Characterization of Female Reproductive Proteases in a Butterfly from Functional and Evolutionary Perspectives

Melissa S. Plakke[1,*]
Jennifer L. Walker[1]
Jeffrey B. Lombardo[1]
Breanna J. Goetz[1]
Gina N. Pacella[1]
Jacob D. Durrant[1]
Nathan L. Clark[2,†]
Nathan I. Morehouse[3]
[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; [2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; [3]Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221

Accepted 8/1/2019; Electronically Published 10/4/2019

Online enhancements: supplemental tables.

ABSTRACT

Molecules that mediate reproductive interactions are some of the most rapidly evolving traits. Researchers have often suggested that this is due to coevolution at key physiological interfaces. However, very few of these interfaces are well understood at the functional level. One such interface is the digestion of the spermatophore in Lepidoptera. Female Lepidoptera have a specialized reproductive organ called the bursa copulatrix that receives and processes the male spermatophore, a complex proteinaceous ejaculate. In the cabbage white butterfly, *Pieris rapae*, the bursa secretes a mixture of proteases hypothesized to digest the spermatophore. However, these proteases remain biochemically uncharacterized. Using a zymogram approach, we identified six proteases in bursal extracts at sufficiently high concentrations to characterize their in vitro activity. We assessed the modes of action of these bursal enzymes by quantifying their activity following exposure to diagnostic protease inhibitors. A serine protease–specific inhibitor failed to reduce bursal protease digestion of casein. However, a cysteine protease–specific inhibitor did decrease the activity of some proteases. To explore the possible molecular mechanisms responsible for these responses, we created protease homology models. The models mirrored the results of our in vitro experiments, indicating that protease homology models may offer insight into underlying functional mechanisms. Whether the observed bursal protease resistance to known inhibitors is important in the context of spermatophore digestion remains to be tested. However, our results suggest the exciting possibility that bursal protease specificity may have evolved in response to interactions with various proteins and inhibitors present within the female tract during the reproductive process.

Keywords: protease, spermatophore, bursa copulatrix, coevolution, reproduction, Lepidoptera.

Introduction

Reproductive characteristics are among the most rapidly evolving traits currently known. Researchers first noted this rapid evolution in reproductive morphologies (Eberhard 1985). However, burgeoning evidence indicates that this rapid evolution extends to the molecular level (Lee et al. 1995; Wyckoff et al. 2000; Torgerson et al. 2002; Clark et al. 2006; Sirot et al. 2014). For example, proteins involved in reproduction consistently exhibit high levels of adaptive and divergent evolution (Torgerson et al. 2002; Vicens et al. 2014). Many reproductive proteins have evolutionary rates that are equal to or greater than those observed for traits involved in other evolutionarily dynamic contexts, such as immunity-related proteins (Jansa et al. 2003).

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 (Swanson et al. 2001; Galindo et al. 2003; Findlay et al. 2014; Brennan and Prum 2015). The rapid and adaptive evolution seen in both sexes suggests that male and female traits may be coevolving with each other, with such coevolution potentially concentrated at specific male-female interfaces. This coevolution can be driven by cooperation between the sexes and/or by sexual conflict (Chapman et al. 2003; Dougherty et al. 2017). To better understand the nature of such coevolutionary interfaces, research must focus on characterizing the identities and functions of reproductive traits in both sexes.

In addition to playing a key role in the reproductive biology of a given species, sexual coevolution may play an important part in population divergence and speciation. Coevolution

*Corresponding author. Present address: Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045; email: mep115@pitt.edu.
†Present address: Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112.

Physiological and Biochemical Zoology 92(6):579–590. 2019. 1522-2152/2019/9206-8149$15.00. DOI: 10.1086/705722

between the sexes may lead to mismatches between populations at morphological, behavioral, or molecular levels, potentially leading to reproductive dysfunctions between populations on secondary contact. These mismatches could thus lead to reproductive barriers during allopatric speciation (Lande 1981; Knowles and Markow 2001; Orr 2005; Larson et al. 2012; Sirot et al. 2015; Garlovsky and Snook 2018; Turissini et al. 2018).

Such reproductive barriers can occur at different stages of the reproductive process and therefore involve different suites of behavioral, morphological, and/or molecular traits. To facilitate focused inquiry, researchers have classically split these reproductive interactions into two time frames: prezygotic and postzygotic. Work investigating prezygotic reproductive isolation has typically concentrated on premating isolation (e.g., differences in courtship behavior or mating preferences) or postmating prezygotic (PMPZ) interactions. Studies of the latter have focused primarily on the process of fertilization, including the proteins involved in interactions between the sperm and the egg. However, PMPZ interactions before fertilization can play a critical role in reproductive physiology and subsequent reproductive interactions (Ahmed-Braimah 2016; McDonough et al. 2016; Avila and Wolfner 2017; Villarreal et al. 2018). For example, in polyandrous species in which females may mate with multiple males, PMPZ interactions can potentially affect the outcome of current and future matings by the female (Wolfner 2009) via mechanisms such as sperm competition, male manipulation of females, and cryptic female choice (Perry et al. 2013; Rowe et al. 2015; Firman et al. 2017). This suggests a potentially important but understudied role for PMPZ traits in sexual selection and speciation.

Here, we describe a series of studies aimed at characterizing the 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 that includes both sperm and a large bolus of proteins, carbohydrates, lipids, and other substances to the female reproductive tract (Marshall 1985; Meslin et al. 2017). During copulation, the male forms this spermatophore inside the female within a specialized reproductive organ called the bursa copulatrix (hereafter, "bursa") in the female reproductive tract (Rogers and Wells 1984; Meslin et al. 2017). Shortly after mating, the sperm migrate out of the bursa into the spermatheca, the sperm storage organ (Rutowski and 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 because of 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 and 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 (Karlsson 1998; Arnqvist and Nilsson 2000). 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 the spermatophore digestion rate may provide males fitness benefits by delaying female remating and therefore extending the period of time where his sperm 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 rates, 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 and 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 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 that 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 predictions remained untested experimentally.

