# Post-mating, pre-zygotic interactions and their potential to drive speciation in the Cabbage White butterfly, *Pieris rapae*

by

**Melissa Sue Plakke**

B.S. in Zoology, Colorado State University, 2012

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

**Melissa Sue Plakke**

It was defended on

July 9, 2019

and approved by

Dr. Anne Carlson, Assistant Professor, Department of Biological Sciences

Dr. Mark Rebeiz, Associate Professor, Department of Biological Sciences

Dr. Corinne Richards-Zawacki, Associate Professor, Department of Biological Sciences

Dissertation Co-advisor: Dr. Nathan Morehouse, Associate Professor, Department of Biological Sciences, University of Cincinnati

Dissertation Co-advisor: Dr. Nathan Clark, Associate Professor, Department of Computational and Systems Biology

**Post-mating, pre-zygotic interactions and their potential to drive speciation in the Cabbage White butterfly, *Pieris rapae***

Melissa Sue Plakke, PhD

University of Pittsburgh, 2019

Research on the evolution of reproductive barriers during incipient speciation has typically focused on either pre-copulatory or post-zygotic processes. However post-mating, pre-zygotic (PMPZ) interactions between males and females may play an important role in the early stages of speciation. PMPZ interactions are shaped by both sexual conflict and selection for reproductive cooperation, and preliminary observations suggest they evolve rapidly. I use butterflies to explore the PMPZ interactions between the female reproductive tract and the male ejaculate. Female butterflies have a specialized reproductive organ, the bursa copulatrix, which receives and digests the male ejaculate, or spermatophore. Spermatophore proteins aid in the cooperative venture of egg production, but also function to manipulate female remating rate, resulting in conflict between the sexes over remating frequency. However, it was unknown how the bursa digests the spermatophore, what spermatophore proteins are targeted for digestion by the bursa, or how these proteins interact and evolve over time. In the Cabbage White butterfly, *Pieris rapae*, I discovered that females of the European subspecies, *Pieris rapae rapae*, experience difficulty degrading spermatophores from males of the Japanese subspecies, *Pieris rapae crucivora*. I hypothesized that this mismatch is due to rapid evolution at the interface between the proteases of the bursa and the proteins of the spermatophore. I first identified spermatophore proteins and bursal enzymes responsible for spermatophore digestion using bioinformatic and biochemical techniques. I

classified spermatophore and bursal protein functions and how the proteins from both sexes interact with each other within the context of the female reproductive tract. I then investigated divergence of spermatophore and bursal proteins using RNA-seq and Pool-seq in both subspecies. I found a total of 40 bursal proteases and 66 spermatophore proteins that likely contribute to the interaction and digestion of the spermatophore. The proteases directly digest the spermatophore at a rapid rate and without prejudice. I also uncovered that female proteases exhibit higher expression and genomic differentiation than the corresponding male proteins involved at the interface between the spermatophore and female reproductive tract. I discuss these results in the context of rapid evolution of PMPZ interactions and their potential to lead to speciation.

## Table of contents

## List of tables

## List of figures

## Preface

It is my great honor and pleasure to dedicate my dissertation to two of the most amazing men I know. First, to Dr. Ronald Plakke, my Pop-pop. He started me on my journey as a scientist and I am proud to be his little critter. And second, to my Douglas Hall, who encouraged me and ensured I had every resource I needed to continue and finish my journey. I would not be where I am today without either of them.

The following dissertation represents a work of love. It not only represents my love for science, but the love and support of everyone who came before me and who helped me throughout the years. I have been exceedingly fortunate to have worked with so many influential and bright minds. I am so thankful for the constant love and support of my parents, Syrena and Russell Plakke, and occasional phone calls from my brother Kevin. I further have to thank my friends who kept me sane throughout the years, especially Elizabeth, Sebastian, Ivona, Katie, and Tamara.

I also have to thank all of the labs I have had the pleasure to work with. Thank you to Dr. Nathan Clark for accepting me with open arms into his lab and being extraordinarily patient with me as I began my soiree into bioinformatics. Thank you to Dr. Kentaro Arikawa for hosting me for a summer and sharing your passion for all aspects of science. And thank you to Dr. Nathan Morehouse for starting it all. In the Morehouse lab I was granted the opportunity to work with so many talented and enthusiastic undergraduates, but have to specifically thank Aaron, Erin, Matt, Breanna, Kevin, Kaylee, Keith, Nick, Sean, Jeff, Gina, and Kelly. I have also been remarkably fortunate to work with such inspirational women. And so I extend my greatest gratitude to Dr. Wynn Meyer, Dr. Michiyo Kinoshita, Dr. Nancy Kaufmann, and Dr. Kathryn Gardner.

.

## 1.0 Introduction

According to the Biological Species Concept, two individuals represent one species if they are able to successfully reproduce and produce fertile offspring (Mayr, 1942). By this definition, there are predicted to be 8.7 million species on planet Earth (Mora, Tittensor, Adl, Simpson, & Worm, 2011). How did these species arise? How are they maintained?

One mechanism thought to give rise to new species is allopatric divergence resulting in reproductive barriers upon secondary contact (Dobzhansky, 1951; Mayr, 1963). This process starts with one population of freely interbreeding individuals. This population is then separated by some physical barrier, be it mountains, bodies of water, or roadways. In the newly isolated populations, genetic changes occur independently, and may be influenced by selection or neutral processes such as drift (Slatkin, 1978). If these two populations then re-establish contact with each other, there are two potential outcomes: 1) the populations freely interbreed again and collapse back into one population again, or 2) the populations interbreed but experience a reduction in fitness and therefore evolve reproductive barriers to prevent matings between the populations to maximize fitness (Grant & Grant, 2002; Hoskin, Higgie, Mcdonald, & Moritz, 2005). These potential barriers can occur at any point during the reproductive process, but are most likely to involve traits that are rapidly evolving are directly involved in the reproductive process.

One set of traits that has the potential to develop as a reproductive barrier are traits of the reproductive tracts themselves. Reproductive traits, such as male genitalia, are some of the most rapidly evolving of all characteristics known across organisms (Eberhard, 1985). In fact, in many species complexes, genitalia are the only characteristic that can be used to discern species from each other (Nagarkatti & Nagaraja, 1971). This rapid evolution is not limited to the morphological

level, but can extend to molecular traits as well (N. L. Clark, Aagaard, & Swanson, 2006). Reproductive tissues and cells, such as sperm, express proteins that experience the highest levels of amino acid substitutions of any tissue or cell type (Torgerson, Kulathinal, & Singh, 2002). Therefore, reproductive traits exhibit multifaceted and dynamic potential to develop barriers to reproduction under certain evolutionary pressures.

Traits, behaviors, and proteins that can lead to the formation of reproductive barriers can develop at any number of stages along the process of reproduction. The various timeframes of reproduction and the interactions that occur throughout the reproductive process are generally categorized in relation to either the time of copulation, or the time of fertilization (Figure 1). Within the following chapters, I focus on the interactions that occur after mating, but before fertilization, also known as post-mating, pre-zygotic (PMPZ) interactions. This timeframe has traditionally received less attention than other phases along the reproductive process (Ahmed-Braimah, 2016). However, increasing evidence suggests that outcomes from this timeframe can have long reaching fitness consequences for both males and females (McDonough, Whittington, Pitnick, & Dorus, 2016; Wolfner, 2009). Additionally, recent evidence suggests that genes underlying PMPZ reproductive barriers evolve at a more rapid rate than those involved in either pre-mating isolation, or post-zygotic events (Turissini, McGirr, Patel, David, & Matute, 2018).

For my dissertation, I focused on how PMPZ interactions and the genes underlying them can contribute to the process of speciation. I chose to approach this question using a lepidopteran species, the Cabbage White butterfly *Pieris rapae*, due to the unique traits the female reproductive tract presents. In the female reproductive tract, the ejaculate and sperm are deposited into a specialized female organ called the bursa copulatrix (Rogers & Wells, 1984). Shortly after the cessation of copulation, the sperm migrates to the sperm storage organ, the spermatheca (Rutowski

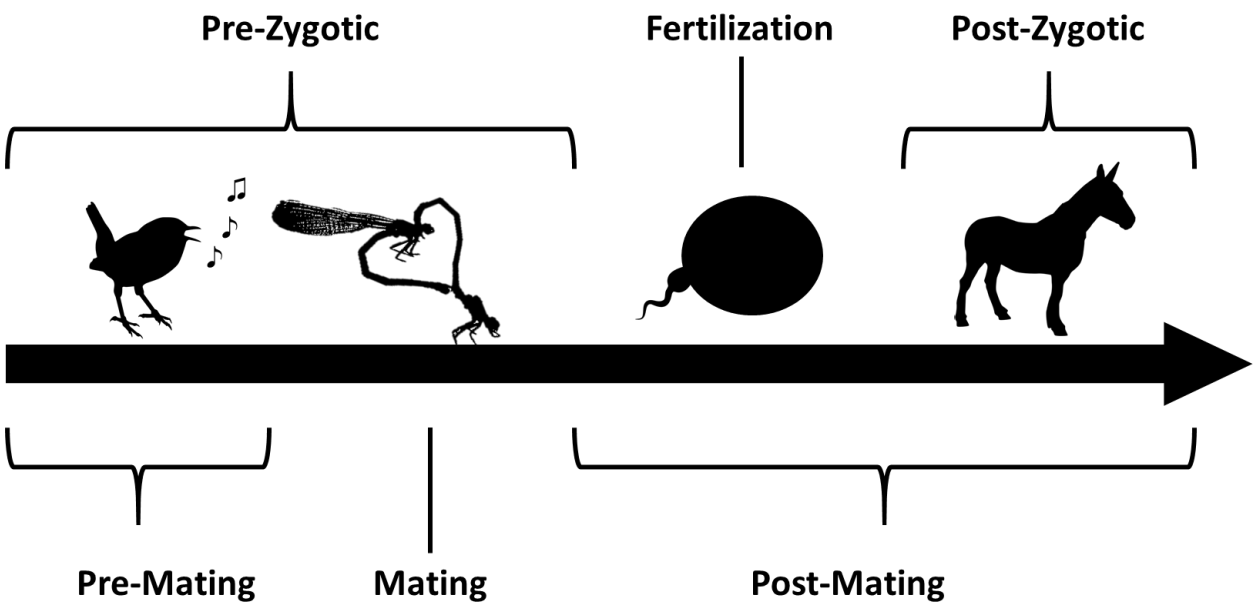


**Figure 1: Phases of reproduction**

Reproductive interactions are generally described in relation to either fertilization or copulation. Barriers to reproduction have the potential to develop at any point along the process, such as mate choice (e.g. bird song), genital compatibility (e.g. damselfly mating wheel), gamete recognition (e.g. egg and sperm fusion), and hybrid sterility (e.g. mule infertility). Images adapted from PhyloPic.

& Gilchrist, 1986). This separation of the sperm from the majority of ejaculate proteins allows for the independent analysis of the effects of the two components (sperm and ejaculate proteins). Interactions pertaining to only one component can be characterized in the absence of the other, an uncommon feature of reproductive systems. It has been long known that the ejaculate proteins that are deposited within the bursa copulatrix are digested and used by the female (Boggs & Gilbert, 1979), but work to date has focused on the mechanical digestion imparted by the bursa copulatrix (Cordero, 2005; Sugawara, 1981). This narrow view of digestion has surely missed an important component of the digestive process, specifically chemical digestion. Biochemical interactions are likely to play a role in the degradation of the spermatophore. The interface between digestive enzymes and their ejaculate proteins provides a targeted interaction that not only can affect the fitness of both sexes, but also has the potential to rapidly diverge across populations.

In the following chapters, I explore the biochemical interaction between the male spermatophore proteins and the female bursal proteases in the context of speciation. I begin by demonstrating the protein digesting capacity of the bursa copulatrix and explore the conditions that influence a female's ability to digest a general protein. I next characterize the proteases predicted to be present within the bursa copulatrix and measure their specificity in relation to their conserved active site homology. Following the functional assays of bursal proteases, I identify which of the spermatophore proteins they target for digestion. Finally, I compare this digestive interaction across subspecies of *P. rapae* at both phenotypic and molecular levels and assess the potential this interaction has to contribute to the process of speciation.

## 2.0 Dynamic digestive physiology of a female reproductive organ in a polyandrous butterfly

The contents of this chapter are adapted from a recently published article of the same name:
 Originally published in The Journal of Experimental Biology https://doi.org/10.1242/jeb.118323

### 2.1 Introduction

Reproductive characteristics are some of the most rapidly evolving traits (N. L. Clark et al., 2006; Torgerson et al., 2002). This rapid evolution is thought to often be the result of sexual co-evolution as each sex aims to increase their reproductive success. Such co-evolution may involve adaptive changes that increase the fitness of both sexes through reproductive cooperation. Alternatively, antagonistic co-evolution can occur when selection favors adaptations that increase the fitness of one sex at the expense of the opposite sex (Pitnick, Miller, Schneider, & Markow, 2003; W. Swanson & Vacquier, 2002). This latter situation, called sexual conflict, has been the subject of intensive research effort over the past two decades, with a particular focus on male reproductive traits that impose fitness costs on females during or following copulation (Arnqvist & Rowe, 1995; L. Rowe & Day, 2006). However, work on related female adaptations has lagged significantly behind, despite repeated calls for increased research attention to female reproductive traits (Ah-King, Barron, & Herberstein, 2014; Mendez & Cordoba-Aguilar, 2004; L. W. Simmons, 2014). In fact, over the past decade, the male bias in the study of primary reproductive traits has worsened, rather than improved (Ah-King et al., 2014).

There are a number of compelling reasons to focus more attention on female reproductive adaptations. First, they should directly inform our conception of how males and females interact over reproduction. The implicit view provided by our male-biased knowledge base is that females are passive or less-active participants in key reproductive interactions. However, there is no clear argument for why this might always or even often be the case. Rather, female reproductive adaptations have been identified in all systems where females have been rigorously studied (Holman & Snook, 2006; Kelleher, Swanson, & Markow, 2007; Knowles & Markow, 2001; L. Simmons & Gwynne, 1991). However, more case studies are required to better inform how and why females influence reproductive outcomes using specific adaptations. An additional benefit to studying female reproductive traits is that this should enable me to identify reproductive interfaces subject to male-female co-evolution. Critical tests that parse between putative mechanisms of reproductive co-evolution (e.g., sexual conflict versus cooperative co-evolution) are much needed. However, in the absence of knowledge of both male and female traits, co-evolutionary explanations for reproductive diversity remain in the realm of speculation. Lastly, characterization of female reproductive adaptations should allow me to better understand reproductive incompatibilities that play a role in pre- and/or post-zygotic isolation during speciation (Orr, 2005; W. Swanson & Vacquier, 2002).

One promising interface for identifying female reproductive adaptations is the processing of male ejaculates by the female reproductive tract. During copulation, males often transfer complex mixtures of ejaculatory proteins to the female alongside their sperm (Perry, Sirot, & Wigby, 2013; Vahed, 1998). The female reproductive system interacts with these diverse male proteins in ways that I are only just beginning to understand (Findlay et al., 2014; Ram & Wolfner, 2007; Wolfner, 2009). However, it is clear that these interactions form a key interface for both

male and female reproductive fitness. Ejaculate proteins can directly influence male reproductive success via their role in both fertilization and sperm competition (Fiumera, Dumont, & Clark, 2005, 2006; Reinhart & Carney, 2014). For example, ejaculate proteins have been shown to provide energetic substrates and aid in sperm mobility, resulting in increased fertilization rates and male paternity share (Gillott, 2003). Ejaculate proteins have also been implicated in a wide range of effects on female post-copulatory phenotypes, including reduced female receptivity to subsequent mates, increased female reproductive output, and reduced female lifespan (Perry et al., 2013; Ram & Wolfner, 2007). In addition, male ejaculates often form copulatory plugs that reduce female remating rate (Baer, Morgan, & Schmid-Hempel, 2001; Bretman, Lawniczak, Boone, & Chapman, 2010; Shine, 2000), although they can also be important for general fertility (Dean, 2013). Such effects may often benefit males at the expense of their female mates (Karen S. Oberhauser, 1989; Wolfner, 1997), resulting in the potential for antagonistic co-evolution between manipulative compounds in male ejaculates and counter-adaptations in the female reproductive tract.

Identifying specific female adaptations to manipulative male ejaculate substances has proven to be challenging. This is in part due to the fact that many key interactions between male ejaculates and female reproductive adaptations occur within the main channel of the female reproductive tract (i.e. the vagina and oviducts (Kelleher et al., 2007; Knowles & Markow, 2001)). Because a wide range of reproductive processes occur in this reproductive region, the specific function of focal female traits is often not clear. Nevertheless, researchers have begun identifying a number of female physiological traits that may serve as counter-adaptations to male ejaculatory substances. These include secreted proteases that may function to dislodge male copulatory plugs and/or de-activate manipulative ejaculate compounds (Kelleher, Clark, & Markow, 2011; Kelleher

et al., 2007), female receptors and associated hormonal processes that respond to male ejaculate constituents (Adams, Ratto, Huanca, & Singh, 2005; Yapici, Kim, Ribeiro, & Dickson, 2008), and morphological features of the female reproductive tract that could play a role in cryptic female choice via their effect on male fertilization success and sperm competition (Pitnick et al., 2003).

These female traits present promising leads for understanding reproductive co-evolution between male ejaculates and the female reproductive tract. However, with the exception of recent work on hormonal changes in the female reproductive tract following copulation (Heifetz, Lindner, Garini, & Wolfner, 2014), I know almost nothing about how these female traits change dynamically in response to female state or male ejaculatory secretions. For example, are female reproductive adaptations modulated by age or temperature? Are female reproductive adaptations sensitive to social cues such as the presence of conspecifics or potential mates? How do specific female adaptations respond to mating and how do they change post-copula? Research on this front is critical for understanding how these traits mediate male-female interactions as well as how they function across individual female life histories. However, with the exception of a few recent studies in *Drosophila melanogaster* (Bono, Matzkin, Kelleher, & Markow, 2011; Kelleher & Pennington, 2009; Mack, Kapelnikov, Heifetz, & Bender, 2006; McGraw, Clark, & Wolfner, 2008), such information is largely lacking for even well-studied organisms.

I sought to answer these questions by focusing on a specific reproductive interaction in the polyandrous butterfly *Pieris rapae*: the digestion of the male ejaculate or spermatophore by a specialized organ in the female reproductive tract called the bursa copulatrix (hereafter bursa). In insects, bursae are common features of female reproductive tracts. Bursae that play a role in spermatophore processing have been described in the Coleoptera (Reijden, Monchamp, & Lewis, 1997), Trichoptera (Khalifa, 1949), and Lepidoptera (Arnqvist & Nilsson, 2000; Oberhauser,

1989; Vahed, 1998). In the Lepidoptera, the bursa copulatrix serves specifically to receive and break down the male spermatophore (Engelmann, 1970; Mamoru Watanabe, Wiklund, & Bon, 1998; Wiklund, Karlsson, & Leimar, 2001). Following spermatophore transfer, sperm migrate to a specialized sperm storage organ called the spermatheca (Rutowski & Gilchrist, 1986), leaving the bursa to process the remaining ejaculatory compounds.

Subsequent processing of the spermatophore by the bursa has important consequences for male and female fitness in the Lepidoptera. Females utilize ejaculate proteins for somatic maintenance and egg production (Boggs & Gilbert, 1979). Thus, in polyandrous lineages, females often gain fitness benefits from mating multiply via increases in their lifespan and fecundity (Wiklund et al., 2001) However, the spermatophore can also function to reduce female remating rate. Females typically do not remate until the spermatophore has been absorbed enough to allow space for another spermatophore (Oberhauser, 1989; Sugawara, 1979). Thus, male traits that delay spermatophore processing are likely to benefit male paternity share in polyandrous lineages (Sánchez, Hernández-Baños, & Cordero, 2011). Researchers have begun to identify male and female traits that influence bursal processing of the spermatophore. These include tough outer spermatophore envelopes that delay female access to the softer material inside the spermatophore (Sánchez & Cordero, 2014a) and toothed, muscularized devices attached to the bursal wall called signa that serve as counter-adaptations via their role in mechanically abrading the spermatophore envelopes (Sánchez et al., 2011). The presence of tough envelopes and bursal signa both appear to be favored in more polyandrous lineages (Sánchez & Cordero, 2014a; Sánchez et al., 2011).

Although researchers have often described bursal processing of the spermatophore as a digestive process, little is known about how this is accomplished physiologically or whether it involves enzymatic digestion at all. I therefore sought to first establish whether females produce

protein-digesting enzymes in the bursa, or simply absorb male ejaculate proteins without enzymatic processing. I then asked whether female enzymatic activity was influenced by female state. More specifically, I evaluated changes in female bursal proteolysis related to female age, adult temperature and social experience. Dynamics of bursal proteolysis may help to provide insights into the control of bursal physiology as well as the consequences of phenotypic plasticity for male-female reproductive interactions.

I first examined the effect of age. I expected to either see high levels of digestive activity at eclosion, implying that females eclose fully sexually mature and prepared to engage in spermatophore digestion immediately, or alternatively that females eclose with low proteolytic capacity but increase their proteolytic capacity with age. This latter pattern would imply that adult females must produce and actively secrete proteolytic enzymes into the bursal lumen in advance of their first mating. Second, I tested the effect of the pre-mating social environment on digestive activity. It is well established that male traits, such as ejaculate composition, can vary in response to social cues (Cornwallis & Birkhead, 2007; Ramm & Stockley, 2009; L. K. Sirot, Wolfner, & Wigby, 2011; Smith & Ryan, 2011; Wigby et al., 2009). However, female responses to social cues remain largely unknown. Because digestive enzymes can present a physiological liability at high concentrations (Hirota, Ohmuraya, & Baba, 2006; van Hoef et al., 2011), I expected that exposure to courting males might stimulate females to increase protease production in anticipation of imminent mating. However, if males are constantly present in a virgin female's environment or females rarely have the opportunity to realize a benefit from such phenotypic plasticity, I would expect little to no effect of male exposure on bursal physiology. Finally, I examined the effect of mating upon bursal enzymatic activity. I predicted that mating would stimulate increased bursal proteolytic activity, potentially leading to increased levels of proteolytic activity post-copula.

Alternatively, a female's digestive activity might decrease, either due to adaptive regulation of bursal physiology by the female or as a result of enzymatic quenching by male ejaculate constituents (Dean et al., 2009).

I followed these physiological studies with work to preliminarily establish the identities of female proteases that underlie the proteolytic activity I observed in the bursal lumen. Using a combination of transcriptomic and proteomic approaches, I sought to identify proteases either highly transcribed in bursal tissue and/or expressed in detectable quantities in the bursal lumen of sexually mature females.

## 2.2 Methods

### 2.2.1 Experimental Animals

*Pieris rapae rapae* were raised in dedicated climate chambers that maintained a 16:8 light:dark photoperiod at a constant 24°C and 60% relative humidity. Larvae were fed on a diet of kale leaves (*Brassica olercea*) grown on site, fertilized twice a week with Peter's Profession General Purpose 20-20-20. Individuals were all the descendants of wild-caught females collected at local agricultural sites in Rochester, PA (40°44'44.4"N 80°09'49.0"W) and Irwin, PA (40°26'34.4"N 79°74'78.3"W). For the Social Environment Study, I used F1 generation individuals from field-caught females. For the remainder of the studies below, I used individuals from a continuous laboratory population established from wild-caught females in October 2012.

### 2.2.2 Enzyme Collection

Enzyme solutions were collected from tissues of interest by micro-vivisection in ice-cold 10mM NaCl. After isolation, excess fluid was removed from the tissues and they were massed in 1.5 mL microcentrifuge tubes and placed on ice. 100 µL of the NaCl solution was added before the tissues were homogenized with 50 turns of a manual pestle. The homogenized solutions were then cooled in ice before centrifugation at 10,000 x G for 15 minutes. Samples were stored at -20°C until assayed.

Bursal tissues were collected from virgins at specific time points post-eclosion and from mated females at either 1, 3, or 5 days post-mating. For bursa samples, extra fat bodies and extraneous tissues were removed. For bursae from mated females, the male's spermatophore was removed by peeling open the bursa and carefully pulling out the spermatophore to remove all male ejaculate contributions contained within the spermatophore mass. Bursae were not rinsed internally to avoid the loss of female enzymatic material from the lumen. For all experiments, the caterpillar intestine was used as a positive control, due to its known protein digesting properties, and the butterfly leg was used as a negative control due to the low expected levels of proteolytic activity in this body part. For larval intestine sampling, the midgut of the $4^{th}$ instar intestinal tract was used. Larval guts were isolated along with the contents in order to adequately capture the proteases within the lumen of the midgut (R. M. Broadway & Duffey, 1986). Leg tissue was collected from adult butterflies.

### 2.2.3 Proteolytic Activity

I used a modified azocasein assay from previously described protocols (Ajamhassani, Zibaee, Sendi, Askary, & Farrar, 2012). 10 µL of each enzyme solution was added to 100 µL of Tris-HCl buffer pH 7 (20 mM) and 50 µL of 2% azocasein (Sigma-Aldrich, St. Louis, MO). The solutions were incubated at 37°C for 60 minutes before 400 µL of 10% trichloroacetic acid was added. The solutions were then placed on ice for 5 minutes to allow precipitation of the excess protein-dye complex before centrifuging at 10,000 x G for 10 minutes. 400 µl of the supernatant was added to an equal amount of 2 M NaOH. Concurrently, a second set of enzymes were run in an identical fashion, with the exception of skipping the incubation step in order to record a standard time zero for enzymatic activity. The absorbance of the resulting solutions were measured in triplicate using an Epoch microplate absorbance spectrophotometer (BioTech, Winooski, VT, USA) at 450 nm with wells filled to 200 µL. Blanks were run in an identical manner, except instead of the enzyme, only the dissecting solution (10 mM NaCl) was used. Enzyme activity is reported in units, defined as the amount of enzyme required to result in a change of 0.01 absorbance per 60 minutes at 37°C. I confirmed linear responses of the azocasein assay to extracted bursal enzymes both across a range of incubation times (linear $r^2 = 0.988$, $p < 0.01$) and enzymes concentrations (linear $r^2 = 0.996$, $p < 0.01$). Incubations occurred at a pH of 7 based on previous work looking at pH-dependent enzymatic activity in the midgut. The enzymes of the late instar larval gut in Lepidoptera peak in activity at a pH of 7 and maintain this high level of activity through a pH of 9 (Ajamhassani et al., 2012; Berenbaum, 1980). Lepidopteran bursa have been found to possess a pH of 7 as well (Khalifa, 1950).

### 2.2.4 Virgin Age Study

I evaluated changes in proteolytic activity in the bursae of virgin females following eclosion into adulthood. Individuals were collected immediately after eclosion, at 1 day post-eclosion, or at 3 days post-eclosion. Bursae from these virgin females were dissected as described above, and the proteolytic activity assayed.

### 2.2.5 Social Environment Study

I evaluated the effect of social environment on bursal physiology by measuring bursal proteolytic capacity following exposure to different social scenarios. Upon pupation, individual females were isolated to prevent exposure to other individuals prior to the social exposure treatments. All focal females used for trials had eclosed 2 days prior to the beginning of the treatment. Each trial consisted of 3 replicates of focal females for each treatment. Focal females were split between treatments designed to provide them with one of three different social environments: 1) isolated (i.e. no social exposure), 2) female-only social exposure, or 3) male and female social exposure. These social environments were accomplished by placing focal females in hanging cylindrical cages with proportional numbers of other butterflies. For the isolated treatment, focal females were housed alone in cages visually isolated from all other treatments (n = 3 for each trial). For the female-only social exposure, 3 focal females were housed with 3 other virgin females. These non-focal females ranged in age but had never been exposed to males. The female and male social exposure treatment contained 3 focal females plus 3 experimentally-castrated males. Male castration was accomplished by dripping unscented wax (Mainstays,

Bentonville, AR) onto the male claspers and genitalia to prevent successful mating attempts while still permitting normal behavior by the male. To ensure that castrated males responded normally to virgin females and vice versa, I monitored male courtship behaviors including approaches and copulation attempts. Castrated males courted females at statistically equivalent rates to those observed for non-castrated males ($F = 3.571$, $df = 1$, $p = 0.132$). I did not observe any other salient differences in male behavior or female responses to male courtship. Each trial was run for 24 hours within a climate-controlled greenhouse. Temperature, humidity and light levels were monitored and treated as a random variable in statistical analyses. I conducted a total of 11 trials. An initial set of trials ($n = 7$) was run early in the season (May and early June) when greenhouse temperatures were cooler (18.86°C ± 0.55). I ran a second set of trials ($n = 4$) later in the summer (July), at which time greenhouse temperatures were notably higher despite climate control systems remaining operational (25.19°C ± 0.04). I consider this difference in temperature in my data analyses. Before being placed in trial enclosures, all individuals were fed 20% honey solution and cages were misted with water every two hours during the daylight hours of the experiment in order to minimize death by dehydration. Following the 24 hour trial period, females were removed, their bursae promptly vivisected, and bursal proteolysis assayed as described above.

### 2.2.6 Mating Study

To analyze the effect of mating on bursal proteolytic activity, I collected bursae from unmated females and mated females 1 day post-copula and 3 days post-copula. Females were mated by housing them with males in a 60 cm x 60 cm x 90 cm insect mating enclosure in direct sunlight. Because matings typically last 30-45 minutes, mating enclosures were checked every 20 minutes to ensure no matings were missed. Males and females found in copula were removed to

an individual cup until separation. For females that were analyzed at one day post mating, the females remained in this cup at 24°C until dissection. For females that were analyzed at three days post mating, the females were also kept at 24°C but provided with a cotton pad soaked with a 20% honey solution until dissection.

