp. 131–138.

Sagisaka, A., Miyanoshita, A., Ishibashi, J. & Yamakawa, M. (2001). Purification, characterization and gene expression of a glycine and proline-rich antibacterial protein family from larvae of a beetle, *Allomyrina dichotoma*. *Insect Mol Biol*. **10** (4). pp. 293–302.

Sánchez, V., Hernández-Baños, B.E. & Cordero, C. (2011). The evolution of a female genital trait widely distributed in the Lepidoptera: Comparative evidence for an effect of sexual coevolution. *PLoS ONE*. **6** (8). pp. e22642.

Schittek, B., et al. (2001). Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nature Immunology*. **2** (12). pp. 1133–1137.

Scott, J.G. & Wen, Z. (2001). Cytochromes P450 of insects: the tip of the iceberg. *Pest Manag Sci*. **57** (10). pp. 958–967.

Scott, J.G., Liu, N. & Wen, Z. (1998). Insect cytochromes P450: diversity, insecticide resistance and tolerance to plant toxins. *Com Biochem Physiol C*. *121* (1-3). pp. 147–155.

Scriber JM, Tsubaki Y, Lederhouse RC. (1995). Swallowtail Butterflies: Their Ecology and Evolutionary Biology. Scientific Publishers, Inc.

Seth, R.K., Kaur, J.J., Rao, D.K. & Reynolds, S.E. (2004). Effects of larval exposure to sublethal concentrations of the ecdysteroid agonists RH-5849 and tebufenozide (RH-5992) on male reproductive physiology in *Spodoptera litura*. *J Insect Physiol*. **50** (6). pp. 505–517.

Sevener, J.D., Dennard, N.N. & Grimnes, K.A. (1992). Histological and histochemical evidence for an additional cell type in the male accessory reproductive glands of *Tribolium brevicornis* (Coleoptera: Tenebrionidae). Tribolium Information Bulletin. **34**. pp. 72-74.

Shapiro, A.B., et al. (1986). Juvenile hormone and juvenile hormone esterase in

adult females of the mosquito *Aedes aegypti*. *J Insect Physiol.* **32** (10). pp. 867–877.

Shemshedini, L., Lanoue, M. & Wilson, T.G. (1990). Evidence for a juvenile hormone receptor involved in protein synthesis in *Drosophila melanogaster*. *J Biol Chem.* **265** (4). pp. 1913–1918.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G., et al. (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem*. **276** (36). pp. 33293–33296.

Simmons LW. (2001). *Sperm competition and its evolutionary consequences in the insects*. New Jersey: Princeton University Press.

Simmons, L.W., Tan, Y.F. & Millar, A.H. (2013). Sperm and seminal fluid proteomes of the field cricket *Teleogryllus oceanicus*: identification of novel proteins transferred to females at mating. *Insect Mol Biol*. **22** (1). pp. 115–130.

Sirot LK, Poulson RL, Caitlin McKenna M, Girnary H, Wolfner MF, Harrington LC. (2008). 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*. **38**. pp.176–189.

Sirot, L., LaFlamme, B., Sitnik, J., Rubinstein, C., Avila, F., Chow, C. & Wolfner, M. (2009). Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. *Adv Genet*. pp. 23–56.

Sirot, L.K., Hardstone, M.C., Helinski, M.E.H., Ribeiro, J.M.C., Kimura, M., Deewatthanawong, P., Wolfner, M.F. & Harrington, L.C. (2011). Towards a semen proteome of the dengue vector mosquito: Protein identification and potential functions. *PLoS Negl Tropl Dis*. **5** (3). p. e989.

Sitnik JL, Francis C, Hens K, Huybrechts R. (2014). Neprilysins: an evolutionarily conserved family of metalloproteases that play important roles in reproduction in *Drosophila*. *Genetics*. **196** (3). pp. 781-797.

Smith RL. (1984). *Sperm competition and the evolution of animal mating systems*. Massachusetts: Academic Pr.

Sokoloff, A. (1974). *The biology of Tribolium: with special emphasis on genetic aspects*. Oxford University Press.

Sonnhammer, E., Heijne, von, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*. **6** (6912). pp. 175–182.