To address these gaps in our knowledge, we 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), we determined protein identities via mass spectrometry. We then investigated their modes of action using diagnostic protease inhibitors. We followed these functional assays with homology modeling to explore possible enzyme structures and their implications for protease function and inhibition. By combining our experimental findings with these modeling approaches, we were able to identify putative mechanisms of proteolytic activity that will serve as working hypotheses for future studies.

Material and Methods

Experimental Animals

Experimental animals were the F_1 female offspring of wild female *Pieris rapae* Linnaeus 1758 collected from an agricultural site in Rochester, Pennsylvania (40°44′44.4″N, 80°09′49.0″W), in the summers of 2016 and 2018. F_1 offspring were reared in climate-controlled chambers that maintained a 16L : 8D photoperiod with a constant temperature (24°C) and relative humidity (60%). Larvae were fed ad lib. on young *Brassica oleracea*

leaves. Upon eclosion, females were housed in individual containers within the chambers until they were used for experiments.

Extract Preparation

Proteases were collected from the bursae of 3-d-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 min. Supernatant was removed from the resulting pellet and subsequently stored at −20°C until assayed.

Zymogram Identification of Active Proteases

We used a modified zymogram technique to identify the active proteases within each bursal sample (Raser et al. 1995). We combined 10 μL of homogenized enzyme solution with 10 μL of native gel loading buffer (150 mM Tris-HCl, pH 6.8, 20% glycerol, 0.004% bromophenol blue) before incubating at 37°C for 15 min. Before samples were loaded, 8% resolving native PAGE gels with 5% stacking gel, both lacking sodium dodecyl sulfate, were prerun for 15 min at 125 V. Gels with samples were run at 125 V for 3 h at 4°C in native running buffer (25 mM Tris-base, 192 mM glycine, 1 mM DTT, 1 mM EDTA). After running, gels were rinsed in deionized (DI) water before being soaked in casein solution (500 μM casein, 3.6 mM $CaCl_2$, pH 7) for 60 min. 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) for 20 min, then destained (20% isopropanol, 7% acetic acid) for 40 h. 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, we characterized the proteins present in the gel area directly above the highest protease band (fig. A1). In total, 10 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, we randomly chose bursal extracts to run on the gels and collected bands until the target number for each band was reached. We used four bursal extracts in total, with each contributing at least two bands to the analysis (table S1; tables S1, S2 are available online). The extracted bands were then trypsinized and subjected to liquid chromatography followed by tandem mass spectrometry (Shevchenko et al. 2006; Granvogl et al. 2007*a*, 2007*b*), in conjunction with previously acquired transcriptomic sequences comprising 15,773 unique components (Meslin et al. 2015; Plakke et al. 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 et al. 2003). Tandem mass spectrometry data were visualized using Scaffold (Proteome Software, Portland, OR), with subsequent annotation of identified proteins using BLASTP (Altschul et al. 1990). We annotated the top hits identified by the National Center for Biotechnology Information BLASTP that had E value scores less than 1×10^{-20}. Additionally, we identified protein family (Pfam) domains using HMMER version 3.2.1 (http://hmmer.org) and converted domains with E values less than 0.01 to gene ontology terms using Pfam2go (Mitchell et al. 2015; table S2). The protein and domain identities obtained were compared with the proteases identified previously in this species (Plakke et al. 2015). Uncorrected spectra counts were used to calculate the proportion of each protein in each gel slice.

Protease Inhibitor Assay

To determine whether the identified proteases have biochemical activity that is consistent with their predicted motifs, we modified the above zymogram methods to include exposing focal proteases to class-specific protease inhibitors. To test for serine-like activity, we 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, we exposed the proteases to the papain inhibitor leupeptin (Aoyagi et al. 1969). Samples were prepared by mixing 10 μL of enzyme extract with 5 μL 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 active enzyme units [U], Fisher Scientific, Hampton, NH) was prepared in an identical fashion alongside bursal samples. Samples were incubated at 37°C for 15 min to allow for inhibition by protease inhibitors (when present) before running in conditions identical to the methods described above. After destaining, gels were imaged at 600 dpi (Canon 9000F Mark II scanner, Tokyo). We quantified the cleared areas from gel scans using ImageJ (Schneider et al. 2012). The extent of clearing caused by commercial trypsin was measured in active enzyme units based on the standard amount of known enzyme and activity loaded into the gel. The cleared areas of sample bands were converted to units of activity based on their measured intensities relative to the standard trypsin band. This method allowed us to observe the effects of commercial trypsin down to a concentration of 1.5 mg/mL in PBS equivalent to 1.94 U. Because of the irreversible nature of the chosen protease inhibitors, increases in inhibitor concentration did not increase inhibition, indicating saturation (data not shown).

Homology Modeling

Using the program I-TASSER (Yang et al. 2014), we 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 our modeled protease structures, we used PyMOL (open-source, ver. 2.1.0) to align a crystal structure of PMSF-bound *Fusarium oxysporum* trypsin (PDB ID: 1PQA; Schmidt et al. 2003) to the BTLP1 homology model. To study the leupeptin binding pose, we used PyMOL to align a crystal structure of leupeptin-bound *Carica papaya* papain (PDB ID: 1POP; Schröder et al. 1993) to our BPLP1 and BPLP2 homology models. Figures of these models were generated using BlendMol (Durrant 2019). We selected structures of *F. oxysporum* and *C. papaya* proteases as references because they are the classic structures associated with their respective classes and are particularly well characterized.

Statistical Analysis

All statistical analyses were conducted using the statistical program SPSS (ver. 25.0, IBM, Armonk, NY). We performed *t*-tests to compare treatments with their appropriate control for each bursal band following Levene's test for equality of variances.