### 2.2.7 Statistical Analyses

Statistical analyses were calculated using the statistical program IBM SPSS Statistics, Version 22.0. A one-way ANOVA was performed for all experiments, with the exception of the social environment experiment, where an ANCOVA was performed with ambient temperature as a co-variate. All datasets were evaluated for the assumptions of parametric statistics using Levene's test for normality and spread-versus-level plots for homoscedascity. Two datasets required natural log transformation to achieve data normality (virgin age study and mating study). These were evaluated statistically as transformed data, but are plotted in the manuscript as untransformed data to aid in cross comparisons between studies. Tukey's-b was used for all post-hoc analyses to determine significance groupings. For the social exposure analysis, caterpillar intestine and leg were not included in the statistical analysis because they did not have an associated temperature value for the ANCOVA analysis.

### 2.2.8 Protein Identification

Putative protein identities were determined using both RNA sequencing techniques as well as proteomic analyses. RNA sequencing methods and transcriptional quantitation are described in detail in Meslin, *et al.*, 2015. In brief, bursas were dissected into 100 µL of RNAlater RNA

Stabilizing Reagent (Qiagen, Valencia, CA, USA). RNA was then extracted using TRIzol (Life Technologies, Grand Island, NY, USA) and samples sent to the Genomics Resources Core Facility of Weill Cornell Medical College (New York, NY, USA) for sequencing. Assembly of transcriptomes was accomplished using the Trinity Software Package (version r2013-02-25) (Haas et al., 2013). Genes coding putative bursa proteases were identified based on high quantitative levels of transcription in bursal tissue as well as the presence of secretion signals, which suggest a high likelihood of protease secretion into the bursal lumen.

For proteomics, bursas from three day old virgin females were vivisected, removed from the abdominal cavity, and their outsides rinsed with PBS. Bursas were then cut open and the contents of the lumen suspended in PBS. This lumen extract was then combined with loading buffer **(**Urea 8 M, 200 mM Tris-HCl pH 6.8, 0.1 mM EDTA pH 8.0, DTT 100 mM, Tris Base 100 mM**)** and incubated at 37°C for 15 minutes. The solution was then run on a 12% SDS-PAGE gel until the band measured 0.25 $cm^2$. After staining with Coomassie blue (Amresco, Solon, OH, USA**)**, the protein band was excised and subsequently submitted to the Biomedical Mass Spectrometry Center at the University of Pittsburgh where liquid chromatography tandem mass spectrometry was used in conjunction with previously acquired transcriptomic sequences in order to determine protein identities within the bursal lumen (Granvogl, Gruber, & Eichacker, 2007; Granvogl, Plöscher, & Eichacker, 2007; Shevchenko, Tomas, Havlis, Olsen, & Mann, 2006). Tandem mass spectrometry data were visualized using Scaffold (Proteome Software; Portland, Oregon, USA), with subsequent annotation of highly abundant proteins using BLASTP.

## 2.3 Results

In all assays, the bursa exhibited high levels of protein digestive activity. Bursal proteolytic activity also appears to be influenced by the female's age, abiotic environment and mating status. Summary statistics for all studies are reported in Table 1.

### 2.3.1 Virgin Age Study

Digestive activity increased with increasing age of virgin females. In newly eclosed virgin females (n = 10), digestive activity in the bursa was very low and not statistically different from that of my negative control tissue, the adult butterfly leg (n = 10). However, as females aged, the mean amount of protein digesting activity in the bursa copulatrix increased at 1 day following eclosion (n = 10) and 3 days post-eclosion (n = 9) to levels substantially higher than the leg, and equivalent to, if not higher than, the entire intestine of $4^{th}$ instar larvae (n = 10) ($F = 11.440$, $df = 4$, $p < 0.001$) (Figure 2).

### 2.3.2 Social Environment Study

Pre-mating social environment, namely exposure to males or other females, did not affect the level of digestive enzymatic activity within a female's bursa ($F = 0.363$, $df = 2$, $p = 0.698$). However, there was a significant difference in enzymatic activity between virgin females that were exposed to cooler temperatures (19°C) versus warmer temperatures (25°C) during the social environment experiment ($F = 19.777$, $df = 1$, $p < 0.001$) (Figure 3). No interaction was observed

**Table 1: Summary statistics for proteolytic activity of all tissues and treatment groups.** Experimental treatments are reported within the table with sample sizes and mean changes in absorbance/hour. Shading differentiates specific experiments.

| Tissue | Sample Size | Proteolytic Activity (mean Units per organ ± 95% CI) |
|---|---|---|
| Leg | 10 | 0.53 ± 0.3 |
| Caterpillar Intestine (gut) | 10 | 7.68 ± 4.8 |
| Virgin 0 day | 10 | 2.21 ± 1.4 |
| Virgin 1 day | 10 | 13.03 ± 8.1 |
| Virgin 3 day | 9 | 19.98 ± 13.1 |
| Social exposure: alone, cool | 6 | 20.76 ± 5.0 |
| Social exposure: alone, warm | 4 | 38.38 ± 11.7 |
| Social exposure: same sex, cool | 10 | 24.46 ± 4.5 |
| Social exposure: same sex, warm | 6 | 35.25 ± 4.8 |
| Social exposure: both sexes, cool | 9 | 25.82 ± 4.6 |
| Social exposure: both sexes, warm | 10 | 38.77 ± 9.5 |
| Unmated females | 11 | 9.57 ± 5.7 |
| Mated 1 day | 9 | 6.94 ± 4.5 |
| Mated 3 day | 9 | 2.31 ± 1.5 |
| Mated 5 day | 4 | 6.58 ± 2.5 |

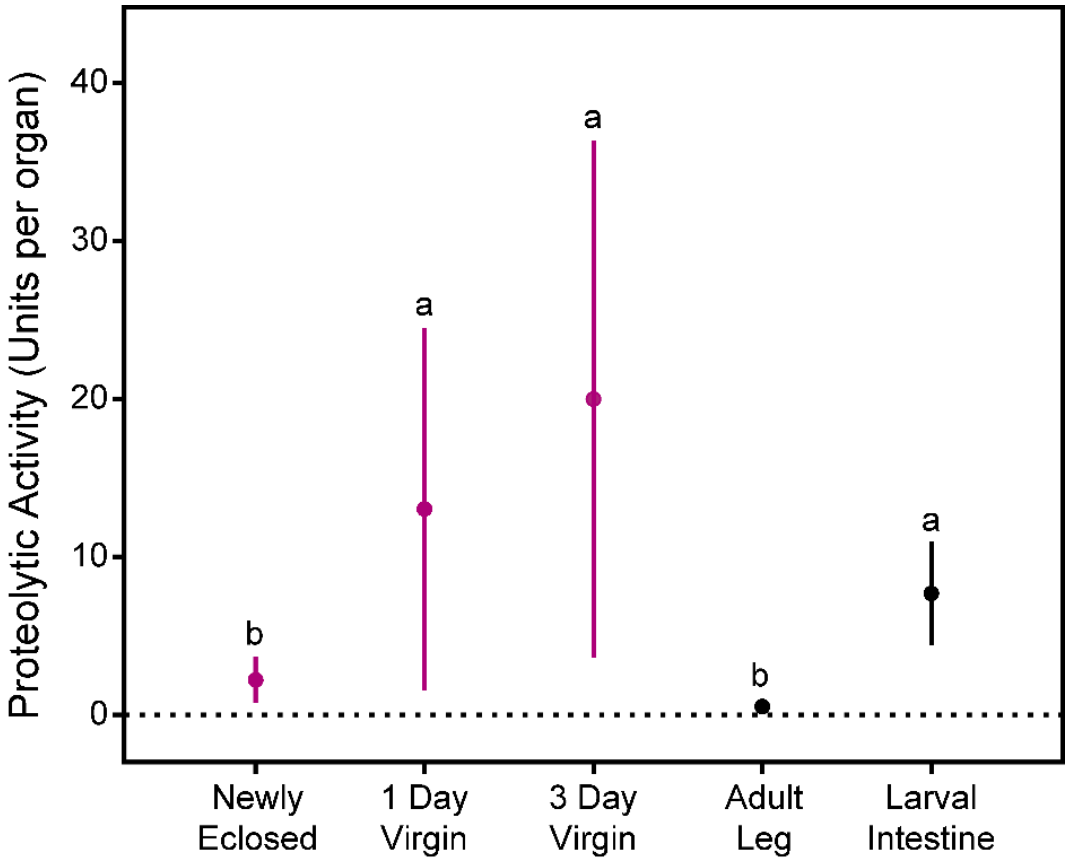


**Figure 2: In virgin butterflies, bursal proteolytic activity increases with female age**

Data are presented as means +/- 95% confidence intervals. Lowercase letters indicate statistical groupings.

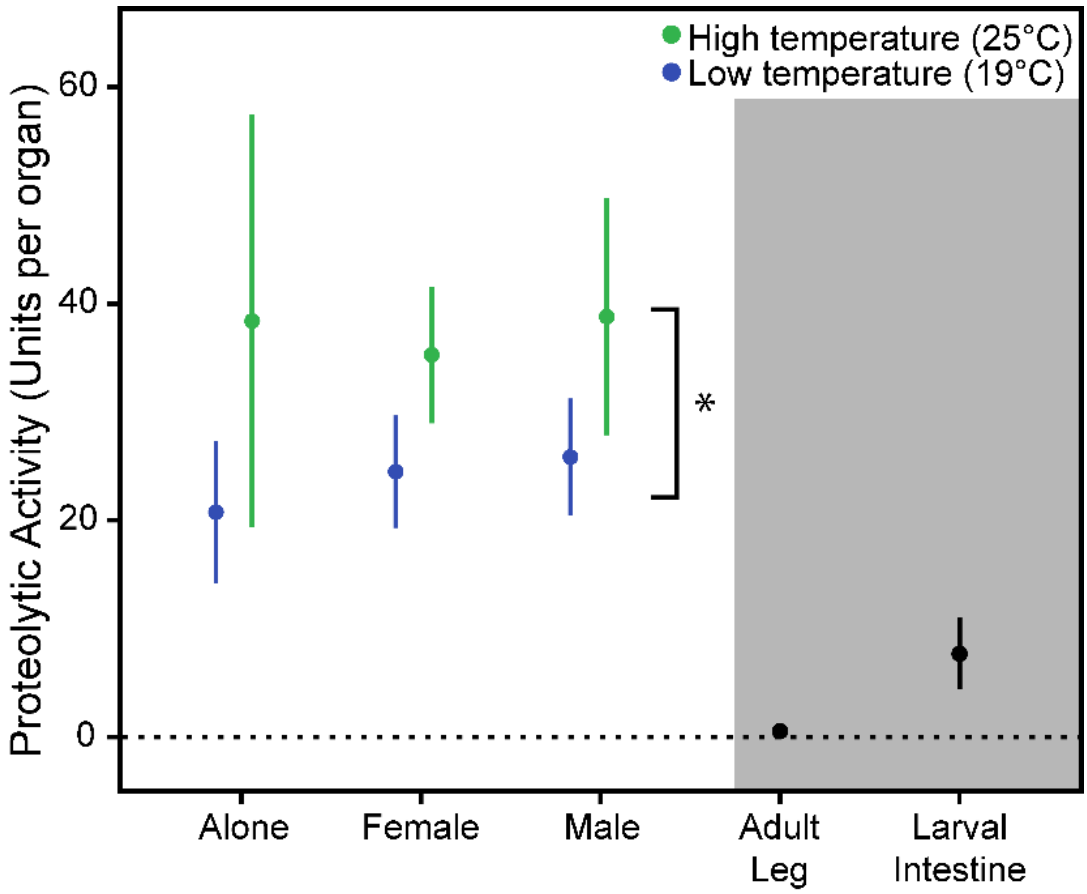


**Figure 3: Bursal proteolytic activity of virgins is not affected by exposure to courting males**

Higher ambient temperatures (25$^{o}$C, plotted in green) do significantly increase the digestive activity of the bursa as compared to lower ambient temperatures (19$^{o}$C, plotted in blue). Data are presented as means +/- 95% confidence intervals. Proteolytic activity of adult leg and caterpillar intestine are presented for comparison, but were not included in the statistical analysis.

between temperature and social environment ($F = 0.357$, $df = 2$, $p = 0.702$). The high interaction density in these constructed social environments combined with reduced female nectar foraging during trials led to some female mortality during the experiment, but mortality rates were not dependent on treatment (chi-square $df = 2$, $p = 0.421$).

### 2.3.3 Mating Study

I measured protein digestive activity in the bursa for unmated females ($n = 11$), females 1 day after mating ($n = 9$), females 3 days after mating ($n = 9$), and females 5 days after mating ($n = 4$) to determine the effect of mating on bursal proteolysis. Protein digestive activity in the virgin bursa begins at a significantly higher activity level than the butterfly leg ($F = 10.946$, $df = 5$, $p < 0.001$). Following mating, bursal proteolysis remains detectably higher than levels found in the adult butterfly leg. However, proteolysis shows a declining trend in both mean and variance during the first 3 days post-copula before increasing again on day 5 (Figure 4), although differences between these time points were not statistically significant.

### 2.3.4 Protein Identification

By extracting the proteins present in the bursal lumen, I were able to identify potential proteases that may act in the bursal digestive process. Mass spectrometry of bursal lumen identified 5 unique proteases at protein identification probabilities exceeding 99% (Table 2). Using bursal transcriptomic data also collected from *P. rapae,* I identified an additional 4 proteases that are highly transcribed in the bursa. Their protein products contain secretion signals, making them also

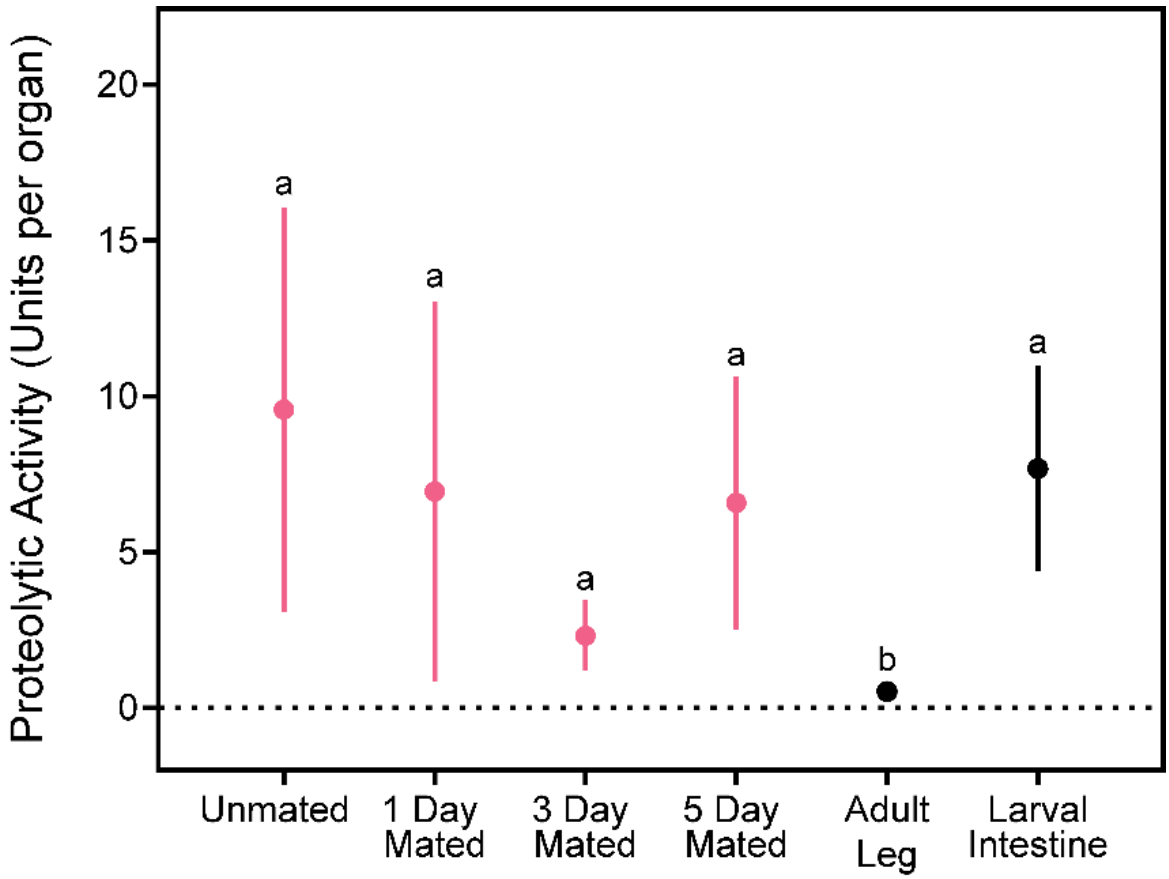


**Figure 4: After mating, bursal proteolytic activity decreases before rebounding at the end of a typical female refractory period**

Data are presented as means +/- 95% confidence intervals. Lowercase letters indicate statistical groupings.

**Table 2: Summary of putative proteases identified in the bursal lumen.**

Sequence identities are listed with information on protease domains and method used to detect them.

| Component Number | Domain(s) Found | Method of Detection |
|---|---|---|
| comp93091_c0 | CLIP, trypsin-like serine protease | RNA |
| comp94445_c1 | Cathepsin propeptide inhibitor, papain family cysteine protease (Pept_C1) | RNA |
| comp95264_c1 | Cystatin-like domain (CY), papain family cysteine protease (Pept_C1) | RNA |
| comp98020_c0 | Peptidase_MA_2, ERAP1_C | RNA |
| comp97068_c0 | Peptidase_S28 | Proteomics |
| comp91676_c0 | Cathepsin propeptide inhibitor domain, papain family cysteine protease (Pept_C1) | Proteomics |
| comp83824_c0 | Trypsin-like serine protease | Proteomics |
| comp85455_c0 | Papain family cysteine protease (Pept_C1) | Proteomics |
| comp83827_c1 | Cathepsin propeptide inhibitor domain, papain family cysteine protease (Pept_C1) | Proteomics |

likely to act in the bursal lumen (C. Meslin et al., 2015). Thus, my transcriptomic and proteomic analyses identify 9 proteases with a putative role in bursal proteolysis (Table 2). These included two trypsin-like serine proteases, five papain family cysteine proteases, and two proteases with poorly classified peptidase domains.

## 2.4 Discussion

I find that the female bursa copulatrix is a dynamic and highly proteolytic organ system. Our study represents the first clear identification of active protein digestion by the lepidopteran bursa, and also reveals the identities of 9 proteases that are likely to contribute to bursal proteolysis. Although standard accounts of bursal function generally describe it as the site of spermatophore digestion (Engelmann, 1970) researchers have yet to directly quantify the proteolytic capacity of the bursa itself. Rather, studies of spermatophore processing in the silkworm *Bombyx mori* have suggested that male-donated enzymes are responsible for spermatophore breakdown (M Osanai & Kasuga, 1990; Minoru Osanai, Kasuga, & Aigaki, 1987). In contrast, I find that the bursa of *P. rapae* is extremely proteolytic, achieving levels of digestive activity equivalent to, if not higher than, the region of the larval gut responsible for protein digestion, the midgut. While this finding is notable in absolute terms, it is even more impressive when considered in relation to the size of these two digestive organ systems. The larval midgut is roughly 20 mg, while the bursa is on average 1 mg. This means that an organ $1/20^{th}$ the size of the larval midgut is able to produce equivalent amounts of proteolytic enzymes, and implies that the bursal lumen exhibits substantially higher enzyme concentrations than midgut. The extent to which this relatively extreme level of proteolysis is representative of bursal function across the Lepidoptera is unknown.

I also find that bursal proteolytic capacity is dynamic. Virgin females eclose with low levels of proteolytic activity in their bursae, but bursal proteolytic activity then increases steadily with age. This pattern is consistent with gradual secretion of proteolytic enzymes following adult eclosion, or alternatively, a gradual release of these enzymes from storage in the bursal tissue. However, I find no detectable levels of proteolytic activity in virgin bursal tissue following rinsing of the bursal lumen (data not shown). This implies that proteolytic enzymes are not stored in detectable quantities in bursal tissue, and thus that active secretion of enzymes into the bursal lumen following synthesis is more likely. I do not know whether this process of gradual accrual of enzymes in the bursal lumen presents a liability for older unmated females due to autodigestion. Intestinal tissues and other digestive organs exhibit preservative or inhibitory mechanisms that reduce the risk of autophagy, such as storing enzymes in inactive forms or secreting specialized inhibitor-like proteins (Hirota et al., 2006; van Hoef et al., 2011). Whether such mechanisms are also present in the bursa is not known, but could present a fruitful avenue for further study. However, it is possible that females rarely experience such potential detrimental effects of high proteolytic activity in the wild, because nearly all females mate within the first several days of adult life (Watanabe & Ando, 1993).

In contrast to enzymatic changes associated with female age, I find no evidence that bursal proteolysis is modulated by the social experience of virgin females. I postulated that females might upregulate their proteolytic activity upon exposure to courting males as a "priming" strategy prior to mating. This could be particularly relevant if bursal enzymes are costly, either due to enzyme synthesis costs or increased risk of autophagy. However, I observed no effect of either the presence of conspecific females or conspecific males on female bursal proteolytic activity. This result may indicate that the rate at which wild females encounter prospective mates is sufficiently high that

there is little fitness benefit to such phenotypic plasticity. However, I did observe that females experiencing higher ambient temperatures in these social exposure trials exhibit detectably higher levels of proteolytic activity. Again, this is consistent with my hypothesis above that females gradually synthesize and secrete enzymes into the bursal lumen. Increased ambient temperatures should raise the metabolic rate of these females, leading to increases in the rate of a number of biochemical processes, including the synthesis of bursal enzymes.

Following female mating, I found a trend that bursal proteolytic activity declines over the first several days post-mating. This decline may be the result of several processes. First, bursal proteases may be absorbed or incorporated into the male spermatophore during or following its deposition in the bursa, resulting in a decline in proteolytic activity in the bursal lumen. In addition, protease inhibitors are not uncommon in male ejaculate cocktails (Dean et al., 2009). Thus, male spermatophore constituents may act to directly reduce female proteolytic activity. Finally, females may themselves down-regulate bursal proteolysis following mating, a possibility supported by my data quantifying post-mating changes in bursal transcriptional profiles (C. Meslin et al., 2015). Following mating, transcription of a number of bursal proteases decreases (C. Meslin et al., 2015). This rather counterintuitive result may indicate that females initially focus on enzymatically digesting the outside of the spermatophore, but then transition to mechanical digestion and subsequent absorption of the spermatophore contents. Interestingly, at 5 days post-mating, I observe an increasing trend in proteolytic activity to approximately the levels seen 1 day post-mating. This corresponds roughly to the length of the typical "mating cycle" in this species (i.e., the length of the refractory period, beyond which females are willing to remate, (Suzuki, 1979)). Thus, this potential increase in proteolytic activity could indicate that females ramp up digestive

enzyme production to prepare for the next spermatophore, or alternatively, that female protease production begins to exceed the rate of quenching by the male spermatophore at this stage.

The identities of the proteases further advance my knowledge of the potential modes of action involved in enzymatic digestion of spermatophore proteins. Three main protease classes were observed, including trypsin-like serine proteases and papain family cysteine proteases. While the value of this protease diversity to female reproduction is not clear, it is possible that employing proteases with different modes of action may increase the digestive efficiency of the bursa regardless of spermatophore substrate. Alternatively, these modes of action may reflect matching diversity in male ejaculate proteins, suggesting coevolution between female enzymes and male substrates. Whatever the case, information about female protease identities offers a critical first step in understanding both the function and evolution of female reproductive physiology in these animals. Future work should explore the evolutionary histories and current functions of these intriguing reproductive enzymes. In addition, the physiological dynamics I report here may be underlain by changes in the titers of a specific subset of proteases. Evidence for independent regulation of these proteases across female reproductive state would provide additional clues into their reproductive role and evolutionary significance.

In conclusion, I find that the female bursa exhibits remarkably high proteolytic capacity, and that bursal physiology is dynamic over a female's life experience. These results highlight the importance of attending to female reproductive adaptations and their responsiveness to female state. Future work should focus on the evolutionary importance of male-female interactions within the bursa, the mechanisms driving diversity in these physiological traits, and their fitness consequences for both males and females. In addition, future research should explore how females dynamically regulate the suite of proteolytic enzymes I have identified to maximize their digestion

of male ejaculate proteins. By expanding this work across a range of lepidopteran species with diverse mating strategies, from monandry to high levels of polyandry, I may begin to better understand the role of sexual conflict and reproductive cooperation in driving male-female reproductive interactions.

## 2.5 Acknowledgements

This project used the Biomedical Mass Spectrometry Center and University of Pittsburgh Cancer Institute Cancer Biomarker Facility, which are supported, in part, by NCI Grant P30CA047904. I would like to thank the University of Pittsburgh for providing funding to support this project. I would also like to thank Nicholas Saleh, Matthew Huff, Erin Moss, Breanna Goetz, and the rest of the Morehouse Lab for animal rearing and sample prep, and Kretschmann and Morris farms for their hospitality during field collection of butterflies.

# 3.0 Protease characterization and specificity in the reproductive tract of female butterflies

## 3.1 Introduction

Reproductive characteristics are among the most rapidly evolving traits currently known (N. L. Clark et al., 2006; Eberhard, 1985; Lee, Ota, & Vacquier, 1995; Laura K Sirot et al., 2014; Torgerson et al., 2002; Wyckoff, Wang, & Wu, 2000). In the study of reproductive trait evolution, female traits have traditionally received less attention than their male counterparts (Ah-King et al., 2014), yet recent evidence suggests that female reproductive morphologies and proteins evolve alongside male traits (Brennan & Prum, 2014; Findlay et al., 2014; Galindo, Vacquier, & Swanson, 2003; W. J. Swanson, Yang, Wolfner, & Aquadro, 2001). To better understand the nature of such coevolutionary interfaces between male and female traits, research must focus on characterizing the identities and functions of reproductive traits in both sexes.

Stages of the reproductive process involve different suites of behavioral, morphological, and/or molecular traits. Work investigating pre-zygotic reproductive isolation has typically concentrated on pre-mating isolation (e.g., differences in courtship behavior or mating preferences) or post-mating pre-zygotic (PMPZ) interactions. PMPZ interactions prior to fertilization can play a critical role in reproductive physiology and subsequent reproductive interactions (Ahmed-Braimah, 2016; Avila & Wolfner, 2017; McDonough et al., 2016; Villarreal et al., 2018), through mechanisms such as sperm competition, male manipulation of females, and cryptic female choice (Firman, Gasparini, Manier, & Pizzari, 2017; Perry et al., 2013; M. Rowe et al., 2015). This suggests a potentially important but understudied role for PMPZ traits in sexual selection and speciation.

Here, I describe a series of studies aimed at characterizing female traits involved in a key PMPZ interaction common to butterflies and moths (Lepidoptera): the digestion of the male ejaculate by the female reproductive tract. During mating, lepidopteran males transfer a complex ejaculate called a spermatophore to the female reproductive tract that includes both sperm and a large bolus of proteins, carbohydrates, lipids, and other substances (Marshall, 1985; C. Meslin et al., 2017). During copulation, the male forms this spermatophore inside the female within a specialized reproductive organ in the female reproductive tract called the bursa copulatrix (hereafter bursa) (C. Meslin et al., 2017; Rogers & Wells, 1984). Shortly after mating, the sperm migrate out of the bursa into the spermatheca, the sperm storage organ (Rutowski & Gilchrist, 1986), leaving the bursa to digest the spermatophore proteins without jeopardizing the viability of the male gametes. The resulting spatial separation of gametes from the rest of the spermatophore is particularly useful in the context of investigating PMPZ physiology due to the ease of isolating interactions that occur between specific male and female proteins within the female reproductive tract, independent of other processes such as fertilization. Following copulation, females process the spermatophore and use the proteins contained within to fund egg production and somatic maintenance (Boggs & Gilbert, 1979), a cooperative interaction that increases fitness for both the male and the female. However, spermatophore digestion is also the subject of conflict between the sexes over female remating (Arnqvist & Nilsson, 2000; Karlsson, 1998). Females will not remate until they have sufficiently digested a spermatophore, a process that they monitor with dedicated stretch receptors on the bursal wall (Sugawara, 1979, 1981). Thus, male spermatophore traits that reduce spermatophore digestion rate may provide males fitness benefits by delaying female remating and therefore extending the period of time where his sperm may preferentially fertilize her eggs. Conversely, female traits that increase the rate of spermatophore digestion allow females

to remate more quickly, allowing females greater control over their reproductive rate, increased access to additional spermatophore nutrition, and increased genetic diversity in their offspring.

Although it is clear from past studies that the female absorbs the contents of the male spermatophore (Boggs & Gilbert, 1979), the female traits involved in spermatophore digestion remain an area of active investigation (Meslin, *et al.* 2015, Plakke *et al.* 2015). One mechanism that females may use to access the stored protein of the spermatophore, proposed by Plakke *et. al.* (2015), involves protein digestion by proteases secreted into the bursa. Plakke *et. al.* (2015) detected protease activity in the bursa copulatrix of the Cabbage White butterfly, *Pieris rapae,* and used a combination of proteomic, transcriptomic, and bioinformatic methods to identify nine putative proteases. Two of the proteases possessed sequence motifs similar to trypsin-like enzymes, which are serine-class proteases. Another set of five proteases possessed papain-like sequence motifs, which suggested they might be cysteine-class proteases. However, the contribution of each of these proteases to observed spermatophore digestion remained unknown. In addition, although sequence homology offers bioinformatic predictions of how these bursal enzymes might function as proteases, these bioinformatic predications remained untested experimentally.