South A, Stanger-Hall K, Jeng M-L, Lewis SM. (2010). Correlated evolution of female neoteny and flightlessness with male spermatophore production in fireflies (COLEOPTERA: LAMPYRIDAE). *Evolution*. **65**. pp. 1099–1113.

South A, Lewis SM. (2011). The influence of male ejaculate quantity on female fitness: a meta-analysis. *Biol Rev*. **86**. pp. 299–309.

South, A., Sirot, L.K. & Lewis, S.M. (2011) Identification of predicted seminal fluid proteins in *Tribolium castaneum*. *Insect Mol Biol.* **20.** pp.447–456.

South, A. & Lewis, S.M. (2012). Effects of male ejaculate on female reproductive output and longevity in *Photinus* fireflies. *Can J Zool*. **90** (5). pp. 677–681.

Spehr, M., Schwane, K., Riffell, J.A., Zimmer, R.K. & Hatt, H. (2006). Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. *Moland Cell Endocrinol*. **250** (1-2). pp. 128–136.

Suarez SS. (2005). Sperm transport in the female reproductive tract. *Human Repro Update.* **12**. pp. 23–37.

Sugawara, T. (1979). Stretch reception in the bursa copulatrix of the butterfly, *Pieris rapae crucivora*, and its role in behaviour. *J Comp Physiol [A]*. **130** (3). pp. 191–199.

Sun, X., Barrett, B.A. & Biddinger, D.J. (2000). Fecundity and fertility reductions in adult leafrollers exposed to surfaces treated with the ecdysteroid agonists tebufenozide and methoxyfenozide. *Entomol Exp Appl*. **94** (1). pp. 75–83.

Sun, Y.-L., et al. (2012). Expression in antennae and reproductive organs suggests a dual role of an odorant-binding protein in two sibling *Helicoverpa* species. *PLoS ONE*. **7** (1). pp. e30040–11.

Surtees, G. (1961). Spermathecal structures in some coleoptera associated with stored products. *Proc A*. **36** (10-12). pp. 144–152.

Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F. & Aquadro, C.F. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Nat Acad Sci USA*. **98** (13). pp. 7375–7379.

Swanson, W.J. & Vacquier, V.D. (2002). The rapid evolution of reproductive proteins. *Nature Rev Genet*. **3** (2). pp. 137–144.

Swanson, W.J. (2004). Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics*. **168** (3). pp. 1457–1465.

Thomas, J.E., Rylett, C.M., Carhan, A., Bland, N.D., Bingham, R.J., Shirras,

A.D., Turner, A.J. & Isaac, R.E. (2005). *Drosophila melanogaster* NEP2 is a new soluble member of the neprilysin family of endopeptidases with implications for reproduction and renal function. *Biochem*. **386** (Pt 2). pp. 357–366.

Thornhill, R. & Alcock, J. (2013). *The Evolution of Insect Mating Systems*. Harvard University Press.

Ting, C. T., et al. (2000). The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus. Proc B*. **97**. pp. 5313–5316.

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B. & Turk, D. (2012). Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*. **1824** (1). pp. 68–88.

Valanne, S., Wang, J.H. & Ramet, M. (2011). The *Drosophila* toll signaling pathway. *J Immunol*. **186** (2). pp. 649–656.

Vencl, F.V. & Carlson, A.D. (1998). Proximate mechanisms of sexual selection in the Firefly *Photinus pyralis* (Coleoptera: Lampyridae). *J Insect Behav*. **11** (2). pp. 191–207.

Vigoreaux, J.O. (1994). The muscle Z band: lessons in stress management. *J Muscle Res Cell Motil*. **15** (3). pp. 237–255.

Viviani, V.R. (2001). Fireflies (Coleoptera: Lampyridae) from southeastern Brazil: Habitats, Life history, and bioluminescence. *Ann Entomol Soc Am*. **94** (1). pp. 129–145.

Wade, M.J., Patterson, H., Chang, N.W. & Johnson, N.A. (1994). Postcopulatory, prezygotic isolation in flour beetles. *Heredity*. **72** (2). pp. 163–167.

Wagstaff, B.J. (2005). Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics*. **171** (3). pp. 1083–1101.

Walters, J.R. & Harrison, R.G. (2010). Combined EST and Proteomic Analysis Identifies Rapidly evolving seminal fluid proteins in *Heliconius* butterflies. *Mol Biol Evol*. **27** (9). pp. 2000–2013.