Results

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 (fig. 1). All samples exhibited at least one of the bands ($n = 19$), though the number and combination of bands varied between individuals. Of the 19 individuals sampled for activity, six exhibited all three bands, 11 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, 52 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 (mean ± standard error) for bursa α, bursa β, and bursa γ, respectively.

The observed proteases correspond to proteases that our research team has previously identified and classified (Plakke et al. 2015). Predicted classes were further supported by Pfam domain identification (table S2). From the bands sampled, we observed one trypsin-like protease (BTLP1), two general peptidases (BGP1 and BGP2), and three papain-like proteases (BPLP1, BPLP2, and BPLP3). Bursa α comprises primarily BGP1 and BPLP1 (table 1). Bursa β comprises mainly BTLP1 and BPLP1 (table 1). Bursa γ comprises 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 (table S2). 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 our samples as a result of our nonspecific bursal extraction techniques (table S1).

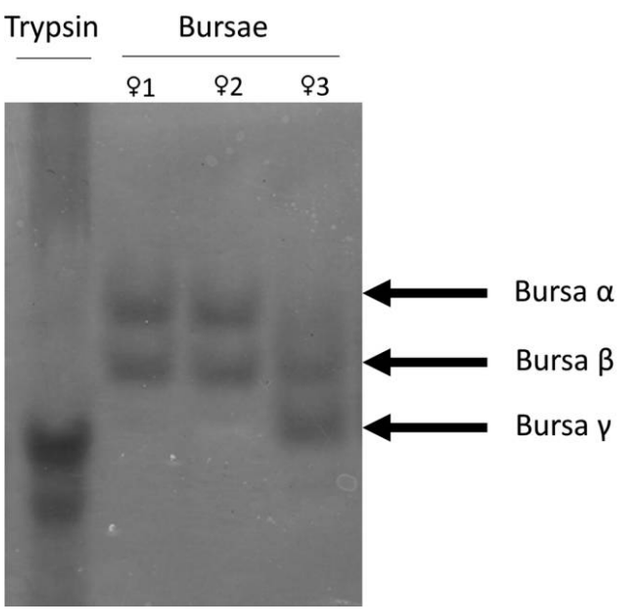

Figure 1. Bursal extracts and commercial trypsin (positive control) resolved using native polyacrylamide gel electrophoresis. Three consistent bands of active bursal protease digestion were identified, labeled as bursa α, bursa β, and bursa γ. Image inverted for visual clarity (i.e., dark bands in this figure would be clear bands in the zymogram itself).

Response of Bursal Proteases to Common Protease Inhibitors

Though we detected 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 $= 5$, $P = 0.718$; fig. 2). 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 $= 8$, $P = 0.532$; fig. 2). We detected far more BPLP1 than BTLP1 in bursa α, suggesting that the observed reductions in protease activity in bursa α are most likely the result of BPLP1 inhibition. In contrast, we also detected considerable amounts of BPLP2 in bursa γ, yet leupeptin did not affect that band (table 1). These results suggest that BPLP2—but not BPLP1—is resistant to leupeptin activity. Interestingly, though we detected BPLP1 in bursa β, we saw no leupeptin inhibition.

Modeling Bursal Protease/Inhibitor Binding

When compared with *Fusarium oxysporum* trypsin, the active site of BTLP1 showed distinct differences that are partially

Table 1: Proteases recovered through proteomic analysis of excised bands of activity on zymograms

Protease	Comp (Meslin et al. 2015)	Bursal sample: 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 ± .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 ± .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 ± 3.4	0	.39 ± .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	.97 ± .97	2.68 ± 1.50	0	General peptidase-like
BGP2	Comp97068_c0	2.29 ± 1.36	3.37 ± 1.12	14.35 ± 7.32	0	General peptidase-like

Note. The nine proteases were previously identified in Plakke et al. (2015). Percentages represent the proportion of all identified protease peptides by spectra counts within the respective excised band. Of the nine proteases previously identified, six were recovered.

localized to a key pocket-adjacent loop connecting two beta strands. In *F. oxysporum* trypsin, this loop spans 12 residues (W212-G223) and does not occlude the active site. In contrast, the homologous BTLP1 loop region contains 14 residues (fig. 3*A*, 3*B*). Our *Pieris rapae* BTLP1 homology model suggests that this longer loop protrudes into the binding pocket, limiting access to the catalytic triad (fig. 3*A*, 3*B*). The crystallographic PMSF pose does in fact clash with the modeled BTLP1 loop when the proteins are superimposed (fig. 3*A*, 3*B*). Thus, consistent with our experimental results, our computational model predicts that PMSF should not inhibit the proteolytic activity of BTLP1.

We next considered the *P. rapae* proteases BPLP1 and BPLP2. Structural differences between these two proteases (fig. 3*C*, 3*D*) may explain why leupeptin inhibits BPLP1 but not BPLP2. To visualize leupeptin in the context of the BPLP1 homology model, we aligned our 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 (fig. 3*C*, 3*D*). In contrast, the homology model of BPLP2 is not compatible with leupeptin binding (fig. 3*C*, 3*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 and Kierstan 1973; Roy et al. 2012) before activation. It is curious that the predominant papain-like protease in bursa γ is zymogen-like. We note that BPLP2 is not merely pro-BPLP1, as there are other amino acid differences between these two bursal papain-like proteases. Deleting preprotein domain from the model eliminates the occlusion of the BPLP2 active site, suggesting that BPLP2 may respond to leupeptin upon activation (fig. A2). The BPLP1 and BPLP2 models thus predict modes of action consistent with our zymogram experiments, potentially explaining why only BPLP1 is susceptible to leupeptin inhibition.