To address these gaps in my knowledge, I used a zymogram approach to identify active proteases in the bursa and to investigate their proteolytic modes of action. Following separation by native polyacrylamide gel electrophoresis (PAGE), I determined protein identities via mass spectrometry. I then investigated their modes of action using diagnostic protease inhibitors. I followed these functional assays with homology modeling to explore possible enzyme structures and their implications for protease function and inhibition. By combining my experimental

findings with these modeling approaches, I were able to identify putative mechanisms of proteolytic activity that will serve as working hypotheses for future studies.

## 3.2 Methods

### 3.2.1 Experimental Animals

Experimental animals were the F1 female offspring of wild female *Pieris rapae rapae* Linnaeus 1758 collected from an agricultural site in Rochester, PA (40°44′44.4″N 80°09′49.0″W) in the summers of 2016 and 2018. F1 offspring were reared in climate-controlled chambers that maintained a 16h:8h light:dark photoperiod with a constant temperature (24°C) and relative humidity (60% RH). Larvae were fed *ad libitum* on young *Brassica oleracea* leaves. Upon eclosion, females were housed in individual containers within the chambers until they were used for experiments.

### 3.2.2 Extract Preparation

Proteases were collected from the bursae of 3-day-old virgin females. Bursae were removed by vivisection in phosphate buffered saline (PBS). Individual bursae were then placed whole in 100 µl of PBS and homogenized with 50 turns of a clean, disposable pestle. Pestles were not reused. Homogenized solutions were centrifuged at 10,000 g for 15 minutes. Supernatant was removed from the resulting pellet and subsequently stored at -20°C until assayed.

### 3.2.3 Zymogram Identification of Active Proteases

I used a modified zymogram technique to identify the active proteases within each bursal sample (Raser, Posner, & Wang, 1995). 10 µl of homogenized enzyme solution was combined with 10 µl of native gel loading buffer (150 mM Tris-HCl 6.8 pH, 20% glycerol, 0.004% Bromophenol blue) before being incubated at 37°C for 15 minutes. 8% resolving native PAGE gels with 5% stacking gel, both lacking sodium dodecyl sulfate (SDS), were pre-run for 15 minutes at 125 V before samples were loaded. Gels with samples were run at 125 V for 3 hours at 4°C in native running buffer (25 mM Tris-Base, 19 2mM glycine, 1 mM DTT, 1mM EDTA). After running, gels were rinsed in de-ionized (DI) water before being soaked in casein solution (500 µM casein, 3.6 mM $CaCl_2$, pH 7) for 60 minutes. This impregnated the gels with casein. Gels were then rinsed in DI water again, stained with Coomassie Blue (50% methanol, 10% acetic acid, 1 g Coomassie Brilliant Blue R-250, Amresco, Solon, OH, USA) for 20 minutes, then de-stained (20% isopropanol, 7% acetic acid) for 40 hours. Clear areas of the resulting stained gels represented locations where active proteolysis had digested all casein (and therefore reduced Coomassie Blue staining). Clear (unstained) 0.25 $cm^2$ bands caused by proteolytic activity were excised and submitted to the Biomedical Mass Spectrometry Center at the University of Pittsburgh. As a negative control, I characterized the proteins present in the gel area directly above the highest protease band. In total, ten bands were extracted in this manner, including three representatives of each bursal band plus one control. As bursal extracts from different individuals did not always exhibit all three bands, I randomly chose bursal extracts to run on the gels and collected bands until the target number for each band was reached. I used four bursal extracts in total, with each contributing at least two bands to the analysis (Table 3).

**Table 3: Individual identities for bands extracted for proteomic analysis with the number of spectra recorded for each sample.**

| | α1 | α2 | α3 | β1 | β2 | β3 | γ1 | γ2 | γ3 | Control |
|---|---|---|---|---|---|---|---|---|---|---|
| **Individual #** | 1 | 2 | 3 | 1 | 3 | 4 | 3 | 4 | 2 | 1 |
| **Total spectrum counts from band** | 346 | 297 | 119 | 172 | 144 | 85 | 291 | 90 | 256 | 20 |

The extracted bands were then trypsinized and subjected to liquid chromatography followed by tandem mass spectrometry (Granvogl, Gruber, et al., 2007; Granvogl, Plöscher, et al., 2007; Shevchenko et al., 2006) in conjunction with previously acquired transcriptomic sequences comprised of 15,773 unique components (C. Meslin et al., 2015; Plakke, Deutsch, Meslin, Clark, & Morehouse, 2015) to determine protein identities within the bursal lumen. Proteins were considered present within a band if spectra met a minimum protein identification threshold of 99% and had at least two mapped peptides with a minimum peptide threshold of 90%. False Discovery Rates were calculated using the probabilistic method implemented through the ProteinProphet algorithm (Nesvizhskii, Keller, Kolker, & Aebersold, 2003). Tandem mass spectrometry data were visualized using Scaffold (Proteome Software, Portland, Oregon, USA), with subsequent annotation of identified proteins using BLASTP (Altschul, Gish, Miller, Myers, & Lipman, 1990). I annotated the top hits identified by NCBI BLASP that had E-value scores less than $1 \times 10^{-20}$. Additionally, I identified PFAM domains using HMMER v3.2.1 (hmmer.org) and converted domains with E-values less than 0.01 to Gene Ontology terms using pfam2go (Mitchell et al., 2015). The protein and domain identities obtained were compared to the proteases identified previously in this species (Plakke et al., 2015). Uncorrected spectra counts were used to calculate the proportion each protein comprised of each gel slice.

Additionally, I identified PFAM domains using HMMER v3.2.1 (hmmer.org) and converted domains with E-values less than 0.01 to Gene Ontology terms using pfam2go (Mitchell et al., 2015) (Appendix A). The protein and domain identities obtained were compared to the proteases identified previously in this species (Plakke et al., 2015). Uncorrected spectra counts were used to calculate the proportion each protein comprised of each gel slice.

### 3.2.4 Protease Inhibitor Assay

To determine whether the identified proteases have biochemical activity consistent with their predicted motifs, I modified the above zymogram methods to include exposing focal proteases to class-specific protease inhibitors. To test for serine-like activity, I used phenylmethane sulfonyl fluoride (PMSF), a broad inhibitor of serine-class proteases such as trypsin and chymotrypsin (Gold, 1965). To test for cysteine-like activity, I exposed the proteases to the papain inhibitor leupeptin (Aoyagi et al., 1969). Samples were prepared by mixing 10 µl of enzyme extract with 5µl of native gel loading buffer and 5 µl of one of the following treatments: 100 mM PMSF in dimethyl sulfoxide (DMSO) (serine protease inhibitor treatment), 10 µM leupeptin in water (papain protease inhibitor treatment), DMSO only (PMSF control), or water (leupeptin control). Commercial trypsin (6 mg/ml in PBS equivalent to 7.75 U, Fisher Scientific, Hampton, NH, USA) was prepared in an identical fashion alongside bursal samples. Samples were incubated at 37°C for 15 minutes to allow for inhibition by protease inhibitors (when present) before running in identical conditions to the methods above. After destaining, gels were imaged at 600 dpi (Canon 9000F Mark II scanner, Tokyo, Japan). I quantified the cleared areas from gel scans using ImageJ (Schneider, Rasband, & Eliceiri, 2012). The extent of clearing caused by commercial trypsin was measured in Active Enzyme Units (U) based on the standard amount of known enzyme and activity

loaded into the gel. Clearing of sample bands were converted to units of activity based on their measured intensities relative to the standard trypsin band. This method allowed me to observe the effects of commercial trypsin down to a concentration of 1.5 mg/ml in PBS equivalent to 1.94 U. Due to the irreversible nature of the chosen protease inhibitors, increases in inhibitor concentration did not increase inhibition, indicating saturation (data not shown).

### 3.2.5 Homology Modeling

Using the program I-TASSER (Yang et al., 2014), I created homology models of three of the bursal proteases identified via mass spectroscopy: one trypsin-like protease (BTLP1) and two papain-like proteases (BPLP1 and BPLP2). To structurally assess PMSF binding to my modeled protease structures, I used PyMOL (open-source, version 2.1.0) to align a crystal structure of PMSF-bound *Fusarium oxysporum* trypsin (PDB ID: 1PQA) (Schmidt, Jelsch, Østergaard, Rypniewski, & Lamzin, 2003) to the BTLP1 homology model. To study the leupeptin binding pose, I used PyMOL to align a crystal structure of leupeptin-bound *Carica papaya* papain (PDB ID: 1POP) (Schröder, Garman, Harlos, & Crawford, 1993) to my BPLP1 and BPLP2 homology models. Figures of these models were generated using BlendMol (Durrant, 2018). I selected structures of *F. oxysporum* and *C. papaya* proteases as references because they are the classic structures associated with their respective classes and so are particularly well characterized.

### 3.2.6 Statistical Analysis

All statistical analyses were conducted using the statistical program SPSS (version 25.0, IBM Corp, Armonk, NY). *t*-tests were performed to compare treatments to their appropriate

control for each bursal band following Levene's Test for Equality of Variances.

## 3.3 Results

### 3.3.1 Identification of Active Proteases within the Bursa

Resolved zymograms of bursal samples showed proteolytic activity against a general protein substrate, casein, in three distinct bands, henceforth designated Bursa α, Bursa β, and Bursa γ (Figure 5). All samples exhibited at least one of the bands (n = 19), though the number and combination of bands varied between individuals. Of the nineteen individuals sampled for activity, six exhibited all three bands, eleven exhibited only two bands, and two exhibited only one band. For those with two bands, nine exhibited Bursa α/β, and two exhibited Bursa β/γ. Both individuals with solitary bands exhibited Bursa α. From the extracted bands, fifty-two distinct proteins, including six predicted proteases, were recovered. Proteases represent 52.94 ± 6.869%, 68.89 ± 5.38%, and 84.49 ± 3.66% of total protein spectra (means ± standard error) for Bursa α, Bursa β, and Bursa γ, respectively.

The observed proteases correspond to proteases that my research team has previously identified and classified (Plakke et al., 2015). Predicted classes were further supported by PFAM domain identification. From the bands sampled, I observed one trypsin-like protease (Bursal Trypsin-Like Protease 1, or BTLP1), two general peptidases (Bursal General Peptidase 1 and 2, or BGP1, BGP2), and three papain-like proteases (Bursal Papain-Like Protease 1, 2, and 3, or BPLP1, BPLP2, BPLP3). Bursa α was comprised primarily of general and papain-like proteases, BGP1 and BPLP1 (Table 4). Bursa β was comprised mainly of trypsin- and papain-like proteases, BTLP1

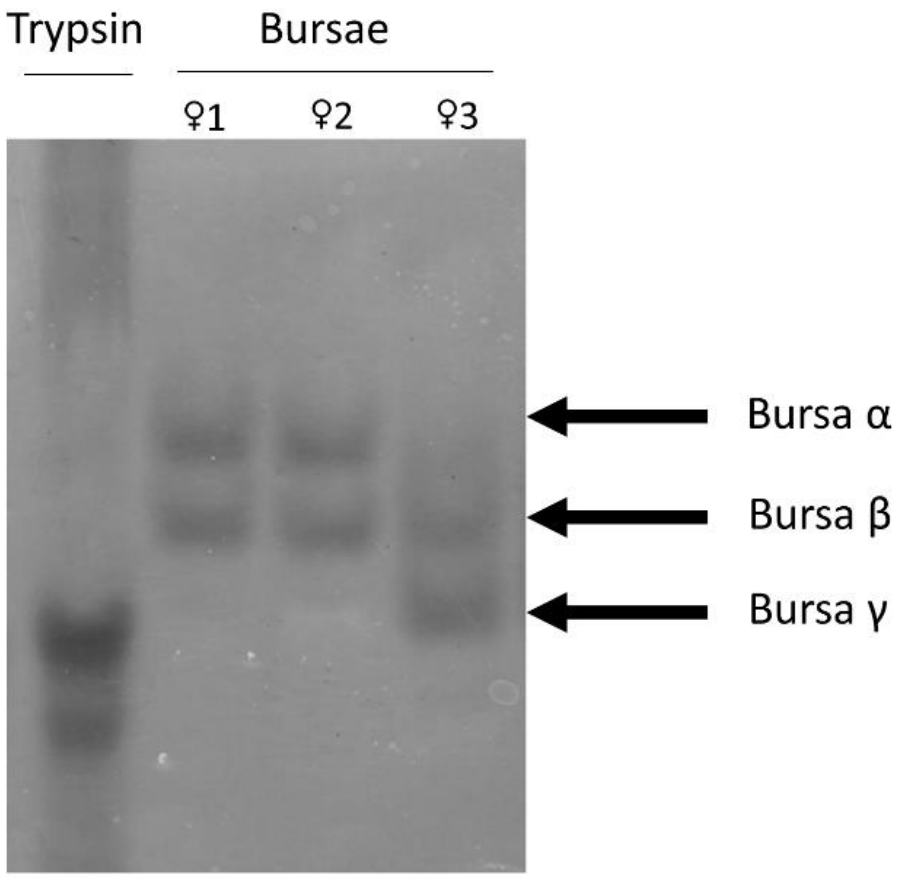


**Figure 5: In virgin butterflies, bursal proteolytic activity increases with female age**

Data are presented as means +/- 95% confidence intervals. Lowercase letters indicate statistical groupings.

and BPLP1 (Table 4). Bursa γ was comprised of a combination of BTLP1 and BPLP2. Mass spectrometry analysis of the control area on the gels returned six proteins. None of these had identifiable protease domains according to PFAM and gene ontology analysis. Other proteins recovered along with the proteases included cytoskeletal and muscle-related proteins as well as proteins involved in general cellular structure and function. These proteins are all expected to be present in my samples as a result of my non-specific bursal extraction techniques.

### 3.3.2 Response of Bursal Proteases to Common Protease Inhibitors

Though I detected a trypsin-like protease (BTLP1) in Bursa β and Bursa γ, PMSF did not significantly reduce the protease activity of any Bursa band (α: $t = 0.635$, $df = 20$, $p = 0.533$; β: $t = 1.808$, $df = 22$, $p = 0.084$; γ: $t = 0.382$, $df = 5$, $p = 0.718$) (Figure 6A). This is in stark contrast to the commercial trypsin control, which was inhibited significantly ($t = 29.512$, $df = 5$, $p < 0.001$). These results suggest that bursal trypsin BTLP1 is not detectably inhibited by PMSF. Leupeptin almost entirely abolished the protease activity of Bursa α, with little effect on Bursa β or Bursa γ (α: $t = 5.011$, $df = 3.147$, $p = 0.014$; β: $t = -0.273$, $df = 8$, $p = 0.792$; γ: $t = 0.654$, $df = 8$, $p = 0.532$) (Figure 6B). I detected far more papain-like protease (BPLP1) in Bursa α than trypsin-like protease (BTLP1), suggesting that the observed reductions in protease activity in Bursa α are most likely the result of BPLP1 inhibition. In contrast, I also detected considerable amounts of a second papain-like protease (BPLP2) in Bursa γ, yet leupeptin did not affect that band (Table 4). These results suggest that BPLP2—but not BPLP1—is resistant to leupeptin activity. Interestingly, though I detected BPLP1 in Bursa β, I saw no leupeptin inhibition.

**Table 4: Protease identities of zymogen bands.**

Of the nine proteases identified in (Plakke et al., 2015), the following six were recovered through proteomic analysis of excised bands of activity on the zymograms. Percentages represent the proportion of all identified protease peptides by spectra counts within the respective excised band.

| **Protease** | **Comp** (C. Meslin et al., 2015) | **Bursa α** (N = 3) Mean ± SE | **Bursa β** (N = 3) Mean ± SE | **Bursa γ** (N = 3) Mean ± SE | **Control** (N = 1) Mean ± SE | **Predicted Mode of Action** |
|---|---|---|---|---|---|---|
| BTLP1 | Comp83824_c0 | 4.95 ± 0.52% | 31.41 ± 5.64% | 10.70 ± 2.51% | 0% | Trypsin-like |
| BTLP2 | Comp93091_c0 | 0% | 0% | 0% | 0% | Trypsin-like |
| BPLP1 | Comp91676_c0 | 25.78 ± 2.96% | 21.54 ± 1.39% | 7.66 ± 0.60% | 0% | Papain-like |
| BPLP2 | Comp85455_c0 | 3.65 ± 1.12% | 11.60 ± 2.42% | 48.72 ± 11.77% | 0% | Papain-like |
| BPLP3 | Comp83827_c1 | 4.94 ± 3.40% | 0% | 0.39 ± 0.39% | 0% | Papain-like |
| BPLP4 | Comp94445_c1 | 0% | 0% | 0% | 0% | Papain-like |
| BPLP5 | Comp95264_c1 | 0% | 0% | 0% | 0% | Papain-like |
| BGP1 | Comp98020_c0 | 11.33 ± 1.10% | 0.97 ± 0.97% | 2.68 ± 1.50% | 0% | Unknown |
| BGP2 | Comp97068_c0 | 2.29 ± 1.36% | 3.37 ± 1.12% | 14.35 ± 7.32% | 0% | Unknown |

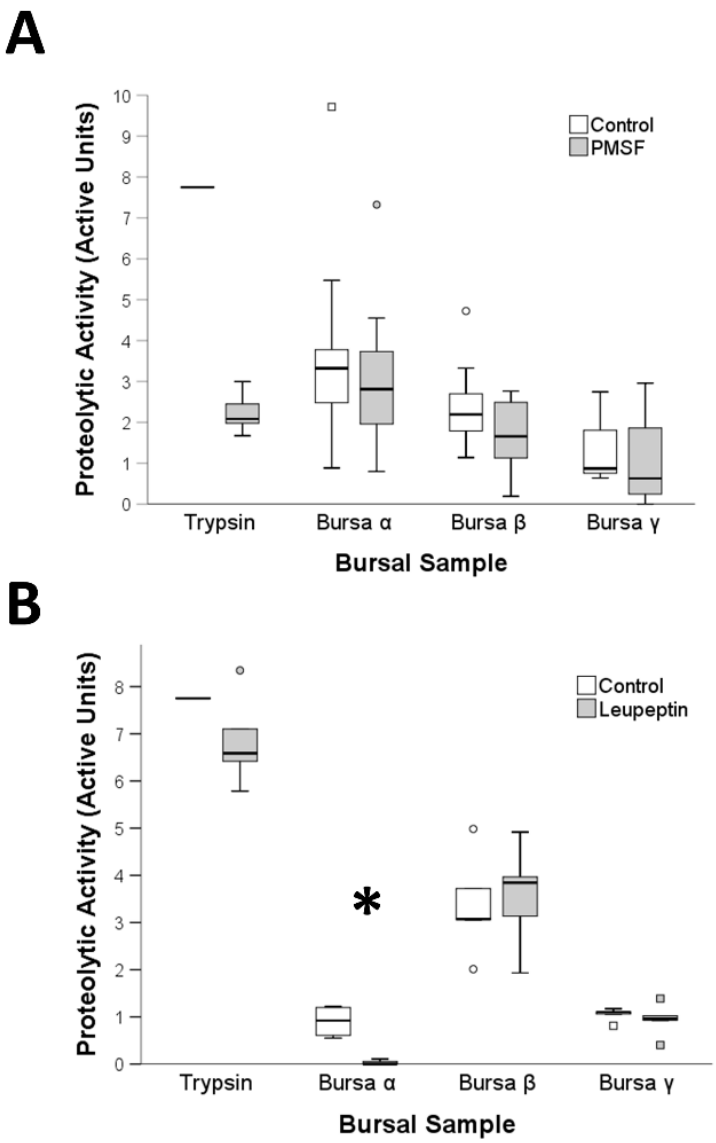


**Figure 6: Inhibition of zymogen bands by commercial inhibitors**

Cleared bands in the zymograms were quantified in the presence and absence of the protease inhibitors leupeptin and PMSF. (A) The trypsin-specific protease inhibitor (PMSF) inhibits the activity of commercial trypsin, but does not inhibit the activity of bursal proteases in Bursa α, Bursa β, or Bursa γ. (B) The papain-specific inhibitor (leupeptin) does inhibit the activity of Bursa α, but not the activity of Bursa β or Bursa γ.

### 3.3.3 Modeling Bursal Protease/Inhibitor Binding

When compared to *F. oxysporum* trypsin, the active site of BTLP1 showed distinct differences that are partially localized to a key pocket-adjacent loop connecting two beta strands. In *F. oxysporum* trypsin, this loop spans twelve residues (W212-G223) and does not occlude the active site. In contrast, the homologous BTLP1 loop region contains fourteen residues (Figure 7A-B). Our *P. rapae* BTLP1 homology model suggests that this longer loop protrudes into the binding pocket, limiting access to the catalytic triad (Figure 7A-B). The crystallographic PMSF pose does in fact clash with the modeled BTLP1 loop when the proteins are superimposed (Figure 7A-B). Thus, consistent with my experimental results, my computational model predicts that PMSF should not inhibit the proteolytic activity of BTLP1.

I next considered the *P. rapae* papain-like proteases (BPLP1 and BPLP2). Structural differences between these two proteases (Figure 7C-D) may explain why leupeptin inhibits BPLP1 but not BPLP2. To visualize leupeptin in the context of the BPLP1 homology model, I aligned my BPLP1 model to a crystal structure of the papain/leupeptin complex (PDB ID: 1POP) (Schröder et al., 1993). The open active site of the BPLP1 model can accommodate leupeptin binding (Figure 9C-D). In contrast, the homology model of BPLP2 is not compatible with leupeptin binding (Figure 7C-D). An extended loop runs along the catalytic cleft, occupying the region that normally binds leupeptin. This model resembles the inactive zymogen form of papain (e.g., PDB ID: 3TNX) (Brocklehurst & Kierstan, 1973; Roy, Choudhury, Aich, Dattagupta, & Biswas, 2012), prior to activation. It is curious that the predominant papain-like protease in Bursa γ is zymogen-like. I note that BPLP2 is not merely pro-BPLP1, as there are other amino-acid differences between these two bursal, papain-like proteases. Deleting pre-protein domain from the model eliminates the

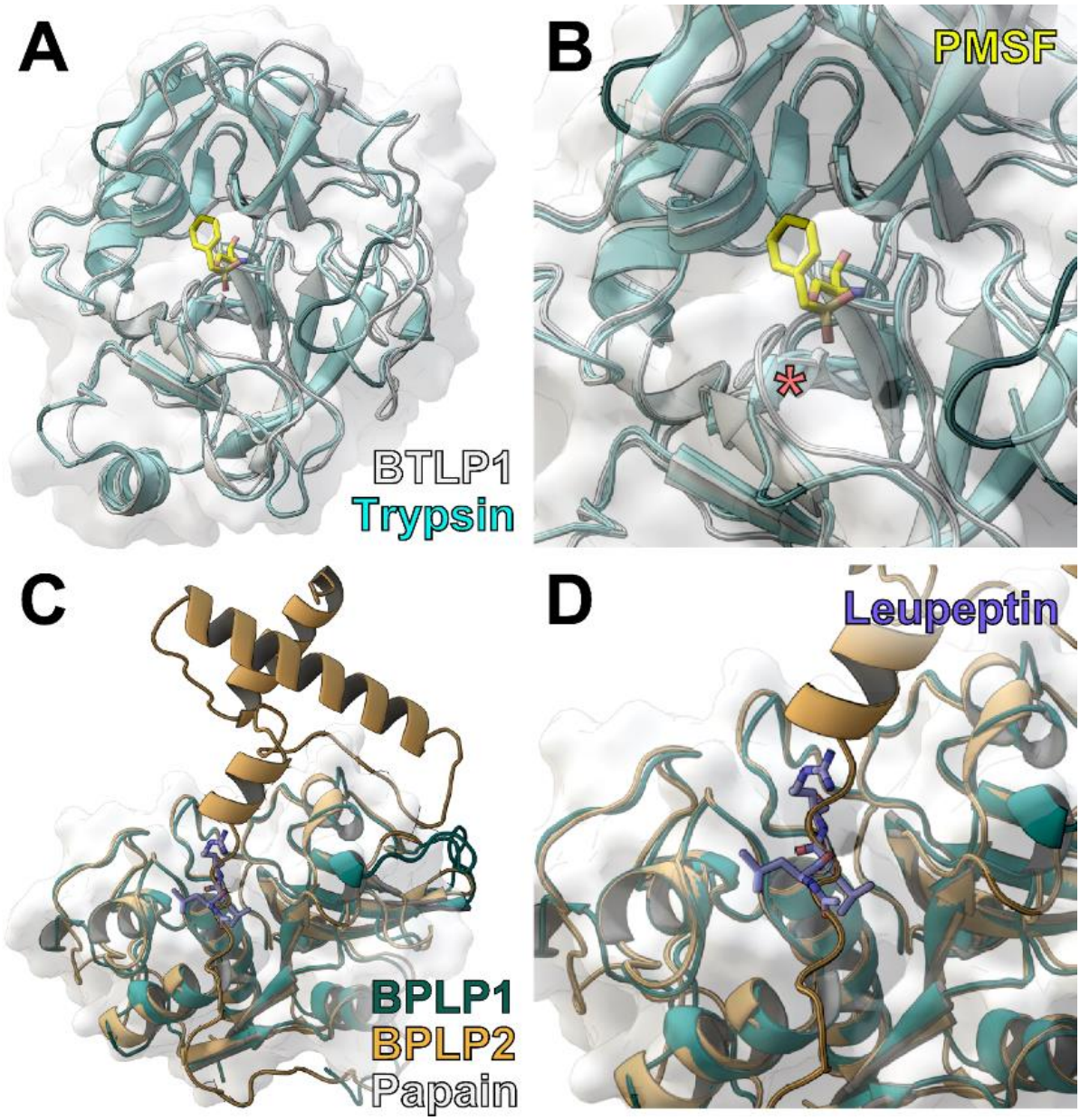


**Figure 7: Predicted binding of inhibitors to bursal proteases**

(A) The crystal structure of *F. oxysporum* trypsin, 1PQA, bound to PMSF (cyan ribbon and yellow sticks, respectively), superimposed on the BTLP1 homology model (white ribbon). (B) A zoomed in view of the active site of BTLP1. The red asterisk marks a loop that is extended in the BTLP1 model. In BTLP1, this loop sterically clashes with the crystallographic ligand pose.

(C) Homology models of BPLP1 and BPLP2 are shown in dark-teal and gold ribbon, respectively. To position the leupeptin inhibitor (purple sticks), I aligned the 1POP holo structure of *C. papaya* papain (protein not shown). The BPLP1 model includes an open cleft that can accommodate the inhibitor. The BPLP2 model has an inhibitory domain typical of inactive (zymogen) papain, which occupies the cleft and is incompatible with leupeptin binding. (D) A zoomed in view of the active sites for BPLP1 and BPLP2 with leupeptin superimposed.

occlusion of the BPLP2 active site, suggesting that BPLP2 may respond to leupeptin upon activation (Figure 8). The BPLP1 and BPLP2 models thus predict modes of action consistent with my zymogram experiments, potentially explaining why only BPLP1 is susceptible to leupeptin inhibition.

## 3.4 Discussion

This study characterizes the proteases active within the bursa copulatrix of virgin *P. rapae*. Building on previous work (Plakke et al., 2015), I characterized the activity and modes of action of the proteases that exhibited *in vitro* activity in my bursal extracts. Through zymogram analysis, I observed band variation across females, potentially explaining the variation in total activity observed in Plakke *et al* (2015). Of the nine previously predicted proteases in the bursa, six were recovered in my native PAGE experiments. These proteases are predicted to belong to families of serine-, cysteine-, and general peptidase-class proteases (C. Meslin et al., 2015; Plakke et al., 2015). Trypsin-like proteases are commonly described as being present in the female reproductive tracts of a variety of organisms (i.e. Diptera (Alfonso-Parra et al., 2016; Kelleher & Pennington, 2009; Lawniczak & Begun, 2007), Lepidoptera (Al-Wathiqui, Lewis, & Dopman, 2014; C. Meslin et al., 2015; Plakke et al., 2015), Mammals (Ou et al., 2012), etc.) based on predictions from sequence homology. However, these proteases and their modes of action are rarely studied biochemically in a reproductive context. I show that these bursal proteases are present *in vivo* and functional *in vitro*.