Walters JR, Harrison RG. (2008). EST analysis of male accessory glands from *Heliconius* butterflies with divergent mating systems. *BMC Genomics.* **9**. pp. 592.

Wang, L., Wang, S., Li, Y., Paradesi, M.S.R. & Brown, S.J. (2007). BeetleBase: the model organism database for *Tribolium castaneum*. *Nucleic Acids Research*. 35. pp. D476–9.

Wang X-W, Luan J-B, Li J-M, Bao Y-Y, Zhang C-X, Liu S-S. (2010). De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics.* **11.** pp. 400.

Warnes, G.R., Bolher, B., & Lumley, T. (2008). gplots: various R programming tools for plotting data, v2.6.0. *R Package Version.*

Wedell N. (2005). Female receptivity in butterflies and moths. *J Exp Biol*. **208**. pp. 3433–3440.

Weidinger S. (2003). Mast cell-sperm interaction: evidence for tryptase and proteinase-activated receptors in the regulation of sperm motility. *Human Repro.* **18**. pp. 2519–2524.

West-Eberhard MJ. (1983). Sexual selection, social competition, and speciation. *Qy Rev Biol*.

Wicher, D., Schäfer, R., Bauernfeind, R., Stensmyr, M.C., Heller, R., Heinemann, S.H. & Hansson, B.S. (2008). *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature*. **452** (7190). pp. 1007–1011.

Wilburn, D.B. & Swanson, W.J. (2015). From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J Proteomics*. pp. 1–14.

Wilson, T.G., DeMoor, S. & Lei, J. (2003). Juvenile hormone involvement in *Drosophila melanogaster* male reproduction as suggested by the Methoprene-tolerant27 mutant phenotype. *Insect Biochem Mol Biol*. **33** (12). pp. 1167–1175.

Wolfner, M. F., Harada, H. A., Bertram, M. J., Stelick, T. J., Kraus, K. W., Kalb, J. M., Lung, Y. O., Neubaum, D. M., Park, M., & Tram, U. (1997). New genes for male accessory gland proteins in *Drosophila melanogaster*. *Insect Biochem Mol Biol*. **27**, pp. 825-834.

Wolfner MF. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. *Heredity.* **88**. pp. 85–93.

Wolfner, M.F. (2009). Battle and ballet: Molecular interactions between the sexes in *Drosophila*. *Heredity*. **100** (4). pp. 399–410.

Wong A, Turchin MC, Wolfner MF, Aquadro CF. (2008). Evidence for Positive Selection on Drosophila melanogaster Seminal Fluid Protease Homologs. *Mol Biol Evol.* **25**. pp. 497–506.

Xu, J., Tan, A. & Palli, S.R. (2010). The function of nuclear receptors in regulation of female reproduction and embryogenesis in the red flour beetle, *Tribolium castaneum*. *J Insect Physiol*. **56** (10). pp. 1471–1480.

Xu, J., Baulding, J. & Palli, S.R. (2013). Proteomics of *Tribolium castaneum* seminal fluid proteins: Identification of an angiotensin-converting enzyme as a key player in regulation of reproduction. *J Proteomics*. **78** (C). pp. 83–93.

Yapici, N., Kim, Y.-J., Ribeiro, C. & Dickson, B.J. (2008). A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature*. **451** (7174). pp. 33–37.

Yeates, S.E., Diamond, S.E., Einum, S., Emerson, B.C., Holt, W.V. & Gage, M.J.G. (2013). Cryptic choice of conspecific sperm controlled by the impact of ovarian fluid on sperm swimming behavior. *Evolution*. **67** (12). pp. 3523–3536.

Zdobnov, E.M. & Apweiler, R. (2001). InterProScan-an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. **17** (9). pp. 847–848.

Zera, A.J. & Denno, R.F. (1997). Physiology and ecology of dispersal polymorphism in insects. *Annu Rev Entomol*. **42** (1). pp. 207–230.

Zhong W, McClure CD, Evans CR, Mlynski DT, Immonen E, Ritchie MG, Priest NK. (2013). Immune anticipation of mating in *Drosophila*: Turandot M promotes immunity against sexually transmitted fungal infections. *Proc B*. **280**. pp. 20132018–20132018.