Discussion

This study characterizes the proteases active within the bursa of virgin *Pieris rapae*. Building on previous work (Plakke et al. 2015), we characterized the activity and modes of action of the proteases that exhibited in vitro activity in our bursal extracts. Through zymogram analysis, we 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 our native PAGE experiments. These proteases are predicted to belong to families of serine-class, cysteine-class, and general peptidase-class proteases (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 [Lawniczak and Begun 2007; Kelleher and Pennington 2009; Alfonso-Parra et al. 2016], Lepidoptera [Al-Wathiqui et al. 2014; 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. We 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 1). However, activity of this protease was largely unaffected by PMSF, a serine-specific protease inhibitor. Our homology modeling suggests that an extended BTLP1 loop blocks inhibitor access to the active site, potentially contributing to PMSF resistance (fig. 3*A*, 3*B*). It is possible that this structural difference renders the protease incapable of digesting casein and therefore not responsible for the activity reported in our assay. However, this seems highly unlikely. Casein lacks tertiary structure and is generally digestible by all classes of

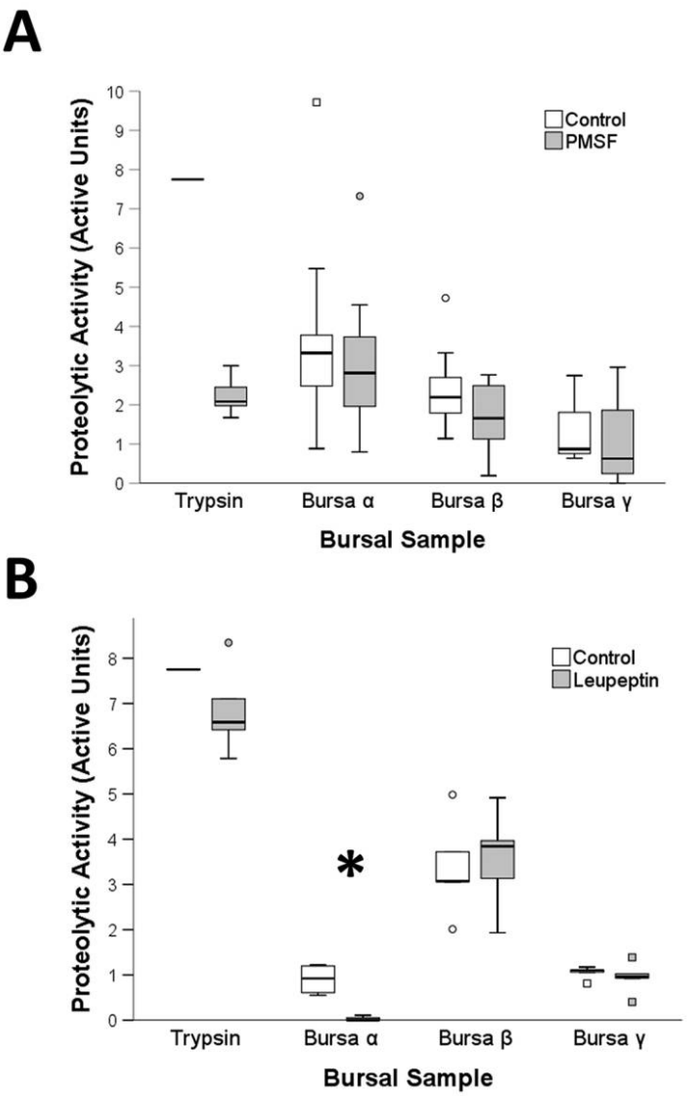

Figure 2. Cleared bands in the zymograms were quantified in the presence and absence of the protease inhibitors leupeptin and phenylmethane sulfonyl fluoride (PMSF). *A*, The trypsin-specific protease inhibitor (PMSF) inhibits the activity of commercial trypsin but does not inhibit the activity of bursal 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.

proteases. Furthermore, high concentrations of BTLP1 were recovered within the bursa β band of activity. We propose instead that BTLP1 is resistant to some modes of serine-specific protease inhibition and hypothesize that this resistance may be biologically relevant. While we did not recover any protease inhibitors from the bursal bands, such inhibitors are known to be present within female reproductive tracts (Prokupek et al. 2010; Al-Wathiqui et al. 2014; Plakke et al. 2015; Dong et al. 2016). Further, males are known to transfer protease inhibitors, together with the ejaculate and sperm, to the female (LaFlamme and Wolfner 2013), and protease inhibitors have been documented across lepidopteran ejaculates (Dong et al. 2016; Al-Wathiqui et al. 2017). 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 1). 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. We hypothesize that posttranslational modifications may confer BPLP1 resistance to leupeptin in bursa β. These modifications can alter protein