Sequence homology suggested that the protease BTLP1, found in Bursa β and γ bands, should exhibit serine-protease-like activity (Table 4). However, activity of this protease was largely unaffected by PMSF, a serine-specific protease inhibitor. Our homology modeling suggests

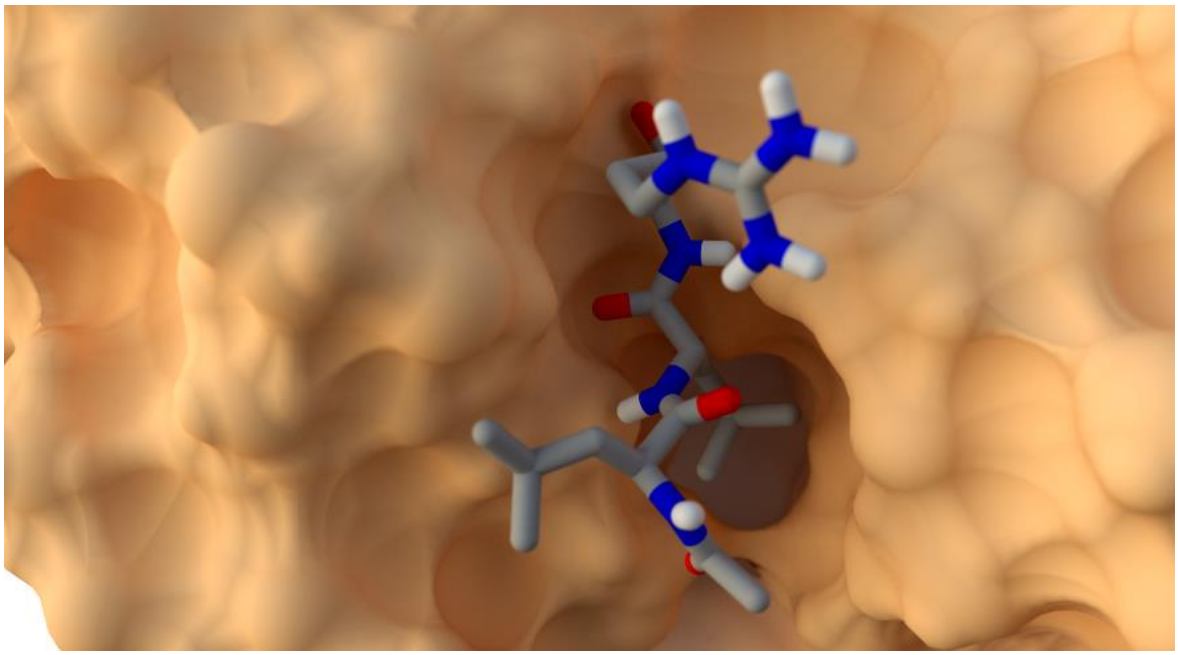

**Figure 8: Potential binding of leupeptin to modified BPLP2**

The BPLP2 homology model in the active form. The extended loop that otherwise occupies the catalytic cleft was removed. The crystallographic pose of the leupeptin inhibitor was taken from the 1POP structure.

that an extended BTLP1 loop blocks inhibitor access to the active site, potentially contributing to PMSF resistance (Figure 7A-B). It is possible that this structural difference renders the protease incapable of digesting casein and therefore not responsible for the activity reported in my assay. However, this seems highly unlikely. Casein lacks tertiary structure and is generally digestible by all classes of proteases. Furthermore, high concentrations of BTLP1 were recovered within the Bursa β band of activity. I propose instead that BTLP1 is resistant to some modes of serine-specific protease inhibition and hypothesize that this resistance may be biologically relevant. While I did not recover any protease inhibitors from the bursal bands, such inhibitors are known to be present within female reproductive tracts (Al-Wathiqui et al., 2014; Dong et al., 2016; Plakke et al., 2015; Prokupek, Eyun, Ko, Moriyama, & Harshman, 2010). Further, males are known to transfer protease inhibitors to the female, together with the ejaculate and sperm (LaFlamme & Wolfner, 2013) and have been documented across lepidopteran ejaculates (Al-Wathiqui, Lewis, & Dopman, 2017; Dong et al., 2016). The female proteases described here may be under selective pressure to counter this inhibition to retain control over spermatophore digestion, causing resistance to classically described inhibitors as well.

Sequence homology also identified the proteases BPLP1 and BPLP2 as being papain-like cysteine proteases. BPLP1 was prevalent in Bursa α and Bursa β (Table 4). The papain-specific inhibitor leupeptin did in fact decrease the proteolytic activity of Bursa α, suggesting that BPLP1 may be leupeptin sensitive and that it may contribute to active digestion. However, Bursa β, paradoxically, did not respond to leupeptin, despite the fact that it also includes large amounts of BPLP1. I hypothesize that post-translational modifications may confer BPLP1 resistance to leupeptin in Bursa β. These modifications can alter protein charge and size and so may also explain why the protease is found at multiple locations along the native PAGE gel. I also considered the

possibility that Bursa β BPLP1 is in the inactive zymogen form (Brocklehurst & Kierstan, 1973), but mass spectrometry recovered no peptide spectra consistent with a papain-like zymogen inhibitory domain (Figure 9). Indeed, the BPLP1 spectra across bands were nearly identical. I also considered the possibility that leupeptin migrates with the Bursa β group, actively inhibiting BPLP1 in Bursa β but not Bursa α. However, the mass spectrometry methodology I used is unable to detect leupeptin, as it is not comprised of amino acids and is thus invisible to such tests. Additional work is needed to further explore these possibilities.

The activity of BPLP2, a papain-like protease that is prevalent in Bursa γ, is unaffected by leupeptin. Our BPLP2 homology model suggests a potential explanation for this resistance. The BPLP2 model resembles the inactive zymogen form of papain, in which a loop occludes leupeptin (and substrate) binding (Figure 7C-D). Additional studies are required to determine whether BPLP2 is converted to an active from in *P. rapae*, as in other species (Yamamoto, Kurata, Watabe, Murakami, & Takahashi, 2002). Alternatively, BPLP2 might be activated at a different time in the mating process. All samples used for this study were acquired from virgin bursal tissues, but the possibility exists that the loop is cleaved upon contact with male ejaculate proteins after mating, by proteins of either male or female origin.

The proteases of the bursa are directly responsible for the degradation of male-derived spermatophore proteins (Plakke, unpublished data). Given that these spermatophore proteins are rapidly evolving (C. Meslin et al., 2017), bursal protease characteristics may be the result of dynamic coevolution with their ejaculate protein substrates, which may have been shaped by either cooperation or sexual conflict. Sexual conflict over spermatophore digestion arises because female *P. rapae* are polyandrous and exhibit last male sperm precedence (Bissoondath & Wiklund, 1997; Suzuki, 1979). Thus, male traits that reduce female remating rate are favored by selection on male

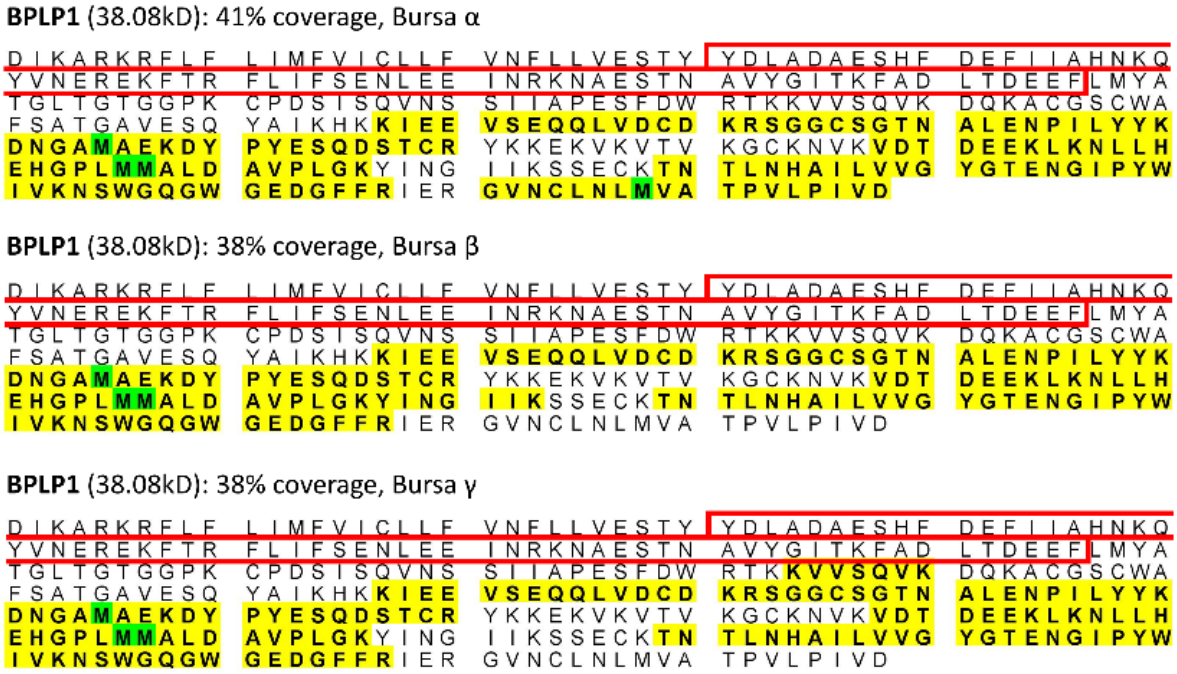


**Figure 9: Spectra recovered for BPLP1**

BPLP1 was identified in all three Bursa band groups through mass spectrometry. The peptides recovered (yellow highlight) corresponded to 38-41% of all amino acids in the sequence. No spectra corresponding to the pro-protein domain (red box) were recovered from any sample.

fitness, including traits that reduce spermatophore digestion rate (Sánchez & Cordero, 2014b). However, females are likely to benefit from increases in remating rates because mating represents a source of both valuable nutrition and additional gametes that increase the genetic diversity of their offspring. This inherent tension over female remating rate may thus be mediated by interactions between the spermatophore and the bursa, with antagonistic coevolution favoring male traits that reduce the digestibility of the spermatophore, and female traits that increase the rate of proteolysis within the bursa. On the other hand, female proteases may coevolve cooperatively with the spermatophore proteins if the rapid evolution of the male proteins were driven by non-antagonistic forces, such as changes to the nutritional ecology of a particular species or population. If either dynamic were the case, I would expect the female to adopt one of two strategies to respond to the constantly changing male proteins provided in the ejaculate: 1) the bursal proteases could evolve very general activity to digest any substrates the male provides, or 2) the proteases could rapidly evolve their specificity in conjunction with the rapidly evolving ejaculate proteins. In the second scenario, specialization of each of the bursal proteases could enable overall broad proteolytic activity while increasing proteolytic rates, as is commonly observed in other digestive organs (Patankar et al., 2001). Our current data cannot differentiate between these two hypotheses, but future studies could address these questions using targeted co-incubations of bursal proteases and spermatophore proteins.

The current study reveals that bursal proteases vary in susceptibility to different inhibitors. This suggests that the protease active sites may exhibit specificity for their target substrates. Alternatively, they may have evolved resistance to protease inhibitors, either of male or female origin, with mechanisms of action similar to those of the commercial inhibitors I tested. These interpretations, however, hinge on the point in evolutionary time when the proposed resistance-

conferring structural changes evolved. Several of the bursal proteases evolved from duplicated proteases expressed in digestive tissues such as the caterpillar gut (C. Meslin et al., 2015), where the proteases may have evolved in response to inhibitors presented through the diet (R. Broadway, 1996). It is currently unknown whether the proposed structure-mediated resistance evolved before or after these proteases were first expressed in the bursa, and therefore the evolutionary pressures behind the altered structure is currently unknown.

Regardless of the origin of the bursal protease structures, their varied specificity could have long-reaching evolutionary consequences. If female protease specificity evolves with male spermatophore proteins, then isolated populations may, by either selection or chance alone, evolve varying responses at this digestive interface. Secondary contact by such populations could result in reproductive mismatches between the sexes, leading to decreased fitness and the development of reproductive barriers.

In conclusion, I have successfully identified and quantified the female half of a male/female reproductive interface important to key reproductive outcomes in the butterfly *P. rapae*. By characterizing the active proteases that mediate a PMPZ interaction, this study paves the way for future manipulative and comparative studies of reproductive protein evolution and coevolution between the sexes. In particular, it provides a targeted set of proteases with known activity. By integrating biochemical and homology modeling approaches, my work unveils important molecular details beyond what bioinformatic surveys alone can reveal. Future studies focusing on antagonistic coevolution and reproductive isolation from the perspective of genic, regulatory, and population-level variation could benefit from such approaches.

## 3.5 Acknowledgements

I would like to thank Kretschmann Farms for allowing me to collect butterflies in their cabbage fields. Additionally, a very heartfelt thank you to Roni Lahr for providing protease inhibitors and Kelly Dulin for help with method development. I would like to thank Sigma Xi, the Scientific Research Society, for funding this project. I also thank HHMI for funding to BJG and JBL. This project used the Biomedical Mass Spectrometry Center and University of Pittsburgh Cancer Institute Cancer Biomarker Facility, which are supported, in part, by NCI Grant P30CA047904.

# 4.0 Targets of female-mediated digestion in a protein-rich ejaculate

## 4.1 Introduction

Ejaculates perform a number of key functions during the reproductive process (Perry et al., 2013). Because ejaculates are often directly tied to fitness outcomes, ejaculates are expected to evolve, sometimes rapidly, under selection for improved function (N. L. Clark et al., 2006; Wolfner, 2002). Observations of variability in ejaculate protein composition support this expectation, and are suggestive of the evolution of novel functions (Dean et al., 2009; W. Swanson & Vacquier, 2002). Likewise, recent work investigating ejaculate structural complexity indicates that this axis of variation also evolves rapidly, implying that ejaculate structure may likewise be under functional selection (C. Meslin et al., 2017).

The Cabbage White butterfly, *Pieris rapae*, has recently been identified as an organism whose ejaculate is not only comprised of rapidly evolving and diverse proteins, but also exhibits structural complexity (C. Meslin et al., 2017). In addition to the male ejaculate proteins, the proteins present in the female reproductive tract have also been described. Many of the female proteins that interact with the male ejaculate have now been characterized both functionally and biochemically (Camille Meslin et al., 2017; Plakke et al., n.d., 2015). During the process of reproduction, male *P. rapae* transfer and deposit a large proteinaceous spermatophore within a specialized organ within the female reproductive tract. This female organ is called the bursa copulatrix (Engelmann, 1970). The spermatophore is comprised of primarily protein and can be separated into three distinct structural regions: the tough outer envelope (73.3% protein by dry

mass), the soft inner matrix (48.8% protein), and the bolus of sperm (Marshall, 1985; C. Meslin et al., 2017; Minoru Osanai et al., 1987).

The different layers of the spermatophore are hypothesized to be under divergent selective pressures that are related to the main functions of the spermatophore (McNamara, Dougherty, Wedell, & Simmons, 2019). The ease of digestion of the inner matrix is expected to support a cooperative venture between the sexes. The spermatophore provides a large nutrient gift to the female while mating, which the female uses to increase egg production and support her cellular maintenance (Boggs & Gilbert, 1979; M Watanabe & Ando, 1993; Wiklund et al., 2001). In contrast, the spermatophore envelope proves so resistant to digestion that a female is never able to digest the outer envelope in its entirety (Burns, 1968). The indigestibility of the spermatophore functions to limit female remating. After receiving a spermatophore, a female is unable to mate until she has adequately reduced the volume of the spermatophore (K. Oberhauser, 1992; Karen S. Oberhauser, 1989; Sugawara, 1981; Wiklund et al., 2001). Due to *P. rapae* experiencing last sperm precedence, the longer it takes the female to digest a spermatophore, the longer a male will have monopolized paternity of her eggs (Wedell & Cook, 1998). The third structural region, the bolus of sperm, is situated in close proximity to the spermatophore neck. Within twenty minutes after mating ends, the sperm will migrate to the sperm storage organ, the spermatheca (Rutowski & Gilchrist, 1986). This rapid migration from a hostile environment is mediated by the proximity of the bolus to the main female reproductive tract.

Both the cooperation over funding egg production and the conflict over the envelope's digestion involve the degradation and digestion of the spermatophore proteins. It has been suggested in the domestic silkmoth, *Bombyx mori*, that the spermatophore proteins self-catalyze in a time-release fashion (Minoru Osanai et al., 1987). Previous studies in *P. rapae* have identified

several highly active proteases in the female's bursa (Plakke et al., 2015), leading me to hypothesize that the digestion of the spermatophore is not entirely driven by male enzymes, but in fact primarily female-mediated. Further, I would expect that male spermatophore proteins that are functioning to delay digestion would be selected to resist degradation by female proteases. Therefore, I would expect spermatophore structural proteins of the envelope to be digested at a slower rate than proteins that are beneficial to both sexes. Of particular interest are the two proteins that comprise a large proportion of the outer spermatophore envelope, PRSP1 and PRSP2 (C. Meslin et al., 2017). These two proteins contribute to the structure and insolubility of the spermatophore envelope and are therefore likely to be under selection to resist degradation by bursal proteases.

In this current study, I aim to 1) identify the role female enzymes that contribute to the digestion of the spermatophore, 2) identify which ejaculate proteins are targeted for degradation by bursal proteases, and 3) characterize the rate of digestion for each ejaculate protein. In order to answer these three questions, I incubated spermatophore proteins in isolation as well as in combination with female bursal extracts over a time series. The amount of each spermatophore protein present at each time point were assessed using differential mass spectrometry analysis for proteomic identification. Digestion rates of spermatophore proteins were then compared across the layers of the spermatophore in order to assess proteins targeted for degradation by the female.

## 4.2 Methods

### 4.2.1 Animal Rearing

Experimental animals were the F1 offspring of wild female *Pieris rapae rapae* Linnaeus 1758, collected from an organic farm in Rochester, PA (40°44′44.4″N 80°09′49.0″W) in the summer of 2016. F1 offspring were housed in chambers that maintained a 16h:8h light:dark photoperiod with a constant temperature (24°C) and relative humidity (60%). Larvae were fed *ad libitum* on young *Brassica oleracea* leaves. Upon eclosion, males and females were housed in individual containers within the climate chambers until they were used for experiments. Matings to collect spermatophores were conducted in a 60 cm × 60 cm × 90 cm insect mating enclosure placed in a greenhouse providing direct sunlight. Mating enclosures were checked every 20 minutes to ensure no matings were missed. Upon copulation, male-female pairs were removed from the mating enclosure, removed from direct sunlight, and placed in a separate cup until separation. Females were frozen at -80°C within one hour of separation in order to preserve the spermatophores for subsequent experiments.

### 4.2.2 Protein Collection

Bursal proteases were collected as previously described (Plakke et al., 2015) with slight modification. In brief, three-day-old virgin females were micro-vivisected in PBS to remove the bursal tissue. Five bursas were homogenized in 300 µL of PBS to extract bursal proteases using clean, single-use manual pestles and stored at -20°C until needed for experiments. To collect spermatophore proteins, I dissected twenty-seven intact spermatophores from frozen, mated

females. Spermatophores were pooled in groups of three with 50 µl of PBS per spermatophore before homogenization. Spermatophores were homogenized manually using a single-use pestle and then additionally exposed to sonication by a Tissue-Tearor (BioSpec, Bartlesville, OK, USA) at 32,000 rpm. Sonication occurred by eight separate pulses of sonication for fifteen seconds interspersed with thirty seconds of non-sonication in between each pulse for a total of two minutes of sonication. Samples were kept on ice through the entire process in order to reduce heat and degradation by the sonication process. After homogenization and sonication, all spermatophores were pooled together and frozen at -20°C until needed for further experiments.

### 4.2.3 Digestion of Spermatophore Protein

780 µl of the spermatophore homogenate was combined with 275 µl of bursal extracts, prepared as described above, and incubated at 37°C. Digestion was halted at 0 (immediately after mixing), 15, 60, and 180 minutes post-mixing by removing 225 µl of the incubation supernatant, combing the solution with equal parts 1% Laemmli sample buffer (1% SDS, 120 mM Tris·HCl), and heating the mixture at 95°C for five minutes. A pooled sample which acted as a control for downstream mass spectrometry analyses was created by combining 50 µl from each time point sampled. All five samples representing digestion (T = 0 min, T = 15 min, T = 60 min, T = 180 min, and pool) were stored at -20°C until further analysis.

### 4.2.4 Visualization of Spermatophore Digestion

In addition to the digestion samples, mentioned above, remaining unmixed bursal and spermatophore samples were prepared alone in an identical fashion, with each type sampled at 0

and 180 minutes of incubation. Samples from all time points used for digestion, in addition to the unmixed samples, were visualized using 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) run at 110V for 90 minutes. Gels were then stained with Coomassie Blue (50% methanol, 10% acetic acid, 1 g Coomassie Brilliant Blue R-250, Amresco, Solon, OH, USA) for 20 minutes, followed de-staining (20% isopropanol, 7% acetic acid) overnight. The de-stained gels were imaged using a Cannon 9000F Mark II scanner at 600 dpi. Gel images were analyzed using the gel analysis feature in ImageJ software (v1.47, (Schneider et al., 2012)). Briefly, each gel lane, representing a distinct sample, was plotted by optical density. The area under the resulting peaks was calculated through manual integration. All statistical analyses were conducted in R (v3.4.1, The R Foundation for Statistical Computing). A Student's *t*-test was used in order to assess differences in relative spermatophore digestion in the presence and absence of bursal extract.

### 4.2.5 Differential Mass Spectrometry

Each time point for mixed and incubated samples (T = 0 min, T = 15 min, T = 60 min, T = 180 min, and pool) was split into nine separate technical replicates (total of 45 samples). Each replicate was comprised of 20 μg total protein and was prepared by filter-aided sample preparation (FASP) followed by an overnight trypsin digestion (Zougman, Nagaraj, Mann, & Wiśniewski, 2009). Samples were treated in a randomized order to control for any difference arising from sample preparation order. Replicates were then desalted with a C18 Saturator Column (Supelco, Bellefonte, PA, USA) and dried using a SpeedVac (Thermo Fisher Scientific, Waltham, MA, USA). Samples were re-solubilized in 20 μL of 0.1% formic acid. An instrument pool sample was created by combining 3 μL from each of the 45 samples into one sample. The samples were re-

ordered and analyzed in a balanced block design with nine samples flanked by two instrument pools before and one after followed by an angiotensin standard. All samples were analyzed by nano-flow liquid chromatography tandem mass spectrometry (nLC-MS/MS) using an LTQ Orbitrap-XL instrument (Thermo Scientific, Waltham, MA, USA). A volume of 1µLwas injected onto a 25cm $C_{18}$ PicoChip Column (New Objective, Woburn, MA, USA) and separation occurred across an LC gradient of 300nL/min with 4-32% acetonitrile over 60 minutes. Peaks were identified using Top4 dd-$MS^2$ using a high resolution full scan. Relative peak area (peptide abundance), was measured by integration under identified peaks using Skyline (Maclean et al., 2010).

### 4.2.6 Protein Identification

Spectra were aligned with peptides via a MASCOT database search using previously acquired transcriptomic sequences comprised of 15,773 unique components (C. Meslin et al., 2015; Perkins, Pappin, Creasy, & Cottrell, 1999). Proteins were considered present within a sample if spectra met a minimum protein identification threshold of 99% and had at least two mapped peptides with a minimum peptide threshold of 95%. False discovery rates (FDR) were calculated using the probabilistic method implemented through the ProteinProphet algorithm (Nesvizhskii et al., 2003). Tandem mass spectrometry data were visualized using Scaffold (Proteome Software, Portland, Oregon, USA).

### 4.2.7 Peptide Digestion Rate

Digestion rates for each peptide recovered were calculated by averaging the technical replicates for each timepoint (T = 0 min, T = 15 min, T = 60 min, T = 180 min) in order to avoid pseudoreplication. Due to the balanced nature of the sampling scheme, linear regressions were conducted in R (v3.4.1, The R Foundation for Statistical Computing) for each peptide over time (Murtaugh, 2007). To calculate for protein abundance and digestion rate, peptides were assigned to their corresponding protein with MASCOT (see above). Peptides, proteins, and their associated digestion rates were evaluated based on their typical location within the spermatophore (i.e., outer envelope, inner matrix, or both), based on locational annotation by Meslin et al. (2017). To determine differences in rates of digestion across spermatophore regions of origin, I used the Wilcoxon Rank Sum Test, due to the large differences in sample size (i.e., number of peptides) across categories.

## 4.3 Results

### 4.3.1 Cause of Digestion

Proteins localized to the spermatophore decrease in abundance over time when exposed to bursal extracts. A subset of spermatophore proteins also decrease in the absence of bursal extracts when exposed to heat. However, for a vast majority of proteins visualized through SDS-PAGE, this decrease occurs only in the presence of bursal extract (Figure 10). When I compare the total

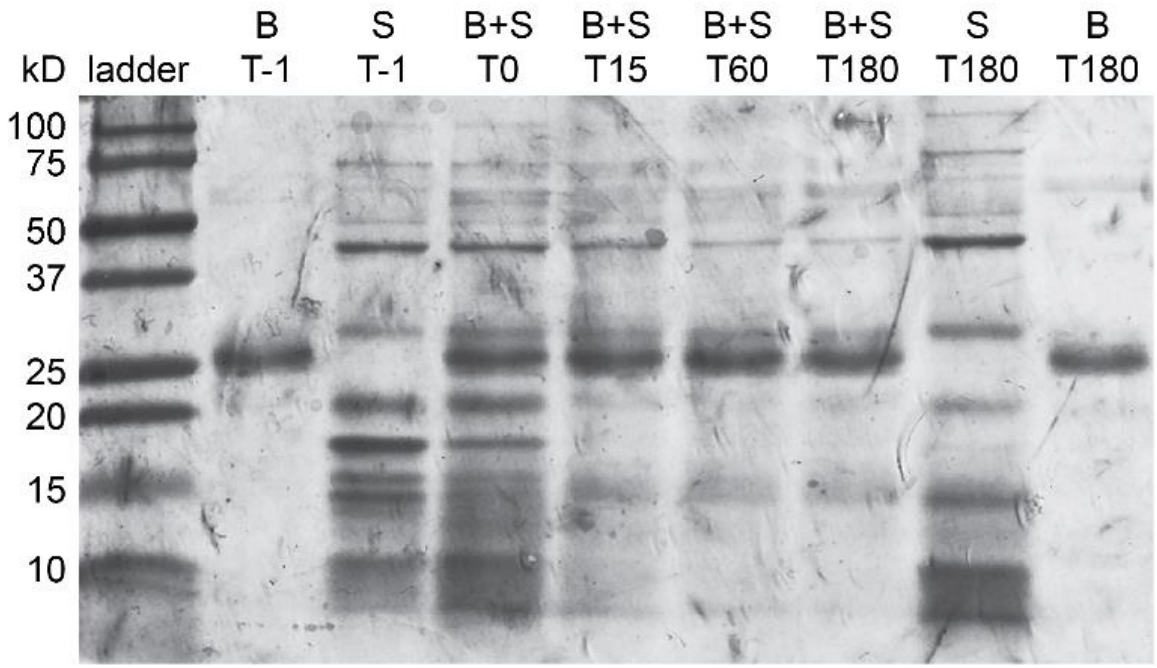


**Figure 10: Spermatophore proteins exposed to female reproductive proteases**

SDS-PAGE shows the decrease of spermatophore (S) proteins in the presence of bursal (B) extracts over time. Length of incubation (T) in minutes sampled ranged from before mixing (-1) until 180 minutes post-mixing.

amount of protein present after 180 minutes of incubation, the total amount of protein remaining from the spermatophore does not decrease with heat alone. The darkness of T = 180 gel lanes remains at 100.3% ± 8.56SE the darkness of T = 0 bands for total spermatophore protein. Conversely, the total amount of protein recovered post-incubation decreases dramatically when bursal extracts are included in combination with heat. With bursal extracts, only 39.4% ± 19.67SE of the starting amount of spermatophore proteins were recovered after incubation for 180 minutes (Figure 11). This difference is significant ($t = 2.984$, $df = 4$, $p = 0.041$), indicating that the bursa is largely responsible for the digestion of the spermatophore, though a subset of proteins do also decrease in the absence of bursal extracts.

### 4.3.2 Protein Targets of Digestion

Mass spectrometry of the spermatophore proteins incubated in the presence of bursal extracts over the digestion series recovered 48 of the 66 previously identified spermatophore proteins at the T = 0 timepoint (C. Meslin et al., 2017). Of note, I fail to recover PRSP1 or PRSP2. These two proteins are hypothesized to be responsible for decreasing digestion rate of the spermatophore by females. Additionally, they are known to be highly insoluble which likely explains their absence from this data set. After the 180-minute digestion period, the abundance of all 48 proteins decreases, with peptides from 28 unique proteins showing a significant decrease across time (Figure 12, Appendix B).

While all proteins are digested by the bursal proteases, the layers of the spermatophore are digested at different rates. After 180 minutes of digestion, peptides corresponding to proteins from the inside layer of the spermatophore disappeared more rapidly than the peptides either from the

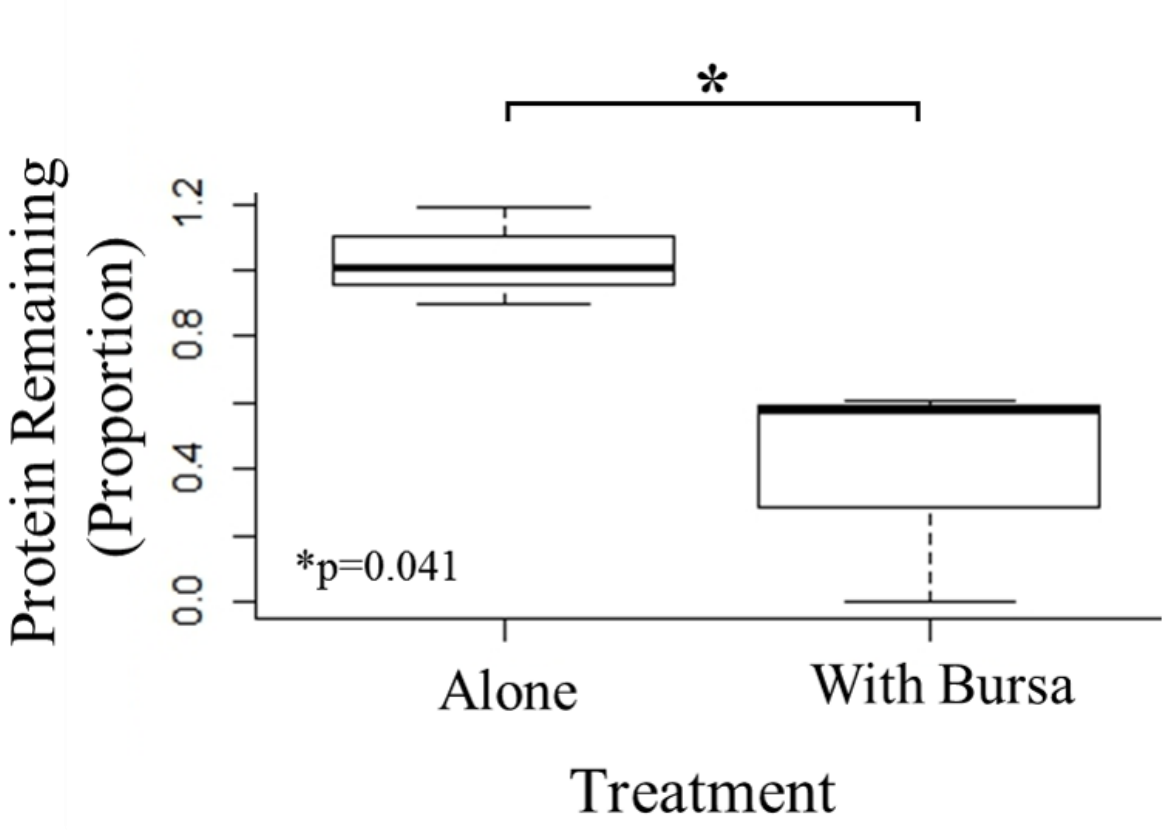


**Figure 11: Bursal contribution to spermatophore digestion**

Quantification of SDS-PAGE lanes loaded with spermatophore protein after 180 minutes of exposure to bursal extracts shows that the total amount of spermatophore protein significantly decreases.