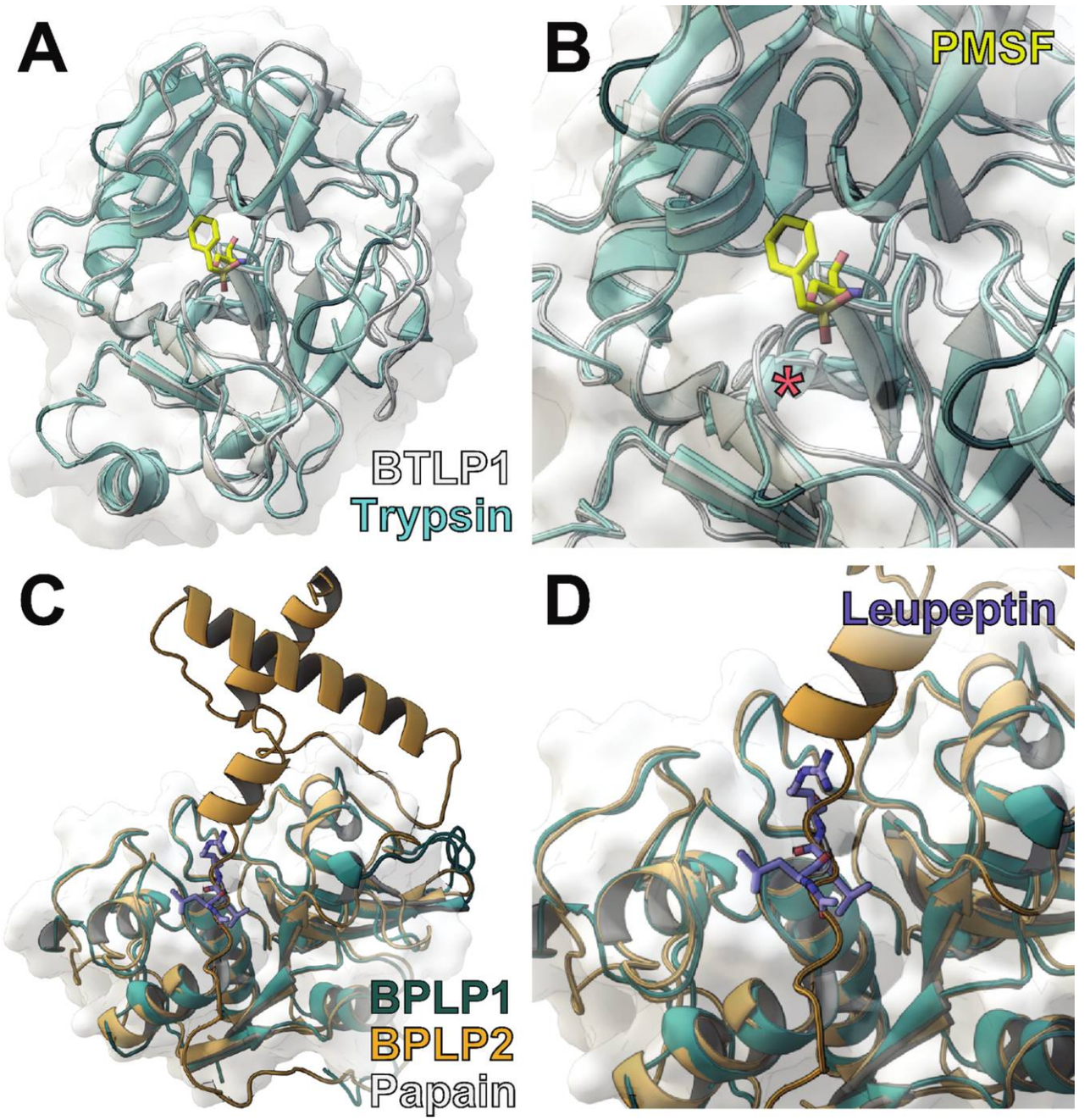

Figure 3. *A*, 1PQA crystal structure of *Fusarium oxysporum* trypsin bound to phenylmethane sulfonyl fluoride (PMSF; cyan ribbon and yellow sticks, respectively) superimposed on the bursal papain-like protease (BTLP1) homology model (white ribbon). *B*, Zoomed-in view of the active site of BTLP1. The 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 shown in dark teal and gold ribbon, respectively. To position the leupeptin inhibitor (purple sticks), we aligned the 1POP holo structure of *Carica 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 that occupies the cleft and is incompatible with leupeptin binding. *D*, Zoomed-in view of the active sites for BPLP1 and BPLP2 with leupeptin superimposed.

charge and size and so may also explain why the protease is found at multiple locations along the native PAGE gel. We also considered the possibility that bursa β BPLP1 is in the inactive zymogen form (Brocklehurst and Kierstan 1973), but mass spectrometry recovered no peptide spectra consistent with a papain-like zymogen inhibitory domain (fig. A3). Indeed, the BPLP1 spectra across bands were nearly identical. We also considered the possibility that leupeptin migrates with the bursa β group, actively inhibiting BPLP1 in bursa β but not in bursa α. However, the mass spectrometry methodology we used is unable to detect leupeptin, as it is not composed of amino acids and is thus invisible to such tests (table S1, available online). 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 (fig. 3*C*, 3*D*). Additional studies are required to determine whether BPLP2 is converted to an active from in *P. rapae*, as in other species (Yamamoto et al. 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 (M. S. Plakke, N. L. Clark, and N. I. Morehouse, unpublished data). Given that these spermatophore proteins are rapidly evolving (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 (Suzuki 1979; Bissoondath and Wiklund 1997). Thus, male traits that reduce the female remating rate are favored by selection on male fitness, including traits that reduce the spermatophore digestion rate (Sánchez and Cordero 2014). 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 rates 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 nonantagonistic forces, such as changes to the nutritional ecology of a particular species or population. If either dynamic were the case, we 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 coincubations 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, of either male or female origin, with mechanisms of action similar to those of the commercial inhibitors we 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 (Meslin et al. 2015), where the proteases may have evolved in response to inhibitors presented through the diet (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 are 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 evolve, by either selection or chance alone, 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, we have successfully identified and quantified the female half of a male/female reproductive interface that is 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, our 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.

Acknowledgments

We would like to thank Kretschmann Farms for allowing us 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. We would like to thank Sigma Xi, the Scientific Research Society, for funding this project. We also thank the Howard Hughes Medical Institute for funding to B.J.G. and J.B.L. This project used the Biomedical Mass Spectrometry Center and University of Pittsburgh Cancer Institute Cancer Biomarker Facility, which are supported, in part, by National Cancer Institute grant P30CA047904.

APPENDIX

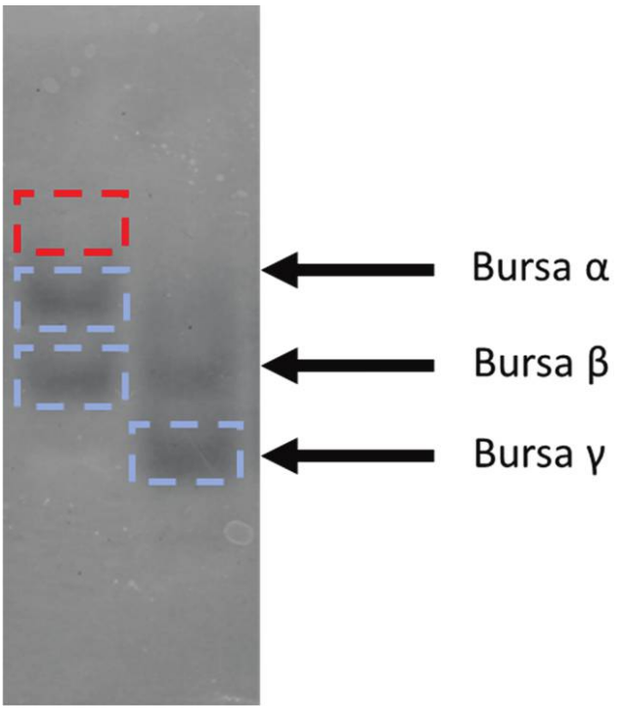

Figure A1. Bands extracted for mass spectrometry analysis were excised from the indicated regions on a gel (modified from fig. 1). The red rectangle indicates where the control band was excised, and blue rectangles represent excision areas for bursal sample bands.

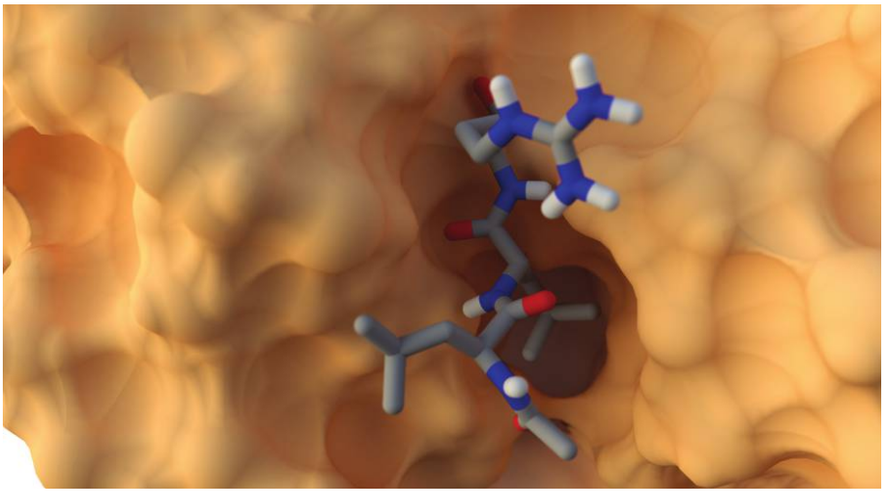

Figure A2. Bursal papain-like protease (BPLP2) homology model in the active form. The extended loop that otherwise occupies the catalytic cleft was removed. The crystallographic pose of the leupeptin inhibitor was taken from the 1POP structure.

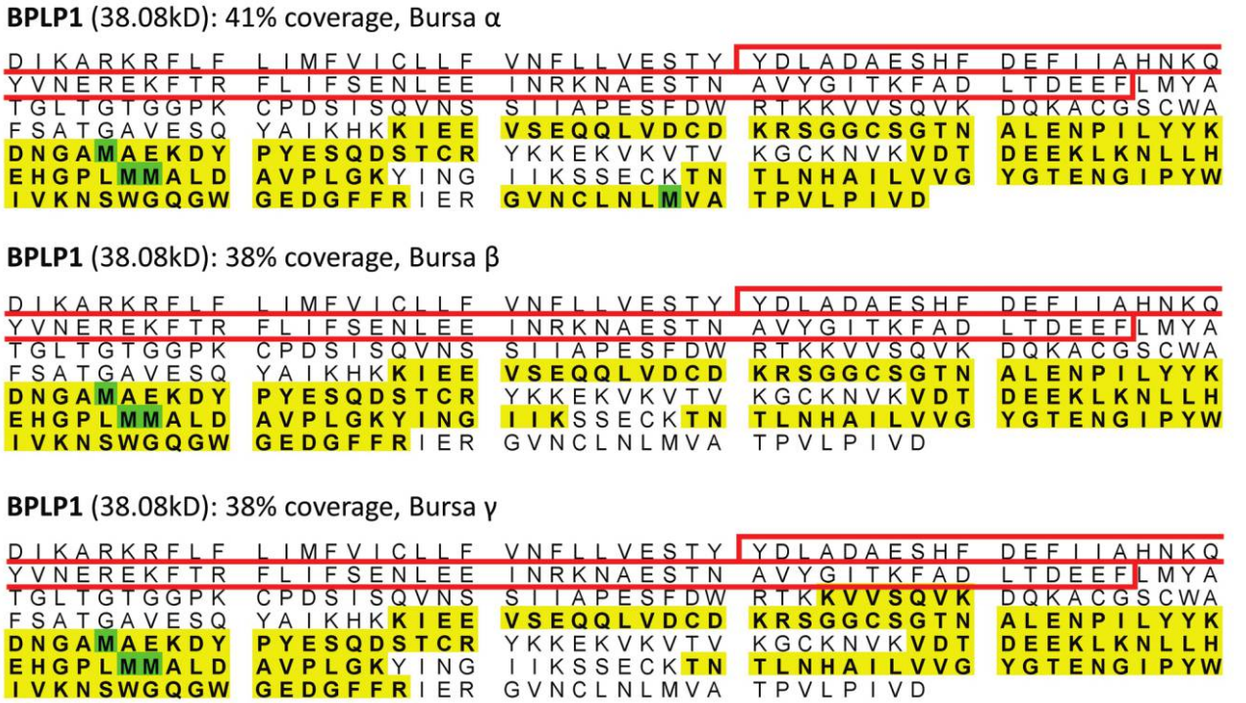

Figure A3. Bursal papain-like protease (BPLP1) was identified in all three bursa bands through mass spectrometry. The peptides recovered (yellow highlighting) corresponded to 38%–41% of all amino acids in the sequence. No spectra corresponding to the proprotein domain (red boxes) were recovered from any sample.

Literature Cited

Ah-King M., A. Barron, and M. Herberstein. 2014. Genital evolution: why are females still understudied? PLoS Biol 12: e1001851.