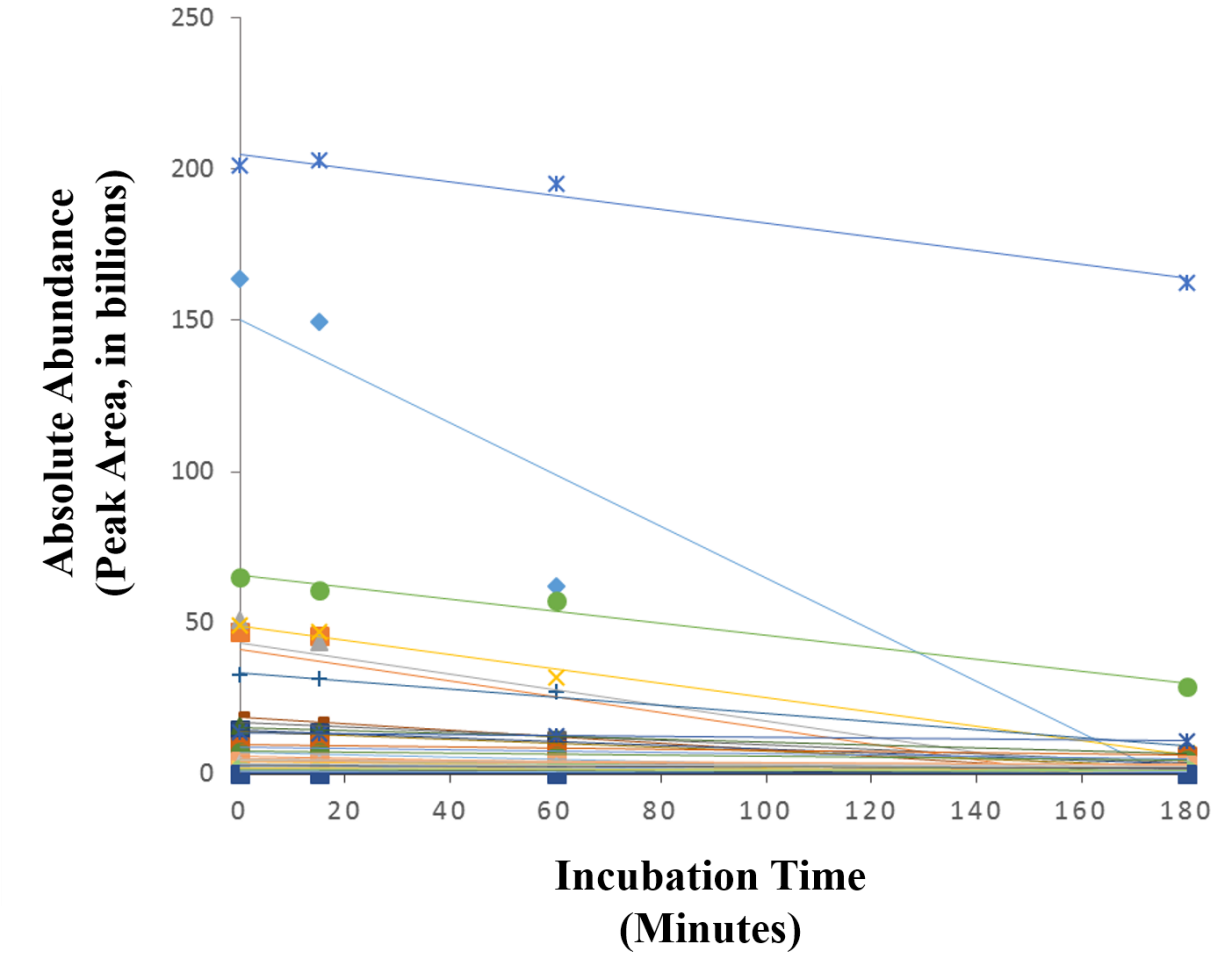


**Figure 12: Digestion of spermatophore proteins**

Spermatophore proteins were measured after incubation with to bursal extracts immediately after mixing, or at 15, 60, and 180 minutes after mixing. Raw peak abundances of each protein are plotted and show that all proteins decrease in abundance over 180 minutes.

outer spermatophore envelope (W = 320, p = 0.012) or those found in both layers of the spermatophore (W = 2079, p = 0.004). Peptides originating from the envelope, on the other hand, are digested at the same rate as peptides that are found in both layers of the spermatophore (W = 2172, p = 0.4207) (Figure 13).

## 4.4 Discussion

Through a time-series digestion assay in conjunction with proteomic analyses, I have established that the digestion of the male spermatophore proteins in *P. rapae* is primarily carried out by female-produced proteases. I find that, with few exceptions, male spermatophore proteins decrease in abundance solely in the presence of bursal extracts. Over half of the spermatophore proteins identified and quantified through differential mass spectrometry showed significant, and in many cases complete, degradation within three hours of exposure to female enzymes. Spermatophore proteins found in the inner matrix of the spermatophore are shown to be digested more rapidly than other proteins of the spermatophore.

The spermatophore of *P. rapae* is a multifaceted, complex structure that is not only responsible for increasing the fitness of both sexes through increased egg production (K.S. Oberhauser, 1997; M Watanabe & Ando, 1993), but also for mediating sexual conflict between the sexes (McNamara et al., 2019). While the spermatophore is important to numerous reproductive interactions and the fitness of both sexes, how the spermatophore is digested has remained largely a mystery. Proteases have been shown to be active and present in high concentrations within the bursa copulatrix of *P. rapae* (Plakke et al., n.d., 2015). Further, the proteins comprising the spermatophore layers have been characterized (C. Meslin et al., 2017).

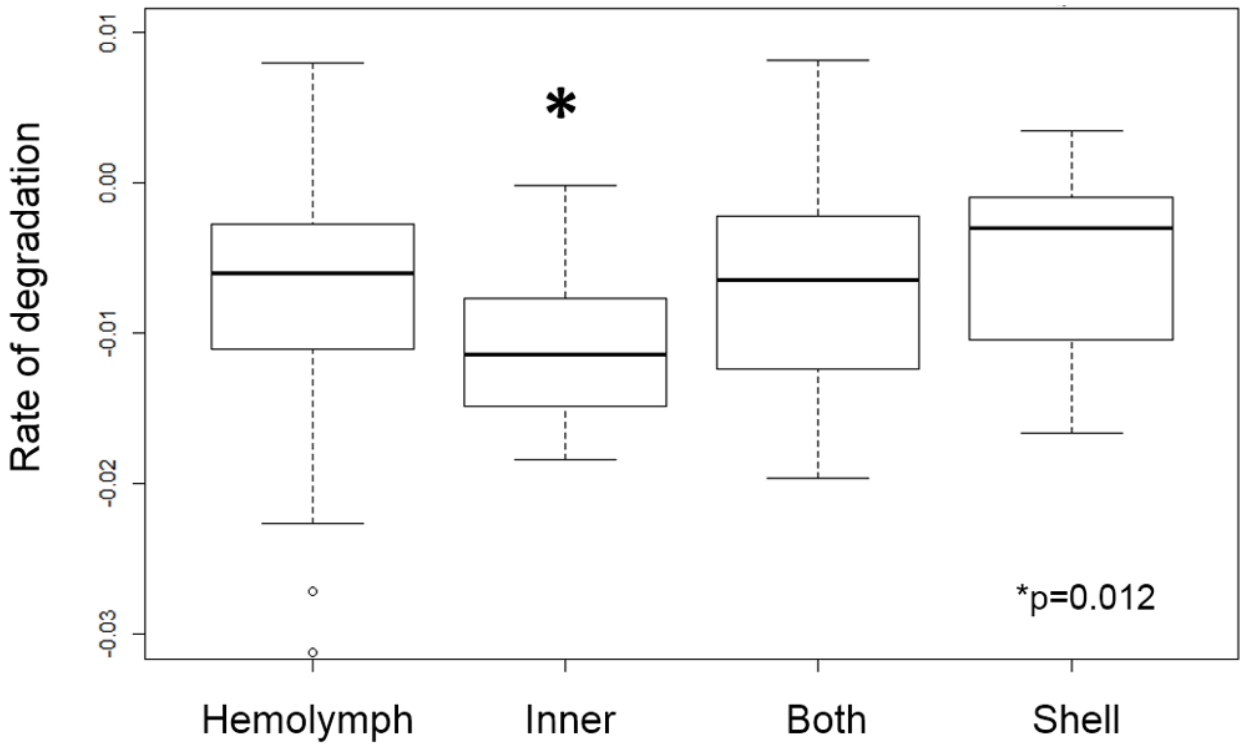


**Figure 13: Digestion rate of spermatophore proteins by origin**

Separating the proteins of the spermatophore based on their location of origin, proteins found only within the inner matrix of the spermatophore are digested more rapidly than proteins found in both layers or only in the envelope. Hemolymph proteins recovered from the spermatophore provided as a general protein reference.

However, no connection had previously been made between how the male and female components interact in order to mediate this post-mating, prezygotic interaction.

In contrast to the suspected auto-degradation of the spermatophore in *Bombyx mori* (Minoru Osanai et al., 1987), the *P. rapae* female is responsible for spermatophore digestion. In the absence of bursal extracts, very few spermatophore proteins degraded over time. These limited cases of digestion in isolation may be due to either digestion from male-derived proteases, the female proteases incorporated into the spermatophore during spermatophore transfer, or potentially due to the small amount of heat used during the incubation. The addition of bursal extracts resulted in rapid and universal degradation of spermatophore proteins.

This potential difference across species over which sex mediates the digestion of the spermatophore could occur for a number of reasons. As it has not been explicitly tested the exact mechanism by which the spermatophore of *B. mori* breaks down, it is still possible that the female degrades male-derived spermatophore proteins across all Lepidoptera. Alternatively, life history traits of species, such as female remating rate, may determine which sex takes on the role of spermatophore digestion. *B. mori* is effectively monogamous in sericulture (Biram Saheb, Singh, Kalappa, & Saratchandra, 2005) whereas *P. rapae* is highly polyandrous for a lepidopteran species, with females mating on average between 2-3 times in their short lifetime. These hypotheses could easily be differentiated with future studies incorporating multiple species which experience a range of female remating rates.

The indiscriminant digestion of the spermatophore by the bursa was not surprising. Previous work has shown the enzymatic activity present within the bursa is due to a highly concentrated mixture of functionally active enzymes (Plakke et al., 2015). These proteases represent a variety of modes of action and, in turn, potential protein substrates (Plakke et al., n.d.).

However, previous proteins of interest, PRSP1 and PRSP2, were not recovered through mass spectrometry data. This is most likely an artifact of sample preparation, as these two proteins are highly insoluble, which make them unlikely to have been extracted into the spermatophore homogenate, despite my best efforts to capture all proteins of the spermatophore.

If the spermatophore outer envelope is functioning to delay mating and the female proteases have evolved to digest the outer layer, one might expect the outer layer to be more difficult to digest due to this antagonistic coevolutionary arms race. On the other hand, if the inner matrix were to function primarily for funding eggs and increasing fitness of both sexes, there would be no pressure to decrease digestibility of the proteins, and rapid digestion would be favored by both sexes. The evidence I present from my mass spectrometry analyses is consistent with these expectations. The proteins found on the innermost layer of the spermatophore are digested at a more rapid rate than the outer envelope, suggesting that the outer envelope is a more resistant target of digestion by bursal enzymes. This lends support to the hypothesis that the inside layer reflects more of a history of cooperation, while the outer envelope may have evolved under sexual conflict. While these data in no way conclusively determine that these are the driving mechanisms behind the evolution of the two layers, the data are consistent with the hypothesis.

Identification of spermatophore proteins and their corresponding digestion rates by bursal proteases now allows for exploration of questions surrounding co-evolution between the sexes. By determining which proteins interact between the sexes at this reproductive interaction, larger evolutionary questions are now possible to evaluate. To date, limited systems have the interacting components on both the male and female sides of a reproductive interaction fully classified. This research provides the opportunity to study the evolution of a reproductive interaction. Further, it

allows for the study of how the interaction varies across time, space, and mating history. These factors will be pivotal in connecting the drivers of evolution to the phenotypic and genotypic levels.

## 4.5 Acknowledgements

I would like to thank Kretschmann Farms for allowing me to collect butterflies in their vegetable fields. I would additionally like to thank Dr. Nathan Yates and the members of the University of Pittsburgh Biomedical Mass Spectrometry Center, particularly Dr. Pamela Cantrell, Dr. Xuemei Zeng, Jackie Jones-Laughner for assistance with the differential mass spectrometry experiment. Finally, I would like to thank Dr. Nathan Brouwer for statistical advice.

# 5.0 Divergence at a post-mating, pre-zygotic interface

## 5.1 Introduction

Species are traditionally defined based on the ability for individuals to successfully mate and produce offspring (Mayr, 1942). Barriers that prevent mating can develop at any point along the reproductive process, starting with behavior and timing of reproduction, and carrying through into the effects on offspring, such as hybrid inviability (Lowry, Modliszewski, Wright, Wu, & Willis, 2008; Tennessen, 1982). When incompatibilities arise, selection acts on the reduced fitness between individuals. This reduction in fitness has the potential to then reinforce species boundaries between two population in order to prevent gene flow and detrimental effects (Dobzhansky, 1951; Mayr, 1963).

Reproductive barriers in relation to speciation have been studied and classified at several timeframes during the reproductive process, though the timeframe after mating but before the formation of the zygote (post-mating, pre-zygotic, PMPZ) has received relatively less attention (Ahmed-Braimah, 2016). This timeframe is vital for reproduction to succeed and involves numerous biochemical, physical, and molecular interactions between the male and female (Perry et al., 2013). Previous work in the field has been restricted on this front due to the limited number of systems in which both the male and female side of a PMPZ interaction had been classified. Recent work on the Cabbage White butterfly, *Pieris rapae*, allows for an in depth exploration of how PMPZ interactions have the potential to develop into reproductive barriers and contribute to the process of speciation.

Previous work in *P. rapae* has uncovered several specific interactions that occur in the PMPZ timeframe between the sexes at a molecular, biochemical, and functional level. In *P. rapae* and other Lepidoptera, females possess a specialized organ in their reproductive tract called the bursa copulatrix. The bursa functions to accept and digest the male ejaculate, termed spermatophore (Rogers & Wells, 1984). The spermatophore plays a dual role within the female reproductive tract, as both a nuptial gift and a modified mating plug. Consistent with its role as a nuptial gift, the spermatophore is high in protein content, which has been shown to directly fund egg production of the female (Boggs & Gilbert, 1979; M Watanabe & Ando, 1993). Conversely, the spermatophore also mediates female remating. Females are unable to remate until they have adequately digested the male's spermatophore (K. Oberhauser, 1992; Sugawara, 1979, 1981). This forced refractory period of the female allows the male a temporary monopoly over paternity of the female's offspring, owing to last male sperm precedence observed in this species (Wedell & Cook, 1998).

The spermatophore itself is comprised of at least 66 seminal fluid proteins (C. Meslin et al., 2017). These proteins are regionalized to one of two main layers of the spermatophore: either the insoluble outer envelope, or the soluble inner matrix (C. Meslin et al., 2017; Minoru Osanai et al., 1987). While the inner matrix is eventually absorbed by the female once the envelope is breached, the outer envelope of the spermatophore is never completely digested and remnants remain in the bursa for the entirety of a female's life (Suzuki, 1979). The structural integrity of the envelope is due to the polymerization of two highly abundant, proline-rich structural proteins: PRSP1 and PRSP2 (C. Meslin et al., 2017).

All proteins of the spermatophore are digested by a cocktail of proteases present in the bursa (see §4.0 for details). Nine proteases have been predicted to digest the spermatophore, and

six of the nine have been shown to be active and behave as predicted (Plakke et al., 2015). These proteases represent two main classes of proteases: serine-like and cysteine-like, as distinguished by their active site homology and predicted specificity (see §3.0).

While the identities and functions of the interacting components at this PMPZ interface are known, it is still unclear how these traits co-evolve with one another and how their evolution can contribute to reproductive isolation. I sought to classify the effects of protease and spermatophore evolution using two isolated populations of *P. rapae* that potentially represent an incipient speciation event. The subspecies *Pieris rapae rapae* is found in Europe and North America, while the subspecies *Pieris rapae crucivora* is found on the islands of Japan (Fukano, Satoh, Hirota, Nishide, & Obara, 2012). These subspecies have only been isolated for 1,200 years yet already show substantial differentiation at a phenotypic level (e.g., pheromones, coloration, UV patterning, size, etc) (Mcqueen & Morehouse, 2018; Obara & Majerus, 2000; Ryan et al., 2018). Since proteins related to reproduction are some of the most rapidly evolving known to date (N. L. Clark et al., 2006; Torgerson et al., 2002), it is likely that the divergence between the subspecies extends to reproductive protein interfaces as well. To test how the digestion of the spermatophore by the bursal proteases could contribute to reproductive isolation, I conducted a series of mating trials to assess spermatophore digestion. I then measured differential abundance of male and female proteins involved in the PMPZ interaction. Finally, I identified sequence divergence across the subspecies for the proteins interacting in the processes of spermatophore digestion by the female enzymes.

## 5.2 Methods

### 5.2.1 Experimental Animals

#### 5.2.1.1 Spermatophore Digestion

Animals used for the spermatophore digestion experiment included individuals of *Pieris rapae rapae* and *Pieris rapae crucivora* from laboratory populations. Stocks were established from females collected in 2012 from Rochester, PA (40°44’44.4”N 80°09’49.0”W) (*P. r. rapae*) and Hayama, Japan (35°15'41.0"N 139°36'30.3"E) (*P. r. crucivora*). Individuals were reared in climate controlled chambers with a 16h:8h light:dark photoperiod, 24°C, and 60% relative humidity. Larvae were fed *ad libitum* from leaves of *Brassica oleracea* grown on site.

#### 5.2.1.2 RNA Samples

Tissues used for RNA sequencing were collected from F1 individuals reared in the lab from wild caught females. Wild females were collected in the summer of 2017 from Rochester, PA (40°44’44.4”N 80°09’49.0”W) (*P. r. rapae*) and Hayama, Japan (35°15'41.0"N 139°36'30.3"E) (*P. r. crucivora*). Larvae were fed *ad libitum* from leaves of *B. oleracea* grown on site.

#### 5.2.1.3 DNA Samples

Wild caught females were collected in May 2017 from Rochester, PA, USA, (40°44’44.4”N 80°09’49.0”W) (*P. r. rapae*), Versailles, France (47°53'32.2"N 1°53'53.6"E) (*P. r. rapae*), and Hayama, Japan (35°15'41.0"N 139°36'30.3"E) (*P. r. crucivora*).

### 5.2.2 Rate of Spermatophore Digestion

Virgin males and females of both subspecies were placed into one of four separate 60 cm x 60 cm x 90 cm insect mating containers, each representing a different treatment: 1) male *P. r. rapae* and female *P. r. rapae*, 2) male *P. r. rapae* and female *P. r. crucivora*, 3) male *P. r. crucivora* and female *P. r. rapae*, or 4) male *P. r. crucivora* and female *P. r. crucivora*. The four mating container treatments were placed in a greenhouse with direct sunlight. Containers were checked every 20 minutes to ensure no copulation events were missed. When a copulation was observed, the male and female in copula were removed from the mating container and allowed to separate naturally. Mated females were randomly assigned to a timepoint and frozen at -80°C either immediately after separation, 1 day post-separation, or 3 days post-separation. Frozen females were dissected and spermatophore removed. Spermatophores were lyophilized (FreeZone, Labconco, Kansas City, MO, USA) for 3 days and dry mass measured. For each timepoint measured (T = 0, 1, or 3 days), a univariate ANOVA was used to determine statistical differences across the mating treatments, with group differences evaluated using post-hoc Tukey's b.

### 5.2.3 RNA Sequencing and Analysis

Males and females three days post-eclosion were microvivisected in PBS and reproductive tracts with the gamete producing organs removed were stored in RNAlater (Life Technologies, Grand Island, NY) at -20°C. The female reproductive tract was further separated with the bursa copulatrix stored in isolation. Each sample consisted of a pool of five siblings, with a total of eight families sampled for each subspecies.

RNA was extracted using Trizol (Life Technologies, Grand Island, NY) and chloroform. 50µL of purified RNA was then combined with GenTegra RNA (Pleasanton, CA, USA) and dried in a sterile fume hood for 72 hours before storage at room temperature. To reconstitute samples, 50µL of water was combined with the dried samples. All samples were processed and sequenced by Novogene (Sacramento, CA, USA). mRNA libraries were enriched using oligo(dT) beads and cDNA constructed using random hexamer primers. Libraries were then sequenced on Illumina HiSeq 4000 with paired-end reads of 150bp. Two female reproductive tracts were unable to be sequenced due to sample degradation, and were therefore excluded from further analyses.

Resulting reads were filtered using FASTX-Toolkit (version 0.0.14) for base quality (minimum of 30) and length (minimum of 20bp). In order to construct the reference transcriptome, I mapped all reads to the previously generated transcriptome (C. Meslin et al., 2015) using NextGenMap (version 0.5.2) (Sedlazeck, Rescheneder, & Von Haeseler, 2013) which was comprised of 15,773 genes. Reads that remained unmapped from *P. r. crucivora* were constructed into a pseudo-transcriptome using Trinity (version 2.8.4) (Haas et al., 2014) using default parameters. ORFs with a minimum length of 100 amino acids were then identified within each resulting component using TransDecoder (version 5.3.0), resulting in an additional 16,640 components. All reads were then mapped to this new pseudo-transcriptome with NextGenMap2. Any component which had fewer than 20 total reads map to it was discarded. Remaining components were combined with the previous transcriptome (C. Meslin et al., 2015) for a total of 19,944 genes for the total, combined reference transcriptome. Only the longest isoform was used for each gene.

Cleaned and filtered reads were mapped to the combined transcriptome using NextGenMap2 and differential expression of genes assessed through the edgeR Bioconductor

package (Robinson, Mccarthy, & Smyth, 2010). Expression counts were normalized across samples by converting read counts into trimmed means of M values (TMM) (Robinson & Oshlack, 2010) and differential expression required a minimum of a 2-fold expression difference between the tissues. PFAM domain annotations were assigned using hmmer (version 3.2.1) (Eddy, 2009) with a minimum E-value of 0.01. Protein family (Pfam) domains were then converted to Gene Ontology by using PFAM2GO (Mitchell et al., 2015). Domain enrichments were determined using an in-house python code, available upon request.

Signal peptides were identified using SignalP (v5.0). The PCA comparing sequenced tissues across subspecies used TMM-normalized reads for each gene across the three tissues and two subspecies was constructed using the Trinity software.

### 5.2.4 DNA Sequencing and Analysis

40 females were collected from each population sampled. Total DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). 100µL of DNA was combined with Gentegra DNA (Pleasanton, CA, USA) and dried in a sterile fume hood for 72 hours before being stored at room temperature. Samples were reconstituted in 100µL of water and then incubated with 10µl of RNase A (Thermo Fisher, Waltham, MA, USA) for 60 minutes at 37°C to degrade RNA contaminants. To remove the glycerol introduced with the RNase, the samples were subjected to an ethanol precipitation before reconstitution in 100µl of water. Cleaned DNA products were then quantified by a Qubit 2.0 fluorometer (Invetrogen, Carlsbad, CA, USA) and each population pooled with 0.1µg of DNA from each individual sample. Whole genomes of pooled samples were sequenced by Novogene with Illumina 150PE. Raw reads were filtered using the FASTX-Toolkit in an identical manner to the RNA sequences (above). I used NextGenMap to map all resulting

reads to the published genome. Gene-wise $F_{ST}$'s were calculated using the PoPoolation2 pipeline (version 1.2.2) (Kofler et al., 2011).

## 5.3 Results

### 5.3.1 Rate of Spermatophore Digestion

Regardless of the identity of the individuals involved in the copulation event, males provide the same sized spermatophore to females ($F_{2, 30} = 1.738$, $p = 0.194$). While individuals of the same subspecies mate with each other readily, hetero-subspecific crosses were much rarer, with the cross between *P. r. crucivora* females and *P. r. rapae* males nearly non-existent (data not shown). Three days after a mating event, *P. r. rapae* females had digested all soluble portions of the spermatophores from their own subspecies, leaving only the insoluble portion of the envelope. However, the females did not digest the opposing subspecies' spermatophore to the same extent when measured at one day post-copula ($F_{2, 35} = 10.098$, $p = 0.000397$) or at three days post-copula ($F_{2, 30} = 2.654$, $p = 0.088$), indicating a delay in digestion had occurred (Figure 14).

### 5.3.2 Expression Differences across Subspecies

Considering sequence expression across all samples, PCA indicates that tissues and samples group in a straightforward manner, with PC1 explaining 41.99% of variation and separating samples by tissue type. PC2 further separates tissues, and PC3 separates samples by subspecies, with all three PCs explaining 68.64% of all variation across samples (Figure 15).

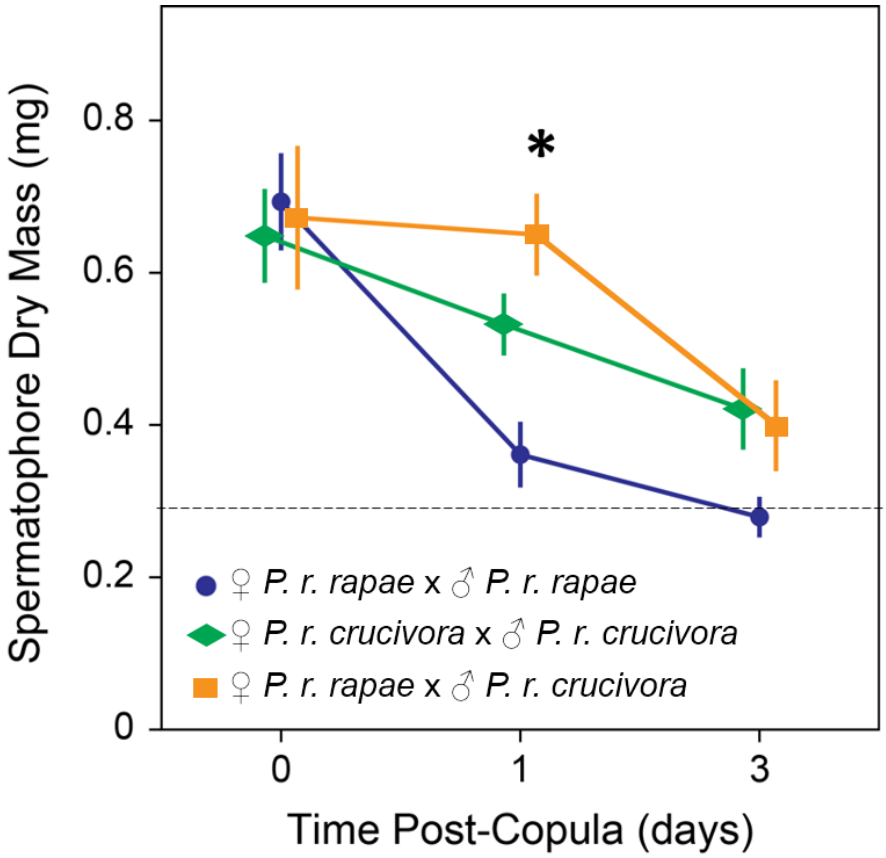


**Figure 14: Rate of spermatophore digestion**

Dry mass of spermatophores recovered from females immediately, 1 day, and 3 days after mating are plotted for the three crosses conducted. The blue line with circles represents a *P. r. rapae* female crossed with a *P. r. rapae* male, the green line with diamonds represents a *P. r. crucivora* female crossed with a *P. r. crucivora* male, and the orange line with rectangles represents a *P. r. rapae* female crossed with a *P. r. crucivora* male. Mean values of spermatophore size are plotted with 1 standard error. The dotted line represents insoluble portion of the spermatophore.