Ahmed-Braimah Y.H. 2016. Multiple genes cause postmating prezygotic reproductive isolation in the *Drosophila virilis* group. Genes Genomes Genet 6:4067–4076.

Alfonso-Parra C., Y.H. Ahmed-Braimah, E.C. Degner, F.W. Avila, S.M. Villarreal, J.A. Pleiss, M.F. Wolfner, and L.C. Harrington. 2016. Mating-induced transcriptome changes in the reproductive tract of female *Aedes aegypti*. PLoS Negl Trop Dis 10:e0004451.

Altschul S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

Al-Wathiqui N., S.M. Lewis, and E.B. Dopman. 2014. Using RNA sequencing to characterize female reproductive genes between Z and E strains of European corn borer moth (*Ostrinia nubilalis*). BMC Genom 15:189.

———. 2017. Molecular dissection of nuptial gifts in divergent strains of *Ostrinia* moths. Physiol Entomol 43:10–19.

Aoyagi T., S. Miyata, M. Nanbo, F. Kojima, M. Matsuzaki, M. Ishizuka, T. Takeuchi, and H. Umezawa. 1969. Biological activities of leupeptins. J Antibiot 22:558–568.

Arnqvist G. and T. Nilsson. 2000. The evolution of polyandry: multiple mating and female fitness in insects. Anim Behav 60:145–164.

Avila F.W. and M.F. Wolfner. 2017. Cleavage of the *Drosophila* seminal protein Acp36DE in mated females enhances its sperm storage activity. J Insect Physiol 101:66–72.

Bissoondath C. and C. Wiklund. 1997. Effect of male body size on sperm precedence in the polyandrous butterfly *Pieris napi* L. (Lepidoptera: Pieridae). Behav Ecol 8:518–523.

Boggs C. and L. Gilbert. 1979. Male contribution to egg production in butterflies: evidence for transfer of nutrients at mating. Science 206:83–84.

Brennan P.L.R. and R.O. Prum. 2015. Mechanisms and evidence of genital coevolution: the role of natural selection, mate choice, and sexual conflict. Cold Spring Harb Perspect Biol 7:1–21.

Broadway R. 1996. Dietary proteinase inhibitors alter complement of midgut proteases. Arch Insect Biochem Physiol 53:39–53.

Brocklehurst K. and M.P.J. Kierstan. 1973. Propapain and its conversion to papain: a new type of zymogen activation mechanism involving intramolecular thiol-disulphide interchange. Nat New Biol 242:167–170.

Chapman T., G. Arnqvist, J. Bangham, and L. Rowe. 2003. Sexual conflict. Trends Ecol Evol 18:41–47.

Clark N.L., J.E. Aagaard, and W.J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131:11–22.

Dong Z., X. Wang, Y. Zhang, L. Zhang, Q. Chen, X. Zhang, P. Zhao, and Q. Xia. 2016. Proteome profiling reveals tissue-specific protein expression in male and female accessory glands of the silkworm, *Bombyx mori*. Amino Acids 48:1173–1183.

Dougherty L.R., E. van Lieshout, K.B. McNamara, J.A. Moschilla, G. Arnqvist, and L.W. Simmons. 2017. Sexual conflict and correlated evolution between male persistence and female resistance traits in the seed beetle *Callosobruchus maculatus*. Proc R Soc B 284:20170132.

Durrant J.D. 2019. BlendMol: advanced macromolecular visualization in Blender. Bioinformatics 35:2323–2325.

Eberhard W.G. 1985. Sexual selection and animal genitalia. Harvard University Press, Cambridge.

Findlay G.D., J.L. Sitnik, W. Wang, C.F. Aquadro, N.L. Clark, and M.F. Wolfner. 2014. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. PLoS Genet 10:e1004108.

Firman R.C., C. Gasparini, M.K. Manier, and T. Pizzari. 2017. Postmating female control: 20 years of cryptic female choice. Trends Ecol Evol 32:368–382.

Galindo B.E., V.D. Vacquier, and W.J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci USA 100:4639–4643.

Garlovsky M.D. and R.R. Snook. 2018. Persistent postmating, prezygotic reproductive isolation between populations. Ecol Evol 17:9062–9073.

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: 897–901.

Granvogl B., P. Gruber, and L.A. Eichacker. 2007*a*. Standardisation of rapid in-gel digestion by mass spectrometry. Proteomics 7:642–654.

Granvogl B., M. Plöscher, and L.A. Eichacker. 2007*b*. Sample preparation by in-gel digestion for mass spectrometry-based proteomics. Anal Bioanal Chem 389:991–1002.

Jansa S.A., B.L. Lundrigan, and P.K. Tucker. 2003. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. J Mol Evol 56:294–307.

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

Kelleher E.S. and J.E. Pennington. 2009. Protease gene duplication and proteolytic activity in *Drosophila* female reproductive tracts. Mol Biol Evol 26:2125–2134.

Knowles L.L. and T.A. Markow. 2001. Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. Proc Natl Acad Sci USA 98:8692–8696.

LaFlamme B.A. and M.F. Wolfner. 2013. Identification and function of proteolysis regulators in seminal fluid. Mol Reprod Dev 80:80–101.

Lande R. 1981. Models of speciation by sexual selection on polygenic traits. Proc Natl Acad Sci USA 78:3721–3725.

Larson E.L., G.L. Hume, J.A. Andrés, and R.G. Harrison. 2012. Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. J Evol Biol 25:174–186.

Lawniczak M.K.N. and D.J. Begun. 2007. Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. Mol Biol Evol 24:1944–1951.

Lee Y.H., T. Ota, and V.D. Vacquier. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol 12:231–238.

Marshall L. 1985. Protein and lipid composition of *Colias philodice* and *C. eurytheme* spermatophores and their changes over time (Pieridae). J Res Lepid 24:21–30.