When comparing general differential expression across the subspecies for each sex, I find that for the most part, expression groups by subspecies (Figure 16). The notable exception is observed in the bursal samples across *P. r. crucivora* where female samples are split across two groupings. This disparity in grouping appears to be due to variation in transposable element expression (data not shown). Gene ontology (GO) enrichment analyses indicate that an overabundance in RNA-DNA hybrid ribonuclease activity (p = 1.12E-04 for females, and p = 2.79E-06 for males) and DNA integration (p = 1.39E-03 for females) within *P. r. rapae.* These GO terms are associated with retroviral and transposable element (TE) functions. Indeed, targeted analysis of the most highly expressed genes in either subspecies returns BLAST hits strongly supporting differing TE and retroviral expression between the subspecies and consistently between tissues within one subspecies.

#### 5.3.2.1 Bursal Proteases

A total of 2,194 genes were found to be differentially expressed between the bursal samples of the two subspecies (Figure 16A). PFAM analysis identified 34 proteases that were differentially expressed between the two subspecies, 22 of which were more highly expressed in *P. r. rapae*, and 12 more highly expressed in *P. r. crucivora.* Of these 34 proteases, 20 were found to have secretion signals (14 in *P. r. rapae* and 6 in *P. r. crucivora*) thus indicating that the proteases are likely not intracellular and are secreted into the lumen of the bursa. Of the nine proteases previously identified and characterized in the bursa of *P. r. rapae* (Plakke et al., n.d., 2015), three were more highly expressed in *P. r. rapae*, but the other six were found at equal levels across the subspecies. A gene tree indicates that the proteases found in the bursas of both subspecies were diverse, representing multiple protease classes. Additionally, the predicted proteases represent

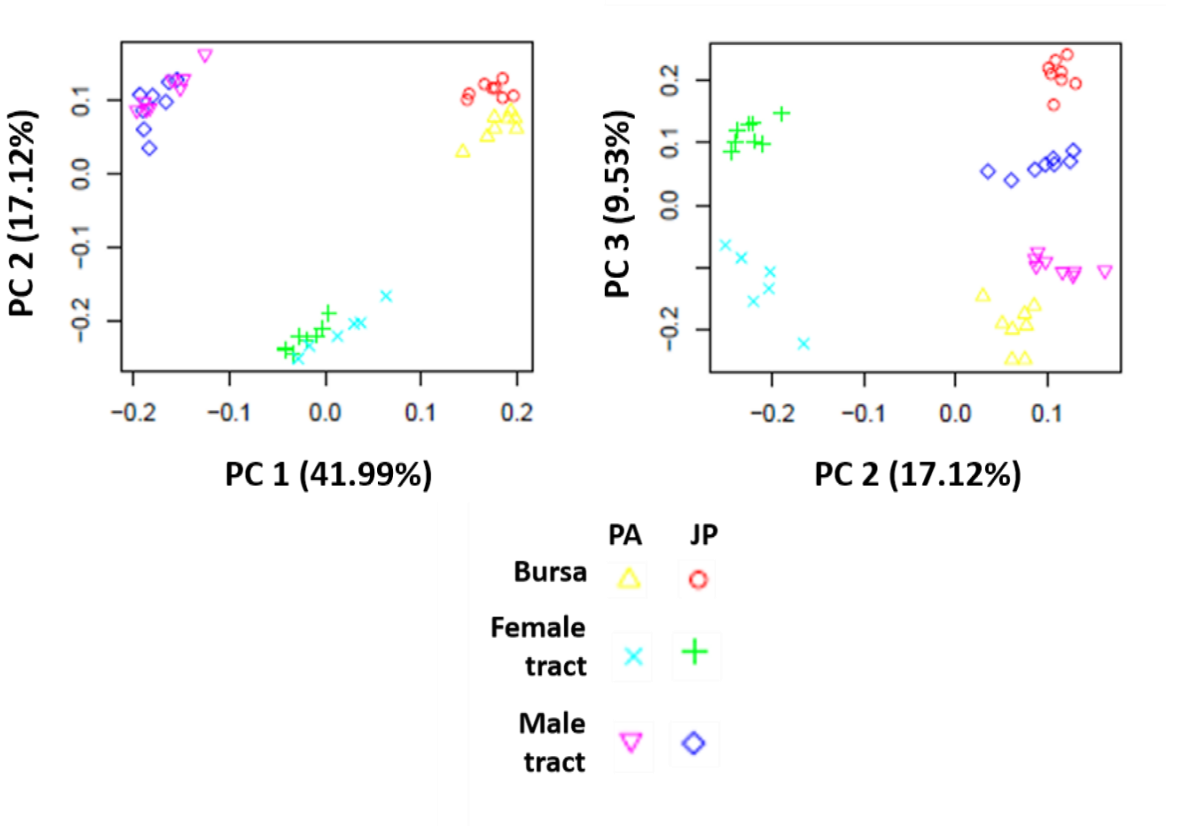


**Figure 15: PCA of tissue expression**

Bursa, female reproductive tract, and male reproductive tract group as expected by tissue type and subspecies. The first three PC's explain 68.64% of variation across samples. No samples appear as outliers and are all therefore used for subsequent steps

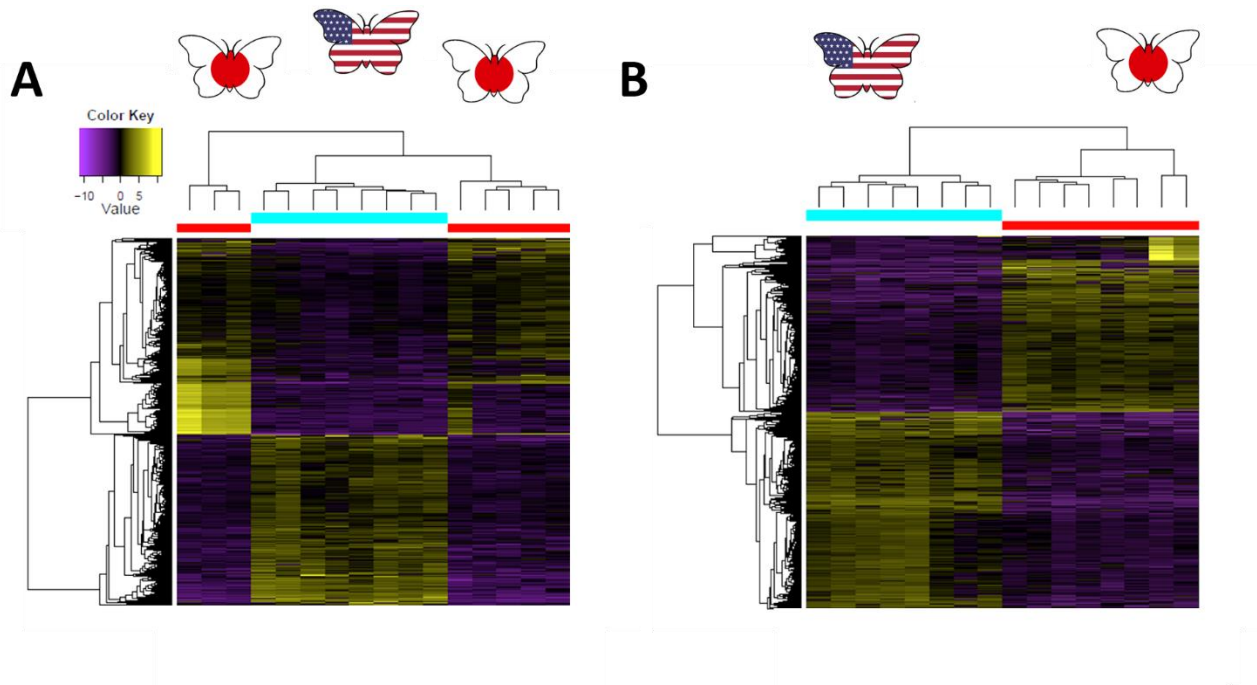


**Figure 16: Differential expression of transcripts**

Transcripts show variable expression across tissues and subspecies. Yellow indicates higher expression in a sample, whereas blue indicates lower expression levels for (A) in bursal samples and for (B) male reproductive tract samples. Butterflies with flags indicate samples from corresponding country (Japan = *P. r. crucivora*, USA = *P. r. rapae*.)

distinct protein sequences and are likely not recently duplicated genes nor different isoforms of the same gene (Figure 17). Cysteine-like proteases were found to be more highly expressed only within the bursa of *P. r. rapae*, yet various serine-like proteases were highly expressed in both subspecies. A general observation on the distribution of differential expression in the serine-like proteases indicates that many of the most closely similar sequence pairs have one sequence in the pair more highly expressed in one subspecies, while the other sequence of the pair is more highly expressed in the other subspecies.

From the 34 new proteases found to be differentially expressed between the bursae of *P. r. rapae* and *P. r. crucivora,* some proteases are expressed in radically different patterns across the subspecies. Of particular note, comp90430 is highly expressed in the *P. r. crucivora* bursa but was not expressed at all in the bursa of *P. r. rapae* (Figure 18A). However, comp90430 has been previously shown to be expressed at high levels in the caterpillar body of *P. r. rapae* (C. Meslin et al., 2015), indicating a potentially unique co-option and co-expression event in one subspecies but not the other.

#### 5.3.2.2 Spermatophore Proteins

A total of 1,617 genes were differentially expressed between the male reproductive tracts of the two subspecies (Figure 16B). For the 66 spermatophore proteins of interest, 21 were differentially expressed between the subspecies, though to a much lower extent than the bursal proteases (Figure 18B). Of particular note are the two proteins that comprise the envelope of the spermatophore. Both PRSPs are highly expressed in both subspecies and among the most highly expressed of all spermatophore proteins. However, they are not differentially expressed across the subspecies (Figure 18B). For the proteins that were differentially expressed, there was no observable difference based on the layer of the spermatophore that the proteins comprise.

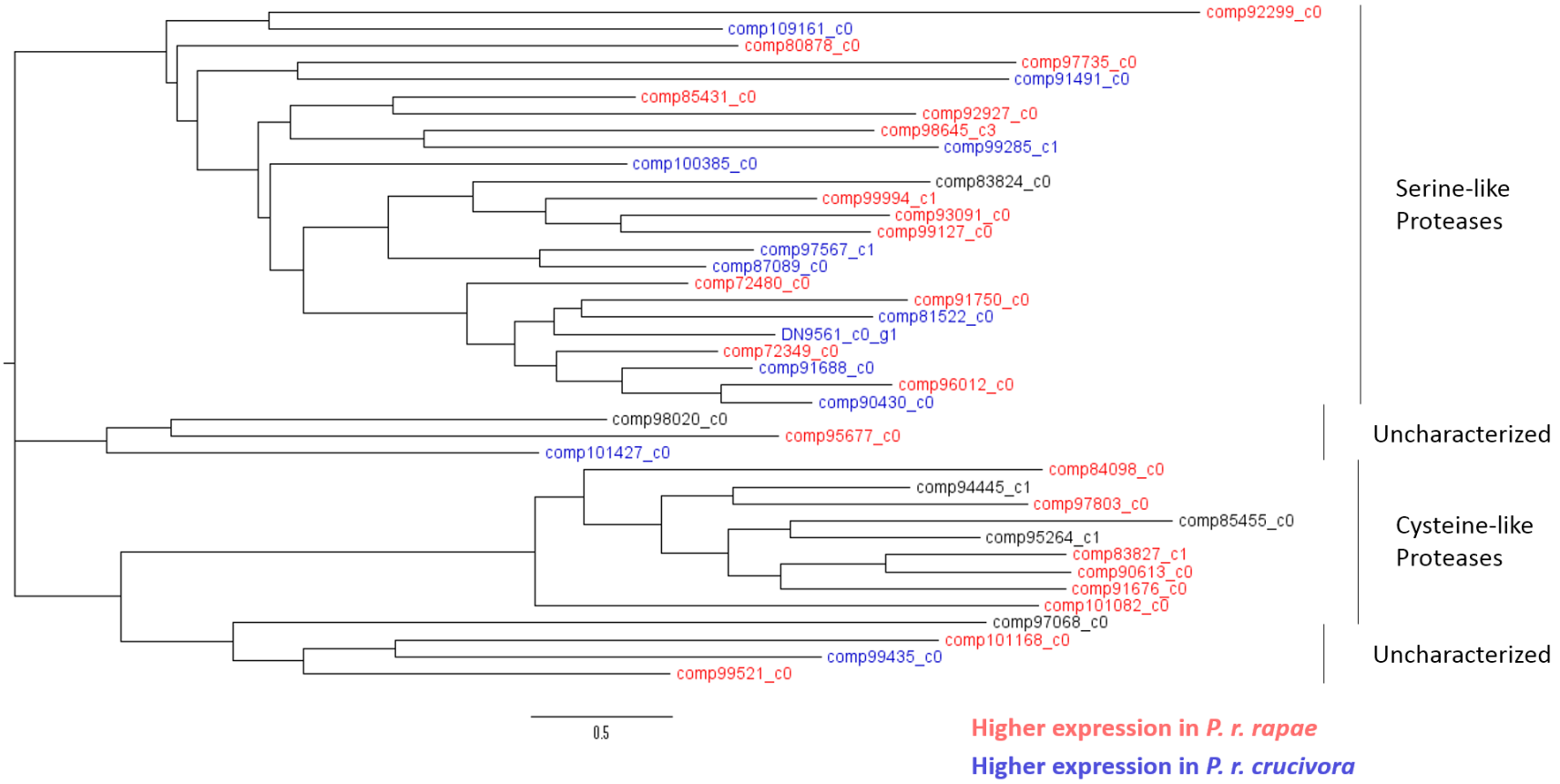


**Figure 17: Gene tree of bursal proteases**

Each protease identified previously in the bursa copulatrix of *P. rapae* was aligned and plotted in relation to other proteases. Class of protease is indicated to the right. Color indicated the subspecies the protease is more highly expressed in: red for *P. r. rapae* and blue for *P. r. crucivora.*

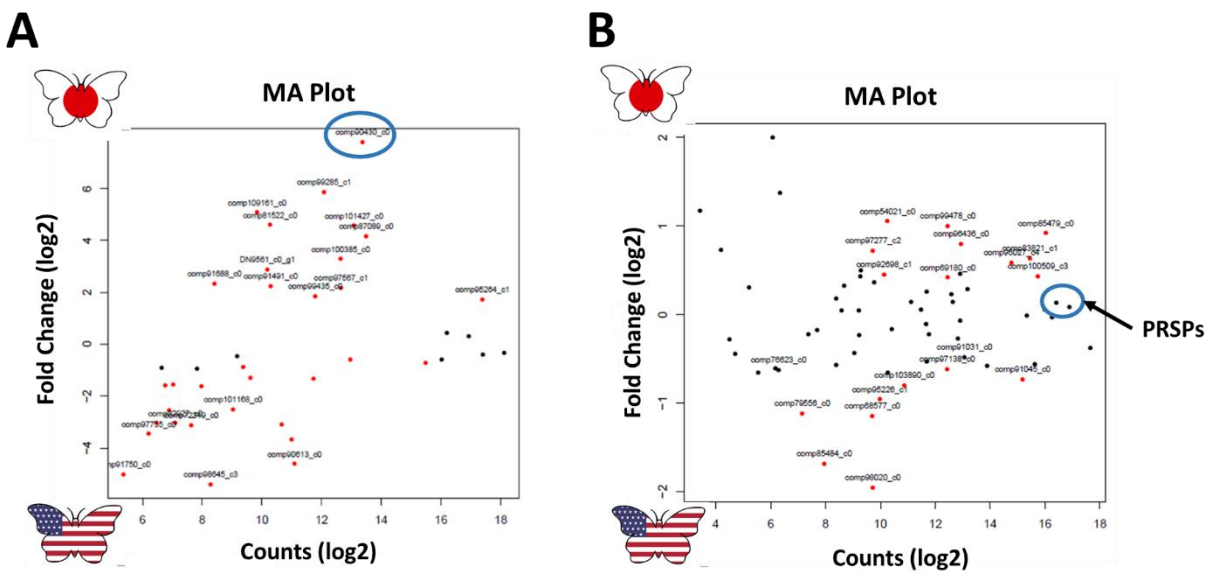


**Figure 18: Expression patterns of proteins involved in spermatophore digestion**

MA plots indicating the relationships between the expression levels for each protein compared to the level that each protein is differentially expressed for (A) bursal proteases and (B) spermatophore proteins. Proteins that are differentially expressed at a significant level are denoted with a red dot. The proteases circled in (A) represents a case study (comp90430) in which a protease is expressed only in the bursa of *P. r. crucivora*. The proteins circled in (B) are PRSP1 and PRSP2 which are responsible for the insolubility of the spermatophore envelope, yet are not differentially expressed. Butterflies with flags indicate samples from corresponding country (Japan = *P. r. crucivora*, USA = *P. r. rapae*.)

### 5.3.3 Sequence Divergence across Subspecies

Across the three populations samples for nucleotide variation, I recovered 327x coverage for European *P. r. rapae* (~8x coverage per individual), 408x coverage for American *P. r. rapae* (~10x coverage per individual), and 333x coverage for *P. r. crucivora* (~8x coverage per individual). Comparing the variation in genomes, I find an overall nucleotide diversity (pi) for European *P. r. rapae* as 0.016, American *P. r. rapae* as 0.015 and *P. r. crucivora* as 0.018. On the whole, bursal proteases had a higher $F_{ST}$ than spermatophore proteins when comparing across subspecies ($t = 2.556$, $df = 46.084$, $p = 0.014$). $F_{ST}$ values were also higher for female protease inhibitors than male protease inhibitors ($t = 2.127$, $df = 18$, $p = 0.048$) (Figure 19).

## 5.4 Discussion

Through my research, I have demonstrated that the evolution of bursal proteases, spermatophore proteins, and their interactions have the potential to lead to reproductive barriers across populations of the same species. In *P. rapae,* the two subspecies experience a delay in spermatophore digestion when mated to the opposite subspecies. I explored the expression level and genomic level differences between the populations that may contribute to this mismatch resulting in slower digestion of the spermatophore. I observed proteases potentially contributing to the digestion of the spermatophore to be more variable across populations than the proteins of the spermatophore itself. I found that females may be responsible for this reproductive mismatch resulting in the delay in spermatophore digestion. The proteases of the bursa were expressed in

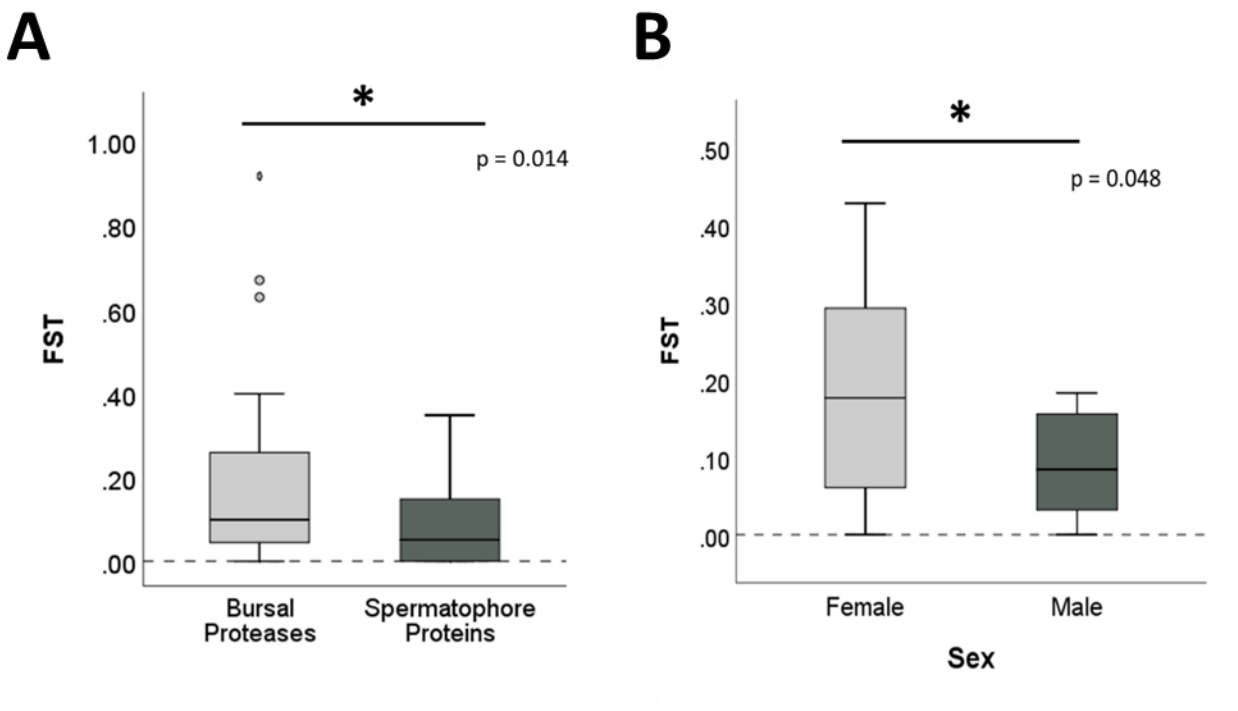


**Figure 19: $F_{ST}$ for reproductive proteins**

Comparisons of $F_{ST}$ for (A) proteins directly involved in spermatophore digestion and (B) protease inhibitors contributed by either sex during reproduction that might interact with and affect spermatophore digestion. Both sets of proteins show the common pattern that male proteins have significantly lower $F_{ST}$ than female derived proteins.

various levels across the subspecies and the $F_{ST}$ of bursal proteases were elevated in comparison to spermatophore proteins.

Independent of reproductive proteins on the male and female sides, I found that the most highly differentially expressed genes between the subspecies were related to retrovirus and transposable element function. Some of the domains annotated in the *P. rapae* samples, such as TRAS3 and Gag, had been previously identified in the telomeric regions of other species (Kubo, Okazaki, Anzai, & Fujiwara, 2001). Some have hypothesized that the expression of certain retroviral agents, like bracoviruses, may act as a counter to infection by other retroviruses, like baculoviruses (Gasmi et al., 2015). This could act as a potentially strong selecting agent, because baculoviruses are used as a biological pesticide worldwide (Moscardi, Souza, Castro, Moscardi, & Szewczyk, 2011) and could be a further explanation for rapid, recent divergence between these two subspecies.

The observed delay in digestion across the two subspecies (*P. r. rapae* female x *P. r. crucivora* male) equates to roughly a two-day delay in access to protein by the female derived from the spermatophore. While this study did not directly test for fitness consequences, the digestion rate potentially influences fecundity and/or genetic. A female takes longer to remate and acquire additional protein and sperm from additional males if digestion is inhibited. Additionally, a female will not be able to use the protein provided by the first male until half-way through her adult life. Due to the fact that females only live on average 10 days or less as adults (Gilbert, 1984), two days represents a large portion of her life. Of particular note was the lack of matings observed in the opposite hetero-subspecific cross. Several potential differences across the populations might help explain this pre-mating reproductive barrier, such as female coloration, male pheromones,

and size differences (Mcqueen & Morehouse, 2018; Obara, Ozawa, Fukano, Watanabe, & Satoh, 2008).

Looking at the differences in expression of interacting players in the reproductive tracts of the two subspecies, I find larger differences on the female side as compared to the male side. Numerous proteases were found to be differentially expressed between the subspecies, with one particular protease that is not found in the bursa of *P. r. rapae* being highly expressed in the bursa of *P. r. crucivora*, providing an interesting target for future studies. When comparing the overall level of differential expression between the subspecies across sexes, I find that proteases of the females exhibit a much higher level of differential expression than any of the proteins of the spermatophore. This could potentially indicate that females may deploy different proteases in different amounts across the subspecies to digest the spermatophore, while the spermatophores that the males provide are fairly static across the populations. One caveat to these analyses are the methods by which proteases and spermatophore proteins were identified. While proteases could be identified through domain analysis, proteins of the spermatophore were limited to previous proteomic work conducted on *P. r. rapae.* As such, there is the possibility that proteins could comprise the spermatophore of *P. r. crucivora* that are not found in the spermatophore of *P. r. rapae*. Surprisingly, on the male side I hypothesized that the PRSP proteins, which comprise the insoluble portion of the spermatophore, would be expressed differentially between the two subspecies. Instead, they were expressed at comparable levels between the two subspecies, further emphasizing the lack of differential expression in spermatophore proteins between the subspecies.

When examining the fixation of alleles based on population, I find a similar pattern to the expression data, in that genes relating to the digestion of the spermatophore show greater divergence in allele frequencies across populations than the proteins of the spermatophore itself. I

find higher $F_{ST}$ for both female proteases and female protease inhibitors than for either spermatophore proteins or male protease inhibitors. While genes encoding female proteins show greater allele frequency differences between populations, I also find that the allele frequency differences are indistinguishable from that of the entire genome, suggesting that instead of acceleration by the female, the opposite might be true. Lower allele frequency difference explained by population for male-specific sequences might indicate that the spermatophore proteins are not undergoing directional selection in any particular population, and therefore may not be what is driving the evolution of the digestion mismatch between the subspecies.

The apparent pattern of female proteases experiencing larger differences at both the RNA and DNA levels in comparison to male proteins of interest lends itself to two alternative explanations. One potential explanation is that the female side of the digestive interaction is driving the system and evolving in response to reproductive pressures. This has been observed in other systems where female reproductive environments were under selection to benefit conspecific sperm in a hybrid zone. Due to the asymmetric fitness consequences of hybrid matings in relation to gamete investment, males did not experience a large detrimental effect and therefore sperm were not under selection (Cramer, Alund, Mcfarlane, Johnsen, & Qvarnstrom, 2016). Alternatively, female proteases might be evolving in conjunction with a separate target not explored by this study, or even in response to pleiotropic effects. For example, the proteases may be more tightly connected to the migration and activation of sperm proteins. Future studies would benefit from exploring this possibility. The lack of genomes from species closely related to *P. rapae* limits the tools that are available, but once more genomes have been sequenced and annotated, analyses including Evolutionary Rate Covariation (N. Clark, Alani, & Aquadro, 2012) can be employed to

test for potentially interacting and co-evolving proteins for both the identified proteases and proteins of the spermatophore.

Reproductive barriers leading to the divergence of populations into separate species have the potential to develop when isolated populations come back into contact and any fitness consequences results. In *P. rapae*, I suggest that the two subspecies found in Japan and North America represent an incipient speciation event due to a mismatch in the reproductive process. My research suggests that the protein digesting enzymes employed by females of the species are likely responsible for this developing reproductive barrier. Future work exploring the functional consequences by experimentally manipulating levels of proteins across populations would verify and expand the role each individual protein plays in contributing to the delay in spermatophore digestion across subspecies.

## 5.5 Acknowledgements

I would like to thank Kretschmann farms and the farms surrounding SOKENDAI for allowing sample collection, Dr. Kevin Kohl for use of growth chamber space, and members of the Arikawa and Morehouse labs for animal rearing assistance. This research was supported by NSF EAPSI #1713955 in conjunction with the JSPS Summer Fellow Program, the University of Pittsburgh Graduate Student Organization Summer Research Award, and the Arthur and Barbara Pape Endowment Award.

# 6.0 Conclusions

Through the culmination of the above research, I have established a system in which post-mating, pre-zygotic interactions can be implicated in the process of driving speciation. Very few systems up to this point have demonstrated how interactions between the female reproductive tract and the male ejaculate interact and co-evolve with one another, with a majority of the research to this point having been conducted on model systems such as *Drosophila*. Many of the claims about evolution between the sexes has been limited by the lack of characterization on female components due to a historical bias, citing females as a passive participant to the reproductive process (Ah-King et al., 2014). However, increasing evidence is changing this perspective, and the research presented above further supports this necessary change of view.

Through my research, I aimed to not only identify and characterize interacting male and female components interacting during the PMPZ timeframe, but to also establish their evolutionary history and consequences of divergence across isolated populations. I began by investigating the potential that proteases can provide a physiological and biochemical environment within the female reproductive tract which may come in direct contact with the male ejaculate. In the Cabbage White butterfly, *Pieris rapae*, the female reproductive tract is sectioned off into separate compartments with the sperm being separated from the remainder of the male ejaculate. In the organ tasked with housing the ejaculate proteins, I determined that the female utilized a cocktail of protein digesting enzymes in levels comparable to the entirety of the caterpillar gut, an organ 20 times larger than the female's bursa copulatrix. I followed this study with an in-depth investigation into the mechanisms of digestion and the specificity the proteases exhibit when digesting protein substrates. Through a combination of structural modelling and biochemical

analyses, I found that the proteases of the bursa do not necessarily act as would be predicted solely from sequence domain analysis, but have fine-scale variability that affect which substrates and sequences the bursal proteases are able to target.

With the bursal proteases classified to a functional level beyond most other systems to date, I turned to assessing the exact targets of digestion. By quantifying the amount of spermatophore proteins present at various points after introduction to a bursal mixture of proteases, I was able to concretely establish the general and near ubiquitous digestion of the spermatophore by the female. This severe degradation of the male proteins may be due to general action of each individual protease, or could be due to a more fined tuned, specific mechanism of digestion. As this experiment took all bursal proteases as a mixture, there is still the possibility that each protease in the bursa acts in isolation on very specific protein targets and that only in combination do all proteins of the spermatophore degrade. This would allow for co-evolution between specific bursal proteases and spermatophore proteins to contribute to population divergence, though would require further studies in order to verify.

With both the male and female sides of the PMPZ interaction classified and the interaction further verified through biochemical assays, I was able to finally assess how this interface between the bursa and spermatophore could contribute to reproductive isolation and the development of reproductive barriers. The observation that females take longer to digest the spermatophore from a divergent population from her own provided evidence that there was a mismatch between the two subspecies at the PMPZ interface surrounding the spermatophore digestion by the female. By examining the transcript-level variation and the genetic sequence variation for the bursal proteases and spermatophore proteins, the trend emerged that female derived proteins exhibit higher levels of differences across populations than males. This runs contrary to traditional views on

reproduction, but indeed it appears that female evolution at sequence and expression levels are driving the mismatch in spermatophore digestion.