McDonough C.E., E. Whittington, S. Pitnick, and S. Dorus. 2016. Proteomics of reproductive systems: towards a molecular understanding of postmating, prezygotic reproductive barriers. J Proteom 135:26–37.

Meslin C., T.S. Cherwin, M.S. Plakke, B.S. Small, B.J. Goetz, N. Morehouse, and N.L. Clark. 2017. Structural complexity and molecular heterogeneity of a butterfly ejaculate reflect a complex history of selection. Proc Natl Acad Sci USA 114:1–8.

Meslin C., M.S. Plakke, A.B. Deutsch, B.S. Small, N.I. Morehouse, and N.L. Clark. 2015. Digestive organ in the female reproductive tract borrows genes from multiple organ systems to adopt critical functions. Mol Biol Evol 32:1567–1580.

Mitchell A., H.Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, et al. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 43:D213–D221.

Nesvizhskii A.I., A. Keller, E. Kolker, and R. Aebersold. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75:4646–4658.

Orr H. 2005. The genetic basis of reproductive isolation: insights from *Drosophila*. Proc Natl Acad Sci USA 102:6522–6526.

Ou C.M., J.B. Tang, M.S. Huang, P.S. Sudhakar Gandhi, S. Geetha, S.H. Li, and Y.H. Chen. 2012. The mode of reproductive-derived Spink (serine protease inhibitor Kazal-type) action in the modulation of mammalian sperm activity. Int J Androl 35:52–62.

Patankar A.G., A.P. Giri, A.M. Harsulkar, M.N. Sainani, V.V. Deshpande, P.K. Ranjekar, and V.S. Gupta. 2001. Complexity in specificities and expression of *Helicoverpa armigera* gut proteinases explains polyphagous nature of the insect pest. Insect Biochem Mol Biol 31:453–464.

Perry J.C., L. Sirot, and S. Wigby. 2013. The seminal symphony: how to compose an ejaculate. Trends Ecol Evol 28:414–422.

Plakke M.S., A.B. Deutsch, C. Meslin, N.L. Clark, and N.I. Morehouse. 2015. Dynamic digestive physiology of a female reproductive organ in a polyandrous butterfly. J Exp Biol 218: 1548–1555.

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

Raser K.J., A. Posner, and K.K. Wang. 1995. Casein zymography: a method to study u-calpain, m-calpain, and their inhibitory agents. Arch Biochem Biophys 319:211–216.

Rogers S.H. and H. Wells. 1984. The structure and function of the bursa copulatrix of the monarch butterfly (*Danaus plexippus*). J Morphol 221:213–221.

Rowe M., T. Albrecht, E.R.A. Cramer, A. Johnsen, T. Laskemoen, J.T. Weir, and J.T. Lifjeld. 2015. Postcopulatory sexual selection is associated with accelerated evolution of sperm morphology. Evolution 69:1044–1052.

Roy S., D. Choudhury, P. Aich, J.K. Dattagupta, and S. Biswas. 2012. The structure of a thermostable mutant of pro-papain reveals its activation mechanism. Acta Crystallogr D 68:1591–1603.

Rutowski R. and G. Gilchrist. 1986. Copulation in *Colias eurytheme* (Lepidoptera: Pieridae): patterns and frequency. J Zool 209:115–124.

Sánchez V. and C. Cordero. 2014. Sexual coevolution of spermatophore envelopes and female genital traits in butterflies: evidence of male coercion? PeerJ 2:e247.

Schmidt A., C. Jelsch, P. Østergaard, W. Rypniewski, and V.S. Lamzin. 2003. Trypsin revisited: crystallography at (SUB) atomic resolution and quantum chemistry revealing details of catalysis. J Biol Chem 278:43357–43362.

Schneider C.A., W.S. Rasband, and K.W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675.

Schröder E., E. Garman, K. Harlos, and C. Crawford. 1993. X-ray crystallographic structure of a papain-leupeptin complex. FEBS Lett 315:38–42.

Shevchenko A., H. Tomas, J. Havlis, J.V. Olsen, and M. Mann. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 1:2856–2860.

Sirot L.K., G.D. Findlay, J.L. Sitnik, D. Frasheri, F.W. Avila, and M.F. Wolfner. 2014. Molecular characterization and evolution of a gene family encoding both female- and male-specific reproductive proteins in *Drosophila*. Mol Biol Evol 31:1554–1567.

Sirot L.K., A. Wong, T. Chapman, and M.F. Wolfner. 2015. Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. Cold Spring Harb Perspect Biol 7:a017533.

Sugawara T. 1979. Stretch reception in the bursa copulatrix of the butterfly, *Pieris rapae crucivora*, and its role in behaviour. J Comp Physiol 130:191–199.

———. 1981. Fine structure of the stretch receptor in the bursa copulatrix of the butterfly, *Pieris rapae crucivora*. Cell Tissue Res 217:23–36.

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

Swanson W.J., Z. Yang, M.F. Wolfner, and C.F. Aquadro. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98:2509–2514.

Torgerson D.G., R.J. Kulathinal, and R.S. Singh. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol 19:1973–1980.

Turissini D.A., J.A. McGirr, S.S. Patel, J.R. David, and D.R. Matute. 2018. The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. Mol Biol Evol 35:312–334.

Vicens A., L. Lüke, and E.R.S. Roldan. 2014. Proteins involved in motility and sperm-egg interaction evolve more rapidly in mouse spermatozoa. PLoS ONE 9:e91302.

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

Wolfner M. 2009. Battle and ballet: molecular interactions between the sexes in *Drosophila*. J Hered 100:399–410.

Wyckoff G.J., W. Wang, and C.I. Wu. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309.

Yamamoto Y., M. Kurata, S. Watabe, R. Murakami, and S.Y. Takahashi. 2002. Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. Curr Protein Pept Sci 3:231–238.

Yang J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. 2014. The I-TASSER suite: protein structure and function prediction. Nat Methods 12:7–8.