These findings leave many questions still unexplored and various future directions for exploration within the system. For example, if not the spermatophore proteins, what could the bursal proteases be evolving in response to? Other explanations might include diet and environmental habitats, as several of the proteases were originally co-opted and co-expressed from other digestive organs. Could it be that the proteases are constrained or driven by pleiotropic effects based on the host plants in the environment? Alternatively, the bursal proteases may be evolving in response to a male-derived protein set that has not been captured by the genes of previous interest. For example, the bursa may evolve and interact with proteins that contribute to sperm function. Butterflies have the unique situation of possessing two types of sperm, one of which is anucleate (M Osanai & Kasuga, 1990; Minoru Osanai, Kasuga, & Aigaki, 1989; Mamoru Watanabe et al., 1998; Wedell & Cook, 1999). Perhaps the bursal proteases are evolving in conjunction with these sperm associated proteins. It is unclear what the exact function of these apyrene sperm might be, but future studies would benefit from determining the potential interactions and associations between any and all proteins of the male with bursal enzymes.

Through my studies, while I have recorded a mismatch between the subspecies and hypothesized that secondary contact between the subspecies may eventually lead to the development of reproductive barriers, I have failed to demonstrate if this mismatch does in fact lead to a fitness consequence. The delay in spermatophore digestion does limit the rate at which females may access proteins to use for egg production and cell maintenance, but whether this leads to a decrease in lifetime fecundity is unknown. Future studies should focus on the number of eggs laid by these hetero-subspecific crosses in addition to how long it takes a female to begin laying

eggs. *P. rapae* do not emerge from their pupal state with mature eggs, and so the eggs develop at the same time that the first spermatophore is being broken down within her reproductive tract. If the spermatophore is not breached during this time, the first clutch a female lays may then suffer due to less available protein from the female. Alternatively, a female may have to invest more heavily in the first clutch, effectively decreasing her lifespan. Any of these fitness consequences, if demonstrated, would provide evidence that if these two subspecies were to come back into contact, hybrid matings should be selected against and barriers to gene flow would evolve.

Interestingly, when looking at overall nucleotide diversity for *P. r. crucivora*, I found a higher measure than in either of the *P. r. rapae* populations. This, combined with the extremes in phenotypic divergence across the subspecies hints at an alternative explanation: hybridization and introgression from co-occurring pierids. In Japan, several species of the genus *Pieris* coexist, and it has been previously hypothesized that the original migrants onto the islands hybridized with the local *Pieris melete* (Obara & Majerus, 2009). Introgression from a separate species would provide a hypothesis for the reproductive mismatch in the absence of reproductive co-evolution. Unfortunately, *P. melete* has not been sequenced yet, and so tests for introgression or hybridization are lacking. With the growing repository of genetic tools and genomes, the possibility may soon be tested.

As a whole, my research argues for the need to study the female side of reproductive interactions. Instead of being viewed as passive, or not contributing to rapid evolution of reproductive interactions, the female reproductive tract and associated traits should be embraced as a burgeoning frontier of research. Such scientific explorations will enlighten and expand current understandings of how reproductive co-evolution occurs. Further expanding on the research beyond an evolutionary or basic science perspective, applications of female reproductive biology

can be applied to human health. A growing number of couples suffer from unexplained infertility and part of this diagnosis comes strictly from the lack of knowledge surrounding what is required for successful reproduction to occur in the first place. In order to be able to develop treatments, and understand how reproductive barriers develop, the female must be included in the equation and no longer ignored.

# Appendix A

## Proteins identified through zymogram band analysis

Mass spectrometry analysis identified a total of 52 proteins across the three bursal bands and 5 additional proteins in the control band. Proportions are reported by protein for each individual sample. α, β, and γ indicate the respective band with the replicate number indicated. C designates the control band taken from a non-band area (see §3.0 for more details).

**Table 5: Proteins identified through zymogram band analysis.**

| | | Proportion of spectra counts (by sample) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Comps (Meslin et al 2015)** | **Annotation** | **α1** | **α2** | **α3** | **β1** | **β2** | **β3** | **γ1** | **γ2** | **γ3** | **C1** |
| comp100001_c0 | *fatty acid synthase* | 0 | 0 | 0.04202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp100014_c0 | *putative Trehalase-1A* | 0 | 0 | 0 | 0 | 0 | 0 | 0.02405 | 0.03321 | 0 | 0 |
| comp101124_c0 | *basement membrane-specific heparan sulfate proteoglycan core protein isoform X5* | 0.05946 | 0.03367 | 0 | 0.0523 | 0.0486 | 0.0234 | 0 | 0 | 0.02344 | 0 |
| comp101499_c1 | *spectrin alpha chain isoform X3* | 0.05946 | 0.01347 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00781 | 0 |
| comp101621_c1 | *muscle M-line assembly protein unc-89-like* | 0 | 0 | 0 | 0.0523 | 0.0486 | 0.0234 | 0 | 0.01107 | 0.02344 | 0 |
| comp101729_c1 | *putative Spectrin beta chain* | 0.04865 | 0.0101 | 0 | 0.0058 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp101894_c0 | *twitchin-like* | 0.03784 | 0.0202 | 0 | 0.0058 | 0 | 0.0352 | 0 | 0 | 0 | 0 |
| comp101896_c0 | *muscle-specific protein 300* | 0.02162 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp1912784_c0 | *glyceraldehyde-3-phosphate dehydrogenase isoform 2* | 0.01081 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Comps (Meslin et al 2015) | Annotation | α1 | α2 | α3 | β1 | β2 | β3 | γ1 | γ2 | γ3 | C1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| comp25696_c0 | *bilin-binding protein* | 0.01622 | 0.01347 | 0.01681 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp414698_c0 | *apolipoprotein E isoform* | 0.00541 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| comp54021_c0 | *myosin heavy chain* | 0.04324 | 0.03367 | 0 | 0 | 0 | 0.0234 | 0 | 0 | 0 | 0 |
| comp54036_c0 | *myosin light chain* | 0 | 0.00673 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00391 | 0 |
| comp54049_c0 | *abnormal wing disc-like protein* | 0 | 0 | 0.01681 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp81421_c0 | *tropomyosin* | 0 | 0 | 0 | 0 | 0 | 0 | 0.00687 | 0.03321 | 0 | 0 |
| comp81444_c0 | *actin, beta* | 0.02703 | 0.00673 | 0 | 0.0174 | 0 | 0 | 0 | 0 | 0.01953 | 0 |
| comp83824_c0 | *Trypsin-like serine protease* | 0.05946 | 0.04714 | 0.04202 | 0.2209 | 0.3056 | 0.4159 | 0.06185 | 0.11071 | 0.14843 | 0 |
| comp83845_c1 | *tropomyosin* | 0 | 0 | 0 | 0 | 0 | 0 | 0.01031 | 0.03321 | 0 | 0 |
| comp83827_c1 | *Cathepsin propeptide inhibitor domain, papain family cysteine protease(Pept_C1)* | 0 | 0.11448 | 0.03361 | 0 | 0 | 0 | 0 | 0 | 0.01172 | 0 |
| comp83918_c0 | *putative citrate synthase* | 0 | 0.0101 | 0 | 0.0174 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp84295_c0 | *mucin-2-like isoform X1* | 0.02703 | 0.0202 | 0.02521 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp85455_c0 | *Papain family cysteine protease (Pept_C1)* | 0.02703 | 0.02357 | 0.05882 | 0.0814 | 0.1042 | 0.1626 | 0.70447 | 0.3 | 0.45704 | 0 |

| Comps (Meslin et al 2015) | Annotation | α1 | α2 | α3 | β1 | β2 | β3 | γ1 | γ2 | γ3 | C1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| comp87729_c0 | *thiol peroxiredoxin* | 0.01081 | 0.00673 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp88065_c0 | *puromycin-sensitive aminopeptidase* | 0 | 0 | 0 | 0 | 0.0208 | 0 | 0 | 0.02214 | 0 | 0 |
| comp88877_c0 | *Calponin* | 0 | 0.00673 | 0 | 0.0407 | 0.0417 | 0.0352 | 0 | 0 | 0.04297 | 0 |
| comp89682_c0 | *alpha-N-acetylgalactosaminidase* | 0.01081 | 0 | 0 | 0.064 | 0.0278 | 0 | 0 | 0 | 0 | 0 |
| comp91250_c1 | *fatty acid-binding protein* | 0 | 0 | 0 | 0 | 0 | 0 | 0.00687 | 0.02214 | 0 | 0 |
| comp91664_c0 | *arylphorin-type storage protein* | 0 | 0 | 0 | 0 | 0 | 0 | 0.00687 | 0 | 0 | 0 |
| comp91676_c0 | *Cathepsin propeptide inhibitor domain, papain family cysteine protease (Pept_C1)* | 0.22162 | 0.3165 | 0.23529 | 0.2209 | 0.2361 | 0.189 | 0.06529 | 0.07857 | 0.08594 | 0 |
| comp91678_c0 | *Hsp70* | 0 | 0 | 0.05882 | 0 | 0.0069 | 0 | 0 | 0 | 0 | 0 |
| comp91744_c1 | *glyceraldehyde-3-phosphate dehydrogenase* | 0.03784 | 0.0303 | 0.01681 | 0.0058 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp92490_c0 | *acid phosphatase* | 0 | 0.0202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp93076_c0 | *triosephosphate isomerase* | 0.01622 | 0.00673 | 0.01681 | 0.0233 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp93823_c0 | *aldehyde dehydrogenase X* | 0 | 0 | 0.05042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp94073_c0 | *yellow-b* | 0.02162 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Comps (Meslin et al 2015) | Annotation | α1 | α2 | α3 | β1 | β2 | β3 | γ1 | γ2 | γ3 | C1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| comp94677_c1 | *ATP-binding cassette sub-family G member 1-like* | 0.01081 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp94819_c0 | *heat shock protein 70* | 0.01622 | 0.01684 | 0.05042 | 0.0058 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp95492_c0 | *protein takeout-like* | 0 | 0 | 0 | 0.0116 | 0 | 0 | 0 | 0 | 0.00781 | 0 |
| comp95939_c1 | *very long-chain specific acyl-CoA dehydrogenase* | 0 | 0 | 0 | 0.0233 | 0 | 0 | 0 | 0 | 0.00781 | 0 |
| comp96031_c0 | *puromycin-sensitive aminopeptidase* | 0 | 0 | 0 | 0 | 0.0208 | 0 | 0.00344 | 0.02214 | 0 | 0 |
| comp97068_c0 | *Peptidase_S28* | 0.02162 | 0.04714 | 0 | 0.0407 | 0.0486 | 0.0117 | 0.08247 | 0.28929 | 0.05859 | 0 |
| comp97151_c0 | *myosin heavy chain* | 0.01622 | 0.03367 | 0.22689 | 0.0058 | 0 | 0 | 0 | 0 | 0.00781 | 0 |
| comp97269_c0 | *Sarco/endoplasmic reticulum Calcium ATPase* | 0 | 0.00673 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp97681_c0 | *hexamerin-like* | 0 | 0 | 0 | 0 | 0 | 0 | 0.02062 | 0 | 0 | 0 |
| comp97762_c0 | *Chromatin modification-related protein YNG2* | 0 | 0 | 0 | 0.0291 | 0.0278 | 0.0234 | 0.00344 | 0 | 0.01953 | 0 |
| comp98020_c0 | *aminopeptidase N-like* | 0.12973 | 0.11785 | 0.09244 | 0.0291 | 0 | 0 | 0.00344 | 0.02214 | 0.05469 | 0 |
| comp98343_c1 | *hexokinase* | 0 | 0 | 0 | 0.0291 | 0.0208 | 0.0567 | 0 | 0.02214 | 0.01172 | 0 |

| Comps (Meslin et al 2015) | Annotation | α1 | α2 | α3 | β1 | β2 | β3 | γ1 | γ2 | γ3 | C1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| comp98432_c0 | *adenylyl cyclase-associated protein 1 isoform X1* | 0 | 0 | 0 | 0.0116 | 0 | 0 | 0 | 0 | 0.00391 | 0 |
| comp99116_c0 | *2-(3-amino-3-carboxypropyl)histidine synthase subunit 1* | 0 | 0.00673 | 0 | 0.0058 | 0.0139 | 0 | 0 | 0 | 0.00391 | 0 |
| comp99229_c0 | *2-amino-3-ketobutyrate coenzyme A ligase* | 0 | 0 | 0.01681 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp99503_c1 | *charged multivesicular body protein 5* | 0.04324 | 0.0303 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| comp99602_c0 | *dipeptidyl peptidase 3 isoform X1* | 0 | 0 | 0 | 0 | 0.0278 | 0 | 0 | 0 | 0 | 0 |
| comp11962_c0 | *pigment epithelium-derived factor isoform 1 precursor* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.15 |
| comp15078_c0 | *clusterin isoform X1* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| comp1887748_c0 | *complement C4-A* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.4 |
| comp3999878_c0 | *apolipoprotein E* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| comp4639105_c0 | *C1-inhibitor* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.15 |

# Appendix B

## Statistical analysis for digestion rates of spermatophore proteins when exposed to bursal extracts

Mass spectrometry analysis identified a total of 78 peptides that significantly decreased over time when exposed to bursal extracts. The peptides correspond to 28 unique proteins. Values provided for each time point represent the average peak integration for that peptide over 9 technical replicates (see §4.0 for more details).

**Table 6: Statistical analysis for digestion rates of spermatophore proteins when exposed to bursal extracts.**

| Protein of origin | Spermatophore layer | T0 | T15 | T60 | T180 | p-value |
|---|---|---|---|---|---|---|
| comp98020_c0 | inner and envelope | 12178463978.00 | 11675583776.00 | 10384784100.00 | 6593896232.00 | 0.000148 |
| comp68577_c0 | inner and envelope | 1142396053.00 | 1086181577.00 | 954939586.20 | 558773442.70 | 0.000245 |
| comp91045_c0 | inner and envelope | 18604092867.00 | 17202783714.00 | 13821633400.00 | 5223628511.00 | 0.000378 |
| comp100125_c0 | inner | 13159370978.00 | 11839896967.00 | 9083561282.00 | 16871326.89 | 0.000651 |
| comp54052_c0 | inner and envelope | 5610277262.00 | 5236692291.00 | 4398478180.00 | 1643335667.00 | 0.000663 |
| comp93096_c0 | inner and envelope | 8159077520.00 | 7471577849.00 | 5693764177.00 | 1766787193.00 | 0.001465 |
| comp96307_c0 | inner and envelope | 135000000000.00 | 132000000000.00 | 98799496333.00 | 11911454811.00 | 0.001647 |
| comp83827_c1 | inner and envelope | 10565825822.00 | 9617699678.00 | 7665913411.00 | 3042627489.00 | 0.001715 |
| comp54021_c0 | envelope | 226773920.00 | 203705308.40 | 145003475.80 | 22218288.89 | 0.002366 |
| comp99478_c0 | inner and envelope | 3869839318.00 | 3791569834.00 | 2675041169.00 | 494943974.40 | 0.002495 |
| comp96436_c0 | inner | 39435584078.00 | 37609919422.00 | 24522390389.00 | 154682305.30 | 0.002629 |
| comp54052_c0 | inner and envelope | 102000000000.00 | 98164222478.00 | 77170519311.00 | 38764914200.00 | 0.002908 |
| comp54021_c0 | envelope | 613683629.80 | 558544949.60 | 389197664.40 | 68931796.00 | 0.003396 |
| comp96307_c0 | inner and envelope | 16802938533.00 | 16978977544.00 | 12851802900.00 | 4005918238.00 | 0.003825 |
| comp54021_c0 | envelope | 1253724579.00 | 1268243069.00 | 1020459708.00 | 358502627.60 | 0.004759 |
| comp83824_c0 | inner and envelope | 20479323033.00 | 18967099244.00 | 14866481433.00 | 7345093106.00 | 0.004762 |
| comp81462_c0 | inner and envelope | 722601824.00 | 645594767.60 | 558193813.80 | 299111424.00 | 0.004926 |
| comp95226_c1 | inner and envelope | 18730057500.00 | 17636873333.00 | 16723886989.00 | 12693863478.00 | 0.005344 |
| comp96027_c4 | inner and envelope | 7345798253.00 | 7549716482.00 | 5259779038.00 | 905820504.00 | 0.005717 |
| comp95226_c1 | inner and envelope | 5162470053.00 | 5272967631.00 | 4344105500.00 | 2257599558.00 | 0.005811 |
| comp99478_c0 | inner and envelope | 86308124111.00 | 81730318778.00 | 71315959700.00 | 19723713544.00 | 0.006367 |
| comp91045_c0 | inner and envelope | 13050389889.00 | 11538884469.00 | 8761891869.00 | 3353878828.00 | 0.007005 |
| comp99478_c0 | inner and envelope | 8237350158.00 | 8288160984.00 | 7661214411.00 | 6625972478.00 | 0.007431 |
| comp54021_c0 | envelope | 669132959.10 | 550001269.80 | 469108198.70 | 93825874.67 | 0.007982 |
| comp96027_c4 | inner and envelope | 124000000000.00 | 129000000000.00 | 100000000000.00 | 13993825733.00 | 0.009048 |
| comp95226_c1 | inner and envelope | 1298456490.00 | 1152984743.00 | 1073801084.00 | 593091064.90 | 0.009499 |
| comp93096_c0 | inner and envelope | 1266017734.00 | 1323967966.00 | 847769225.80 | 111639716.90 | 0.00992 |

| Protein of origin | Spermatophore layer | T0 | T15 | T60 | T180 | p-value |
|---|---|---|---|---|---|---|
| comp85484_c0 | envelope | 751912745.60 | 777319439.80 | 603534357.10 | 6006393.94 | 0.010068 |
| comp91045_c0 | inner and envelope | 41738020400.00 | 39189090344.00 | 26241022144.00 | 8575425478.00 | 0.011145 |
| comp90489_c1 | inner | 211504139.60 | 177778922.20 | 119874514.20 | 16974821.78 | 0.01126 |
| comp98020_c0 | inner and envelope | 586742968.90 | 542480955.30 | 519951644.90 | 309843174.00 | 0.011625 |
| comp95226_c1 | inner and envelope | 4839998900.00 | 4503245209.00 | 3534621928.00 | 2119057529.00 | 0.011706 |
| comp96027_c4 | inner and envelope | 8354337078.00 | 8341503013.00 | 7060258139.00 | 717326923.30 | 0.012764 |
| comp81135_c0 | inner and envelope | 2671176946.00 | 2553103512.00 | 1613411946.00 | 435340489.30 | 0.013619 |
| comp96307_c0 | inner and envelope | 63661082889.00 | 63972581444.00 | 38757897433.00 | 7181383992.00 | 0.013791 |
| comp25711_c0 | inner and envelope | 170783.67 | 227094.67 | 479171.28 | 799876.22 | 0.014488 |
| comp91676_c0 | inner and envelope | 25060389133.00 | 20688512444.00 | 17469361937.00 | 7752429907.00 | 0.015968 |
| comp95226_c1 | inner and envelope | 20346298033.00 | 19669964800.00 | 16708661189.00 | 13150904778.00 | 0.017229 |
| comp83824_c0 | inner and envelope | 46707098800.00 | 40911680500.00 | 39775926300.00 | 19259254056.00 | 0.018107 |
| comp96027_c4 | inner and envelope | 157000000000.00 | 148000000000.00 | 138000000000.00 | 31511238856.00 | 0.018354 |
| comp83827_c1 | inner and envelope | 22999568856.00 | 20834868333.00 | 13814495728.00 | 5308038210.00 | 0.019075 |
| comp95226_c1 | inner and envelope | 2071508978.00 | 1977928307.00 | 1944149617.00 | 1349892942.00 | 0.019135 |
| comp97244_c1 | inner | 4814124523.00 | 4811607397.00 | 4358758239.00 | 1209735781.00 | 0.019921 |
| comp85455_c0 | inner and envelope | 555000000000.00 | 567000000000.00 | 540000000000.00 | 453000000000.00 | 0.021039 |
| comp54021_c0 | envelope | 4226224342.00 | 3981716330.00 | 3937433357.00 | 2670808773.00 | 0.021127 |
| comp98020_c0 | inner and envelope | 674620549.80 | 564018079.10 | 514188431.10 | 291438866.90 | 0.021918 |
| comp98020_c0 | inner and envelope | 31497895156.00 | 29396568356.00 | 27500626844.00 | 23388701978.00 | 0.022403 |
| comp54021_c0 | envelope | 3218170018.00 | 3392885304.00 | 2983699032.00 | 1276479871.00 | 0.022523 |
| comp90489_c1 | inner | 660127278.00 | 481554748.70 | 459638116.20 | 10301135.11 | 0.024242 |
| comp96027_c4 | inner and envelope | 16653155800.00 | 16825723456.00 | 15131745211.00 | 1441958450.00 | 0.02482 |
| comp98020_c0 | inner and envelope | 2103339093.00 | 2045142385.00 | 1353439502.00 | 674354172.90 | 0.025026 |
| comp100125_c0 | inner | 91454538889.00 | 85031788256.00 | 82982399889.00 | 24215332533.00 | 0.026119 |
| comp96307_c0 | inner and envelope | 60936737556.00 | 55662205811.00 | 33303972633.00 | 10556181567.00 | 0.02627 |
| comp98020_c0 | inner and envelope | 880946531.60 | 795342872.70 | 543435447.60 | 273414127.80 | 0.027014 |
| comp54021_c0 | envelope | 2129902234.00 | 1782066693.00 | 1801317867.00 | 606978036.00 | 0.028535 |
| comp54021_c0 | envelope | 876069505.90 | 933290151.10 | 792034289.30 | 596312644.00 | 0.029396 |

| Protein of origin | Spermatophore layer | T0 | T15 | T60 | T180 | p-value |
|---|---|---|---|---|---|---|
| comp96027_c4 | inner and envelope | 68619626444.00 | 78662259289.00 | 62149756778.00 | 10535491411.00 | 0.030172 |
| comp96307_c0 | inner and envelope | 17261530678.00 | 16245310497.00 | 16304556722.00 | 9293150211.00 | 0.031828 |
| comp95236_c0 | inner and envelope | 953348942.70 | 763146596.20 | 431973761.60 | 37554729.89 | 0.031897 |
| comp96436_c0 | inner | 55878118333.00 | 49301196567.00 | 23798534922.00 | 75542924.56 | 0.032489 |
| comp81370_c0 | inner | 1986603790.00 | 1948508074.00 | 902277542.00 | 12249652.06 | 0.03306 |
| comp83824_c0 | inner and envelope | 15874627300.00 | 13582092178.00 | 13552215833.00 | 8565882700.00 | 0.033831 |
| comp95226_c1 | inner and envelope | 8583726889.00 | 9507742100.00 | 7801590044.00 | 4927038904.00 | 0.035416 |
| comp75578_c0 | inner | 15820779878.00 | 15969307878.00 | 15720428422.00 | 13124165589.00 | 0.037756 |
| comp97068_c0 | inner and envelope | 4804786702.00 | 5290686144.00 | 4595159593.00 | 2679034888.00 | 0.037795 |
| comp92111_c1 | inner and envelope | 6044418378.00 | 5309198926.00 | 5474211454.00 | 3088810143.00 | 0.038783 |
| comp81135_c0 | inner and envelope | 2891999616.00 | 3096098510.00 | 1716962427.00 | 593333645.10 | 0.038916 |
| comp54021_c0 | envelope | 825391615.80 | 861075650.20 | 815976101.80 | 569091753.80 | 0.039186 |
| comp75578_c0 | inner | 4192285672.00 | 3933510786.00 | 4008221761.00 | 2919218856.00 | 0.039858 |
| comp96436_c0 | inner | 3006048572.00 | 3607161386.00 | 1776532572.00 | 11178424.22 | 0.040407 |
| comp96307_c0 | inner and envelope | 49023589444.00 | 43398083611.00 | 22617045411.00 | 5364667432.00 | 0.040549 |
| comp91676_c0 | inner and envelope | 18952342644.00 | 20792691044.00 | 18032530433.00 | 12148867933.00 | 0.04067 |
| comp54052_c0 | inner and envelope | 163000000000.00 | 156000000000.00 | 158000000000.00 | 79589089333.00 | 0.041906 |
| comp95226_c1 | inner and envelope | 7401551071.00 | 8221213589.00 | 6964695153.00 | 4545735596.00 | 0.042444 |
| comp97068_c0 | inner and envelope | 40807102878.00 | 37005286607.00 | 32524110078.00 | 26783539933.00 | 0.043039 |
| comp68577_c0 | inner and envelope | 3925280799.00 | 3171228280.00 | 2601282090.00 | 1576539657.00 | 0.045165 |
| comp95236_c0 | inner and envelope | 15534919578.00 | 13219719733.00 | 11686010367.00 | 8634673600.00 | 0.047117 |
| comp103890_c0 | inner and envelope | 1468927182.00 | 1290786979.00 | 544978172.70 | 11418662.50 | 0.049701 |

## Bibliography


Adams, G. P., Ratto, M. H., Huanca, W., & Singh, J. (2005). Ovulation-inducing factor in the seminal plasma of alpacas and llamas. *Biology of Reproduction*, *73*(3), 452–7. http://doi.org/10.1095/biolreprod.105.040097

Ah-King, M., Barron, A., & Herberstein, M. (2014). Genital evolution: why are females still understudied? *PLoS Biology*, *12*(5), 1–7. http://doi.org/10.1371/journal.pbio.1001851

Ahmed-Braimah, Y. H. (2016). Multiple genes cause postmating prezygotic reproductive isolation in the Drosophila virilis group. *G3*, *6*(October), 4067–4076. http://doi.org/10.1534/g3.116.033340

Ajamhassani, M., Zibaee, A., Sendi, J., Askary, H., & Farrar, N. (2012). Proteolytic Activity in the Midgut of the Crimson Speckled Moth Utethesia Pulchella L. (Lepidoptera: Arctiidae). *Journal of Plant Protection Research*, *52*(3), 3–8. http://doi.org/10.2478/v10045-012-0061-0

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 moth (Ostrinia nubilalis). *BMC Genomics*, *15*(1), 189. http://doi.org/10.1186/1471-2164-15-189

Al-Wathiqui, N., Lewis, S. M., & Dopman, E. B. (2017). Molecular dissection of nuptial gifts in divergent strains of *O strinia* moths. *Physiological Entomology*. http://doi.org/10.1111/phen.12220

Alfonso-Parra, C., Ahmed-Braimah, Y. H., Degner, E. C., Avila, F. W., Villarreal, S. M., Pleiss, J. A., … Harrington, L. C. (2016). Mating-Induced Transcriptome Changes in the Reproductive Tract of Female Aedes aegypti. *PLoS Neglected Tropical Diseases*, *10*(2), 1–24. http://doi.org/10.1371/journal.pntd.0004451

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10. http://doi.org/10.1016/S0022-2836(05)80360-2

Aoyagi, T., Miyata, S., Nanbo, M., Kojima, F., Matsuzaki, M., Ishizuka, M., … Umezawa, H. (1969). Biological activities of leupeptins. *The Journal of Antibiotics*, *22*(11), 558–68. http://doi.org/10.7164/antibiotics.22.558

Arnqvist, G., & Nilsson, T. (2000). The evolution of polyandry: multiple mating and female fitness in insects. *Animal Behaviour*, *60*(2), 145–164. http://doi.org/10.1006/anbe.2000.1446

Arnqvist, G., & Rowe, L. (1995). Sexual Conflict and Arms Races between the Sexes: A Morphological Adaptation for Control of Mating in a Female Insect. *Proceedings of the Royal*

*Society B: Biological Sciences*, *261*(1360), 123–127. http://doi.org/10.1098/rspb.1995.0126

Avila, F. W., & Wolfner, M. F. (2017). Cleavage of the Drosophila seminal protein Acp36DE in mated females enhances its sperm storage activity. *Journal of Insect Physiology*, *101*(52), 66–72. http://doi.org/10.1016/j.jinsphys.2017.06.015

Baer, B., Morgan, E. D., & Schmid-Hempel, P. (2001). A nonspecific fatty acid within the bumblebee mating plug prevents females from remating. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 3926–3928. http://doi.org/10.1073/pnas.061027998

Berenbaum, M. (1980). Adaptive significance of midgut pH in larval Lepidoptera. *American Naturalist*, *115*(1), 138–146.

Biram Saheb, N. M., Singh, T., Kalappa, H. K., & Saratchandra, B. (2005). Mating behaviour in Mulberry Silkworm, Bombyx mori (L). *International Journal of Industrial Entomology2*, *10*(2), 87–94.

Bissoondath, C., & Wiklund, C. (1997). Effect of male body size on sperm precedence in the polyandrous butterfly Pieris napi L.(Lepidoptera: Pieridae). *Behavioral Ecology*, *8*(5), 518–523.

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

Bono, J., Matzkin, L., Kelleher, E. S., & Markow. (2011). Postmating transcriptional changes in reproductive tracts of con-and heterospecifically mated Drosophila mojavensis females. *Proceedings of the National Academy of Sciences*, *108*(19), 7878–7883. http://doi.org/10.1073/pnas.1100388108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1100388108

Brennan, P. L. R., & Prum, R. O. (2014). Mechanisms and evidence of genital coevolution: the role of natural selection, mate choice, and secual conflict. *The Genetics and Biology of Sexual Conflict*, 385–413. http://doi.org/10.1101/cshperspect.a017749

Bretman, A., Lawniczak, M. K. N., Boone, J., & Chapman, T. (2010). A mating plug protein reduces early female remating in Drosophila melanogaster. *Journal of Insect Physiology*, *56*(1), 107–13. http://doi.org/10.1016/j.jinsphys.2009.09.010

Broadway, R. (1996). Dietary proteinase inhibitors alter complement of midgut proteases. *Archives of Insect Biochemistry and Physiology*, *53*(1 996), 39–53.

Broadway, R. M., & Duffey, S. S. (1986). The effect of dietary protein on the growth and digestive physiology of larval Heliotis zea and Spodoptera exigua. *Journal of Insect Physiology*, *32*(8), 673–680.

Brocklehurst, K., & Kierstan, M. P. J. (1973). Propapain and its conversion to papain: A new type of zymogen activation mechanism involving intramolecular thiol-disulphide interchange.

*Nature New Biology*, *242*(119), 167–170. http://doi.org/10.1038/newbio242167a0

Burns, J. M. (1968). Mating frequency in natural populations of skippers and butterflies as determined by spermatophore counts. *Zoology*, *61*, 852–859.

Clark, N., Alani, E., & Aquadro, C. (2012). Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Research*, *22*, 714–720. http://doi.org/10.1101/gr.132647.111.Guerois

Clark, N. L., Aagaard, J. E., & Swanson, W. J. (2006). Evolution of reproductive proteins from animals and plants. *Reproduction (Cambridge, England)*, *131*(1), 11–22. http://doi.org/10.1530/rep.1.00357

Cordero, C. (2005). The evolutionary origin of signa in female Lepidoptera: natural and sexual selection hypotheses. *Journal of Theoretical Biology*, *232*(3), 443–9. http://doi.org/10.1016/j.jtbi.2004.08.031

Cornwallis, C. K., & Birkhead, T. R. (2007). Changes in sperm quality and numbers in response to experimental manipulation of male social status and female attractiveness. *The American Naturalist*, *170*(5), 758–70. http://doi.org/10.1086/521955

Cramer, E. R. A., Alund, M., Mcfarlane, S. E., Johnsen, A., & Qvarnstrom, A. (2016). Females discriminate against heterospecific sperm in a natural hybrid zone. *Evolution*, *70*(8), 1844–1855. http://doi.org/10.1111/evo.12986

Dean, M. D. (2013). Genetic disruption of the copulatory plug in mice leads to severely reduced fertility. *PLoS Genetics*, *9*(1), e1003185. http://doi.org/10.1371/journal.pgen.1003185

Dean, M. D., Clark, N. L., Findlay, G. D., Karn, R. C., Yi, X., Swanson, W. J., … Nachman, M. W. (2009). Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (Mus domesticus). *Molecular Biology and Evolution*, *26*(8), 1733–43. http://doi.org/10.1093/molbev/msp094

Dobzhansky, T. (1951). Genetics and the Origin of Species (3rd edn), *94*(July), 7768–7775.

Dong, Z., Zhang, Y., Xia, Q., Wang, X., Zhao, P., Chen, Q., … Zhang, L. (2016). Proteome profiling reveals tissue-specific protein expression in male and female accessory glands of the silkworm, Bombyx mori. *Amino Acids*, *48*(5), 1173–1183. http://doi.org/10.1007/s00726-015-2141-8

Durrant, J. D. (2018). Structural bioinformatics BlendMol : advanced macromolecular visualization in Blender. *Bioinformatics*, 1–3. http://doi.org/10.1093/bioinformatics/bty968

Eberhard, W. G. (1985). *Sexual Selection and Animal Genitalia*. Cambridge: Harvard University Press.

Eddy, S. R. (2009). A NEW GENERATION OF HOMOLOGY SEARCH TOOLS BASED ON PROBABILISTIC INFERENCE. *Genome Informatics*, *23*, 205–211.

Engelmann, F. (1970). *The physiology of insect reproduction. Int. Ser. Monogr. pure appl. Biol.* New York: Pergamon Press.

Findlay, G. D., Sitnik, J. L., Wang, W., Aquadro, C. F., Clark, N. L., & Wolfner, M. F. (2014). Evolutionary rate covariation identifies new members of a protein network required for Drosophila melanogaster female post-mating responses. *PLoS Genetics*, *10*(1), e1004108. http://doi.org/10.1371/journal.pgen.1004108

Firman, R. C., Gasparini, C., Manier, M. K., & Pizzari, T. (2017). Postmating Female Control: 20 Years of Cryptic Female Choice. *Trends in Ecology and Evolution*, *32*(5), 368–382. http://doi.org/10.1016/j.tree.2017.02.010

Fiumera, A. C., Dumont, B. L., & Clark, A. G. (2005). Sperm competitive ability in Drosophila melanogaster associated with variation in male reproductive proteins. *Genetics*, *169*(1), 243–57. http://doi.org/10.1534/genetics.104.032870

Fiumera, A. C., Dumont, B. L., & Clark, A. G. (2006). Natural variation in male-induced "cost-of-mating" and allele-specific association with male reproductive genes in Drosophila melanogaster. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1466), 355–61. http://doi.org/10.1098/rstb.2005.1791

Fukano, Y., Satoh, T., Hirota, T., Nishide, Y., & Obara, Y. (2012). Geographic expansion of the cabbage butterfly (Pieris rapae) and the evolution of highly UV-reflecting females. *Insect Science*, *19*(2), 239–246. http://doi.org/10.1111/j.1744-7917.2011.01441.x

Galindo, B. E., Vacquier, V. D., & Swanson, W. J. (2003). Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(8), 4639–43. http://doi.org/10.1073/pnas.0830022100

Gasmi, L., Boulain, H., Gauthier, J., Hua-Van, A., Musset, K., Jakubowska, A. K., … Drezen, J. M. (2015). Recurrent Domestication by Lepidoptera of Genes from Their Parasites Mediated by Bracoviruses. *PLoS Genetics*, *11*(9), 1–32. http://doi.org/10.1371/journal.pgen.1005470

Gilbert, N. (1984). Control of fecundity in Pieris rapae: I. The problem. *The Journal of Animal Ecology*, *53*(2), 581–588.

Gillott, C. (2003). Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annual Review of Entomology*, *48*, 163–84. http://doi.org/10.1146/annurev.ento.48.091801.112657

Gold, A. M. (1965). Sulfonyl Fluorides as Inhibitors of Esterases. III. Identification of Serine as the Site of Sulfonylation in Phenylmethanesulfonyl α-Chymotrypsin. *Biochemistry*, *4*(5), 897–901. http://doi.org/10.1021/bi00881a016

Grant, B. R., & Grant, P. R. (2002). Simulating secondary contact in allopatric speciation : an empirical test of premating isolation, 545–556.

Granvogl, B., Gruber, P., & Eichacker, L. A. (2007). Standardisation of rapid in-gel digestion by

mass spectrometry. *Proteomics*, *7*(5), 642–54. http://doi.org/10.1002/pmic.200600607

Granvogl, B., Plöscher, M., & Eichacker, L. A. (2007). Sample preparation by in-gel digestion for mass spectrometry-based proteomics. *Analytical and Bioanalytical Chemistry*, *389*(4), 991–1002. http://doi.org/10.1007/s00216-007-1451-4

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–512. http://doi.org/10.1038/nprot.2013.084

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., … Pochet, N. (2014). *reference generation and analysis with Trinity*. *Nat Protocols* (Vol. 8). http://doi.org/10.1038/nprot.2013.084.De

Heifetz, Y., Lindner, M., Garini, Y., & Wolfner, M. F. (2014). Report Mating Regulates Neuromodulator Ensembles at Nerve Termini Innervating the Drosophila Reproductive Tract. *Current Biology*, *24*(7), 731–737. http://doi.org/10.1016/j.cub.2014.02.042

Hirota, M., Ohmuraya, M., & Baba, H. (2006). The role of trypsin, trypsin inhibitor, and trypsin receptor in the onset and aggravation of pancreatitis. *Journal of Gastroenterology*, *41*(9), 832–6. http://doi.org/10.1007/s00535-006-1874-2

Holman, L., & Snook, R. R. (2006). Spermicide, cryptic female choice and the evolution of sperm form and function. *Journal of Evolutionary Biology*, *19*(5), 1660–70. http://doi.org/10.1111/j.1420-9101.2006.01112.x

Hoskin, C. J., Higgie, M., Mcdonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation, *437*(October), 10–13. http://doi.org/10.1038/nature04004

Karlsson, B. (1998). Nuptial gifts, resource budgets, and reproductive output in a polyandrous butterfly. *Ecology*, *79*(8), 2931–2940.

Kelleher, E. S., Clark, N. L., & Markow, T. A. (2011). Diversity-enhancing selection acts on a female reproductive protease family in four subspecies of Drosophila mojavensis. *Genetics*, *187*(3), 865–76. http://doi.org/10.1534/genetics.110.124743

Kelleher, E. S., & Pennington, J. E. (2009). Protease gene duplication and proteolytic activity in Drosophila female reproductive tracts. *Molecular Biology and Evolution*, *26*(9), 2125–2134. http://doi.org/10.1093/molbev/msp121

Kelleher, E. S., Swanson, W. J., & Markow, T. A. (2007). Gene duplication and adaptive evolution of digestive proteases in Drosophila arizonae female reproductive tracts. *PLoS Genetics*, *3*(8), e148. http://doi.org/10.1371/journal.pgen.0030148

Khalifa, A. (1949). Spermatophore production in Trichoptera and some other insects. *Transactions of the Royal Entomological Society of London*, *100*(19), 449–71.

Khalifa, A. (1950). Spermatophore production in Galleria Mellonella L. (Lepidoptera). *Proceedings of the Royal Entomological Society of …*, *25*, 33–42.

Knowles, L. L., & Markow, T. A. (2001). Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 8692–6. http://doi.org/10.1073/pnas.151123998

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., … Schlötterer, C. (2011). Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*, *6*(1). http://doi.org/10.1371/journal.pone.0015925

Kubo, Y., Okazaki, S., Anzai, T., & Fujiwara, H. (2001). Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. *Molecular Biology and Evolution*, *18*(5), 848–857. http://doi.org/10.1093/oxfordjournals.molbev.a003866

LaFlamme, B. A., & Wolfner, M. F. (2013). Identification and function of proteolysis regulators in seminal fluid. *Mol Reprod Dev*, *80*(2), 80–101. http://doi.org/10.1002/mrd.22130

Lawniczak, M. K. N., & Begun, D. J. (2007). Molecular population genetics of female-expressed mating-induced serine proteases in Drosophila melanogaster. *Molecular Biology and Evolution*, *24*(9), 1944–1951. http://doi.org/10.1093/molbev/msm122

Lee, Y. H., Ota, T., & Vacquier, V. D. (1995). Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol Biol Evol*, *12*(2), 231–238. http://doi.org/10.1093/oxfordjournals.molbev.a040200

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B*, *363*, 3009–3021. http://doi.org/10.1098/rstb.2008.0064

Mack, P., Kapelnikov, A., Heifetz, Y., & Bender, M. (2006). Mating-responsive genes in reproductive tissues of female Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *103*(27), 10358–10363.

Maclean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., … Maccoss, M. J. (2010). Skyline : an open source document editor for creating and analyzing targeted proteomics experiments, *26*(7), 966–968. http://doi.org/10.1093/bioinformatics/btq054

Marshall, L. (1985). Protein and Lipid Composition of Colias philodice and C. eurytheme Spermatophores and Their Changes Over Time (Pieridae). *Journal of Reasearch on the Lepidoptera*, *24*(1), 21–30.

Mayr, E. (1942). *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. *Nature* (Vol. 151). New York: Columbia University Press. http://doi.org/10.1038/151347a0

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press.

McDonough, C. E., Whittington, E., Pitnick, S., & Dorus, S. (2016). Proteomics of reproductive systems: Towards a molecular understanding of postmating, prezygotic reproductive barriers. *Journal of Proteomics*, *135*, 26–37. http://doi.org/10.1016/j.jprot.2015.10.015

McGraw, L. A., Clark, A. G., & Wolfner, M. F. (2008). Post-mating gene expression profiles of female Drosophila melanogaster in response to time and to four male accessory gland proteins. *Genetics*, *179*(3), 1395–408. http://doi.org/10.1534/genetics.108.086934

McNamara, K. B., Dougherty, L. R., Wedell, N., & Simmons, L. W. (2019). Experimental evolution reveals divergence in female genital teeth morphology in response to sexual conflict intensity in a moth. *Journal of Evolutionary Biology*, *32*(5), 519–524. http://doi.org/10.1111/jeb.13428

Mcqueen, E. W., & Morehouse, N. I. (2018). Rapid Divergence of Wing Volatile Profiles Between Subspecies of the Butterfly Pieris rapae ( Lepidoptera : Pieridae ). *Journal of Insect Science*, *18*(2). http://doi.org/10.1093/jisesa/iey026

Mendez, V., & Cordoba-Aguilar, A. (2004). Sexual selection and animal genitalia. *Trends in Ecology & Evolution*, *19*(5), 224–5. http://doi.org/10.1016/j.tree.2004.03.005

Meslin, C., Cherwin, T. S., Plakke, M. S., Small, B. S., Goetz, B. J., Morehouse, N., & Clark, N. L. (2017). Structural complexity and molecular heterogeneity of a butterfly ejaculate reflecct a complex history of selection. *Proceedings of the National Academy of Sciences*, *114*(27), 1–8. http://doi.org/10.1073/pnas.1707680114

Meslin, C., Cherwin, T. S., Plakke, M. S., Small, B. S., Goetz, B. J., Morehouse, N. I., & Clark, N. L. (2017). Structural complexity and molecular heterogeneity of a butterfly ejaculate reflect a complex history of selection. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(27). http://doi.org/10.1073/pnas.1707680114

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. *Molecular Biology and Evolution*, *32*(6), 1567–1580. http://doi.org/10.1093/molbev/msv048

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., … Finn, R. D. (2015). The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Research*, *43*, D213–D221. http://doi.org/10.1093/nar/gku1243

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean ?, *9*(8), 1–8. http://doi.org/10.1371/journal.pbio.1001127

Moscardi, F., Souza, M. L. De, Castro, M. E. B. de, Moscardi, M. L., & Szewczyk, B. (2011). *Baculovirus Pesticides : Present State and Future Perspectives*. http://doi.org/10.1007/978-1-4419-7931-5

Murtaugh, P. A. (2007). Simplicity and complexity in ecological data analysis. *Ecology*, *88*(1), 56–62.

Nagarkatti, S., & Nagaraja, H. (1971). Redescriptions of some known species of Trichogramma (Hym., Trichogrammatidae), showing the importance of the male genitalia as a diagnostic character. *Bulletin of Entomological Research*, *61*(1), 13–31.

Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry abilities that proteins are present in a sample on the basis. *Anal. Chem.*, *75*(17), 4646–4658. http://doi.org/10.1021/AC0341261

Obara, Y., & Majerus, M. (2000). Initial Mate Recognition in the British Cabbage Butterfly, Pieris rapae rapae. *Zoological Science*, *17*(6), 725–730. http://doi.org/10.2108/zsj.17.725

Obara, Y., & Majerus, M. E. N. (2009). Inaccurate mate recognition as a mating strategy of a "pioneer male". *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *85*(6), 198–203. http://doi.org/10.2183/pjab.85.198

Obara, Y., Ozawa, G., Fukano, Y., Watanabe, K., & Satoh, T. (2008). Mate preference in males of the cabbage butterfly, Pieris rapae crucivora, changes seasonally with the change in female UV color. *Zoological Science*, *25*(1), 1–5. http://doi.org/10.2108/zsj.25.1

Oberhauser, K. (1992). Rate of ejaculate breakdown and intermating intervals in monarch butterflies. *Behavioral Ecology and Sociobiology*, *31*(5), 367–373.

Oberhauser, K. S. (1989). Effects of spermatophores on male and female monarch butterfly reproductive success. *Behavioral Ecology and Sociobiology*, *25*(4), 237–246.

Oberhauser, K. S. (1997). Fecundity , lifespan and egg mass in butterflies : effects of male-derived nutrients and female size. *Functional Ecology*, *11*(2), 166–175.

Orr, H. (2005). The genetic basis of reproductive isolation: insights from Drosophila. *Proceedings of the National Academy of Sciences*, *102*, 6522–6526.

Osanai, M., & Kasuga, H. (1990). Role of endopeptidase in motility induction in apyrene silkworm spermatozoa; micropore formation in the flagellar membrane. *Experientia*, *46*, 261–264.

Osanai, M., Kasuga, H., & Aigaki, T. (1987). The spermatophore and its structural changes with time in the bursa copulatrix of the silkworm, Bombyx mori. *Journal of Morphology*, *11*, 1–11.

Osanai, M., Kasuga, H., & Aigaki, T. (1989). Induction of motility of apyrene spermatozoa and dissociation of eupyrene sperm bundles of the silkmoth, Bombyx mori, by: Initiatorin and trypsin. *Invertebrate Reproduction and Development*, *15*(2), 97–103. http://doi.org/10.1080/07924259.1989.9672029

Ou, C. M., Tang, J. B., Huang, M. S., Sudhakar Gandhi, P. S., Geetha, S., Li, S. H., & Chen, Y. H. (2012). The mode of reproductive-derived Spink (serine protease inhibitor Kazal-type)

action in the modulation of mammalian sperm activity. *International Journal of Andrology*, *35*(1), 52–62. http://doi.org/10.1111/j.1365-2605.2011.01159.x

Patankar, A. G., Giri, A. P., Harsulkar, A. M., Sainani, M. N., Deshpande, V. V., Ranjekar, P. K., & Gupta, V. S. (2001). Complexity in specificities and expression of Helicoverpa armigera gut proteinases explains polyphagous nature of the insect pest. *Insect Biochemistry and Molecular Biology*, *31*(4–5), 453–464. http://doi.org/10.1016/S0965-1748(00)00150-8

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, *20*(18), 3551–67. http://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Perry, J. C., Sirot, L., & Wigby, S. (2013). The seminal symphony: how to compose an ejaculate. *Trends in Ecology & Evolution*, *28*(7), 414–22. http://doi.org/10.1016/j.tree.2013.03.005

Pitnick, S., Miller, G. T., Schneider, K., & Markow, T. A. (2003). Ejaculate-female coevolution in Drosophila mojavensis. *Proceedings. Biological Sciences / The Royal Society*, *270*(1523), 1507–12. http://doi.org/10.1098/rspb.2003.2382

Plakke, M. S., Deutsch, A. B., Meslin, C., Clark, N. L., & Morehouse, N. I. (2015). Dynamic digestive physiology of a female reproductive organ in a polyandrous butterfly. *Journal of Experimental Biology*, *218*(10), 1548–1555. http://doi.org/10.1242/jeb.118323

Plakke, M. S., Walker, J. L., Lombardo, J. B., Goetz, B. J., Pacella, G. N., Durrant, J. D., … Morehouse, N. I. (n.d.). Characterization of female reproductive proteases in a butterfly from functional and evolutionary perspectives. *PBZ*, 1–29.

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. *Journal of Evolutionary Biology*, *23*(7), 1386–1398. http://doi.org/10.1111/j.1420-9101.2010.01998.x

Ram, K. R., & Wolfner, M. M. F. (2007). Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. *Integrative and Comparative Biology*, *47*(3), 427–45. http://doi.org/10.1093/icb/icm046

Ramm, S. a, & Stockley, P. (2009). Adaptive plasticity of mammalian sperm production in response to social experience. *Proceedings. Biological Sciences / The Royal Society*, *276*(1657), 745–51. http://doi.org/10.1098/rspb.2008.1296

Raser, K. J., Posner, A., & Wang, K. K. (1995). Casein zymography: a method to study u-calpain, m-calpain, and their inhibitory agents. *Archives of Biochemistry and Biophysics*. http://doi.org/10.1006/abbi.1995.1284

Reijden, E., Monchamp, J., & Lewis, S. (1997). The formation, transfer, and fate of spermatophores in Photinus fireflies (Coleoptera: Lampyridae). *Canadian Journal of Zoology*, *75*, 1202–1207.

Reinhart, M., & Carney, T. (2014). Characterizing Male–Female Interactions Using Natural Genetic Variation in Drosophila melanogaster. *Journal of Heredity*, 1–13. http://doi.org/10.5061/dryad.2856m

Robinson, M. D., Mccarthy, D. J., & Smyth, G. K. (2010). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

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

Rogers, S. H., & Wells, H. (1984). The Structure and Function of the Bursa Copulatrix of the Monarch Butterfly ( Danaus plexippus ). *Journal of Morphology*, *221*, 213–221.

Rowe, L., & Day, T. (2006). Detecting sexual conflict and sexually antagonistic coevolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1466), 277–85. http://doi.org/10.1098/rstb.2005.1788

Rowe, M., Albrecht, T., Cramer, E. R. A., Johnsen, A., Laskemoen, T., Weir, J. T., & Lifjeld, J. T. (2015). Postcopulatory sexual selection is associated with accelerated evolution of sperm morphology. *Evolution*, *69*(4), 1044–1052. http://doi.org/10.1111/evo.12620

Roy, S., Choudhury, D., Aich, P., Dattagupta, J. K., & Biswas, S. (2012). The structure of a thermostable mutant of pro-papain reveals its activation mechanism. *Acta Crystallographica Section D: Biological Crystallography*, *68*(12), 1591–1603. http://doi.org/10.1107/S0907444912038607

Rutowski, R., & Gilchrist, G. (1986). Copulation in Colias eurytheme (Lepidoptera: Pieridae): patterns and frequency. *Journal of Zoology*, *209*, 115–124.

Ryan, S. F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dincă, V., … Shoemaker, D. (2018). Global invasion history of the world's most abundant pest butterfly: a citizen science population genomics study. *BioRxiv*, 506162. http://doi.org/10.1101/506162

Sánchez, V., & Cordero, C. (2014a). Sexual coevolution of spermatophore envelopes and female genital traits in butterflies: Evidence of male coercion? *PeerJ*, *2*, e247. http://doi.org/10.7717/peerj.247

Sánchez, V., & Cordero, C. (2014b). Sexual coevolution of spermatophore envelopes and female genital traits in butterflies: Evidence of male coercion? *PeerJ*, *2*, e247. http://doi.org/10.7717/peerj.247

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), e22642. http://doi.org/10.1371/journal.pone.0022642

Schmidt, A., Jelsch, C., Østergaard, P., Rypniewski, W., & Lamzin, V. S. (2003). Trypsin Revisited: Crystallography at (SUB) atomic resolution and quantum chemistry revealing

details of catalysis. *Journal of Biological Chemistry*, *278*(44), 43357–43362. http://doi.org/10.1074/jbc.M306944200

Schneider, C. a, Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. http://doi.org/10.1038/nmeth.2089

Schröder, E., Garman, E., Harlos, K., & Crawford, C. (1993). X-ray crystallographic structure of a papain-leupeptin complex. *FEBS Letters*, *315*(1), 38–42.

Sedlazeck, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, *29*(21), 2790–2791. http://doi.org/10.1093/bioinformatics/btt468

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V, & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, *1*(6), 2856–60. http://doi.org/10.1038/nprot.2006.468

Shine, R. (2000). Chastity belts in gartersnakes: the functional significance of mating plugs. *Biological Journal of the Linnean Society*, *70*(3), 377–390. http://doi.org/10.1006/bijl.1999.0427

Simmons, L., & Gwynne, D. (1991). The refractory period of female katydids (Orthoptera: Tettigoniidae): sexual conflict over the remating interval? *Behavioral Ecology*, *2*(4), 276–282.

Simmons, L. W. (2014). Sexual selection and genital evolution. *Austral Entomology*, *53*(1), 1–17. http://doi.org/10.1111/aen.12053

Sirot, L. K., Findlay, G. D., Sitnik, J. L., Frasheri, D., Avila, F. W., & Wolfner, M. F. (2014). Molecular characterization and evolution of a gene family encoding both female- and male-specific reproductive proteins in Drosophila. *Molecular Biology and Evolution*, *31*(6), 1554–67. http://doi.org/10.1093/molbev/msu114

Sirot, L. K., Wolfner, M. F., & Wigby, S. (2011). Protein-specific manipulation of ejaculate composition in response to female mating status in Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *108*(24), 9922–9926. http://doi.org/10.1073/pnas.1100905108

Slatkin, M. (1978). Spatial patterns in the distributions of polygenic characters. *Journal of Theoretical Biology*, *70*(2), 213–228.

Smith, C., & Ryan, M. (2011). Tactic-dependent plasticity in ejaculate traits in the swordtail Xiphophorus nigrensis. *Biology Letters*, (March), 16–18.

Sugawara, T. (1979). Stretch reception in the bursa copulatrix of the butterfly, Pieris rapae crucivora, and its role in behaviour. *Journal of Comparative Physiology*, *130*, 191–199.

Sugawara, T. (1981). Fine structure of the stretch receptor in the bursa copulatrix of the butterfly,

Pieris rapae crucivora. *Cell and Tissue Research*, *217*(1), 23–36.

Suzuki, Y. (1979). Mating frequency in females of the small cabbage white, Pieris rapae crucivora Boisduval (Lepidoptera: Pieridae). *Kontyu*, *47*(3), 335–339.

Swanson, W. J., Yang, Z., Wolfner, M. F., & Aquadro, C. F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences*, *98*(5), 2509–2514. http://doi.org/10.1073/pnas.051605998

Swanson, W., & Vacquier, V. (2002). The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, *3*(February), 137–144. http://doi.org/10.1038/nrg/733

Tennessen, K. J. (1982). Review of reproductive isolating barriers in Odonata. *Advances in Odonatology*, *1*, 251–265.

Torgerson, D. G., Kulathinal, R. J., & Singh, R. S. (2002). Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Molecular Biology and Evolution*, *19*(11), 1973–80.

Turissini, D. A., McGirr, J. A., Patel, S. S., David, J. R., & Matute, D. R. (2018). The rate of evolution of postmating-prezygotic reproductive isolation in drosophila. *Molecular Biology and Evolution*, *35*(2), 312–334. http://doi.org/10.1093/molbev/msx271

Vahed, K. (1998). The function of nuptial feeding in insects: a review of empirical studies. *Biological Reviews*, *73*(1), 43–78.

van Hoef, V., Breugelmans, B., Spit, J., Simonet, G., Zels, S., Billen, J., & Vanden Broeck, J. (2011). Functional analysis of a pancreatic secretory trypsin inhibitor-like protein in insects: silencing effects resemble the human pancreatic autodigestion phenotype. *Insect Biochemistry and Molecular Biology*, *41*(9), 688–95. http://doi.org/10.1016/j.ibmb.2011.04.012

Villarreal, S. M., Pitcher, S., Helinski, M. E. H., Johnson, L., Wolfner, M. F., & Harrington, L. C. (2018). Male contributions during mating increase female survival in the disease vector mosquito Aedes aegypti. *Journal of Insect Physiology*, *108*(August 2017), 1–9. http://doi.org/10.1016/j.jinsphys.2018.05.001

Watanabe, M., & Ando, S. (1993). Influence of mating frequency on lifetime fecundity in wild females of the small white Pieris rapae (Lepidoptera, Pieridae). *Japanese Journal of Entomology*, *61*(4), 691–696.

Watanabe, M., Wiklund, C., & Bon, M. (1998). The Effect of Repeated Matings on Sperm Numbers in Successive Ejaculates of the Cabbage White Butterfly Pieris rapae ( Lepidoptera : Pieridae ). *Journal of Insect Behavior*, *11*(4), 559–570.

Wedell, N., & Cook, P. (1998). Determinants of paternity in a butterfly. *Proceedings of the Royal Society of London B*, *265*, 625–630.

Wedell, N., & Cook, P. A. (1999). Butterflies tailor their ejaculate in response to sperm competition risk and intensity. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1423), 1033–1039. http://doi.org/10.1098/rspb.1999.0740

Wigby, S., Sirot, L. K., Linklater, J. R., Buehner, N., Calboli, F. C. F., Bretman, A., … Chapman, T. (2009). Seminal Fluid Protein Allocation and Male Reproductive Success. *Current Biology*, *19*(9), 751–757. http://doi.org/10.1016/j.cub.2009.03.036

Wiklund, C., Karlsson, B., & Leimar, O. (2001). Sexual conflict and cooperation in butterfly reproduction: a comparative study of polyandry and female fitness. *Proceedings of the Royal Society of London B*, *268*, 1661–7. http://doi.org/10.1098/rspb.2001.1719

Wolfner, M. (1997). Tokens of love: functions and regulation of Drosophila male accessory gland products. *Insect Biochemistry and Molecular Biology*, *27*(3), 179–192.

Wolfner, M. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. *Heredity*, *88*, 85–93. http://doi.org/10.1038/sj/hdy/6800017

Wolfner, M. (2009). Battle and ballet: molecular interactions between the sexes in Drosophila. *Journal of Heredity*, *100*(4), 399–410. http://doi.org/10.1093/jhered/esp013

Wyckoff, G. J., Wang, W., & Wu, C. I. (2000). Rapid evolution of male reproductive genes in the descent of man. *Nature*, *403*(6767), 304–309. http://doi.org/10.1038/35002070

Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R., & Takahashi, S. Y. (2002). Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Curr.Protein Pept.Sci.*, *3*(1389–2037), 231–238. http://doi.org/10.2174/1389203024605331

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. http://doi.org/10.1038/nmeth.3213

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), 33–7. http://doi.org/10.1038/nature06483

Zougman, A., Nagaraj, N., Mann, M., & Wiśniewski, J. R. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–62. http://doi.org/10.1038/nmeth.1322